



**Universidade do Estado do Rio de Janeiro**  
Centro de Tecnologia e Ciências  
Instituto Politécnico

Geovane Pacheco da Rocha

**Estimação de energia baseada em filtragem casada para um calorímetro operando em condições de empilhamento de sinais**

Nova Friburgo

2021

Geovane Pacheco da Rocha

**Estimação de energia baseada em filtragem casada para um calorímetro  
operando em condições de empilhamento de sinais**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Bernardo Sotto-Maior Peralva

Nova Friburgo

2021

CATALOGAÇÃO NA FONTE  
UERJ/REDE SIRIUS/BIBLIOTECA CTC/E

R672 Rocha, Geovane Pacheco da.  
Estimação de energia baseada em filtragem casada para um calorímetro operando em condições de empilhamento de sinais / Geovane Pacheco da Rocha. – 2021.  
97 f. : il.

Orientador: Bernardo Sotto-Maior Peralva.  
Dissertação (mestrado) - Universidade do Estado do Rio de Janeiro, Instituto Politécnico.

1. Processamento de sinais – Métodos de simulação - Teses. 2. Estimativa de parâmetros - Teses. 3. Calorimetria - Teses. 4. Aceleradores de partículas – Teses. I. Peralva, Bernardo Sotto-Maior. II. Universidade do Estado do Rio de Janeiro. Instituto Politécnico. III. Título.

CDU 536.6:621.391

Bibliotecária Cleide Sancho CRB7/5843

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Geovane Pacheco da Rocha  
Assinatura

08/11/2021  
Data

Geovane Pacheco da Rocha

**Estimação de Energia Baseada em Filtragem Casada para um Calorímetro  
Operando em Condições de Empilhamento de Sinais**

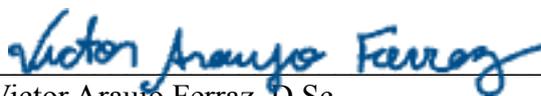
Dissertação apresentada como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Modelagem Computacional do Instituto Politécnico, da Universidade do Estado do Rio de Janeiro.

Aprovado em 27 de outubro de 2021.

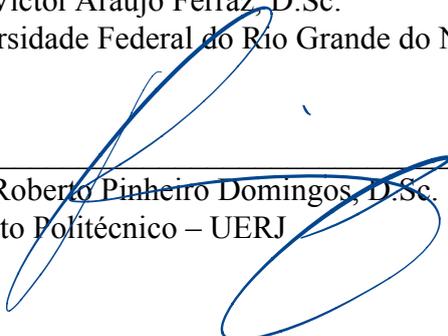
Banca examinadora:



Prof. Bernardo Sotto-Maior Peralva, D.Sc. - Orientador  
Instituto Politécnico – UERJ



Prof. Victor Araujo Ferraz, D.Sc.  
Universidade Federal do Rio Grande do Norte



Prof. Roberto Pinheiro Domingos, D.Sc.  
Instituto Politécnico – UERJ

Nova Friburgo

2021

## DEDICATÓRIA

À minha esposa que me acompanha, me motiva e não me deixa desistir nem perante os desafios mais difíceis.

## AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, pelo livramento da morte durante o COVID-19 este ano e por cuidar de mim mostrando seu amor infinito todos os dias.

Agradeço à minha esposa Priscila por estar ao meu lado com seu apoio incondicional durante estes 14 anos. Se não fosse por ela, eu jamais chegaria aqui.

Agradeço aos meus pais por me darem educação e me mostrar que a vida não é fácil desde criança.

Agradeço ao meu orientador, prof. Bernardo pelo apoio e por acreditar neste projeto mesmo sabendo que eu não poderia ser um bolsista de dedicação exclusiva.

Agradeço à UERJ por ter me aceitado em 2012 como aluno e ter me tornado um Engenheiro de alto nível. Graças à excelência no ensino, hoje além de estar finalizando o Mestrado também sou realizado profissionalmente trabalhando para uma empresa Norte Americana.

Agradeço à BriteCore por ser uma empresa incrível que além do apoio, me permitiu ter a flexibilidade durante estes anos para finalizar o mestrado.

Agradeço à FAPERJ, FAPEMIG, CNPq e RENAFAP pelo suporte para a realização deste trabalho. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

... O meu clã é o mais pobre da tribo de Manassés, e eu sou a pessoa menos importante da minha família! Mas o Senhor afirmou: “Eis que Eu estarei contigo e tu vencerás os midianitas como se fossem um só homem!”

*Juízes 6 15-16*

## RESUMO

ROCHA, G. P. *Estimação de energia baseada em filtragem casada para um calorímetro operando em condições de empilhamento de sinais*. 2021. 97 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2021.

A descrição e reprodução computacional de fenômenos são um desafio comum na área da Modelagem Computacional. No contexto da calorimetria de altas energias, a energia da partícula proveniente de colisões é absorvida e amostrada em forma de um sinal digitalizado. Em operação de alta taxa de eventos e luminosidade, o empilhamento de sinais pode ocorrer devido à alta ocupação dos canais de leitura do detector, tendo como consequência a distorção do sinal esperado. Neste contexto, este trabalho avalia o desempenho do método conhecido como Filtro Casado, aplicando-o em duas abordagens distintas chamadas de Filtro Casado Determinístico e Filtro Casado Estocástico. Tais abordagens serão comparadas ao método atualmente aplicado no Calorímetro de Telhas TileCal, chamado Filtro Ótimo. Para a análise de eficiência, foi criado um ambiente composto de dados simulados considerando diferentes condições de empilhamento de sinais no TileCal do ATLAS. Além deste ambiente, também foi criado o ambiente híbrido, que consiste em sinais cuja a amplitude é previamente conhecida e então adicionada em diferentes condições de ruídos reais. Para ambos ambientes, diferentes relações sinal-ruído foram consideradas a fim de se entender como cada método se comporta em tais situações. Os resultados mostraram que o Filtro Casado Estocástico se mostrou capaz de estimar a amplitude dos sinais de forma mais precisa se apresentando como uma alternativa para a estimação de energia do TileCal.

Palavras-chave: Filtro casado. Filtro ótimo. Estimação de sinais. Calorimetria de altas energias.

## ABSTRACT

ROCHA, G. P. *Energy estimation based on matched filtering for a calorimeter operating under pile-up conditions*. 2021. 97 f. Dissertação (Mestrado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2021.

The description and reproduction of physical phenomena are a common challenge in the field of computational modeling. In the context of high-energy calorimetry, the particle energy produced from collisions is absorbed and sampled in the form of a digitized signal. In a high event-rate and luminosity conditions, the signal pile-up effect may arise due to the high occupancy of the detector's readout channels, causing the distortion of the expected signal. In this context, this work evaluates the performance of the method known as Matched Filter, by applying it in two distinct approaches called Deterministic Matched Filter and Stochastic Matched Filter. These approaches will be compared to the current method applied in the ATLAS Tile Calorimeter (TileCal), known as Optimal Filter. For the efficiency analysis, a computational environment was created to contain simulated data considering different signal pile-up conditions in the TileCal. Furthermore, a hybrid environment was also created, consisting in signals whose amplitudes are previously known and added to different conditions of pile-up noise. For both analysis, different signal-noise ratio conditions were considered. The results show that the Stochastic Matched Filter presented high accuracy in estimating the amplitude, becoming an alternative for the TileCal energy estimation task.

Keywords: Matched filter. Optimal filter. Signal estimation. High energy calorimetry.

## LISTA DE FIGURAS

Figura 1	- Visão geral do CERN . . . . .	18
Figura 2	- Aceleradores do CERN . . . . .	20
Figura 3	- Detectores do LHC . . . . .	21
Figura 4	- Sistema de Filtragem Avançada do ATLAS . . . . .	23
Figura 5	- Detectores do ATLAS . . . . .	24
Figura 6	- Diagrama do caminho das partículas . . . . .	25
Figura 7	- Sistema de Calorimetria do ATLAS . . . . .	26
Figura 8	- Partições do TileCal . . . . .	27
Figura 9	- Partições do TileCal . . . . .	28
Figura 10	- Seção transversal do barril e de uma das extensões do TileCal . . . . .	29
Figura 11	- Pulso de referência do TileCal. . . . .	30
Figura 12	- Linha do tempo e atualizações previstas para o LHC até 2040. . . . .	30
Figura 13	- Empilhamento de sinais no TileCal. . . . .	31
Figura 14	- Típica resposta de um calorímetro para sinais digitais sem ruído. . . . .	37
Figura 15	- Resposta de um calorímetro considerando sinais originais, analógicos e amostras digitais. . . . .	37
Figura 16	- Exemplo de pulso utilizado. . . . .	50
Figura 17	- Ruídos simulados. . . . .	52
Figura 18	- Covariância dos ruídos simulados. . . . .	53
Figura 19	- Células do TileCal selecionadas. . . . .	55
Figura 20	- Comparação entre células para diferentes ruídos reais. . . . .	56
Figura 21	- Covariância dos ruídos reais. . . . .	57
Figura 22	- Comparação entre amplitudes simuladas considerando $SNR = 3$ . . . . .	59
Figura 23	- Comparação entre amplitudes simuladas considerando $SNR = 1$ . . . . .	60
Figura 24	- Comparação entre amplitudes simuladas considerando $SNR = 0,1$ . . . . .	61
Figura 25	- Comparação entre erros considerando $SNR = 3$ . . . . .	62
Figura 26	- Comparação entre erros considerando $SNR = 1$ . . . . .	63
Figura 27	- Comparação entre erros considerando $SNR = 0,1$ . . . . .	64
Figura 28	- Comparação entre as médias e RMSs entre todas $SNRs$ para métodos OF, DMF e SMF em função das ocupações. . . . .	66
Figura 29	- Comparação entre amplitudes para a célula $E4$ . . . . .	72
Figura 30	- Comparação entre amplitudes para a célula $A9$ . . . . .	73
Figura 31	- Comparação entre erros de estimação para a célula $E4$ . . . . .	74
Figura 32	- Comparação entre erros de estimação para a célula $A9$ . . . . .	75
Figura 33	- Comparação entre erros relativos de estimação para $E4 <\mu> \approx 90$ . . . . .	78
Figura 33	- Comparação entre erros relativos de estimação para $A9 <\mu> \approx 30$ . . . . .	80

Figura 33 - Componentes da PCA para o SMF. . . . .	81
Figura 34 - Comparação entre $PCA_{comps}$ considerando diferentes ocupações para $SNR = 1$ . . . . .	83
Figura 35 - Comparação entre erros de estimação de cada $PCA_{comp}$ em função de todas as ocupações. . . . .	84
Figura 36 - Comparação entre erros de estimação de cada $PCA_{comp}$ para a célula $E4$ . . . . .	85
Figura 37 - Comparação entre erros de estimação de cada $PCA_{comp}$ para a célula $A9$ . . . . .	86
Figura 38 - Comparação entre as médias e RMSs para cada $PCA_{comp}$ em função das amplitudes. . . . .	87

## LISTA DE TABELAS

Tabela 1	- Quantidade de eventos para cada célula para suas respectivas $\langle \mu \rangle$ . . .	54
Tabela 2	- Porcentagem negativa da amplitude estimada do OF para diferentes ocupações e $SNRs$ . . . . .	65
Tabela 3	- médias e RMSs de cada método para suas respectivas ocupações para $SNR = 3$ . . . . .	68
Tabela 4	- médias e RMSs de cada método para suas respectivas ocupações para $SNR = 1$ . . . . .	69
Tabela 5	- médias e RMSs de cada método para suas respectivas ocupações para $SNR = 0,1$ . . . . .	70
Tabela 6	- médias e RMSs de cada método para $E4 \langle \mu \rangle \approx 90$ e suas respectivas amplitudes. . . . .	76
Tabela 7	- médias e RMSs de cada método para a célula $A9 \langle \mu \rangle \approx 30$ e suas respectivas amplitudes. . . . .	76
Tabela 8	- Valores numéricos das componentes da PCA para o SMF. . . . .	79
Tabela 9	- médias e RMSs para cada $PCA_{comp}$ em função das ocupações. . . . .	82
Tabela 10	- médias e RMSs para cada $PCA_{comp}$ em função das amplitudes. . . . .	88

## LISTA DE ABREVIATURAS E SIGLAS

AD	<i>Antiproton Decelerator</i>
ALICE	<i>A Large Ion Collider Experiment</i>
ATLAS	<i>A Toroidal LHC Apparatus</i>
CERN	<i>Centre Européenne pour La Recherche Nucléaire</i>
CMS	<i>Compact Muon Solenoi</i>
CTP	<i>Central Trigger Processor</i>
DMF	<i>Deterministic Matched Filter</i>
EB	<i>Extended Barrel</i>
ELENA	<i>Extra Low ENergy Antiproton</i>
FIR	<i>Finite Impulse Response</i>
LB	<i>Long Barrel</i>
LEIR	<i>Low Energy IonRing</i>
LHC	<i>Large Hadron Collider</i>
LHCb	<i>Large Hadron Collider beauty</i>
LS	<i>Long Shutdown</i>
HIE-ISOLDE	<i>High-Intensity and Energy upgrade of ISOLDE</i>
HL-LHC	<i>High-LuminosityLarge Hadron Collide</i>
MF	<i>Matched Filter</i>
MLE	<i>Maximum Likelihood Estimator</i>
OF	<i>Optimal Filter</i>
PCA	<i>Principal Component Analysis</i>
PDF	<i>Probability Density Function</i>
PMT	<i>Photo-multiplier Tube</i>
PS	<i>Proton Synchrotron</i>
RF	<i>Cavidades de Radiofrequência</i>
RMS	<i>Root Mean Square</i>
SC	<i>Synchrocyclotron</i>
SMF	<i>Stochastic Matched Filter</i>
SNR	<i>Signal-Noise Ratio</i>
SPS	<i>Super ProtonSynchrotron</i>
TDAQ	<i>Trigger and Data Acquisition</i>
WLS	<i>Wavelength Shifiting</i>
WWW	<i>World Wide Web</i>

## SUMÁRIO

	<b>INTRODUÇÃO</b>	13
1	<b>ACELERADOR DE PARTÍCULAS DO CERN</b>	17
1.1	<b>CERN</b>	17
1.2	<b>Aceleradores de Partículas</b>	18
1.3	<b>LHC</b>	20
1.4	<b>ATLAS</b>	22
1.4.1	<u>Calorímetros do ATLAS</u>	24
1.4.1.1	Calorímetro Hadrônico de Telhas	26
1.4.2	<u>Atualização do LHC</u>	29
2	<b>RECONSTRUÇÃO DE ENERGIA EM CALORIMETRIA DE ALTAS ENERGIAS</b>	32
2.1	<b>Estimação de Parâmetros</b>	32
2.2	<b>Estimador de Máxima Verossimilhança (MLE)</b>	34
2.3	<b>Estimadores de energia em calorimetria</b>	36
2.4	<b>Filtro Ótimo</b>	38
3	<b>FILTRO CASADO</b>	42
3.1	<b>Detecção do Sinal</b>	42
3.2	<b>Filtro Casado Determinístico</b>	44
3.2.1	<u>Amplitude do Filtro Casado Determinístico</u>	45
3.3	<b>Filtro Casado Estocástico</b>	46
3.3.1	<u>Amplitude do Filtro Casado Estocástico</u>	47
3.4	<b>Aplicação em alta taxas de eventos</b>	47
4	<b>RESULTADOS</b>	49
4.1	<b>Estratégia Utilizada</b>	49
4.2	<b>Conjunto de Dados</b>	50
4.2.1	<u>Ambiente Simulado</u>	51
4.2.2	<u>Ambiente Híbrido</u>	54
4.3	<b>Análise dos Resultados</b>	57
4.3.1	<u>Ambiente Simulado</u>	58
4.3.2	<u>Ambiente Híbrido</u>	67
4.4	<b>Análise da PCA para o SMF</b>	77
4.4.1	<u>Ambiente Simulado</u>	82
4.4.2	<u>Ambiente Híbrido</u>	84
	<b>CONCLUSÃO</b>	89
	<b>REFERÊNCIAS</b>	91
	<b>APÊNDICE A – Publicações científicas</b>	97

## INTRODUÇÃO

*A ciência serve para nos dar uma ideia de quão extensa é a nossa ignorância.*

*Félicité Robert de Lamennais*

Durante a história da humanidade ocorreram diversas evoluções que propiciaram o desenvolvimento da vida humana, dentre estas destacam-se a Revolução Agrícola aproximadamente 10000 anos atrás e a Revolução Industrial iniciada após a invenção da máquina a vapor em 1776 (TOFFLER, 1991). Agora há outra grande mudança em curso, a Revolução da Informação, motivada inicialmente pelo lançamento do primeiro computador ENIAC (*Electronic Numerical Integrator and Calculator*) em 1947 (CAVALCANTI, 1995) que ampliou ainda mais o desenvolvimento da ciência.

Na mesma década acontecia a Segunda Guerra mundial, e por isso, visando interromper a migração de profissionais e cientistas para a América, o desenvolvimento da ciência, e o fortalecimento da Europa no pós guerra, um pequeno número de cientistas visionários na Europa e na América do Norte identificaram a necessidade da Europa ter um centro de pesquisa em física de classe mundial.

Durante uma reunião intergovernamental da UNESCO (*United Nations Educational, Scientific and Cultural Organization*) em Dezembro de 1951 em Paris, fora adotada a primeira resolução sobre o estabelecimento de um Conselho Europeu de Pesquisa Nuclear, ou como é conhecido hoje CERN (*Conseil Européen pour la Recherche Nucléaire*) (CERN, 2020c).

Em 1953, Genebra, cidade localizada na Suíça, foi escolhida para sediar a construção do CERN devido a sua posição geográfica na Europa e também à sua neutralidade durante a guerra. Com objetivo puramente científico e não militar, determinou-se que todo trabalho realizado deveria ser publicado ou disponibilizado publicamente. O foco na pesquisa era tão grande que, durante a guerra fria o CERN foi um dos poucos lugares onde cientistas de ambos lados do mundo trabalhavam em conjunto (CERN, 2020c).

Desde então diversas importantes contribuições e construções ocorreram. Como exemplo de contribuições têm-se a descoberta das partículas W e Z em 1983, *World Wide Web* (WWW) em 1989 por Tim Berners-Lee (CERN, 1998) e Bóson de Higgs em 2012 (COLLABORATION, 2012), e construções como SPS (*Super Proton Synchrotron*) em 1976 e LHC (*Large Hadron Collider*) (EVANS; BRYANT, 2008) em 2008. O LHC é o maior e mais poderoso acelerador de partículas do mundo e possui vários detectores, como por exemplo o ATLAS e seus sub-sistemas de detecção.

Os Calorímetros fazem parte dos sub-sistemas de detecção do ATLAS, e, atualmente, calorímetros é uma das formas mais utilizadas de obtenção dos dados dos experi-

mentos físicos de colisão de partículas, fornecendo a informação sobre a energia detectada por colisão (WIGMANS, 2000). Calorímetros são utilizados em diversas camadas subdivididas em milhares de células responsáveis por efetuarem a leitura da energia absorvida, que é então amostrada e processada eletronicamente (PERALVA, 2013).

Em ambientes complexos como os encontrados no LHC e cujos dados produzidos precisam ser analisados de alguma forma, são utilizados sensores a fim de se captar dados. Em tais dados, podem haver informações relevantes, informações não relevantes (geralmente chamadas de ruídos) ou ambas ocorrendo simultaneamente. Por isso, sistemas de seleção de eventos são comumente utilizados junto de conversores digitais a fim de que apenas os dados de interesse sejam processados computacionalmente (PASTORE, 2016).

Através de sensores, cada célula consegue realizar a leitura da energia depositada na mesma e, ao analisar o somatório das energias de todas as células pode-se reconstruir os sinais originais e inferir a natureza da partícula incidente (KNOLL, 2010). No entanto, devido ao alto nível de segmentação das células, é possível que algumas não possuam dados relevantes para o experimento, afetando assim a reconstrução dos sinais originais. Além destas, deve-se também considerar outras incertezas nos dados coletados, como ruídos eletrônicos, depreciação dos equipamentos de leitura, e desvios de fase na leitura das informações (PERALVA, 2012).

Quando estas incertezas estão presentes, deve-se buscar um meio de descartá-las ou tratá-las baseando-se em alguns critérios ou métodos. Tais métodos, além de efetuar a filtragem dos dados devem também levar em consideração a eficiência de estimação, visto que a taxa de coleta de dados é extremamente alta, podendo impactar na análise dos mesmos.

Dentre estes, têm-se os métodos baseados na minimização da variância do ruído (CLELAND; STERN, 1994) que podem ser utilizados para estimar a amplitude em cenários cujo ruído possui um modelo do tipo gaussiano. Em demais cenários pode ser interessante analisar o comportamento de outras técnicas como o Filtro Casado (SEIXAS; CALOBA; PINTO, 1996a), que por sua vez não é limitado a ruídos gaussianos.

Portanto, este trabalho explora a aplicação do Filtro Casado no Calorímetro de Telhas (ADRAGNA, 2006) do experimento ATLAS (COLLABORATION, 2008) no LHC (EVANS; BRYANT, 2008) considerando as abordagens determinística e estocástica, onde estas serão comparadas com o método conhecido como Filtro Ótimo 2 que é atualmente utilizado no Calorímetro de telhas do ATLAS.

## Objetivo

Os aceleradores de partículas são responsáveis por acelerar feixes de partículas para colidirem com objetos estáticos ou com outras partículas. Um dos objetivos destas

colisões é a descoberta de novas partículas ou o estudo das já conhecidas. Uma forma de se fazer isso é através da reconstrução da colisão e para isso é necessário a estimação correta da energia, que também pode ser interpretada como amplitude proporcional do sinal.

Como mencionado anteriormente, é comum a presença de diversos tipos de ruídos, principalmente os ruídos eletrônicos. Durante colisões de partículas, uma quantidade muito grande de dados é gerada, e como consequência disso, um novo ruído chamado de empilhamento de sinais pode surgir, sendo capaz de distorcer os sinais recebidos.

Empilhamentos de sinais ocorrem quando a taxa de ocupação de um canal do detector está próxima de seu limite, ou seja, ocorrem quando a taxa de leitura de seus equipamentos é mais lenta que a taxa dos sinais recebidos. Este cenário de alta ocupação acontecerá com maior frequência quando o LHC receber sua nova atualização que prevê um acréscimo substancial na taxa de colisão de partículas, tendo como consequência o aumento da complexidade de todo sistema de leitura de sinais.

Na literatura, encontram-se trabalhos relacionados ao uso do Filtro Casado em cenários comuns, ou seja, cenários que não ocorrem empilhamentos. Um exemplo é o artigo "*A matched filter based approach for high-energy estimation in calorimetry*" de (PERALVA et al., 2021).

