

Universidade Federal de Juiz de Fora
Faculdade de Engenharia
Programa de Pós-Graduação em Engenharia Elétrica

Bernardo Sotto-Maior Peralva

Reconstrução de Energia para Calorímetros Finamente Segmentados

Tese de Doutorado

Juiz de Fora
2015

Bernardo Sotto-Maior Peralva

Reconstrução de Energia para Calorímetros Finamente Segmentados

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, na área de concentração em Sistemas Eletrônicos, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Orientadores: Prof. Augusto Santiago Cerqueira, D.Sc.
Prof. José Manoel de Seixas, D.Sc.

Juiz de Fora
2015

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Sotto-Maior Peralva, Bernardo.

Reconstrução de Energia para Calorímetros Finamente Segmentados / Bernardo Sotto-Maior Peralva. -- 2015. 141 f.

Orientador: Augusto Santiago Cerqueira

Coorientador: José Manoel de Seixas

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2015.

1. Estimaco de Energia. 2. Mxima Verossimilhana. 3. Filtragem Ótima. 4. Fsica de Altas Energias. 5. Empilhamento de Sinais. I. Santiago Cerqueira, Augusto, orient. II. de Seixas, Jos Manoel, coorient. III. Ttulo.

RECONSTRUÇÃO DE ENERGIA PARA CALORÍMETROS FINAMENTE
SEGMENTADOS

Bernardo Sotto-Maior Peralva

TESE SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA (PPEE) DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Augusto Santiago Cerqueira, D.Sc.

Prof. José Manoel de Seixas, D.Sc.

Prof. Carlos Augusto Duque, D.Sc.

Prof. Rafael Antunes Nóbrega, Ph.D.

Prof. Márcia Begalli, Ph.D.

Prof. Sérgio Lima Netto, Ph.D.

JUIZ DE FORA, MG – BRASIL
SETEMBRO DE 2015

Agradecimentos

Primeiramente a Deus pela oportunidade e por ter proporcionado satisfação para a realização deste trabalho.

Ao meu orientador prof. Augusto, pela formação de alto nível e com quem tenho o privilégio de trabalhar desde a graduação, quando ingressei na pesquisa há 10 anos. Apesar dos inúmeros compromissos pessoais e profissionais, em todo instante se mostrou disposto e disponível para discussões. Meu eterno obrigado pela paciência, confiança, ensinamentos e amizade.

Ao prof. Seixas pela co-orientação, confiança e comprometimento com a educação. Muito obrigado pelo aprendizado e incentivo à pesquisa.

Em especial, ao prof. Luciano Manhães que foi o grande precursor deste trabalho e que sempre batalhou junto, com muita energia e sugestões brilhantes, para que este trabalho se consolidasse. Agora colhe os primeiros frutos deste esforço.

À minha querida esposa Elisa, que acompanhou toda essa trajetória com muito amor, paciência e aprendizado mútuo constante. Dedico esta tese a você, meu amor!

Aos meus pais e irmãos pelo incentivo e apoio nas minhas escolhas, além da educação e ensinamentos para que eu chegasse até este momento. Sei que o caminho também não foi fácil para vocês. Obrigado!

Aos colegas e amigos do Laptel (UFJF) e LPS (UFRJ), pelos momentos de descontração, companheirismo e troca de informações. Também agradeço àqueles com quem tive a oportunidade de conviver no CERN durante o período sanduíche e dividir experiências: Dênis, Davis, Andressa, Júlio, Marcos, Marisílvia e Yara.

À colaboração ATLAS no CERN, em especial ao TileCal e minha co-orientadora no exterior, Ana Henriques, por acreditar e apoiar este trabalho. Agradeço também aos colegas do grupo de validação de sinal, Sasha Solodkov, Giulio Usai e Alberto Valero pelos conhecimentos de calorimetria e da infraestrutura de software.

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro dado a este trabalho.

Ao Povo Brasileiro que indiretamente ou diretamente financiou este trabalho. Espero poder retribuir na mesma moeda.

Resumo da Tese apresentada à PPEE como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

RECONSTRUÇÃO DE ENERGIA PARA CALORÍMETROS FINAMENTE SEGMENTADOS

Bernardo Sotto-Maior Peralva

Setembro/2015

Orientadores: Augusto Santiago Cerqueira
José Manoel de Seixas

Programa: Engenharia Elétrica

Esta tese apresenta técnicas de processamento de dados para a detecção de sinais e estimação de energia usando calorimetria de altas energias. Os calorímetros modernos possuem milhares de canais de leitura e operam sob alta taxa de eventos. Tipicamente, a reconstrução da energia envolve etapas de detecção e estimação, e é baseada na medida da amplitude do sinal (digitalizado) recebido. Os métodos empregados, atualmente, em experimentos de altas energias são baseados em técnicas de minimização da variância e selecionam os sinais de interesse a partir da estimação da energia. Este trabalho explora o uso de filtros casados para a detecção de sinais e faz uso de uma calibração para a estimação da energia dos sinais. Na abordagem proposta, os parâmetros aleatórios do pulso processado (fase e deformação) e a estatística do ruído de fundo são considerados no projeto do filtro digital, aumentando seu desempenho. No caso particular de experimentos onde a probabilidade de empilhamento de sinais é alta, uma outra solução, baseada na desconvolução linear de sinais para estimação de energia, é discutida. As técnicas propostas nesta tese foram implementadas *offline* e aplicadas no calorímetro de telhas (TileCal) do ATLAS no LHC. Foram utilizados sinais simulados, assim como dados reais adquiridos durante a operação nominal do LHC. Os estimadores propostos apresentaram menor erro quando comparados aos métodos empregados em calorímetros modernos e estão, atualmente, sendo validados para serem utilizados no TileCal.

Palavras-chave: Detecção e Estimação de Sinais, Filtro Casado, Empilhamento de Sinais, Física de Altas Energias.

Abstract of Thesis presented to PPEE/UFJF as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

ENERGY RECONSTRUCTION FOR FINELY SEGMENTED CALORIMETERS

Bernardo Sotto-Maior Peralva

September/2015

Advisors: Augusto Santiago Cerqueira

José Manoel de Seixas

Department: Electrical Engineering

This thesis presents data processing techniques of signal detection and energy estimation for high energy calorimetry. Modern calorimeters have thousands of readout channels and operate at high event rate conditions. Typically, the energy reconstruction involves both detection and estimation tasks, and it is based on the amplitude estimation of the received digitized signal. The current methods employed by high energy experiments are based on variance minimization techniques, and the valid signals are selected based on the energy estimation. This work explores the use of a technique based on Matched Filter for signal detection, and it makes use of a calibration factor to estimate the energy. In the proposed approach, the stochastic parameters of the pulse (phase and deformation) and the statistics from the background are considered for the filter design in order to increase performance. In particular cases, where the signal pile-up is likely to occur, another promising technique, based on linear signal deconvolution is discussed. The techniques proposed in this thesis were implemented offline and applied on the ATLAS Tile Calorimeter (TileCal) at LHC. Both simulated signals and real data acquired during nominal LHC operation were used. The proposed estimators presented smaller error with respect to the methods currently used in modern calorimeter systems, and they have been extensively tested to be used in TileCal.

Keywords: Signal Detection and Estimation, Matched Filter, Signal Pile-up, High Energy Physics.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiv
1 Introdução	1
1.1 Motivação e Objetivo	2
1.2 Principais Contribuições da Tese	3
1.3 O Ambiente Experimental de Desenvolvimento	4
1.4 Organização do texto	4
2 O Ambiente de Desenvolvimento: LHC, ATLAS e TileCal	6
2.1 O acelerador LHC no CERN	6
2.1.1 Os principais detectores do LHC	8
2.2 O detector ATLAS	9
2.2.1 Sistema de filtragem <i>online</i> de eventos do ATLAS	9
2.2.2 Calorimetria do ATLAS	10
2.2.3 O Calorímetro de Telhas do ATLAS	12
2.3 Estabilidade do sinal do TileCal	17
2.4 Reconstrução da Energia do TileCal	18
2.5 O programa de atualização do TileCal para o aumento da luminosidade do LHC	21
3 Reconstrução de Energia em Calorímetros para Física de Altas Energias	23
3.1 Estimação de Parâmetros	23
3.2 Tendência e Variância de um Estimador	25
3.3 BLUE - <i>Best Linear Unbiased Estimator</i>	25
3.4 Estimador de Máxima Verossimilhança (MLE)	27
3.5 Estimadores de energia em calorimetria	29
3.5.1 Estimação de energia no TileCal	31
3.5.2 Estimação de energia no calorímetro eletromagnético do ATLAS	33
3.5.3 Estimação de energia no calorímetro eletromagnético do CMS	34

4	Reconstrução de Energia para Calorimetria Baseada em Filtragem Casada	36
4.1	Filtragem Casada para Reconstrução de Energia em Calorimetria . . .	36
4.1.1	Detecção do sinal	37
4.1.2	Estimação da Amplitude	40
4.2	Implementação	42
5	Resultados	44
5.1	Simulação do Sinal de um Calorímetro	44
5.1.1	Conjunto de Dados Básico	45
5.1.2	Projeto do detector de sinais por filtragem casada	46
5.1.3	Resultados de detecção de sinais	48
5.1.4	Resultados de estimação da amplitude	50
5.2	Resultados de simulação para o TileCal	53
5.2.1	Descrição do ruído de fundo	55
5.2.2	Projeto dos filtros e variância	55
5.2.3	Detecção de sinal	57
5.2.4	Estimação da amplitude	59
5.3	Resultados com dados reais do TileCal	60
5.3.1	Dados de baixa luminosidade	62
5.3.2	Dados de alta luminosidade	62
5.4	Proposta para o Cenário de Empilhamento de Sinais no TileCal Durante a Segunda Tomada de Dados do LHC	66
5.4.1	Reconstrução <i>online</i> de energia no TileCal para o Run2 do LHC	66
5.4.2	Projeto dos estimadores	67
5.4.3	Resultados de simulação de eventos no TileCal previsto para o Run2	73
5.4.4	Generalização dos coeficientes do MF	76
6	Reconstrução de Energia para Ambiente com Alto Nível de Empilhamento de Sinais	82
6.1	Recuperação de fontes através de desconvolução	84
6.2	A Matriz de Desconvolução (DM)	85
6.2.1	Projeto do estimador para apenas um sinal central	86
6.2.2	Projeto do estimador para N sinais	87
6.3	Resultados de Simulação	88
6.4	Método DM combinado	91
6.4.1	COF, <i>Constrained Optimal Filter</i>	91
6.4.2	Dependência da simetria do pulso e de condições diferentes de SPR	93

6.5	COF aplicado a dados experimentais do TileCal	95
6.5.1	Análise do erro de estimação	96
6.5.2	Linearidade do COF	97
6.5.3	Desempenho de reconstrução do sinal	100
7	Conclusões	103
7.1	Perspectivas para o cenário de empilhamento de sinais	105
7.2	Trabalhos futuros	105
7.2.1	Impacto na física	106
7.2.2	Novos métodos para cenário de alto nível de empilhamento de eventos	106
	Referências Bibliográficas	108
A	Sistema de Coordenadas do ATLAS	115
B	Publicações	117
B.1	Trabalhos publicados em periódicos	117
B.2	Trabalhos publicados em anais de congressos	119

Lista de Figuras

2.1	Vista do globo de ciência e inovação no CERN.	7
2.2	O complexo do acelerador.	8
2.3	O anel do LHC e seus quatro detectores principais.	9
2.4	O detector ATLAS e seus sub-sistemas.	10
2.5	Sistema de trigger do ATLAS e seus níveis.	11
2.6	(a) Esquemático de uma fatia transversal do detector ATLAS mostrando os padrões de deposição de energia nos calorímetros do ATLAS, e (b) sistema de calorimetria do ATLAS.	13
2.7	Segmentação do calorímetro eletromagnético do ATLAS.	14
2.8	Diagrama esquemático mostrando os elementos que compõem a leitura do sinal físico pelo TileCal, correspondente a um módulo.	14
2.9	Vista esquemática da segmentação de células de um módulo do barril central ($\eta > 0$) e um módulo do barril estendido.	15
2.10	Pulso de referência do TileCal.	16
2.11	Cadeia de processamento do sinal do TileCal.	17
2.12	(a) Sinal reconstruído utilizando dados reais e o pulso de referência armazenado no banco de dados e (b) Distribuição de fase dos sinais adquiridos durante calibração com feixes reais em 2015.	19
2.13	Estabilidade do pedestal medido pela diferença entre o valor médio da primeira amostra digital e seu valor armazenado em banco de dado, para a partição LBA.	20
2.14	Ilustração do efeito de empilhamento de sinais no TileCal. O sinal de interesse está presente na quarta amostra enquanto um outro sinal, de outra colisão (50 ns atrasado) é observado distorcendo o pulso final.	22
3.1	Resposta típica de um calorímetro de altas energias, em que as amostras digitais estão destacadas para (a) o sinal puro (sem a presença de ruído) e (b) sob condições de baixa SNR (com a presença de ruído).	30
3.2	Resposta do calorímetro eletromagnético do CMS. As amostras digitais estão destacadas sobre a curva.	34

3.3	Resposta do calorímetro eletromagnético do CMS. As amostras digitais estão destacadas sobre a curva.	35
4.1	Operação de filtragem casada entre o sinal recebido e os coeficientes do Filtro Casado.	40
4.2	Diagrama de blocos do filtro casado estocástico.	41
4.3	Reconstrução do sinal recebido utilizando o Filtro Casado. Os pontos em vermelho na forma de losangos representam as amostras do sinal recebido e a curva, o fitting do pulso de referência nas amostras baseado na amplitude estimada.	42
4.4	Implementação da abordagem determinística do filtro casado para o caso de $N=5$	42
5.1	Pulso unipolar utilizado nas análises.	45
5.2	Distribuição das amostras do ruído de fundo.	46
5.3	Eficiência da matriz de branqueamento no sub-conjunto de teste de ruído.	47
5.4	Curva de carga acumulada do processo aleatório de sinal.	48
5.5	Desempenho de detecção em condições de baixa SNR: (a) Curvas ROC para o método SMF considerando números diferentes de componentes principais e (b) curvas ROC para os métodos avaliados.	50
5.6	(a) Distribuições do erro de estimação (número de entradas em escala linear) para cada método considerado e (b) distribuições de energia para os métodos baseados em filtros casado (número de entradas em escala logarítmica).	52
5.7	Erro relativo do método SMF utilizando sinais simulados de alta amplitude (energia).	53
5.8	Distribuição do ruído experimental do TileCal e o fitting utilizando uma distribuição gaussiana.	54
5.9	Matriz de covariância do ruído do TileCal.	55
5.10	Coefficientes dos estimadores (a) MF e (b) OF2.	56
5.11	Distribuição de energia para (a) MF e (b) OF2.	57
5.12	Curvas ROC para os métodos MF e OF2.	58
5.13	Células selecionadas de um evento utilizando um corte de 300 MeV para as estimativas (a) MF e (b) OF2. Células com cores mais escuras indicam maior deposição de energia.	59
5.14	Distribuição de erros para o MF ($RMS = 17,6$ MeV) e para o OF2 ($RMS = 21,0$ MeV).	60
5.15	Linearidade do método MF.	61
5.16	Análise para dados reais de baixa luminosidade.	63

5.17	Ilustração das posições dos cruzamento de feixes (BC) dentro da janela de leitura do TileCal.	64
5.18	Análise para dados reais de alta luminosidade.	65
5.19	Distribuições das amostras de ruído de fundo para (a) $\langle \mu \rangle = 0$, (b) $\langle \mu \rangle = 40$ e (c) mostra o <i>fitting</i> extraído de $\langle \mu \rangle = 0$ e sobreposto no histograma de $\langle \mu \rangle = 40$ para a célula A13 do TileCal.	69
5.20	(a) Matriz de covariância estimada pelo método tradicional para uma célula de baixa ocupância e (b) seus respectivos coeficientes do método MF determinístico. Célula A3 do TileCal.	70
5.21	(a) Matriz de covariância estimada pelo método robusto (MCDE) para uma célula de baixa ocupância e (b) seus respectivos coeficientes do método MF determinístico. Célula A3 do TileCal.	71
5.22	Distribuição de energia utilizando o MF determinístico com matriz de covariância estimada por abordagens diferentes. (a) célula A3 e (b) célula A13 do TileCal.	72
5.23	Melhora no erro (resolução) através do uso do MF utilizando a matriz robusta contra o método OF2 utilizando a matriz identidade.	73
5.24	Distribuição de energia em células de (a) baixa e (b) alta ocupância utilizando matrizes de covariância diferentes para os métodos MF determinístico e OF2. Células A3 e A13 do TileCal.	74
5.25	Distribuições de energia em células de alta ocupância utilizando matrizes de covariância diferentes (identidade e MCDE).	75
5.26	Melhora na resolução utilizando o método MF com a matriz de covariância em relação à matriz identidade.	76
5.27	Característica típica do comportamento da luminosidade de uma run do LHC.	77
5.28	Coefficientes do método MF para a célula A13 no cenário de 25 ns de espaçamento entre as colisões.	79
5.29	Distribuições de energia na célula A13 utilizando os coeficientes ótimos e generalizados para os cenários de (a) $\langle \mu \rangle = 10$ e (b) $\langle \mu \rangle = 30$, ambos considerando 25 ns de espaçamento entre colisões.	80
6.1	Sistema LTI para dados digitais de calorímetros. $a[k]$ modela a energia depositada em uma dada colisão, $g[k]$ corresponde à resposta ao impulso do sistema, $x[k]$ são os valores adquiridos e $n[k]$ representa o ruído aditivo.	84
6.2	Pulso de referência utilizado para análise de desempenho.	89

6.3	Reconstrução da energia para simulação de empilhamento de sinais. A informação da luminosidade é utilizada nos estimadores MF e OF2.	89
6.4	Reconstrução da energia para simulação de empilhamento de sinais. A informação da luminosidade não é utilizada nos estimadores MF e OF2 ($\mathbf{C} = \mathbf{I}$).	90
6.5	Correlação entre os estimadores DM a_{-1} e a_{+1} . Note que nenhuma correlação é observada para valores acima do ruído, devido a ortogo- nalidade dos estimadores DM.	90
6.6	Diagrama de blocos do método COF.	92
6.7	Distribuição de energia dos métodos COF e MF.	93
6.8	Pulso assimétrico utilizado para análise de dependência da simetria.	94
6.9	Erro de estimação (RMS) em função da ocupância para o conjunto de sinal considerando um valor fixo de $SPR=10$. O desempenho para ambos os pulsos simétrico (na cor preta) e assimétrico (na cor cinza) é mostrado.	96
6.10	Erro de estimação em função do SPR para o conjunto de sinal con- siderando uma ocupância fixa de 10%. O desempenho para ambos os pulsos simétrico (na cor preta) e assimétrico (na cor cinza) é mostrado.	97
6.11	Espectro de energia reconstruído pelo COF e OF2 em que o eixo vertical está na escala (a) linear e (b) logarítmica.	98
6.12	Correlação entre as energias estimadas pelos métodos COF e OF2.	99
6.13	Pesos (coeficientes) do método OF2.	99
6.14	Reconstrução do sinal a partir das amostras digitais em condições de empilhamento de sinais.	100
6.15	Correlação entre os fatores de qualidade dos métodos COF e OF2.	102
A.1	O sistema de coordenadas do ATLAS.	116

Lista de Tabelas

5.1	Número de multiplicações por número de componentes principais (PC) para o sinal com 7 amostras. O número de multiplicações é dado por $C(N + 2C + 1)$, onde C é o número de componentes principais e N é o número de amostras do pulso de referência.	48
5.2	Probabilidade de detecção (PD) para 10% de falso alarme.	49
5.3	Desvio padrão dos histogramas (GeV)	51
5.4	Resumo dos valores de média e RMS das distribuições de energia para as células da camada A (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$	77
5.5	Resumo dos valores de média e RMS das distribuições de energia para as células da camada BC (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$	78
5.6	Resumo dos valores de média e RMS das distribuições de energia para as células da camada D (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$	78
5.7	Resumo dos valores de média e RMS das distribuições de energia para as células da camada E (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$	79
5.8	Resumo dos valores de média e RMS (em MeV) para o cenário de 25 ns de espaçamento entre colisões na célula A13 do TileCal.	81

Lista de Abreviaturas e Siglas

ADC *Analog-to-Digital Converter*

ALICE *A Large Ion Collider Experiment*

ATLAS *A Toroidal LHC ApparatuS*

BC *Bunch Crossing*

BLUE *Best Linear Unbiased Estimator*

CERN *Conseil Européen pour la Recherche Nucléaire*

CIS *Charge Injection System*

CMS *Compact Muon Solenoid*

CNGS *CERN Neutrinos to Gran Sasso*

COF *Constrained Optimal Filter*

DM *Deconvolution Matrix*

DSP *Digital Signal Processor*)

EF *Event Filter*

FIR *Finite Impulse Response*

FPGA *Field-Programmable Gate Array*)

HLT *High Level Trigger*

ISOLDE *Isotope Separation Online Device*

LAr *Liquid Argon*

LHC *Large Hadron Collider*

LHCb *Large Hadron Collider beauty*

LVL1 *Level-1 trigger*

MCDE *Minimum Covariance Determinant Estimator)*

MF *Matched Filter*

MLE *Maximum Likelihood Estimator*

MMSE *Minimal Mean Square Error*

MVU *Minimum Variance Unbiased*

OF *Optimal Filter*

OOT *Out-Of-Time*

PCA *Principal Component Analysis*

PDF *Probability Density Function*

PMT *Photo Multiplier Tube*

RE *Reconstruction Error*

RI *Resolution Improvement)*

ROC *Receiver Operating Curve*

ROD *Read-Out Driver*

SCT *Semi-Conductor Tracker*

SMF *Stochastic Matched Filter)*

SNR *Signal-to-Noise Ratio*

TileCal *Tile Calorimeter*

TDAQ *Trigger and Data Acquisition*

TRT *Transition Radiation Tracker*

RMS *Root Mean Square*

RoI *Region of Interest*

WGN *White Gaussian Noise*

Capítulo 1

Introdução

As áreas da instrumentação eletrônica e processamento digital de sinais têm contribuído de forma significativa para o desenvolvimento tecnológico. Como exemplos podemos citar os grandes avanços na automação industrial e residencial, o crescente aumento da eletrônica embarcada nos veículos automotores, o advento das redes inteligentes na área de sistemas de potência, os avanços na medicina relacionados à instrumentação biomédica, os avanços nas ciências básicas com a utilização de sistemas de instrumentação avançados, entre outros.

O aumento da capacidade de processamento e armazenamento de dados pelos processadores e memórias, o desenvolvimento da tecnologia dos sensores, aliado ao desenvolvimento de técnicas avançadas de processamento digital de sinais, são alguns dos fatores que impulsionaram a grande difusão dos sistemas eletrônicos e, conseqüentemente, o desenvolvimento tecnológico atual.

Estes avanços na área de instrumentação, além dos benefícios já citados, trazem enormes desafios para o campo do processamento digital de sinais, já que existe uma tendência na utilização de um número crescente de sensores atuando em taxas cada vez maiores, resultando em uma quantidade cada vez maior de dados a serem processados em pequenos intervalos de tempo.

A instrumentação para a física experimental de altas energias apresenta-se como uma área extremamente desafiadora, envolvendo justamente o processamento e a aquisição de milhares de sinais a taxas extremamente altas. Esta tese se insere neste contexto, apresentando contribuições no desenvolvimento de técnicas de processamento de sinais para a estimação de energia em calorimetria de altas energias.

Os experimentos modernos em física de altas energias se apoiam fortemente nos seus sistemas de calorimetria, que exercem tarefas cruciais, tais como absorção e amostragem da energia das partículas incidentes, sendo sua informação utilizada para reconstrução das partículas resultantes da colisão e também para seleção de eventos. As informações fornecidas pelos sistemas de calorimetria têm impacto direto na identificação de diferentes partículas, bem como nas diferentes análises

físicas desenvolvidas no experimento.

Do ponto de vista físico, um calorímetro é um bloco de matéria que intercepta partículas primárias, e que deve ser suficientemente espesso para que toda a energia das partículas seja depositada em seu volume. Parte da energia inicial é dissipada na forma de calor, mas uma fração é amostrada de forma mais prática para o processamento, sendo proporcional à energia inicial.

Com o objetivo de determinar as propriedades relevantes das partículas, o calorímetro é segmentado em células que fornecem resolução espacial ao detector. Os calorímetros atuais podem possuir milhares de células. O material pesado (número atômico alto) utilizado pelo calorímetro é escolhido de acordo com a partícula de interesse a qual interage e deposita sua energia.

Quando partículas são absorvidas pelo calorímetro, suas propriedades podem ser reconstruídas a partir dos sinais produzidos por todas as células que foram envolvidas no processo da geração do chuveiro. A energia depositada em cada célula é medida individualmente através de sua eletrônica de leitura dedicada. Se a luz for a fonte de informação, a cadeia de leitura eletrônica pode consistir de um transdutor que converte a luz em sinal elétrico, um amplificador, um circuito de conformação de sinal e um conversor analógico-digital. Do ponto de vista do sistema de aquisição de dados, cada célula corresponde a um canal de leitura individual.

Tipicamente, os sinais utilizados para determinar a energia são fornecidos por um circuito eletrônico de conformação do pulso analógico. Tal processamento visa fornecer um pulso estável e de forma fixa o qual tem sua amplitude proporcional à energia do sinal. Desta forma, através da estimativa da amplitude do pulso, o sinal e sua energia podem ser reconstruídos. Idealmente, os parâmetros do pulso são determinísticos exceto a amplitude, a qual depende da energia depositada pela partícula.

1.1 Motivação e Objetivo

Conforme abordado acima, nos calorímetros, a reconstrução de energia pode ser dividida em duas etapas, a identificação dos canais com informação relevante e a estimação da energia depositada nesses canais. A energia é a informação mais relevante nos sistemas de calorimetria, e esta informação é fundamental para a reconstrução final do evento e identificação das partículas, lembrando que é através dessa identificação que os modelos para fenômenos físicos conhecidos, e desconhecidos, são testados e validados.

Os métodos comumente utilizados nos calorímetros modernos para reconstrução de energia, formulam o problema como sendo de estimação da amplitude de um pulso imerso em ruído aditivo, realizando a identificação dos canais com informação

relevante através de um corte em energia na saída do estimador. Adicionalmente, algumas hipóteses para a abordagem do problema são comumente utilizadas como a consideração da forma do pulso fixo e do ruído aditivo como sendo gaussiano. Desta forma, estes são os pontos principais de motivação desta tese, já que tanto a abordagem do problema de reconstrução de energia em calorimetria quanto as hipóteses consideradas podem ser estudadas e avaliadas com mais profundidade, visando a obtenção de métodos com melhor desempenho.

Em experimentos que operam em condições ainda mais extremas (e.g. maiores taxas de eventos), o problema da reconstrução de energia ganha em complexidade pois pode ocorrer o problema de empilhamento de sinais, ou seja, a janela de aquisição pode possuir, além do sinal de interesse (aquele em que se deseja estimar a amplitude), outros sinais sobrepostos. Como estes cenários ainda são raros, não existem ainda métodos para reconstrução de energia em calorimetria que tratam especificamente este problema, servindo também de motivação para o estudo e desenvolvimento de novos métodos para reconstrução de energia.

Portanto, o principal objetivo desta tese é o desenvolvimento de estudos e métodos para a reconstrução de energia em calorimetria de altas energias.

1.2 Principais Contribuições da Tese

1. Foi proposto um método para reconstrução de energia em calorimetria baseado em filtragem casada. A estimação da energia é obtida através de uma simples calibração da saída do algoritmo de detecção. Portanto, a identificação dos canais com informação relevante e a estimação da energia são realizadas concomitantemente. Desta forma, o problema é formulado do ponto de vista da detecção, sendo a estimação obtida como consequência do processo de detecção.
2. Visando a utilização do método proposto em cenários aonde existem flutuações na forma do pulso do calorímetro, foi proposta a aproximação do processo randômico descrito pelo sinal recebido pela série de Karhunen-Loève, gerando assim um banco de filtros para detecção e estimação da energia nos canais do calorímetro.
3. O método proposto foi avaliado no contexto de um dos principais experimentos de física de partículas do mundo.
4. Para o cenário de empilhamento de eventos, foi proposto um estimador, chamado de COF (*Constrained Optimal Filter*), que trata o sinal recebido como uma mistura linear de sinais defasados.

5. O COF também foi avaliado no contexto de um dos principais experimentos de física de partículas do mundo, em condições de empilhamento de sinais.

1.3 O Ambiente Experimental de Desenvolvimento

Este trabalho foi desenvolvido no contexto do experimento ATLAS, que é um dos principais detectores do maior acelerador de partículas já construído, o LHC (do inglês *Large Hadron Collider*), localizado no CERN (do francês *Centre Européenne pour La Recherche Nucleaire*). O ATLAS é composto por diversos subdetectores, entre eles os calorímetros eletromagnético e hadrônico. Os métodos propostos nesta tese foram avaliados no calorímetro hadrônico do ATLAS e comparados com o método padrão até então utilizado.

1.4 Organização do texto

No Capítulo 2, será apresentado o ambiente no qual esta proposta de tese foi desenvolvida. O experimento LHC e seu detector ATLAS serão detalhados, dando uma atenção especial aos calorímetros deste detector. A cadeia de processamento do sinal no calorímetro hadrônico será apresentada, assim como as características de seu sinal de interesse. Ao final, será brevemente discutido o programa de atualização para o cenário de alta luminosidade do LHC.

O Capítulo 3 apresenta uma revisão bibliográfica sobre o problema de reconstrução de energia em calorimetria para física de altas energias. O problema de estimação de parâmetros e técnicas clássicas para solução deste problema serão apresentadas. Em seguida, o algoritmo utilizado atualmente para detecção e estimação da energia no TileCal será descrito em detalhes. Por fim, a reconstrução de energia em outros calorímetros modernos será discutida.

No Capítulo 4, o método proposto para reconstrução de energia em calorimetria baseado na razão da máxima verossimilhança é apresentado. Inicialmente, o problema de detecção de um sinal imerso em ruído será descrito utilizando a abordagem proposta. Será descrito o procedimento para fornecer a estimativa da energia a partir da saída dos detectores de sinal.

O Capítulo 5 apresenta a avaliação da utilização do método proposto dentro do ambiente de desenvolvimento desta tese. Serão apresentados resultados de detecção de sinais e estimação de energia utilizando sinais simulados e dados reais adquiridos pelo LHC. O desempenho da técnica proposta é comparada com métodos atualmente utilizados em calorímetros de altas energias.

No Capítulo 6, a teoria de desconvolução de sinais será apresentada. O método COF, proposto para o ambiente de empilhamento de sinais, será introduzido destacando-se suas vantagens em relação às abordagens clássicas. A análise de desempenho será feita utilizando dados simulados e experimentais.

Finalmente, no Capítulo 7, serão apresentadas as discussões sobre o trabalho realizado, dando ênfase às contribuições e perspectivas de trabalhos futuros.

Capítulo 2

O Ambiente de Desenvolvimento: LHC, ATLAS e TileCal

A presente tese de doutorado foi desenvolvida dentro do grupo do calorímetro hadrônico do detector ATLAS, no Centro Europeu para a Pesquisa Nuclear (CERN). Portanto, neste capítulo, serão descritos o experimento LHC, o detector ATLAS, o sistema de seleção de eventos *online* e seus calorímetros.

Como o objetivo principal deste trabalho é o desenvolvimento de técnicas para reconstrução de energia para calorímetros, e a aplicação das técnicas forma feitas no calorímetro hadrônico do ATLAS (chamado de TileCal), o problema da reconstrução de energia no TileCal é apresentado.

2.1 O acelerador LHC no CERN

O CERN [1] é o maior centro de física de partículas do mundo e se localiza na fronteira franco-suíça. Ao total, o CERN possui 22 estados membros, mas vários outros países estão envolvidos com CERN em diversos tipos de atividades.

No CERN, pesquisadores estudam, através de instrumentos científicos complexos, a estrutura fundamental da matéria explorando a física além do Modelo Padrão [2]. Dentre os experimentos realizados no CERN, os aceleradores de partículas [3] estão entre os mais complexos, envolvendo normalmente grandes quantidades de recursos financeiros e grandes colaborações para o desenvolvimento, operação e análise dos resultados.

Diversas descobertas se originaram no CERN, tais como a corrente neutra das interações fracas em 1974 pelo experimento Gargamelle [4], a observação das partículas W e Z em 1983 pelos experimentos UA1 e UA2 [5], e as medidas mais precisas da partícula Z em 1989, as quais mostraram que existem somente três famílias de partículas na natureza. Em reconhecimento pelas suas descobertas junto ao CERN,

alguns pesquisadores receberam o prêmio nobel tais como Carlo Rubbia e Simon Van der Meer. Além disso, alguns desenvolvimentos tecnológicos criados no CERN podem ser destacados como a invenção da WWW (*World Wide Web*) nos anos 80.

A Figura 2.1 mostra uma foto recente do globo de ciência e inovação no CERN, situado num lugar de destaque e que abriga exposições e palestras, informando os visitantes sobre as descobertas mais relevantes no CERN.



Figura 2.1: Vista do globo de ciência e inovação no CERN (extraído de [6]).

A infraestrutura do principal acelerador do CERN é ilustrada pela Figura 2.2. Diversas etapas são sucessivamente realizadas pelas várias máquinas, onde os feixes de partículas são injetados de uma para a outra, aumentando suas energias. Atualmente, o Grande Colisionador de Hadrons (LHC) [7] é o principal acelerador do CERN. Este colisionador de partículas é o mais energético do mundo e vem operando com sucesso desde 2010, colidindo prótons com energia de centro de massa variando entre 0,9 e 13 TeV, com intervalo entre colisões de 50 e 25 ns. O LHC foi construído, no CERN, em um túnel circular de 27 km de circunferência a aproximadamente 100 metros abaixo da superfície terrestre, como ilustrado na Figura 2.3. Neste túnel, dois feixes de prótons são acelerados próximo à velocidade da luz e colididos em pontos específicos. Detectores são posicionados estrategicamente no túnel e coletam as informações provenientes das colisões.

Além do LHC, o programa de pesquisa do CERN inclui outros experimentos como, por exemplo, o experimento de neutrinos CNGS (do inglês, *CERN Neutrinos*

CERN's Accelerator Complex

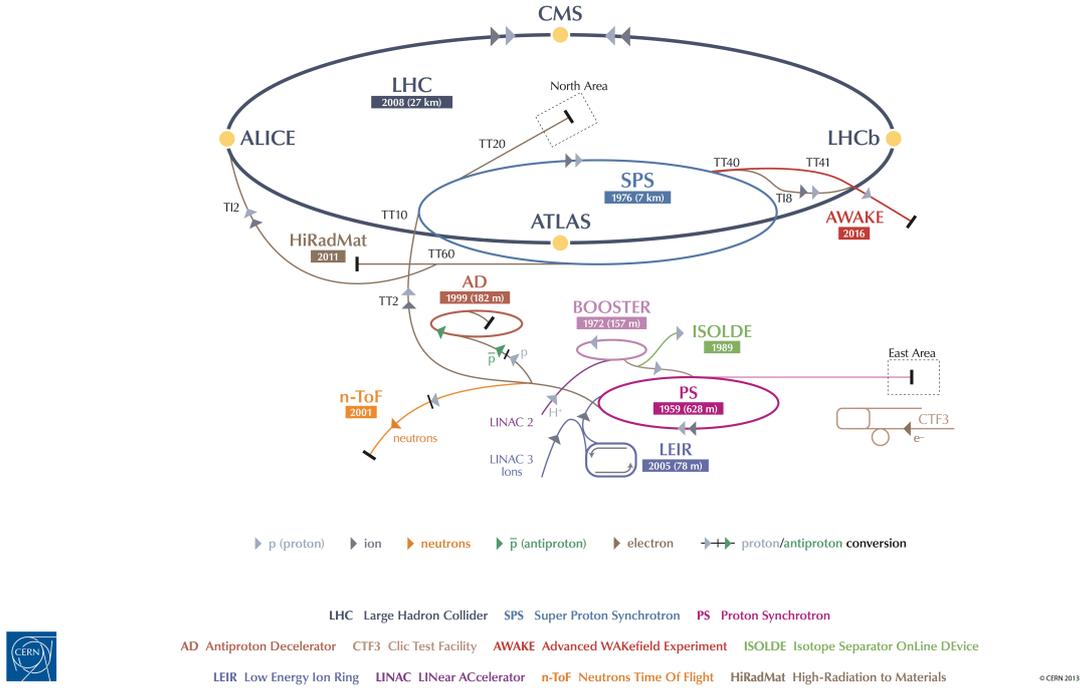


Figura 2.2: O complexo do acelerador (extraído de [8]).

to *Gran Sasso*) [9] e desenvolvimento de tecnologias para experimentos em física de altas energias, tais como o projeto GRID [10], criado para lidar com a enorme quantidade de dados adquirida pelo LHC.

2.1.1 Os principais detectores do LHC

Quatro experimentos grandes estão instalados ao longo do túnel no LHC (ver Figura 2.3). Dois detectores de propósito geral, o ATLAS (*A Toroidal LHC ApparatuS*) [11] e o CMS (*Compact Muon Solenoid*) [12], que foram projetados e otimizados para estudar a física na escala de energia de TeV. Os outros dois experimentos, o LHCb (*Large Hadron Collider-beauty*) [13] e o ALICE (*A Large Ion Collider Experiment*) [14], foram projetados para estudar fenômenos específicos. O LHCb foi desenvolvido para estudos da violação da simetria CP e decaimentos raros de mésons. Já o ALICE, destina-se ao estudo do plasma de quarks e glúons obtido pela colisão de íons pesados (colisões Pb-Pb).

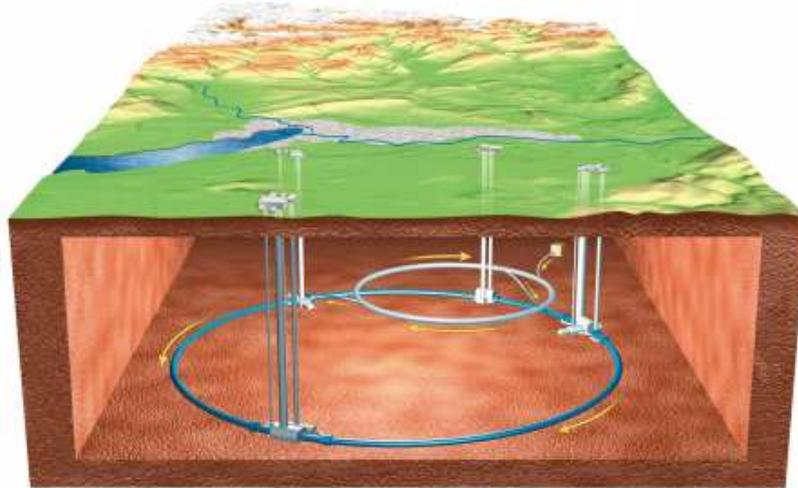


Figura 2.3: O anel do LHC e seus quatro detectores principais (extraído de [15]).

2.2 O detector ATLAS

O ATLAS é o maior detector do LHC, possuindo aproximadamente 45 metros de comprimento, 25 metros de altura e em torno de 7000 toneladas (ver Figura 2.4). Os componentes principais do ATLAS são: o Detector Central (do inglês, *Inner Detector*) [16], o Solenóide (do inglês, *Solenoidal Magnets*) que envolve o Detector Central, os calorímetros Eletromagnético e Hadrônico [17], o Toróide (do inglês, *Toroid Magnets*) e o Espectrômetro de Múons (do inglês, *Muon Spectrometer*) [18]. O sistema de coordenadas do ATLAS está detalhado no Apêndice A.

O detector ATLAS tem um programa de física extenso sendo projetado para explorar todo o potencial de descoberta do LHC. Depois de um intenso período sendo comissionado, o ATLAS está em operação desde 2010 e comprovações da existência do bóson de Higgs, partícula prevista pelo Modelo Padrão [2], já foram observadas pelo experimento [20].

Devido à alta taxa de colisões do LHC, ao grande número de informações geradas no ATLAS e à imensa quantidade de ruído de fundo gerado nas colisões, um sistema de filtragem online de eventos é fundamental para o operação do experimento.

2.2.1 Sistema de filtragem *online* de eventos do ATLAS

O sistema de filtragem *online* seleciona eventos em três estágios diferentes, nível 1 (LVL1) [21], nível 2 (LVL2) [21] e o Filtro de Eventos (EF, do inglês, *Event Filter*) [22]. Os dois últimos estágios, juntos, são chamados de Trigger de Alto Nível (HLT, do inglês, *High Level Trigger*) sendo implementados em *software* utilizando ferramentas de computação e *hardware* disponíveis comercialmente. O LVL1 é im-

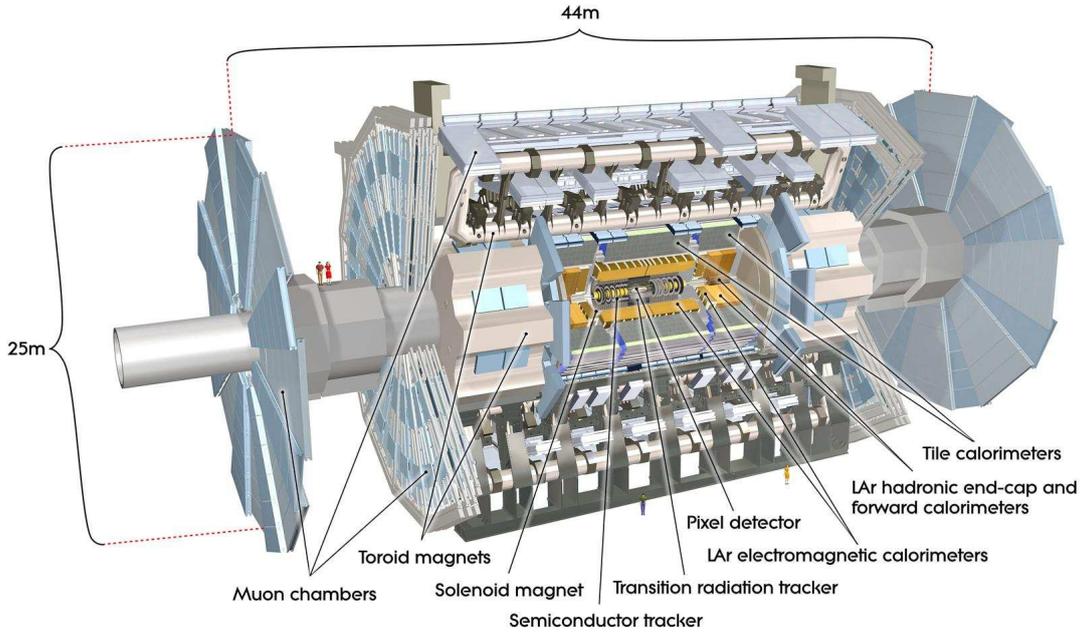


Figura 2.4: O detector ATLAS e seus sub-sistemas (extraído de [19]).

plementado utilizando-se eletrônica dedicada, e reduz a taxa de eventos de 40 MHz (taxa de eventos do LHC) para 100 kHz.

Para cada trigger aceito pelo LVL1, uma região de interesse associada a este evento é propagada ao LVL2, alimentando os algoritmos deste nível e reduzindo a taxa de eventos para 2 kHz [23]. Ao fim do processo de filtragem de dados, apenas 200 Hz são gravados em disco para serem reconstruídos e analisados *offline*. A Figura 2.5 mostra o diagrama de blocos com o sistema de aquisição de dados e os níveis de trigger do ATLAS.

Vale ressaltar que a seleção de sinais do LVL1 é feita baseada na informação proveniente dos sistemas de calorimetria e múons. O LVL1 foi projetado para processar a informação compactada enviada por estes sistemas. Nos calorímetros, esta informação corresponde à soma dos sinais presentes em uma mesma região de η do calorímetro. Esta informação é utilizada pelo LVL1 para localizar elétrons e fótons no plano transversal ao eixo do feixe (momento transversal, p_T) e transferir esta informação ao HLT.