O objetivo deste trabalho é além de analisar o cenário comum supracitado, também analisar o problema de empilhamento de sinais em um dos detectores do ATLAS chamado TileCal.

Para isso, será implementado duas abordagens do Filtro Casado conhecidas como Filtro Casado Determinístico e Filtro Casado Estocástico, ambas comparadas à atual versão do Filtro Ótimo usada atualmente no TileCal. A escolha do Filtro Casado como método base para este trabalho se dá devido ao fato deste não ser limitado a ruídos gaussianos, podendo então ser um bom candidato à análise do problema de empilhamento de sinais.

Para isso, serão elaborados três conjuntos simulados de testes, onde estes serão compostos de sinais cujas as amplitudes esperadas variam a fim de considerar diferentes relações sinal-ruído. Os conjuntos serão criados com componentes aleatórias, gaussianas e com empilhamento de sinais.

Por fim, também será avaliado os desempenhos destes métodos nos ambientes híbridos, ou seja, ambientes compostos de ruídos reais acrescidos de diferentes amplitudes conhecidas a fim de também avaliar a relação sinal-ruído.

## Organização Textual

O Capítulo 1 introduzirá o contexto o qual esta dissertação foi elaborada, o CERN. Definições importantes a respeito da física de alta energia e seu uso nos aceleradores do CERN serão apresentadas. Além da definição de aceleradores de partículas e suas propriedades, será abordado em maiores detalhes o maior acelerador de partículas do mundo, o LHC, e seus experimentos, com foco maior ao experimento ATLAS e seus sistemas de calorimetria. Por fim, será descrito a atualização prevista para o LHC e os desafios que tal atualização trará tanto para o ponto de vista físico quanto para o computacional.

O Capítulo 2 abordará a teoria a respeito da Reconstrução de Energia para o calorímetro TileCal. Esta teoria é fundamentada na Estimação de Parâmetros e estimadores de energia. Ao final será apresentado o estimador Filtro Ótimo que é utilizado atualmente no TileCal.

O Capítulo 3 irá descrever o Filtro Casado que é o principal método proposto nesta dissertação para o cenário de empilhamento de sinais. O Filtro Casado é uma técnica de detecção não limitada a ruído gaussiano e será aplicada para selecionar os canais do TileCal que possuem informações relevantes e também para filtrar sinais imersos em diferentes tipos de ruídos. Por fim, duas abordagens deste método serão apresentadas para a estimação de energia, o Filtro Casado Determinístico e o Filtro Casado Estocástico.

O Capítulo 4 apresentará a estratégia utilizada para a elaboração dos conjuntos de dados para dois ambientes, o ambiente simulado e o ambiente híbrido. Cada ambiente será constituído de diferentes conjuntos de dados os quais serão responsáveis por analisar a eficiência de cada método em diferentes relações sinal-ruído. Após, resultados e gráficos serão apresentados seguidos de suas análises.

Finalmente, as conclusões serão apresentadas seguidas dos possíveis trabalhos futuros.

# 1 ACELERADOR DE PARTÍCULAS DO CERN

*É uma questão de compreensão!  
Compreensão do mundo!*

*Peter Ware Higgs*

Este capítulo contextualizará o estudo da física de altas energias realizado pelo CERN e seus pesquisadores através dos experimentos envolvendo colisões de partículas realizadas no acelerador de partículas chamado LHC. No escopo do LHC e devido ao foco desta dissertação, será abordado em maiores detalhes o detector ATLAS e seu calorímetro hadrônico conhecido como TileCal.

## 1.1 CERN

Ao final da segunda guerra mundial, as pesquisas Europeias não eram mais de nível mundial. Seguindo exemplos de organizações internacionais, cientistas visionários aspiravam um laboratório Europeu de física atômica, dentre estes pioneiros estava Niels Bohr.

Então, em uma reunião intergovernamental da UNESCO em Dezembro de 1951 em Paris, a primeira resolução sobre um Conselho Europeu de Pesquisa Nuclear fora adotada. Em dois meses onze países assinaram um acordo estabelecendo o conselho que hoje é conhecido como CERN (*Conseil Européen pour la Recherche Nucléaire*).

Durante a terceira reunião do conselho, diversos países estavam em discussão sobre onde sediar o laboratório Europeu, dentre as opções estavam Dinamarca, Holanda, França e Suíça. Genebra na Suíça, foi selecionada, e em Junho de 1953 foi aprovada pelo conselho possuindo aproximadamente 69% dos votos. Tal cidade foi eleita devido a sua posição geográfica na Europa, a neutralidade durante a guerra e também devido ao fato de já hospedar diversas organizações internacionais.

O primeiro acelerador do CERN foi o Synchrocyclotron (*SC*) construído em 1957 com energia de 600 MeV que foi descontinuado em 1990 após 33 anos operando. Desde então o CERN vem realizando diversas outras construções como o SPS (*Super Proton Synchrotron*) em 1976 e LHC (*Large Hadron Collider*) em 2008, e, contribuições como a tela sensível ao toque na década de 70 (CERN, 2010), *World Wide Web* (WWW) em 1989 por Tim Berners-Lee (CERN, 1998), e Bóson de Higgs em 2012 (COLLABORATION, 2012).

De forma geral, o CERN tem como missão:

- Fornecer uma gama única de aceleradores de partículas que permitem pesquisas na

vanguarda do conhecimento humano.

- Realizar pesquisas de classe mundial em física fundamental.
- Unir pessoas de todo o mundo para expandir as fronteiras da ciência e da tecnologia, para o benefício de todos.

Atualmente o CERN possui 23 países membros e mais de 17500 colaboradores ao redor do mundo, onde destes, destacam-se aproximadamente 12 mil pesquisadores de mais de 70 países.

A Figura 1 mostra a visão geral do CERN com o globo da ciência e inovação no centro. Em tal espaço, é possível agendar visitas internas que permitem aos visitantes conhecer em detalhes como o CERN e seus aceleradores de partículas funcionam através de exposições interativas.

Figura 1 - Visão geral do CERN



Fonte: CERN, 2021a.

## 1.2 Aceleradores de Partículas

Aceleradores de partículas foram inventados em 1930 para investigar a estrutura atômica dos núcleos, e, desde então, são utilizados para o estudo em diversos aspectos da física de partículas. Durante este período ocorreu uma revolução na compreensão da estrutura da matéria, pois, acreditava-se que a mesma era composta apenas de prótons, nêutrons e elétrons, onde estes eram considerados como blocos básicos de toda a matéria existente (PERALVA, 2012).

Os aceleradores, também conhecidos como colisionadores de partículas, consistem em impulsionar partículas (como por exemplo prótons e elétrons) em velocidade próxima a da luz a fim de as colidirem contra algum alvo específico ou outra partícula também acelerada em direção oposta. Com este processo, partículas ainda mais elementares são geradas (PERALVA, 2012).

Aceleradores podem ser lineares ou circulares. No primeiro, feixes de partículas são lançados em linha reta, ao passo que no segundo as partículas percorrem uma trajetória circular em *loop* (EDWARDS; SYPHERS, 2008; GRIFFITHS, 2008). No acelerador linear, caso deseja-se aumentar a energia (ou velocidade) da partícula, é necessário aumentar o comprimento do mesmo, já na trajetória circular, a partícula repetirá o mesmo trajeto até que obtenha a energia necessária para realizar a colisão.

No geral, o funcionamento de um acelerador circular consiste em gerar campos elétricos e magnéticos onde estes são responsáveis respectivamente por acelerar e direcionar o feixe de partículas. Os pulsos elétricos são gerados por Cavidades de Radiofrequência (RF) enquanto mais de cinquenta tipos de ímãs são necessários para manter o feixe estável, alinhado, e o direcionar corretamente para outros aceleradores ou até mesmo para o ponto de colisão (EVANS; BRYANT, 2008; WIEDEMANN, 2015).

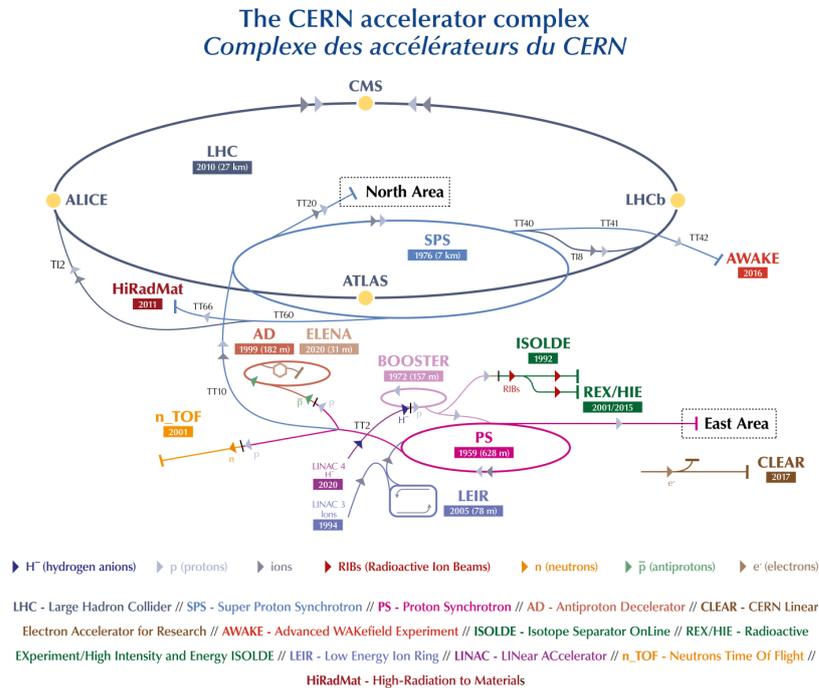
A energia da colisão é transformada em matéria na forma de novas partículas. Esse fenômeno é descrito pela equação de Albert Einstein  $E = mc^2$ , que define que a matéria é uma forma concentrada de energias onde estas não são intercambiáveis. Por meio destas colisões, foi possível descobrir outras partículas como Quarks e Bóson de Higgs. Através destas descobertas cada vez mais detalhadas, cientistas são capazes de melhor entender a matéria e a origem do universo.

Atualmente o CERN opera 8 aceleradores e 2 desaceleradores, dentre os aceleradores têm-se: LHC (*Large Hadron Collider*), Linac 3, Linac 4, LEIR (*Low Energy Ion Ring*), PS (*Proton Synchrotron*), SPS (*Super Proton Synchrotron*), HIE-ISOLDE (*High-Intensity and Energy upgrade of ISOLDE*). Dentre os desaceleradores têm-se: ELENA (*Extra Low ENergy Antiproton*) e AD (*Antiproton Decelerator*).

A Figura 2 mostra o quão complexa é toda a estrutura geral de aceleradores do CERN e outras informações como por exemplo os tipos de partículas. De baixo para cima, é possível ver o caminho feito pelas partículas e por onde estas são aceleradas até chegar no maior anel em azul escuro, denominado LHC, que será abordado a seguir.

Devido ao escopo de pesquisa desta dissertação, o acelerador que será analisado em detalhes será apenas o LHC, mais precisamente seu detector ATLAS a ser apresentado.

Figura 2 - Aceleradores do CERN



Fonte: MOBS, 2016.

### 1.3 LHC

O LHC é o maior e mais poderoso acelerador de partículas do mundo. Localizado a 100 metros de profundidade do solo, consiste em um anel supercondutor magnético de 27 quilômetros onde dois feixes de partículas são impulsionados no vácuo em sentidos opostos próximos da velocidade da luz a fim de se colidirem (EVANS; BRYANT, 2008; WIEDEMANN, 2015).

Colisões no LHC geram temperaturas 100 mil vezes mais quente que o centro do sol, fazendo necessário um incrível projeto de arquitetura e engenharia, como por exemplo o ambiente de vácuo. Este sistema possui três subsistemas separados: um para os tubos do feixe, um para isolar os ímãs resfriados criogenicamente e um para isolar a linha de distribuição de hélio. Isso além de possibilitar baixas temperaturas, também evita que as partículas colidam com as moléculas de gás dentro do acelerador. Por possuir o total de 104 quilômetros, a tubulação de vácuo do LHC é considerada uma das maiores do mundo (CERN, 2020a; CERN, 2020b).

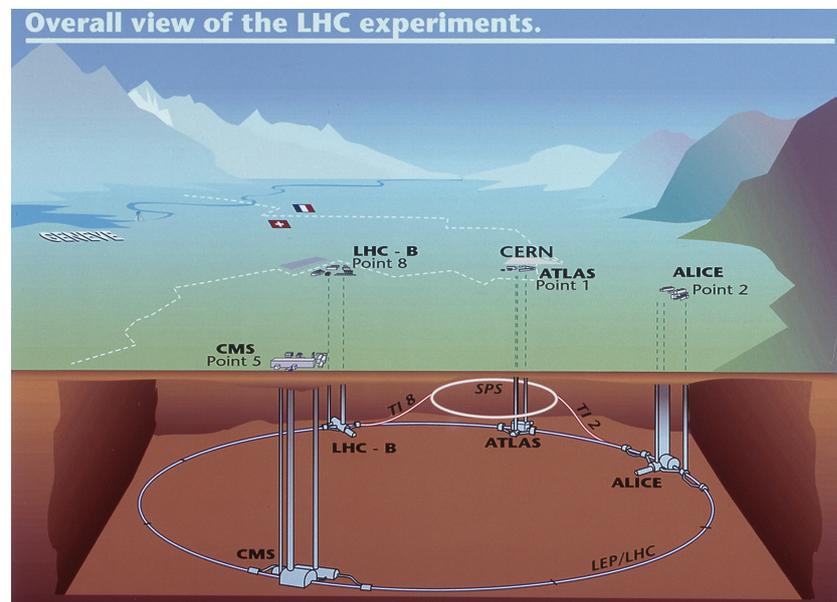
Antes de chegarem ao LHC, as partículas são aceleradas em uma série de aceleradores lineares e circulares interconectados. Uma vez que atingem a velocidade almejada em uma parte da cadeia, elas são lançadas para a próxima, até que cheguem na última etapa, o LHC, com energia máxima (CERN, 2018).

Estas partículas são impulsionadas por Cavidades de Radiofrequência (RF) e direcionadas por mais de 50 supercondutores eletromagnéticos. O LHC possui 16 RFs, sendo capaz de gerar uma energia de 450 GeV de partículas. A energia máxima é alcançada aproximadamente após 20 minutos percorrendo as RFs mais de 10 milhões de vezes. Neste caso, produzindo feixes com energia de aproximadamente 7 TeV, gerando então colisões de 14 TeV (EVANS; BRYANT, 2008; WIEDEMANN, 2015).

O feixe de partículas do LHC é constituído por milhares de pacotes de prótons onde cada um é lançado com o intervalo 25 ns ocasionando colisões a uma taxa constante de 40 MHz nos detectores do LHC (PERALVA, 2012). Estes detectores são: ATLAS, CMS, ALICE e LHCb.

A Figura 3 exemplifica a estrutura do LHC assim como seus detectores.

Figura 3 - Detectores do LHC



Fonte: CERN, 2021b.

Sucintamente, o propósito de cada um é:

- ATLAS (*A Toroidal LHC ApparatuS*) (COLLABORATION, 2008)
  - Experimento de propósito geral otimizado para ter o alcance máximo da física proposta pelo LHC.
- CMS (*Compact Muon Solenoid*) (COLLABORATION, 2010)
  - Outro experimento de propósito geral para o estudo do Bóson de Higgs, de partículas supersimétricas e de física de íons pesados.
- ALICE (*A Large Ion Collider Experiment*) (AAMODT et al., 2008)

- Experimento dedicado à física de colisões nucleares a fim de estudar a formação do plasma de quarks e glúons.
- LHCb (*Large Hadron Collider beauty*) (ALVES JR. et al., 2008)
  - Experimento dedicado ao estudo do decaimento méson-B e da violação CP.

Dentre os detectores supracitados, o ATLAS será analisado em detalhes devido ao escopo deste trabalho.

O LHC tem uma grande atualização chamada de HL-LHC (*High-Luminosity Large Hadron Collider*) que visa aumentar a performance do LHC e possibilitar novas descobertas após 2027. O objetivo deste projeto é aumentar a atual luminosidade por um fator de 10 além do projeto inicial do LHC (COLLABORATION, 2010; ERIKSSON et al., 2012). Com este aumento, a quantidade de dados processados e armazenados aumentará consideravelmente, o que acarretará em ainda mais empilhamento de sinais em seus detectores. Este problema será abordado em detalhes na Seção 1.4.2.

## 1.4 ATLAS

Como já mencionado, ATLAS (*A Toroidal LHC ApparatuS*) é um dos detectores do LHC e é um dos focos de estudo deste trabalho. Pesando 7000 toneladas, é o maior detector do LHC medindo aproximadamente 44 metros de comprimento e 25 metros de altura. Em 2012 mais de 3 mil cientistas de diferentes países trabalhavam neste experimento. Também é responsável por investigar uma ampla gama física, desde Bóson de Higgs até Matéria Escura, e, apesar de ter objetivos semelhantes ao CMS, ele faz uso de um design magnético diferente (COLLABORATION, 2008).

Os feixes de partículas provenientes do LHC colidem no centro do ATLAS gerando *terabyte* de dados. Tal colisão provê novas partículas que são dispersadas em todas as direções. Como exemplo de nova partícula, tem-se a notória contribuição já mencionada do Bóson de Higgs em 2012 e seu recente uso no estudo da Matéria escura. Além disso, em Fevereiro de 2021 foi anunciado a descoberta da primeira evidência do decaimento do Bóson de Higgs durante colisões para um sistema de dilepton de baixa massa e um fóton (AAD et al., 2021).

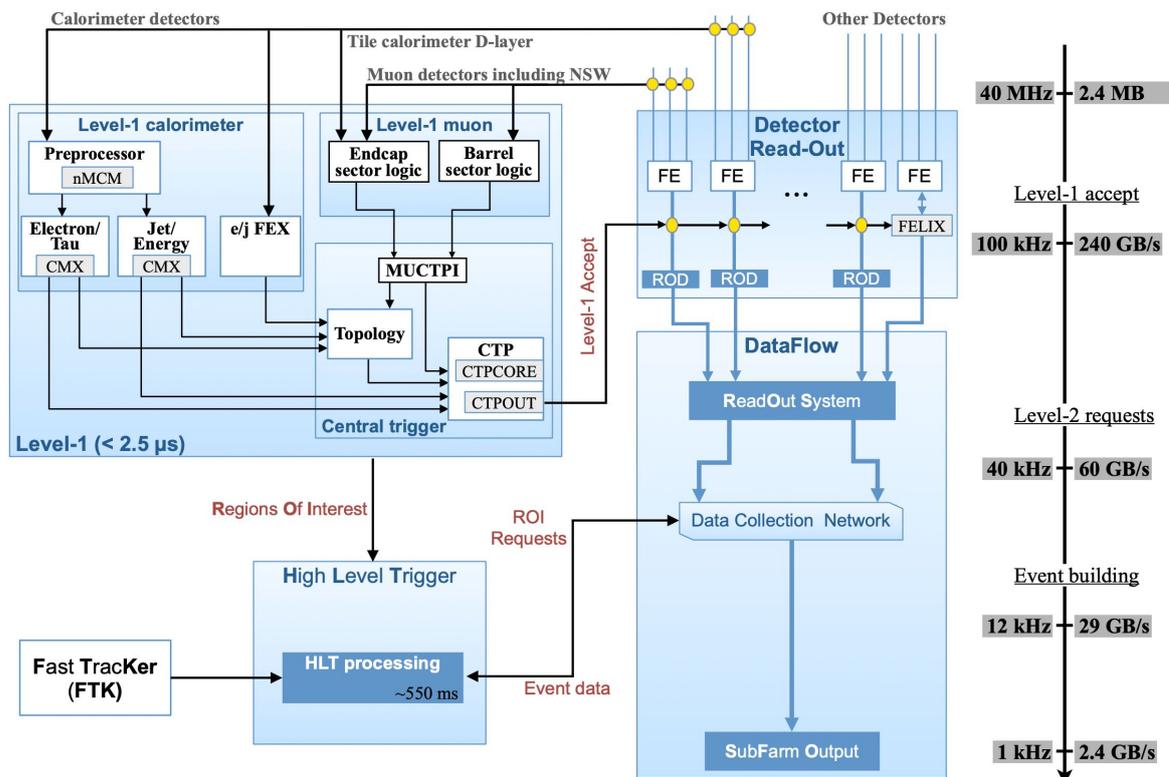
Com colisões ocorrendo a uma taxa de 40 MHz, uma imensa quantidade de dados é gerada, fazendo-se necessário um sistema de leitura avançado (SÁNCHEZ, 2010). Para isso, ATLAS faz uso do sistema TDAQ (*Trigger and Data Acquisition*) que são gatilhos responsáveis por selecionar quais eventos devem ser armazenados e quais devem ser ignorados em três níveis sequenciais, ou seja, os dados filtrados em cada nível são transferidos para o nível seguinte (JENNI et al., 2003).

A Figura 4 mostra o sistema de filtragem do ATLAS e seus níveis chamados de *Level 1* e *High-Level Trigger (HLT)*. O *level 1* é um sistema de Baixo Nível implementado em *hardware* enquanto o HLT pode ser considerado de Alto Nível, visto que é implementado em *software*.

O *Level 1*, é implementado em hardware dedicado que utiliza informações dos sistemas de calorimetria e dos detectores de muons, sendo capaz de selecionar eventos em um intervalo inferior a  $2,5 \mu s$ , sendo capaz de selecionar até 100000 eventos por segundo como filtro para o segundo nível, o HLT. O nível HLT, é responsável por refinar ainda mais a seleção de eventos realizada anteriormente através de software, baseando-se em regiões de interesse selecionadas em cada detector, sendo capaz de selecionar cerca de 1000 eventos por segundo a fim de serem analisados posteriormente (COLLABORATION, 2013).

O Level 1 é projetado para reduzir a taxa de eventos de 40 MHz para 100 kHz, ao passo que o HLT reduz este valor para apenas 1kHz. Ao fim de todo este processo de filtragem e seleção de eventos, apenas 2,4 GB de eventos são armazenados em disco por segundo para a análise *offline* (COLLABORATION, 2013).

Figura 4 - Sistema de Filtragem Avançada do ATLAS



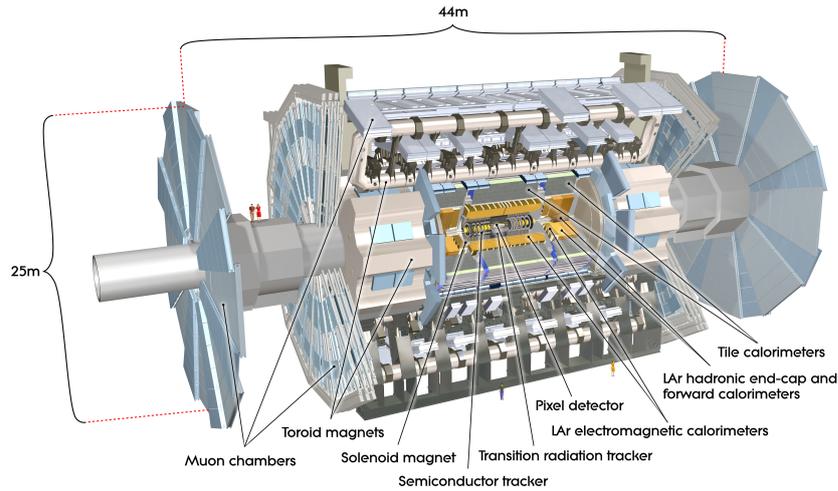
Fonte: Adaptado de COLLABORATION, 2013.

Ao fim, devido ao processo de filtragem e seleção de eventos, é possível reduzir aproximadamente 60 milhões de *megabytes* por segundo para apenas aproximadamente 2457 *megabytes* de eventos por segundo que serão armazenados em disco para posterior

reconstrução e análise *offline* (COLLABORATION, 2013).

Em termos de estrutura, ATLAS possui seis subsistemas de detecção distribuídos em camadas ao redor do ponto de colisão com o objetivo de registrarem o momento e a energia das partículas, permitindo que sejam individualmente identificadas para análises mais detalhadas. A Figura 5 mostra o ATLAS e seus detectores.

Figura 5 - Detectores do ATLAS



Fonte: PEQUENAO, 2008b.

A Figura 6 mostra três sub-detectores do ATLAS e os respectivos caminhos que as partículas fazem. Em ordem do mais interno para o mais externo tem-se: o detector de trajetórias, os calorímetros eletromagnético e hadrônico e o detector de múon.

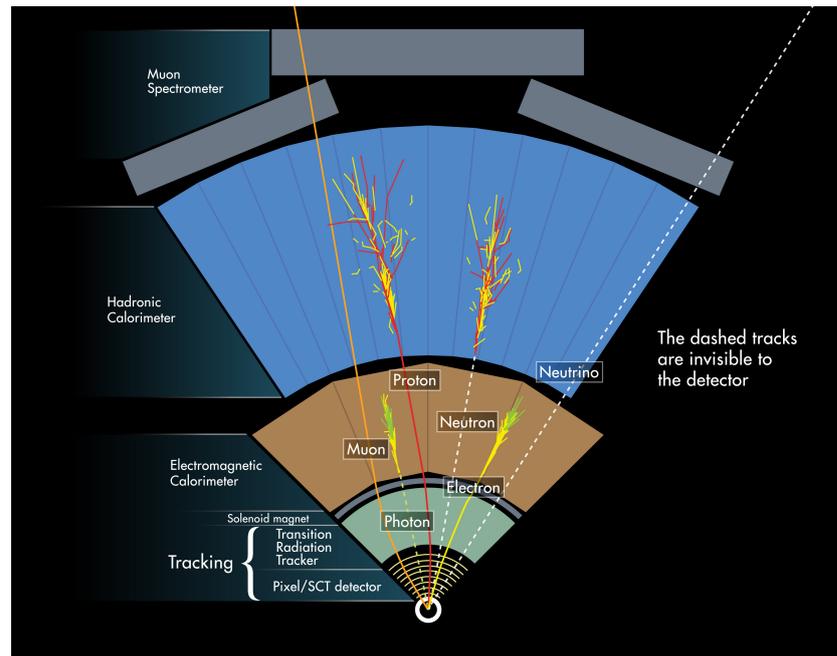
O calorímetro hadrônico de telhas é um destes detectores e é o principal objeto de estudo desta dissertação. Este será abordado em detalhes na próxima seção.

#### 1.4.1 Calorímetros do ATLAS

Calorímetros são utilizados em física de altas energias, e no cenário de colisão de partículas tem o objetivo de absorver, amostrar e medir as partículas nele incidentes (LEO, 1994). Geralmente são transversalmente segmentados a fim de se obter informações das direções e energias depositadas das partículas (PERALVA, 2012).

O projeto de um calorímetro geralmente consiste em parar ou absorver a maioria das partículas provenientes de uma colisão, depositando suas energias no interior do detector (WIGMANS, 2017). Sua estrutura é frequentemente disposta em camadas de material passivo, ou absorvente, de alta densidade como por exemplo aço e chumbo, intercalando-os com um meio ativo, como argônio líquido ou polímetros (COLLABORATION, 2008).

Figura 6 - Diagrama do caminho das partículas



Fonte: PEQUENAO; SCHAFFNER, 2013.

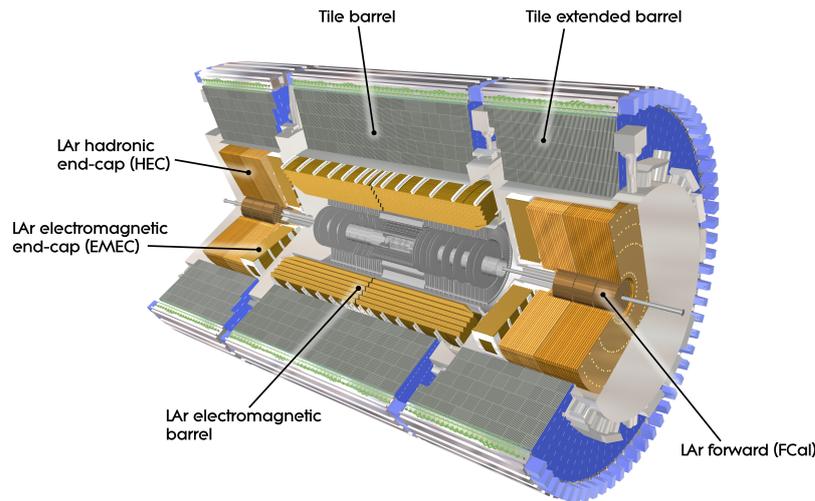
Durante a incidência de partículas, inicia-se um processo chamado de Chuveiro de Partículas, que consiste nestas atravessando diferentes materiais e permitindo que suas energias possam ser coletadas e então amostradas (DAS; FARBEL, 2003). A Figura 6 mostra este efeito ocorrendo para diferentes materiais e partículas.

Calorímetros são considerados como detectores chave, e de acordo com (PERALVA, 2012) alguns destes motivos são:

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

O sistema de calorimetria do ATLAS mostrado na Figura 7, é composto por dois calorímetros: Eletromagnético e Hadrônico. O calorímetro eletromagnético foi projetado para determinar a energia de partículas que interagem eletro-magneticamente. Já o calorímetro hadrônico que será discutido em detalhes na próxima seção, é dedicado para medir a energia de hádrons.

Figura 7 - Sistema de Calorimetria do ATLAS



Fonte: PEQUENAO, 2008a.

Este sistema se inicia após o detector de trajetórias, a 1,5 metros do feixe de colisão. Como mencionado, o propósito é determinar a energia total de uma partícula, e isso é realizado através da absorção da mesma e de seus produtos de decaimento, os quais são medidos através da energia depositada nos calorímetros supracitados.

#### 1.4.1.1 Calorímetro Hadrônico de Telhas

O calorímetro de telhas do ATLAS, também conhecido como TileCal, é o objeto de estudo desta dissertação, e por isso, será abordado em maiores detalhes.

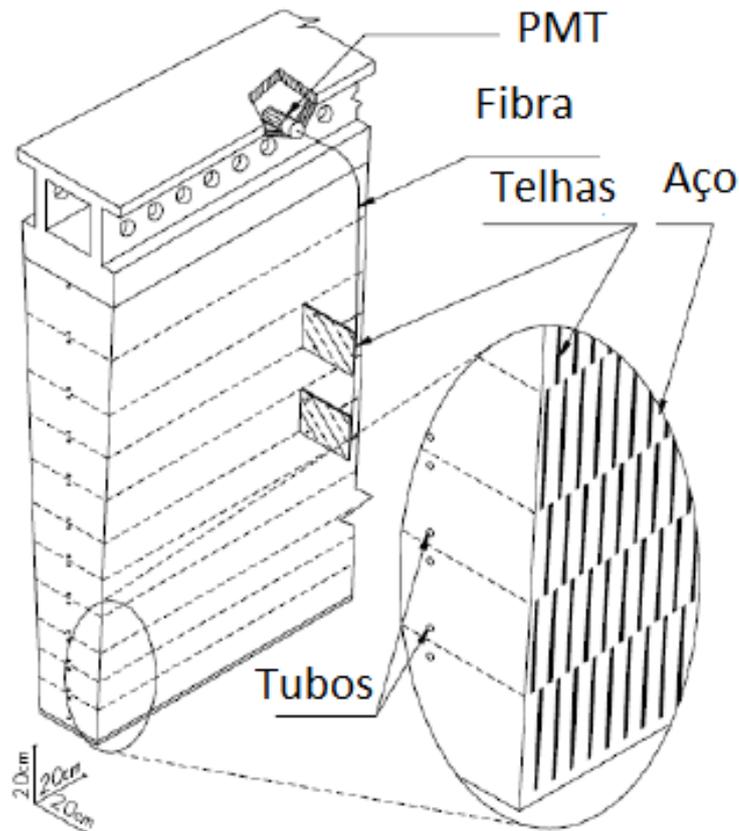
O TileCal é o principal calorímetro Hadrônico do ATLAS e seu objetivo é medir precisamente hádrons, jatos e taus. O TileCal também foi capaz de atingir suas expectativas durante os primeiros anos de operação do LHC (FRANCAVILLA, 2012).

Sua estrutura é dada por placas de aço como material absorvedor e telhas plásticas cintilantes como material ativo. Ambos os lados de cada telha são conectados a um tipo específico de fibra ótica, chamada WLS (*Wavelength Shifting*), composta por um material foto-fluorescente que absorve um fóton de alta frequência e emite múltiplos fótons de baixa frequência. Múltiplas telhas são agrupadas em células e cada célula é conectada a duas

fotomultiplicadoras, ou PMTs (*photo-multiplier tube*) (COLLABORATION, 1996).

Em outras palavras, as partículas hadrônicas provenientes das colisões passam pelo material pesado, armazenando a energia que é então parcialmente amostrada pelas telhas cintilantes na forma de luz. Através de fibras óticas, a luz gerada é transmitida e lida através das PMTs, as quais são responsáveis por converter a luz recebida em sinais elétricos que serão processados. A Figura 8 mostra a captação do sinal luminoso de um módulo e sua conversão para sinal elétrico através da PMT.

Figura 8 - Partições do TileCal



Fonte: COLLABORATION, 2008.

O TileCal não faz uso do sistema tradicional de coordenadas  $x$ ,  $y$  e  $z$ , invés disso, usa-se a transformação não linear:

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

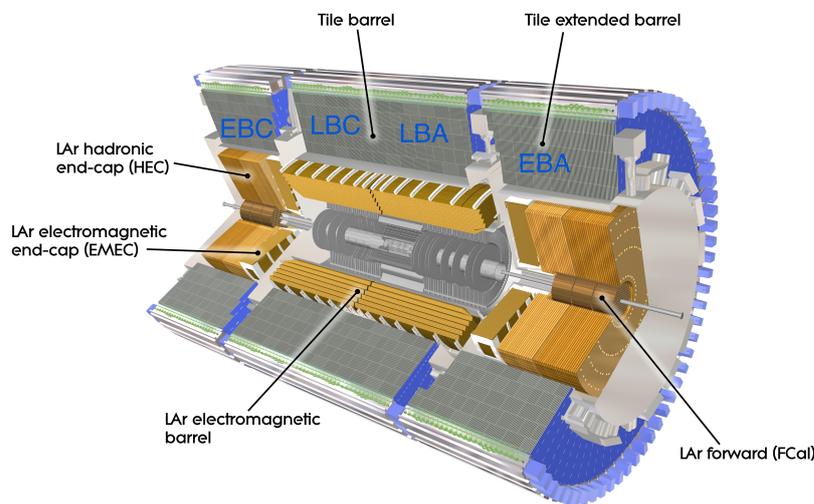
$$\eta = -\log\left(\tan \frac{\phi}{2}\right) \quad (2)$$

Onde  $\phi$  representa a rotação e  $\eta$  (também chamada de pseudo-rapidez) representa

a direção de projeção das partículas após colisão (BARBOSA, 2017). A pseudo-rapidez do TileCal é considerada ampla  $|\eta| < 1,7$  com cobertura total do azimute (SÁNCHEZ, 2010).

A Figura 7 já mostrou que o TileCal é cilíndrico e sua dimensão do raio interno é 2,28 metros e do raio externo é de 4,23 metros. Internamente é dividido em três partes: um Barril Central, ou LB (*Long Barrel*) de 5,64 metros de comprimento cobrindo  $|\eta| < 1$  e dois Barris Estendidos, ou EB (*Extended Barrel*) de 2,91 metros cada, cobrindo  $0,8 < |\eta| < 1$ . Os barris são separados em dois lados, lado A e lado C. Desta forma, as partições do Barril Central são identificadas por LBA e LBC, e as partições do Barril Estendido por EBA e EBC (PERALVA, 2012). A Figura 9 mostra todas as partes alinhadas.

Figura 9 - Partições do TileCal

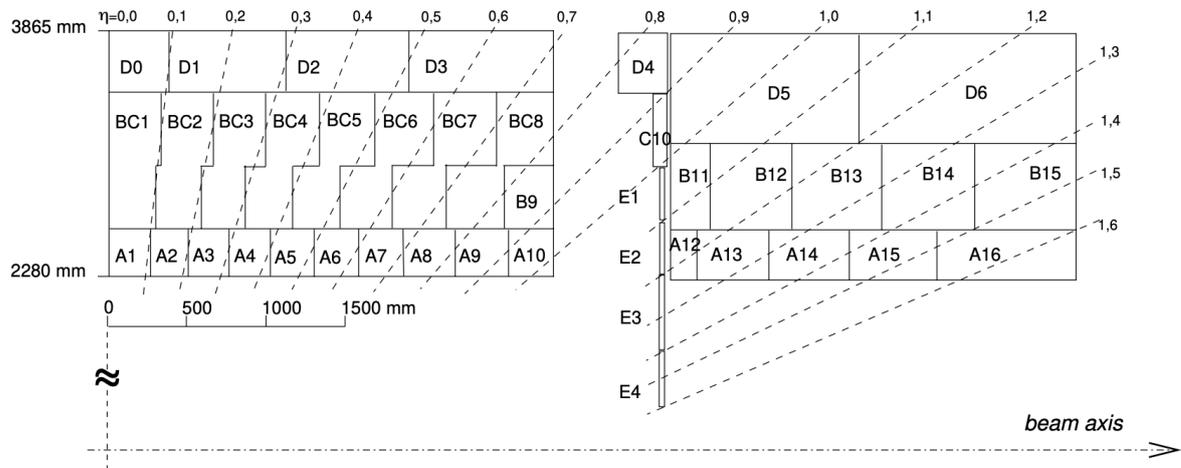


Fonte: Adaptado de PEQUENAO, 2008a.

Cada barril é composto por 64 módulos dispostos na direção  $\phi$ . Os módulos do TileCal são segmentados por três camadas de células, A, BC e D. A granularidade das células  $\Delta\eta \times \Delta\phi$  é de  $0,1 \times 0,1$  para a camada A e BC, e de  $0,2 \times 0,1$  para a camada D, como ilustrado na Figura 10. Além disso, o calorímetro possui uma camada especial de células (camada E) composta apenas por cintiladores, sem material pesado. Cada módulo do barril central e do barril estendido está dividido respectivamente em 23 e 16 células com dupla leitura, resultando em aproximadamente 10000 canais de leitura.

Quando uma partícula atravessa o calorímetro, esta gera um chuva de partículas e perde energia ao atravessar o material passivo (aço) e interage com o material ativo (telhas cintiladoras), emitindo luz. Como mencionado, essa luz é conduzida pelas fibras óticas e lidas por PMTs, que geram um pulso elétrico em resposta ao sinal luminoso. O pulso gerado pela PMT é condicionado por um circuito de conformação (*shaper*), o qual fornece um pulso com um formato conhecido e com uma amplitude proporcional à energia

Figura 10 - Seção transversal do barril e de uma das extensões do TileCal



Legenda: Distribuição das células para metade de um módulo do barril central (à esquerda) e de um módulo do barril estendido (à direita). Linhas tracejadas mostram a abrangência da pseudo-rapidez.

Fonte: CERN, 2013.

depositada (FULLANA et al., 2006). Este pulso analógico é digitalizado por um conversor Analógico-Digital (ADC) com frequência de amostragem de 40 MHz e uma janela com 7 amostras digitais (150 ns) cobre aproximadamente todo o pulso.

A energia depositada em cada célula do calorímetro pode ser calculada através da estimativa correta da amplitude do pulso de resposta, sendo esta uma abordagem comumente observada nos calorímetros modernos. A Figura 11 mostra o pulso de referência.

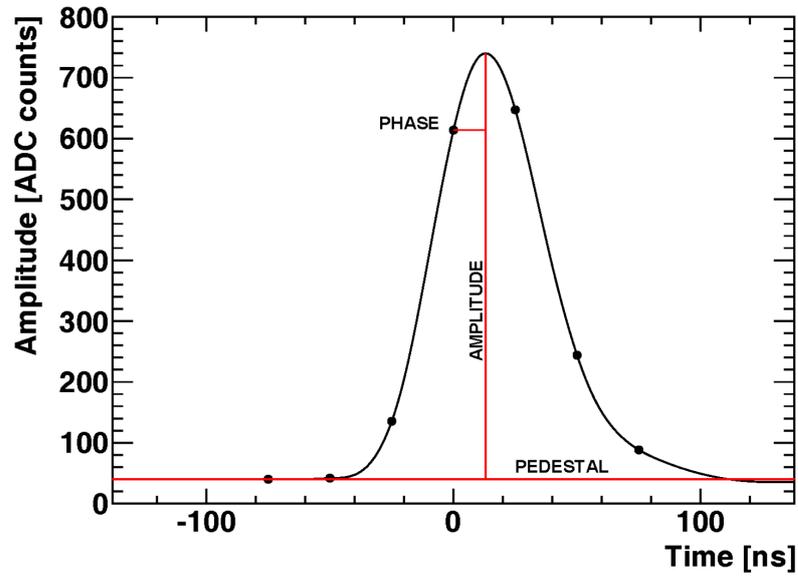
Fisicamente, apenas as células situadas na trajetória de uma partícula absorvem sua energia, portanto, devido à grande quantidade de canais, apenas as células com informação relevante são selecionadas e utilizadas na reconstrução de energia (PASTORE, 2016).

#### 1.4.2 Atualização do LHC

Luminosidade é uma grandeza que mede a habilidade de um acelerador de partículas produzir um número determinado de interações (BARBOSA, 2017), e, como a física de interesse é rara, é necessário uma quantidade muito grande de dados para se inferir algum comportamento nos dados (PERALVA, 2016).

A luminosidade é definida como um fator proporcional entre o número de eventos por segundo e a área do feixe de partículas, tendo unidade de  $cm^2s^{-1}$  (HERR; MURATORI, 2006). Sua equação, considerando o fator geométrico de redução igual a 1, é dada

Figura 11 - Pulso de referência do TileCal.



Legenda: A linha contínua representa o sinal analógico e os pontos representam as amostras digitais do pulso.

Fonte: MEYER, 2013.

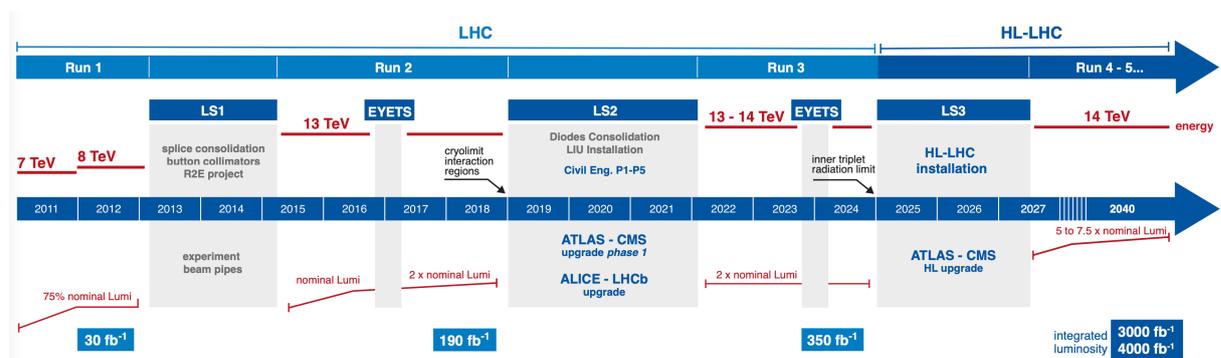
por

$$L \sim \frac{N^2}{t \cdot S'} \quad (3)$$

onde  $N$  corresponde ao número de prótons em cada feixe,  $t$  ao tempo entre colisões e  $S$  à seção transversal do feixe.

A Figura 12 mostra a linha do tempo do LHC para os seus respectivos LS (*Long Shutdown*) e atualizações para os próximos anos.

Figura 12 - Linha do tempo e atualizações previstas para o LHC até 2040.



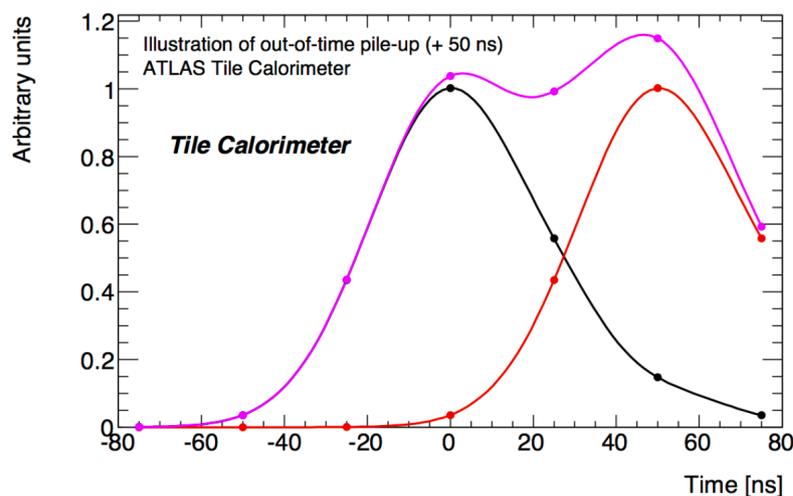
Fonte: WESTON, 2019.

Como pode-se observar pela linha do tempo, as atualizações previstas irão aumentar consideravelmente a luminosidade do LHC. Durante o *Run 3*, o LHC voltará a realizar colisões com uma expectativa de ocorrência de 50 a 80 interações por colisão ( $\langle\mu\rangle$ ), em média, a cada 25 ns (GONÇALVES, 2020).