2.2.2 Calorimetria do ATLAS

Em física de altas energias, o calorímetro é responsável por absorver, amostrar e medir a energia das partículas que o atravessam [24]. Ao penetrarem no calorímetro, inicia-se o processo de chuva de partículas [25] em que a energia das

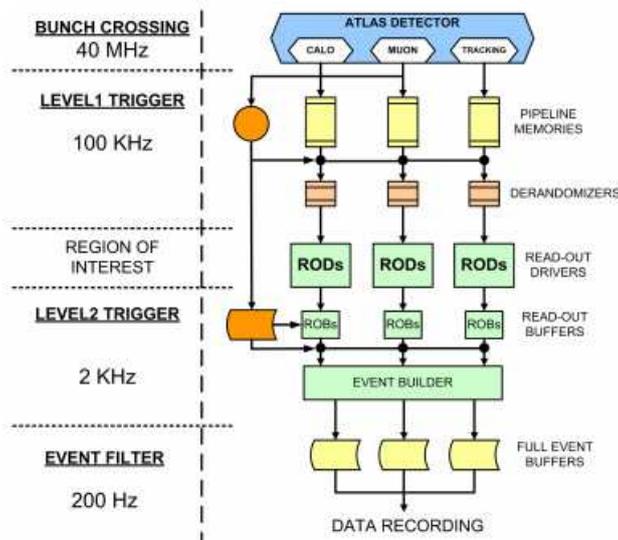


Figura 2.5: Sistema de trigger do ATLAS e seus níveis (extraído de [15]).

partículas é depositada e amostrada. Tipicamente, calorímetros são transversalmente segmentados para obter informação da direção das partículas, bem como da energia depositada. Uma segmentação longitudinal também pode ser feita, para obter informação da identidade da partícula, baseada na forma do chuveiro que ela produz [26].

Algumas razões pelas quais os calorímetros se tornaram sistemas importantes em diversos experimentos em física de partículas são:

1. Calorímetros podem ser sensíveis tanto a partículas neutras quanto a carregadas.
2. Devido às diferenças na forma de deposição de energia, a identificação de partículas pode ser feita com alta eficiência.
3. Para conter o desenvolvimento de cascatas dos objetos a serem medidos, a profundidade dos calorímetros aumenta logaritmicamente com a energia, o que permite o projeto de detectores mais compactos.
4. Calorímetros podem ser segmentados, o que permite tanto medida da energia quanto de trajetória das partículas.
5. Resposta rápida (menor que 50 ns) pode ser atingida com os calorímetros, o que é importante num ambiente com alta taxa de eventos.
6. A informação de energia obtida dos calorímetros pode ser usada para filtrar eventos interessantes com alta seletividade.

O sistema de calorimetria do ATLAS inicia-se a 1,5 metros do feixe de colisão. Uma vez que as interações de hádrons, léptons e fótons com a matéria acontece de formas diferentes, o sistema de calorimetria do ATLAS é dividido em duas partes: o calorímetro eletromagnético e o calorímetro hadrônico. O calorímetro eletromagnético foi projetado para medir a energia de partículas que interagem eletromagneticamente, enquanto o calorímetro hadrônico é dedicado à medir a energia de hádrons. Exceto pelos múons e pelos neutrinos, todas partículas são absorvidas pelos calorímetros [27]. A Figura 2.6(a) ilustra, de forma esquemática, os padrões de deposição de energia nos diferentes sub-sistemas do ATLAS que permitem identificar tipos diferentes de partículas produzidas nas colisões. O sistema de calorimetria do ATLAS é mostrado na Figura 2.6(b).

2.2.2.1 Calorímetro Eletromagnético

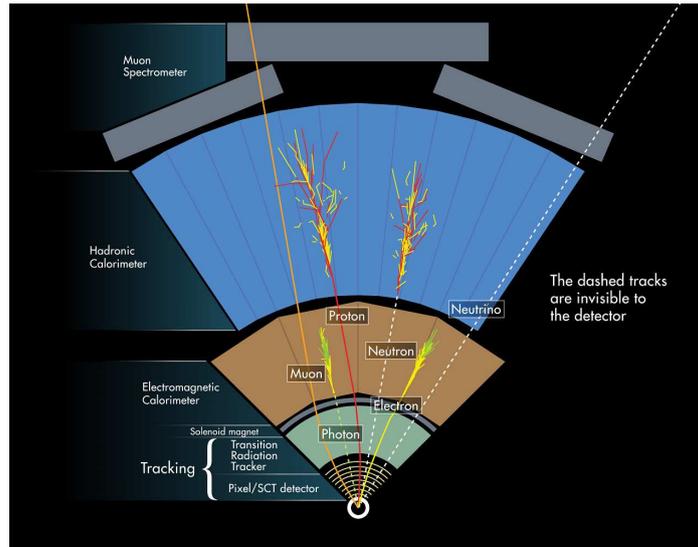
O Calorímetro de Argônio Líquido (LAr, do inglês *Liquid Argon*) [29] é o calorímetro eletromagnético (EM) do ATLAS que utiliza o chumbo como material absorvedor, e os materiais ativos (amostradores) são compostos de eletrodos de chumbo imersos em argônio líquido, com a forma de acordeões. Os eletrodos amostram a energia da partícula através da ionização das células do calorímetro. Este detector é dividido em duas partes fundamentais: o barril (do inglês, *barrel*) e as tampas (do inglês, *end-caps*). Juntas, estas partes completam o espaço ao redor do ponto de colisão até um valor de $\eta = 3,2$. A porção do barril de tal calorímetro estende-se de $\eta = 0$ até $\eta = 1,475$. Em $\eta = 1,375$ o barril começa a sobrepor a tampa, que é dividida entre tampa exterior (até $\eta = 2,5$) e interior ($\eta = 2,5$ até $\eta = 3,2$).

No que tange à sua segmentação, este calorímetro é dividido em três camadas, das quais a segunda é a mais profunda. Cada camada possui granularidade específica, que ajuda a determinar alguns aspectos dos objetos que interagem com este detector. O calorímetro eletromagnético do ATLAS possui uma granularidade constante com relação à rotação (eixo ϕ), mas variável com relação a η . Esta diversificação da granularidade ao longo do plano $\eta \times \phi$, pode ser observada na Figura 2.7.

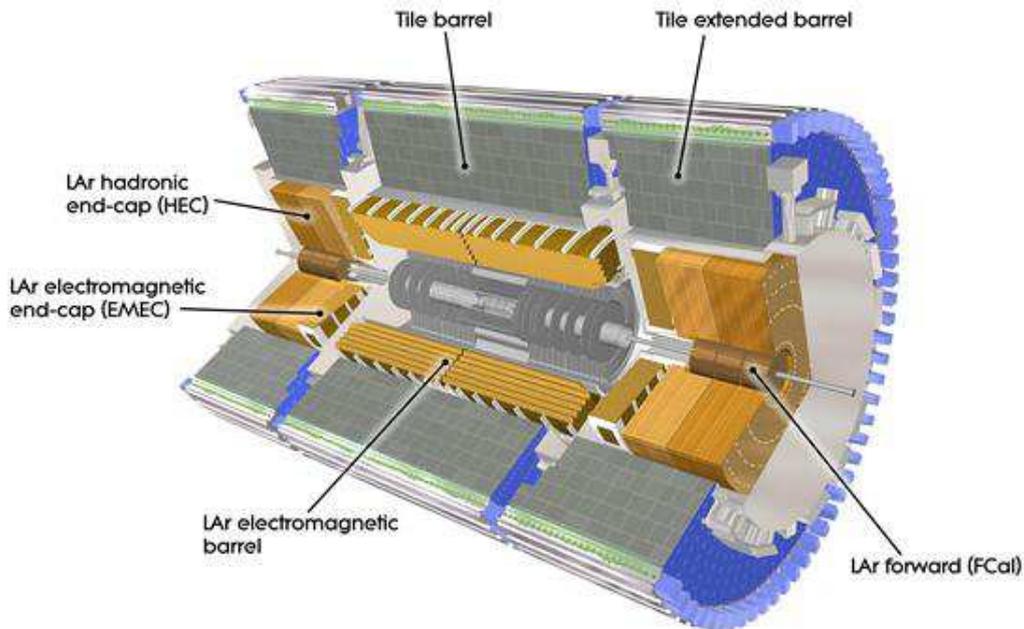
2.2.3 O Calorímetro de Telhas do ATLAS

O calorímetro de telhas do ATLAS, ou TileCal, servirá como o contexto de aplicação desta tese de doutorado e, sendo assim, será descrito em detalhes nesta seção.

O TileCal é o principal calorímetro hadrônico do detector ATLAS e atingiu suas expectativas durante os primeiros anos de operação do LHC [30]. O objetivo do TileCal é realizar medidas precisas de hádrons, jatos e taus. Além disso, a informação proveniente do TileCal é utilizada para a tomada de decisão realizada



(a)



(b)

Figura 2.6: (a) Esquemático de uma fatia transversal do detector ATLAS mostrando os padrões de deposição de energia nos calorímetros do ATLAS (extraído de [28]), e (b) sistema de calorimetria do ATLAS (extraído de [28]).

pelo sistema de filtragem *online* de eventos do ATLAS [22].

O TileCal é formado por três cilindros, um barril longo central dividido em duas partições independentes, LBA e LBC, e dois barris estendidos, EBA e EBC. Ao todo, o TileCal cobre a maior parte da região central do ATLAS, com $\eta < 1.7$ (na Figura 2.6(b) o TileCal é o calorímetro mais externo).

No que tange o seu princípio de funcionamento, o TileCal utiliza placas de aço

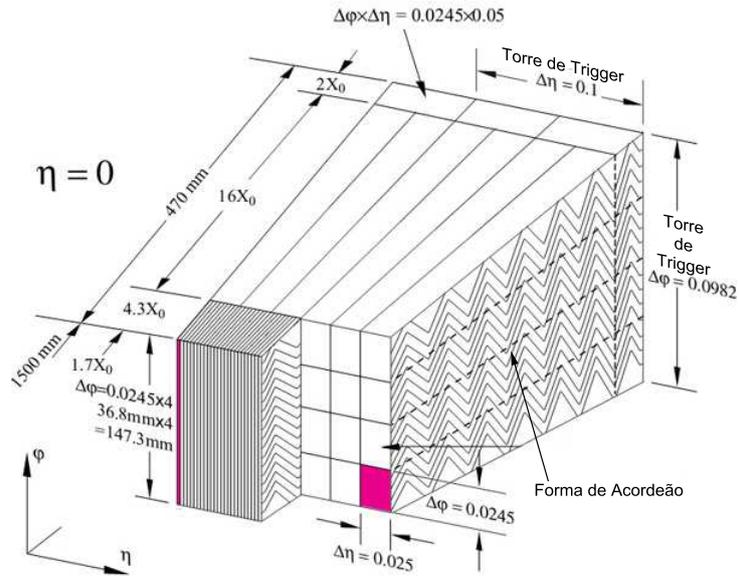


Figura 2.7: Segmentação do calorímetro eletromagnético do ATLAS (extraído de [29]).

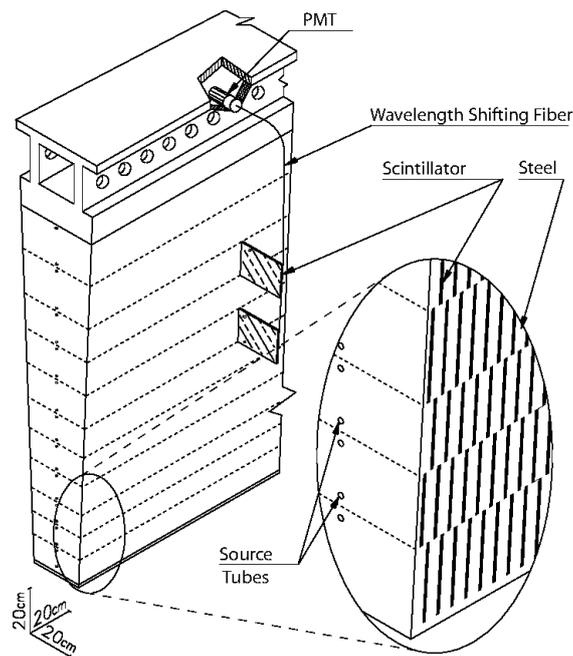


Figura 2.8: Diagrama esquemático mostrando os elementos que compõem a leitura do sinal físico pelo TileCal, correspondente a um módulo (extraído de [11]).

como material absorvedor e telhas plásticas cintilantes como material ativo [31]. As partículas hadrônicas provenientes das colisões interagem com o material pesado, depositando a energia que é parcialmente amostrada pelas telhas cintilantes na forma de luz. A luz é em seguida transmitida por fibras óticas e lida por tubos

fotomultiplicadores (PMTs), os quais geram os sinais elétricos que devem ser processados. A Figura 2.8 mostra o desenho esquemático do sistema de coleta do sinal correspondente a um módulo do TileCal.

Cada módulo do barril central e do barril estendido é dividido, respectivamente, em 23 e 16 células com dupla leitura, resultando em aproximadamente 10.000 canais (sinais). Além disso, cada barril é dividido em 64 módulos em azimute, ϕ , resultando numa granularidade de $\Delta\phi = 0,1$. Cada módulo é segmentado radialmente (profundidade) em três camadas de células de leitura com uma granularidade de $\Delta\eta = 0,1$ para as duas primeiras camadas e $\Delta\eta = 0,2$ para a terceira camada. As camadas pertencentes a cada $\Delta\eta$ são somadas e esta informação compactada, chamada de Torre de Trigger, é enviada ao primeiro nível de seleção de eventos (LVL1) do ATLAS. A segmentação de células do TileCal é representada pelo Figura 2.9 para um módulo do barril central (apenas $\eta > 0$) e um módulo do barril estendido.

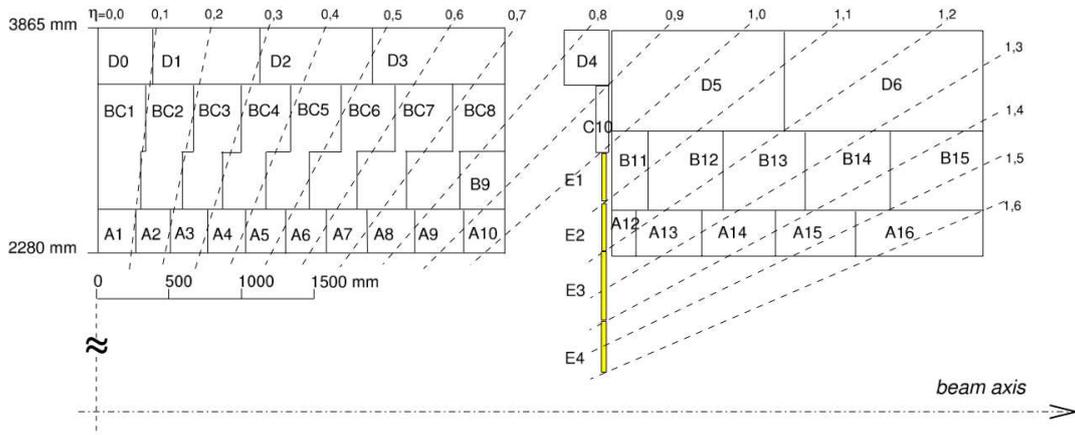


Figura 2.9: Vista esquemática da segmentação de células de um módulo do barril central ($\eta > 0$) e um módulo do barril estendido (extraído de [32]).

O pulso rápido gerado na saída da PMT é condicionado por um circuito de conformação (*shaper*) [33], o qual fornece um pulso com amplitude proporcional à energia depositada. Consequentemente, a forma do pulso pode ser considerada aproximadamente como invariante de canal a canal [34]. A energia depositada pela partícula em um dado canal pode então ser calculada através da estimação da amplitude do pulso, sendo uma situação comumente observada nos calorímetros modernos. Portanto, uma das etapas para reconstrução de energia no TileCal é realizar a estimação da amplitude de um pulso, entretanto, devido à grande quantidade de canais, apenas os canais com informação relevante devem ser selecionados e utilizados na reconstrução de energia.

Com a finalidade de cobrir toda a faixa de energia requerida (220 MeV até 1,3 TeV) o pulso conformado é amplificado por dois amplificadores com razão de ganho de 64. Os sinais analógicos são convertidos em digitais por um conversor

Analógico-Digital (ADC) com frequência de amostragem de 40 MHz e uma janela com 7 amostras (150 ns) é suficiente para cobrir todo o pulso. O pulso analógico e suas amostras, representadas por pontos, podem ser vistos na Figura 2.10.

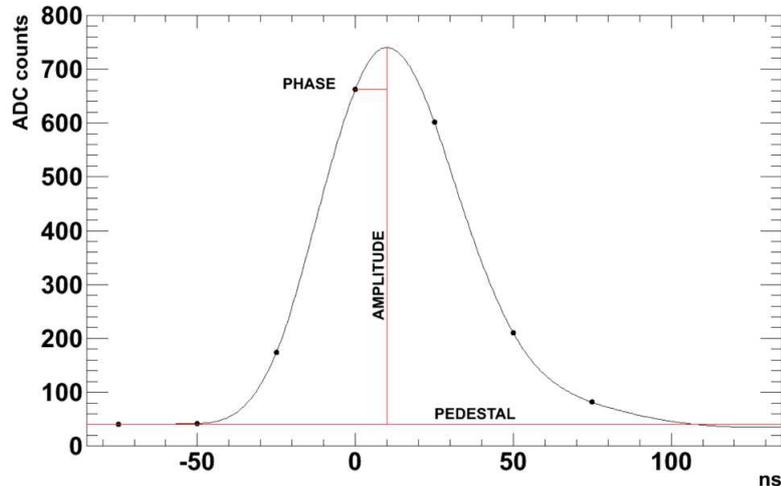


Figura 2.10: Pulso de referência do TileCal (extraído de [35]).

Na Figura 2.10, o pedestal é definido como sendo a linha de base do sinal e a amplitude é a altura do sinal medido a partir do pedestal. A fase do sinal pode ser medida como a diferença de tempo entre a amostra central (quarta amostra) e o pico do pulso. As sete amostras são finalmente enviadas através de fibras óticas, para os *Read Out Drivers* (RODs), aonde a energia é estimada para os eventos aceitos pelo LVL1.

Todo o processamento do sinal, desde a sua coleta nas telhas cintilantes até a transmissão do sinal digital dos RODs para o HLT, pode ser visto no diagrama de blocos da Figura 2.11. A informação analógica compactada do TileCal, que é enviada ao LVL1 para a seleção das regiões de interesse, está representada pelo bloco ($\Sigma cell$).

Ainda, na Figura 2.11, podem ser notados dois dos três principais sistemas de calibração desenvolvidos e utilizados no TileCal, e que são destacados a seguir:

- Sistema de Injeção de Carga (CIS, do inglês, *Charge Injection System*) [37]: é utilizado para derivar os fatores de conversão de contagens de ADC para pC. Este sistema simula pulsos físicos nos canais do TileCal a partir da descarga de capacitores de 5,2 e 100 pF. Este sistema fornece uma medida que relaciona os sinais analógicos e sua respectiva resposta na eletrônica de leitura.
- Sistema de Laser [38]: este sistema mede o ganho da PMT no TileCal. Uma quantidade controlada de luz é enviada para monitorar os foto-diodos de cada

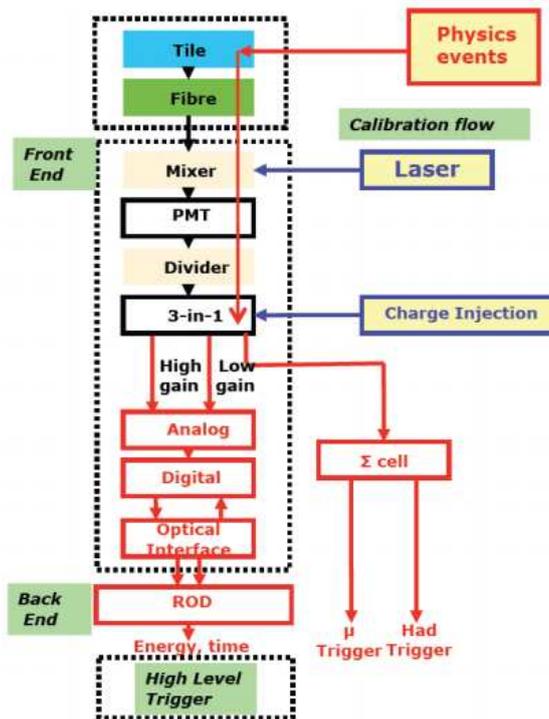


Figura 2.11: Cadeia de processamento do sinal do TileCal (extraído de [36]).

PMT do TileCal simultaneamente. As medidas deste sistema são utilizadas para calibração de fase dos pulsos e corrigir variações no ganho das PMTs

- Sistema de Cesium [39]: este sistema é baseado numa fonte móvel radioativa controlada por dispositivos hidráulicos. Uma fonte γ ^{137}C atravessa todas as telhas para equalizar a resposta dos canais e monitorar a estabilidade dos componentes óticos.

2.3 Estabilidade do sinal do TileCal

No TileCal, os sistemas de calibração, mencionados na seção anterior, fornecem uma calibração de alto desempenho nos pulsos adquiridos pela eletrônica de leitura. Um dos objetivos da calibração do pulso é verificar se a forma do pulso é fixa e uniformizar os ganhos dos milhares de canais para que os algoritmos de reconstrução da energia operem de forma eficiente, uma vez que estes se apoiam na forma do sinal de interesse. A Figura 2.12 representa a estabilidade do pulso de interesse do TileCal utilizando dados reais.

No que tange a deformação na forma do pulso, a curva superior da Figura 2.12(a) mostra o desvio do pulso normalizado reconstruído por todos canais do TileCal, e o pulso parametrizado armazenado em banco de dados. Pode-se observar que a tolerância da eletrônica do TileCal gera pequenas variações na forma do pulso. En-

tretanto, tais variações não apresentam deformações significativas quando observados os desvio padrão da distribuição de cada ponto da curva, como mostra a curva inferior da mesma figura.

A Figura 2.12(b) mostra a distribuição do desvio de fase nos canais de leitura do calorímetro. Apesar da distribuição não possuir um perfil gaussiano fiel, pode-se inferir que a maioria dos eventos apresenta valores de desvio abaixo de 2 ns. Estes eventos foram adquiridos recentemente, em 2015, durante o início da segunda tomada de dados do LHC.

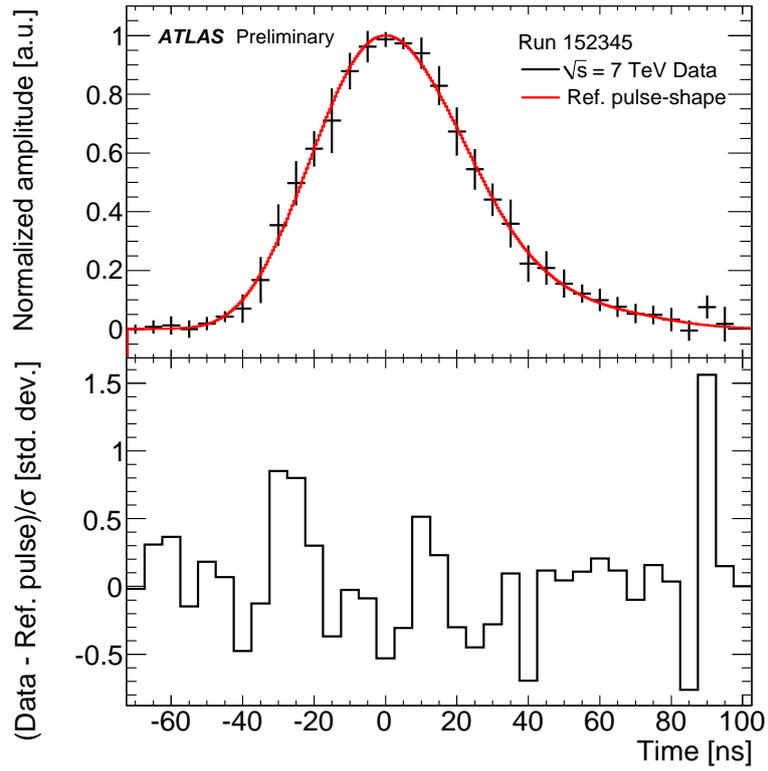
Outro parâmetro importante do sinal do TileCal, o qual impacta diretamente no desempenho da estimação de energia, é a estabilidade da linha de base, ou pedestal. No TileCal, o valor do pedestal é calculado através de tomada de dados dedicadas a medir este parâmetro. Neste tipo de análise, o detector é colocado em funcionamento, mas sem a ocorrência de colisões. Desta forma, a resposta do detector será caracterizada principalmente por ruído eletrônico do experimento e, através da aquisição destes dados, pode-se estimar o valor da linha de base calculando-se, por exemplo, a média da primeira amostra digital de um conjunto com milhares de observações.

Os valores de pedestal são armazenados para cada canal de leitura em banco de dados para uso durante a reconstrução do sinal. Vale destacar, que alguns algoritmos de estimação da energia como, por exemplo, os métodos propostos nesta tese, fazem uso deste parâmetro no processo de estimação da energia. Nestes casos, a estabilidade e a qualidade da estimação deste parâmetro é de máxima relevância.

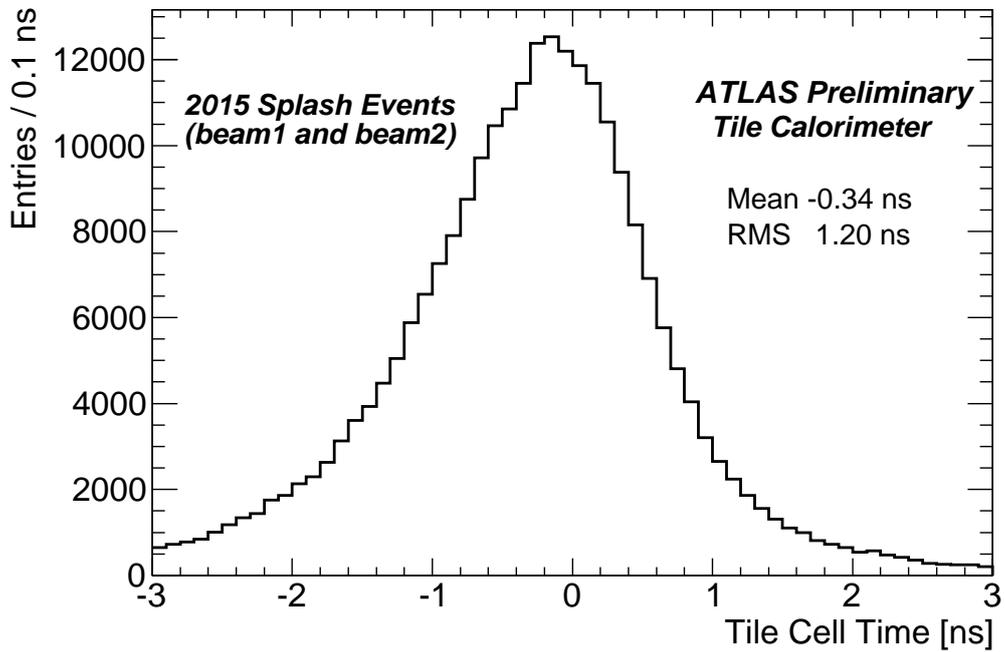
Uma forma de avaliar a estabilidade da linha de base é conferindo se o valor da linha de base armazenado no banco de dados se aproxima do valor medido em tomadas de dados de calibração. A Figura 2.13 mostra as distribuições da diferença entre a linha de base presente no banco de dados e o valor calculado da tomada de dados para a partição LBA do TileCal. Esta análise se refere a uma tomada de dados recente (número 265964), realizada em 26 de maio de 2015, utilizando a sistema de calibração por injeção de carga, onde aproximadamente 100.000 eventos foram adquiridos. Ou seja, cada entrada do histograma da Figura 2.13 corresponde à média calculada sobre 100.000 eventos, subtraída pelo valor no banco de dados (DB). Pode-se observar que o valor medido da tomada de dados possui um desvio muito pequeno quando comparado com o valor armazenado no banco, caracterizando assim, a boa estabilidade da linha de base do sinal do TileCal.

2.4 Reconstrução da Energia do TileCal

A reconstrução da energia no TileCal é baseada na estimação da amplitude do pulso de interesse. O procedimento de estimação é baseado em uma soma ponderada



(a)



(b)

Figura 2.12: (a) Sinal reconstruído utilizando dados reais e o pulso de referência armazenado no banco de dados e (b) Distribuição de fase dos sinais adquiridos durante calibração com feixes reais em 2015 (extraído de [40]).

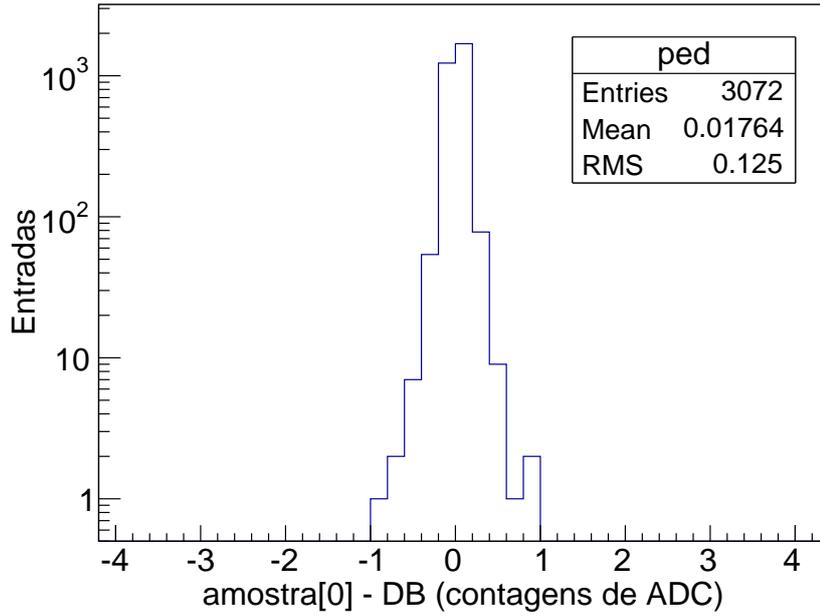


Figura 2.13: Estabilidade do pedestal medido pela diferença entre o valor médio da primeira amostra digital e seu valor armazenado em banco de dado, para a partição LBA.

das amostras digitais recebidas:

$$\hat{A}_{OF} = \sum_{k=0}^{K-1} w[k]r[k], \quad (2.1)$$

onde $w[k]$ são os coeficientes do filtro e $r[k]$ a amostra do sinal recebido no instante k . Os pesos $w[k]$ são obtidos a partir do pulso de referência do canal e da matriz de covariância do ruído neste respectivo canal. O procedimento tem como objetivo minimizar a variância da distribuição da amplitude estimada.

Este método é simples e apropriado para aplicações *online* pois pode ser implementado através de um filtro de resposta ao impulso finita. No TileCal, tal técnica é empregada tanto nos dispositivos DSP (do inglês, *Digital Signal Processor*), que realizam a estimação *online* da energia, quanto nas análises *offline*. Este algoritmo será melhor detalhado no capítulo seguinte. Vale ressaltar que o TileCal armazena todos os sinais (sete amostras) de todos os canais para o processamento *offline*, desde que a diferença entre a amostra de maior valor e a de menor valor seja superior a 5 contagens de ADC.

A seleção dos canais que irão entrar na reconstrução de energia para um determinado evento é realizada através de um corte na energia estimada. Desta forma, a energia de todos os canais pertencentes à região de interesse (RoI, *Region of Interest*) destacada pelo LVL1 é passada para o LVL2, onde o corte em energia é realizado

para seleção dos canais onde ocorreu deposição de energia relevante.

2.5 O programa de atualização do TileCal para o aumento da luminosidade do LHC

A física de interesse é rara e necessita de uma quantidade imensa de dados para se poder inferir algum comportamento nos dados. Visando aumentar a estatística e, conseqüentemente, aumentar a probabilidade de se observar um fenômeno desconhecido, o LHC prevê um aumento na luminosidade dos feixes de prótons, nos níveis de radiação e na taxa de eventos.

A luminosidade é proporcional ao número de interações por segundo dividido pela seção de choque do feixe. O aumento da luminosidade significa que o feixe será mais denso e, assim sendo, mais interações próton-próton ocorrerão quando os feixes se cruzarem (colisão) [41]. A luminosidade tem unidade $\text{cm}^{-2} \text{s}^{-1}$ e pode ser calculada através de

$$L \sim \frac{N^2}{t \cdot S}, \quad (2.2)$$

considerando o fator geométrico de redução da luminosidade igual a 1. Nesta expressão, N corresponde ao número de prótons em cada feixe, considerando que cada partícula de um dos feixes pode colidir com qualquer outra do outro feixe. O parâmetro t representa o tempo entre colisões e S , a seção transversal do feixe.

Como consequência do aumento da luminosidade, aumenta a probabilidade da ocorrência de eventos em um mesmo canal dos calorímetros do ATLAS em colisões subseqüentes, ocasionando o efeito de empilhamento de sinais. Este efeito tem impacto direto nos algoritmos de reconstrução de energia, uma vez que estes ainda não prevêem tal efeito em seus modelos. A Figura 2.14 ilustra o empilhamento de um sinal no TileCal. Vale ressaltar que a informação de interesse está presente no pulso centralizado na janela, ou seja, o objetivo do algoritmo de reconstrução do TileCal é estimar com a maior precisão possível a energia do evento centralizado na janela de aquisição.

O LHC prevê três fases de atualização para os seus diferentes sistemas [42], com três longas paradas do colisionador. A primeira iniciou-se no final de 2012 e se estende até o início de 2015. A segunda está prevista para começar em 2019 e tem duração de dois anos e a terceira longa parada acontecerá de 2024 a meados de 2026. Para manter a mesma taxa de eventos (100 kHz) no primeiro nível de seleção de eventos do ATLAS, um novo sistema de seleção *online* deverá ser implantado em que a informação proveniente dos calorímetros deverá ser transmitida de forma digital e com granularidade total.

Ao fim da terceira longa parada do LHC, a eletrônica do TileCal será substi-

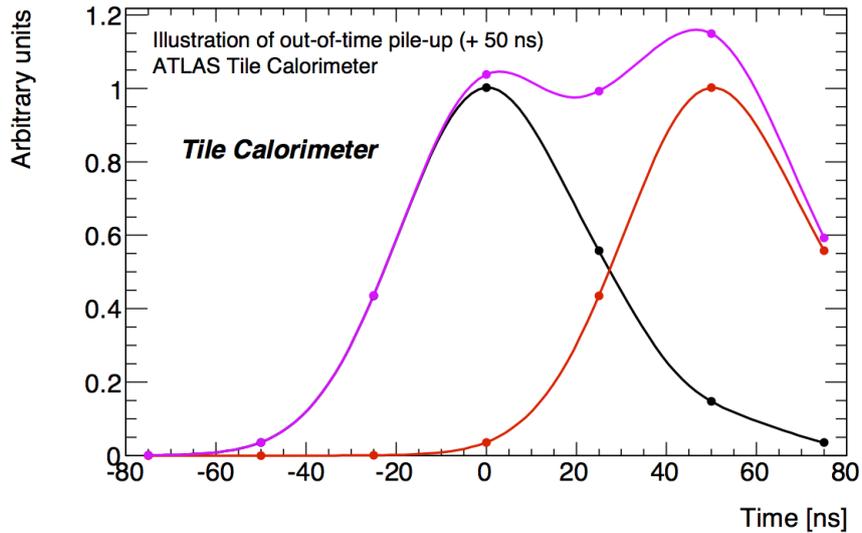


Figura 2.14: Ilustração do efeito de empilhamento de sinais no TileCal. O sinal de interesse está presente na quarta amostra enquanto um outro sinal, de outra colisão (50 ns atrasado) é observado distorcendo o pulso final (extraído de [43]).

tuída por componentes mais robustos à radiação, com menos conectores e de melhor precisão, fornecendo uma informação com maior granularidade e resolução para o primeiro nível de trigger do ATLAS. Um protótipo deste novo sistema, chamado de Demonstrador [44], está em desenvolvimento e foi instalado, em 2015, em um módulo de testes do TileCal para avaliação das opções da nova eletrônica que estão sendo propostas. Adicionalmente, a eletrônica para o processamento digital, atualmente baseada em dispositivos DSP, será substituída por componentes mais modernos, baseado em FPGA (do inglês, *Field-Programmable Gate Array*) [45]. Em tais dispositivos, algoritmos mais sofisticados de reconstrução de energia poderão ser, eventualmente, implementados.

Capítulo 3

Reconstrução de Energia em Calorímetros para Física de Altas Energias

O problema da reconstrução da energia em calorímetros é comumente abordado como um problema de estimação de parâmetros de um sinal recebido. Desta forma, neste capítulo, são descritas algumas técnicas de estimação de parâmetros de sinais, para em seguida, apresentar alguns métodos tipicamente utilizados em calorímetros modernos.

3.1 Estimação de Parâmetros

A maioria dos processos físicos pode ser representada na forma de algum modelo matemático. O uso de modelos nos permite analisar o comportamento de um sistema sob diversas condições. Tais modelos podem ser classificados como modelos determinísticos ou estocásticos. Os modelos determinísticos são mais simples de representar e analisar. Por outro lado, tais modelos podem não fornecer informações suficientes para representar o comportamento completo do sistema. Por outro lado, os modelos estocásticos, os quais envolvem as densidades de probabilidade dos processos envolvidos, podem revelar comportamentos e propriedades importantes do sistema, sendo mais complexos que os modelos determinísticos.

A teoria de estimação é utilizada em muitas aplicações na engenharia, como: radar, sonar, áudio, vídeo, biomedicina, telecomunicações, controle e instrumentação. Estes sistemas compartilham um problema em comum: a necessidade de estimar um, ou vários parâmetros, de um processo aleatório muitas vezes corrompido por ruído, tipicamente aditivo. Em alguns casos, o homem é capaz de identificar, interpretar e discernir a informação desejada em meio ao ruído, mas na maioria das

vezes o problema vai além da capacidade humana e necessita de sistemas capazes de realizar a mesma tarefa [46].

Para formular o problema da estimação de parâmetros em sinais, considere o sinal $x(t)$, que representa a saída de um determinado sistema, como a combinação do sinal $g(t; \boldsymbol{\theta})$, onde $\boldsymbol{\theta}$ é um vetor de parâmetros a ser estimado e um ruído aditivo $n(t)$. Desta forma

$$x(t) = g(t; \boldsymbol{\theta}) + n(t). \quad (3.1)$$

Com o crescente aumento da utilização de sistemas micro-processados e da utilização de processamento digital de sinais, é interessante passar a formulação do problema para o tempo discreto, aonde

$$x[k] = g[k; \boldsymbol{\theta}] + n[k], \quad (3.2)$$

em que $x[k]$ é uma sequência com N valores $\mathbf{x} = \{x[0], \dots, x[N - 1]\}$ dependentes dos parâmetros desconhecidos $\boldsymbol{\theta}$. Para determinar $\boldsymbol{\theta}$ através do conjunto de dados observados, pode-se definir um estimador através da função $f(x)$, tal que

$$\hat{\boldsymbol{\theta}} = f(x[0], \dots, x[N - 1]). \quad (3.3)$$

Devido ao caráter aleatório dos dados, para se obterem bons estimadores, o primeiro passo é ter um bom modelo matemático dos dados. Portanto, pode-se descrever a sequência aleatória por sua função densidade de probabilidade (PDF, do inglês *Probability Density Function*). Esta função é parametrizada por $\boldsymbol{\theta}$, ou seja, temos um conjunto de funções onde uma é diferente da outra, devido aos diferentes valores de $\boldsymbol{\theta}$.

O desempenho obtido pelo estimador será dependente dos pressupostos escolhidos para a PDF. Esse tipo de estimação que se baseia na PDF é denominada de estimação clássica, na qual os parâmetros de interesse assumem-se determinísticos mas desconhecidos [46].

Se, *a priori*, conhecemos alguma característica a respeito do parâmetro a ser estimado, como por exemplo o alcance deste parâmetro, podemos incorporar este conhecimento na solução do problema. Para atribuir este conhecimento na resolução do problema, podemos assumir que $\boldsymbol{\theta}$ não é mais determinístico, mas sim uma variável aleatória e atribuir a ela uma PDF dentro do intervalo de alcance conhecido. Sendo assim, todo estimador produzirá valores dentro deste intervalo. O parâmetro que estamos tentando estimar será então visto como uma realização da variável $\boldsymbol{\theta}$ [46],[47].

Porém, estimadores baseados em PDF muitas vezes não possuem solução analítica e quando possuem ou são aproximados, normalmente levam a estimadores

não-lineares, o que torna a implementação do mesmo muito complexa.

Em aplicações aonde existem restrições na complexidade computacional, operando em tempo real ou *online*, pode ser interessante forçar o projeto de estimadores lineares, pois são de fácil implementação, porém geralmente não-ótimos.

Os estimadores lineares têm sido bastante utilizados para o problema de estimação de energia em calorimetria [48], [49], [50] devido, principalmente, à sua simplicidade de implementação.

3.2 Tendência e Variância de um Estimador

Para que um determinado estimador seja considerado não tendencioso, a média das estimativas ($\hat{\theta}$) deve ser igual ao valor verdadeiro do parâmetro a ser estimado (θ)

$$E\{\hat{\theta}\} = \theta, \quad (3.4)$$

aonde $E\{.\}$ representa o operador de valor esperado. Portanto, um estimador não tendencioso resulta, em média, no valor verdadeiro do parâmetro a ser estimado, garantindo a exatidão da estimativa em termos médios.

Uma outra característica importante a se avaliar em um estimador é a sua precisão. Para isto, pode-se utilizar a variância do estimador, conforme

$$\text{var}(\hat{\theta}) = E\{(\theta - \hat{\theta})^2\}. \quad (3.5)$$

3.3 BLUE - *Best Linear Unbiased Estimator*

Partindo-se de um caso particular, onde se deseja estimar um único parâmetro θ , o BLUE (do inglês, *Best Linear Unbiased Estimator*) restringe o estimador a ser uma função linear, resultando em

$$\hat{\theta} = \sum_{k=0}^{N-1} w[k]x[k], \quad (3.6)$$

onde $w[k]$ são valores fixos, também chamadas de pesos, a serem determinadas. Pode-se gerar um grande número de diferentes estimadores de θ pois este depende diretamente dos valores escolhidos para $w[k]$. Entretanto, o BLUE visa a obtenção dos pesos ótimos que resultam na minimização da variância do estimador com a restrição que ele seja não tendencioso.

Para que o estimador seja não tendencioso

$$E\{\hat{\theta}\} = \sum_{k=0}^{N-1} w[k]E\{x[k]\} = \theta \quad (3.7)$$

e a variância do estimador pode ser obtida por

$$\text{var}(\hat{\theta}) = E \left\{ \left(\sum_{k=0}^{N-1} w[k]x[k] - E \left\{ \sum_{k=0}^{N-1} w[k]x[k] \right\} \right)^2 \right\}. \quad (3.8)$$

Para obtenção dos pesos ótimos do estimador, é necessário assumir um modelo para $x[k]$. No contexto de calorimetria, conforme foi visto na Seção 2.2.3, frequentemente a estimação de energia pode ser realizada através da estimação da amplitude de um pulso de forma fixa. Portanto, pode-se modelar $x[k]$ como

$$x[k] = Ag[k] + n[k], \quad (3.9)$$

em que A corresponde à amplitude a ser estimada, $g[k]$ é conhecido e $n[k]$ representa o ruído aditivo com média zero.

Reescrevendo a Equação (3.6) na forma matricial e assumindo $x[k]$ como o modelo apresentado na Equação (3.9) com o ruído possuindo média nula

$$\begin{aligned} \text{var}(\hat{\theta}) &= E\{(\mathbf{w}^T \mathbf{x} - \mathbf{w}^T E\{\mathbf{x}\})^2\} \\ &= E\{(\mathbf{w}^T (\mathbf{x} - E\{\mathbf{x}\}))^2\} \\ &= E\{\mathbf{w}^T (\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^T \mathbf{w}\} \\ &= \mathbf{w}^T \mathbf{C} \mathbf{w}, \end{aligned} \quad (3.10)$$

onde $\mathbf{w} = \{w[0], w[1], w[2], \dots, w[N-1]\}$ corresponde ao vetor de pesos do estimador. O parâmetro \mathbf{C} corresponde à matriz de covariância do ruído, calculada através de

$$C[k, j] = E\{(N[k] - E\{N[k]\})(N[j] - E\{N[j]\})\}, \quad (3.11)$$

em que N corresponde à variável aleatória associada ao processo ruído no instante k . O vetor \mathbf{w} é achado pela minimização da Equação (3.10) sujeita à restrição imposta pela Equação (3.7).

Para obtenção do estimador BLUE, a variância do estimador, dada pela Equação (3.10), deve ser minimizada sujeita à restrição de tendência. Substituindo o modelo da Equação (3.9) na condição de tendência do estimador da Equação (3.7),

tem-se

$$\begin{aligned}\sum_{k=0}^{N-1} w[k] E\{x[k]\} &= A \\ \sum_{k=0}^{N-1} w[k] g[k] A &= A \\ \sum_{k=0}^{N-1} w[k] g[k] &= 1\end{aligned}$$

ou

$$\mathbf{w}^T \mathbf{g} = 1$$

, onde $\mathbf{g} = \{g[0] \ g[1] \ \dots \ g[N-1]\}^T$ é conhecido.