Outra atualização, denominada *Phase II Upgrade*, ocorrerá entre os anos 2024 e 2026, antes dos períodos das próximas campanhas de aquisição de dados *Run 4* e *Run 5* (2026 a 2038). Até lá, o LHC planeja alcançar valores ainda maiores de luminosidade, podendo chegar a 200 interações por colisão (GONÇALVES, 2020). Com isso, mais interações próton-próton ocorrerão quando os feixes se cruzarem, o que significa uma maior luminosidade sendo gerada, produzindo então mais dados e consequentemente aumentando a probabilidade de ocorrer eventos físicos de interesse.

Este aumento introduz novos desafios para a eletrônica e para a área de processamento de sinais, visto que a probabilidade da ocorrência de eventos em um mesmo canal de leitura aumentará a sua ocupação. Tal efeito é chamado de empilhamento de sinais e é ilustrado pela Figura 13. Este fenômeno ocorre quando a resposta do calorímetro é mais lenta que o intervalo entre as colisões (CERQUEIRA, 2013).

Figura 13 - Empilhamento de sinais no TileCal.



Legenda: Empilhamento acontecendo quando dois sinais sendo captados em um curto período de tempo, gerando o sinal distorcido na cor magenta.

Fonte: KLIMEK, 2011.

## 2 RECONSTRUÇÃO DE ENERGIA EM CALORIMETRIA DE ALTAS ENERGIAS

Por séculos, a humanidade tem estudado e desenvolvido novas formas de filtragem, como por exemplo as de partículas sólidas, luz solar e energia elétrica. No entanto, após a revolução da informação a palavra "Filtro" ganhou um significado abstrato que vai além de seu significado convencional.

Assim como em seu uso convencional, usam-se filtros quando tem-se o interesse de barrar algo, como se fosse uma barreira. Tal conceito fora trazido para elétrica e computação através do design de filtros conhecidos como *Butterworth Filter*, *Chebyshev Filter*, e *Elliptical Filter* (ANDERSON; MOORE, 1979). No contexto de física de altas energias, diversos métodos e filtros podem ser implementados em cada experimento. Comumente, tais métodos possuem como base a solução do problema de estimação de parâmetros de um processo aleatório.

A estimação de energia dos calorímetros é um dos exemplos onde estimação de parâmetros é utilizada (PERALVA, 2012). E por isso, o Calorímetro de Telhas TileCal abordado na seção anterior faz uso do método conhecido como Filtro Ótimo 2, ou do Inglês *Optimal Filter 2* (OF2), para estimar amplitudes. Além deste, também para a estimação da amplitude, esta dissertação abordará no capítulo 3 o método denominado Filtro Casado, ou do Inglês *Matched Filter* (MF).

### 2.1 Estimação de Parâmetros

A maioria dos processos físicos podem ser representados ou aproximados por algum modelo matemático a fim de se analisar seus comportamentos em diferentes condições. Tais modelos podem ser classificados como determinísticos ou estocásticos.

Objetivamente, modelos determinísticos possuem saídas determinadas pelos valores de seus parâmetros e suas condições iniciais, ao passo que, modelos estocásticos são compostos de certas aleatoriedades de forma a possuírem saídas distintas mesmo usando os mesmos parâmetros e condições iniciais. Geralmente, os fenômenos físicos tendem a ser mais estocásticos que determinísticos, no entanto, a complexidade dos estocásticos e as respostas aceitáveis dos determinísticos muitas vezes fazem que quando possível este seja mais utilizado como modelo em diversas áreas.

A estimação de parâmetros, como o nome sugere, tem o objetivo de estimar um ou vários parâmetros de um processo aleatório e é utilizada em áreas como: processamento de sinais, áudio, vídeo, telecomunicações, sonar, e instrumentação. A informação de interesse pode estar corrompida por ruídos tipicamente aditivos onde em alguns casos o ser humano

é capaz de distingui-los, mas na maioria das vezes o problema em questão vai além da capacidade humana, necessitando de sistemas para tal (KAY, 1993).

Sua formulação para sinais é dada por um sinal contínuo no tempo  $x(t)$  que representa a saída de um determinado sistema como sendo a combinação do sinal  $g(t; \boldsymbol{\theta})$  de forma conhecida, onde  $\boldsymbol{\theta}$  é um vetor de parâmetros a ser estimado com um ruído aditivo dentro do sistema  $n(t)$ . A Equação (4) mostra tal modelo

$$x(t) = g(t; \boldsymbol{\theta}) + n(t). \quad (4)$$

Como os sistemas de computação trabalham de forma discreta, é conveniente transformar esta equação em sua forma discreta para  $k$  onde o tempo contínuo  $x(t)$  é representado de forma finita em  $N$  elementos como  $x(k) = \{x[0], \dots, x[N-1]\}$ . Com tal transformação, a Equação 4 se torna

$$x[k] = g[k; \boldsymbol{\theta}] + n[k]. \quad (5)$$

Sendo assim, o parâmetro que definirá o melhor ajuste ao conjunto de dados será  $\boldsymbol{\theta}$ , e este pode ser definido como um *estimador* em função de  $f(x)$

$$\hat{\boldsymbol{\theta}} = f(x[0], \dots, x[N-1]). \quad (6)$$

Estimadores são geralmente modelados matematicamente para o conjunto de dados alvo, e, devido a aleatoriedade dos dados, estes podem ser descritos por meio do uso de Funções de Densidade de Probabilidade (PDF). PDFs são funções parametrizadas pelo parâmetro  $\boldsymbol{\theta}$ , ou seja, diferentes  $\boldsymbol{\theta}$  irão gerar diferentes PDFs.

A eficiência do estimador está relacionado com a escolha da PDF. Esta por sua vez nem sempre é conhecida, e neste caso o ideal é buscar uma que seja consistente com as condições do problema e que sua solução matemática seja convergente. Logo, tal estimação baseada em PDFs é denominada estimação clássica quando os parâmetros de interesse são determinísticos, porém desconhecidos (KAY, 1993).

Em cenários que algumas propriedades do parâmetro de interesse são conhecidas a *priori*, como por exemplo o alcance deste parâmetro, este conhecimento pode ser incorporado na solução do problema. Desta forma, pode-se assumir que  $\boldsymbol{\theta}$  não é mais determinístico, mas sim uma variável aleatória, atribuindo a ela uma PDF dentro de um intervalo conhecido. Assim, cada valor produzido por este estimador estará contido neste intervalo e, portanto, o parâmetro desconhecido será visto como uma realização da variável aleatória  $\boldsymbol{\theta}$  (KAY, 1993; THEODORIDIS, 2015).

No entanto, muitas vezes os estimadores baseados em PDFs não possuem soluções analíticas ou são não-lineares, aumentando a dificuldade de sua implementação computacional. Nestes cenários, pode ser viável o uso de estimadores alternativos que possuem resultados sub-ótimos, mas que permitem sua implementação computacional. Um exemplo são os estimadores lineares que além de serem de fácil implementação, também possuem baixo custo computacional. Por esses motivos, tais estimadores são amplamente utilizados para a estimação de energia em calorimetria (CLELAND; STERN, 1994; PAGANINI; VULPEN, 2004; FULLANA et al., 2006). Um exemplo é o estimador de Máxima Verossimilhança que será abordado na próxima seção.

## 2.2 Estimador de Máxima Verossimilhança (MLE)

O estimador de Máxima Verossimilhança é utilizado como base para outros métodos de calorimetria do ATLAS e é uma das alternativas para se obter a expressão do Filtro Casado Determinístico que será apresentado posteriormente no Capítulo 3. Sua teoria se dá em torno do problema de estimação de parâmetros e da maximização da probabilidade  $P(\boldsymbol{\theta}|\mathbf{x})$ , ou seja, obter o valor de  $\hat{\boldsymbol{\theta}}$  que resulta na maior probabilidade condicional de  $P(\boldsymbol{\theta}|\mathbf{x})$ . Esta probabilidade geralmente não é conhecida, no entanto, com o auxílio do teorema de Bayes sabe-se que a probabilidade de um evento pode ser determinada a partir do conhecimento a *priori* das condições do mesmo

$$P(\boldsymbol{\theta}|x) = \frac{P(\mathbf{x}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{x})}, \quad (7)$$

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

Considerando que este processo possui suas variáveis aleatórias independentes e igualmente distribuídas, a maximização da probabilidade  $P(\boldsymbol{\theta}|\mathbf{x})$  pode ser obtida através da maximização da densidade condicional  $P(\mathbf{x}|\boldsymbol{\theta})$ , chamada função de verossimilhança. Logo, a melhor estimativa de  $\boldsymbol{\theta}$  é o valor que maximiza  $P(\mathbf{x}|\boldsymbol{\theta})$ , dada pela equação

$$\frac{\partial P(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \quad (8)$$

Como mencionado, o MLE (Maximum Likelihood Estimation) requer conhecimentos prévios a respeito do processo aleatório que envolve o parâmetro de interesse. Na prática, o problema costuma não ter solução analítica ou a mesma possui solução demasiadamente complexa ou inviável, resultando também em estimadores complexos. Por isso,

aproximações são usadas nos projetos dos estimadores.

Na área da calorimetria é comum ter um pulso eletrônico de forma conhecida, na qual permite os sistemas de calibração garantir a estabilidade dos parâmetros, podendo assim assumi-los como sinais de aspectos determinísticos. Com isso, a estimação de energia pode ser realizada através da estimação da amplitude de um pulso de forma fixa. Partindo do problema de estimação definido pela Equação 9, pode-se modelar o pulso  $x$  como

$$x[k] = Ag[k] + n[k] \quad k = 0, 1, \dots, N - 1, \quad (9)$$

em que  $A$  representa o parâmetro desconhecido, neste caso podendo ser interpretado como amplitude ou energia a ser estimada,  $g[k]$  o sinal conhecido e  $n[k]$  representa o ruído aditivo. A estimativa da amplitude  $\hat{A}_{mle}$  pode ser encontrada resolvendo a seguinte equação

$$\frac{dP(\mathbf{x}|\hat{A}_{mle})}{d\hat{A}_{mle}} = 0. \quad (10)$$

Supondo o caso particular onde a Equação 9 pode ser modelada por uma distribuição gaussiana, e o processo multivariável possui matriz de covariância  $\mathbf{C}$ , a função densidade de probabilidade conjunta é

$$P(\mathbf{x}|\hat{A}_{mle}) = \frac{1}{\sqrt{2\pi \det(\mathbf{C})}} \exp\left(-\frac{(\mathbf{x} - A\mathbf{g})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{g})}{2}\right), \quad (11)$$

onde  $\det$  representa a determinante da matriz.

Como  $P(\mathbf{x}|\hat{A}_{mle})$  é uma função estritamente convexa, pode-se aplicar o logaritmo, resultando em

$$\ln[P(\mathbf{x}|\hat{A}_{mle})] = \frac{-1}{2} \frac{(\mathbf{x} - A\mathbf{g})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{g})}{\sqrt{2\pi \det(\mathbf{C})}}. \quad (12)$$

Derivando-o em função de  $\hat{A}_{mle}$  e igualando a zero tem-se

$$\frac{d \ln[P(\mathbf{x}|\hat{A}_{mle})]}{d\hat{A}_{mle}} = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{g} - A \mathbf{g}^T \mathbf{C}^{-1} \mathbf{g} = 0. \quad (13)$$

Explicitando  $\hat{A}_{mle}$

$$\hat{A}_{mle} = \frac{\mathbf{x}^T \mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}. \quad (14)$$

Considerando  $\mathbf{w}$  como um vetor de coeficientes dado por

$$\mathbf{w} = \frac{\mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}}, \quad (15)$$

e substituindo na Equação 14 obtém-se o estimador MLE

$$\hat{A}_{mle} = \sum_{k=0}^{N-1} x[k]w[k]. \quad (16)$$

O MLE pode ser implementado em processadores digitais como um estimador através de um filtro FIR (Finite Impulse Response) de ordem  $N - 1$  e coeficientes  $\mathbf{w}$  (MITRA; KUO, 2006).

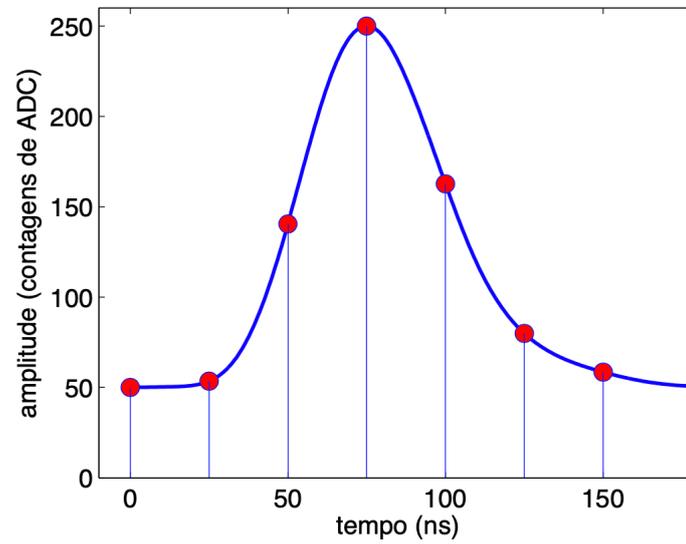
### 2.3 Estimadores de energia em calorimetria

Os calorímetros de física de altas energias normalmente possuem dois sinais envolvidos durante o processo de estimação de energia, o pulso de interesse geralmente como a Figura 14 e o ruído de fundo que é majoritariamente composto por ruído eletrônico (PERALVA, 2015a) e é capaz de distorcer o sinal de interesse como mostra a Figura 15. Como este é composto de diversas fontes, pelo teorema do limite central (PEEBLES, 1987) pode ser aproximado por uma distribuição gaussiana, o que simplifica o projeto de estimadores de energia, uma vez que apenas a covariância do ruído é utilizada para descrever todo o processo aleatório.

O circuito de condicionamento de sinais presente nos calorímetros modernos fornece um pulso de forma fixa, estável e tipicamente unipolar (CALDWELL et al., 1992; D0, 2005; TYLMAD, 2009).

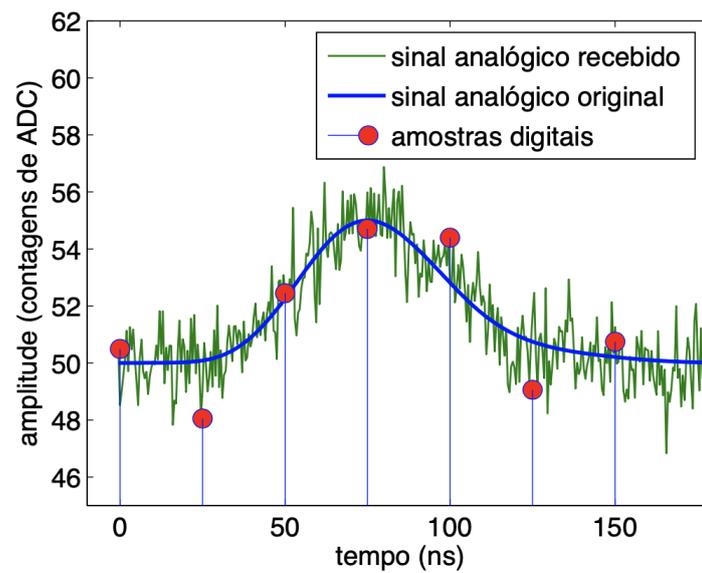
Em cenário comuns a quantidade de dados gerados é alta, no entanto, há experimentos com feixes de partículas mais densos e com maior quantidade de energia que geram cenários cujas colisões proporcionam alta luminosidade devido a alta taxa de eventos. Quando isso acontece, o calorímetro é sobrecarregado e a velocidade de leitura dos dados é inferior a velocidade que as colisões ocorrem, podendo então haver sobreposição de sinais provenientes de colisões subsequentes. Tal efeito, também chamado de empilha-

Figura 14 - Típica resposta de um calorímetro para sinais digitais sem ruído.



Fonte: PERALVA, 2015a

Figura 15 - Resposta de um calorímetro considerando sinais originais, analógicos e amostras digitais.



Fonte: PERALVA, 2015a

mento de sinais, distorce a forma do pulso de interesse, comprometendo sua estimação.

Uma característica dos calorímetros de altas energias é o fato de usarem algoritmos de estimação de amplitude baseados na soma ponderada das  $N$  amostras digitais recebidas (CLELAND; STERN, 1994; PAGANINI; VULPEN, 2004; FULLANA et al., 2006). Esta é uma boa abordagem *online* para sistemas de alta taxa de eventos onde uma das principais necessidades é a resposta rápida de um algoritmo de estimação.

Além do processamento e análise *online* citada, também há o mesmo de forma *offline*, ou seja, análise dos dados que foram previamente armazenados. O valor da energia referente a cada canal armazenado é então utilizado como patamar para decidir quais canais de leitura contém informação relevante para o processamento *offline* e reconstrução do evento.

Apesar de possuírem respostas rápidas, estes métodos tradicionais possuem a eficácia de seus resultados afetados no cenário de empilhamentos de sinais pois os consideram como fontes adicionais de ruído as quais são tratadas utilizando sua informação da estatística de segunda ordem. Ou seja, tais estimadores permanecem ótimos para condições em que o ruído pode ser modelado por uma distribuição gaussiana, no entanto, no cenário de empilhamento de sinais a característica do ruído é afetada e componentes não gaussianas são incorporadas ao ruído, o que leva estes métodos à condições de operação sub-ótimas (KAY, 1993; PERALVA, 2015b).

## 2.4 Filtro Ótimo

O Filtro Ótimo (OF) é um estimador de energia utilizado nos sistemas de calorimetria do ATLAS para a estimação de energia e se enquadra nos métodos descritos na seção anterior. O OF é uma técnica de minimização da variância do erro que usa o conhecimento da forma do pulso para estimar a amplitude que é proporcional à energia do sinal. O ruído de fundo do ATLAS é predominantemente constituído de ruído eletrônico cuja distribuição é gaussiana, e, por esse motivo, é bastante empregado em tais sistemas de calorimetria (BERTUCCIO et al., 1992; CLELAND; STERN, 1994; DELMASTRO, 2003).

A versão deste estimador implementado no TileCal é chamado de OF2 e está em operação desde 2014, sendo utilizado para reconstrução *online* e *offline* de energia (FULLANA et al., 2006). Um critério importante para algoritmos de estimação *online* no ATLAS é que precisam ser simples e eficientes para atender aos requisitos de tempo de processamento (LUNDBERG, 2010). Com base nisto, a amplitude do sinal de entrada é estimada através de um processo de baixo custo computacional, que utiliza uma operação

de soma ponderada dada por

$$\hat{A}_{OF} = \sum_{k=0}^{N-1} x[k]w[k], \quad (17)$$

onde  $w[k]$  são os coeficientes do filtro e  $x[k]$  as amostras do sinal recebido. Já para o processamento *offline*, outros algoritmos podem ser utilizados, visto que performance e baixa complexidade computacional não são tão importantes se comparados à abordagem *online*.

No TileCal, uma amostra digital  $x$  no instante  $k$  pode ser modelada por

$$x[k] = Ag[k - \tau] + n[k] + ped, \quad (18)$$

onde  $A$  é a amplitude do pulso,  $g[k]$  são as amostras do pulso de referência (circuito de *shaper*) normalizado,  $n[k]$  o ruído eletrônico,  $\tau$  o desvio de fase do sinal e  $ped$  o pedestal ou linha de base, um parâmetro constante adicionado ao sinal analógico antes de sua digitalização.

Para calcular os coeficientes  $w[k]$  do estimador OF2, o sinal digital recebido do TileCal pode ser aproximado por uma série de Taylor de primeira ordem, dada por

$$x[k] = Ag[k] - A\tau\dot{g}[k] + n[k] + ped, \quad (19)$$

onde  $\dot{g}[k]$  representa a derivada do pulso de referência  $g$ , e  $k = 0, 1, 2, \dots, N - 1$ .

Para garantir a imparcialidade do estimador, é exigido que o valor esperado de  $\hat{A}_{OF}$  seja igual a  $A$ . Logo, substituindo a Equação 19 na Equação 17 e considerando que a média do ruído é zero ( $E\{n[k]\} = 0$ ), tem-se

$$\begin{aligned} E\{\hat{A}_{OF}\} &= \sum_{k=0}^{N-1} (Aw[k]g[k] - A\tau w[k]\dot{g}[k] + w[k]ped) \\ &= A \sum_{k=0}^{N-1} w[k]g[k] - A\tau \sum_{k=0}^{N-1} w[k]\dot{g}[k] + ped \sum_{k=0}^{N-1} w[k]. \end{aligned} \quad (20)$$

A fim de que o estimador seja independente do pedestal e da fase, as seguintes

restrições são estabelecidas

$$\sum_{k=0}^{N-1} w[k]g[k] = 1, \quad (21)$$

$$\sum_{k=0}^{N-1} w[k]\dot{g}[k] = 0, \quad (22)$$

$$\sum_{k=0}^{N-1} w[k] = 0. \quad (23)$$

A primeira restrição garante a imparcialidade do estimador, enquanto a segunda e terceira restrições garantem, respectivamente, que o estimador seja imune à fase e à flutuações da linha de base, ou pedestal.

Vale ressaltar que a variância de um estimador linear é dada pela expressão:

$$E\{(\hat{A}_{OF} - A)^2\} = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} w[k]w[j]C[k, j] = \mathbf{w}^T \mathbf{C} \mathbf{w}, \quad (24)$$

onde  $\mathbf{w}$  é o vetor de pesos do estimador e  $\mathbf{C}$  a matriz de covariância do ruído dada por

$$C[k, j] = E\{(M[k] - E\{M[k]\})(M[j] - E\{M[j]\})\}, \quad (25)$$

sendo  $M$  a variável aleatória associada ao ruído no instante  $k$ .