Este problema de otimização pode ser resolvido utilizando o método dos multiplicadores de Lagrange, resultando em [46]

$$\mathbf{w}_{opt} = \frac{\mathbf{g}^T \mathbf{C}^{-1}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}, \quad (3.12)$$

onde \mathbf{w}_{opt} representa o conjunto de pesos ótimos do estimador BLUE. Portanto, a estimativa da amplitude pode ser obtida através da expressão

$$\hat{A}_{BLUE} = \frac{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}. \quad (3.13)$$

Se as hipóteses assumidas na formulação do problema forem atendidas, o estimador BLUE apresentará variância mínima dada por

$$\text{var}(\hat{\theta}) = \frac{1}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}. \quad (3.14)$$

A tendência do estimador pode ser verificada, substituindo a Equação (3.13) na Equação (3.7)

$$E\{\hat{A}_{BLUE}\} = \frac{\mathbf{g}^T \mathbf{C}^{-1} E\{\mathbf{x}\}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}} = \frac{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{A} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}} = A. \quad (3.15)$$

3.4 Estimador de Máxima Verossimilhança (MLE)

O problema de estimação de parâmetros pode ser abordado através da maximização da probabilidade $P(\boldsymbol{\theta}|\mathbf{x})$, ou seja, deseja-se obter o valor de $\hat{\boldsymbol{\theta}}$ que fornece a maior probabilidade condicional $P(\boldsymbol{\theta}|\mathbf{x})$. Como esta probabilidade normalmente não é conhecida, pode-se reescrevê-la em função da densidade de probabilidade con-

dicional do vetor observado dado o vetor de parâmetros, utilizando a regra de Bayes

$$P(\boldsymbol{\theta}|\mathbf{x}) = p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})P(\boldsymbol{\theta}), \quad (3.16)$$

onde $p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$ é a função densidade de probabilidade condicional do processo observado dado o vetor de parâmetros e $P(\boldsymbol{\theta})$ é a probabilidade *a priori* do vetor de parâmetros.

Observando a Equação (3.16), é possível notar que a maximização da probabilidade $P(\boldsymbol{\theta}|\mathbf{x})$ pode ser obtida através da maximização da densidade condicional $p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$, chamada neste caso de função de verossimilhança. Considerando que o processo observado $\mathbf{X} = \{X_1, X_2, X_3, \dots, X_N\}$ corresponde a variáveis aleatórias independentes e igualmente distribuídas (i.i.d.) com densidades de probabilidade $p_{X_k|\boldsymbol{\theta}}(x_k|\boldsymbol{\theta})$, a função de verossimilhança resulta em

$$p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta}) = \prod_{k=1}^N p_{X_k|\boldsymbol{\theta}}(x_k|\boldsymbol{\theta}). \quad (3.17)$$

Com isso, o primeiro passo para obtenção do MLE é conhecer a função densidade de probabilidade das variáveis aleatórias, $p_{X_k|\boldsymbol{\theta}}(x_k|\boldsymbol{\theta})$, dado que este possui um determinado parâmetro a ser estimado $\boldsymbol{\theta}$. Em seguida, se as variáveis forem i.i.d. (do inglês, *independent and identically distributed*), extrai-se a função de verossimilhança através da Equação (3.17). A melhor estimativa de $\boldsymbol{\theta}$ é o valor que maximiza $p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})$. Desta forma, a estimativa da amplitude pode ser encontrada resolvendo a seguinte equação para a variável θ

$$\frac{\partial p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta} = 0. \quad (3.18)$$

Como pode ser observado, para o desenvolvimento do MLE, é necessário o conhecimento *a priori* sobre o processo aleatório que envolve o parâmetro de interesse. Em muitos casos práticos, o problema não apresenta solução analítica ou se torna um procedimento complicado, que resulta em estimadores complexos. Desta forma, muitos projetistas utilizam aproximações para simplificar o projeto de estimadores.

No problema de estimação de amplitude formulado na seção anterior, aonde o sinal $x[k]$ é modelado através da Equação (3.9), se supusermos o caso particular em que as variáveis podem ser modeladas por uma distribuição gaussiana, e o processo multivariável possui matriz de covariância \mathbf{C} , a função densidade de probabilidade conjunta é dada por

$$p_{\mathbf{x}|A}(\mathbf{x}|A) = \frac{1}{\sqrt{2\pi \det(\mathbf{C})}} \exp\left(\frac{-(\mathbf{x} - A\mathbf{g})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{g})}{2}\right), \quad (3.19)$$

aonde $\det(\cdot)$ representa o determinante da matriz.

A solução pode ser obtida aplicado-se o logaritmo da Equação (3.20), derivando, e igualando a zero

$$\frac{\partial p_{\mathbf{x}|A}(\mathbf{x}|A)}{\partial A} = \frac{-1}{\sqrt{2\pi \det(\mathbf{C})}} \frac{(\mathbf{x} - A\mathbf{g})^T \mathbf{C}^{-1} (\mathbf{x} - A\mathbf{g})}{2} = 0. \quad (3.20)$$

A estimativa \hat{A}_{MLE} é descrita na Equação (3.21). Assim como o método BLUE, este estimador MLE pode ser implementado em processadores digitais através de um filtro FIR (do inglês, *Finite Impulse Response*) de ordem $N-1$ e coeficientes \mathbf{w} .

$$\hat{A}_{MLE} = \frac{\mathbf{x}^T \mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}} = \sum_{k=0}^{N-1} x[k]w[k] \quad (3.21)$$

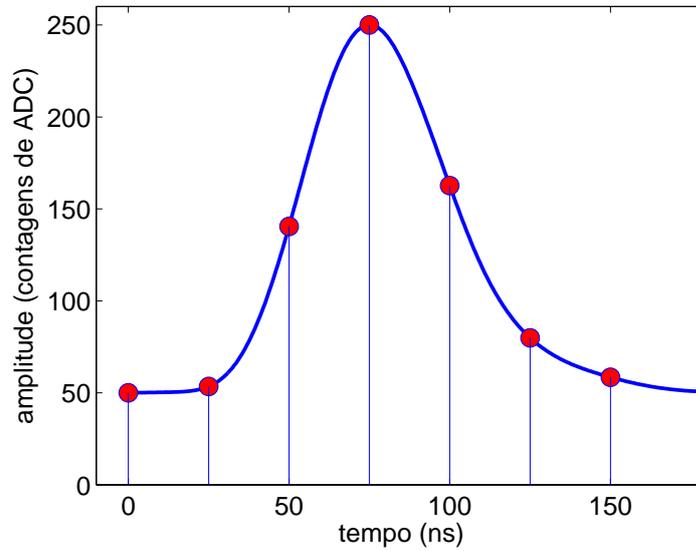
Pode ser observado que, neste caso particular, $\hat{A}_{MLE} = \hat{A}_{BLUE}$ e os estimadores MLE e BLUE apresentam os mesmos conjuntos de pesos.

3.5 Estimadores de energia em calorimetria

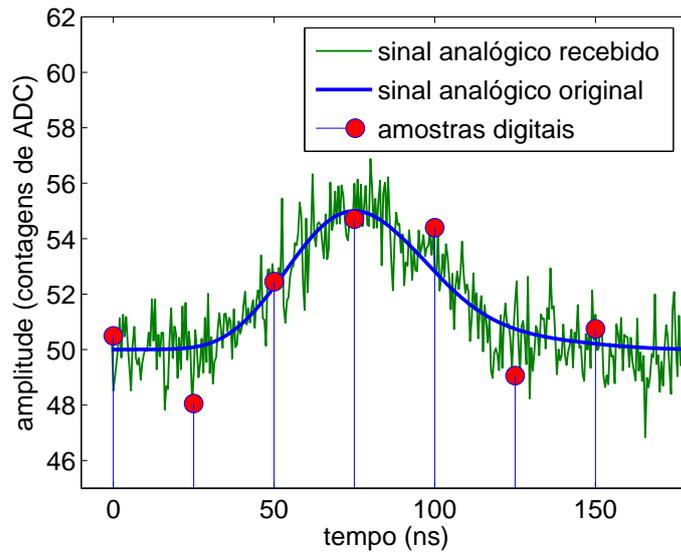
Normalmente, os sinais envolvidos no processo de estimação da energia de calorímetros em experimentos de física de partículas são o pulso de interesse e o ruído de fundo, que é composto principalmente por ruído eletrônico do experimento. Este tipo de ruído é normalmente originado por contribuições de diversas fontes independentes e, segundo o teorema do limite central [51], pode ser aproximado por uma distribuição gaussiana. O uso do modelo gaussiano na descrição do ruído de fundo simplifica o projeto e análise de estimadores, uma vez que a covariância pode ser usada para descrever todo o processo aleatório. Além do mais, se as variáveis da densidade de probabilidade conjunta de um processo aleatório são gaussianas decorrelacionadas, pode-se assumir independência estatística entre as mesmas.

O circuito de condicionamento de calorímetros modernos fornece um pulso fixo e estável. Tipicamente, os pulsos utilizados em calorimetria são unipolares [52],[53],[54],[14], cuja largura pode cobrir diversos períodos de colisão. A Figura 3.1(a) ilustra uma resposta típica encontrada em calorímetros modernos, onde as amostras digitais que são utilizadas no processamento digital são destacadas. O problema de estimação de energia se resume em medir a amplitude deste sinal digital recebido, normalmente corrompido por ruído eletrônico (gaussiano). A Figura 3.1(b) ilustra um sinal de baixa amplitude imerso em ruído e suas amostras digitais, neste caso, espaçadas 25 ns uma das outras. O sinal analógico ideal, proveniente da deposição de energia no calorímetro, gera o sinal recebido (em verde). O objetivo consiste em estimar a amplitude do sinal de interesse (em azul) a partir das amostras digitais

recebidas (em vermelho).



(a)



(b)

Figura 3.1: Resposta típica de um calorímetro de altas energias, em que as amostras digitais estão destacadas para (a) o sinal puro (sem a presença de ruído) e (b) sob condições de baixa SNR (com a presença de ruído).

A maioria dos calorímetros de altas energias utiliza um algoritmo de estimação da amplitude (reconstrução da energia) baseado na soma ponderada (ou um filtro de resposta ao impulso finita) das amostras digitais recebidas dentro de uma janela de leitura de N amostras [48], [49], [50]. Este tipo de abordagem é conveniente para sistemas que necessitam de uma resposta rápida do algoritmo de estimação, como é o caso de experimentos de alta taxa de eventos, unindo simplicidade e desempenho. A decisão de qual canal contém informação relevante é baseada no valor da ampli-

tude estimada. Ou seja, somente os sinais acima de um determinado patamar de amplitude são utilizados durante as análises *offline* de reconstrução do evento.

Em condições de alta luminosidade, onde a taxa de eventos é alta e o feixe de partículas denso, o sinal de interesse ainda pode sofrer a sobreposição de sinais provenientes de colisões subsequentes. Este efeito distorce a forma do pulso de interesse, comprometendo assim a estimação da amplitude. Os métodos tradicionais de reconstrução da amplitude em calorimetria consideram o empilhamento de sinais como uma fonte adicional de ruído, a qual é tratada utilizando a informação da estatística de segunda ordem do ruído. Desta forma, os estimadores utilizam a matriz de covariância do ruído a qual está intrínseca no projeto do filtro digital [46]. Assim, tais estimadores permanecem ótimos para condições em que o ruído pode ser modelado por uma distribuição gaussiana, entretanto, o empilhamento de sinais modifica a característica do ruído, levando estes métodos a condições de operação sub-ótimas.

As próximas seções descrevem os métodos utilizados em alguns calorímetros modernos de altas energias e que servirão de comparação com os métodos propostos nesta tese, nos capítulos 4 e 6.

3.5.1 Estimação de energia no TileCal

O algoritmo chamado de Filtro Ótimo (OF, do inglês, *Optimal Filter*) [55] é o método utilizado nos sistemas de calorimetria do ATLAS para estimação de energia. É uma técnica de minimização de variância que utiliza o conhecimento da forma do pulso (saída do circuito de *shaper*) para estimar a amplitude do sinal e consequentemente a energia [56]. Como o ruído de fundo dos calorímetros do ATLAS é composto principalmente pelo ruído eletrônico, o qual pode ser modelado por uma distribuição gaussiana, o OF é bastante empregado em tais experimentos [57].

A versão do OF implementada no TileCal é chamada de OF2 [58] e foi o algoritmo utilizado até meados de 2014 para estimar a amplitude do sinal, tanto *online* quanto *offline*, e reconstruir a energia do mesmo nos RODs. A amplitude do sinal de entrada é estimada através de uma operação de soma ponderada, de acordo com a Equação (3.22)

$$\hat{A}_{OF} = \sum_{k=0}^{N-1} w[k]x[k], \quad (3.22)$$

onde $w[k]$ são os coeficientes do filtro e $x[k]$ a amostra do sinal recebido no instante k . Os pesos $w[k]$ são obtidos a partir do pulso de referência do canal e da matriz de covariância do ruído neste respectivo canal. O procedimento de otimização tem como objetivo minimizar a variância da distribuição da estimação da amplitude. Assim, o método é ótimo para sinais determinísticos corrompidos por ruído gaussi-

ano. Os pesos são calculados através da minimização da variância do estimador na reconstrução da amplitude, da fase e do pedestal (ver Figura 2.10) [58].

Para calcular os coeficientes do estimador OF2, o sinal recebido do TileCal pode ser modelado como uma aproximação de primeira ordem da sua série de Taylor

$$x[k] = Ag[k] - A\tau\dot{g}[k] + n[k] + ped \quad k = 0, 1, 2, \dots, N - 1, \quad (3.23)$$

onde $x[k]$ representa a amostra digital recebida no instante k e N corresponde ao número de amostras disponíveis (sete no caso do TileCal). A amplitude A é o parâmetro a ser estimado enquanto $n[k]$ representa o ruído de fundo. Os parâmetros $g[k]$ e $\dot{g}[k]$ correspondem, respectivamente, às amostras do pulso de referência do TileCal e sua derivada (aproximação linear para a fase do pulso), enquanto o parâmetro τ é a fase do sinal. A variável ped corresponde ao pedestal do sinal e é um parâmetro constante adicionado ao sinal analógico antes da digitalização.

Para um estimador não tendencioso, é exigido que o valor esperado de \hat{A}_{OF} seja igual a A . Consequentemente, a Equação (3.24) pode ser derivada

$$E\{\hat{A}_{OF}\} = \sum_{k=0}^{N-1} (Aw[k]g[k] - A\tau w[k]\dot{g}[k] + w[k]ped) = A. \quad (3.24)$$

Uma vez que a média do ruído é zero ($E\{n[k]\} = 0$), as seguintes restrições são deduzidas para o estimador se tornar independente do pedestal e da fase

$$\sum_{k=0}^{N-1} w[k]g[k] = 1, \quad (3.25)$$

$$\sum_{k=0}^{N-1} w[k]\dot{g}[k] = 0, \quad (3.26)$$

$$\sum_{k=0}^{N-1} w[k] = 0. \quad (3.27)$$

A variância do estimador é dada por

$$E\{(\hat{A}_{OF} - A)^2\} = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} w[k]w[j]C[k, j] = \mathbf{w}^T \mathbf{C} \mathbf{w}. \quad (3.28)$$

Com isso, para encontrar os pesos do OF2, a Equação (3.28) é minimizada sujeita às restrições expressas pelas Equações (3.25), (3.26) e (3.27). Para solução, o método dos multiplicadores de Lagrange foi utilizado [55], como mostra a Equação (3.29)

$$I_w = \sum_{k,j=0}^{N-1} w[k]w[j]C[k, j] - \lambda \left(\sum_{k=0}^{N-1} w[k]g[k] \right) - \xi \left(\sum_{k=0}^{N-1} w[k]\dot{g}[k] \right) - v \left(\sum_{k=0}^{N-1} w[k] \right). \quad (3.29)$$

Os elementos λ , ξ e v correspondem aos multiplicadores de Lagrange. A etapa seguinte é encontrar o mínimo da Equação (3.29), derivando-a em função de $w[k]$ e igualando-a a zero, como mostra a Equação (3.30)

$$\frac{\partial I_w}{\partial w[k]} = 2 \sum_{k=0}^{N-1} w[k] E\{N[k]N[j]\} - \lambda g[k] - \xi \dot{g}[k] - v = 0. \quad (3.30)$$

Os termos $n[k]$ e $n[j]$ são amostras pertencentes ao ruído, que se considerarmos como um processo estacionário [59], temos $E\{N[k]\} = E\{N[j]\} = 0$ e $E\{N[k]^2\} = E\{N[j]^2\}$ para qualquer k e j . Desta forma, as equações em função de $w[k]$ podem ser representadas pela Equação (3.31)

$$\begin{aligned} \sum_{k=0}^{N-1} w[k]g[k] &= 1 \\ \sum_{k=0}^{N-1} w[k]\dot{g}[k] &= 0 \\ \sum_{k=0}^{N-1} w[k] &= 0 \\ \sum_{j=0}^{N-1} w[j]C[k,j] - \lambda g[k] - \xi \dot{g}[k] - v &= 0, \forall k. \end{aligned} \quad (3.31)$$

Este sistema de equações pode ser escrito na forma matricial representada pela Equação (3.32)

$$\begin{pmatrix} C[1,1] & C[1,2] & \dots & C[1,N] & -g[1] & -\dot{g}[2] & -1 \\ C[2,1] & C[2,2] & \dots & C[2,N] & -g[2] & -\dot{g}[2] & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \\ C[N,1] & C[N,2] & \dots & C[N,N] & -g[N] & -\dot{g}[N] & -1 \\ g[1] & g[2] & \dots & g[N] & 0 & 0 & 0 \\ \dot{g}[1] & \dot{g}[2] & \dots & \dot{g}[N] & 0 & 0 & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w[1] \\ w[2] \\ \vdots \\ w[N] \\ \lambda \\ \xi \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}. \quad (3.32)$$

A solução do sistema (3.32) resulta nos pesos $w[k]$ do método OF. Devido à baixa correlação entre as amostras do sinal e à aproximação das distribuições das mesmas como gaussiana, a matriz \mathbf{C} pode ser aproximada por um matriz identidade. Os pesos $w[k]$ obtidos através da resolução do sistema (3.32) são programados nos DSPs e estimam, de formas *online* e *offline*, a amplitude do sinal.

3.5.2 Estimação de energia no calorímetro eletromagnético do ATLAS

Neste calorímetro, o método OF também é utilizado para estimação de energia, mas sem a restrição expressa pela Equação (3.27), a qual torna a estimação insensível

a variações da linha de base. Neste caso, o pedestal é estimado através de tomadas de dados dedicadas e armazenado em um banco de dados, sendo este valor subtraído de cada amostra digital recebida. A forma do pulso de resposta do LAr é mostrada na Figura 3.2, onde 32 amostras são necessárias para representar todo o pulso. Para a estimação da energia, somente cinco amostras, localizadas ao redor do pico, são utilizadas.

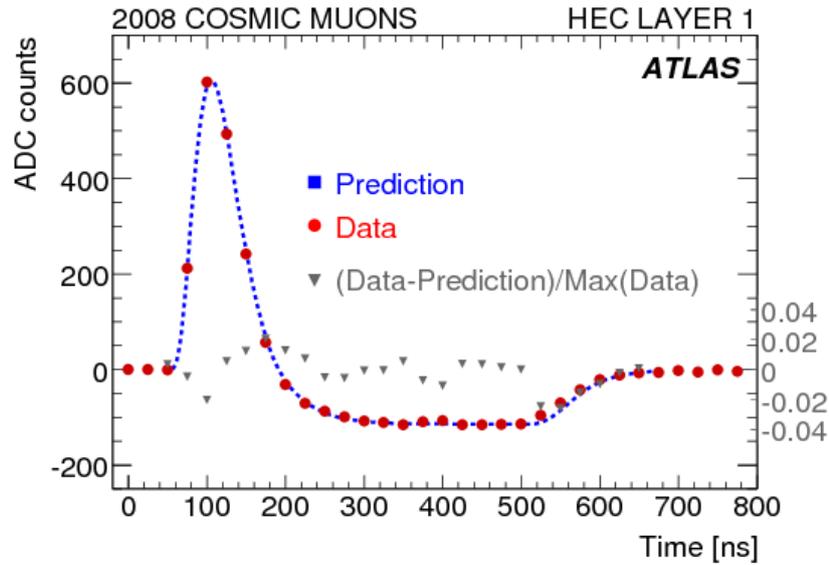


Figura 3.2: Resposta do calorímetro eletromagnético do CMS. As amostras digitais estão destacadas sobre a curva (extraído de [60]).

3.5.3 Estimação de energia no calorímetro eletromagnético do CMS

O calorímetro eletromagnético do CMS, também tem como método base o OF sem a restrição a imunidade ao pedestal, mas este valor é estimado através de amostras adicionais que antecedem o pulso de interesse. Ou seja, a estimação da linha de base dos sinais recebidos é feita evento-a-evento. A Figura 3.3 ilustra o pulso de referência deste calorímetro e as amostras disponíveis para o processamento digital. Pode-se perceber a presença de amostras antes da subida do pulso de interesse, que são utilizadas para a estimação da linha de base do sinal.

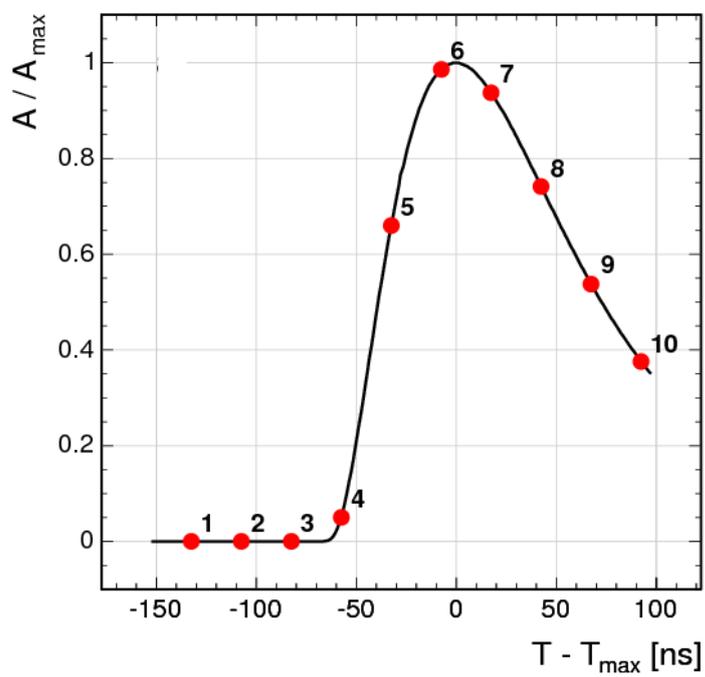


Figura 3.3: Resposta do calorímetro eletromagnético do CMS. As amostras digitais estão destacadas sobre a curva (extraído de [61]).

Capítulo 4

Reconstrução de Energia para Calorimetria Baseada em Filtragem Casada

O objetivo principal desta tese é propor uma nova técnica baseada em filtragem casada para a seleção de sinais e estimação da energia em calorimetria de altas energias. Neste capítulo, o método proposto é apresentado em detalhes, sendo ainda consideradas duas propostas para sua implementação.

Inicialmente, a descrição teórica do método é apresentada, para, em seguida, serem apresentadas duas propostas para implementação. A primeira é para o caso em que o sinal de interesse possui flutuações estatísticas. A segunda proposta é adequada para calorímetros em que o sinal de interesse possui forma constante, resultando no filtro casado.

Em um primeiro momento, o projeto do método proposto considera o sinal original como um processo aleatório gaussiano, e seu desempenho e complexidade são discutidos. Em seguida, uma aproximação determinística é desenvolvida. Por fim, soluções para a implementação são apresentadas, assim como algumas discussões do uso da técnica proposta em condições de empilhamento de sinais.

4.1 Filtragem Casada para Reconstrução de Energia em Calorimetria

Como apresentado no Capítulo 3, em calorímetros modernos, os métodos de reconstrução de energia apóiam-se inicialmente na estimação da amplitude do sinal de interesse para, posteriormente, realizar a seleção de canais através de um corte em energia. Neste contexto, métodos de estimação lineares baseados na minimização da variância do estimador têm sido muito utilizados.

Diferentemente desta abordagem clássica, o método proposto formula o problema a partir da teoria de detecção visando, inicialmente, resolver o problema da seleção ótima dos canais com deposição de energia. A estimação de energia é realizada a partir da saída do algoritmo de detecção, através de calibração.

4.1.1 Detecção do sinal

Tendo em vista a fina segmentação típica dos calorímetros modernos (milhares de canais de leitura), a maioria dos seus canais não possuem informação relevante para ser processada. Desta forma, um algoritmo para a detecção do sinal imerso ao ruído pode ser desenvolvido visando a seleção de canais com deposição de energia relevante nos calorímetros.

O problema de detecção pode ser descrito como um teste de hipóteses sobre o sinal recebido

$$\begin{aligned} H_0 : x[k] &= n[k] & k &= 0, 1, 2, \dots, N - 1 \\ H_1 : x[k] &= g[k] + n[k] & k &= 0, 1, 2, \dots, N - 1, \end{aligned} \quad (4.1)$$

onde H_0 representa a hipótese em que somente ruído foi recebido, enquanto H_1 corresponde à hipótese em que o sinal de interesse $g[k]$ corrompido com o ruído foi recebido. Considerando $\mathbf{x} = \{x[0], x[1], x[2], \dots, x[N - 1]\}$ um dado sinal recebido, e baseando-se na teoria de detecção de Bayes, a relação que maximiza a eficiência de detecção é dada pela razão de verossimilhança [62], como mostrado na Equação (4.2)

$$\Lambda(\mathbf{x}) = \frac{p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)}{p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \quad (4.2)$$

Os termos $p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$ e $p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$ representam, respectivamente, as funções de densidade de probabilidade do sinal recebido (processo \mathbf{X}) para as hipóteses “ruído”, H_0 , e “sinal mais ruído”, H_1 . O parâmetro γ corresponde ao patamar de detecção, que pode ser obtido teoricamente para minimizar a probabilidade de erro, caso sejam conhecidas as probabilidades *a priori*, ou pode ser definido por projeto. Decide-se a favor da hipótese H_1 se a razão de verossimilhança do sinal recebido for maior que o patamar de detecção, e a favor de H_0 caso contrário. Pode-se mostrar que a razão de verossimilhança maximiza a razão sinal-ruído (SNR, do inglês, *Signal-to-Noise Ratio*) [62].

Normalmente, as funções densidade de probabilidade são desconhecidas para a maioria dos problemas de detecção. Com isso, $p_{\mathbf{X}|H_1}(\mathbf{x}|H_1)$ e $p_{\mathbf{X}|H_0}(\mathbf{x}|H_0)$ podem ser estimadas pelo algoritmo de detecção a partir de um conjunto de dados. Entretanto, tal abordagem apresenta-se como uma solução aproximada, não possuindo mais as características ótimas da razão de máxima verossimilhança quando resolvida

analiticamente.

Desta forma, visando a implementação do método proposto em calorimetria, nesta tese, são propostas duas abordagens que serão descritas a seguir.

4.1.1.1 Filtro Casado Estocástico

Um processo estocástico pode ser representado pela sua expansão de Karhunen-Loève. Desta forma, caso o sinal $g[k]$ da Equação (4.1) forme uma sequência estocástica (G), a expansão de Karhunen-Loève pode ser obtida através da Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*), sendo computada através de um conjunto de dados. Assim, o vetor de amostras digitais recebidas pode ser mapeado sobre os autovetores da matriz de covariância de G sem perda de informação, de acordo com a seguinte expressão

$$x_{pca}[k] = \mathbf{x}\phi_{G_k} \quad k = 0, 1, 2, \dots, N - 1, \quad (4.3)$$

em que ϕ_{G_k} corresponde aos autovetores da matriz de covariância do processo G, e $\Phi_G = \{\phi_{G_0}, \phi_{G_1}, \phi_{G_2}, \dots, \phi_{G_{N-1}}\}$. Esta operação resulta em amostras $x_{pca}[k]$ decorrelacionadas. Considerando a sequência aleatória formada pelo sinal $g[k]$ como sendo gaussiana e estatisticamente independente da sequência do ruído (branco gaussiano), após algumas manipulações, o teste da razão de máxima verossimilhança pode ser decomposto em dois termos, I_D e I_R [62]

$$\Lambda(\mathbf{x}) = I_D + I_R, \quad (4.4)$$

$$I_D = \mathbf{m}_G \left[\sum_{k=0}^{N-1} \frac{1}{\lambda_{G_k} + N_0/2} \phi_{G_k} \phi_{G_k}^T \right] \mathbf{x}_{pca} \Phi_G, \quad (4.5)$$

$$I_R = \mathbf{x}_{pca} \Phi_G \left[\sum_{k=0}^{N-1} \frac{\lambda_{G_k}}{\lambda_{G_k} + N_0/2} \phi_{G_k} \phi_{G_k}^T \right] \mathbf{x}_{pca} \Phi_G, \quad (4.6)$$

em que \mathbf{m}_G corresponde ao vetor médio do processo de sinal mapeado sobre os autovetores do processo G, λ_{G_k} é o autovalor associado ao autovetor ϕ_{G_k} , e $N_0/2$ é a variância do ruído.

Nesta tese, esta abordagem é denominada SMF (do inglês, *Stochastic Matched Filter*).

4.1.1.2 Filtro Casado Determinístico

Para o caso em que o ruído de fundo é gaussiano e o sinal de interesse $\mathbf{g} = \{g[0], g[1], g[2], \dots, g[N - 1]\}$ é determinístico, a estimativa da matriz de covariância \mathbf{C} do ruído de fundo é suficiente para caracterizar os sinais. Com isso, a

distribuição gaussiana multivariada é utilizada para descrever as funções de verossimilhança tanto para a hipótese H_0 quanto para H_1 . Como resultado, as funções $p_{\mathbf{x}|H_1}(\mathbf{x}|H_1)$ e $p_{\mathbf{x}|H_0}(\mathbf{x}|H_0)$ são definidas pelas Equações (4.7) and (4.8), respectivamente

$$p_{\mathbf{x}|H_1}(\mathbf{x}|H_1) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{g})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{g})\right), \quad (4.7)$$

$$p_{\mathbf{x}|H_0}(\mathbf{x}|H_0) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1}\mathbf{x}\right). \quad (4.8)$$

Desta maneira, a razão de verossimilhança se torna

$$\Lambda(\mathbf{x}) = \frac{\exp\left(-\frac{(\mathbf{x}-\mathbf{g})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{g})}{2}\right)}{\exp\left(-\frac{\mathbf{x}^T \mathbf{C}^{-1}\mathbf{x}}{2}\right)} \underset{H_0}{\overset{H_1}{\geq}} \gamma. \quad (4.9)$$

Como pode ser visto, a única condição necessária para decidir entre as duas hipóteses é dada pela seguinte expressão

$$\mathbf{x}^T \mathbf{C}^{-1} \mathbf{g} \underset{H_0}{\overset{H_1}{\geq}} \gamma'. \quad (4.10)$$

Este resultado é conhecido como Filtro Casado, ou filtro correlator, para detecção de sinais determinísticos corrompidos por ruído gaussiano de média nula. Para ruído gaussiano e branco, em que as amostras estão descorrelacionadas umas das outras (e conseqüentemente independentes), a matriz \mathbf{C} é diagonal e pode ser desprezada. Deste modo, os coeficientes do Filtro Casado correspondem ao próprio sinal determinístico de interesse \mathbf{g} . Para o caso em que o ruído de fundo é gaussiano e correlacionado, a inversa da matriz de covariância do ruído tem a função de descorrelacionar as amostras do sinal. Neste caso, os coeficientes do Filtro Casado correspondem ao vetor $\mathbf{C}^{-1}\mathbf{g}$.

Em muitos calorímetros modernos, a saída do circuito de conformação (*shaper*) da eletrônica de leitura fornece um pulso analógico fixo e estável cuja energia é calculada pela estimativa da amplitude. Aproveitando-se desta característica do sinal, o problema de detecção pode ser aproximado como determinístico, o que simplifica a implementação do algoritmo de detecção, resultando no Filtro Casado.

O sinal recebido pode ser descrito pelos componentes mostrados na Equação (4.11), em que $x[k]$ é o sinal recebido, ped é a linha de base, $n[k]$ é o ruído de fundo, A é a amplitude e $g[k]$ representa as amostras da resposta (pulso) de referência de um dado calorímetro numa dada amostra k

$$x[k] = ped + n[k] + Ag[k]. \quad (4.11)$$

Antes da operação do Filtro Casado, o valor ped é subtraído de cada amostra do

sinal recebido $x[k]$. Assim, a operação de filtragem casada y_{MF} entre o sinal recebido \mathbf{x} e os coeficientes do Filtro Casado é mostrada na Equação (4.12). O diagrama de blocos do processo de filtragem casada pode ser visto na Figura 4.1.

$$y_{MF} = (\mathbf{x} - ped)^T \mathbf{C}^{-1} \mathbf{g} \quad (4.12)$$

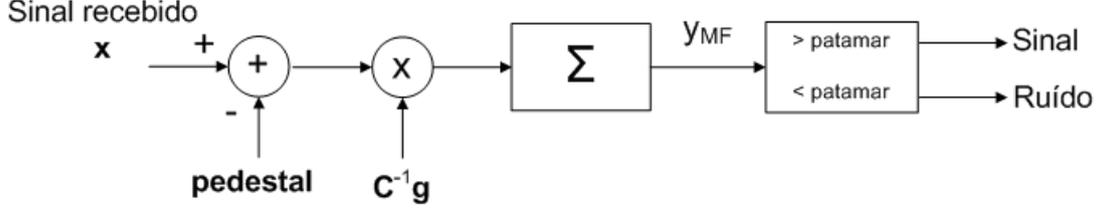


Figura 4.1: Operação de filtragem casada entre o sinal recebido e os coeficientes do Filtro Casado.

Nesta tese, esta abordagem é denominada MF (do inglês, *Matched Filter*).

4.1.2 Estimação da Amplitude

A saída da operação do Filtro Casado, mostrada na Equação (4.4) para a versão estocástica e na Equação (4.12) para a aproximação determinística, pode ser utilizada para selecionar sinais com informação relevante através da aplicação de um simples patamar. Entretanto, este valor não estima a amplitude do pulso que é a informação de interesse para a reconstrução final de um determinado evento.

Filtro Casado Estocástico

A operação de filtragem casada utilizando a abordagem estocástica pode ser representada por

$$y_{SMF} = (\mathbf{B}_2 \mathbf{B}_1 \mathbf{A} \mathbf{g})^T (\mathbf{B}_1 \mathbf{A} \mathbf{g}) + (\mathbf{B}_3 \mathbf{B}_1 \mathbf{A} \mathbf{g})^T, \quad (4.13)$$

em que y_{SMF} corresponde à saída do detector MF estocástico

$$y_{SMF} = (\mathbf{B}_2 \mathbf{B}_1 \mathbf{x})^T (\mathbf{B}_1 \mathbf{x}) + (\mathbf{B}_3 \mathbf{B}_1 \mathbf{x})^T. \quad (4.14)$$

As matrizes \mathbf{B}_1 , \mathbf{B}_2 e \mathbf{B}_3 representam os blocos presentes na Figura 4.2, em que:

$$\mathbf{B}_1 = \mathbf{\Phi}_G \mathbf{\Phi}_G^T,$$

$$\mathbf{B}_2 = \sum_{k=0}^{N-1} \frac{\lambda_{G_k}}{\lambda_{G_k} + N_0/2} \phi_{G_k} \phi_{G_k}^T,$$

$$\mathbf{B}_3 = \mathbf{m} \left[\sum_{k=0}^{N-1} \frac{1}{\lambda_{G_k} + N_0/2} \phi_{G_k} \phi_{G_k}^T \right]. \quad (4.15)$$

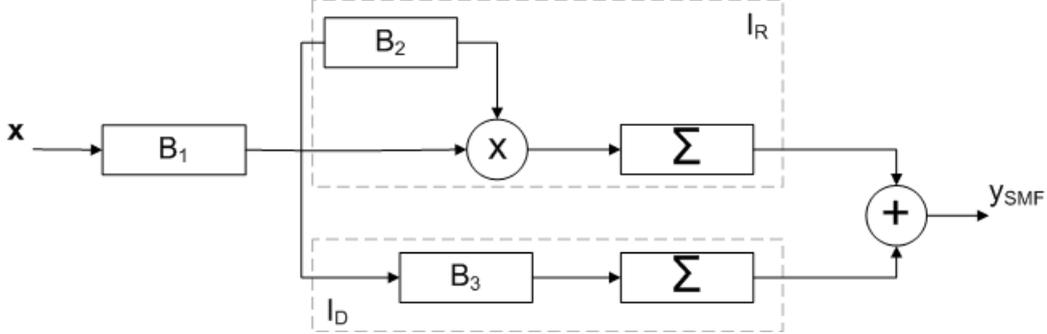


Figura 4.2: Diagrama de blocos do filtro casado estocástico.

A estimativa da amplitude do pulso recebido utilizando a saída do detector SMF, pode ser encontrada igualando-se as equações (4.13) e (4.14), e resolvendo a equação resultante para a variável A

$$\hat{A}_{SMF} = \frac{-\mathbf{B}_1 \mathbf{B}_3 \mathbf{g} + \sqrt{(\mathbf{B}_3 \mathbf{B}_1 \mathbf{g}) + 4(\mathbf{B}_1 \mathbf{B}_1^T \mathbf{B}_2 \mathbf{g} \mathbf{g}^T y_{SMF})}}{2\mathbf{B}_1 \mathbf{B}_1^T \mathbf{B}_2 \mathbf{g} \mathbf{g}^T}. \quad (4.16)$$

Vale ressaltar que somente a parte positiva da função de segunda ordem é utilizada, uma vez que a estimação da amplitude, na aplicação proposta (energia sempre positiva), necessariamente tem que respeitar a igualdade $A \geq 0$.

Filtro Casado

A saída y_{MF} da Equação (4.12) pode ser substituída pela Equação (4.17) e a amplitude do sinal pode ser estimada isolando-se a variável A

$$y_{MF} = (\mathbf{n} + A\mathbf{g})^T \mathbf{C}^{-1} \mathbf{g}. \quad (4.17)$$

Este procedimento resulta na expressão mostrada pela Equação (4.18), a qual é a estimativa da amplitude do pulso

$$\hat{A}_{MF} = \frac{(\mathbf{x} - ped)^T \mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}. \quad (4.18)$$

O numerador $(\mathbf{x} - ped)^T \mathbf{C}^{-1} \mathbf{g}$ corresponde à operação de filtragem casada e o denominador $\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}$ é uma constante que normaliza a saída do Filtro Casado recuperando a amplitude do sinal na unidade correta. A Figura 4.3 mostra o diagrama de blocos da reconstrução do sinal utilizando o Filtro Casado.

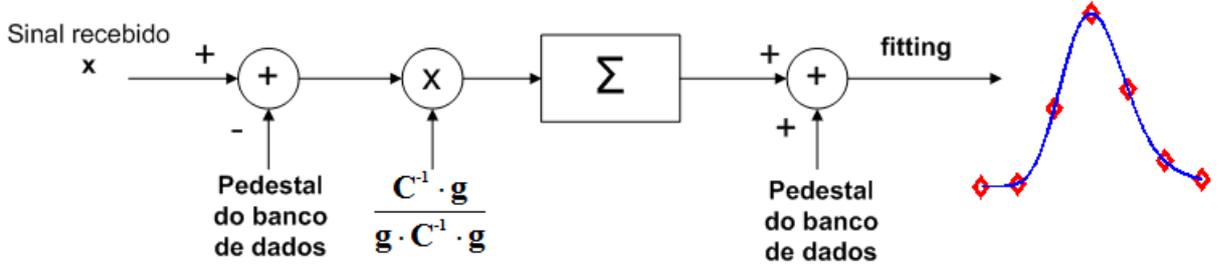


Figura 4.3: Reconstrução do sinal recebido utilizando o Filtro Casado. Os pontos em vermelho na forma de losangos representam as amostras do sinal recebido e a curva, o fitting do pulso de referência nas amostras baseado na amplitude estimada.

4.2 Implementação

A implementação do SMF pode ser simplificada quando são desprezadas componentes de baixa energia na decomposição PCA. Os autovalores da matriz de covariância da sequência estocástica G são ordenados de acordo com suas respectivas variâncias (autovalores), e os autovetores com informação insignificante podem ser descartados, reduzindo a dimensão do problema. Consequentemente, o número de operações necessárias para a discriminação do sinal é reduzido.

O diagrama de blocos da Figura 4.2 descreve uma implementação do filtro casado estocástico, em que \mathbf{B}_1 é uma matriz de ordem $N \times C$ que inclui o pre-processamento por branqueamento e o mapeamento de \mathbf{x} nos autovetores Φ_G . A variável N representa o número de amostras temporais disponíveis e C corresponde ao número de componentes principais utilizado (onde $C \leq N$). O bloco \mathbf{B}_2 é uma matriz de ordem $C \times N$ que compõe a parte aleatória I_R do filtro estocástico, enquanto \mathbf{B}_3 corresponde a parte determinística I_D (matriz de ordem $C \times 1$). Desta forma, o número de multiplicações necessárias na implementação do SMF é $C(N + 2C + 1)$.

A implementação da abordagem determinística, assim como os métodos clássicos de estimação descritos na Seção 3.5, são mostradas na Figura 4.4. Note que esta implementação sugerida corresponde a um filtro de resposta ao impulso finita [63] de N coeficientes, de simples implementação. Assim, somente N multiplicações são necessárias para efetuar a operação de filtragem (estimação da amplitude).

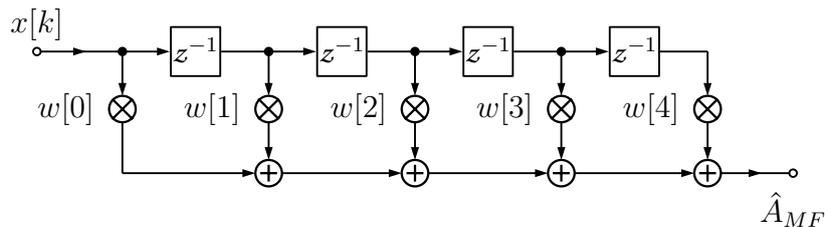


Figura 4.4: Implementação da abordagem determinística do filtro casado para o caso de $N=5$.

Uma observação que deve ser destacada é o fato do projeto do Filtro Casado considerar o ruído de fundo gaussiano. Esta aproximação é válida para calorímetros operando em cenários aonde o efeito do empilhamento de sinais não é observado na maioria dos canais. Entretanto, como o filtro é derivado a partir da razão de verossimilhança, o modelo correto do empilhamento de sinais pode ser incorporado ao projeto. Já para o caso de estimadores clássicos utilizados em calorimetria (ver Seção 3.5), isto não é possível, pois esses são baseados em uma técnica de minimização da variância. Por outro lado, o uso da estatística do empilhamento de sinais pode levar a filtros complexos não-lineares e de difícil implementação.

Uma forma de atacar o problema de empilhamento de sinais utilizando estimadores simples, como o MF determinístico, visando aplicações *online*, é através do uso da matriz de covariância do ruído de fundo no projeto do estimador. Este parâmetro pode ser utilizado com o intuito de reduzir as incertezas introduzidas pelo empilhamento de sinais. O uso desta informação no projeto do detector (e estimador) MF não introduz nenhuma complexidade adicional, podendo ser implementado *online* também através de um filtro de resposta ao impulso finita.

Vale destacar que a estimação ótima de energia sob condições de empilhamento de sinais requer abordagens mais sofisticadas, que modelem corretamente o ruído de fundo ou que apresentem soluções baseadas no conhecimento especialista sobre o fenômeno de sobreposição de sinais. No Capítulo 6, um método alternativo, baseado na desconvolução linear de sinais, é apresentado e sua implementação é discutida.

Capítulo 5

Resultados

Neste capítulo, os métodos propostos para detecção de sinais e estimação da amplitude baseados em filtragem casada são avaliados em simulações e com dados reais. Inicialmente, as análises da aplicação dos métodos propostos em dados simulados são apresentadas utilizando um pulso típico de calorímetros modernos. Em seguida, os métodos são aplicados ao calorímetro de telhas (TileCal) do ATLAS. Dados de Monte Carlo simulados pelo ATLAS [64] e dados reais são utilizados. Uma discussão sobre o desempenho dos métodos propostos aplicados ao Tilecal, em condições de empilhamento de sinais prevista para a segunda tomada de dados (do LHC), também é abordada.

5.1 Simulação do Sinal de um Calorímetro

Os métodos apresentados nesta tese podem ser projetados para qualquer pulso utilizado em calorimetria de altas energias.

Para avaliar o desempenho dos métodos propostos para reconstrução de energia em calorimetria, foi inicialmente considerado um pulso unipolar com duração de 180 ns, já que a maioria dos calorímetros modernos fornece pulsos unipolares quando ocorre deposição de energia em algum dos seus canais [65], [66], [54].

Com o intuito de cobrir os principais algoritmos utilizados em calorimetria, e para fins de comparação, três algoritmos diferentes são considerados. Todos algoritmos são baseados em métodos de minimização da variância, descritos na Seção 3.5. O primeiro, chamado de Método 1, é a técnica padrão de reconstrução da energia no calorímetro eletromagnético do ATLAS e, como descrito na Seção 3.5.2, não utiliza a restrição contra flutuações na linha de base. Ao invés disso, a linha de base é estimada através de tomadas de dados dedicadas para medir este parâmetro. O segundo método, chamado de Método 2, também não utiliza a restrição da linha de base mencionada e, conseqüentemente, os coeficientes são os mesmos calculados para o Método 1. Entretanto, estimativa da linha de base é realizada evento-a-evento,

utilizando-se algumas amostras que antecedem o pulso de interesse. Finalmente, o Método 3 (OF2, apresentado na Seção 3.5.1) utiliza a restrição de linha de base imposta durante o processo de otimização que calcula os coeficientes. Este método foi utilizado no TileCal até 2014.

5.1.1 Conjunto de Dados Básico

Para avaliar o desempenho do método MF para detecção e estimação de energia, uma simulação de sinais foi utilizada. Dois conjuntos de dados foram produzidos, sendo o primeiro contendo somente eventos de ruído e outro contendo sinais corrompidos por ruído gaussiano. Cada conjunto possui 50.000 eventos. Sete amostras digitais foram utilizadas para cobrir todo o pulso, simulando uma taxa de amostragem de 40 MHz, típica do LHC [57], [50]. Para fins de análise de desempenho, três amostras adicionais, que antecedem o pulso, também são consideradas. Entretanto, vale ressaltar que estas amostras não são utilizadas no projeto dos estimadores e detectores, e servem somente para avaliar o desempenho do Método 2, o qual utiliza tal informação para estimação da linha de base do sinal. A Figura 5.1 ilustra o pulso utilizado na simulação assim com suas sete amostras digitais destacadas.

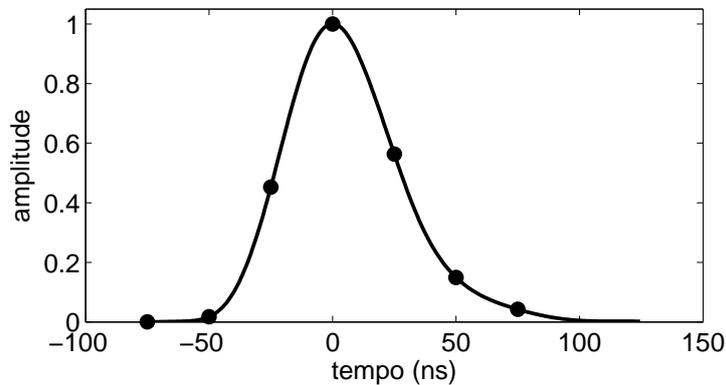


Figura 5.1: Pulso unipolar utilizado nas análises.

As amostras de ruído foram geradas com variância de aproximadamente 30 MeV, comumente encontrada em sistemas modernos de calorimetria [50], [67]. Vale ressaltar que consideramos que uma contagem de ADC corresponde à 12 MeV nas simulações realizadas. Uma distribuição gaussiana foi escolhida para modelar o ruído de fundo (ver Figura 5.2). Com o objetivo de simular condições de baixa relação sinal-ruído, a amplitude dos eventos válidos (sinal de interesse) simulados variam uniformemente, entre 0 e 60 MeV, inclusive. Um desvio de fase, modelado por uma distribuição gaussiana de média zero e desvio padrão de aproximadamente 10 ns, também foi introduzido no sinal de interesse, simulando incertezas da fase na

aquisição. Adicionalmente, uma pequena deformação no pulso também foi introduzida no sinal de interesse, modelada por uma distribuição gaussiana de média zero e desvio padrão de 6%. Este último parâmetro corresponde ao erro introduzido pela tolerância dos componentes eletrônicos de condicionamento do pulso, deformando o mesmo.

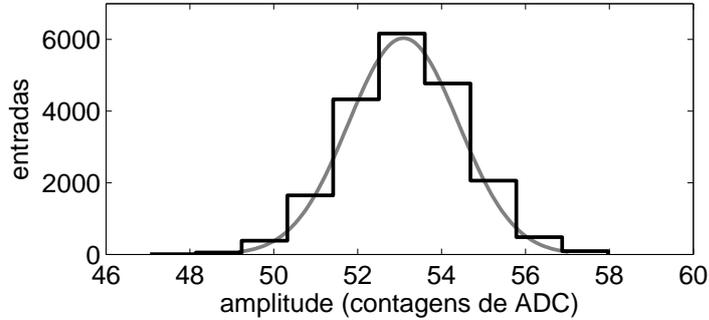


Figura 5.2: Distribuição das amostras do ruído de fundo.

Os conjuntos de sinal e ruído foram divididos, igualmente, em dois, formando o sub-conjunto de desenvolvimento e o sub-conjunto de teste. O sub-conjunto de desenvolvimento de ruído foi utilizado para estimar a matriz de covariância \mathbf{C} (para a transformação de branqueamento) e o valor de linha de base do sinal. A linha de base é estimada extraindo-se a média da primeira amostra digital deste conjunto.

O sub-conjunto de desenvolvimento de sinal não possui a presença do ruído. Este processo aleatório foi utilizado para calcular as componentes principais através da extração de sua matriz de covariância. Os autovalores desta matriz correspondem às constantes λ_G enquanto os autovetores representam as funções de base ortonormais (ϕ_G). Ambos sub-conjuntos de teste de ruído e sinal foram dedicados à avaliação de desempenho dos métodos comparados.

5.1.2 Projeto do detector de sinais por filtragem casada

Os métodos de minimização da variância, descritos nas seções 3.5.1, 3.5.2 e 3.5.3, realizam a detecção do sinal de interesse baseando-se na estimativa da amplitude. Os parâmetros necessários são: o conhecimento da forma do sinal de interesse, a estimativa do pedestal e a matriz de covariância do ruído de fundo. O método MF também faz uso dos mesmos parâmetros para o projeto do detector e estimador.

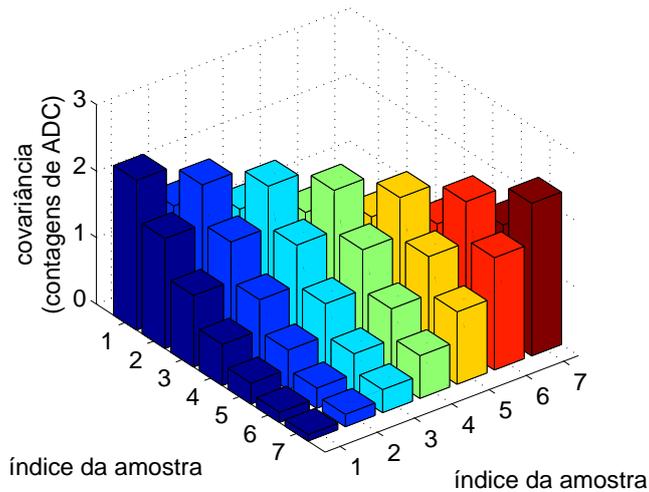
Já para o projeto do método SMF, o primeiro passo é estimar a matriz de branqueamento \mathbf{W} utilizando o sub-conjunto de desenvolvimento de ruído. Este procedimento tem o objetivo de descorrelacionar as amostras de ruído do sinal recebido. Esta matriz de transformação linear possui ordem $N \times N$ e é estimada pela seguinte

equação:

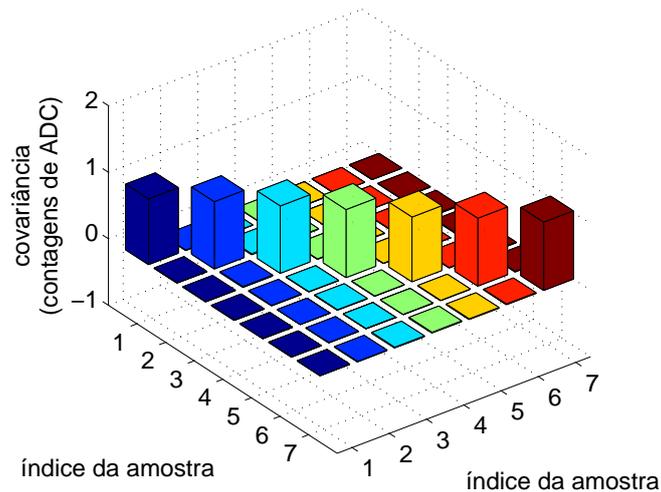
$$\mathbf{W} = \boldsymbol{\lambda}_N^{-1} \boldsymbol{\Phi}_N^T \quad (5.1)$$

em que $\boldsymbol{\lambda}_N$ e $\boldsymbol{\Phi}_N$ são matrizes contendo os autovalores e autovetores da matriz de covariância do sub-conjunto de desenvolvimento de ruído, respectivamente.

Com o objetivo de avaliar a eficiência de decorrelação das amostras, a matriz de branqueamento é aplicada no sub-conjunto de teste de ruído. As figuras 5.3(a) e 5.3(b) mostram a matriz de covariância do sub-conjunto de teste de ruído antes e depois da etapa de branqueamento, respectivamente.



(a) Matriz de covariância do ruído antes da etapa de branqueamento.



(b) Matriz de covariância normalizada do ruído depois da etapa de branqueamento.

Figura 5.3: Eficiência da matriz de branqueamento no sub-conjunto de teste de ruído.

No método SMF, após a etapa branqueamento, deve-se obter as componentes

principais do sub-conjunto de desenvolvimento de sinal, o qual não possui a presença do ruído. A partir dos autovalores e autovetores da matriz de covariância deste processo, é possível analisar a curva de carga e determinar o número de componentes que contabilizam a maior parte da variância do processo, descartando-se assim, as componentes com pouca contribuição [68]. A Figura 5.4 mostra a curva de carga do processo aleatório de sinal.

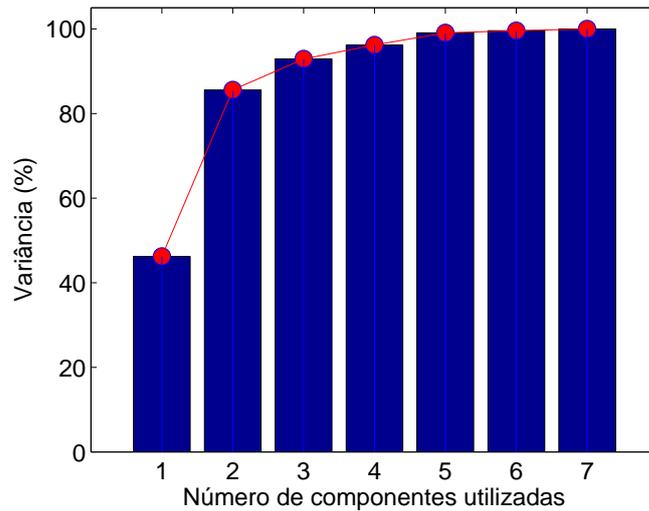


Figura 5.4: Curva de carga acumulada do processo aleatório de sinal.

Note que aproximadamente 94% da variância do processo está contida em somente 3 componentes. A decisão do número de componentes impacta no esforço computacional necessário no cálculo da razão de máxima verossimilhança. Por exemplo, a Tabela 5.1 apresenta um sumário do número de multiplicações necessárias para o SMF em função da escolha do número de componentes.

Tabela 5.1: Número de multiplicações por número de componentes principais (PC) para o sinal com 7 amostras. O número de multiplicações é dado por $C(N + 2C + 1)$, onde C é o número de componentes principais e N é o número de amostras do pulso de referência.

Número de PC	1	2	3	4	5	6	7
Multiplicações	10	24	42	64	90	120	154

5.1.3 Resultados de detecção de sinais

Os eventos dos sub-conjuntos de sinal e ruído foram aplicados aos métodos projetados. Para avaliar o desempenho de detecção em condições de baixa SNR, a curva

ROC (*Receiver Operating Characteristics*) [69] foi utilizada. A análise da curva ROC é apropriada para avaliar a probabilidade de detecção contra a probabilidade de falso alarme (ruído detectado como sinal) de um sistema binário de detecção. A probabilidade de detecção para um dado patamar corresponde ao número de sinais corretamente detectados (acima do patamar) dividido pelo número total de eventos. O falso alarme é calculado como sendo o número de eventos de ruído classificados erroneamente como sinal, dividido pelo número total de eventos de ruído. A Figura 5.5(a) mostra o desempenho de detecção considerando números diferentes de componentes principais no projeto do SMF. Pode-se observar que a partir de 3 componentes o desempenho máximo é atingido, sendo as curvas para os detectores que utilizam 3 ou mais componentes sobrepostas. Já a Figura 5.5(b) mostra a curva ROC para os métodos considerados nesta tese, sendo o SMF projetado com 7 componentes principais.

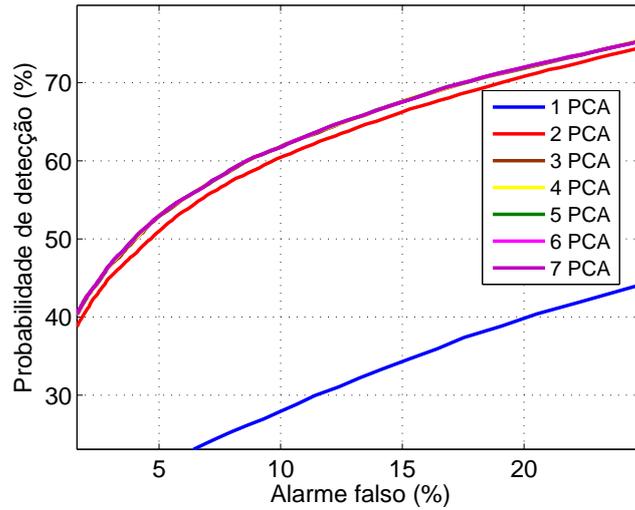
Pode-se observar que o método SMF (utilizando todas as PCs) apresentou o melhor desempenho de detecção. Este resultado pode ser explicado pelo fato do método estocástico absorver a estatística do processo aleatório de sinal, enquanto os outros métodos assumem o sinal de interesse determinístico. A Tabela 5.2 mostra os valores de probabilidade de detecção para 10% de falso alarme.

Tabela 5.2: Probabilidade de detecção (PD) para 10% de falso alarme.

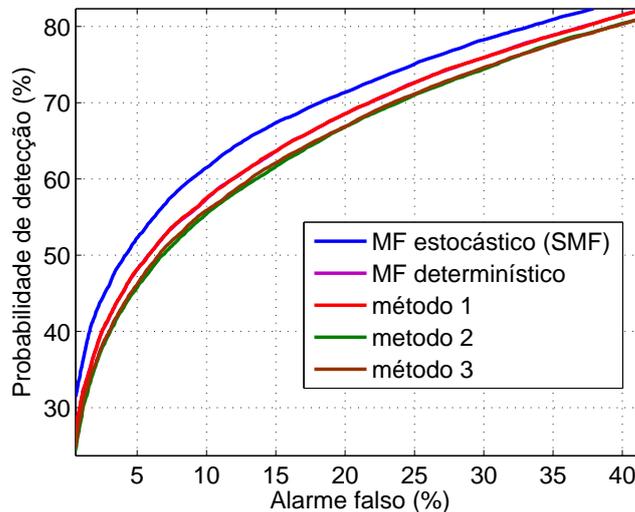
Algoritmo	PD (%)
SMF	61.82
MF	58.44
Método 1	58.39
Método 2	55.16
Método 3	56.94

Os métodos 2 e 3 apresentaram os piores desempenhos de detecção, que pode ser explicado pelo artifício utilizado no tratamento da linha de base do sinal. O Método 3 utiliza uma restrição que o torna imune à flutuações na linha de base, aumentando a variância da medida do parâmetro de interesse (amplitude). Já o Método 2 estima a linha de base do sinal recebido evento-a-evento, através da média entre as três amostras que sucedem o pulso de interesse. Apesar desta ser uma boa estimativa para a linha de base, esta estratégia se mostrou inferior quando comparada ao uso do valor da linha de base estimado por tomadas de dados dedicadas, com milhares de eventos, como utilizado pelo Método 1.

Deve-se salientar que as curvas para o método MF e o Método 1 estão sobrepostas. Embora estes métodos sejam desenvolvidos a partir de teorias diferentes (teste da razão de verossimilhança para o MF e minimização da variância para o Mé-



(a)



(b)

Figura 5.5: Desempenho de detecção em condições de baixa SNR: (a) Curvas ROC para o método SMF considerando números diferentes de componentes principais e (b) curvas ROC para os métodos avaliados.

todo 1), eles fazem exatamente as mesmas aproximações a respeito das estatísticas dos sinais envolvidos. Ambos os métodos utilizam o valor de linha de base calculado *offline*, através do sub-conjunto de desenvolvimento de ruído, e assumem o sinal de interesse como determinístico. Assim, é esperado que eles possuam desempenhos semelhantes.

5.1.4 Resultados de estimação da amplitude

O desempenho na estimação da amplitude é avaliado através das análises de erro de estimação e linearidade. Para tal, um segundo conjunto de dados foi gerado, onde

a amplitude dos sinais variam de 0 a 10 GeV, simulando toda a faixa dinâmica de um conversor analógico-digital de 10 bits, considerando a calibração de uma contagem de ADC representando 12 MeV. Este conjunto de sinais possui 50.000 eventos e, para cada observação, foi adicionado ruído de fundo com as mesmas características descritas na Seção 5.1.1.

A saída dos métodos 1, 2 e 3 já apresentam os valores em amplitude, enquanto os métodos baseados em filtragem casada, utilizam as equações (4.18) e (4.16) para retornar os valores da amplitude para as abordagens determinística e estocástica, respectivamente.

5.1.4.1 Erro de estimação

O erro de estimação é calculado medindo-se o desvio entre os valores de amplitude estimados e os valores de referência (conhecido pelo processo de simulação). Uma maneira simples de realizar tal medida é através da subtração direta entre cada valor de amplitude estimada e o valor de referência associado a tal observação. A Figura 5.6(a) mostra o erro de estimação para os métodos baseado em filtragem casada e os métodos utilizados para comparação. A Tabela 5.3 apresenta um sumário do desvio padrão para cada um dos algoritmos. A Figura 5.6(b) mostra as distribuições do erro de estimação para os métodos propostos SMF e MF, onde o impacto do uso da estocasticidade dos parâmetros do sinal no projeto do estimador pode ser observado, reduzindo o erro de estimação.

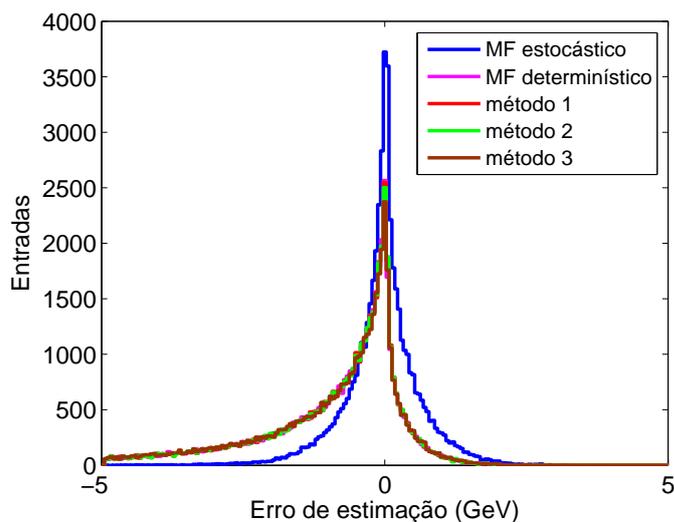
Tabela 5.3: Desvio padrão dos histogramas (GeV)

Algoritmo	Desvio padrão
SMF	0.67
MF	1.90
Método 1	1.90
Método 2	1.90
Método 3	2.00

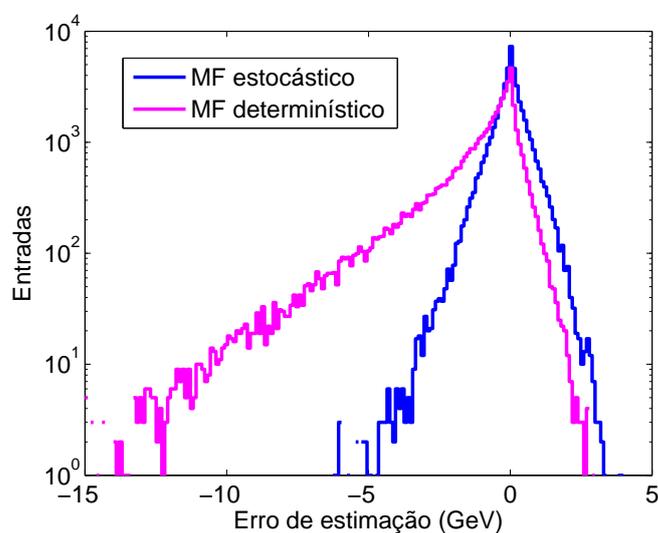
Note que, como esperado, a dispersão do erro de estimação é menor para o método SMF, e ambos os métodos MF e Método 1 atingiram desempenhos parecidos (os histogramas estão sobrepostos).

5.1.4.2 Linearidade

A variância de um estimador linear é independente da amplitude do sinal, como mostrado na Equação (3.28). Em outras palavras, o erro de estimação se mantém constante em toda faixa de operação. Nesta análise, a linearidade é medida



(a)



(b)

Figura 5.6: (a) Distribuições do erro de estimação (número de entradas em escala linear) para cada método considerado e (b) distribuições de energia para os métodos baseados em filtros casado (número de entradas em escala logarítmica).

verificando-se se o erro de estimação é homogêneo ao longo de toda a faixa dinâmica. A Figura 5.7 mostra o erro relativo do método SMF para a faixa de amplitude considerada. Note que a estimação do SMF não apresenta tendências na medida, haja visto que o desvio nas estimativas se mantém constante em relação aos valores de referência.

Do ponto de vista da aplicação dos métodos para reconstrução de energia em calorímetros modernos, o método MF é de simples implementação (filtro FIR), sendo indicado para cenários com restrições na complexidade computacional do método. Já a implementação do método SMF é dificultada por dois fatores: (1) obtenção

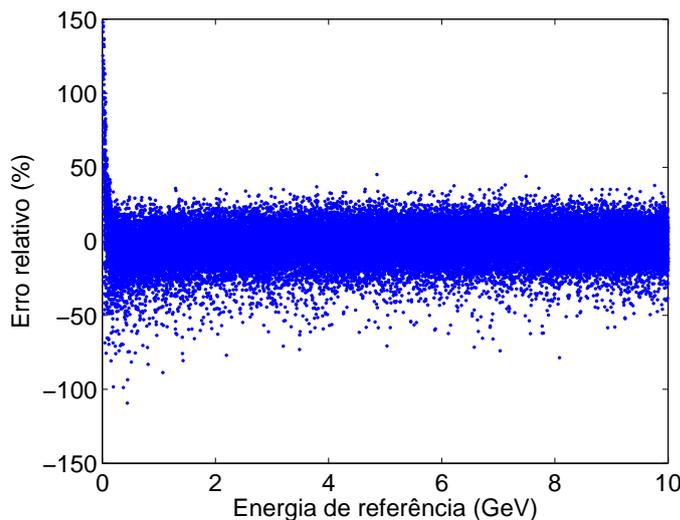


Figura 5.7: Erro relativo do método SMF utilizando sinais simulados de alta amplitude (energia).

da decomposição PCA do processo do sinal; (2) complexidade computacional do método.

A obtenção da decomposição PCA do conjunto de sinal pode ser realizada através de um conjunto de dados com alta relação sinal-ruído, obtido através dos sistemas de calibração dos calorímetros. Já a restrição em relação à maior complexidade computacional pode ser superada através da utilização de dispositivos com maior capacidade de processamento. De qualquer forma, para reconstrução de energia *offline*, a complexidade computacional não se apresenta como restrição. Desta forma, o SMF apresenta-se como boa solução para reconstrução *offline* de energia em calorímetros em que o pulso de referência possua flutuações estatísticas.

5.2 Resultados de simulação para o TileCal

Para avaliar o desempenho dos métodos propostos no ambiente de desenvolvimento do trabalho, dois conjuntos de dados foram utilizados, um contendo somente ruído de fundo adquirido através de tomada de dados de ruído, e outro contendo o sinal de interesse corrompido por ruído aditivo. Cada conjunto contém 50.000 eventos característico da célula A14 localizado na região de $\phi = 4,22$ radianos e $\eta = 1,4$ (ver Figura 2.9). A fim de simular condições de baixa relação sinal-ruído, a amplitude dos sinais variam entre 0 e 5 contagens de ADC (uma contagem de ADC corresponde à 12 MeV), o que corresponde a 3 vezes o desvio padrão do ruído nesta célula. Além disso, outra variável aleatória foi considerada para simular o desvio de fase do pulso, variando uniformemente de -7 a 7 ns. Vale ressaltar que o

ruído utilizado neste conjunto de dados foi adquirido experimentalmente através das tomadas de dados de ruído no TileCal.

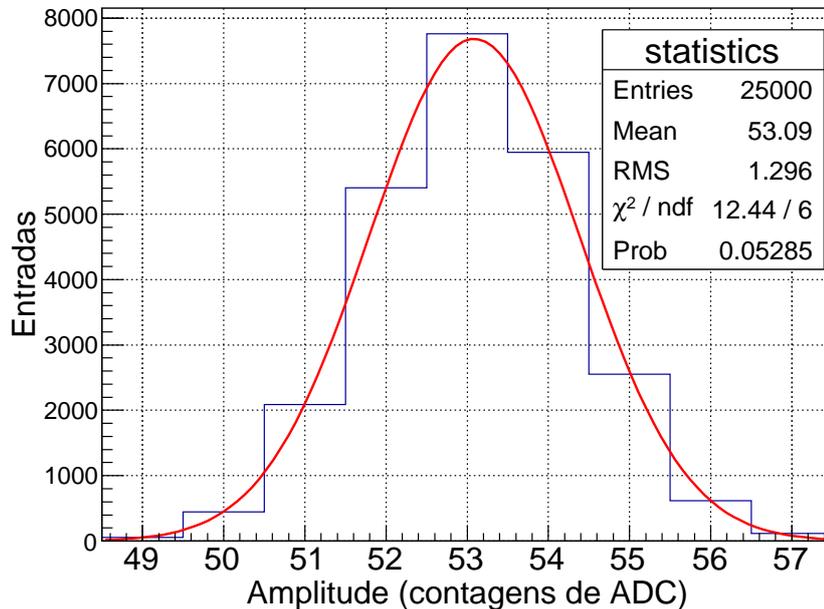


Figura 5.8: Distribuição do ruído experimental do TileCal e o fitting utilizando uma distribuição gaussiana.

O conjunto de ruído foi igualmente dividido em dois, formando os sub-conjuntos de desenvolvimento e teste. O sub-conjunto de desenvolvimento foi utilizado para estimar a matriz de covariância do ruído \mathbf{C} (para ambos MF e OF2) e o pedestal (para o MF apenas). Vale ressaltar que o pedestal foi calculado através da estimativa da média da primeira amostra dos sinais do sub-conjunto de desenvolvimento. A razão pela qual foi escolhido este método de estimação do pedestal é pelo fato da primeira amostra do sinal digitalizado conter somente informação da linha de base do sinal, como pode ser observado na Figura 2.10.

De posse da matriz \mathbf{C} , os coeficientes dos filtros podem ser computados. O valor de pedestal é utilizado somente no processo de estimação da amplitude, quando este valor é subtraído de cada amostra do sinal de entrada. O sub-conjunto de teste e o conjunto de sinal corrompido com ruído foram utilizados para a análise de desempenho dos filtros, que será mostrada nas seções 5.2.3 and 5.2.4.

Devido à estabilidade da forma do pulso do TileCal e às restrições de complexidade computacional para implementação *online*, considerou-se apenas o método MF nas análises para o TileCal. O método OF2, que corresponde ao Método 3, foi utilizado para comparação por ter sido o método utilizado no TileCal até meados de 2014.

Apenas o método MF será avaliado com dados do TileCal, uma vez que a flutu-

ação do pulso do TileCal é pequena, como foi mostrado na Seção 2.3. Desta forma, o método SMF não traz ganhos de desempenho suficientes que justifiquem a sua maior complexidade computacional, quando comparado ao MF.

5.2.1 Descrição do ruído de fundo

Tanto o MF quanto o OF2 utilizam a distribuição e a matriz de covariância do ruído para computarem seus coeficientes. Uma distribuição gaussiana foi utilizada para modelar o histograma do ruído (ver Figura 5.8), concordando com a aproximação assumida no projeto dos estimadores. A matriz de covariância é ilustrada na Figura 5.9.

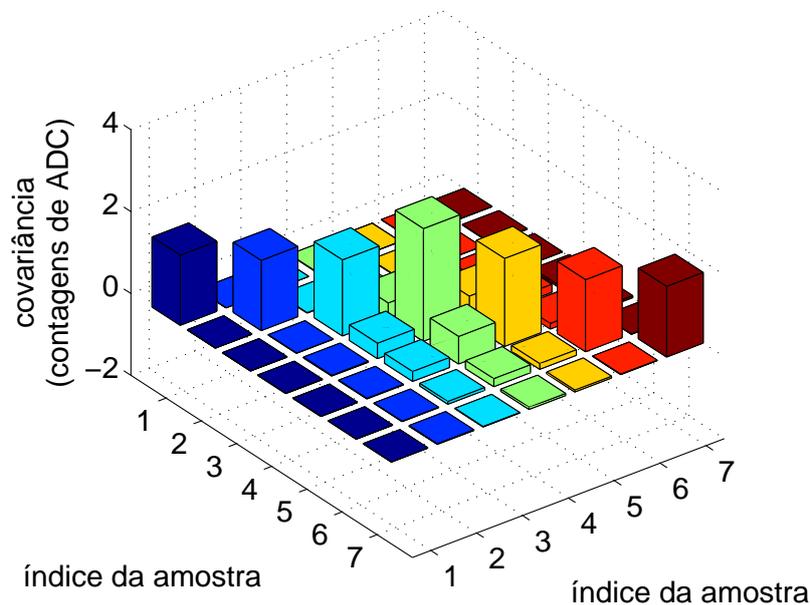


Figura 5.9: Matriz de covariância do ruído do TileCal.

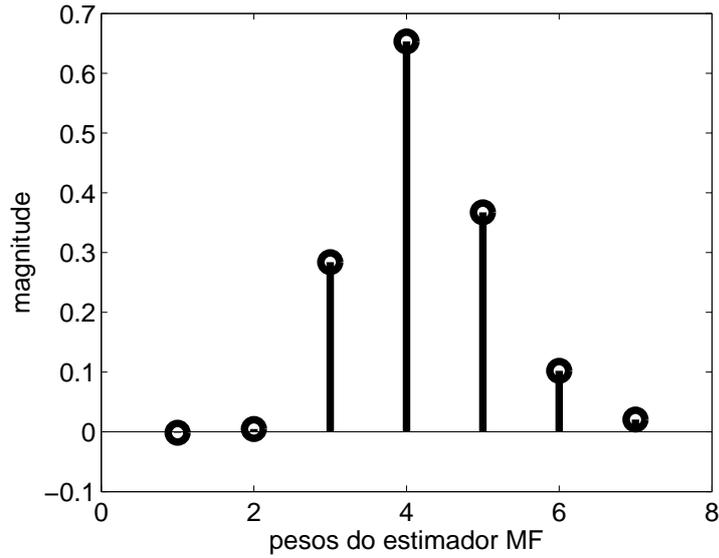
Pode ser observado que as amostras do ruído são correlacionadas e a aproximação gaussiana pode ser aceita. Consequentemente, é esperado que os métodos operem próximo às seus desempenhos ótimos.

5.2.2 Projeto dos filtros e variância

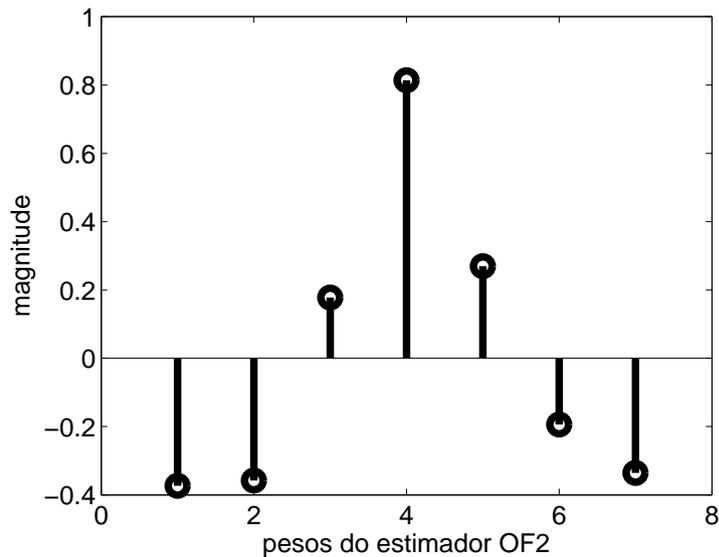
De posse da matriz de covariância do ruído, os pesos dos estimadores MF e OF2 podem ser estimados. As Figuras 5.10(a) e 5.10(b) ilustram os pesos para os estimadores MF e OF2, respectivamente. Pode-se notar que os pesos de ambos os métodos apresentam a forma do pulso de referência com diferentes valores médios. A restrição de insensibilidade à linha de base para o método OF2 faz com que a

soma dos pesos do estimador seja igual a zero. A variância dos estimadores pode ser calculada utilizando a Equação (3.28), sendo 2,2 contagens de ADC para o método MF e 3,1 contagens de ADC para o método OF2.

A Equação (3.28) também mostra que a variância de um estimador linear é independente da amplitude do sinal. Ou seja, ela corresponde ao erro de estimação produzido pelos métodos MF e OF2 independentemente da amplitude do sinal (energia).



(a) Pesos calculados para o estimador MF utilizando a matriz de covariância \mathbf{C} e o pulso de referência \mathbf{g} .

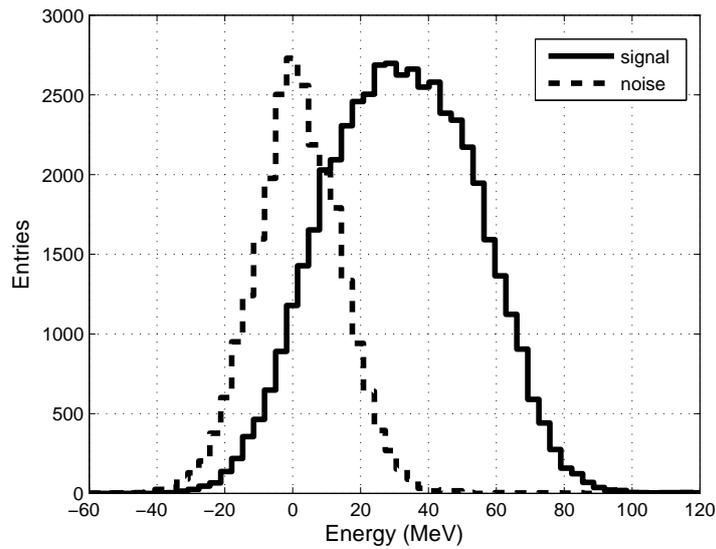


(b) Pesos calculados para o estimador OF2 utilizando a matriz de covariância \mathbf{C} , o pulso de referência \mathbf{g} e sua derivada $\dot{\mathbf{g}}$.

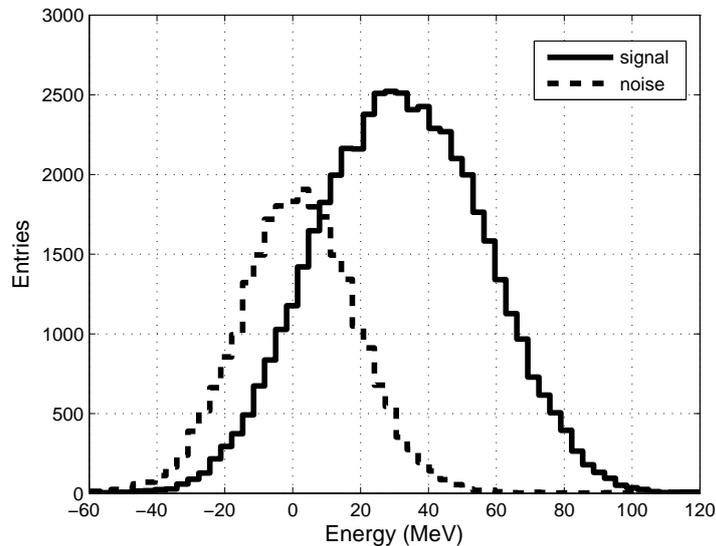
Figura 5.10: Coeficientes dos estimadores (a) MF e (b) OF2.

5.2.3 Detecção de sinal

Para a análise de detecção de sinal, foram utilizados o sub-conjunto de teste para o ruído e o conjunto de sinal corrompido com ruído. As distribuições com as saídas de ambos os conjuntos, utilizando os métodos MF e OF2, podem ser observadas na Figuras 5.11(a) e 5.11(b), respectivamente. Pode-se observar que a região de sobreposição entre as duas distribuições é menor para o MF quando comparado com o OF2, comprovando a melhora na relação sinal-ruído.



(a)



(b)

Figura 5.11: Distribuição de energia para (a) MF e (b) OF2.

Para avaliar o desempenho de detecção, a curva ROC (*Receiver Operating Characteristics*) [69] foi utilizada. A Figura 5.12 mostra a curva ROC para ambos os

métodos. Pode-se observar que o MF supera o método OF2 para qualquer taxa de falso alarme.

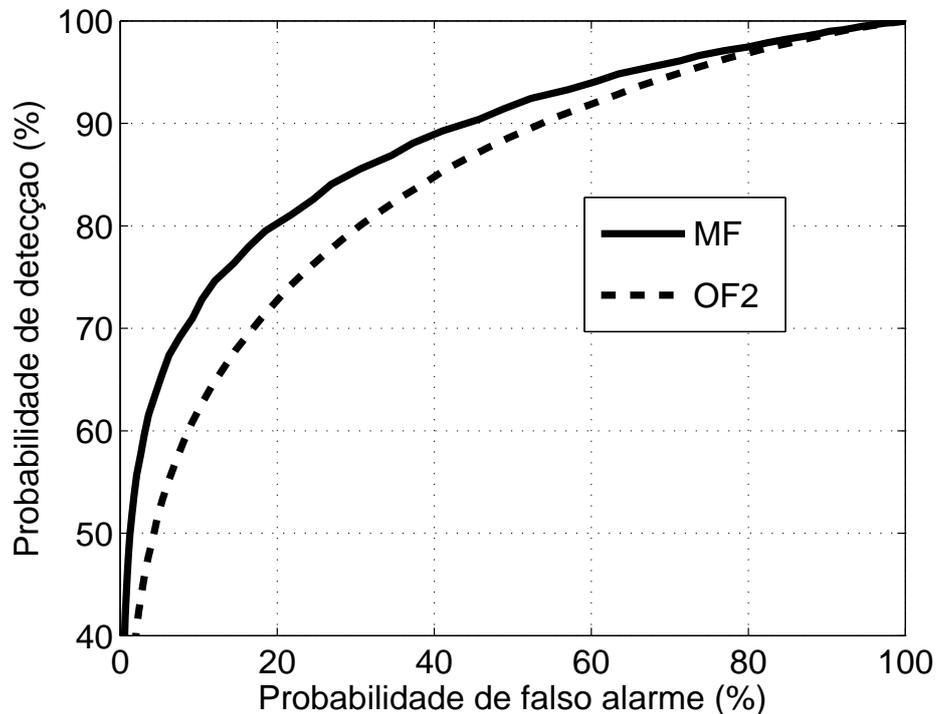
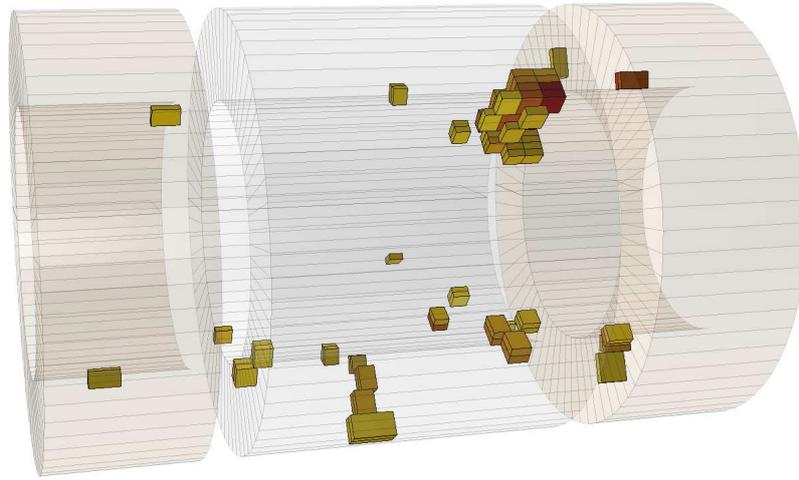


Figura 5.12: Curvas ROC para os métodos MF e OF2.

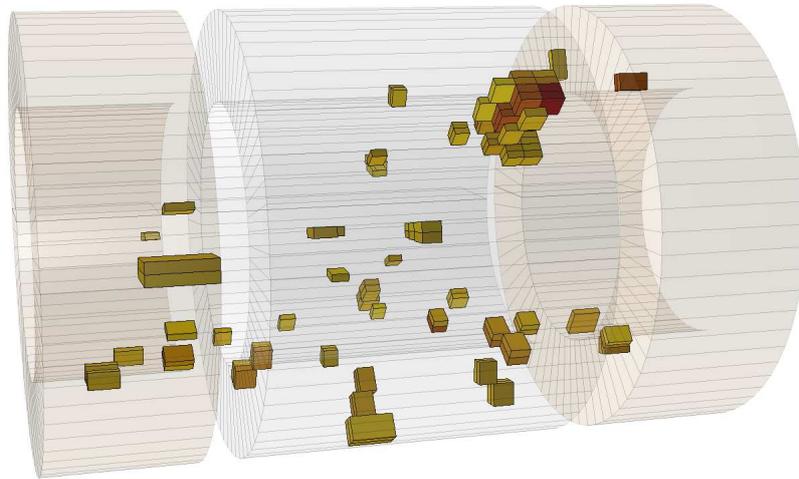
Durante a aquisição (de dados), uma quantidade enorme de dados é coletada e armazenada para ser analisada *offline*. Para uma reconstrução precisa, apenas as células com efetiva deposição de energia devem ser levadas em conta. Além disso, o erro na estimação da energia deve ser o menor possível a fim de evitar o efeito do ruído de fundo e sinais indesejados. Tipicamente, um corte em energia é feito e somente as células com energia estimada acima desde patamar são selecionadas.

Com o propósito de ilustrar o número de células selecionadas pelo MF e OF2 para um determinado patamar de energia, uma ferramenta tridimensional foi utilizada [70], onde quanto mais escura a cor da célula, maior foi a deposição de energia. Um evento aleatório de Monte Carlo foi selecionado e um corte de 300 MeV (valor comumente utilizado nas análises do ATLAS para o TileCal) foi aplicado para selecionar somente células com informação relevante.

Na Figura 5.13, pode ser visto que o método MF apresenta melhor rejeição ao ruído de fundo quando comparado com o OF2, como já mostrado na Figura 5.5(b). Embora ambos os métodos pareçam detectar corretamente um evento (dois jatos representados pelo aglomerado de células), a maioria das células não selecionadas pelo MF mas selecionadas pelo OF2 provavelmente correspondem a ruído de fundo, já que estas células se encontram dispersas pelo calorímetro e não configuram um



(a)



(b)

Figura 5.13: Células selecionadas de um evento utilizando um corte de 300 MeV para as estimativas (a) MF e (b) OF2. Células com cores mais escuras indicam maior deposição de energia.

evento de interesse.

5.2.4 Estimação da amplitude

Na etapa de projeto de um estimador, dois parâmetros importantes devem ser analisados: variância e linearidade.

Uma forma de avaliar o erro de estimação, ou variância do estimador, é subtraindo-se cada estimativa de energia do seu valor de referência associado. A Figura 5.14 mostra o erro de estimação para o MF e o OF2.