Para determinar os pesos  $\mathbf{w}$ , é necessário minimizar a expressão da variância do estimador descrita pela Equação 24 aplicando as restrições descritas pelas Equações 21, 22, e 23. Aplicando os multiplicadores de Lagrange (BERTUCCIO et al., 1992)  $\lambda$ ,  $\xi$  e  $v$ , a solução do problema corresponde ao ponto de mínimo da função

$$I_w = \sum_{k,j=0}^{N-1} w[k]w[j]C[k, j] - \lambda \left( \sum_{k=0}^{N-1} w[k]g[k] \right) - \xi \left( \sum_{k=0}^{N-1} w[k]\dot{g}[k] \right) - v \left( \sum_{k=0}^{N-1} w[k] \right), \quad (26)$$

onde seu mínimo será dado por sua derivada igualada a zero

$$\frac{dI_w}{dw[k]} = 2 \sum_{j=0}^{N-1} w[j]E\{M[k]M[j]\} - \lambda g[k] - \xi \dot{g}[k] - v = 0. \quad (27)$$

Por fim, considerando um processo estacionário (PAPOULIS; PILLAI, 2002), pode-se assumir  $E\{M[k]\} = E\{M[j]\} = 0$  e  $E\{M[k]^2\} = E\{M[j]^2\}$  para qualquer  $k$  e  $j$ , resultando no seguinte sistema de equações

$$\sum_{k=0}^{N-1} w[k]g[k] = 1, \quad (28)$$

$$\sum_{k=0}^{N-1} w[k]\dot{g}[k] = 0, \quad (29)$$

$$\sum_{k=0}^{N-1} w[k] = 0, \quad (30)$$

$$\sum_{k=0}^{N-1} w[k]C[k, j] - \lambda g[k] - \xi \dot{g}[k] - v = 0, \quad \forall k. \quad (31)$$

O sistema pode ser dado em forma matricial

$$\begin{pmatrix} C[1,1] & C[1,2] & \dots & C[1,N] & -g[1] & -\dot{g}[1] & -1 \\ C[2,1] & C[2,2] & \dots & C[2,N] & -g[2] & -\dot{g}[2] & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ C[N,1] & C[N,2] & \dots & C[N,N] & -g[N] & -\dot{g}[N] & -1 \\ g[1] & g[2] & \dots & g[N] & 0 & 0 & 0 \\ \dot{g}[1] & \dot{g}[2] & \dots & \dot{g}[N] & 0 & 0 & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w[1] \\ w[2] \\ \vdots \\ w[N] \\ \lambda \\ \xi \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad (32)$$

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 uma matriz identidade. A solução deste sistema resulta nos pesos  $w[k]$  do método OF2 os quais podem ser utilizados para estimação de energia *online* e *offline*.

### 3 FILTRO CASADO

O objetivo principal desta dissertação é analisar o uso da técnica baseada em filtragem casada para a estimação da energia em calorimetria de altas energias. Sua teoria e implementação para duas abordagens serão apresentadas em detalhes.

Como abordado na seção 2.3, calorímetros modernos usam métodos de reconstrução de energia através da estimação da amplitude do sinal, a fim de que possam selecionar os canais desejados com base em suas respectivas energias. Por este motivo, métodos de estimação lineares baseados na minimização da variância do estimador têm sido utilizados (PERALVA, 2015a).

Diferente desta abordagem clássica, o Filtro Casado ou em inglês Matched Filter (MF) formula o problema a partir da teoria de detecção, visando resolver a seleção ótima dos canais com deposição de energia supracitados. Por fim, através de calibração, a estimação da energia é realizada a partir da saída do algoritmo de detecção.

Um exemplo de utilização deste método foi a aproximação desenvolvida para separação elétron / pión usando o calorímetro de espaguete (SEIXAS; CALOBA; PINTO, 1996b) que foi recentemente aplicada em um gatilho de múon assistido por calorimetria hadrônica (CIODARO; SEIXAS; CERQUEIRA, 2014).

Como mencionado, o MF terá inicialmente sua teoria base abordada seguida por duas abordagens. A primeira, chamada de determinística, será usada para casos onde o sinal de interesse possui forma constante, ao passo que a segunda, chamada de estocástica, será usada para casos onde o sinal de interesse possui flutuações estatísticas.

#### 3.1 Detecção do Sinal

Devido ao fato dos calorímetros serem altamente segmentados, muitas células de leitura não possuem informações relevantes para a reconstrução do sinal. Por isso, faz-se necessário o uso de um algoritmo que seja capaz de detectar sinais imersos em ruídos a fim de que apenas canais relevantes, ou seja, que possuem deposição de energia sejam considerados na estimação.

O Filtro Casado é uma técnica de detecção que pode ser aplicada para diferentes tipos de ruídos, ou seja, não é limitado apenas a ruídos gaussianos geralmente presentes na eletrônica.

Este filtro modela a saída de um processo aleatório  $G$ , que neste caso, será a saída do calorímetro. Tal processo além do sinal de interesse, também é constituído de incertezas como desvio de fases e deformidades conforme descrito na Equação 33. Em resumo, pode-se dizer que o MF é uma técnica de detecção de sinais que visa maximizar

a relação sinal-ruído.

Neste método, sinal recebido é modelado como

$$x_k = Ag_k + n_k + ped \quad k = 0, 1, 2, \dots, N - 1, \quad (33)$$

onde  $g_k$  e  $n_k$  correspondem ao sinal e ao ruído respectivamente, e  $ped$  (pedestal) é a linha de base que é um valor constante e pode ser estimado através de tomadas de dados de calibração (MARJANOVIĆ, 2019). O parâmetro  $A$  é a amplitude do sinal recebido que se deseja estimar.

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

$$\begin{aligned} H_0 : x[k] &= n[k] & k &= 0, 1, \dots, N - 1 \\ H_1 : x[k] &= g[k] + n[k] & k &= 0, 1, \dots, N - 1, \end{aligned} \quad (34)$$

onde  $H_0$  é a hipótese constituída apenas de ruído e  $H_1$  considera a hipótese do sinal recebido constituir sinal mais ruído.

Seja  $x = \{x[0], x[1], \dots, x[N - 1]\}$  um sinal recebido, e de acordo com o teorema de Bayes, a relação que maximiza a eficiência de detecção é dada pela razão de verossimilhança (VANKOV, 2013) dada por

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \quad (35)$$

em que os elementos  $p(\mathbf{x}|H_1)$  e  $p(\mathbf{x}|H_0)$  correspondem a probabilidade de  $x$  pertencer às hipóteses  $H_1$  e  $H_0$ , respectivamente. O parâmetro  $\gamma$  corresponde ao patamar de detecção, que caso sejam conhecidas as probabilidades a *priori*, pode ser obtido para minimizar a probabilidade de erro ou então  $\gamma$  também pode ser definido por projeto. Desta forma, decide-se a favor da hipótese  $H_1$  caso a razão de verossimilhança do sinal recebido for maior ou igual à  $\gamma$  e a favor de  $H_0$  caso contrário. A razão de verossimilhança maximiza a relação sinal-ruído também conhecida como SNR (*Signal-to-Noise Ratio*) (TREES, 2004).

É comum que funções densidade de probabilidade sejam desconhecidas para a maioria dos problemas de detecção e como alternativa o algoritmo pode estimar  $p(\mathbf{x}|H_1)$  e  $p(\mathbf{x}|H_0)$  através de um conjunto de dados, que apesar de possível, não apresentará resultados ótimos quando comparado à solução analítica.

Se as amostras de ruído deste conjunto de dados estiverem correlacionadas, uma etapa de pré-processamento chamada de branqueamento ou em inglês *whitening filter* é necessária antes da operação do MF. O branqueamento tem o objetivo de descorrelacionar as amostras de ruído e permitir que o teste de razão de verossimilhança se beneficie de informações estatísticas ortogonais. Nesse sentido, o sinal de entrada recebido  $\mathbf{x}$  após o

filtro de branqueamento é representado como  $\mathbf{x}'$ . No caso de as amostras de ruído não estarem correlacionadas, a operação de branqueamento pode ser suprimida, uma vez que conduz à matriz de identidade. Vale ressaltar que o design do MF pode ser desenvolvido para qualquer formato de resposta do calorímetro (PERALVA et al., 2021).

O MF será abordado a seguir em duas abordagens mencionadas no início do capítulo, o Filtro Casado Determinístico e o Filtro Casado Estocástico.

### 3.2 Filtro Casado Determinístico

Filtro Casado Determinístico ou do inglês DMF (*Deterministic Matched Filter*) é utilizado para casos onde o calorímetro além de fornecer um pulso fixo e estável em sua saída, também tem-se o ruído de fundo gaussiano, o sinal de interesse  $g = \{g[0], g[1], \dots, g[N-1]\}$  é determinístico e a estimativa da matriz de covariância  $\mathbf{C}$  do ruído de fundo é suficiente para caracterizar os sinais. Nestas condições, a distribuição gaussiana multivariada é utilizada para descrever as funções de verossimilhança para a hipótese  $H_1$  e  $H_0$  (PERALVA et al., 2021). Sendo assim,  $p(\mathbf{x}|H_1)$  e  $p(\mathbf{x}|H_0)$  são definidas respectivamente por

$$p_{\mathbf{x}|H_1}(\mathbf{x}|H_1) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp\left(\frac{-(\mathbf{x} - \mathbf{g})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{g})}{2}\right), \quad (36)$$

$$p_{\mathbf{x}|H_0}(\mathbf{x}|H_0) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}|}} \exp\left(\frac{-\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{2}\right). \quad (37)$$

Aplicando as Equações 36 e 37 na Equação 35 tem-se

$$\Lambda(\mathbf{x}) = \frac{\exp\left(\frac{-(\mathbf{x} - \mathbf{g})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{g})}{2}\right)}{\exp\left(\frac{-\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}}{2}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \quad (38)$$

Logo, a expressão final da abordagem determinística será dada por

$$\mathbf{x}^t \mathbf{C}^{-1} \mathbf{s} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma', \quad (39)$$

onde  $\mathbf{x}$  é um vetor constituído dos sinais recebidos acrescido do pedestal,  $\mathbf{C}^{-1}$  é a inversa da matriz de covariância do ruído, e  $\mathbf{s}$  é um sinal de referência.

A Equação 39 é a expressão final do Filtro Casado para detecção de sinais determinísticos corrompidos por ruído gaussiano de média nula. Para o caso onde o ruído

gaussiano é branco e as amostras estão descorrelacionadas, a matriz  $\mathbf{C}$  se torna diagonal e por isso pode ser desprezada. Desta forma, os coeficientes do MF correspondem ao próprio sinal determinístico de interesse  $\mathbf{g}$ . Já para o caso onde o ruído de fundo é gaussiano e correlacionado, a inversa da matriz de covariância do ruído ( $\mathbf{C}$ ) tem a função de descorrelacionar as amostras do sinal. Sendo assim, os coeficientes do Filtro Casado correspondem ao vetor  $\mathbf{C}^{-1}\mathbf{g}$ .

### 3.2.1 Amplitude do Filtro Casado Determinístico

No que tange a estimação da amplitude (energia), o sinal recebido é representado pela Equação 33. Visto que o ruído  $E[n_k] = 0$ :

$$E[\mathbf{x}] = A\mathbf{s}, \quad (40)$$

onde  $\mathbf{s}$  corresponde ao pulso normalizado de amplitude unitária e  $A$  a amplitude. Portanto, a operação de detecção pode ser escrita por:

$$E[\mathbf{x}^t\mathbf{C}^{-1}\mathbf{s}] = (A\mathbf{s})^t\mathbf{C}^{-1}\mathbf{s}, \quad (41)$$

e resolvendo essa expressão para a variável  $A$ , resultando na expressão da Equação 42, a qual corresponde à estimativa da amplitude do sinal de entrada.

$$\hat{A}_{DMF} = \frac{\mathbf{x}^t\mathbf{C}^{-1}\mathbf{s}}{\mathbf{s}^t\mathbf{C}^{-1}\mathbf{s}}. \quad (42)$$

O numerador  $\mathbf{x}^t\mathbf{C}^{-1}\mathbf{s}$  representa o MF para detecção, enquanto o denominador  $\mathbf{s}^t\mathbf{C}^{-1}\mathbf{s}$  pode ser representado por uma constante que normaliza a operação do MF visando recuperar a amplitude do sinal. Neste caso,  $\mathbf{C}^{-1}$  representa o processo de branqueamento do ruído (KESSY; LEWIN; STRIMMER, 2018).

Vale ressaltar que a Equação 42 do Filtro Casado Determinístico para ruídos gaussianos e sinais determinísticos também pode ser obtida através do MLE, conforme descrito na seção 2.2 e apresentado pela Equação 14.

### 3.3 Filtro Casado Estocástico

O Filtro Casado Estocástico ou do inglês SMF (*Stochastic Matched Filter*) pode ser obtido através de um processo estocástico. Por isso, seja  $G$  um processo estocástico e  $g[k]$  da Equação 34 o sinal de interesse. Supondo que tal sinal forme uma sequência estocástica, este pode ser representado pela sua expansão de Karhunen-Loève que pode ser obtida através da Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*), sendo processada através de um conjunto de dados. A técnica PCA obtém uma nova base ótima através de transformações lineares resultando, em novos eixos ordenados de forma que cada um represente a maior quantidade de informação disponível (JOLLIFFE; CADIMA, 2016).

Sendo assim através da PCA, o vetor de amostras digitais recebidas pode ser mapeado sobre os autovetores da matriz de covariância de  $G$  sem perda de informação

$$x_{pca}[k] = \mathbf{x}\phi_{G_k} \quad k = 0, 1, \dots, N - 1, \quad (43)$$

em que  $\phi_{G_k}$  corresponde aos autovetores da matriz de covariância do processo  $G$ , e  $\Phi_G = \{\phi_{G_0}, \phi_{G_1}, \dots, \phi_{G_{N-1}}\}$ . Como resultado têm-se amostras  $x_{pca}[k]$  descorrelacionadas. Assumindo  $g[k]$  gaussiana e estatisticamente independente da sequência do ruído (branco gaussiano), é possível obter o teste da razão de máxima verossimilhança decomposto em dois termos,  $I_D$  e  $I_R$  (TREES, 2004).

$$\Lambda(\mathbf{x}) = I_D + I_R, \quad (44)$$

em que para o processo estocástico  $G$ :

$$I_D = \mathbf{m} \left[ \sum_{k=0}^{N-1} \frac{1}{\lambda_k + \frac{N_0}{2}} \phi_k \phi_k^t \right] \mathbf{x}^t, \quad (45)$$

$$I_R = \frac{1}{N_0} \mathbf{x} \left[ \sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + \frac{N_0}{2}} \phi_k \phi_k^t \right] \mathbf{x}^t. \quad (46)$$

O parâmetro  $\mathbf{m}$  corresponde ao vetor média normalizado do sinal recebido,  $N_0/2$  representa a energia do ruído. Já os parâmetros  $\lambda_k$  e  $\phi_k$  correspondem aos autovalores e autovetores obtidos através da PCA, respectivamente.

### 3.3.1 Amplitude do Filtro Casado Estocástico

A saída do MF estocástico pode ser utilizada para estimar a amplitude do sinal recebido  $\mathbf{x}$ , e a hipótese  $H_1$  pode ser representada pela seguinte expressão:

$$I_D + I_R = \mathbf{b}_1(\mathbf{x})^t + b_2(\mathbf{x})^t(\mathbf{x}) = \theta, \quad (47)$$

onde  $\theta = \Lambda(\mathbf{x})$ , e os elementos  $b_1$  e  $b_2$  correspondem a:

$$\mathbf{b}_1 = \mathbf{m} \left[ \sum_{k=0}^{N-1} \frac{1}{\lambda_k + N_0/2} \Phi_k \Phi_k^t \right], \quad (48)$$

$$b_2 = \frac{1}{N_0} \sum_{k=0}^{N-1} \frac{\lambda_k}{\lambda_k + N_0/2} \Phi_k \Phi_k^t. \quad (49)$$

Visto que a informação de interesse é a amplitude, o sinal recebido pode ser representado pelo sinal médio normalizado  $\mathbf{m}$  com a correspondente amplitude  $A$  associada, ou  $E[\mathbf{x}] = A\mathbf{m}$ , o qual resulta na seguinte expressão:

$$(\mathbf{b}_1 A\mathbf{m})^t + (b_2 A\mathbf{m})^t(A\mathbf{m}) = \theta. \quad (50)$$

Desta forma, a estimação da amplitude pode ser encontrada resolvendo a seguinte expressão em termos de  $A$ :

$$\hat{A}_{SMF} = \frac{-\mathbf{b}_1 \mathbf{m}^t + \sqrt{(\mathbf{b}_1 \mathbf{m}^t)^2 + 4(b_2 \theta)}}{2b_2}. \quad (51)$$

Vale ressaltar que apenas a parte positiva da função de raiz quadrada é utilizada, visto que a estimação da amplitude requer que  $A \geq 0$ . Assim, a abordagem estocástica do MF se torna robusta contra estimativas negativas indesejadas da amplitude.

## 3.4 Aplicação em alta taxas de eventos

O MF será usado como estimador da Amplitude do sinal em diversos cenários. Um deles, é o cenário comum, onde não há alta taxa de eventos acontecendo nos canais, ou seja, as ocupações dos mesmos são baixas, o que dificulta a ocorrência de empilhamento de sinais. Um cenário similar onde não há a ocorrência de empilhamento de sinais já foi abordado por (PERALVA et al., 2021), onde foi mostrado que o MF é capaz de realizar aproximações satisfatórias.

Além deste, este trabalho fundamentado na atualização do LHC descrita na Seção 1.4.2, se propõe ao uso do MF em cenários de alta taxa de eventos, o que acontecerá com maior frequência após a atualização prevista do LHC. Nestes cenários, os canais de leitura do TileCal tendem a apresentar alta ocupação, e, quando isso acontece, as chances de ocorrer empilhamento de sinais é maior.

## 4 RESULTADOS

Este capítulo abordará as estratégias utilizadas para a elaboração dos conjuntos de dados assim como a análise dos resultados obtidos quando os mesmos foram aplicados aos métodos Filtro Ótimo, Filtro Casado Determinístico e Filtro Casado Estocástico.

Por entender que a ciência é um patrimônio da sociedade e por apoiar o código livre, todo código a respeito da engenharia de software, arquitetura, algoritmos e gráficos gerados nesta dissertação está publicada como código livre, sujeito a constante avaliação e contribuição por parte da comunidade. Este projeto pode ser acessado em: <https://github.com/geovanecom/Masters-Degree>.

### 4.1 Estratégia Utilizada

A fim de avaliar a eficiência de estimação de energia dos métodos supracitados, todos os resultados e análises geradas foram baseadas em dois ambientes.

O primeiro ambiente criado foi o ambiente simulado. Este além de reproduzir condições semelhantes às encontradas em problemas reais, também permite ter total controle dos dados gerados através do algoritmo e parâmetros utilizados. Seu principal objetivo é reproduzir as saídas digitais do TileCal de acordo com seus parâmetros, e com isso simular canais de leitura cujas taxas de ocupação variam de 0 à 100%. Onde 0% significa nenhum empilhamento e 100% significa que toda colisão irá gerar um sinal de interesse no canal de leitura considerado.

Após, é possível avaliar quão boa é a estimativa da amplitude em ambientes cujos os sinais estão sofrendo empilhamento de sinais em função da ocupação dos canais de leitura. Neste ambiente, é possível repetir os testes  $N$  vezes a fim de se extrair os melhores valores médios e entender o quão esparsos estes são em função de cada taxa de ocupação.

O segundo ambiente criado foi o ambiente híbrido. Este ambiente foi criado com o propósito de conter os dados mais próximos possíveis dos dados reais, mas ainda permitir certo controle a fim de que se saiba qual é a amplitude esperada e compará-la com a amplitude aproximada pelos métodos numéricos. Tal controle se dá ao fato dos dados gerados serem compostos de ruídos reais  $\langle \mu \rangle$  acrescidos de um valor previamente conhecido, isto é, a amplitude esperada. Neste caso, maiores valores de  $\langle \mu \rangle$ , ocasionarão maiores números médios de interações por colisão.

Durante a análise do comportamento do MF em situações onde ocorrem empilhamentos de sinais, optou-se por utilizar 100% das componentes da PCA para o SMF para a análise geral da Seção 4.3. Após tais análises, uma nova análise será feita exclusivamente sobre o efeito de cada componente da PCA na eficiência do SMF na Seção 4.4.

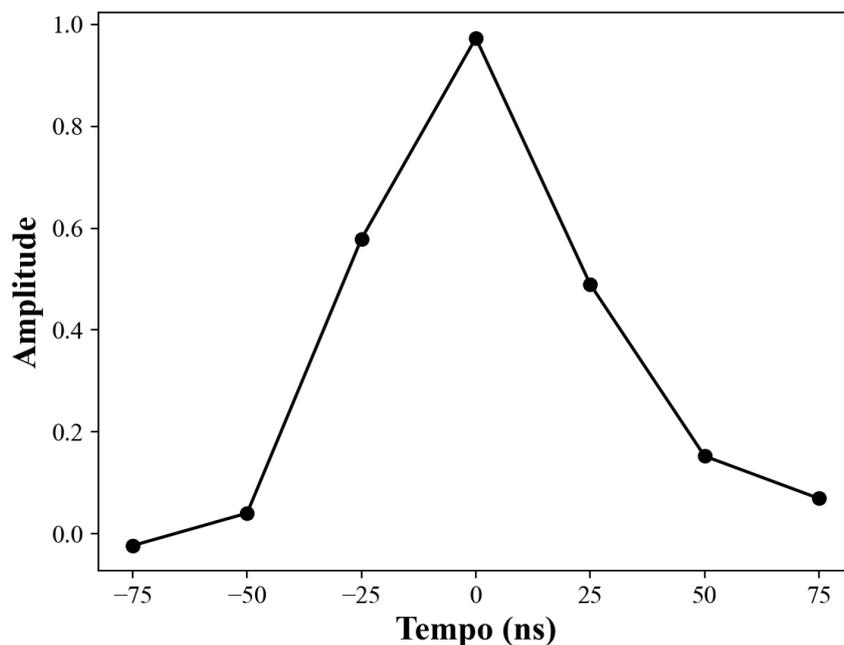
## 4.2 Conjunto de Dados

Conforme mencionado anteriormente, os conjuntos de dados foram elaborados com foco em dois ambientes, o ambiente simulado e o ambiente híbrido. Ambos serão detalhados a seguir. Vale ressaltar que os resultados serão apresentados baseando-se na estrutura do LHC onde sete amostras digitais são utilizadas para cobrir todo o pulso, simulando uma taxa de amostragem de 40 MHz (DELMASTRO, 2003), (PAGANINI; VULPEN, 2004). Além disso, os resultados estarão na unidade de contagens de ADC, que no TileCal, podem ser convertidos para unidade de energia MeV, onde 1 contagem de ADC corresponde a 12 MeV.

Ambos ambientes fazem uso do pulso tipicamente utilizado no TileCal, considerando desvios de fase e deformações que serão adicionados. O desvio de fase foi modelado por uma distribuição gaussiana de média zero e desvio padrão de aproximadamente 12,5 ns. Já a deformação, que pode representar por exemplo o envelhecimento dos equipamentos eletrônicos capazes de deformarem o pulso, foi modelada por uma distribuição gaussiana de média zero e desvio padrão de 2%.

A Figura 16 mostra um pulso extraído aleatoriamente da base de dados para um intervalo de 150 ns com as 7 amostras digitais destacadas. Esta figura tem seu tempo entre [-75, 75] ns devido ao fato da colisão ocorrer no tempo 0 mas a resposta do TileCal levar 75 ns para chegar ao pico.

Figura 16 - Exemplo de pulso utilizado.



Fonte: O autor, 2021.