Pode-se perceber que o erro de estimação para o MF é menor que para o OF2. O desvio padrão da distribuição do erro para o MF é 17,6 MeV, enquanto para o OF2, 21,0 MeV. Estes valores estão de acordo com os valores calculados na Seção 5.2.2,

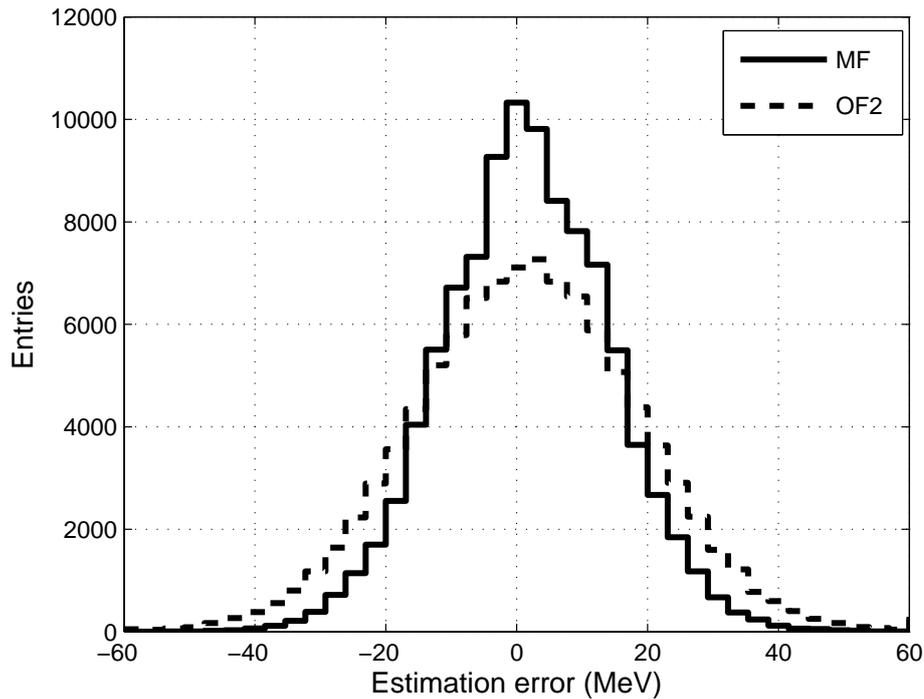


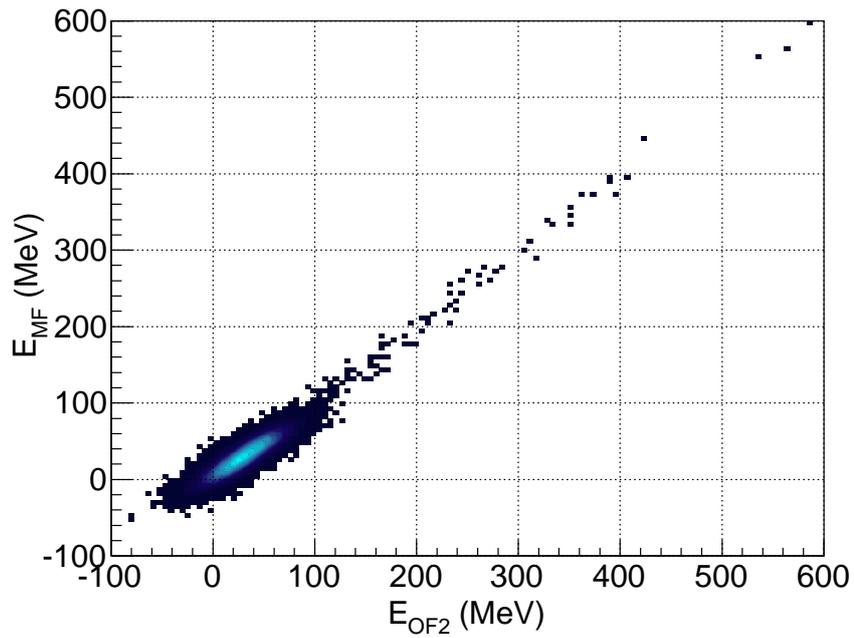
Figura 5.14: Distribuição de erros para o MF ($RMS = 17,6$ MeV) e para o OF2 ($RMS = 21,0$ MeV).

uma vez que o fator de calibração de contagens de ADC para MeV, para este canal, tem seu valor em torno de 12.

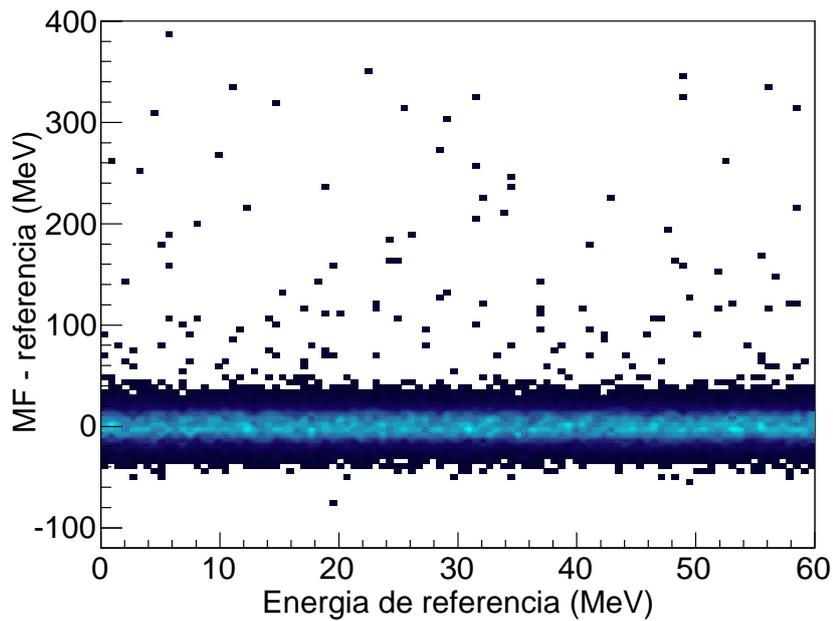
No que tange a linearidade, a Figura 5.15(a) mostra a correlação entre MF e OF2. Alguns eventos de alta energia aparecem devido a partículas espúrias da caverna do detector que são adquiridos na tomada de dados. Pode ser observado que para sinais de altas energias, os métodos se mostram fortemente correlacionados. A Figura 5.15(b) mostra o erro de estimação do MF para sinais de baixa relação sinal-ruído considerados. Pode ser notado que a técnica não é tendenciosa, uma vez que suas estimativas desviam uniformemente dos valores de referência. Os pontos de alto erro de estimação correspondem ao ruído da caverna do detector (partículas espúrias), e podem ser interpretados como empilhamento de sinais devido à radiação da caverna. Estes eventos são raros (aproximadamente 0,3% dos eventos) e de difícil estimação uma vez que não pertencem à nenhuma colisão do LHC.

5.3 Resultados com dados reais do TileCal

Nesta seção, os estimadores MF e OF2 foram aplicados a dados reais adquiridos durante a operação nominal do LHC. Diferentemente das seções anteriores, nesta análise todos os canais do TileCal foram utilizados. Será mostrado primeiramente as



(a) Correlação entre as energias reconstruídas pelos métodos OF2 e MF.



(b) Erro absoluto (MF - referencia) para o método MF.

Figura 5.15: Linearidade do método MF.

análises para dados de baixa luminosidade e depois, o impacto dos filtros para dados de alta luminosidade. Novamente, vale ressaltar que os filtros implementados no pacote oficial do ATLAS foram projetados considerando somente o ruído eletrônico (ver Figura 5.10). Com isso, os métodos não estão otimizados para o cenário de alta luminosidade.

5.3.1 Dados de baixa luminosidade

A Figura 5.16(a) mostra a distribuição de energia reconstruída pelos algoritmos MF e OF2 utilizando dados de colisão adquiridos em 2010 com o LHC operando com energia de colisão no centro de massa $\sqrt{s} = 7$ TeV, 150 ns de espaçamento entre colisões (dT) e média de interações por colisão $\langle \mu \rangle = 3,31$. Sob tais condições o efeito do empilhamento de sinais pode ser desprezado e os métodos operam próximo aos seus desempenhos ótimos. Contudo, a distribuição para o OF2 apresenta maior dispersão quando comparada à distribuição do MF em torno da região de ruído (± 200 MeV). Novamente, esta dispersão pode ser entendida como o erro de estimação (variância) dos métodos. Com o aumento da energia, é observado uma concordância entre os métodos, porém, o erro absoluto se mantém constante, como pode ser visto na Figura 5.16(b), em que a correlação entre os métodos é mostrada.

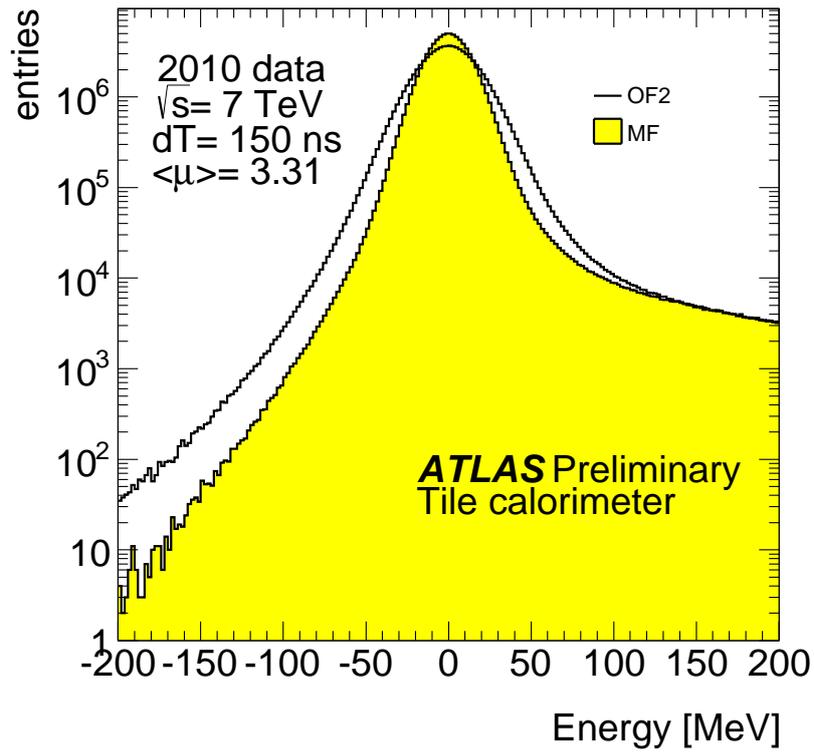
5.3.2 Dados de alta luminosidade

Nesta seção, vale destacar o comportamento do pulso do TileCal no ambiente de alta taxa de colisões do LHC. A Figura 5.17 ilustra as posições dos cruzamentos dos feixes, ou BC, (do inglês, *Bunch Crossing*) dentro da janela de leitura do TileCal. Note que o sinal de interesse tem duração de aproximadamente 150 ns e que o pulso é amostrado na taxa de 40 MHz (mesma taxa de eventos do LHC), resultando em 7 amostras (ou BCs). O sinal de interesse é esperado no BC central, ou seja, BC 0.

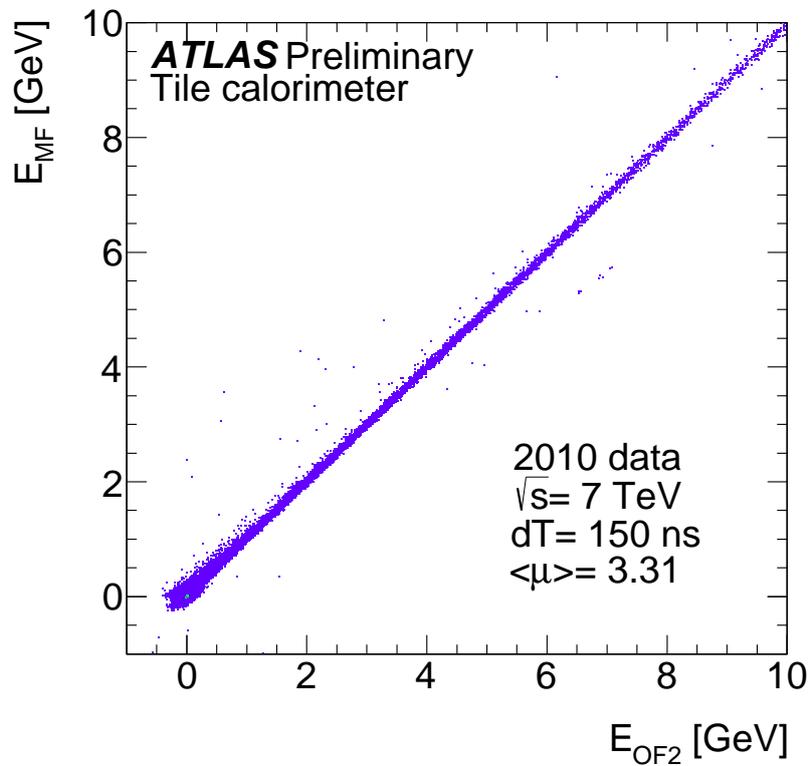
Com o intuito de estudar o comportamento dos estimadores no cenário de alta luminosidade do LHC, foram utilizados dados adquiridos em 2012 durante operação nominal do LHC configurado com energia no centro de massa $\sqrt{s} = 8$ TeV, 25 ns de espaçamento entre colisões (dT) e número médio de interações por colisão $\langle \mu \rangle = 11,3$. Da mesma forma, como foi realizado para os dados de baixa luminosidade, a Figura 5.18(a) mostra as distribuições de energia reconstruídas utilizando os métodos MF e OF2. Sob tais condições, pode-se observar o aumento na dispersão das distribuições de ambos os métodos quando comparado com a análise anterior, para baixa luminosidade. Este aumento no erro é devido à presença do efeito de empilhamento de sinais, que a implementação atual dos métodos não trata corretamente.

No que tange a linearidade dos métodos no cenário de alta luminosidade, a Figura 5.18(b) mostra a correlação entre as energias reconstruídas pelas métodos. Nesta figura, fica evidente a presença do efeito de empilhamento devido à presença de sinais fora de fase (OOT, do inglês, *Out-Of-Time*) como pode ser observado pelas correlações fora da diagonal principal. O algoritmo OF2 reconstrói sistematicamente energias menores que o MF para estes sinais fora de fase.

Em consequência do uso da restrição imposta pelo OF2 para tornar-se imune



(a) Espectro de energia dos estimadores MF e OF2.



(b) Correlação entre as estimativas de energia dos estimadores MF e OF2.

Figura 5.16: Análise para dados reais de baixa luminosidade (extraído de [40]).

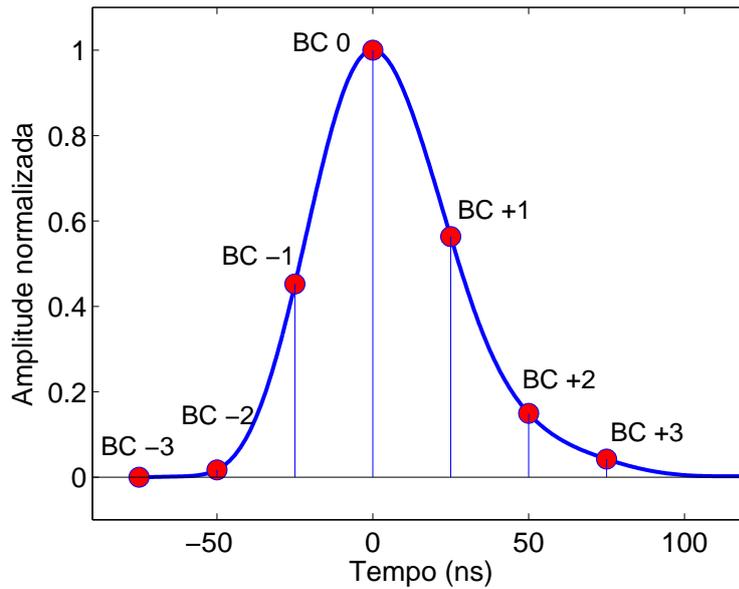


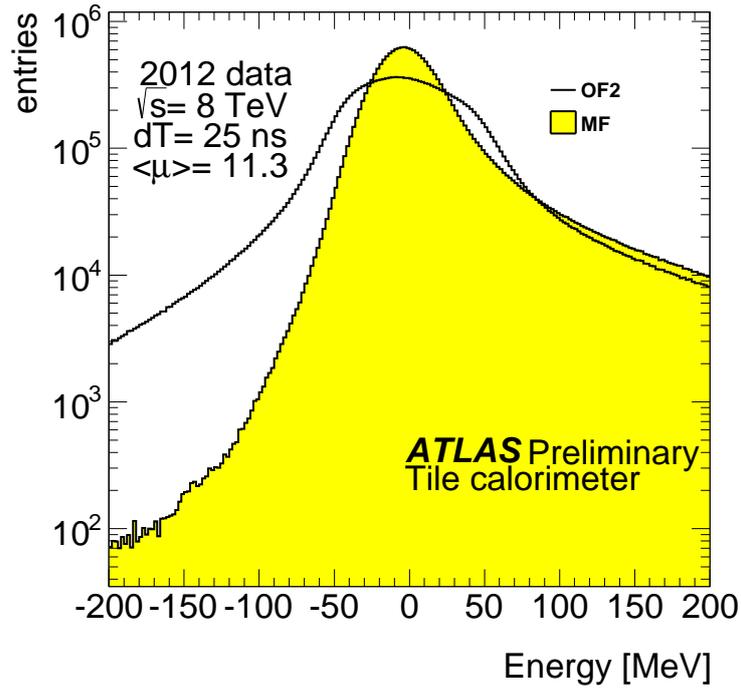
Figura 5.17: Ilustração das posições dos cruzamento de feixes (BC) dentro da janela de leitura do TileCal.

à variação da linha de base do sinal, a qual força a soma dos coeficientes do filtro ser igual a zero (ver Equação (3.27)), alguns de seus coeficientes possuem valores negativos (ver Figura 5.10). Como resultado, a presença de sinais fora de fase localizados nas posições $BC \pm 3$ e $BC \pm 2$ introduz tendência negativa na estimativa final da energia, enquanto sinais fora de fase na posição $BC \pm 1$ introduzem tendência positiva.

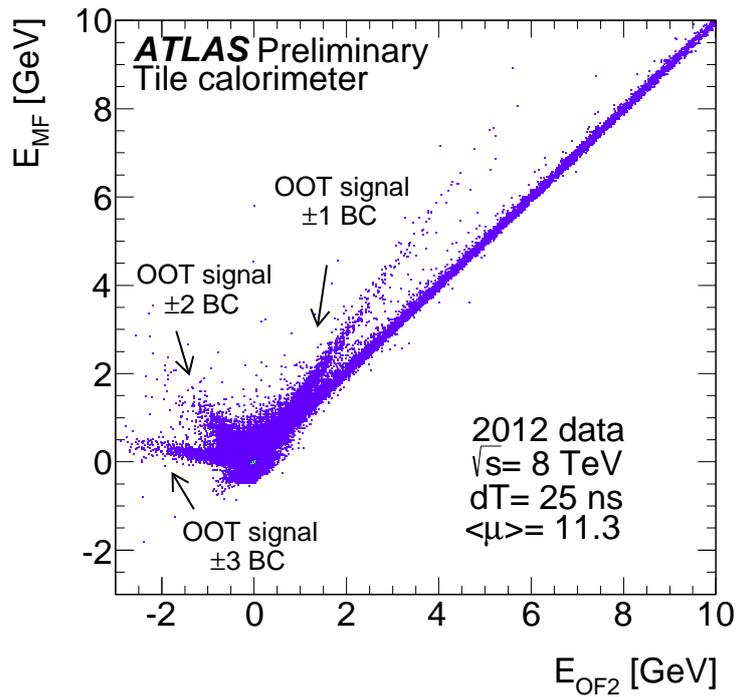
O estimador MF não possui restrição contra variações na linha de base do sinal, e todos os seus coeficientes são positivos. Portanto, em presença de sinais fora de fase, o estimador sempre introduzirá uma polarização positiva na estimativa final. Entretanto, como seus coeficientes associados às posições $BC \pm 3$ e $BC \pm 2$ são menores que os coeficientes do OF2 para as mesmas posições, a polarização introduzida na medida da energia é menor para o MF. Por outro lado, a polarização se torna maior para sinais fora de fase localizados nas posições $BC \pm 1$.

Ambos os métodos ainda podem utilizar a matriz de covariância para reduzir o efeito da polarização na medida introduzido pelos sinais fora de fase. O impacto do uso de tal parâmetro no desempenho dos estimadores é avaliado na seção seguinte. Vale ressaltar que, ainda assim, os métodos não operarão de forma ótima, uma vez que o ruído de fundo neste cenário não pode mais ser modelado por uma distribuição gaussiana.

Além disso, é válido salientar neste momento que o método MF utiliza a razão de verossimilhança para calcular seus coeficientes. Com isso, o método pode ser reprojeto utilizando a descrição correta do ruído de fundo, com contribuições do



(a) Espectro de energia dos estimadores MF e OF2.



(b) Correlação entre as estimativas de energia dos estimadores MF e OF2. Sob condições de alta luminosidade, os métodos se tornam tendenciosos devido à presença de sinais fora de fase (OOT).

Figura 5.18: Análise para dados reais de alta luminosidade (extraído de [40]).

ruído eletrônico e do empilhamento de sinais. Contudo, o uso de uma distribuição não-gaussiana pode conduzir ao projeto de um método não-linear, de difícil imple-

mentação. Já pensando em soluções para a reconstrução de energia em condições de empilhamento de sinais da segunda tomada de dados do LHC, a próxima seção explora o impacto do uso da matriz de covariância no projeto dos estimadores.

5.4 Proposta para o Cenário de Empilhamento de Sinais no TileCal Durante a Segunda Tomada de Dados do LHC

O LHC prevê, em seu programa de atualização, um aumento da luminosidade. Foi mostrado, na Seção 2.5, que este parâmetro é proporcional ao efeito indesejado de empilhamento de sinais. No TileCal, as células de leitura que apresentam maior ocupância (maior probabilidade de incidência de sinais) se localizam na camada A, a qual está mais próxima do feixe (ver Figura 2.9). Além disso, o TileCal possui um conjunto de canais de leitura especiais, chamados de células E ou cintiladores, que se localizam na região entre o barril central e barril estendido, que sofrem ainda mais do efeito de empilhamento de sinais.

A resposta do TileCal (150 ns) é mais lenta que a taxa de eventos do LHC (25 ns) e, conseqüentemente, o sinal da colisão de interesse pode ser deformado pela presença de sinais provenientes de colisões adjacentes, principalmente em células de alta ocupância. Como resultado deste empilhamento de sinais, os métodos que se apoiam na forma do pulso tem seus desempenhos degradados, pois o sinal adicional introduz uma componente não-linear no ruído de fundo, além do ruído eletrônico.

Neste contexto, as próximas seções apresentam formas de reduzir o impacto do efeito de empilhamento de sinais no TileCal previsto pelo LHC durante sua segunda tomada de dados (chamada de Run2). Os métodos propostos são aplicados a dados de simulação completa de Monte Carlo do ATLAS, no cenário previsto para a segunda tomada de dados do LHC.

5.4.1 Reconstrução *online* de energia no TileCal para o Run2 do LHC

Durante o Run1 do LHC, as colisões aconteceram, em sua maioria, numa taxa de 20 MHz (50 ns) e não foi observado alto nível de empilhamento de sinais em grande parte das células do TileCal. Neste cenário, as células do TileCal apresentaram baixa ocupância e o ruído de fundo pôde ser descrito pelo modelo do ruído eletrônico, gaussiano e aproximadamente descorrelacionado (WGN, *White Gaussian Noise*). Assim, durante o Run1, a matriz de covariância do ruído de fundo, utilizada pelo método OF2 (ver Seção 3.5.1) foi descrita por uma matriz identidade.

Já para o Run2, que está atualmente em operação, o feixe é mais denso, e a taxa de colisão foi aumentada para 40 MHz. Portanto, a descrição do ruído de fundo como WGN compromete a estimação da energia em células de maior ocupância. Foi mostrado nas seções 5.2 e 5.3 que o método MF apresenta melhor desempenho do que o método OF2. Desta forma, o método que está atualmente sendo empregado no TileCal para reconstrução *online* de energia é o método MF determinístico. Assim, como o método OF2, o MF também prevê o uso da matriz de covariância em seu projeto como mostrado na Equação (4.18) e repetido a seguir:

$$\hat{A}_{MF} = \frac{(\mathbf{x} - ped)^T \mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}} \quad (5.2)$$

em que o vetor \mathbf{g} corresponde às amostras digitais do pulso de referência, \mathbf{C} é a matriz de covariância do ruído de fundo e \mathbf{x} são as amostras digitais recebidas.

Durante todo o Run1 do LHC, a matriz \mathbf{C} foi aproximada por uma matriz identidade, uma vez que pouco empilhamento de sinais foi observado no TileCal. Entretanto, tal aproximação não é válida em células de alta ocupância, como as localizadas na camada A do barril estendido e nas células E.

Vale ressaltar que, devido às limitações de recursos da eletrônica para o Run2, o TileCal não permite a implementação de métodos mais sofisticados, que demandam um número maior de operações, tais como o SMF. Por isso, os resultados apresentados nas seções seguintes se limitam aos métodos MF e OF2 (utilizado durante o Run1), e mostram o projeto dos estimadores e o impacto do uso da matriz de covariância do ruído de fundo na reconstrução da energia em condições de empilhamento de sinais do Run2.

5.4.2 Projeto dos estimadores

Em ambientes onde o empilhamento de sinais não está presente, o ruído de fundo do TileCal é composto principalmente por ruído eletrônico, normalmente modelado por um processo gaussiano [51]. Sob tais condições, os métodos lineares baseados em minimização da variância e o MF determinístico operam de forma ótima. Já em cenários onde o empilhamento de sinais está presente, o ruído de fundo não pode mais ser representado por uma distribuição gaussiana, e os métodos de estimação da energia operam de forma sub-ótima.

A fim de verificar o modelo do ruído de fundo a partir das amostras digitais adquiridas, as figuras 5.19(a) e 5.19(b) mostram, respectivamente, as distribuições de ruído sem o empilhamento de sinais ($\langle \mu = 0 \rangle$) e no cenário onde o número médio de interações por colisão é 40 ($\langle \mu \rangle = 40$). Para ambos os casos, foram utilizados dados de simulação de MB (*Minimum Bias*) do ATLAS. As tomadas de

dados de MB podem ser entendidas como tomadas de dados para medir o ruído do experimento em condições de empilhamento de sinais, ou seja, nestes dados, o evento de interesse não está presente, mesmo que seja observado um sinal central na janela de leitura.

Com o intuito de avaliar a mudança na descrição do ruído, a Figura 5.19(c) mostra o *fitting* calculado sobre o histograma de $\langle \mu \rangle = 0$ e sobreposto no histograma de $\langle \mu \rangle = 40$. Nesta análise, a célula de maior ocupância da camada A do TileCal (A13) foi utilizada para ilustrar o comportamento do ruído em condições diferentes de empilhamento de sinais. O valor da ocupância é proporcional ao valor de $\langle \mu \rangle$, e no cenário de $\langle \mu \rangle = 40$, a célula A13 do TileCal possui ocupância de aproximadamente 10% [71]. Ou seja, em média um sinal é observado na janela de leitura desta célula a cada 10 colisões.

Como pode ser observado, o empilhamento de sinais introduz uma componente adicional no ruído de fundo, que passa a não possuir mais um comportamento gaussiano. Por outro lado, o uso da matriz de covariância do ruído de fundo pode ser considerado no projeto do MF e OF2 com o objetivo de reduzir as incertezas introduzidas pelo empilhamento de sinais, aumentando o desempenho de reconstrução da energia destes métodos.

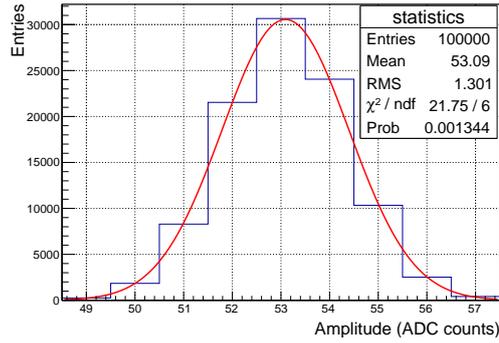
Vale destacar que, no TileCal, a maioria das células possui baixa ocupância, e o empilhamento de sinais pode ser tratado como eventos raros, ou *outliers*. Entretanto, dada a sensibilidade do método comumente utilizado para a obtenção da matriz de covariância a partir de um conjunto de dados na presença de *outliers*, nesta tese são consideradas outras abordagens para estimar tal parâmetro.

O algoritmo clássico de estimação da matriz de covariância [51] utiliza todo o conjunto de dados, e pode ser expresso através da seguinte equação:

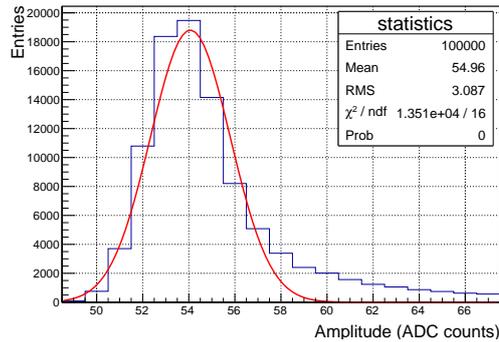
$$\text{cov}(x, y) = \frac{\sum_{i=1}^M (x_i - \bar{x}) \times (y_i - \bar{y})}{M} \quad (5.3)$$

em que \bar{x} e \bar{y} correspondem aos valores médios das variáveis aleatórias X e Y , respectivamente, e M representa o número total de eventos do conjunto de dados considerado. A presença de eventos *outliers* distorce a descrição correta deste parâmetro, resultando no cálculo de coeficientes incorretos e, conseqüentemente, pior desempenho de estimação da energia.

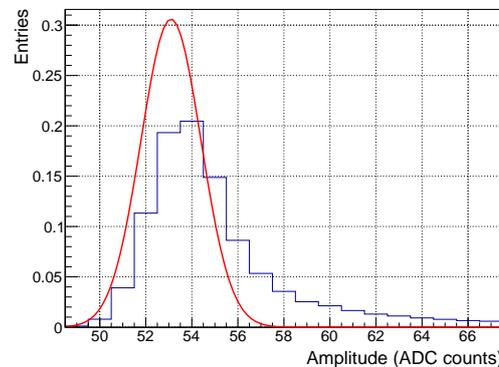
Como alternativa, esta tese utiliza uma técnica mais robusta para a estimação da matriz de covariância, conhecida como Estimador do Mínimo Determinante da Covariância (MCDE, do inglês, *Minimum Covariance Determinant Estimator*) [72]. Tal abordagem seleciona, aleatoriamente, um sub-conjunto e estima a matriz de covariância tradicional (Equação (5.3)) e seu determinante. Este procedimento é repetido diversas vezes (nesta tese foram utilizados aproximadamente 500 iterações)



(a)



(b)



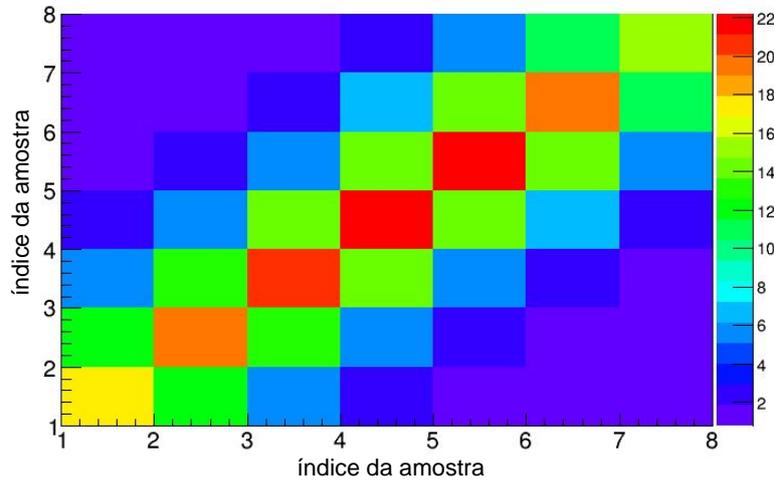
(c)

Figura 5.19: Distribuições das amostras de ruído de fundo para (a) $\langle \mu \rangle = 0$, (b) $\langle \mu \rangle = 40$ e (c) mostra o *fitting* extraído de $\langle \mu \rangle = 0$ e sobreposto no histograma de $\langle \mu \rangle = 40$ para a célula A13 do TileCal.

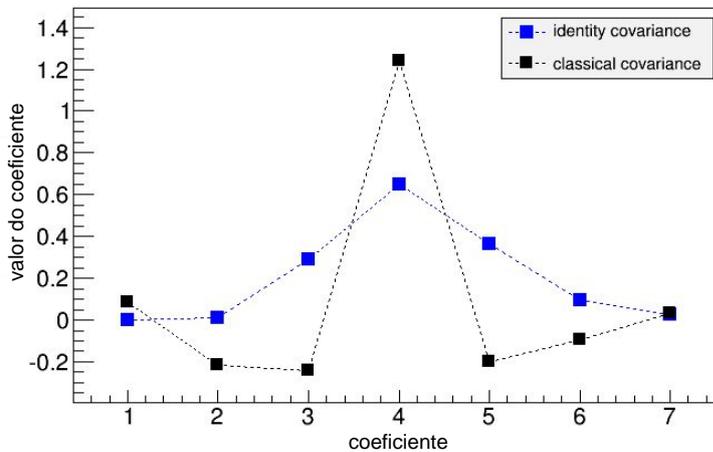
e seleciona a matriz de covariância referente ao menor determinante calculado. O sub-conjunto associado a tal matriz possui os eventos com menor covariância entre as variáveis e descreve a estatística de segunda ordem da maioria dos eventos do conjunto total, uma vez que sinais de alta energia empilhados são descartados.

Para demonstrar e comparar os métodos clássico e MCDE, duas células de baixa e alta ocupância (células A3 e E4, respectivamente) foram utilizadas. O ruído de fundo da célula A3 consiste principalmente de ruído eletrônico, enquanto na célula

E4, o empilhamento de sinais é observado com maior frequência. As figuras 5.20 e 5.21 mostram, respectivamente, a matriz de covariância estimada utilizando o método tradicional e o método robusto (MCDE) para a célula A3. Para efeito de comparação, os coeficientes do método MF determinístico também são mostrados.



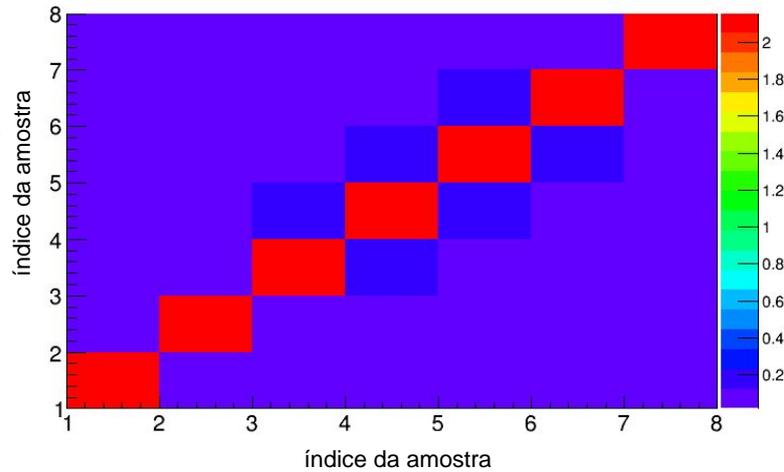
(a)



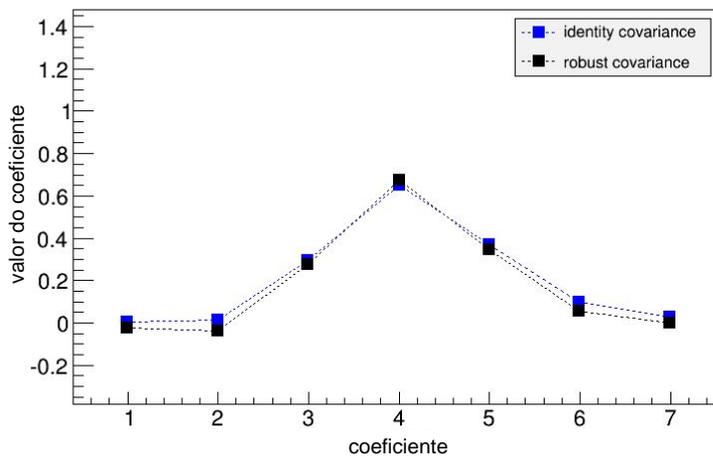
(b)

Figura 5.20: (a) Matriz de covariância estimada pelo método tradicional para uma célula de baixa ocupância e (b) seus respectivos coeficientes do método MF determinístico. Célula A3 do TileCal.

A maioria dos eventos desta célula (A3) é composta essencialmente por ruído eletrônico WGN, o qual pode ser aproximado por uma matriz identidade. Assim, é esperado que os coeficientes calculados utilizando a matriz identidade descrevendo a matriz de covariância sejam os mais adequados para a célula e ocupância em questão.



(a)



(b)

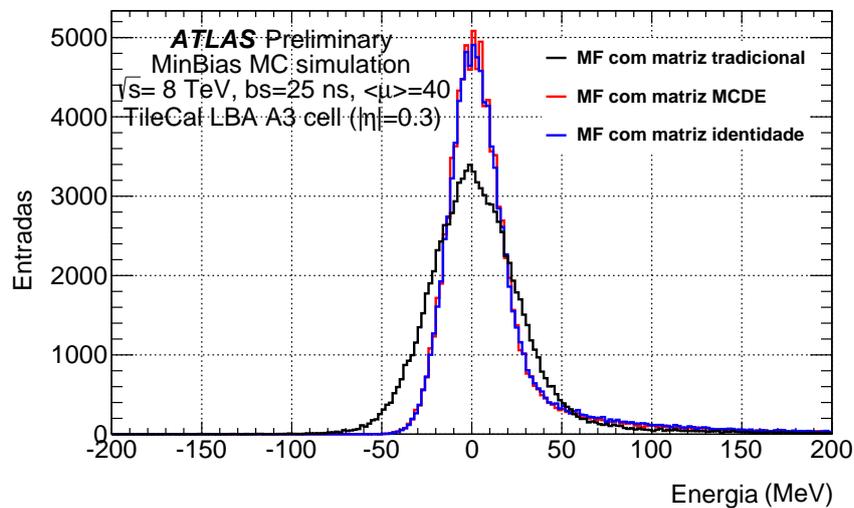
Figura 5.21: (a) Matriz de covariância estimada pelo método robusto (MCDE) para uma célula de baixa ocupância e (b) seus respectivos coeficientes do método MF determinístico. Célula A3 do TileCal.

Como pode ser notado na Figura 5.20(b), os coeficientes calculados utilizando a matriz de covariância estimada pelo método tradicional resulta em um conjunto de coeficientes diferentes. Por outro lado, os coeficientes estimados a partir da matriz de covariância estimada pelo método MCDE fornece uma descrição melhor e mais robusta, representando corretamente a estatística de segunda ordem da maioria dos eventos, o que resulta em coeficientes mais apropriados (ver Figura 5.21).

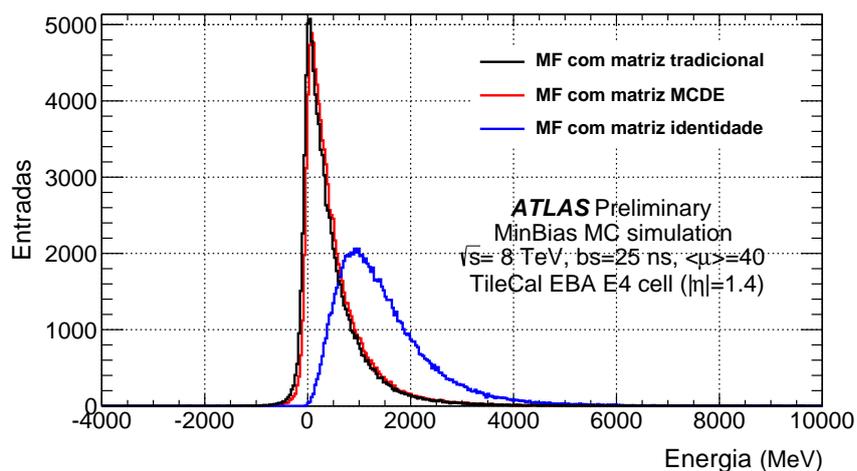
No que tange o impacto dos coeficientes na resolução da energia, a Figura 5.22(a) mostra as distribuições utilizando eventos de MB na célula A3. Esta figura repre-

senta o erro introduzido na medida pelo ruído de fundo, ou seja, quanto menor a dispersão, melhor a resolução da energia. O desempenho utilizando o método tradicional foi pior (maior dispersão) devido ao cálculo incorreto dos coeficientes do estimador. Note que os coeficientes calculados utilizando a matriz de covariância representada pela matriz identidade e estimada pelo método MCDE apresentaram desempenhos semelhantes, como esperado.

Em condições em que a ocupância é muito alta, o empilhamento de sinais deixa de ser considerado como *outliers* e as matrizes estimadas pelos métodos tradicional e MCDE se aproximam, como é mostrado na Figura 5.22(b) para o canal de maior ocupância no TileCal (célula E4). A ocupância desta célula para o cenário de $\mu \geq 40$ é aproximadamente 60% [71].



(a)



(b)

Figura 5.22: Distribuição de energia utilizando o MF determinístico com matriz de covariância estimada por abordagens diferentes. (a) célula A3 e (b) célula A13 do TileCal.

5.4.3 Resultados de simulação de eventos no TileCal previsto para o Run2

Esta seção apresenta os resultados de desempenho para simulação completa de MB do ATLAS [73] em condições de diferentes níveis de empilhamento de sinais (taxa de colisão e $\langle \mu \rangle$). Tanto o método MF determinístico quanto o método OF2 foram aplicados nos dados utilizando a aproximação da matriz de covariância como identidade e como robusta (MCDE). Aproximadamente 10.000 eventos foram utilizados para cada cenário. Os estimadores foram aplicados nos dados e os parâmetros de erro de estimação (resolução) e tendência foram analisados. Vale destacar, novamente, que eventos de MB contêm somente ruído de fundo do experimento (eletrônico mais empilhamento), e a média e dispersão da distribuição da estimação pode ser utilizada para quantificar o erro de estimação e tendência, respectivamente.

Para avaliar a melhora no erro de estimação utilizando o MF contra o OF2, o parâmetro RI (do inglês, *Resolution Improvement*) é apresentado. Este parâmetro quantifica o quanto a dispersão (RMS da distribuição do erro de estimação) é reduzida quando se utiliza o MF em comparação com o OF2 (utilizado durante o Run1). O RI é calculado de acordo com a Equação 5.4. A Figura 5.23 mostra os valores de RI para todas as células do TileCal localizadas em $\phi = 4,2$ rad. O uso do método MF determinístico melhora a resolução de energia em aproximadamente 25% nas células normais (camadas A, BC e D).

$$RI(\%) = 100 - \frac{RMS_{MFrobusto}}{RMS_{OF2identidade}} \times 100 \quad (5.4)$$

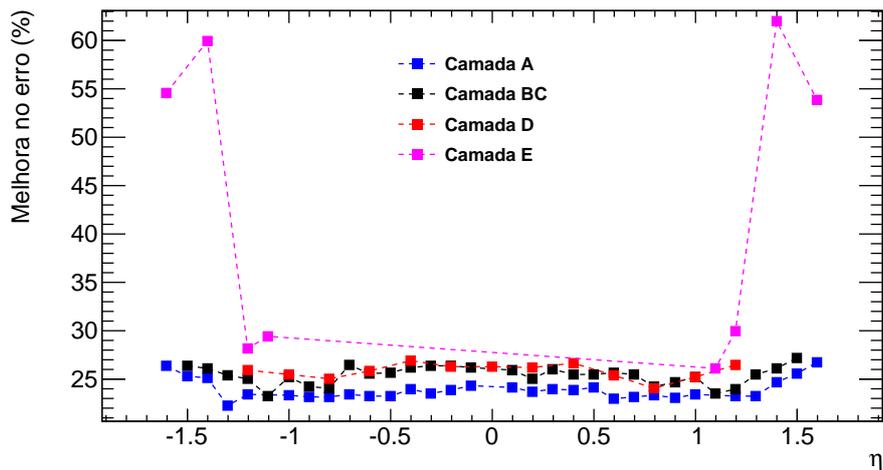
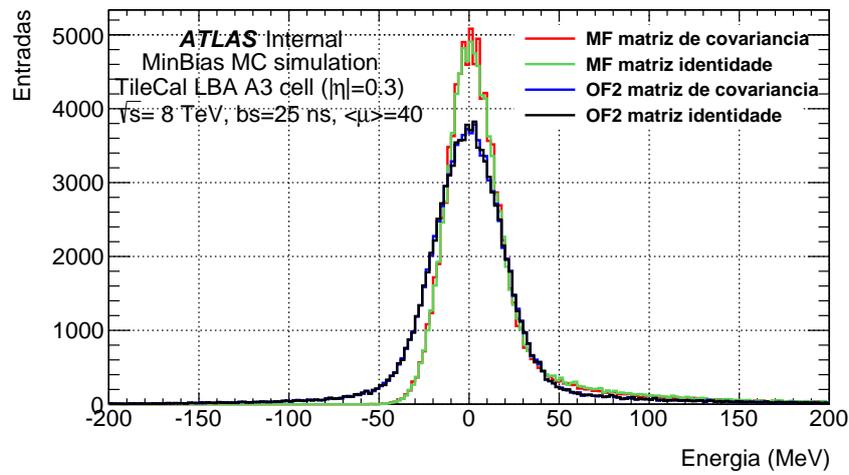


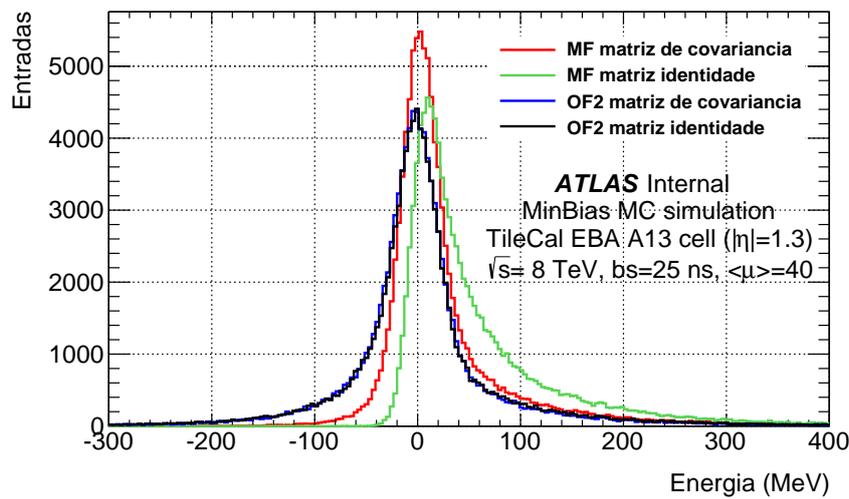
Figura 5.23: Melhora no erro (resolução) através do uso do MF utilizando a matriz robusta contra o método OF2 utilizando a matriz identidade.

Em seguida, o comportamento dos métodos em células de baixa e alta ocupância foi investigado. As Figuras 5.24(a) e 5.24(b) mostram a energia reconstruída a partir

de eventos de MB para as células A3 (baixa ocupância) e A13 (alta ocupância), respectivamente.



(a)

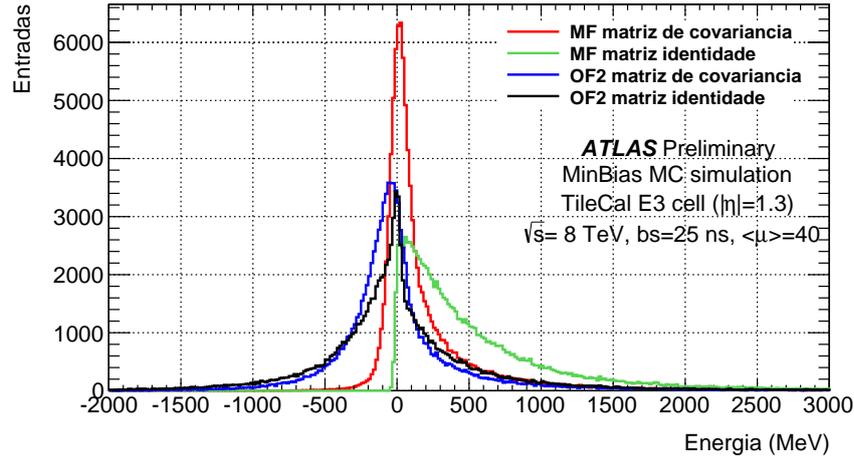


(b)

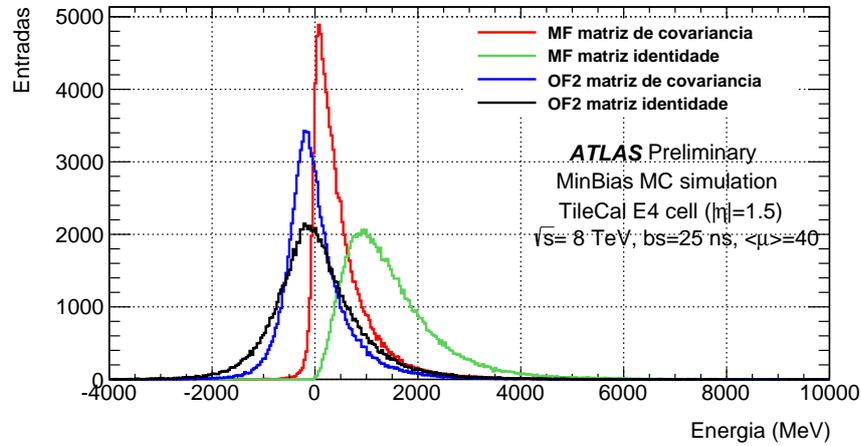
Figura 5.24: Distribuição de energia em células de (a) baixa e (b) alta ocupância utilizando matrizes de covariância diferentes para os métodos MF determinístico e OF2. Células A3 e A13 do TileCal.

Com o objetivo de avaliar o comportamento das distribuições de erro em células de maior ocupância, a Figura 5.25 mostra as estimações de energia para as células E3 e E4. Note que o uso da matriz de covariância melhora consideravelmente a resolução de energia. A distribuição do MF apresenta uma subida rápida negativa e positiva enquanto o método OF2 indica uma dispersão alta tanto para as caudas positiva e negativa da distribuição. Além disso, pode ser observado uma grande tendência para o método MF em condições de alto nível de empilhamento de sinais utilizando a matriz identidade ao invés da matriz de covariância.

A Figura 5.26 mostra a melhora na resolução de energia (RMS da distribuição



(a) Célula E3 do TileCal.



(b) Célula E4 do TileCal.

Figura 5.25: Distribuições de energia em células de alta ocupância utilizando matrizes de covariância diferentes (identidade e MCDE).

do erro) quando se utiliza o MF com a matriz de covariância MCDE contra a matriz identidade para todas as células do barril central e estendido localizadas em $\phi = 4,2$ rad. Como esperado, as células da camada A e as células E do barril estendido foram as que apresentaram melhorias mais substanciais, devido à alta ocupância nessa região. As tabelas 5.4, 5.5, 5.6 e 5.7 resumem os valores de média (tendência) e RMS (erro) dos histogramas para as camadas A, BC, D e E, respectivamente. Como esperado, as células das camadas BC e D não sofrem fortemente do efeito de empilhamento de sinais para o cenário considerado (taxa de colisão de 40 MHz e $\langle \mu \rangle = 40$).

Note que o uso do valor da linha de base do sinal pelo método MF não introduz nenhuma tendência na estimativa final da energia para todas as células, uma vez que a média das distribuições foi menor que 1 contagem de ADC (aproximadamente 12 MeV). As células de maior ocupância (E3 e E4) apresentaram um valor de ten-

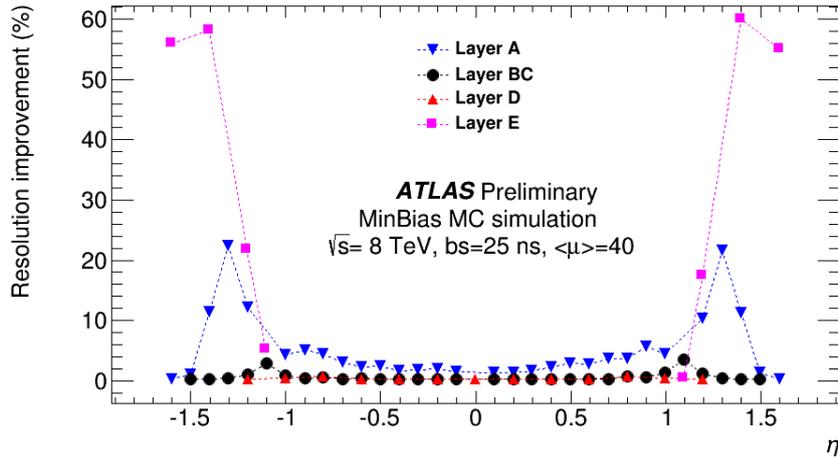


Figura 5.26: Melhora na resolução utilizando o método MF com a matriz de covariância em relação à matriz identidade (extraído de [74]).

dência alto devido à presença mais frequente de sinais localizados no mesmo tempo da colisão de interesse. Tais sinais são responsáveis por deslocar a distribuição para valores mais positivos e, como medida de correção, os valores de tendência para estas células devem ser armazenados em banco de dados e subtraídos da estimação final da energia, para compensar a contribuição destes sinais. Também pode-se observar que os valores de média e RMS para as células das camadas BC e D apresentaram valores semelhantes, devido à baixa ocupância.

5.4.4 Generalização dos coeficientes do MF

A Seção 5.4.3 mostrou os resultados para um valor fixo de luminosidade (taxa de eventos e $\langle \mu \rangle$). Entretanto, durante as tomadas de dados reais, a luminosidade decresce em função do tempo, pois o número de interações p-p é reduzido. A Figura 5.27 ilustra a curva de decaimento da luminosidade ao longo da tomada de dados. Note que ao final da tomada de dados, o nível de luminosidade atinge aproximadamente metade do valor inicial.

Como consequência, a descrição do ruído de fundo também se altera ao longo da tomada de dados. Idealmente, seria necessário carregar os coeficientes do MF associados à luminosidade atual, o que demandaria uma alta frequência de paradas no sistema de aquisição de dados. Neste contexto, esta seção apresenta um estudo de generalização de um conjunto de coeficientes do MF (utilizando a matriz de covariância correta) para ser utilizado numa faixa fixa de luminosidade, evitando assim o recarregamento constante de novos coeficientes e a reestruturação do banco de dados que armazena os coeficientes.

A Figura 5.28 mostra os coeficientes do método MF determinístico sob valores

Tabela 5.4: Resumo dos valores de média e RMS das distribuições de energia para as células da camada A (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$.

	MF covariância		MF identidade		OF2 covariância		OF2 identidade	
	média	RMS	média	RMS	média	RMS	média	RMS
0.1	2.33	14.96	2.30	15.14	-0.22	19.69	-0.20	19.70
0.2	6.34	15.41	7.21	15.62	-0.64	20.19	-0.65	20.19
0.3	0.64	16.22	0.69	16.47	-0.16	21.33	-0.09	21.33
0.4	2.19	15.08	2.21	15.40	-0.43	19.80	-0.31	19.80
0.5	4.74	17.27	5.37	17.75	-0.17	22.74	-0.17	22.76
0.6	2.81	18.24	3.21	18.71	-0.36	23.64	-0.44	23.66
0.7	0.66	14.88	0.61	15.41	-0.26	19.34	-0.29	19.35
0.8	2.62	19.98	2.81	20.68	0.19	26.02	0.26	26.03
0.9	4.57	19.76	5.54	20.93	-0.17	25.63	-0.11	25.66
1.0	4.04	23.87	4.87	24.93	0.18	31.09	0.16	31.14
1.2	6.42	19.98	10.70	22.23	-0.28	26.03	-0.18	26.03
1.3	2.68	26.23	12.71	33.45	-1.68	33.62	-1.29	34.15
1.4	5.19	21.79	8.21	24.53	-0.61	28.79	-0.62	28.90
1.5	3.03	13.09	3.13	13.24	-0.23	17.58	-0.27	17.58
1.6	4.87	12.30	4.91	12.33	0.18	16.78	0.24	16.78

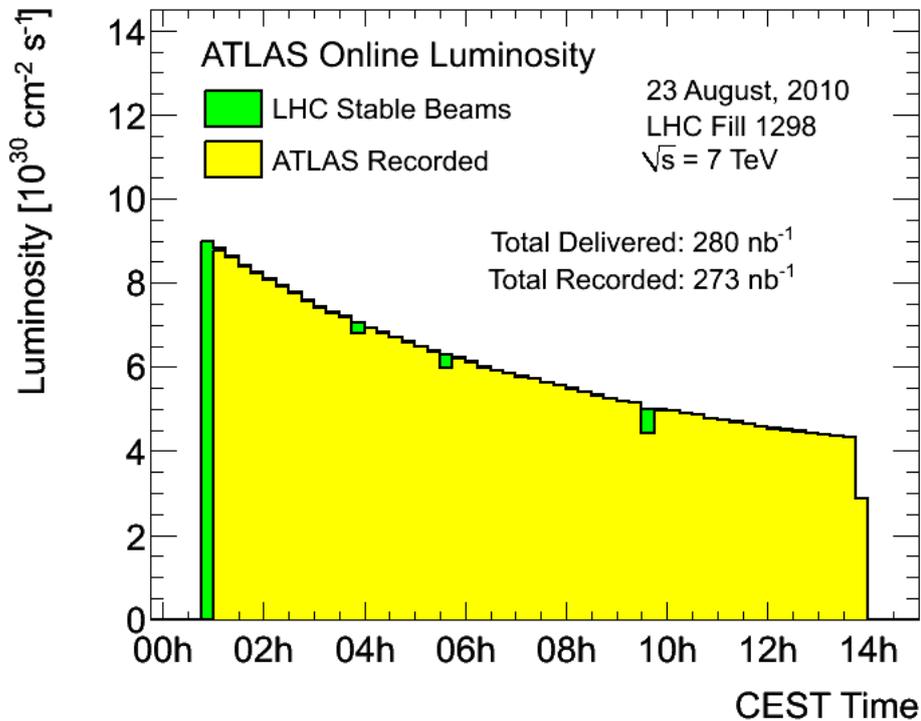


Figura 5.27: Característica típica do comportamento da luminosidade de uma run do LHC (extraído de [75]).

Tabela 5.5: Resumo dos valores de média e RMS das distribuições de energia para as células da camada BC (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$.

	MF covariância		MF identidade		OF2 covariância		OF2 identidade	
	média	RMS	média	RMS	média	RMS	média	RMS
0.1	0.23	13.98	0.22	14.00	-0.13	18.88	-0.14	18.88
0.2	-1.06	12.53	-1.11	12.54	-0.28	16.72	-0.19	16.71
0.3	4.75	13.41	4.85	13.41	0.08	18.10	0.06	18.11
0.4	4.62	12.98	4.67	13.00	-0.07	17.40	-0.08	17.41
0.5	5.86	13.52	5.80	13.53	-0.15	18.13	-0.12	18.13
0.6	6.74	13.90	6.83	13.92	-0.45	18.68	-0.35	18.68
0.7	0.71	14.74	0.79	14.76	-0.13	19.75	-0.11	19.76
0.8	-1.53	17.54	-1.56	17.64	0.01	23.12	0.01	23.13
0.9	2.89	17.16	3.05	17.23	0.18	22.75	0.29	22.76
1.0	6.84	11.52	7.52	11.65	-0.37	15.39	-0.31	15.39
1.2	4.76	16.91	5.32	17.48	-0.31	22.09	-0.32	22.10
1.3	6.68	14.74	7.45	14.88	0.02	19.37	0.03	19.37
1.4	-0.32	15.30	-0.46	15.34	0.07	20.52	0.04	20.52
1.5	1.63	11.03	1.44	11.03	-0.18	14.93	-0.16	14.93
1.6	0.01	13.08	0.01	13.09	-0.24	17.96	-0.25	17.96

Tabela 5.6: Resumo dos valores de média e RMS das distribuições de energia para as células da camada D (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$.

	MF covariância		MF identidade		OF2 covariância		OF2 identidade	
	média	RMS	média	RMS	média	RMS	média	RMS
0.0	1.40	13.87	1.39	13.87	0.14	18.81	0.08	18.81
0.2	7.06	13.76	7.03	13.76	-0.22	18.64	-0.22	18.64
0.4	-4.69	20.57	-4.67	20.57	-0.46	28.03	-0.42	28.03
0.6	-4.52	13.72	-4.67	13.72	-0.42	18.38	-0.43	18.38
0.8	-6.20	15.45	-6.37	15.53	-0.40	20.33	-0.38	20.34
1.0	-0.72	14.01	-0.83	14.05	-0.22	18.72	-0.21	18.72
1.2	0.61	13.62	0.54	13.62	0.11	18.51	0.16	18.51

diferentes de $\langle \mu \rangle$. O impacto da matriz de covariância nos coeficientes pode ser notado, em que o procedimento de otimização compensa os efeitos indesejáveis devido ao empilhamento de sinais.

A informação sobre o comportamento da luminosidade é conhecido antes do início da tomada de dados. Assim, os coeficientes ótimos para a luminosidade em questão podem ser carregados nos dispositivos DSP. Tipicamente, o valor de $\langle \mu \rangle$ decresce

Tabela 5.7: Resumo dos valores de média e RMS das distribuições de energia para as células da camada E (em MeV) considerando o cenário de espaçamento entre colisões de 25 ns e $\langle \mu \rangle = 40$.

	MF covariância		MF identidade		OF2 covariância		OF2 identidade	
	média	RMS	média	RMS	média	RMS	média	RMS
1.1	1.03	15.99	-3.69	16.04	-1.37	21.79	-1.11	21.63
1.2	3.64	60.74	6.70	73.40	-5.10	85.59	-3.76	86.65
1.3	18.28	113.84	74.04	282.69	-55.23	201.06	-43.97	299.55
1.4	145.79	311.26	971.25	687.82	-163.13	424.67	-98.75	674.40

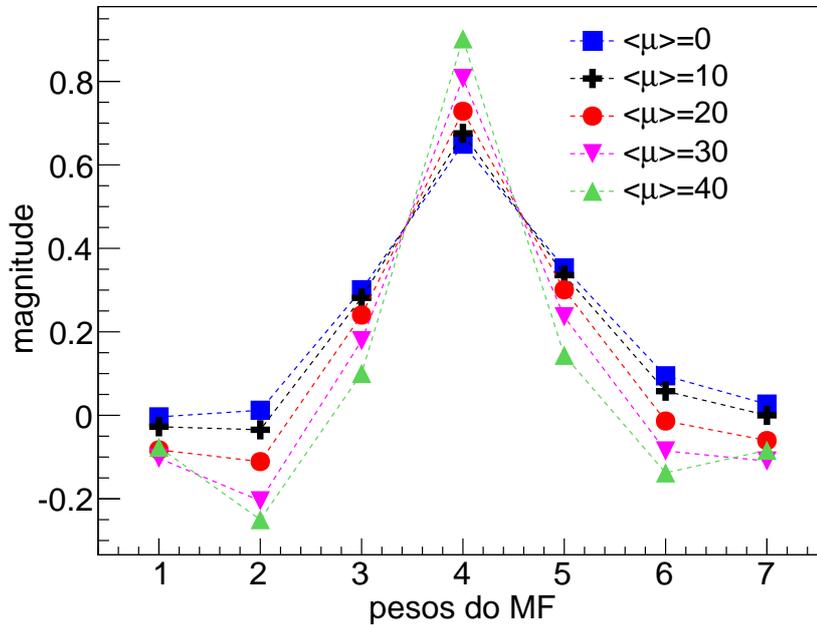
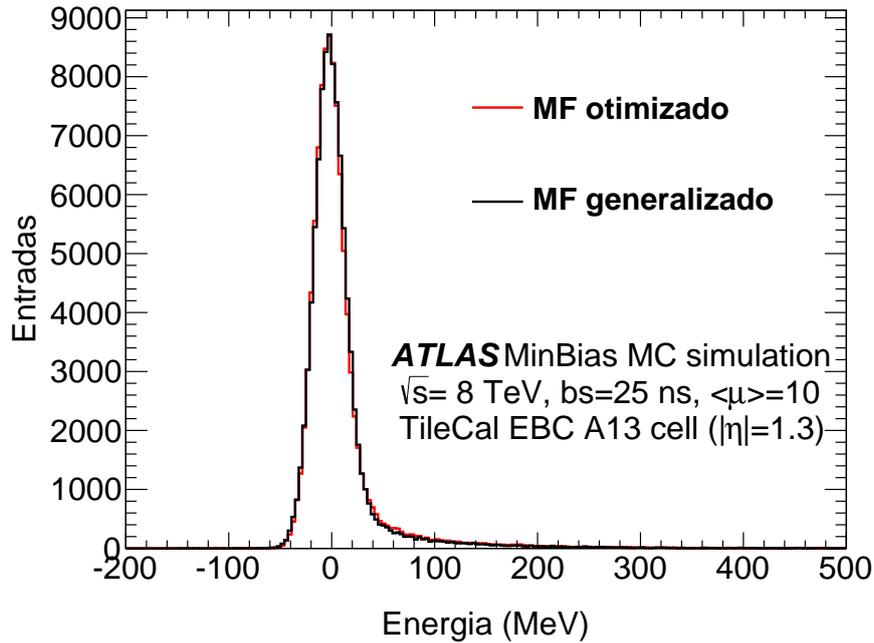


Figura 5.28: Coeficientes do método MF para a célula A13 no cenário de 25 ns de espaçamento entre as colisões.

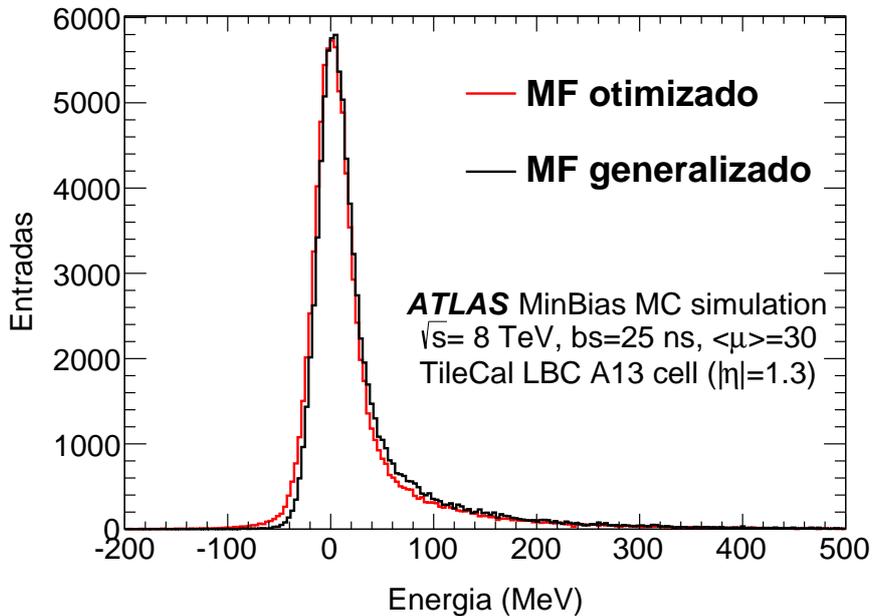
até aproximadamente um terço de seu valor inicial. Embora a luminosidade não decaia linearmente, um conjunto de coeficientes referentes ao valor médio de $\langle \mu \rangle$ pode ser testado para generalização.

Para avaliar a eficiência de generalização, foi considerado o ambiente de 25 ns de espaçamento entre colisões e número médio de interação por colisão variando de $\langle \mu \rangle = 30$ até $\langle \mu \rangle = 10$ (valores típicos previstos durante o Run2 do LHC). Além disso, foi considerada a célula A13, devido à sua alta ocupância. Os coeficientes ótimos correspondem aos coeficiente calculados utilizando os dados de 25 ns de espaçamento entre colisões com $\langle \mu \rangle = 30$ e $\langle \mu \rangle = 10$, separadamente. Já os coeficientes generalizados foram calculados considerando a média entre os dois cenários, ou seja, $\langle \mu \rangle = 20$. As figuras 5.29(b) e 5.29(a) mostram as distribuições

para $\langle \mu \rangle = 10$ e $\langle \mu \rangle = 30$, respectivamente.



(a)



(b)

Figura 5.29: Distribuições de energia na célula A13 utilizando os coeficientes ótimos e generalizados para os cenários de (a) $\langle \mu \rangle = 10$ e (b) $\langle \mu \rangle = 30$, ambos considerando 25 ns de espaçamento entre colisões.

A Tabela 5.8 resume os valores de média e RMS das distribuições de energia utilizando os coeficientes ótimos e generalizados para a célula A13. Pode ser observado que o uso dos coeficientes generalizados não degrada o desempenho do método MF,

uma vez que ambos a média e RMS apresentam valores próximos comparado aos valores pelos coeficientes ótimos. Vale ressaltar que, para as outras células, o impacto da generalização dos coeficientes tende a ser ainda menor, devido à tais células apresentarem ocupância mais baixa que a célula A13, excetuando as células E.

Tabela 5.8: Resumo dos valores de média e RMS (em MeV) para o cenário de 25 ns de espaçamento entre colisões na célula A13 do TileCal.

	coeficientes ótimos		coeficientes generalizados	
	média	RMS	média	RMS
$\langle \mu \rangle = 10$	-3.00	16.31	-2.47	16.34
$\langle \mu \rangle = 20$	0.21	19.32	0.21	19.32
$\langle \mu \rangle = 30$	1.40	22.80	3.16	22.81

Capítulo 6

Reconstrução de Energia para Ambiente com Alto Nível de Empilhamento de Sinais

Os colisionadores atuais e, principalmente, os futuros tendem a aumentar a taxa de eventos e o número de partículas contido nos feixes de colisão com o objetivo de aumentar a probabilidade de se observarem fenômenos raros [7, 76]. O aumento do número de interações por colisão e a taxa de eventos resulta numa quantidade maior de sinais disponíveis para as análises, melhorando a significância estatística. Entretanto, foi mostrado que este ambiente de alta luminosidade traz novos desafios em termos de aquisição de dados e processamento de sinais.

No capítulo anterior, foi mostrado que o método MF e as técnicas baseadas na minimização da variância incorporam o efeito do empilhamento de sinais em seus projetos através da matriz de covariância do ruído de fundo. Embora esta abordagem apresente vantagens, como bom desempenho e simplicidade, ela também introduz algumas limitações em condições de alto nível de empilhamento de sinais, tais como:

1. o projeto do estimador se torna dependente da luminosidade;
2. a distribuição do ruído de fundo não pode ser modelada por uma distribuição gaussiana;
3. a informação extra presente nos sinais das colisões adjacentes é descartada.

Neste capítulo, uma nova técnica baseada na desconvolução de sinais é utilizada para recuperar os sinais misturados, provenientes de colisões diferentes e presentes numa mesma janela de leitura. Técnicas de desconvolução têm sido utilizadas em processamento digital de sinais como, por exemplo, equalização de canais em comunicações [77], entre outras aplicações. O sinal recebido pode ser modelado por uma convolução entre o sinal de interesse e a resposta do canal de transmissão. Um filtro

no receptor, calibrado para cancelar a resposta do canal, realiza o processo de desconvolução e o sinal de interesse é totalmente recuperado. Tal abordagem introduz novas perspectivas para a aplicação de métodos modernos de processamento digital de sinais em calorimetria.

Para que o procedimento de desconvolução se realize completamente, a leitura ininterrupta das amostras seria necessária. Entretanto, em sistemas de calorimetria, normalmente apenas uma janela curta de leitura é disponível e, conseqüentemente, a desconvolução não pode ser realizada completamente. Nesta tese, uma aproximação para a desconvolução é proposta, em que somente algumas amostras (presentes na janela de leitura) são disponíveis. O método é baseado na minimização do erro médio quadrático e sua relação com os métodos descritos nos capítulos anteriores também é destacada.

O objetivo do método é realizar a reconstrução do sinal de interesse de forma que o projeto do estimador seja independente do nível de luminosidade e, ainda, permitir a reconstrução dos sinais adjacentes presentes na mesma janela de leitura. A informação proveniente de colisões vizinhas é utilizada para melhorar a estimação do sinal de interesse. Após a etapa de desconvolução, as amplitudes dos sinais identificados podem ser estimadas através do projeto de estimadores cujos componentes são ortogonais entre si.

Esta abordagem de estimação alternativa possibilita a reconstrução da forma do pulso em condições de alto nível de empilhamento de sinais. O método pode ainda ser utilizado para detectar sinais corrompidos ou saturados. Tipicamente, o Erro de Reconstrução (RE, do inglês, *Reconstruction Error*) é calculado a partir da extração da raiz quadrada do MSE entre as amostras recebidas e as amostras do sinal reconstruído [78, 79]. Em casos em que a forma do pulso pode ser considerada invariante, a informação sobre as amplitudes dos sinais presentes pode ser utilizada para reconstruir o sinal analógico recebido da eletrônica de leitura, e a medida de RE é utilizada para testar a eficiência desta operação.

A próxima seção apresenta a teoria na qual o método de desconvolução é desenvolvido. Em seguida, a implementação do método é descrita e resultados utilizando dados simulados são apresentados. O desempenho de estimação também é testado para pulsos diferentes, simétricos e assimétricos. Ao final, alguns resultados do método proposto aplicado ao TileCal serão apresentados.

6.1 Recuperação de fontes através de desconvolução

Para o método de desconvolução de sinais proposto, a informação do pulso digital do calorímetro é interpretada como a saída $x[k]$, em que k corresponde à amostra temporal, de um sistema discreto linear e invariante no tempo (LTI, *Linear Time Invariant*) [63] de resposta ao impulso $g[k]$, como mostrado na Figura 6.1. A energia amostrada de uma dada colisão ($a[k]$ da Figura 6.1) é modelada por um impulso o qual representa a energia depositada no calorímetro na k -ésima colisão. Desta forma, a saída do sistema LTI $y[k]$ pode ser modelada como o sinal conformado que cobre várias colisões adjacentes, corrompido pelo ruído aditivo ($n[k]$ na Figura 6.1) WGN de variância unitária (matriz identidade representando a matriz de covariância).

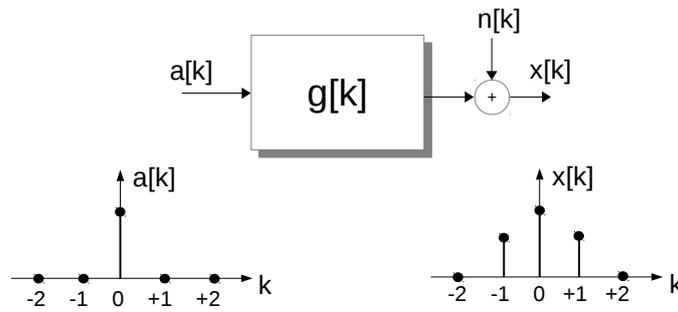


Figura 6.1: Sistema LTI para dados digitais de calorímetros. $a[k]$ modela a energia depositada em uma dada colisão, $g[k]$ corresponde à resposta ao impulso do sistema, $x[k]$ são os valores adquiridos e $n[k]$ representa o ruído aditivo.

Em condições de empilhamento de sinais, a saída contém sinais sobrepostos. A resposta do calorímetro pode ser modelada como a convolução entre o sinal de entrada (interesse) e a resposta ao impulso do sistema LTI [63]

$$x[k] = \sum_i (g[i]a[k - i]) + n[k], \quad (6.1)$$

em que a resposta ao impulso $g[k]$ do calorímetro corresponde à forma do pulso de referência. Assim, para recuperar as componentes do sinal de entrada, técnicas de desconvolução linear podem ser aplicadas.

Para o caso em que se tem disponível todas as amostras de forma ininterrupta, é possível recuperar o sinal de entrada aplicando-se uma versão causal do filtro inverso de resposta em frequência $1/G(z)$, onde $G(z)$ é a resposta em frequência do sistema LTI [80], ou seja, a transformada- z da resposta ao impulso do sistema ($g[k]$). Como em calorímetros os sistemas de filtragem são capazes de armazenar somente uma janela contendo a colisão de interesse e colisões adjacentes, a informação não

está disponível por completo e as componentes do sinal de entrada não podem ser totalmente recuperadas por um filtro inverso.

No caso de o pulso de interesse se estender a D colisões e N amostras de $x[k]$ forem armazenadas, a Equação (6.1) pode ser descrita melhor na forma compacta

$$\mathbf{x}_{N \times 1} = \mathbf{G}_{N \times P} \mathbf{a}_{P \times 1} + \mathbf{n}_{N \times 1}, \quad (6.2)$$

onde $P = D + N - 1$. Nesta equação, $\mathbf{x}_{N \times 1} = \{x[0], x[1], x[2], \dots, x[N]\}$ corresponde ao sinal de saída do calorímetro, $\mathbf{G}_{N \times P}$ é a matriz com P versões deslocadas do pulso de referência do calorímetro, $\mathbf{a}_{P \times 1} = \{a[0], a[1], a[2], \dots, a[P]\}$ é o vetor sinal de entrada do calorímetro (energia depositada em cada colisão), e $\mathbf{n}_{N \times 1} = \{n[0], n[1], n[2], \dots, n[N]\}$ é o vetor de ruído aleatório. A fim de exemplificar a montagem da matriz $\mathbf{G}_{N \times P}$, a Equação (6.3) mostra tal matriz para $N = P = 7$. Pode-se observar que o sinal de saída $\mathbf{x}_{N \times 1}$ terá contribuições de cada sinal deslocado, ponderado pela respectiva amplitude $\mathbf{a}_{P \times 1}$.

$$\mathbf{G}_{N \times P}^T = \begin{pmatrix} g[3] & g[4] & g[5] & g[6] & 0 & 0 & 0 \\ g[2] & g[3] & g[4] & g[5] & g[6] & 0 & 0 \\ g[1] & g[2] & g[3] & g[4] & g[5] & g[6] & 0 \\ g[0] & g[1] & g[2] & g[3] & g[4] & g[5] & g[6] \\ 0 & g[0] & g[1] & g[2] & g[3] & g[4] & g[5] \\ 0 & 0 & g[0] & g[1] & g[2] & g[3] & g[4] \\ 0 & 0 & 0 & g[0] & g[1] & g[2] & g[3] \end{pmatrix} \quad (6.3)$$

O sistema descrito pela Equação (6.2) não possui solução única para um dado \mathbf{a} , uma vez que existem somente N equações para P variáveis ($P > N$). A seção seguinte apresenta algumas modificações para produzir uma versão mais prática da Equação (6.2).

6.2 A Matriz de Desconvolução (DM)

A implementação proposta consiste em recuperar apenas os $p \leq N$ componentes do vetor \mathbf{a} da Equação (6.2). As outras componentes são incorporadas ao vetor de ruído de fundo \mathbf{n}_p . Desta forma, as amplitudes dos p sinais devem ser estimadas simultaneamente e o vetor de ruído \mathbf{n}_p inclui o ruído eletrônico somado com os sinais remanescentes empilhados. Assim, a Equação (6.2) pode ser escrita por

$$\mathbf{x} = \mathbf{G}_p \mathbf{a}_p + \mathbf{n}_p, \quad (6.4)$$

onde a matriz \mathbf{G}_p tem ordem $N \times p$ e possui p versões deslocadas do pulso de referência do calorímetro onde suas amplitudes \mathbf{a}_p devem ser determinadas. O procedimento resume em encontrar a matriz \mathbf{W}_p , a qual tem ordem $N \times p$ e estima as p componentes de entrada $\hat{\mathbf{a}}_p$. Utilizando o estimador BLUE, a equação do vetor de estimação é expressa por

$$\hat{\mathbf{a}}_p = \mathbf{W}_p^T \mathbf{x}. \quad (6.5)$$

Para estimadores imparciais, é necessário que

$$E\{\hat{\mathbf{a}}_p\} = \mathbf{W}_p^T E\{\mathbf{x}\} = \mathbf{a}_p, \quad (6.6)$$

onde $E\{\cdot\}$ é o valor esperado de uma variável aleatória. Assumindo-se que a linha de base é removida como parte de um pré-processamento e o vetor \mathbf{n}_p tem média zero, o valor esperado de \mathbf{x} se torna

$$E\{\mathbf{x}\} = E\{\mathbf{G}_p \mathbf{a}_p + \mathbf{n}_p\} = \mathbf{G}_p \mathbf{a}_p. \quad (6.7)$$

Substituindo-se a Equação (6.7) na Equação (6.6) produz

$$\mathbf{W}_p^T \mathbf{G}_p = \mathbf{I}_p. \quad (6.8)$$

As variâncias dos p estimadores em \mathbf{W}_p devem ser minimizadas respeitando as restrições impostas pela Equação (6.8). O procedimento de minimização simultânea de equações pode ser encontrado em [46] e resulta na fórmula fechada do estimador BLUE

$$\mathbf{W}_p = \mathbf{C}_p^{-1} \mathbf{G}_p (\mathbf{G}_p^T \mathbf{C}_p^{-1} \mathbf{G}_p)^{-1}, \quad (6.9)$$

em que a matriz de covariância \mathbf{C}_p do vetor de ruído \mathbf{n}_p tem dimensão $N \times N$. Conseqüentemente, a matriz \mathbf{C}_p absorve a estatística de segunda ordem das amostras que não estão incluídas em \mathbf{G}_p .

A Equação (6.9) representa uma aproximação do procedimento de desconvolução para as $p \leq N$ componentes selecionadas. O projeto do estimador para $p = 1$ e $p = N$ apresentam propriedades especiais que são destacadas a seguir.

6.2.1 Projeto do estimador para apenas um sinal central

No caso onde $p = 1$, a matriz \mathbf{G}_1 se iguala ao vetor de amostras do pulso de referência \mathbf{g} . A matriz de covariância do ruído absorve a informação do empilhamento de sinais, assim como o método MF apresentado anteriormente. Desta forma, a Equação (6.9) se torna

$$\mathbf{W}_1 = \frac{\mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}, \quad (6.10)$$

o qual é o estimador MF apresentado no Capítulo 4.

A Equação (6.10) é utilizada para recuperar a amplitude somente do sinal central na janela de leitura, o qual representa a informação relevante de acordo com o sistema de filtragem. Para $p = 1$, a Equação (6.8) apresenta o número mínimo de restrições. Assim, este estimador exibe a menor variância entre os estimadores lineares e, para ruído gaussiano, o resultado deste estimador é ótimo (se iguala ao estimador Imparcial de Mínima Variância (MVU, *Minimum Variance Unbiased*) [46]). Porém, este estimador necessita de conhecimento prévio sobre a informação da estatística de empilhamento de sinais, uma vez que a matriz de covariância contém a informação de todas colisões contidas na janela de leitura.

Quanto maior o valor de p , menor é a informação sobre a luminosidade que é absorvida pela matriz de covariância \mathbf{C}_p e o estimador se torna robusto à mudanças na luminosidade.

6.2.2 Projeto do estimador para N sinais

O caso onde $p = N$ representa a melhor aproximação do processo de desconvolução. A fim de simplificar a nomenclatura, p é omitido das equações que se seguem. Tal implementação consiste em recuperar os N componentes do vetor \mathbf{a} da Equação (6.4).

Substituindo-se a Equação (6.9) na Equação (6.5), o estimador BLUE para o vetor de parâmetros \mathbf{a} é

$$\hat{\mathbf{a}} = (\mathbf{G}^T \mathbf{C}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{C}^{-1} \mathbf{x}, \quad (6.11)$$

em que \mathbf{G} é uma matriz de ordem $N \times N$, \mathbf{C} é a matriz de covariância do ruído de ordem $N \times N$ e \mathbf{x} corresponde ao vetor de observação de ordem $N \times 1$.

Quando o pulso inteiro ($D \leq N$) está presente na janela de leitura, a informação do empilhamento de sinais na matriz \mathbf{C} pode ser desprezada (para $p = N$) para considerações práticas. Sendo assim, o ruído de fundo é composto apenas pelo ruído eletrônico WGN, e a matriz \mathbf{C} pode ser eliminada da Equação (6.11), resultando na seguinte simplificação

$$\hat{\mathbf{a}} = \mathbf{G}^{-1} \mathbf{x}. \quad (6.12)$$

Apesar de o estimador da Equação (6.12) ser uma boa aproximação para o estimador BLUE, ele é melhor visualizado como estimador de erro médio quadrático mínimo (MMSE, *Minimal Mean Square Error*), uma vez que a estatística de segunda ordem não é considerada. A matriz \mathbf{G}^{-1} é chamada de matriz de desconvolução (DM, do inglês, *Deconvolution Matrix*) e é calculada *offline* para cada célula de leitura utilizando a forma do pulso referente ao canal (resposta ao impulso do canal). Vale

ressaltar que a mesma matriz é aplicada independente da luminosidade.

De acordo com as restrições impostas pela Equação (6.8), uma propriedade de \mathbf{W}_p é que seu i -ésimo estimador \mathbf{w}_i é ortogonal ao pulso de referência \mathbf{g}_j centrado na j -ésima colisão

$$\mathbf{w}_i^T \mathbf{g}_j = \delta_{ij}, \quad (6.13)$$

em que δ_{ij} corresponde ao delta de Kronecker. Tal condição de ortogonalidade caracteriza a imunidade a empilhamento de sinais do método, destacando sua independência à luminosidade.

6.3 Resultados de Simulação

Objetivando avaliar o desempenho do DM, o pulso unipolar mostrado na Figura 6.2 foi utilizado como referência, observando que o pulso está totalmente contido na janela de leitura. Aproximadamente 1.000.000 de amostras temporais (colisões) representando um sequência de conversões de ADC de um dado canal de leitura foram populadas aleatoriamente com um dado valor de ocupância (por exemplo 10%). Para simular a energia depositada pelo empilhamento de sinais, uma distribuição exponencial de média 300 MeV foi utilizada [73, 81]. Em seguida, o sinal de interesse foi sobreposto. Um desvio de fase de $[-4, +4]\%$ e uma deformação no pulso de $[-1, +1]\%$ foram considerados. Após este processo, o ruído gaussiano decorrelacionado com $\sigma=20$ MeV foi adicionado amostra a amostra, simulando o ruído eletrônico. Finalmente, janelas de 7 amostras foram recortadas sequencialmente, simulando a aquisição de dados.

Além disso, metade do conjunto de dados foi utilizado para calcular os coeficientes dos métodos MF e OF2 (obtenção da matriz de covariância do ruído através do método clássico). A outra metade é utilizada para análise de desempenho dos estimadores considerados.

As distribuições de reconstrução da energia, para uma ocupância de 10%, são mostradas na Figura 6.3. A escala logarítmica é utilizada para melhorar a visualização e comparação entre os métodos. O mesmo comportamento foi observado para todos os estimadores DM mas, para efeito de simplificação, apenas a_0 é mostrado (estimativa da amplitude do sinal centralizado na janela, ou seja, do sinal proveniente da colisão de interesse). O método MF apresenta o melhor desempenho (menor dispersão). A cauda positiva nas distribuições pode ser explicada pela presença de sinais adquiridos na colisão de interesse, mas que pertencem a outras interações da mesma colisão (empilhamento simultâneo de eventos). Desta forma, eles são interpretados como sinais válidos pelos métodos.

A Figura 6.4 mostra a reconstrução da energia para os três estimadores consi-

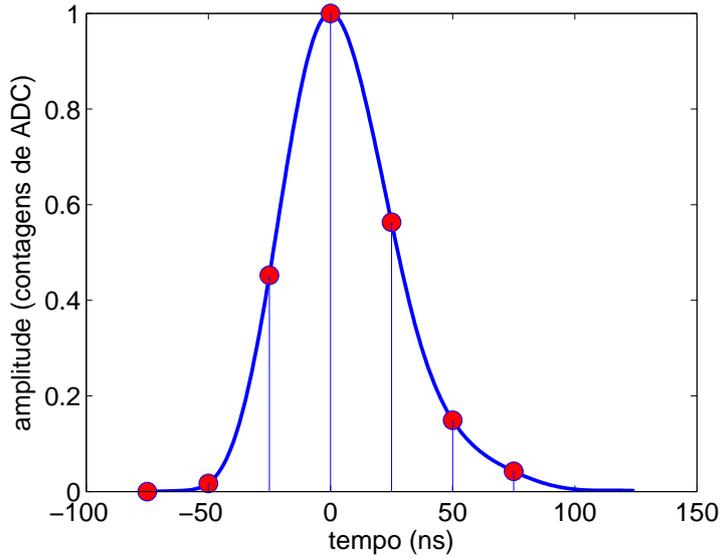


Figura 6.2: Pulso de referência utilizado para análise de desempenho.

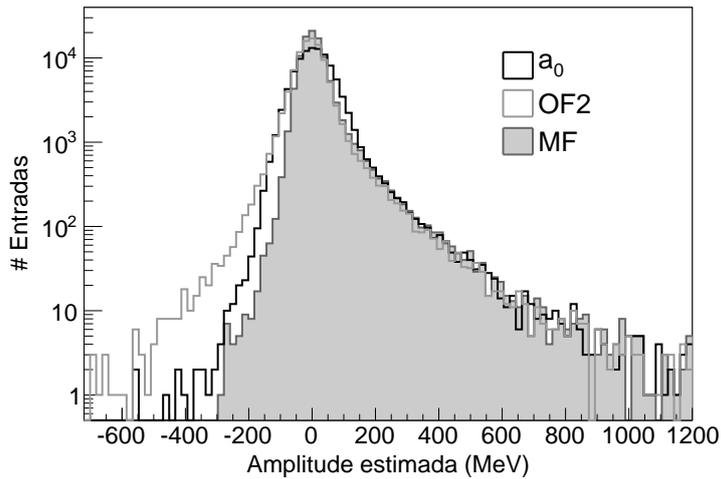


Figura 6.3: Reconstrução da energia para simulação de empilhamento de sinais. A informação da luminosidade é utilizada nos estimadores MF e OF2.

derados quando nenhuma informação sobre a luminosidade é empregada ($\mathbf{C} = \mathbf{I}$). Como já mostrado na Seção 5.4, a variância do método OF2 aumenta consideravelmente, como pode ser notado pela cauda negativa de sua distribuição. Além disso, o MF apresenta uma tendência grande, característica também esperada.