#### 4.2.1 Ambiente Simulado

No TileCal, o ruído eletrônico pode ser descrito por uma função gaussiana de média zero e desvio padrão 1,5 contagens de ADC. Já o ruído de empilhamento, no entanto, depende da posição física do canal de leitura. Quanto mais próximo ao feixe de colisão, maior será a probabilidade de um canal de leitura sofrer o efeito de empilhamento de sinais devido a alta ocupação do mesmo. Para simular o ruído de empilhamento, uma distribuição exponencial foi utilizada (CHAPMAN, 2011) devido ao fato de que os sinais do TileCal possuem assinatura exponencial (KHANDAI et al., 2013).

A fim de testar a eficiência em condições adversas de ruído, foram consideradas condições de ocupação do canal de 0% a 100% onde estas foram discretamente analisadas a cada 10%. Tais condições indicam que quanto maior a ocupação do canal em questão, maiores as chances de ocorrerem empilhamentos de sinais. Os empilhamentos de sinais foram gerados considerando uma distribuição exponencial de média igual a 100 contagens de ADC.

Por fim, três bases de dados foram geradas considerando a estrutura supracitada dos ruídos, onde estes foram adicionados em cada base para as onze ocupações mencionadas anteriormente. A Figura 17 mostra tais ruídos para algumas condições de ocupação do canal de leitura.

A fim de saber o quão relacionados os ruídos estão, a Figura 18 mostra a matriz de correlação dos ruídos para ocupações de 0%, 50% e 100%. Percebe-se através do mapa de cores que devido ao efeito do empilhamento tem-se as diagonais principais e suas adjacentes com valores absolutos maiores, os quais decrescem a medida que se afastam da diagonal principal. Também pode-se destacar que a correlação entre as amostras aumenta em condições de empilhamento de sinais.

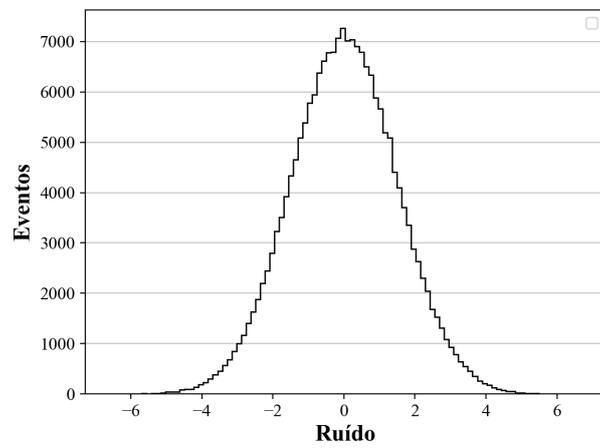
Adicionalmente, cada base de dados foi criada para efetuar o estudo da relação sinal-ruído ( $SNR$ ), que é dada por

$$SNR = \frac{\mu_{sinal}}{\mu_{emp}}, \quad (52)$$

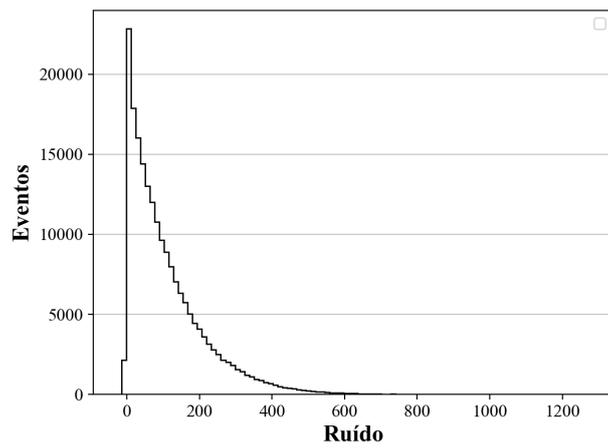
em que  $\mu_{sinal}$  e  $\mu_{emp}$  correspondem respectivamente às médias das distribuições exponenciais utilizadas para as amplitudes do sinal e ruído de empilhamento. Neste caso, como mencionado  $\mu_{emp}$  será composto por uma distribuição exponencial de média 100 ADCs devido ao empilhamento de sinais, ou seja,  $\mu_{emp} = 100 \text{ ADCs}$ . Já  $\mu_{sinal}$  será abordado a seguir.

A primeira base de dados tem o objetivo de avaliar a eficiência dos métodos em em condições de alta  $SNR$ , e para isso a amplitude utilizada para  $\mu_{sinal}$  foi gerada através de uma distribuição exponencial de média 300 ADCs, ou seja,  $SNR = 3$ . A segunda visa

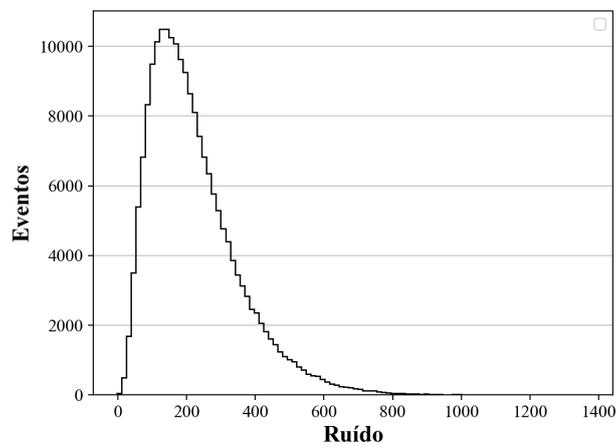
Figura 17 - Ruídos simulados.



(a)



(b)

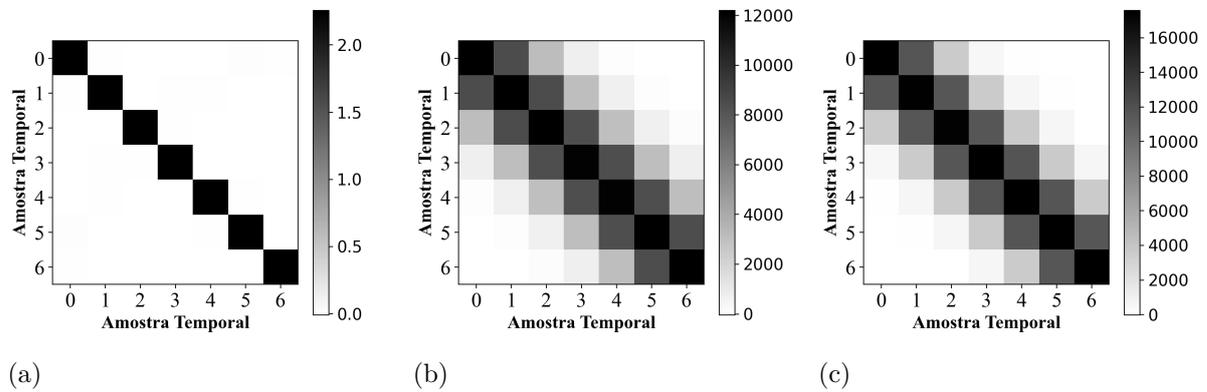


(c)

Legenda: Ruídos (em ADC) com diferentes ocupações adicionados em todas as bases de dados simuladas. (a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

Figura 18 - Covariância dos ruídos simulados.



Legenda: Análise da correlação dos ruídos para três ocupações. (a) 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

avaliar a eficiência dos métodos para  $SNR = 1$ , para isso, a amplitude utilizada foi gerada através de uma distribuição exponencial de média 100 ADCs. Por fim, a terceira base de dados tem o objetivo de avaliar a eficiência dos métodos em uma baixa relação  $SNR$ , neste caso a amplitude utilizada foi gerada através de uma distribuição exponencial de média 10 ADCs tendo então  $SNR = 0,1$ .

Como as bases de dados são fundamentadas em diferentes tipos de distribuições aleatórias, sabe-se que qualquer aproximação baseada em uma única e aleatória amostra poderia ser facilmente contestada, por exemplo, se algum método for melhor que o outro, poderia haver a dúvida se o resultado poderia ser diferente em outro conjunto aleatório.

Para mitigar esta possibilidade e fornecer resultados concisos, optou-se por utilizar uma base de dados maior de 200 mil eventos para cada  $SNR$ , e suas análises foram computadas 10 vezes antes de se obter o *errormédio* e seu respectivo RMS (Root Mean Square). Sendo assim, para cada ocupação, 10 experimentos foram considerados para construir a barra de erro dos valores utilizados para avaliar a eficiência de cada método.

Portanto, para cada experimento, uma base de dados de 200 mil sinais foi gerada sob as mesmas condições de simulação. Estas bases foram divididas em dois conjuntos de 100 mil sinais cada, sendo um conjunto, chamado de conjunto de desenvolvimento, destinado para a estimação dos parâmetros do ruído e extração das componentes principais para o projeto dos estimadores do Filtro Casado Estocástico, e o segundo conjunto, chamado de teste, foi utilizado para verificar a análise de eficiência das amplitudes estimadas de todos os métodos.

### 4.2.2 Ambiente Híbrido

Além da análise simulada, também foi realizada a análise em um ambiente híbrido formado pela combinação de dados reais e simulados. De forma similar à abordada na seção anterior, nesta também teve o objetivo de avaliar a eficiência dos métodos numéricos em diferentes bases de dados.

Os ruídos coletados provenientes de colisões são armazenados junto de seus respectivos pedestais e são então disponibilizados para pesquisadores membros. Para iniciar toda análise de ruídos desta seção, foi necessária a aquisição de ambos e, após, foi realizada uma operação de remoção do pedestal de todos os ruídos, a fim de os deixarem preparados e também otimizar o tempo de execução dos métodos numéricos durante suas aproximações.

Tais ruídos reais foram adquiridos provenientes de colisões próton-próton considerando a última tomada de dados do LHC em 2018. Através destes dados foi possível extrair três conjuntos de dados com número médio de interações por colisão de  $\langle\mu\rangle \approx 30$ ,  $\langle\mu\rangle \approx 50$ , e um conjunto com  $\langle\mu\rangle \approx 90$ , que representa faixas de alta luminosidade previstas para os próximos anos de operação do LHC.

Como o objetivo é a análise da eficiência dos métodos numéricos sob diferentes condições de empilhamento de sinais, foi aplicado um pré-processamento em cada um desses conjuntos a fim de selecionar as células das partições cujas as ocupações são elevadas durante a colisão. Em outras palavras, para a análise de empilhamento foram escolhidas células representativas próximas ao centro de colisão. Por esses motivos as células selecionadas foram as células *A9* referente à partição *LBA* e *E4*, *A13* referentes à partição *EBA*, conforme mostra a Figura 19.

A quantidade de eventos em cada célula varia em função de suas respectivas médias  $\langle\mu\rangle$ , no entanto, como pode-se ver na Tabela 1 a quantidade para cada  $\langle\mu\rangle$  é entre cada célula, varia menos de 5%.

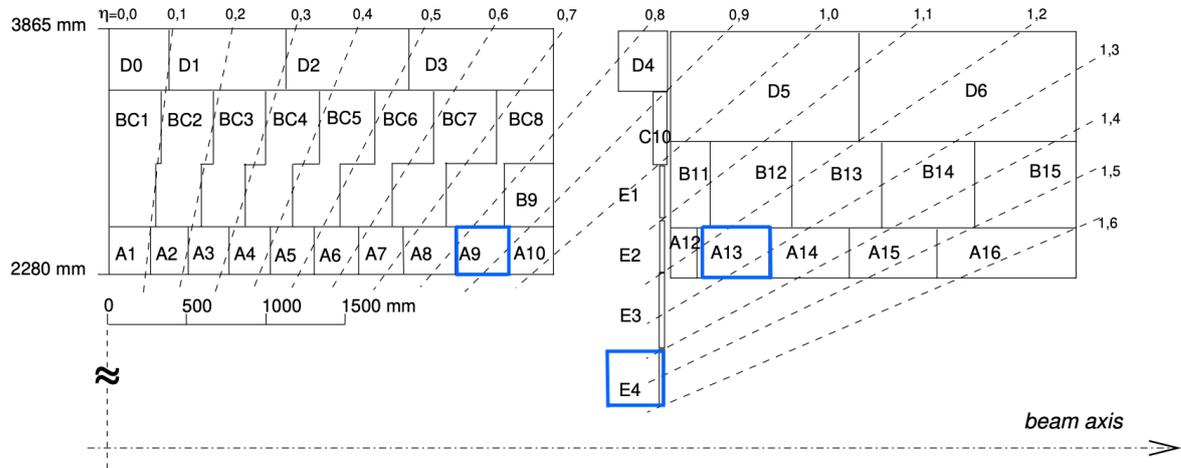
Tabela 1 - Quantidade de eventos para cada célula para suas respectivas  $\langle\mu\rangle$ .

Célula	$\langle\mu\rangle \approx 30$	$\langle\mu\rangle \approx 50$	$\langle\mu\rangle \approx 90$
<i>A9</i>	656510	583217	1051990
<i>A13</i>	666982	592504	1050988
<i>E4</i>	643953	560025	1025844

Fonte: O autor, 2021.

Como foi enunciado na Seção 1.3, o LHC gera colisões de até 13 TeV, e além disso o espaçamento entre colisões é de 25 ns. Os eventos responsáveis por gerar estes dados foram obtidos através de tomadas de dados chamadas *ZeroBias*, onde um tipo especial de *trigger* é utilizado a fim de armazenar apenas dados de *background*, ou seja, apenas

Figura 19 - Células do TileCal selecionadas.



Legenda: Em azul as células  $A9$  da partição LBA e  $E4$ , e  $A13$  da partição EBA selecionadas devido suas posições físicas no barril do TileCal.

Fonte: Adaptado de CERN, 2013.

informações de ruído eletrônico e de empilhamento são coletados. Com este *trigger*, é esperado completa ausência de sinais de interesse, o que fazem desses ruídos candidatos à serem usados nesta abordagem híbrida.

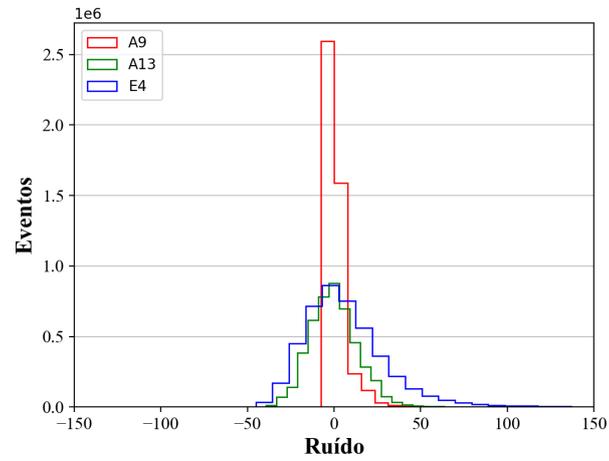
De acordo disposição física de cada célula apresentada pela Figura 19, espera-se que todas possam apresentar alta ocupação, principalmente para  $\langle \mu \rangle \approx 50$  e  $\langle \mu \rangle \approx 90$ . Pela direção das partículas após colisão e a proximidade de cada célula tem-se que  $E4$  tende ser a célula com maior probabilidade de empilhamento severo de sinais ao passo que  $A9$  tende ser a célula com menor ocupação dentre as selecionadas, principalmente para  $\langle \mu \rangle \approx 30$ .

Para a escolha das melhores células e validar as informações do parágrafo anterior, é importante fazer uma análise de como estas estão distribuídas em função de cada erro médio  $\langle \mu \rangle$ . Por isso a Figura 20 mostra as distribuições das amostras temporais de ruído para as diferentes células e diferentes condições de  $\langle \mu \rangle$ .

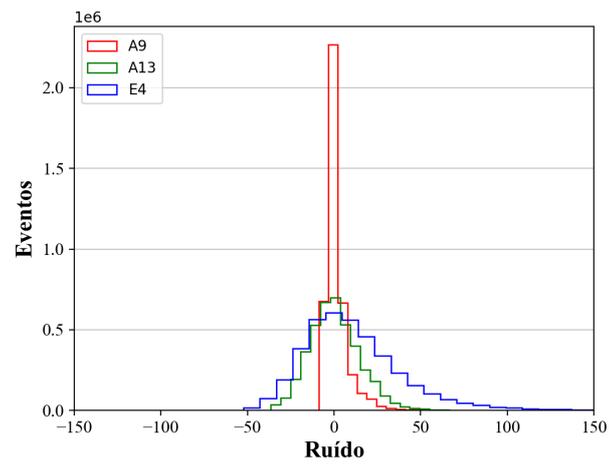
Percebe-se na Figura 20 que, no geral, a célula  $A9$ , a mais distante do ponto de colisão possui dispersão menor que as demais. Isso acontece devido ao seu posicionamento, pois, com menor ocorrência de empilhamento de sinais, o erro que predomina é proveniente do ruído eletrônico que é por sua vez na grande maioria dos casos gaussiano de média zero. Já (c) para  $\langle \mu \rangle \approx 90$ , é o cenário com maior empilhamento de sinais, logo, é esperado que uma célula próxima ao ponto de colisão como a célula  $E4$  fosse possuir o ruído mais severo.

Devido a estes fatos,  $E4$  e  $A9$  serão as células abordadas em detalhes na seção de resultados. Desta forma será possível abordar as aproximações numéricas dos métodos

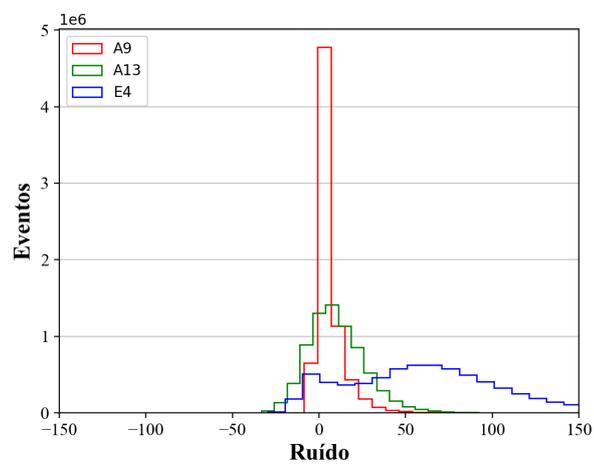
Figura 20 - Comparação entre células para diferentes ruídos reais.



(a)



(b)



(c)

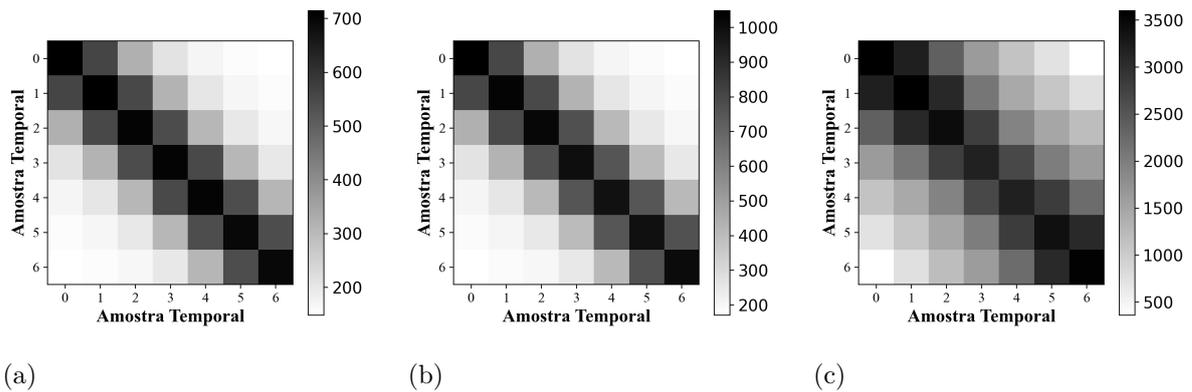
Legenda: Ruídos (em ADC) para (a)  $\langle \mu \rangle \approx 30$ , (b)  $\langle \mu \rangle \approx 50$  e (c)  $\langle \mu \rangle \approx 90$ .

Fonte: O autor, 2021.

em diferentes condições de empilhamentos, isto é, realizar a comparação para casos de empilhamentos severos e empilhamentos normais.

Sabe-se que empilhamento de sinais resulta em uma matriz de covariância correlacionada, por isso, para ter certeza que  $E4$  é uma boa escolha em termos de ruídos de empilhamento severos, gerou-se sua matriz de covariância como mostra a Figura 21 para os ruídos médios de  $\langle \mu \rangle \approx 30$ ,  $\langle \mu \rangle \approx 50$ , e  $\langle \mu \rangle \approx 90$ .

Figura 21 - Covariância dos ruídos reais.



Legenda: Análise da correlação dos ruídos de  $E4$  para três ruídos médios. (a)  $\langle \mu \rangle \approx 30$ , (b)  $\langle \mu \rangle \approx 50$  e (c)  $\langle \mu \rangle \approx 90$ .

Fonte: O autor, 2021.

De acordo com a Figura 21, todos eles possuem correlação, (a)  $\langle \mu \rangle \approx 30$  possui a menor entre eles, (b)  $\langle \mu \rangle \approx 50$  possui um pouco mais e por fim (c)  $\langle \mu \rangle \approx 90$  é o mais correlacionado de todos, como esperado devido a natureza de seus sinais.

O processo de criação das bases de dados mencionadas no início desta seção foram baseadas nas células  $E4$  e  $A9$  considerando respectivamente os ruídos médios  $\langle \mu \rangle \approx 90$  e  $\langle \mu \rangle \approx 30$ . Em seguida, foi gerado uma simulação do sinal de interesse com amplitude conhecida, e este sinal foi somado ao ruído. Desta forma, o sinal resultante possui o ruído real e um sinal o qual a amplitude deve ser estimada. As amplitudes conhecidas foram extraídas aleatoriamente de distribuições exponenciais com as seguintes médias:  $10$ ,  $100$  e  $300$ . Esta abordagem permitirá avaliar a eficiência de cada método em diferentes células,  $\langle \mu \rangle$  e  $SNRs$ .

### 4.3 Análise dos Resultados

Esta seção irá apresentar e discutir os resultados para os ambientes Simulados e Híbridos. Todos os resultados numéricos apresentados serão referentes aos métodos OF (Seção 2.4), DMF (Seção 3.2) e SMF (Seção 3.3). Apesar de apresentar as aproximações

das amplitudes e outras análises referentes às aproximações, a principal figura de mérito para analisar o quão bom cada método é, será a média e RMS do erro de estimação, onde o erro e o RMS correspondem a:

$$\text{erro} = \text{amp\_aproximada} - \text{amp\_verdadeira}, \quad (53)$$

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}. \quad (54)$$

#### 4.3.1 Ambiente Simulado

Apesar de tanto a base de dados quanto os algoritmos estarem computados para todas *SNRs* e ocupações, seria inviável exibir todos os gráficos para as três *SNRs* e onze ocupações. Por esse motivo, os gráficos estarão restritos às três *SNRs* mencionadas e para as ocupações de 0%, 50% e 100%. Os resultados referentes às demais ocupações serão apresentados na Figura 28 e também em forma de tabelas para melhor visualização.