A Figura 6.5 apresenta a correlação entre as estimativas a_{-1} (estimativa da amplitude do sinal empilhado resultante de uma colisão anterior a de interesse) and a_{+1} (estimativa da amplitude do sinal empilhado resultante de uma colisão posterior a de interesse) fornecidas pelo DM. Como esperado, não é observado correlação acima do nível de ruído. O mesmo comportamento é observado em todos os estimadores DM, mostrando que não existe dependência entre as energias depositadas em

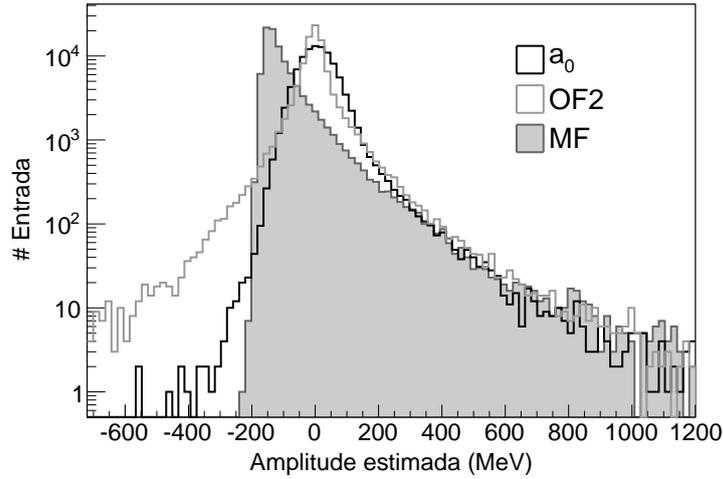


Figura 6.4: Reconstrução da energia para simulação de empilhamento de sinais. A informação da luminosidade não é utilizada nos estimadores MF e OF2 ($\mathbf{C} = \mathbf{I}$).

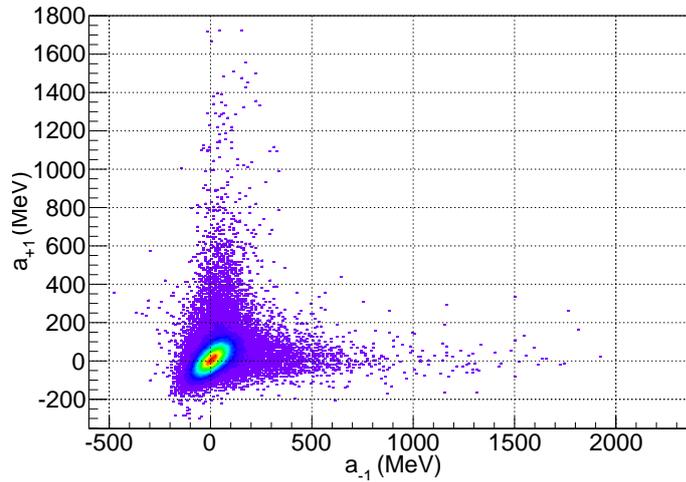


Figura 6.5: Correlação entre os estimadores DM a_{-1} e a_{+1} . Note que nenhuma correlação é observada para valores acima do ruído, devido a ortogonalidade dos estimadores DM.

colisões diferentes.

Esta análise mostrou que a matriz de desconvolução recupera as amplitudes dos sinais adjacentes com boa precisão. A estimação de colisões adjacentes presentes numa mesma janela de leitura traz novas perspectivas para as análises. A próxima seção apresenta uma sugestão eficiente para estimação da amplitude em condições de empilhamento de sinais.

6.4 Método DM combinado

Nesta seção, as características de ortogonalidade do DM são exploradas com o objetivo de melhorar o desempenho do estimador. Novamente, o desempenho é avaliado em termos de erro de estimação e comparado com os métodos MF e OF2. Ao final, o estudo do impacto de diferentes formas de pulso (simétrico e assimétrico) operando em diferentes condições de ocupância e SNR, é considerado.

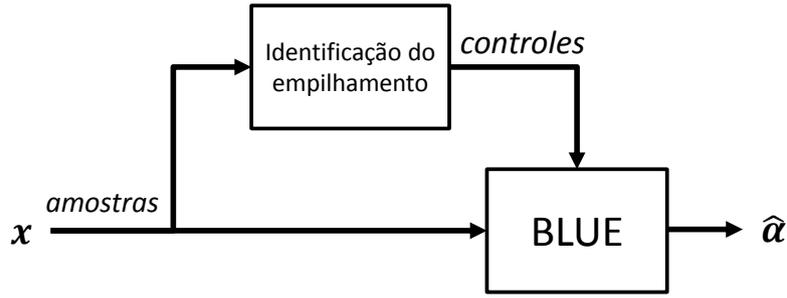
6.4.1 COF, *Constrained Optimal Filter*

No Capítulo 5, foi mostrado que o método MF apresentou melhor desempenho entre os estimadores considerados, devido ao fato de não usar restrições em seu projeto e a possibilidade de usar a matriz de covariância do ruído de fundo para reduzir as incertezas devido ao empilhamento de sinais. Por outro lado, o método MF apresenta algumas particularidades indesejadas:

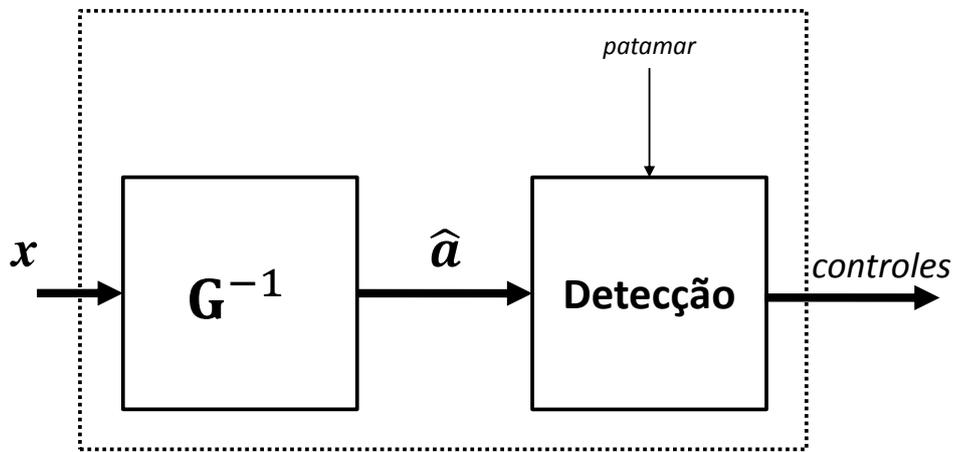
- o projeto dos coeficientes do MF depende da luminosidade, uma vez que a matriz de covariância muda de acordo com o nível de empilhamento de sinais;
- o estimador é tendencioso em condições de empilhamento de sinais, pois seus coeficientes não são ortogonais aos sinais de colisões adjacentes;
- a distribuição do ruído de fundo se torna não-gaussiana, o que torna o MF determinístico sub-ótimo.

Foi mostrado que o DM não sofre dos problemas listados acima. Contudo, dado o número grande de restrições, o desempenho dos estimadores DM decresce em condições de baixa luminosidade. Porém, a característica do DM pode ser explorada com o intuito de detectar em quais colisões (BC), que estão contidas na janela de leitura, ocorreu deposição relevante de energia. A partir do conhecimento dos BC com informação relevante, pode-se projetar um estimador BLUE, o qual estima somente os sinais dos BC identificados pelo DM. Este método proposto é chamado de COF (do inglês, *Constrained Optimal Filter*) e visa minimizar a variância do estimador ao mesmo tempo que permite uma estimação não tendenciosa e independente da luminosidade.

O diagrama de blocos do sistema proposto é mostrado na Figura 6.6(a). Uma etapa de pre-processamento para detecção de sinais indica os BC onde os sinais empilhados estão localizados na janela de leitura. Esta informação é propagada para o bloco COF o qual estima simultaneamente a amplitude do sinal central juntamente com os possíveis sinais empilhados (detectados pelo DM). Ou seja, a identificação de empilhamento de sinais é obtida pelo DM, como mostrado na Figura 6.6(b), em que \mathbf{G}^{-1} corresponde à matriz de desconvolução, como mostrado na Equação (6.12). As



(a) Diagrama do estimador combinado. Os índices dos sinais empilhados são lidos pelo bloco COF.



(b) Processo de identificação de empilhamento de sinais utilizando os estimadores DM. Os sinais são detectados se o valor de suas amplitudes superarem um patamar baseado no RMS do ruído.

Figura 6.6: Diagrama de blocos do método COF.

amplitudes de saída do processo de desconvolução são comparadas à um patamar calibrado por projeto. Tais sinais de controle binário são lidos pelo bloco COF.

O COF é implementado utilizando a Equação (6.9), para $p = K + 1$, onde K corresponde ao número de sinais detectados. A matriz \mathbf{G}_p contém a forma do pulso de referência (sinal de interesse localizado no BC central) e as formas dos pulsos dos sinais detectados, pertencentes à colisões adjacentes. Uma vez que os sinais empilhados são detectados previamente, o ruído de fundo resultante contém somente a componente gaussiana, e o estimador opera em condições ótimas. Nestes caso, a Equação (6.9) pode ser simplificada e expressa por

$$\mathbf{W}_p = \mathbf{G}_p(\mathbf{G}_p^T \mathbf{G}_p)^{-1}, \quad (6.14)$$

o qual representa o estimador de amplitude para os p sinais. Note que \mathbf{W}_p , dado pela Equação (6.14), corresponde à matriz pseudo-inversa [77] de \mathbf{G}_p , que é a solução

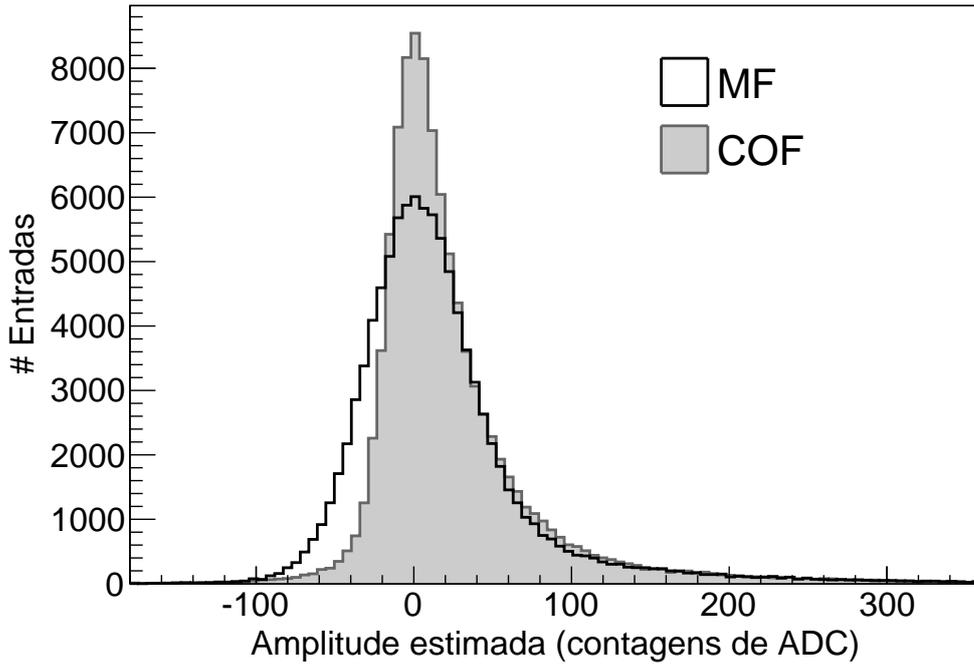


Figura 6.7: Distribuição de energia dos métodos COF e MF.

MMSE da Equação (6.8) para matrizes de convolução não-quadradas. Se nenhum empilhamento de sinais for detectado, a Equação (6.14) se torna

$$\mathbf{W}_1 = \frac{\mathbf{g}}{\mathbf{g}^T \mathbf{g}}, \quad (6.15)$$

que corresponde ao método MF determinístico para $\mathbf{C} = \mathbf{I}$. Desta forma, sob tais condições, o estimador opera em condições ótimas.

A Figura 6.7 mostra os resultados quando o método COF é aplicado no mesmo conjunto de dados descrito na Seção 6.3. Pode ser observado que o método COF apresenta uma cauda negativa mais curta, o que indica melhor desempenho na reconstrução final do evento. A cauda positiva apresentada pelos estimadores é proveniente de sinais localizados no BC central mas que pertencem à colisão de interesse. Além disso, os coeficientes do MF foram otimizados utilizando a matriz de covariância do ruído, enquanto o COF não utiliza nenhuma informação sobre a luminosidade.

6.4.2 Dependência da simetria do pulso e de condições diferentes de SPR

Esta seção avalia o desempenho do método COF para dois tipos de pulsos unipolares: um aproximadamente simétrico, mostrado pela Figura 6.2, e outro assimétrico, ilustrado na Figura 6.8. O pulso simétrico possui uma forma gaussiana e 7 amostras

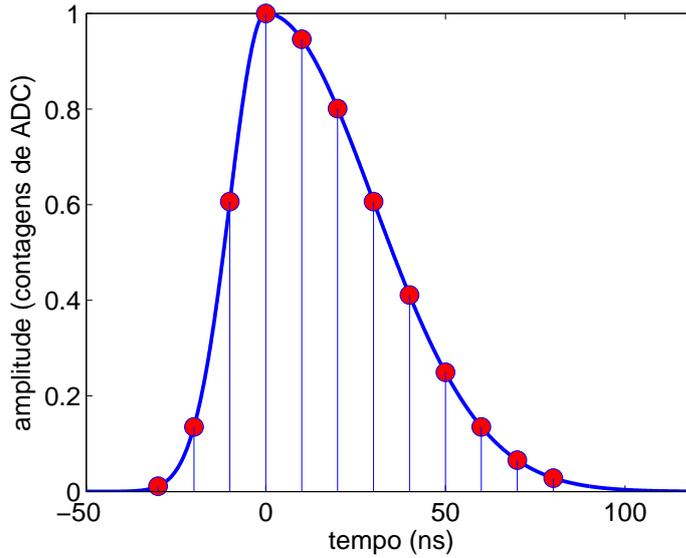


Figura 6.8: Pulso assimétrico utilizado para análise de dependência da simetria.

representam todo o pulso, sendo o valor máximo localizado na quarta amostra. No caso do pulso assimétrico, o pico também está localizado na quarta amostra, mas este possui um decaimento mais lento e são necessárias 12 amostras digitais para cobrir todo o pulso. Vale destacar que, em ambos os casos, o pulso está contido na janela de leitura.

A simulação adotada nesta análise é a mesma descrita na Seção 6.3, onde janelas de M amostras foram sequencialmente recortadas, sendo $M = 7$ e $M = 12$ para os pulsos simétrico e assimétrico, respectivamente. Metade das janelas foi utilizada para compor o conjunto de ruído (eletrônico mais empilhamento). Diferentemente da análise anterior, a outra metade corresponde ao conjunto de sinal, onde um pulso central (junto com os parâmetros de desvio de fase e deformação do pulso) foi adicionado a cada janela representando o sinal de interesse cuja amplitude deve ser estimada. Para tais sinais, a amplitude é um parâmetro aleatório modelado por uma distribuição exponencial de média m que corresponde a cenários diferentes de relação sinal-empilhamento (SPR , *Signal to Pile-up ratio*). O valor SPR é calculado de acordo com a seguinte equação:

$$SPR = \frac{m_{sinal}}{m_{empilhamento}}, \quad (6.16)$$

em que m_{sinal} e $m_{empilhamento}$ são as médias das amplitudes para o sinal e o empilhamento de sinal, respectivamente.

A Figura 6.9 mostra o erro de estimação em função da ocupância para um valor fixo de SPR de 10 utilizando os dois tipos de pulsos. 0% de ocupância significa a ausência de empilhamento de sinais e, uma ocupância de 50%, significa que, em

média, existe deposição de energia em pelo menos um em cada dois BC. Em condições de muito baixa ocupância, o empilhamento de sinais é desprezível e o ruído pode ser modelado por uma distribuição gaussiana. Neste cenário, ambos MF e OF2 operam próximo ao ponto ótimo. Para ocupâncias altas, o método COF apresenta melhor desempenho quando comparado aos algoritmos MF e OF2. Em condições de ocupâncias muito altas, próximas a 50%, os métodos COF e MF apresentaram desempenhos semelhantes. Entretanto, vale ressaltar que nenhuma informação a respeito da luminosidade é necessária para o projeto do COF. Além disso, sob alta ocupância, os sinais fora da janela de leitura começam a contribuir significativamente e a Equação (6.12) deixa de ser válida.

Sobre o uso de formas de pulso diferentes, os resultados utilizando o pulso assimétrico apresentou desempenho pior. Isto pode ser explicado pela presença de uma cauda positiva longa, sofrendo mais do efeito de empilhamento de sinais. Porém, vale ressaltar que o método COF apresentou melhor desempenho independente da forma do pulso utilizado.

No que tange a eficiência de estimação sob condições diferentes de SPR , a Figura 6.10 mostra o resultado para um nível fixo de ocupância de 15%. Pode ser observado que quanto maior o valor de SPR , menor é o erro de estimação para todos os métodos considerados. Novamente, o método COF supera as técnicas MF e OF2 independente do nível de empilhamento de sinais e da forma do pulso.

6.5 COF aplicado a dados experimentais do TileCal

Os métodos COF, MF e OF2 estão atualmente implementados no pacote de reconstrução do ATLAS [82]. O método OF2 foi o algoritmo padrão de reconstrução, *online* e *offline* da energia do TileCal até meados de 2014. Para o Run2, que se iniciou este ano, o método MF passou a ser utilizado para reconstrução de energia *online*, com atualização da matriz de covariância obtida através do método robusto, uma vez que o nível de empilhamento de sinais será maior. Atualmente, o COF está em fase de validação junto à colaboração, podendo ser utilizado para reconstrução *offline* da energia no TileCal e ser testado na eletrônica nova para o program de atualização do ATLAS [83].

Esta seção apresenta resultados da aplicação do COF em dados reais adquiridos durante operação nominal do LHC. Para efeito de comparação, os resultados para o algoritmo OF2, utilizado atualmente para reconstrução *offline* de energia, também são mostrados.

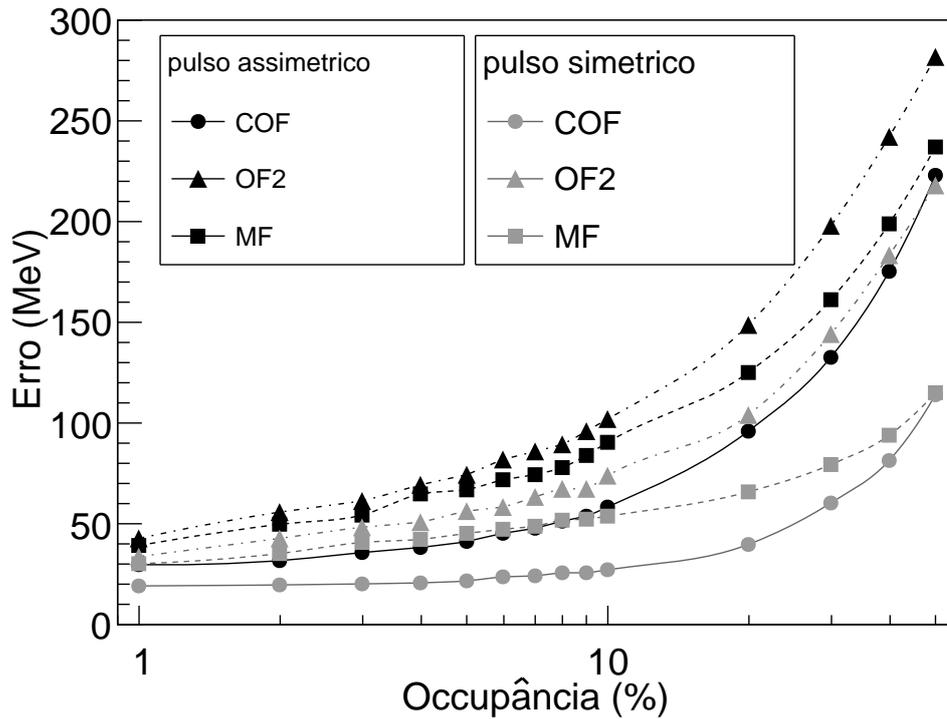


Figura 6.9: Erro de estimação (RMS) em função da ocupância para o conjunto de sinal considerando um valor fixo de $SPR=10$. O desempenho para ambos os pulsos simétrico (na cor preta) e assimétrico (na cor cinza) é mostrado.

6.5.1 Análise do erro de estimação

Quando se utiliza dados reais, a informação dos valores de referência não é acessível, por isso a análise da subtração absoluta entre os valores de referência e os valores estimados não pode ser feita. Entretanto, uma forma de analisar o erro de estimação, é verificando o comportamento dos métodos em torno da região de ruído. Além disso, a forma da cauda negativa da distribuição também fornece indicativos sobre o desempenho do método.

As figuras 6.11(a) e 6.11(b) mostram a distribuição de energia utilizando todas as células de leitura do TileCal para cada um dos métodos em escala linear e logarítmica, respectivamente. Estes dados foram adquiridos em 2012, onde os feixes foram colididos na energia de 8 TeV e com espaçamento entre colisões de 25 ns. O número médio de interações por colisão ($\langle \mu \rangle$) foi de 11,3. Pode ser observado que a distribuição do COF apresenta menor dispersão na região do ruído do TileCal (± 200 MeV).

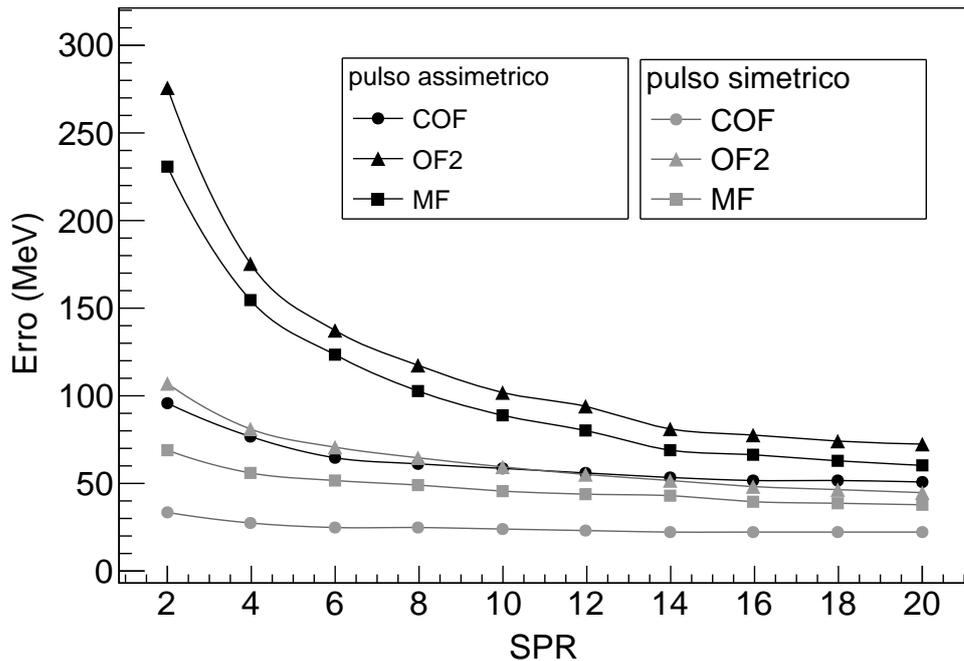


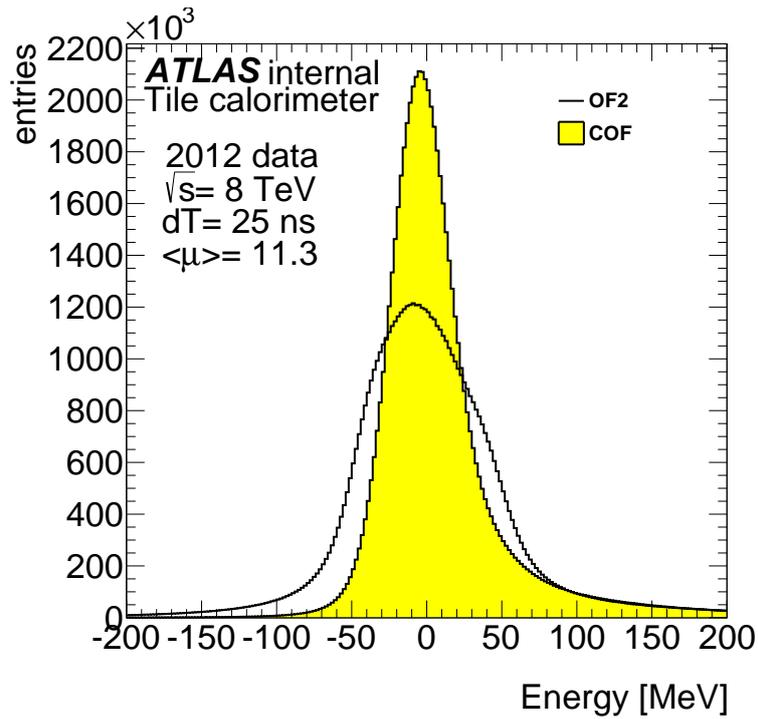
Figura 6.10: Erro de estimação em função do SPR para o conjunto de sinal considerando uma ocupância fixa de 10%. O desempenho para ambos os pulsos simétrico (na cor preta) e assimétrico (na cor cinza) é mostrado.

6.5.2 Linearidade do COF

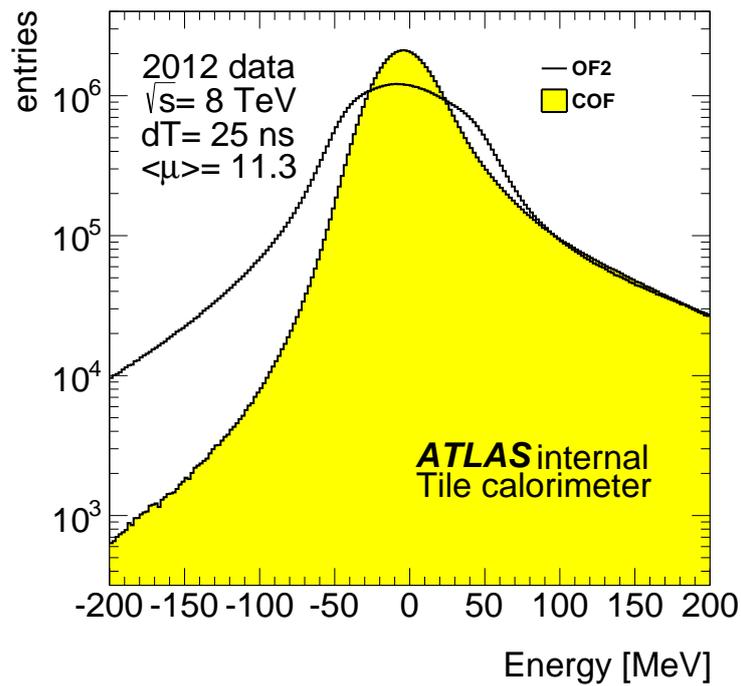
O método OF2 já é bastante estabelecido dentro do TileCal, fornecendo um comportamento linear ao longo de toda a faixa de energia. A Figura 6.12 mostra a correlação entre as energias estimadas pelo OF2 e COF. Pode ser notado que os métodos estão fortemente correlacionados para sinais de alta energia, indicando o comportamento linear do método COF.

Esta figura também fornece uma informação interessante sobre a tendência dos métodos na presença de empilhamento de sinais. Para o caso do COF, uma pequena tendência destacada por $|BC| > 3$ é observada devido à presença de sinais fora da janela de leitura. Já para o OF2, pode ser notado que este método apresenta uma tendência a medir valores de energia negativos sob presença de sinais empilhados localizado nos BC ± 2 e ± 3 . Já na presença de sinais empilhados no BC ± 1 , o método OF2 tende a superestimar a energia.

A característica da tendência do OF2 pode ser explicada pelos valores de seus coeficientes. A Figura 6.13 ilustra os coeficientes do OF2 para um determinado canal do TileCal. Note que para um sinal empilhado, localizado no BC ± 2 e ± 3 (primeiro, segundo, sexto e sétimo coeficientes), este método tende a subestimar a energia, chegando a fornecer valores negativos, o que explica a cauda negativa das distribuições de energia do OF2. Já para sinais empilhados em BC ± 1 (terceiro e



(a)



(b)

Figura 6.11: Espectro de energia reconstruído pelo COF e OF2 em que o eixo vertical está na escala (a) linear e (b) logarítmica (extraído de [84]).

quinto coeficientes) a tendência é positiva.

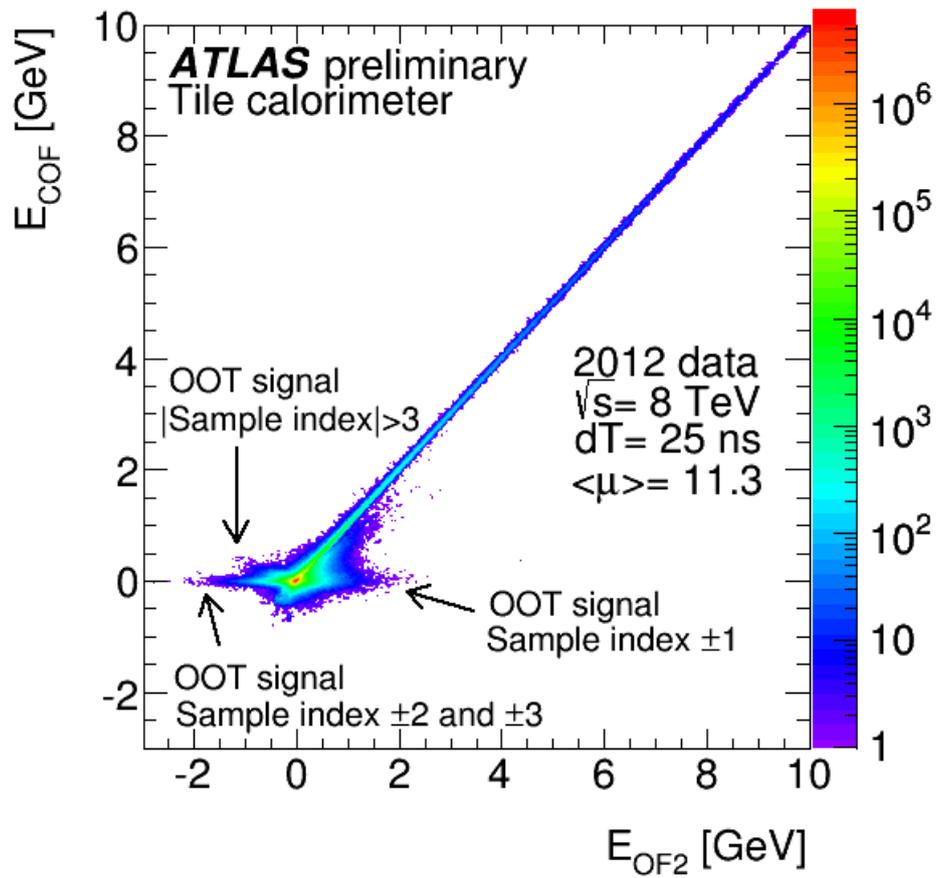


Figura 6.12: Correlação entre as energias estimadas pelos métodos COF e OF2 (extraído de [84]).

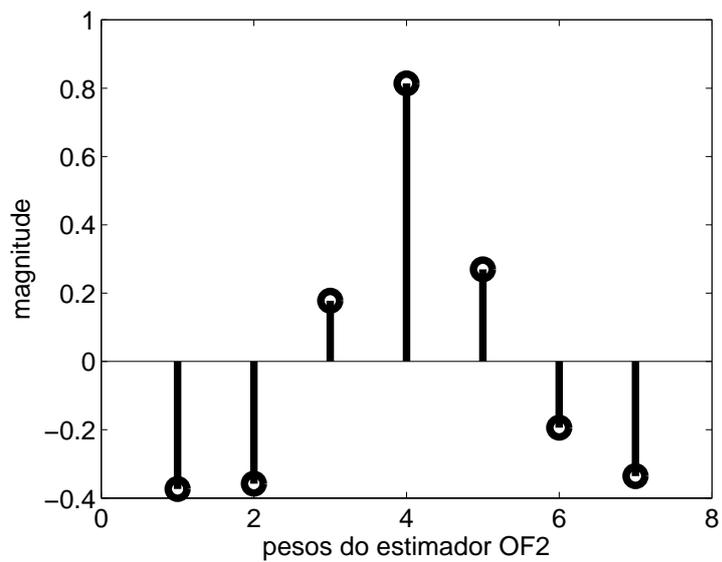


Figura 6.13: Pesos (coeficientes) do método OF2.

6.5.3 Desempenho de reconstrução do sinal

Para os métodos MF e OF2, a reconstrução do sinal é feita baseando-se somente na medida da amplitude do pulso central. No caso do método COF, as amplitudes dos sinais adjacentes são incluídas na reconstrução do sinal original fornecendo, assim, uma reconstrução mais fiel. A Figura 6.14 mostra a reconstrução de um sinal real recebido em condições de empilhamento de sinais. Pode ser observado que o COF apresenta melhor *fitting* nas amostras recebidas que o OF2, devido ao uso das informações (amplitudes) de todos os sinais presentes na janela de leitura.

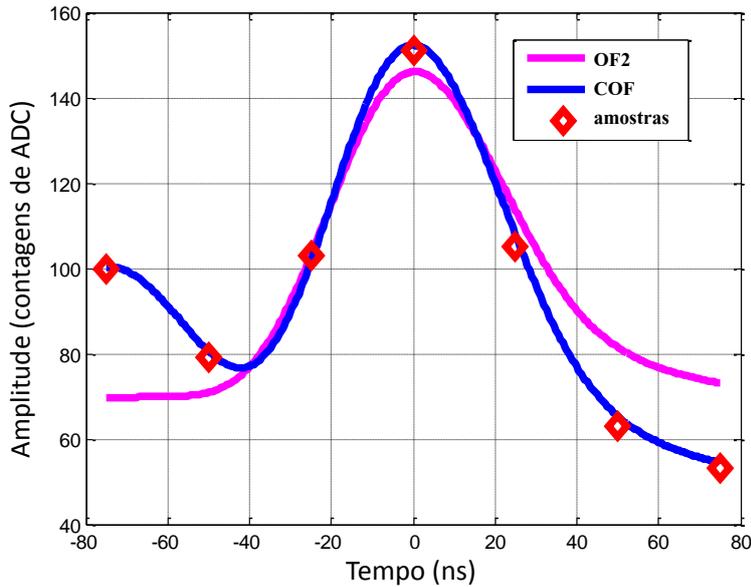


Figura 6.14: Reconstrução do sinal a partir das amostras digitais em condições de empilhamento de sinais.

No caso do COF, as amostras do pulso recebido são reconstruídas através da soma das amostras das formas do pulso normalizadas para cada BC, multiplicada pela amplitude estimada associada ao respectivo BC

$$\mathbf{reco} = \sum_{k=1}^N \hat{a}_k \mathbf{g}_k, \quad (6.17)$$

onde \hat{a}_k corresponde às amplitudes para cada BC k e, \mathbf{g}_k , corresponde às amostras do pulso de referência k . Para o caso do TileCal, $N=7$. No caso do OF2 e MF, somente \hat{a}_4 é disponível, e para as outras estimativas é atribuído o valor zero.

Para quantificar a qualidade na reconstrução do sinal recebido, a raiz quadrada do erro médio quadrático entre as amostras recebidas e as amostras reconstruídas é comumente utilizada como medida de qualidade (QF, *Quality Factor*) [79] [78].

O valor do QF pode ser calculado através da seguinte expressão:

$$QF = \frac{1}{N} \sqrt{\sum_{k=1}^N (x[k] - \text{reco}[k])^2} \quad (6.18)$$

em que $x[k]$ corresponde as amostras temporais recebidas. Os valores de QF para os métodos MF e OF2 tendem a ser altos sob presença de empilhamento de sinais, mesmo que a estimativa da amplitude de interesse (sinal central) seja boa. Desta forma, para estes estimadores, o QF é mais apropriado para a detecção de empilhamento de sinais do que para quantificar a qualidade da reconstrução do sinal.

A Figura 6.15 mostra a correlação entre os valores de QF calculados pelo COF e OF2. Valores altos de QF para a reconstrução pelo OF2 podem ser observados. Estes eventos correspondem a sinais reconstruídos na presença de um ou mais sinais empilhados. Para os mesmos sinais, o COF apresentou valores baixos de QF, indicando a boa reconstrução do sinal recebido.

Além disso, a reconstrução do sinal pelo método COF se mostrou robusta à presença do empilhamento de sinais. Desta forma, o QF do COF pode ser utilizado para detectar sinais corrompidos (*data integrity issues*, como é destacado na figura, função primordial do QF no TileCal. Estes eventos correspondem a pulsos que perderam a forma do sinal de referência devido à algum comportamento indesejado tais como problemas na eletrônica ou saturação. Os sinais identificados como problemáticos podem ser descartados durante a reconstrução do evento, melhorando o desempenho da mesma.

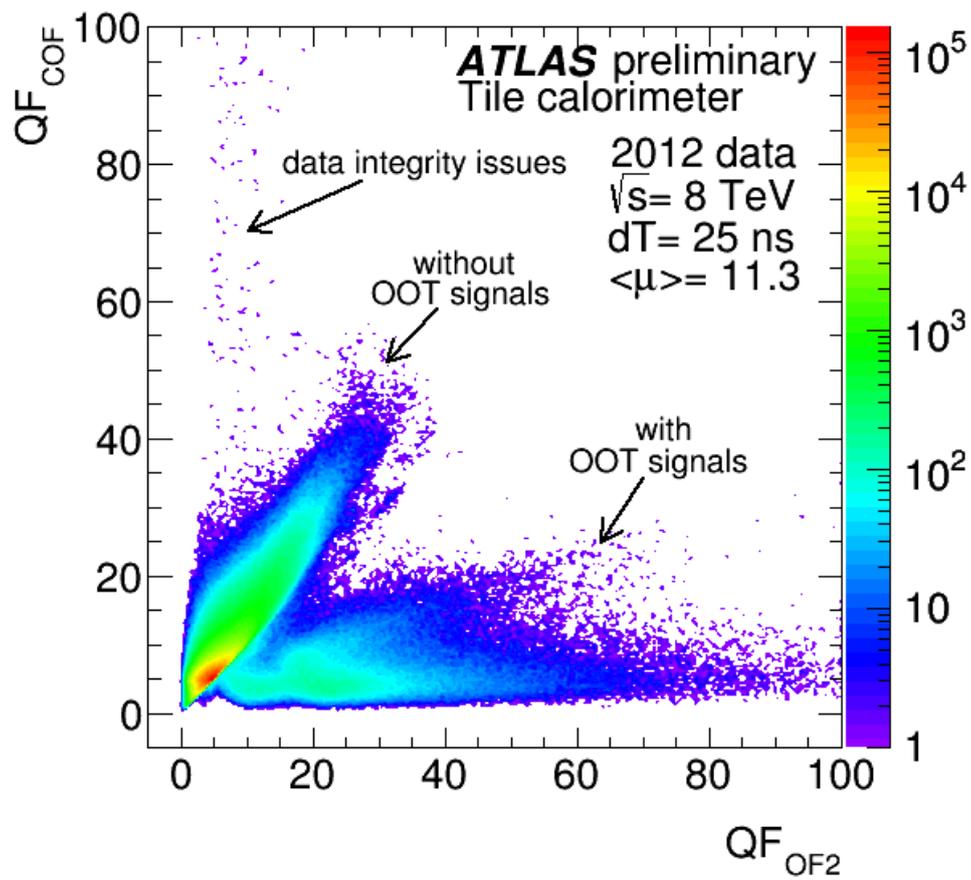


Figura 6.15: Correlação entre os fatores de qualidade dos métodos COE e OF2 (extraído de [84]).

Capítulo 7

Conclusões

Esta tese abordou o problema da reconstrução de energia em calorímetros utilizados em experimentos de física de altas energias. Esta reconstrução pode ser dividida em duas etapas: a identificação dos canais com informação relevante e a estimação da energia depositada nos canais. A energia é a informação mais relevante nos sistemas de calorimetria e esta informação é fundamental para a reconstrução final do evento e identificação das partículas. É através da identificação das partículas que os modelos para fenômenos físicos conhecidos e desconhecidos são testados e validados.

Nesta tese, de forma inovadora, o problema de reconstrução de energia é formulado, inicialmente, como um problema de detecção binária (sinal/ruído) e a estimação é obtida como consequência do processo de detecção. O método proposto é baseado em filtragem casada, sendo ótimo do ponto de vista da maximização da relação sinal-ruído. A estimação da energia é obtida através de uma simples calibração da saída do algoritmo de detecção. Portanto, a identificação dos canais com informação relevante e a estimação da energia são realizadas concomitantemente.

Em condições aonde a forma do pulso proveniente do calorímetro não pode ser considerada fixa, devido à variações ligadas à eletrônica e/ou aos processos físicos de deposição de energia nos canais do calorímetro, a solução analítica para a razão de máxima verossimilhança é muitas vezes impossível de ser obtida. Para estas condições, nesta tese é proposta a aproximação do processo aleatório pela série de Kerhunen-Loève, gerando assim um banco de filtros para detecção e estimação da energia nos canais do calorímetro. Esta abordagem, apesar de não ser ótima caso o processo aleatório não seja gaussiano, apresenta desempenho superior às abordagens que consideram o pulso da calorimetria fixo, conforme foi demonstrado nas simulações realizadas neste trabalho.

Em ambientes onde as deformidades do pulso recebido podem ser desprezadas, e o sinal de interesse é corrompido puramente por ruído gaussiano, o método proposto resulta no conhecido filtro casado de detecção, ou filtro correlator, apresentando

uma implementação simplificada através de um filtro de resposta ao impulso finita. Sob tais condições, o método proposto opera em condições ótimas, se igualando ao estimador de máxima verossimilhança (MLE).

O método MF desenvolvido nesta tese, foi aplicado no calorímetro de telhas (TileCal) do ATLAS no LHC. Para avaliar e comparar a seu desempenho com o método padrão utilizado pelo TileCal (chamado de OF), foram utilizados dados simulados e dados reais, adquiridos durante operação nominal do LHC. O método MF se mostrou superior no que tange a detecção de sinais e o erro na estimação, selecionando melhor os canais com efetiva deposição de energia, e reduzindo o erro de estimação em aproximadamente 20%. Foi mostrado que a perda de desempenho do OF nestas condições pode ser atribuída à inclusão de restrições adicionais no seu processo de otimização.

Vale ressaltar que, devido à alta taxa de eventos do LHC, o TileCal observou o efeito de empilhamento já durante o Run1. Sob tais condições o ruído de fundo pode ser entendido como a convolução das distribuições gaussianas (proveniente do ruído eletrônico) e log-normal (modelo aproximado do efeito de empilhamento de sinais). Foi mostrado que, neste cenário, ambos métodos MF e OF perdem desempenho, uma vez que a estatística de empilhamento de sinais não é incluída no projeto dos estimadores. Diferentemente do método OF, a descrição correta do ruído de fundo pode ser incorporada ao projeto do MF, embora seja provável que esta consideração leve a estimadores complexos, não-lineares e de difícil implementação.

Entretanto, o estimador MF prevê o uso da matriz de covariância em seu projeto sem aumento de complexidade. Embora em condições de empilhamento de sinais a estatística de segunda ordem não descreva mais o ruído de fundo corretamente, o uso da matriz de covariância, corretamente estimada, pode reduzir as incertezas introduzidas pelo empilhamento de sinais nas estimativas da energia. Foi mostrado que, através deste procedimento, o erro de estimação do MF é reduzido em mais de 30% em células de alta ocupância (alta probabilidade de empilhamento de sinais) para as condições atuais de operação do LHC (segunda tomada de dados, chamada de Run2).

Assim, o método MF, através do uso da matriz de covariância, foi validado para ser utilizado *online* durante o período de colisões Run2 do LHC, que se iniciou em Abril de 2015. Tendo em vista o programa de atualização do LHC, que prevê um aumento em uma ordem de grandeza da sua luminosidade (o que aumenta a probabilidade de ocorrer empilhamento de sinais), uma outra abordagem foi proposta nesta tese e está sendo analisada para o cenário de alta luminosidade do LHC.

7.1 Perspectivas para o cenário de empilhamento de sinais

O método proposto MF apresenta uma boa solução de detecção de sinais e estimação de energia para a grande maioria dos calorímetros que não sofrem de empilhamento de sinais. Além disso, sob alta luminosidade, o MF apresenta uma solução razoável para o problema de estimação, através do uso da matriz de covariância.

Entretanto, para casos particulares onde a luminosidade é muito alta, como no caso do programa de atualização do LHC, esta tese de doutorado também abordou a busca de uma solução de baixa complexidade e alto desempenho para a estimação da energia sob tais condições. Em colisionadores de partículas, como o caso do LHC, as colisões acontecem em intervalos iguais de tempo. Desta forma, os sinais são sobrepostos linearmente em tempos conhecidos. O empilhamento de sinais ocorre devido à resposta do sistema de calorimetria (como é o caso do TileCal) ser mais lenta que a taxa de eventos, e mais de um sinal pode ser observado numa mesma janela de leitura.

Desta forma, foi proposto o projeto de um estimador, chamado de COF (*Constrained Optimal Filter*), que trata o sinal recebido como uma mistura linear de sinais defasados. Ou seja, uma matriz de desconvolução linear é projetada baseada no pulso de referência, e recupera os sinais presentes numa determinada janela de leitura. Desta forma, o empilhamento de sinais não é tratado como ruído e o ruído de fundo possui contribuição somente do ruído eletrônico do experimento, o qual, em muitos casos, pode ser modelado por uma distribuição gaussiana. Assim, foi mostrado que o projeto do estimador se torna independente da luminosidade do experimento.

Resultados da aplicação do estimador COF no TileCal mostraram que a presença de sinais empilhados degrada o desempenho de estimação do método OF, e o COF apresentou melhor resolução de energia (erro de estimação) do que o OF. Além disso, foi destacado que a qualidade da reconstrução do sinal recebido é melhorada consideravelmente sob condições de empilhamento de sinais, uma vez que o COF disponibiliza a amplitude de todos sinais presentes na janela de leitura.

7.2 Trabalhos futuros

Como propostas de trabalhos futuros, a avaliação do impacto da estimação da energia pelos métodos propostos nas análises da física se encontra em andamento. Juntamente, outras abordagens mais sofisticadas de estimação de energia em condições de alto nível de empilhamento de eventos estão sendo consideradas.

7.2.1 Impacto na física

A informação final dos sistemas de calorimetria utilizada pelos físicos para estudarem os fenômenos gerados pelas colisões, se baseiam na reconstruções de objetos mais abstratos, como jatos, no caso do TileCal. Um dado evento (colisão) pode conter vários objetos que são reconstruídos durante as análises físicas. Estes objetos são reconstruídos utilizando a informação espacial da distribuição da energia das células dentro do calorímetro finamente segmentado. Desta forma, a eficiência da estimação da energia depositada na célula impacta diretamente na qualidade (resolução) destes objetos, uma vez que células sem informação relevante podem ser consideradas na reconstrução do evento.

7.2.1.1 Avaliação no ambiente de desenvolvimento

Durante a primeira tomada de dados do LHC (chamada de Run1), foi mostrado que, no TileCal, o empilhamento de sinais degradou a eficiência da medida da energia pelo método utilizado, comprometendo a reconstrução destes objetos. Tendo em vista o melhor desempenho dos métodos propostos em condições de empilhamento de sinais, mostrada nesta tese, algumas análises se encontram em andamento com o objetivo de avaliar a qualidade da reconstrução de jatos no TileCal, e comparar com o método utilizado durante o Run1 do LHC.

7.2.2 Novos métodos para cenário de alto nível de empilhamento de eventos

Em ambientes de alto nível de empilhamento de sinais, como proposto para o HL-LHC (do inglês, *High-Luminosity LHC*) com previsão de início de operação para 2026, novas abordagens de detecção e estimação podem ser consideradas para a reconstrução de energia em calorímetros operando sob tais condições. Contudo, métodos mais sofisticados podem demandar um esforço computacional e complexidade mais alto, muitas vezes inviável para aplicações *online*.

Sendo assim, como continuidade desta tese, os seguintes métodos são propostos para a estimação de energia em calorímetros operando sob condições de alto nível de empilhamento de eventos:

- Corretor não-linear: uma rede neural [85], se treinada corretamente, pode ser usada como um corretor na medida linear da amplitude pelo método proposto MF. Tal abordagem já foi utilizada com sucesso em outros contextos[86], [87]. Nesta proposta, o processamento não-linear tem a função de corrigir o modelo linear o qual não trata as não-linearidades presentes no ruído de fundo em condições de empilhamento de sinais. Ou seja, o corretor não estima a amplitude

do pulso de entrada, mas fornece um ajuste fino a estimativa linear já validada. Este tipo de abordagem tem a vantagem de a saída linear, sem correção, ser disponível e ficar a critério do usuário usá-la ou não. Além disso, a capacidade de generalização da rede neural bem treinada torna o método independente da luminosidade sem a necessidade de treinamento para diferentes condições de empilhamento de sinais.

- Representação esparsa: foi mostrado que o uso da matriz de desconvolução do método proposto COF é limitado a sinais que estejam inteiramente contidos na janela de leitura. Ou seja, em ambientes de alto nível de empilhamento de sinais, a presença de sinais empilhados fora da janela de aquisição introduz uma componente adicional ao ruído de fundo degradando o desempenho do algoritmo. Portanto, uma técnica que considere uma janela mais ampla, baseada na representação esparsa de sistemas lineares [88], pode ser testada no contexto de calorimetria de altas energias. Tal abordagem pode ser eficientemente implementada através de algoritmos de programação linear [89].
- Estimador de Máxima Verossimilhança (MLE): se o efeito de empilhamento de sinais for corretamente modelado, este parâmetro pode ser utilizado no projeto de um estimador MLE. Este estimador utiliza a função densidade de probabilidade para calcular seus coeficientes. Entretanto, esta abordagem pode não acarretar em uma fórmula fechada resultando em um processo iterativo de estimação da amplitude, o que dificulta sua implementação. Porém, dada a capacidade e eficiência do uso de redes neurais em estimar uma função densidade de probabilidade complexa [90], um estimador baseado em redes neurais pode ser projetado com a finalidade de aproximar o estimador MLE ótimo para o cenário de empilhamento de sinais. Neste método, puramente neural, a rede neural é projetada para estimar diretamente a amplitude do sinal a partir das amostras digitais recebidas.

Referências Bibliográficas

- [1] <http://www.cern.ch> (acessado em julho de 2015).
- [2] COTTINGHAM, W. N., GREENWOOD, D. A. *An Introduction to the Standard Model of Particle Physics*. Cambridge University Press, 1998.
- [3] EDWARDS, D. A., SYPHERS, M. J. *An Introduction to the Physics of High Energy Accelerators*. Wiley-Interscience, 1992.
- [4] HASERT, F. “Search for Elastic muon-neutrino Electron Scattering”, *Physics Letters B*, v. 46, n. 1, pp. 121–124, 1973.
- [5] ARNISON, G., ET AL. “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{540}$ GeV”, *Physics Letters B*, v. 122, n. 1, pp. 103–116, 1983.
- [6] CERN. “Aerial view of CERN site of Meyrin and Globe of Innovation”. , maio 2012. Disponível em: <<https://cds.cern.ch/record/1476896/>>.
- [7] EVANS, L., BRYANT, P. “LHC Machine”, *Journal of Instrumentation, JINST 3 S08001*, 2008.
- [8] CERN. “The CERN accelerator complex”. , out. 2013. Disponível em: <<https://cds.cern.ch/record/1621894/>>.
- [9] K.ELSENER. *The CERN Neutrino Beam To GRAN SASSO*. Cern-98-02, infn/ae-98/05, CERN, 1998.
- [10] BIRD, I., ET AL. *LHC Computing Grid*. Lcg-tdr-001, cern-lhcc-2005-024, CERN, <http://cern.ch/lcg/tdr>, 2005.
- [11] THE ATLAS COLLABORATION. “The ATLAS Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation, JINST 3 S08003*, 2008.
- [12] THE CMS COLLABORATION. “The CMS experiment at the CERN LHC”, *Journal of Instrumentation, JINST 3 S08004*, 2008.

- [13] THE LHCb COLLABORATION. “The LHCb Detector at the LHC”, *Journal of Instrumentation*, *JINST 3 S08005*, 2008.
- [14] THE ALICE COLLABORATION. “The ALICE experiment at the CERN LHC”, *Journal of Instrumentation*, *JINST 3 S08002*, 2008.
- [15] SÁNCHEZ, C. A. S. *Implementation of the ROD Crate DAQ Software for the ATLAS Tile Calorimeter and a Search for a MSSM Higgs Boson Decaying into Tau Pairs*. Tese de Ph.D., (<https://cds.cern.ch/record/1309926>) Universitat de València - CSIC, Valência, Espanha, 2010.
- [16] ROS, E. “ATLAS inner detector”, *Nuclear Physics B - Proceedings Supplements*, v. 120, pp. 235–238, 2003.
- [17] THE ATLAS COLLABORATION. *Calorimeter Performance*. Technical Design Report, CERN/LHCC/96-40, 1997.
- [18] PALESTINI, S. “The Muon Spectrometer of the ATLAS Experiment”, *Nuclear Physics B*, v. 125, pp. 337–345, 2003.
- [19] CERN. “Computer generated image of the whole ATLAS detector”. , mar. 2008. Disponível em: <<https://cds.cern.ch/record/1095924/>>.
- [20] THE ATLAS COLLABORATION. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Physics Letters B*, v. 716, pp. 1–29, 2012.
- [21] THE ATLAS COLLABORATION. *Level-1 Trigger*. Technical Design Report, ATLAS TDR-12, 1998.
- [22] KORDAS, K., ABOLINS, M. “The ATLAS Data Acquisition and Trigger: concept, design and status”, *Nuclear Physics B - Proceedings Supplements*, v. 172, pp. 178–182.
- [23] ARMSTRONG, S., BAINES, J. T., BEE, C. P., et al. “Algorithms for the ATLAS High-Level Trigger”, *IEEE Transactions on Nuclear Science*, v. 51, 2004.
- [24] LEO, W. R. *Techniques for Nuclear and Particle Physics Experiments*. Springer-Verlag, 1994.
- [25] DAS, A., FERBEL, T. *Introduction to Nuclear and Particle Physics*. John Wiley and Sons, 2003.

- [26] WIGMANS, R. *Calorimetry - Energy Measurement in Particle Physics*. Oxford University Press, 2000.
- [27] THE ATLAS COLLABORATION. *Muon Spectrometer*. Technical Design Report, CERN/LHCC/97-22, 1997.
- [28] CERN. “Computer Generated image of the ATLAS calorimeter”. , mar. 2008. Disponível em: <<https://cds.cern.ch/record/1095927/>>.
- [29] PERRODO, P. “The ATLAS liquid argon calorimetry system”, *Proceedings of ICHEP*, pp. 909–912, 2002.
- [30] FRANCAVILLA, P. “The ATLAS Tile Hadronic Calorimeter performance at the LHC”, *J. Phys.: Conf. Ser.*, v. 404, pp. 012007, 2007.
- [31] TILECAL COLLABORATION. *ATLAS Tile Calorimeter Technical Design Report*. Atlas tdr 3, cern/lhcc/96-42, CERN, Geneva, 1996.
- [32] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTile> (acesado em agosto de 2015).
- [33] ANDERSON, K. “Design of the front-end analog electronics for the ATLAS tile calorimeter”, *Nuclear Instruments and Methods in Physics Research A*, v. 551, pp. 469–476, 2005.
- [34] TYLMAD, M., ET AL. “Pulse shapes for signal reconstruction in the ATLAS Tile Calorimeter”, *16th IEEE-NPSS*, pp. 543–547, 2009.
- [35] PERALVA, B., ET AL. “TileCal Energy Reconstruction for Collision Data Using the Matched Filter”, *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 1–6, 2013.
- [36] CERQUEIRA, A. “ATLAS Tile Calorimeter Readout Electronics Upgrade Program for the High Luminosity LHC”, *Proceedings of LISHEP*, 2013.
- [37] ANDERSON, K., ET AL. “Design of the front-end analog electronics for the ATLAS Tile Calorimeter”, *Nucl. Instrum. Meth.*, p. A(551):469476, 2005.
- [38] VIRET, S. “LASER monitoring system for the ATLAS Tile Calorimeter”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 617, pp. 120 – 122, 2010.
- [39] STARICHENKO, E. “Cesium Monitoring System for ATLAS Tile Hadron Calorimeter”, *Nucl. Instrum. Meth.*, v. A494, pp. 381 – 384, 2002.

- [40] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TileCaloPublicResults>
(acessado em agosto de 2015).
- [41] RUGGIERO, F. “LHC Accelerator R&D and Upgrade Scenarios”, *The European Physical Journal C - Particles and Fields*, v. 34, pp. 433–442, 2004.
- [42] VANKOV, P. “ATLAS Upgrades Towards the High Luminosity LHC: extending the discovery potential”, *European Physical Society Conference on High Energy Physics*, 2013.
- [43] KLIMEK, P. “ATLAS Tile Calorimeter Data Quality Assessment with Commissioning Data”, *15th International Conference on Calorimetry in High Energy Physics*, link: <https://cdsweb.cern.ch/record/1473499>, 2012.
- [44] TANG, F. “Upgrade Analog Readout and Digitizing System for ATLAS TileCal Demonstrator”, *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 2013.
- [45] MEYER-BAESE, U. *Digital Signal Processing with Field Programmable Gate Arrays*. Springer, 2007.
- [46] KAY, S. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Pearson, 1993.
- [47] THEODORIDIS, S. *Machine Learning: A Bayesian and Optimization Perspective*. Academic press, 2015.
- [48] CLELAND, W., STERN, E. “Signal processing considerations for liquid ionization calorimeters in a high rate environment”, *Nucl. Instr. and Meth. in Phys. Res. A*, v. 338, pp. 467–497, 1994.
- [49] FULLANA, E., ET AL. “Digital Signal Reconstruction in the ATLAS Hadronic Tile Calorimeter”, *IEEE Transaction On Nuclear Science*, v. 53, n. 4, pp. 2139–2143, 2006.
- [50] PAGANINI, P., VAN VULPEN, I. “Pulse amplitude reconstruction in the CMS ECAL using the weights method”. , out. 2004. Disponível em: <http://cdsweb.cern.ch/record/801816/files/note04_025.pdf>.
- [51] PEEBLES, P. *Probability, Random Variables, and Random Signal Principles*. McGraw-Hill, 2000.
- [52] CALDWELL, A., ET AL. “Design and Implementation of a High Precision Readout System for the ZEUS Calorimeter”, *Nuclear Instruments and Methods in Physics Research A*, v. 321, n. 1-2, pp. 356–364, 1992.

- [53] ABAZOV, V., ET AL. “The Upgraded D0 Detector”, *Nuclear Instruments and Methods in Physics Research A*, v. 565, n. 2, pp. 463–537, 2006.
- [54] TYLMAD, M., ET AL. “Pulse shapes for signal reconstruction in the ATLAS Tile Calorimeter”, *Proc. 16th IEEE-NPSS Real Time Conference RT’09*, pp. 543–547, 2009.
- [55] BERTUCCIO, G., GATTI, E., SAPIETRO, M. “Sampling and optimum data processing of detector signals”, *Nuclear Instruments and Methods in Physics Research A*, v. 322, pp. 271–279, 1992.
- [56] CLELAND, W., STERN, E. “Signal processing considerations for liquid ionization calorimeters in a high rate environment”, *Nuclear Instruments and Methods in Physics Research A*, v. 338, pp. 467–497, 1994.
- [57] DELMASTRO, M., ET AL. “A stand-alone signal reconstruction and calibration algorithm for the ATLAS electromagnetic calorimeter”, *IEEE Nuclear Science Symposium Conference Record*, v. 2, pp. 1110–1114, 2003.
- [58] FULLANA, E. “Digital Signal Reconstruction in the ATLAS Hadronic Tile Calorimeter”, *IEEE Transaction On Nuclear Science*, v. 53, pp. 2139 – 2143, 2006.
- [59] PAPOULIS, S. *Probability, Random Variable and Stochastic Processes*. McGraw Hill, 1991.
- [60] CERN. “Liquid Argon Calorimeter Public Plots”. , set. 2015. Disponível em: <<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResultsDetStatus>>.
- [61] CHATRCHYAN, S., ET AL. “Time Reconstruction and Performance of the CMS Electromagnetic Calorimeter”, *Journal fo Instrumentation*, v. 5, n. T03011, 2010.
- [62] TREES, H. L. V. *Detection, Estimation and Modulation Theory, Part I*. Wiley, 2001.
- [63] MITRA, S. K. *Digital Signal Processing: A Computer-Based Approach*. 1998.
- [64] CHAPMAN, J. “ATLAS Simulation Computing Performance and Pile-Up Simulation in ATLAS”, *LPCC Detector Simulation Workshop - CERN*, 2011.

- [65] CALDWELL, A., ET AL. “Design and Implementation of a High Precision Readout System for the ZEUS Calorimeter”, *Nuclear Instruments and Methods in Physics Research A*, v. 321, n. 1-2, pp. 356–364, 1992.
- [66] MULLER, H., ET AL. “Front-End Electronics for PWO-based PHOS Calorimeter of ALICE”, *Nuclear Instruments and Methods in Physics Research A*, v. 567, pp. 264–267, 2006.
- [67] LUNDBERG, O., ET AL. “Calibration Systems of the ATLAS Tile Calorimeter”, *Proceedings of 32nd Symposium on Physics in Collision*, 2012.
- [68] JOLLIFFE, I. T. *Principal Component Analysis*. Springer, segunda edição, 2002.
- [69] SHANMUGAN, K. S., BREIPOHL, A. *Random Signals - detection, estimation and data analysis*. Wiley, 2001.
- [70] FILHO, L. M. A., SEIXAS, J. M. “A ROOT Tool for 3D Event Visualization in ATLAS Calorimeters”, *Proceeding of Science, PoS(ACAT)069*, 2007.
- [71] LIMA, G., PROUDFOOT, J. “Studies of cell response using occupancy rates measured in Tile Calorimeter ZeroBias data with 50ns bunch-spacing”. , jun. 2013. Disponível em: <ATL-COM-TILECAL-2013-032(<https://cds.cern.ch/record/1556925>)>.
- [72] ROUSSEEUW, P., DRIESSEN, K. “A Fast Algorithm for the Minimum Covariance Determinant Estimator”, *Technometrics*, v. 41, n. 3, pp. 212–223, 1999.
- [73] CHAPMAN, J. “ATLAS Simulation Computing Performance and Pile-Up Simulation in ATLAS”, *LPCC Detector Simulation Workshop-CERN*, 2011.
- [74] PERALVA, B. “TileCal Public Plots”. , ago. 2014. Disponível em: <<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTileReconstruction>>.
- [75] CERN. “ATLAS Luminosity Public Plots”. , ago. 2010. Disponível em: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults#Luminosity_Plots_for_Specific_LH>.
- [76] WILSON, R. R. *The Tevatron*. Fermilab-tm-0763, Fermilab, 1978.
- [77] HAYKIN, S. *Adaptive Filter Theory*. Prentice Hall, 4th edition, 2001.

- [78] DELMASTRO, M., ET AL. “Quality Factor Analysis and Optimization of Digital Filtering Signal Reconstruction for Liquid Ionization Calorimeters”, *Nuclear Instruments and Methods in Physics Research A*, v. 600, n. 3, pp. 545–554, 2009.
- [79] CLEMENT, C., KLIMEK, P. “Identification of Pile-up Using the Quality Factor of Pulse Shapes in the ATLAS Tile Calorimeter”, *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 1188–1193, 2011.
- [80] OPPENHEIM, A. V. *Discrete-time Signal Processing*. Prentice Hall, 2010.
- [81] BANERJEE, S., ET AL. “CMS Simulation Software”, *Journal of Physics: Conference Series*, v. 396, n. 022003, 2012.
- [82] CRANMER, K. S. “The ATLAS Analysis Architecture”, *Nuclear Physics B - Proceedings Supplements*, v. 177-178, pp. 126–130, 2008.
- [83] CARRIO, F., ET AL. “The sROD demonstrator for the ATLAS Tile Calorimeter Upgrade”, *IEEE Nuclear Science Symposium and Medical Conference*, pp. 1512–1516, 2012.
- [84] PERALVA, B. “TileCal Public Plots”. , mar. 2015. Disponível em: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TileCaloPublicResults#Energy_Reconstruction.
- [85] HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1998.
- [86] MADSEN, P. “Neural network for combining linear and non-linear modelling of dynamic systems”, *IEEE World Congress on Computational Intelligence*, v. 7, pp. 4541–4546, 1994.
- [87] WANG, Z., ET AL. “Application of BP neural networks in non-linearity correction of optical tweezers”, *IEEE World Congress on Computational Intelligence*, v. 3, n. 4, pp. 475–479, 2008.
- [88] ELAD, M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.
- [89] LUENBERGER, D. *Linear and Nonlinear Programming*. Springer, 2003.
- [90] A. SARAJEDINI, R. H.-N., CHAU, P. “Conditional Probability Density Function Estimation with Sigmoidal Neural Networks”, *IEEE Transactions on Neural Networks*, v. 10, n. 2, pp. 231–238, 1999.

Apêndice A

Sistema de Coordenadas do ATLAS

O sistema de coordenadas usados em experimentos com feixes não é o sistema polar. É um sistema adequado ao formato cilíndrico dos detectores dispostos ao redor do ponto de impacto, ou seja, um sistema que acompanha a direção dos feixes de partículas provenientes da colisão. As coordenadas empregadas são η , ϕ e z em contraposição a x , y e z . Os termos η e ϕ seguem a uma transformação não-linear de x e y .

$$\phi = \arctan \frac{x}{y} \quad (\text{A.1})$$

$$\eta = -\log\left(\tan \frac{\phi}{2}\right) \quad (\text{A.2})$$

A Figura A.1 pode ser explicativa quanto ao sistema. Em sua parte superior, é possível ver um esquema do barril e da tampa de um detector, mostrando como se comportam as coordenadas tomando por referência as coordenadas cartesianas x , y e z (marcadas em pontilhado). Nota-se que a variável ϕ representa a rotação e a variável η (também chamada de pseudo-rapidez) representa a direção de projeção das partículas, após a colisão.

Os valores dados das variáveis η e ϕ são apenas para referência do leitor. A variável ϕ , como é possível ver no canto direito da parte superior da figura, possui uma região em que dois valores são possíveis: 0 e 2π . Esta área é chamada de região *wrap-around*. Cálculos utilizando esta variável devem atentar para este fato.

Os detectores são simétricos, com relação ao eixo ϕ . A construção dos dispositivos é feita em gomos.

Repara-se que quando alcança o eixo z , $\eta = 1$, isto significa que objetos com valores grandes em η representam colisões onde as partículas do feixe apenas se desviaram, não havendo, usualmente informações interessantes de análise pois re-

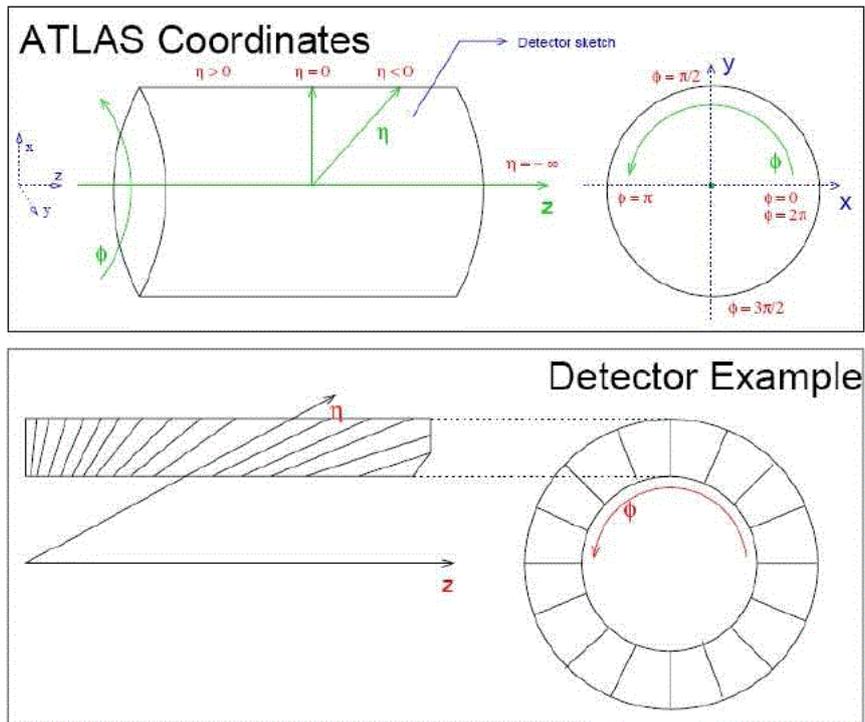


Figura A.1: O sistema de coordenadas do ATLAS.

presentam choques elásticos. É comum utilizar-se detectores com baixa resolução quando $\eta > 3$.

Na parte inferior da Figura B.1, é possível ver um exemplo de como um detector genérico é segmentado, acompanhando as coordenadas η e ϕ , tanto para o barril, quanto para uma tampa.

Apêndice B

Publicações

Os artigos, frutos deste trabalho, são listados juntamente com um breve resumo.

B.1 Trabalhos publicados em periódicos

1. Luciano M. A. Filho, Bernardo S. Peralva, José M. Seixas, Augusto S. Cerqueira. *Calorimeter Response Deconvolution for Energy Estimation in High-Luminosity Conditions*. Artigo aceito para publicação em IEEE Transactions on Nuclear Science.

Resumo: Signal superposition, also named pile-up noise, has become an important issue for calorimeters operating under high luminosity conditions. Current techniques model those superposed signals as a highly correlated noise, whose covariance matrix information is used to develop an Optimal Filter (OF), in order to minimize the pile-up contribution on the in-time signal amplitude estimation. Hence, the OF coefficients depend on the current luminosity. This paper presents an amplitude estimator design independent of the luminosity information. The method consists of unconvolving the front-end electronic shaping response, recovering the impulsive amplitude of each signal component. This technique results in orthogonal estimators for the amplitude of each superposed signal. This orthogonality constraint provides luminosity resilient estimator performance, suitable to high-luminosity-time-variant environments. Comparisons between the OF and the proposed method, using pile-up event simulation for the ATLAS Hadronic Calorimeter, are presented. The results also point out the combining the optimality characteristic of the OF techniques with the orthogonality constraint of the proposed method, an improvement on amplitude estimation performance is achieved and better quality-factor fittings are provided for different luminosity ranges.

2. (apresentado em nome da colaboração ATLAS) Bernardo S. Peralva, on

behalf of the ATLAS Collaboration, *The TileCal Online Energy Estimation for the Next LHC Operation Period*. Advanced Computing and Analysis Techniques, Praga, 2014. Publicado em: Journal of Physics, Conference Series, 2015.

Resumo: The technique to be employed for online energy estimation in TileCal for next LHC operation period must be based on fast filters such as the Optimal Filter (OF) and the Matched Filter (MF). Both the OF and MF methods envisage the use of the background second order statistics in its design, more precisely the covariance matrix. However, the identity matrix has been used to describe this quantity. Although this approximation can be valid for low luminosity LHC, it leads to biased estimators under pile-up conditions. Since most of the TileCal cell present low occupancy, the pile-up, which is often modeled by a non-Gaussian distribution, can be seen as outlier events. Consequently, the classical covariance matrix estimation does not describe correctly the second order statistics of the background for the majority of the events, as this approach is very sensitive to outliers. As a result, the OF (or MF) coefficients are miscalculated leading to a larger variance and biased energy estimator. This work evaluates the usage of a robust covariance estimator, namely the Minimum Covariance Determinant (MCD) algorithm, to be applied in the OF design. The goal of the MCD estimator is to find a number of observations whose classical covariance matrix has the lowest determinant. Hence, this procedure avoids taking into account low likelihood events to describe the background. It is worth mentioning that the background covariance matrix as well as the OF coefficients for each TileCal channel are computed offline and stored for both online and offline use. In order to evaluate the impact of the MCD estimator on the performance of the OF, simulated data sets were used. Different average numbers of interactions per bunch crossing and bunch spacings were tested. The results show that the estimation of the background covariance matrix through MCD improves significantly the final energy resolution with respect to the identity matrix which is currently used. Particularly, for high occupancy cells, the final energy resolution is improved by more than 20%. Moreover, the use of the classical covariance matrix degrades the energy resolution for the majority of TileCal cells.

3. **(apresentado em nome da colaboração ATLAS)** Bernardo S. Peralva, on behalf of the ATLAS Collaboration, *An Alternative Method for the TileCal Signal Detection and Amplitude Estimation*. Advanced Computing and Analysis Techniques, Londres, U.K., 2011. Publicado em: Journal of Physics, Conference Series, 2012.

Resumo: The Tile Barrel Calorimeter (TileCal) is the central section of the hadronic calorimeter of ATLAS. It is a key detector for the reconstruction of hadrons, jets, taus and missing transverse energy and it assists the muon measurements due to a low signal-to-noise ratio. The energy deposited in each cell is read out by two electronic channels for redundancy and is estimated by reconstructing the amplitude of the digitized signal pulse sampled every 25 ns. This work presents an alternative approach for TileCal signal detection and amplitude estimation under low signal-to-noise ratio (SNR) conditions, exploring the applicability of a Matched Filter. The proposed method is compared to the Optimal Filter algorithm, that is currently being used at TileCal for energy reconstruction. The results for a simulated data set showed that for conditions where the signal pedestal could be considered stationary, the proposed method achieves a better SNR performance than the Optimal Filter technique.

B.2 Trabalhos publicados em anais de congressos

1. Bernardo S. Peralva, Luciano M. A. Filho, Augusto S. Cerqueira, José M. Seixas, *Correção Não-linear na Estimação de Energia para um Calorímetro de Altas Energias*, Congresso Brasileiro de Inteligência Computacional (CBIC2015), Curitiba, 2015 (a ser apresentado).

Resumo: Em ambientes de alta taxa de eventos, a resposta do experimento pode ser distorcida devido à presença de informação proveniente de eventos adjacentes. Este efeito de empilhamento de sinais degrada a eficiência de métodos lineares utilizados para estimar os parâmetros de interesse. No contexto de física de altas energias, a energia do sinal de interesse é medida através da estimação da amplitude do pulso recebido do sistema de calorimetria. No momento, o LHC (do inglês, *Large Hadron Collider*) opera numa taxa de eventos que produz o empilhamento de sinais. Este trabalho apresenta um método alternativo para estimar a energia das células do Calorímetro de Telhas (TileCal) do detector ATLAS no LHC nestas condições de empilhamento. A abordagem proposta neste trabalho utiliza uma técnica linear baseada no estimador de máxima verossimilhança combinada com uma rede neural. A rede neural tem a função de corrigir as não-linearidades introduzidas pela estatística do empilhamento de sinais na estimativa linear. Os resultados mostram que o uso da abordagem proposta melhora consideravelmente a resolução de energia em condições de operação em alta taxa.

2. D. Gonçalves, Bernardo S. Peralva, Luciano M. A. Filho, Augusto S. Cerqueira,

Estimação da Energia do Sinal de Múons do Calorímetro Hadrônico do Experimento ATLAS, Simpósio Brasileiro de Telecomunicações (SBrT2015), Juiz de Fora, 2015 (a ser apresentado).

Resumo: Este trabalho apresenta o estudo, o desenvolvimento e a implementação da técnica mais eficiente para estimação de energia no contexto de um projeto relacionado com física experimental de altas energias. Técnicas de estimação encontradas na literatura foram aplicadas no contexto do projeto e comparadas. Os resultados para dados simulados e experimentais mostraram que o estimador de amplitude de pulso de máxima verossimilhança para o caso de ruído Gaussiano atinge a melhor eficiência em termos do erro de estimação, juntamente com a menor complexidade computacional, sendo o mais indicado para a aplicação.

3. Bernardo S. Peralva, Luciano M. A. Filho, Augusto S. Cerqueira, José M. de Seixas, *Reconstrução da Energia do Calorímetro de Telhas do ATLAS para o Run2 do LHC*, Encontro Nacional de Física de Partículas e Campos, Caxambu, Minas Gerais, 2015 (a ser apresentado).

Resumo: Este trabalho apresenta o algoritmo que será utilizado de forma online no TileCal durante o Run2. O método é chamado de Filtro Ótimo (OF), sendo baseado na mesma abordagem utilizada durante o Run1. A diferença entre as duas versões se concentra no tratamento da linha de base do sinal (pedestal) e o uso da estatística de segunda ordem do ruído de fundo, através da matriz de covariância. Na versão proposta para o Run2, o valor do pedestal será subtraído de cada amostra do sinal digitalizado recebido, e não estimado evento a evento, como feito no Run1. Vale ressaltar que o valor do pedestal é estimado através de tomadas de dados específicas, e armazenado em um banco de dados. Além disso, a matriz de covariância do ruído de fundo no projeto do OF pode ser utilizada em canais de alta ocupância com o objetivo de diminuir as incertezas introduzidas pelo empilhamento de sinais, mesmo que este fenômeno introduza componentes não-lineares no ruído de fundo. A eficiência dos métodos OF utilizados para o Run1 e Run2 é comparada sob diferentes condições de luminosidade, utilizando dados de Monte Carlo. Os resultados mostram que o erro de estimação é reduzido em, aproximadamente, 15%, quando o pedestal é subtraído das amostras digitais recebidas (OF proposto para o Run2). Adicionalmente, o uso da matriz da covariância do ruído de fundo melhora consideravelmente a resolução de energia para canais de alta ocupância do TileCal. Os trabalhos futuros se concentram no estudo de técnicas mais sofisticadas que tratam corretamente o efeito de empilhamento de sinais.

4. Bernardo S. Peralva, José M. Seixas *et al.*, *The TileCal Energy Reconstruction for LHC Run2 and Future Perspectives*, LISHEP2015 - SESSION C, Manaus, 2015.

Resumo: The Tile Calorimeter (TileCal) is the main hadronic calorimeter of ATLAS and it covers the central part of the detector ($|\eta| < 1.6$). The energy deposited by the particles in TileCal is read out by approximately 10,000 channels. The signal provided by the readout electronics for each channel is digitized at 40 MHz and its amplitude is estimated by an optimal filtering algorithm. The increase of LHC luminosity leads to signal pile-up that deforms the signal of interest and compromises the amplitude estimation performance. This work presents the proposed algorithm for energy estimation during LHC Run 2. The method is based on the same approach used during LHC Run 1, namely the Optimal Filter (OF). The only difference is that the signal baseline (pedestal) will be subtracted from the received digitized samples, while in Run 1 this quantity was estimated on an event-by-event basis. The pedestal value is estimated through special calibration runs and it is stored in a data base for online and offline usage. Additionally, the background covariance matrix will also be used for the computation of the OF weights for high occupancy channels. The use of such information reduces the bias and uncertainties introduced by signal pile-up. The performance of the OF version used in Run 1 and Run 2 is compared using Monte Carlo data. The efficiency achieved by the methods is shown in terms of error estimation, when different conditions of luminosity and occupancy are considered. Concerning future work, a new method based on linear signal deconvolution has been recently proposed and it is under validation. It could be used for Run 2 offline energy reconstruction and future upgrades.

5. Marcos Teixeira, Luciano M. A. Filho, Bernardo S. Peralva, *Reconstrução Online de Energia para Calorimetria de Altas Energias em Condições de Altas Luminosidades*, Congresso Brasileiro de Automática (CBA2014), Belo Horizonte, 2014.

Resumo: Este trabalho apresenta um algoritmo para reconstrução online de energia em ambiente de alta luminosidade e sua respectiva implementação em FPGA. Baseado no algoritmo Gradiente Descendente, o método implementado tem suas propriedades definidas através do método COF (do inglês, *Constrained Optimal Filter*). O COF visa estimar a amplitude de sinais independentemente das informações de luminosidade, recuperando, além do sinal de interesse, os sinais sobrepostos. Porém o método resulta em inversão de matrizes para a estimativa da amplitude dos sinais. Para evitar a inversão

de matrizes, o algoritmo Gradiente Descendente é aplicado no método COF resultando em operações de soma e produto, facilitando sua implementação em FPGA para reconstrução online. Para este fim, foi desenvolvido um processador RISC (do inglês, *Reduced Instruction Set Computer*) dedicado com arquitetura pipeline, permitindo a otimização de recursos da FPGA. Também foi desenvolvido um compilador C para facilitar o desenvolvimento das subrotinas do algoritmo Gradiente Descendente.

6. Bernardo S. Peralva, Luciano M. A. Filho, Augusto S. Cerqueira, José M. de Seixas. *Reconstrução de Energia do Calorímetro de Telhas do ATLAS Utilizando o Filtro Casado*, Encontro Nacional de Física de Partículas e Campos, Passa Quatro, Minas Gerais, 2013.

Resumo: Este trabalho propõe um método alternativo para reconstruir a energia depositada pelas partículas no TileCal, chamado de Filtro Casado (MF). A técnica explora as características do sinal do TileCal para projetar um detector de sinais baseado na teoria do detector de Bayes, o qual maximiza a relação sinal-ruído. Este detector é ótimo quando o sinal de interesse é determinístico e o ruído é Gaussiano. Durante a etapa de condicionamento do sinal analógico, este é conformado resultando em um pulso estável e de forma fixa, embora a amplitude varie de acordo com a intensidade e identidade da partícula. Além disso, o ruído eletrônico do TileCal possui características Gaussianas onde a correlação entre as amostras pode ser absorvida através da estimação da matriz de covariância do processo. Por fim, a saída do filtro detector pode ser calibrada de forma a apresentar seu resultado em unidade de energia. Como resultado, o MF implementa um filtro FIR rápido em que seus coeficientes são baseados no pulso de interesse do TileCal e na matriz de covariância do ruído. A fim de avaliar a eficiência da técnica proposta, o método MF foi comparado com o algoritmo OF. Para tal, dados experimentais adquiridos durante operação nominal do LHC foram utilizados. O método MF se mostrou superior ao OF no que tange o erro na medida da amplitude. Além disso, as estimativas de energia pelos métodos MF e OF se mostraram fortemente correlacionadas para sinais com alta relação sinal ruído. Os trabalhos futuros se concentram nas análises do comportamento dos métodos no cenário de empilhamento de sinais, problema que cresce com o aumento da luminosidade.

7. **(apresentado em nome da colaboração ATLAS)** Bernardo S. Peralva, on behalf of the ATLAS Collaboration. *The TileCal Energy Reconstruction for Collision Data Using the Matched Filter*, IEEE Nuclear Science Symposium and Medical Imaging Conference, Seoul, South Korea, 2013.

Resumo: The Tile Barrel Calorimeter (TileCal) is the central section of the hadronic calorimeter of ATLAS at LHC. It is divided in a central barrel and two extended barrels, where each part is formed by 64 modules in order to complete the entire cylinder. The central barrel modules are divided in 46 cells each, while the extended barrels modules are divided in 14 cells. The energy deposited in each cell is read out by two electronic channels for redundancy and the analog pulse is conditioned by a shaper circuit. Therefore, TileCal comprises more than 10,000 readout channels. The energy is estimated for each channel by reconstructing the amplitude of the digitized pulse sampled every 25 ns. This work presents the performance of an alternative algorithm for TileCal energy reconstruction, namely the TileCal Matched Filter (MF). The performance of the MF method is compared to the currently implemented algorithm (OF2) using collision data acquired in 2010 during LHC operation period. The results showed that the MF presents smaller error estimation (variance) than the OF2 method. In addition to that, the methods showed to be highly correlated with each other for high SNR signals. Preliminary results using a special ATLAS collision data taken later in 2012, for which LHC operated at 25 ns bunch spacing and ATLAS observed an increase of the pile-up effect, are also provided.

8. **(apresentado em nome da colaboração ATLAS)** Bernardo S. Peralva, *et al.*. *Calibration and Performance of the ATLAS Tile Calorimeter*, XI LISHEP - Workshop on High Energy Physics in the Near Future, Rio de Janeiro, 2013.

Resumo: The Tile Calorimeter (TileCal) is the hadronic calorimeter covering the most central region of the ATLAS experiment at the LHC. It is a key detector for the measurement of hadrons, jets, tau leptons and missing transverse energy. The TileCal calibration system comprises radioactive source, laser and charge injection elements and it allows to monitor and equalize the calorimeter response at each stage of the signal production, from scintillation light to digitization. This contribution presents a brief description of the different TileCal calibration systems as well as the latest results on their performance in terms of calibration factors, linearity and stability. The performance of the Tile Calorimeter with the cosmic muons and collision data is also presented, including the absolute energy scale, time resolution and associated stabilities.

9. Bernardo S. Peralva, Davis P. Barbosa, Luciano, M. A. Filho, Augusto S. Cerqueira. *Implementação em FPGA de Detectores Baseados em Filtragem Casada e Redes Neurais em Ambientes de Física Experimental de Altas Energias*. Congresso Brasileiro de Automática (CBA2012), Campina Grande, 2012.

Resumo: Este trabalho propõe a implementação de duas técnicas de detecção

de sinais em dispositivos FPGA no contexto de experimentos de física de alta energia. Foram abordados algoritmos baseados em filtragem casada e em rede neural artificial. Os resultados para o banco de dados utilizado mostraram um desempenho superior do detector neural.

10. Bernardo S. Peralva, Luciano M. A. Filho, Augusto S. Cerqueira, José M. de Seixas. *Uma Técnica de Estimção de Energia para o Calorímetro de Telhas do ATLAS Baseada em Filtragem Casada*. Publicado em: Conferência de Física da Comunidade de Países de Língua Portuguesa, Rio de Janeiro, 2012.

Resumo: Este trabalho apresenta um método alternativo para estimar a energia depositada pelas partículas nos canais do TileCal. O Filtro Casado (MF), o qual é baseado na teoria do detector de Bayes e que maximiza a SNR, é aproximado e aplicado no contexto do TileCal. Este detector é ótimo quando o sinal de interesse é determinístico e o ruído é branco Gaussiano. Embora o sinal analógico proveniente da deposição de energia pelas partículas nos canais do TileCal possui amplitude variável, este é condicionado para que sua forma seja bem definida. Além disso, o ruído eletrônico do TileCal possui características gaussianas, que pode ser branqueado através da estimação da matriz de branqueamento do processo. Como resultado, o MF implementa um filtro FIR rápido em que seus coeficientes são baseados no pulso de interesse do TileCal. A fim de avaliar a eficiência da técnica proposta, o método MF foi comparado com o algoritmo atual utilizado pelo TileCal para estimação de energia, chamado de Filtro Ótimo (OF). Para tal, dados reais de colisão adquiridos durante operação nominal do LHC foram utilizados. O método MF se mostrou superior ao OF no que tange o erro na medida da amplitude. Além disso, as estimativas de energia pelos métodos MF e OF se mostraram fortemente correlacionadas para sinais com alta relação sinal ruído. Os trabalhos futuros se concentram nas análises do comportamento dos métodos no cenário de empilhamento de sinais, problema que cresce com o aumento da luminosidade.

11. B. S. Peralva, Y. Amaral *et al.*. *Use of the background covariance matrix for the TileCal energy reconstruction under pile-up conditions*. Nota interna do ATLAS. Link CDS (para usuários do CERN): <https://cds.cern.ch/record/1967734>

Resumo: This note presents a study on the use of the background covariance matrix to improve the TileCal energy estimation under pile-up conditions. TileCal uses the Optimal Filter (OF) technique to estimate the amplitude of the readout signals, which envisages the use of the background covariance matrix in the optimization procedure that computes its weights. The correct

description of the covariance matrix is expected to reduce the uncertainties and bias due to the pile-up. The results for different luminosities scenarios foreseen for LHC Run2 show significant improvements in the energy resolution with respect to the identity matrix that was used during LHC Run1.

12. S. Amor Santos, J.P. Araque, N. Castro, L. Cerda Alberich, V. Drugakov, L. Manhães, A. Hrynevich, A. Onofre, C. Roda, C. Santoni, C. Solans, B. Sotto-Maior Peralva, P. Starovoitov, G. Usai, A. Valero. *Comparisons of data recorded at 25 and 50 ns bunch spacings in the Tile calorimeter*. Nota interna do ATLAS. Link CDS (para usuários do CERN): <https://cds.cern.ch/record/1610549>

Resumo: This note summarizes the studies performed in TileCal to evaluate the performance of the detector at 25 ns. Zero-bias stream data used to measure the cell and topo-clusters energy distributions. The results of different reconstruction cell channel energy methods were compared using JetTauEtmis stream data. Online and offline energy reconstruction results were also studied. The observed differences between the 50 and 25 ns are compatible with the expectations.