Inicialmente, é importante avaliar as características e formatos das amplitudes aproximadas (ou energia) pelos métodos numéricos, por isso, as Figuras 22, 23 e 24 mostram as amplitudes aproximadas sendo comparadas à amplitude verdadeira.

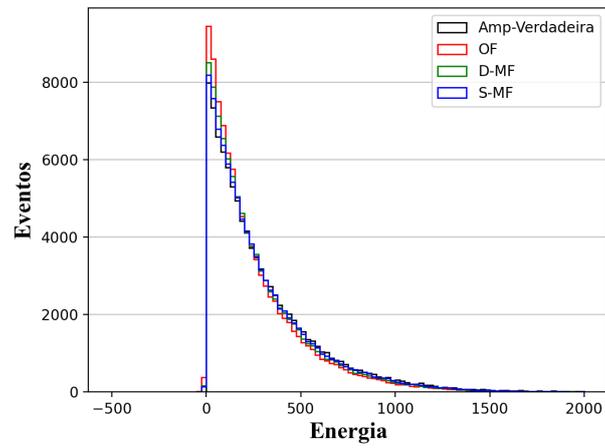
Em todos os casos das Figuras 22, 23 e 24, percebe-se que os métodos numéricos apresentam inicialmente uma boa aproximação, mas o OF apesar de ter seu centro mais alinhado com o centro da amplitude verdadeira, fornece amplitudes negativas, o que representaria uma energia negativa, sendo então, fisicamente impossível. Isto ocorre com maior frequência nos cenários de empilhamento de sinais, ou seja, para ocupações de 50% e 100%. A área negativa fica ainda mais evidente a medida que a relação sinal-ruído diminui. A Tabela 2 mostra a comparação:

Uma informação ainda mais interessante em termos de aproximações é saber o quão boas estas são, e uma boa alternativa para isso é efetuar a análise do erro. Estas análises do erro baseadas na Equação 53 são exibidas pela Figuras 25, 26 e 27 para todas as *SNRs*.

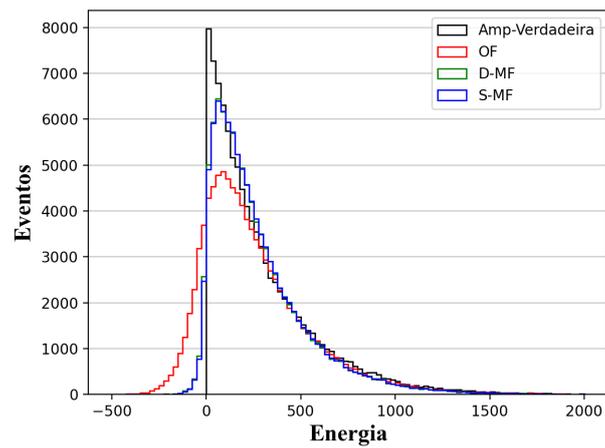
Pelos erros demonstrados, percebe-se novamente a cauda negativa do OF. Caudas negativas podem indicar características não-gaussianas do erro de estimação, o que faz sentido nesta simulação, visto que sinais de empilhamentos (que por sua vez introduzem erros não gaussianos) foram adicionados.

Ademais, apesar de também ocorrer para as demais *SNRs* das Figuras 25, 26, fica mais evidente que para uma relação sinal-ruído baixa como  $SNR = 0,1$  mostrada na

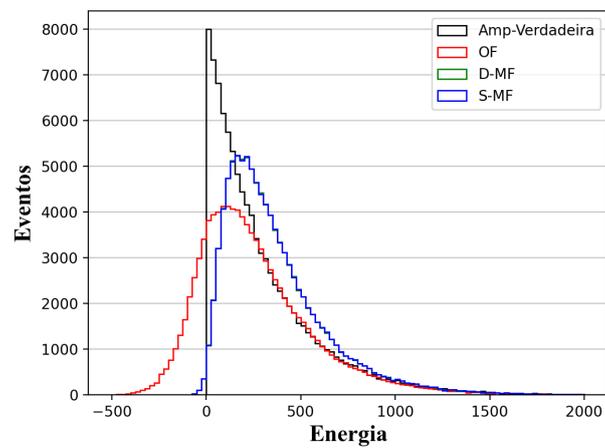
Figura 22 - Comparação entre amplitudes simuladas considerando  $SNR = 3$ .



(a)



(b)



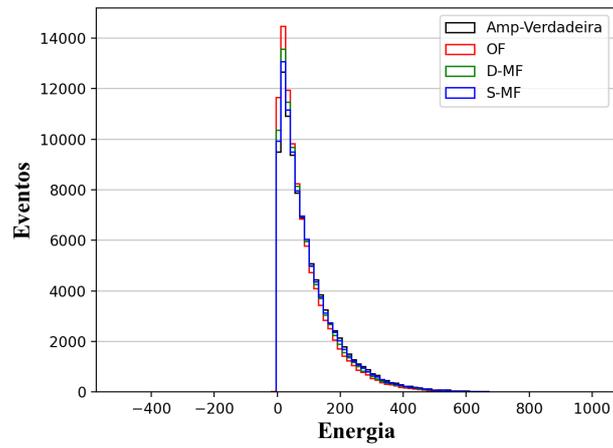
(c)

Legenda: Comparação entre amplitudes geradas (em ADC) com  $SNR = 3$  para diferentes ocupações.

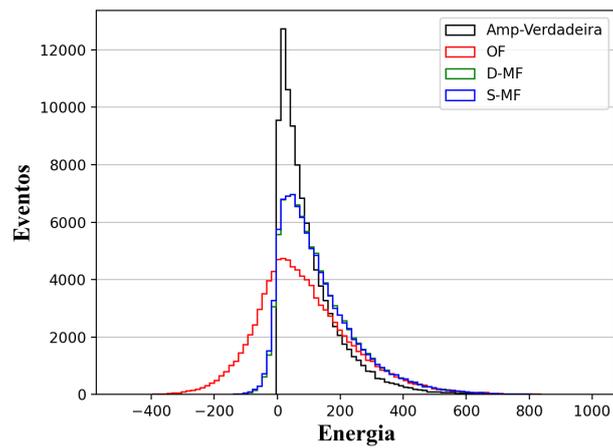
(a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

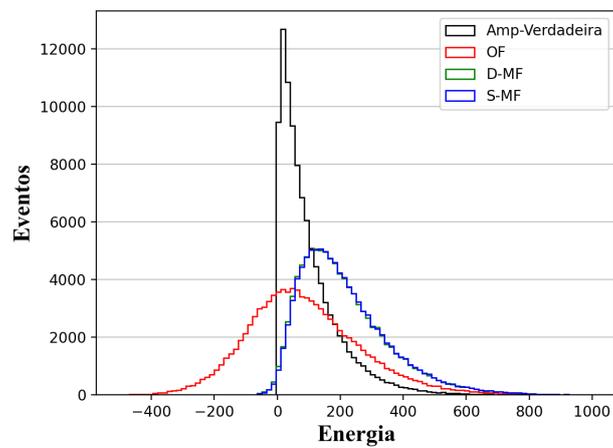
Figura 23 - Comparação entre amplitudes simuladas considerando  $SNR = 1$ .



(a)



(b)



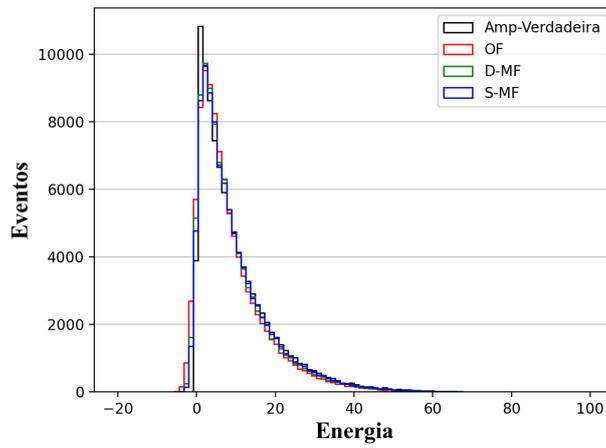
(c)

Legenda: Comparação entre amplitudes geradas (em ADC) com  $SNR = 1$  para diferentes ocupações.

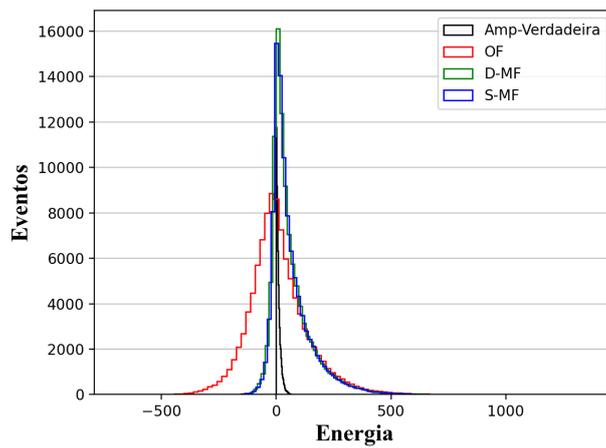
(a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

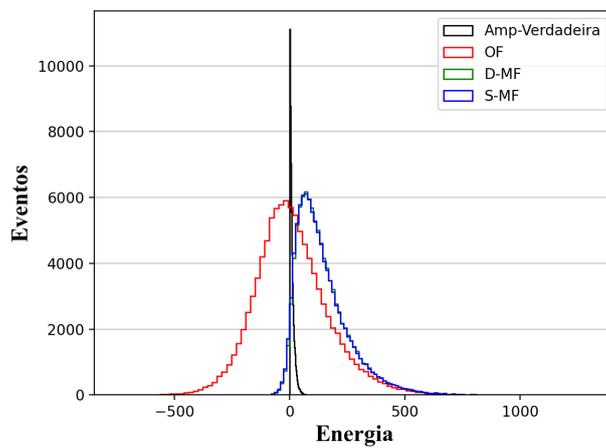
Figura 24 - Comparação entre amplitudes simuladas considerando  $SNR = 0,1$ .



(a)



(b)



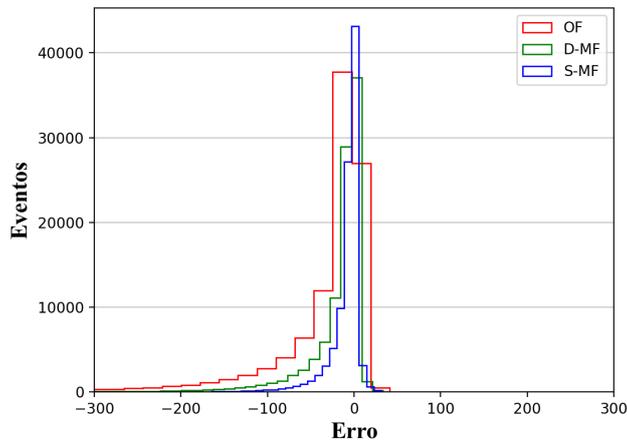
(c)

Legenda: Comparação entre amplitudes geradas (em ADC) com  $SNR = 0,1$  para diferentes ocupações.

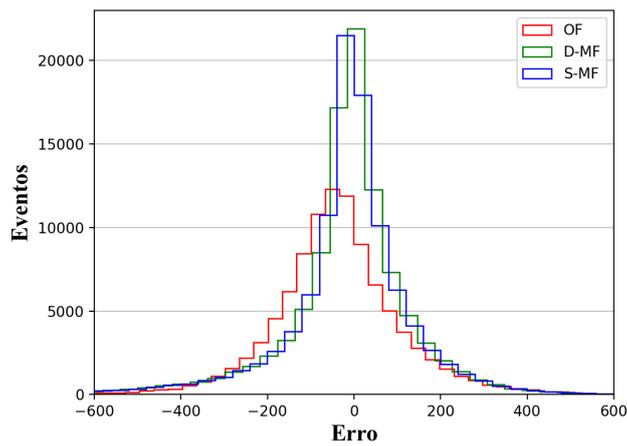
(a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

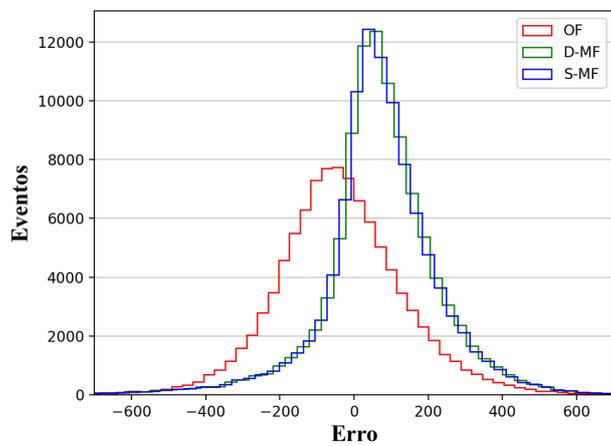
Figura 25 - Comparação entre erros considerando  $SNR = 3$ .



(a)



(b)

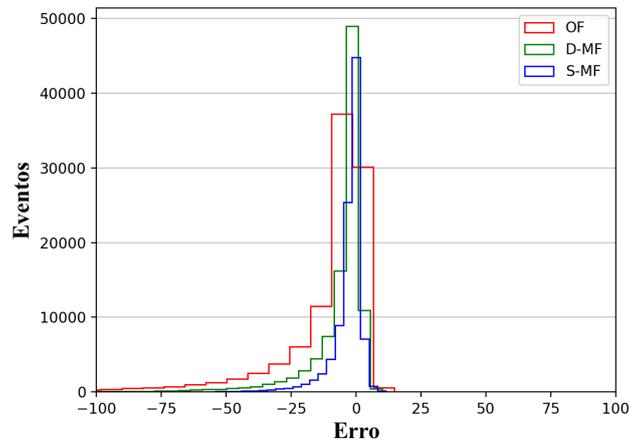


(c)

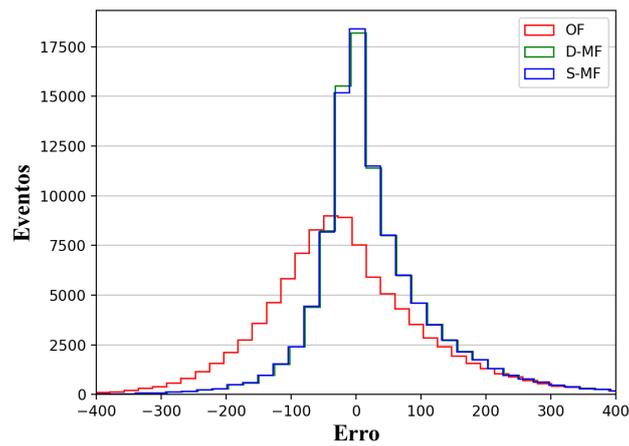
Legenda: Comparação entre erros geradas (em ADC) com  $SNR = 3$  para diferentes ocupações. (a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

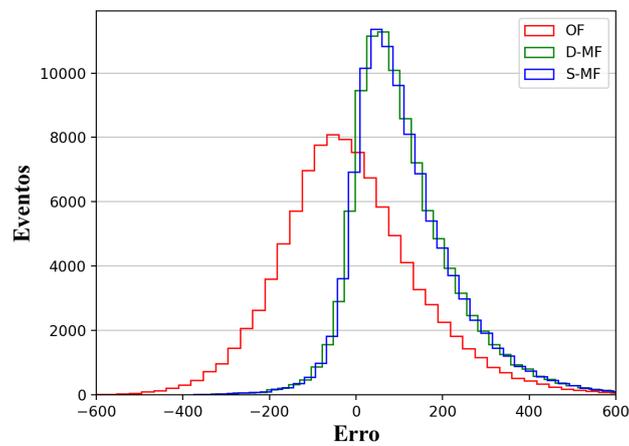
Figura 26 - Comparação entre erros considerando  $SNR = 1$ .



(a)



(b)

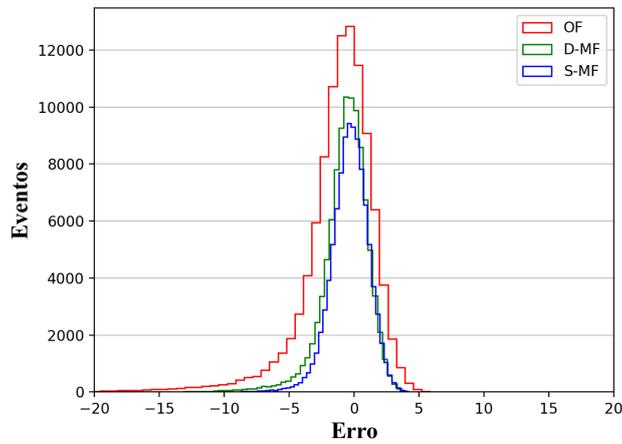


(c)

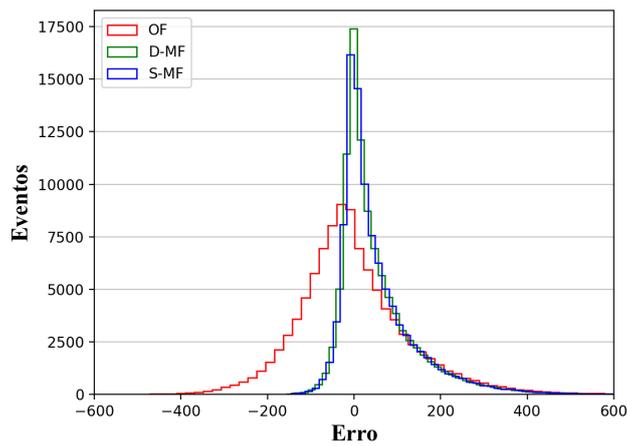
Legenda: Comparação entre erros geradas (em ADC) com  $SNR = 1$  para diferentes ocupações. (a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

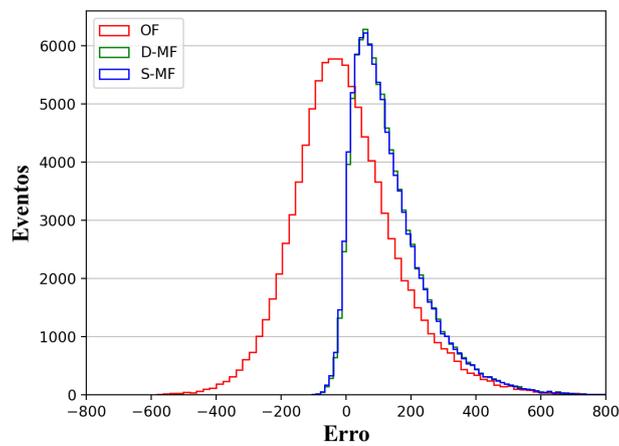
Figura 27 - Comparação entre erros considerando  $SNR = 0,1$ .



(a)



(b)



(c)

Legenda: Comparação entre erros geradas (em ADC) com  $SNR = 0,1$  para diferentes ocupações. (a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

Tabela 2 - Porcentagem negativa da amplitude estimada do OF para diferentes ocupações e  $SNRs$ .

Ocupação (%)	OF: Área Negativa (%)
	SNR = 3
0	0,422
50	14,980
100	17,841
	SNR = 1
0	0,968
50	30,029
100	33,428
	SNR = 0,1
0	7,192
50	52,475
100	51,349

Fonte: O autor, 2021.

Figura 27, o empilhamento ocupa a maior parte do erro para (b) e (c) quando comparado com o erro sem empilhamento (a).

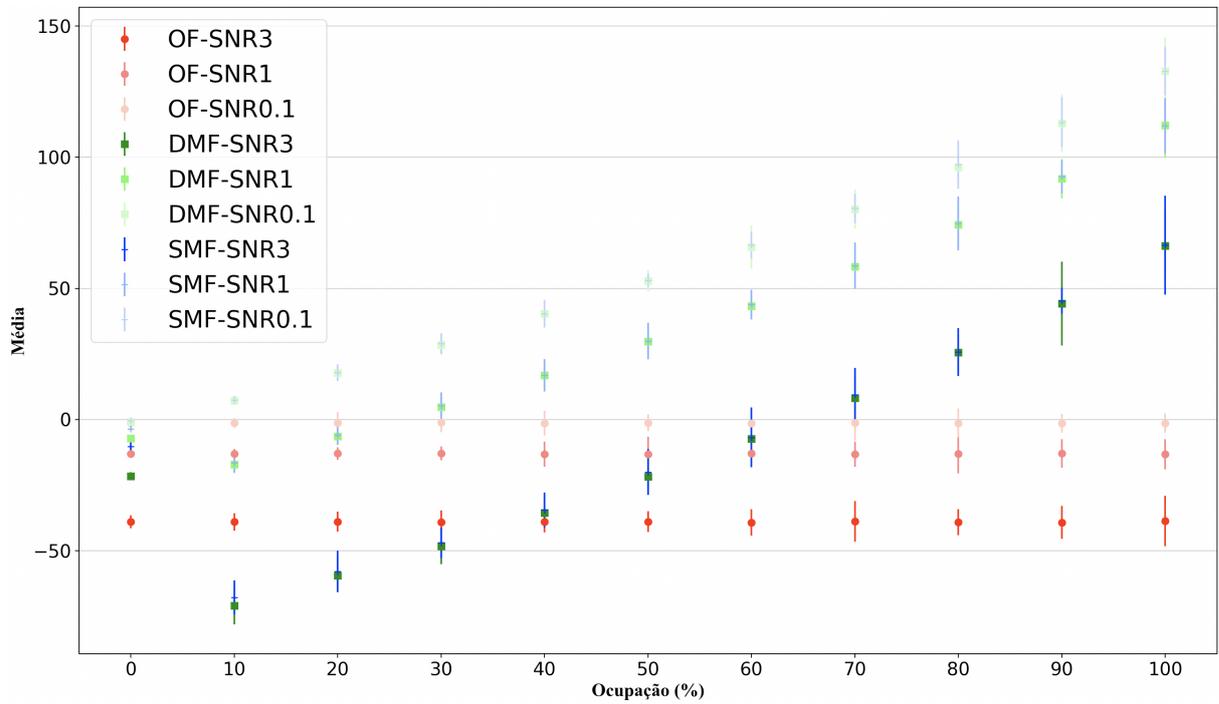
A fim de se entender a tendência associada de cada estimador, e o quão preciso são as estimações destes, optou-se por usar como referência a média e o RMS respectivamente.

A Figura 28 mostra comparações entre as médias e RMSs de todas as  $SNRs$  para os métodos numéricos em função de todas as ocupações. Este é um gráfico que irá fornecer de forma concisa, um panorama geral de todas as simulações. Para tal, todas as informações apresentadas por este é o resultado da média aproximada após 10 execuções para cada cenário mencionado. Os RMSs não seriam visíveis devido seus pequenos valores, por esse motivo, foram multiplicados por um fator de 10. Por fim, neste gráfico, a intensidade das cores dos métodos foram reduzidas proporcionalmente à redução das  $SNRs$ , ou seja, quanto menor a  $SNRs$ , menor a intensidade da cor.

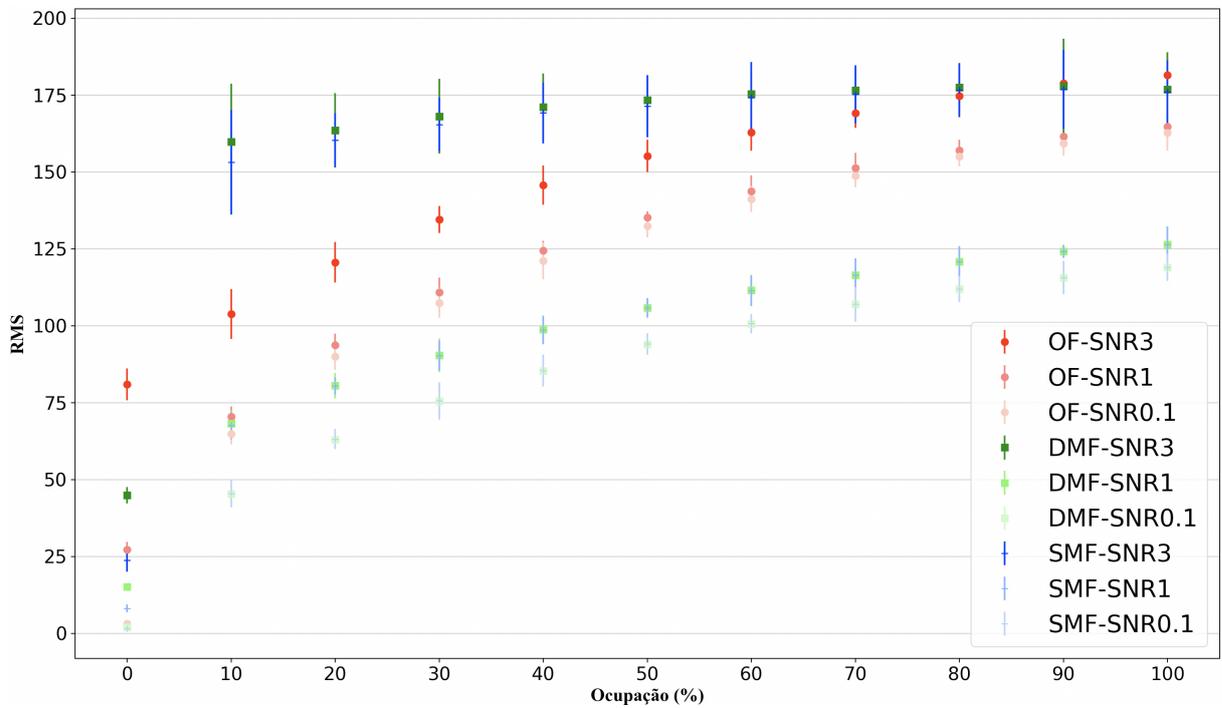
Pela Figura 28 (a), pode-se observar que em todos os casos a média do OF se mantém quase constante independentemente da ocupação do canal, ao passo que DMF e SMF possuem perfil linear crescente em função das ocupações. Em termos da relação sinal-ruído, percebe-se que o OF teve a média do erro absoluto se aproximando de 0 a medida que a relação sinal-ruído diminuiu. Já DMF e SMF mantiverem seu perfil linear crescente a medida que a relação sinal-ruído diminuiu.

O RMS (b) tem um cenário diferente, a dispersão do OF aumenta consideravelmente em função das ocupações, e novamente tem-se melhores aproximações para menores  $SNRs$ . Já o DMF e SMF a partir da ocupação de 10% e para  $SNR = 3$  mantiveram

Figura 28 - Comparação entre as médias e RMSs entre todas  $SNRs$  para métodos OF, DMF e SMF em função das ocupações.



(a)



(b)

Legenda: (a) média, (b) RMS.

Fonte: O autor, 2021.

uma taxa quase constante, no entanto, com RMSs mais altos que o OF. Para  $SNR = 1$  e  $SNR = 0,1$  os métodos DMF e SMF foram muito similares e ambos tiveram valores médios do RMS menores que o OF. Por fim, com exceção da  $SNR = 3$  para o DMF e SMF, os métodos demonstraram seguir um padrão exponencial a medida que a ocupação aumenta.

No geral, mesmo tendo suas médias proporcionalmente afetadas pelo empilhamento, os métodos DMF e SMF se mostraram mais eficazes em termos de dispersão se comparados ao OF para as  $SNR = 1$  e  $SNR = 0,1$ .

As informações numéricas referentes às Figuras 28 (a) e (b) foram dispostas em três tabelas para permitir análise numérica geral de todos os métodos em diferentes condições. Sendo assim, as Tabelas 3, 4 e 5 mostram respectivamente  $SNR = 3$ ,  $SNR = 1$ ,  $SNR = 0,1$  para as aproximações de todos os métodos em função de todas ocupações.

Numericamente, as Tabelas 3, 4 e 5 mostram que a média todas as médias do OF foram negativas, o que seria equivalente a dizer que a energia média da colisão foi negativa. Isso também aconteceu em alguns casos para o DMF e SMF, no entanto, quando aconteceu foi em valores absolutos menores.

Refletindo a mudança abrupta ocorrida para a ocupação de 10% na Figura 28 (a) para DMF e SMF, agora é possível ver que isso representou uma variação na média em torno de 50 ADCs entre 0% e 10% para  $SNR = 3$ .

No geral, DMF e SMF foram muito próximos, no entanto, uma das informações mais importantes que pode ser extraída da Figura 28 e conseqüentemente das tabelas é que o SMF se mostrou mais eficaz que o DMF em cenários sem empilhamento de sinais. Este cenário poderá ser validado na Seção 4.3.2 ao testar ambos os métodos na célula A9 com  $\langle \mu \rangle \approx 30$ .

Como o DMF e SMF teve uma média linear e crescente, é possível aplicar uma correção linear para a média do erro a fim de minimizar este erro médio, como por exemplo a técnica conhecida como Método dos Mínimos Quadrados (MMQ), o que poderia melhorar suas aproximações.

Outro fator importante a se destacar é no impacto que o empilhamento de sinais causa nas aproximações. Em qualquer método, assim que houve a mudança de 0% para 10% ou mais, o RMS passou a ter um valor consideravelmente maior.

### 4.3.2 Ambiente Híbrido

Como apresentado na seção 4.2.2, ao todo são três métodos (OF, DMF, SMF) para serem avaliados em bases de dados compostas por três tipos de ruídos ( $\langle \mu \rangle \approx 30$ ,  $\langle \mu \rangle \approx 50$ ,  $\langle \mu \rangle \approx 90$ ) que poderiam ser testados em três células (A9, A13 e E4) para três amplitudes esperadas (10, 100, 300). Apresentar todas as combinações possíveis destas

Tabela 3 - médias e RMSs de cada método para suas respectivas ocupações para  $SNR = 3$ .

Ocupação (%)	Média	Desvio Padrão
OF		
0	$-39.0031 \pm 0.2492$	$80.9396 \pm 0.5159$
10	$-38.9821 \pm 0.3347$	$103.8419 \pm 0.8089$
20	$-38.8873 \pm 0.3841$	$120.6456 \pm 0.6528$
30	$-39.1068 \pm 0.4537$	$134.5698 \pm 0.4380$
40	$-38.8926 \pm 0.4036$	$145.7701 \pm 0.6375$
50	$-38.8911 \pm 0.3943$	$155.2296 \pm 0.5197$
60	$-39.2145 \pm 0.5018$	$162.9366 \pm 0.6032$
70	$-38.8464 \pm 0.7743$	$169.1868 \pm 0.4796$
80	$-39.1176 \pm 0.4935$	$174.6899 \pm 0.6892$
90	$-39.1962 \pm 0.6218$	$178.8425 \pm 0.8658$
100	$-38.6779 \pm 0.9641$	$181.4716 \pm 0.6788$
DMF		
0	$-21.5544 \pm 0.1542$	$44.8970 \pm 0.2674$
10	$-70.9292 \pm 0.7126$	$159.8085 \pm 1.8852$
20	$-59.4513 \pm 0.6318$	$163.5767 \pm 1.2133$
30	$-48.2361 \pm 0.6894$	$168.1236 \pm 1.2152$
40	$-35.5963 \pm 0.5009$	$171.0665 \pm 1.0963$
50	$-21.7810 \pm 0.3911$	$173.4099 \pm 0.8039$
60	$-7.3312 \pm 0.6047$	$175.4323 \pm 0.7202$
70	$8.2457 \pm 0.8776$	$176.5169 \pm 0.7418$
80	$25.5314 \pm 0.6909$	$177.5636 \pm 0.7932$
90	$44.1710 \pm 1.5963$	$177.9140 \pm 1.5474$
100	$66.2620 \pm 1.2229$	$176.7823 \pm 1.2157$
SMF		
0	$-10.2430 \pm 0.1607$	$23.7227 \pm 0.3720$
10	$-67.7846 \pm 0.6529$	$153.1824 \pm 1.6998$
20	$-57.8876 \pm 0.7949$	$160.3016 \pm 0.8774$
30	$-46.8896 \pm 0.5833$	$165.4570 \pm 0.8740$
40	$-34.5079 \pm 0.6665$	$169.2507 \pm 0.9915$
50	$-19.9597 \pm 0.8736$	$171.3943 \pm 1.0145$
60	$-6.7671 \pm 1.1362$	$174.1409 \pm 1.1617$
70	$9.4354 \pm 1.0205$	$175.2527 \pm 0.9382$
80	$25.7633 \pm 0.9157$	$176.5601 \pm 0.8622$
90	$45.2973 \pm 0.5002$	$176.8846 \pm 1.2859$
100	$66.5274 \pm 1.8856$	$176.0231 \pm 1.0263$

Fonte: O autor, 2021.

Tabela 4 - médias e RMSs de cada método para suas respectivas ocupações para  $SNR = 1$ .

Ocupação (%)	Média	Desvio Padrão
OF		
0	$-13.0255 \pm 0.1001$	$27.1740 \pm 0.2545$
10	$-13.0093 \pm 0.1764$	$70.3738 \pm 0.3370$
20	$-12.9586 \pm 0.2380$	$93.7192 \pm 0.3771$
30	$-12.8465 \pm 0.2619$	$110.8511 \pm 0.4801$
40	$-13.2070 \pm 0.4754$	$124.4389 \pm 0.3167$
50	$-13.1942 \pm 0.6685$	$135.1502 \pm 0.2010$
60	$-12.8752 \pm 0.4715$	$143.7895 \pm 0.5169$
70	$-13.2120 \pm 0.4788$	$151.2489 \pm 0.4960$
80	$-13.0959 \pm 0.7425$	$157.0580 \pm 0.3468$
90	$-12.8663 \pm 0.5457$	$161.6075 \pm 0.3692$
100	$-13.2458 \pm 0.5738$	$164.7134 \pm 0.6060$
DMF		
0	$-7.1827 \pm 0.0519$	$15.0607 \pm 0.1302$
10	$-17.0353 \pm 0.1991$	$68.1725 \pm 0.4197$
20	$-6.4192 \pm 0.2589$	$80.5651 \pm 0.4104$
30	$4.8173 \pm 0.3309$	$90.3887 \pm 0.5406$
40	$16.8916 \pm 0.4146$	$98.7286 \pm 0.4178$
50	$29.8504 \pm 0.5436$	$105.8602 \pm 0.2921$
60	$43.3501 \pm 0.4926$	$111.4736 \pm 0.5162$
70	$58.3336 \pm 0.5921$	$116.4486 \pm 0.5388$
80	$74.3109 \pm 0.6629$	$120.7883 \pm 0.4932$
90	$91.7718 \pm 0.7480$	$124.1969 \pm 0.1987$
100	$112.1627 \pm 1.2324$	$126.4638 \pm 0.5633$
SMF		
0	$-3.5029 \pm 0.0521$	$8.1125 \pm 0.1260$
10	$-16.395 \pm 0.3987$	$67.5856 \pm 0.4409$
20	$-5.9692 \pm 0.3645$	$80.4063 \pm 0.2838$
30	$5.2668 \pm 0.5100$	$90.2753 \pm 0.4845$
40	$16.8595 \pm 0.6232$	$98.6764 \pm 0.4654$
50	$29.9628 \pm 0.6958$	$105.8215 \pm 0.3202$
60	$43.8272 \pm 0.5621$	$111.4474 \pm 0.4919$
70	$58.6771 \pm 0.8817$	$116.4409 \pm 0.5433$
80	$74.7703 \pm 1.0256$	$120.7933 \pm 0.5153$
90	$92.5467 \pm 0.6331$	$124.1536 \pm 0.2108$
100	$112.0152 \pm 1.0616$	$126.4869 \pm 0.5754$

Fonte: O autor, 2021.

Tabela 5 - médias e RMSs de cada método para suas respectivas ocupações para  $SNR = 0, 1$ .

Ocupação (%)	Média	Desvio Padrão
OF		
0	$-1.2981 \pm 0.0088$	$3.1476 \pm 0.0214$
10	$-1.2483 \pm 0.1822$	$64.7938 \pm 0.3410$
20	$-1.1931 \pm 0.4095$	$90.0095 \pm 0.4363$
30	$-1.1260 \pm 0.3639$	$107.4746 \pm 0.4893$
40	$-1.4006 \pm 0.4736$	$121.1357 \pm 0.5954$
50	$-1.2678 \pm 0.3311$	$132.4458 \pm 0.3708$
60	$-1.3491 \pm 0.2823$	$141.2316 \pm 0.4280$
70	$-1.2574 \pm 0.6697$	$148.8386 \pm 0.3814$
80	$-1.3631 \pm 0.5503$	$154.9778 \pm 0.3097$
90	$-1.4018 \pm 0.3590$	$159.3155 \pm 0.3978$
100	$-1.3445 \pm 0.3623$	$162.7938 \pm 0.5870$
DMF		
0	$-0.7157 \pm 0.0060$	$1.9179 \pm 0.0127$
10	$7.2482 \pm 0.1608$	$45.2419 \pm 0.4373$
20	$17.7660 \pm 0.3121$	$62.9981 \pm 0.3219$
30	$28.5473 \pm 0.3664$	$75.4511 \pm 0.6090$
40	$40.3734 \pm 0.2614$	$85.3975 \pm 0.5116$
50	$52.8528 \pm 0.4015$	$94.0068 \pm 0.3439$
60	$65.8405 \pm 0.8217$	$100.6466 \pm 0.3155$
70	$80.2655 \pm 0.7567$	$106.9627 \pm 0.5603$
80	$96.2030 \pm 0.6043$	$111.8918 \pm 0.4260$
90	$112.9310 \pm 1.1007$	$115.6015 \pm 0.5432$
100	$132.8556 \pm 1.3040$	$118.9240 \pm 0.4380$
SMF		
0	$-0.3705 \pm 0.0054$	$1.4935 \pm 0.0084$
10	$7.5038 \pm 0.1481$	$45.4506 \pm 0.4444$
20	$17.9309 \pm 0.3168$	$63.1377 \pm 0.3263$
30	$28.9271 \pm 0.3922$	$75.5622 \pm 0.6074$
40	$40.3634 \pm 0.5317$	$85.4875 \pm 0.5147$
50	$53.0292 \pm 0.2484$	$94.0880 \pm 0.3460$
60	$66.5256 \pm 0.5191$	$100.7176 \pm 0.3150$
70	$80.4957 \pm 0.5661$	$107.0317 \pm 0.5547$
80	$97.1989 \pm 0.9293$	$111.9543 \pm 0.4253$
90	$113.3146 \pm 0.9558$	$115.6587 \pm 0.5437$
100	$132.7811 \pm 0.9101$	$118.9795 \pm 0.4392$

Fonte: O autor, 2021.

bases seria inviável, por isso, como constatado pela Figura 20, optou-se por apresentar graficamente os resultados referentes aos cenários com maior empilhamento e compará-los aos com menor empilhamento, isto é, comparar as células *E4* e *A9*. A célula *E4* terá seus gráficos exibidos para o pior cenário em termos de empilhamento, ou seja,  $\langle \mu \rangle \approx 90$  ao passo que *A9* terá seus gráficos exibidos para o cenário com menor empilhamento, ou seja,  $\langle \mu \rangle \approx 30$ .

De forma similar à seção anterior, é importante avaliar as características e formatos das amplitudes aproximadas (ou energia) pelos métodos numéricos. Desta vez como não se tem controle total do ambiente, não é possível calcular precisamente as relações sinal-ruído, mas pela Equação 52 é possível inferir isto em função da amplitude esperada, ou seja, maiores amplitudes gerarão maiores *SNRs* ao passo que menores amplitudes gerarão menores *SNRs*. A Figura 29 mostra as amplitudes aproximadas para a célula *E4* e a Figura 30 para a célula *A9* considerando diferentes amplitudes esperadas.

As Figuras 29 e 30 deixam claro o impacto de baixas *SNRs* em aproximações para empilhamento. A medida que a *SNR* aumenta, as aproximações também se aproximam da amplitude esperada.

A fim de entender quão boa foram as aproximações, foram computados nas Figuras 31 e 32 os erros absolutos das aproximações utilizando a Equação 53.

Visualmente para ambas células, o erro de aproximação dos métodos independentemente da amplitude utilizada segue o formato esperado, ou seja, uma gaussiana centrada em 0. Maiores informações serão apresentadas nas tabelas a seguir.

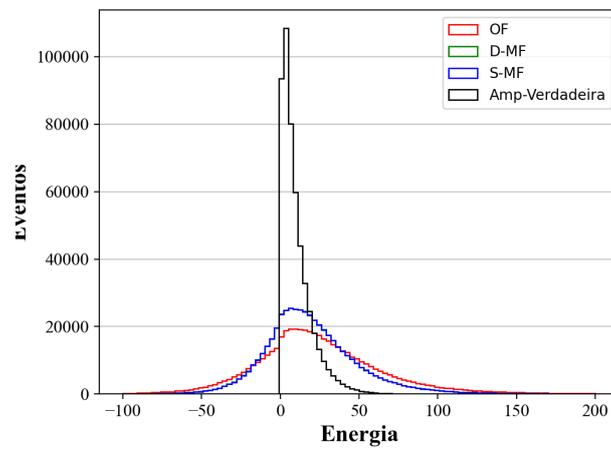
As Tabelas 6 e 7 mostram respectivamente as médias e RMSs de todos os métodos para *E4*  $\langle \mu \rangle \approx 90$  e *A9*  $\langle \mu \rangle \approx 30$ .

De acordo com a Tabela 6, o OF apresenta de fato excelente aproximações, principalmente em cenário cujas as relações sinal-ruído são maiores, ou seja, para casos onde *Amplitude* = 300 e *Amplitude* = 100. O DMF e SMF possuem resultados satisfatórios, no entanto suas melhores aproximações ocorrem no cenário de baixa relação sinal-ruído, ou seja, para casos onde *Amplitude* = 10.

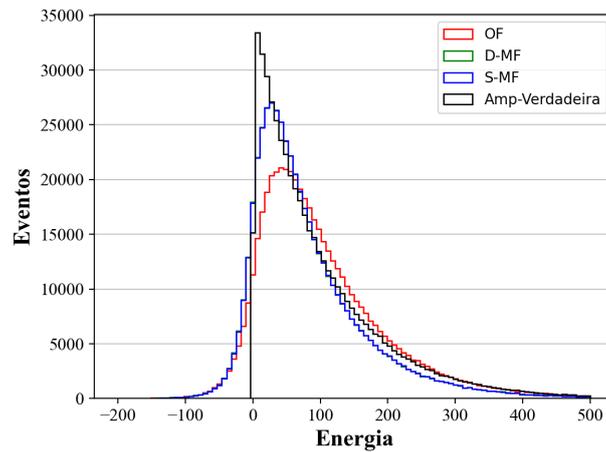
Similarmente ao ocorrido na simulação, a Tabela 7 mostra que SMF é consideravelmente melhor tanto na média quanto no RMS para cenários com baixo empilhamento de sinais como ocorre na célula *A9*  $\langle \mu \rangle \approx 30$ . Os cenários onde as médias são expressivamente melhores são para *SNR* = 1 onde o SMF chega a ser 29,7% melhor que o OF e 55,8% que o DMF e para o *SNR* = 3 onde o SMF é 60,8% melhor que o OF e 75,7% melhor que o DMF.

Em termos absolutos, parece que maiores *SNRs* estão gerando erros piores, mas deve-se levar em consideração que o erro aumentou devido a quantidade maior de energia envolvida na aproximação. Para saber de fato qual aproximação é mais precisa, o ideal é avaliar seu erro relativo. Por isso, a Figura 33 apresentará o erro relativo dos métodos para *E4*  $\langle \mu \rangle \approx 90$  e *A9*  $\langle \mu \rangle \approx 30$ .

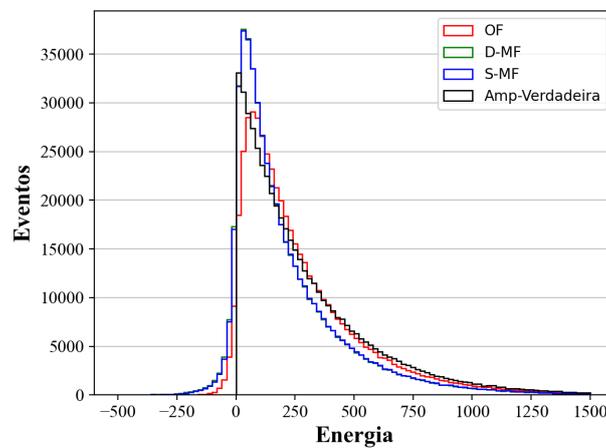
Figura 29 - Comparação entre amplitudes para a célula  $E4$ .



(a)



(b)

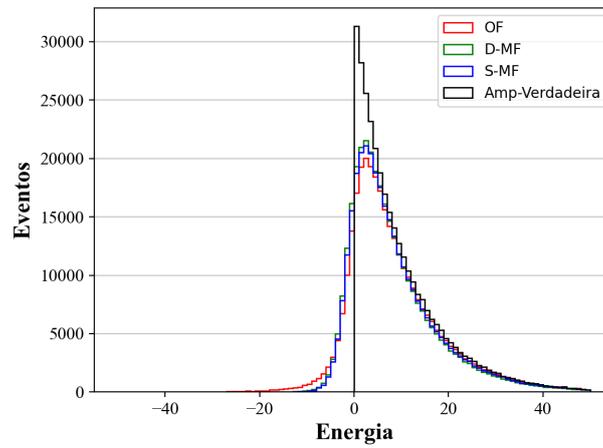


(c)

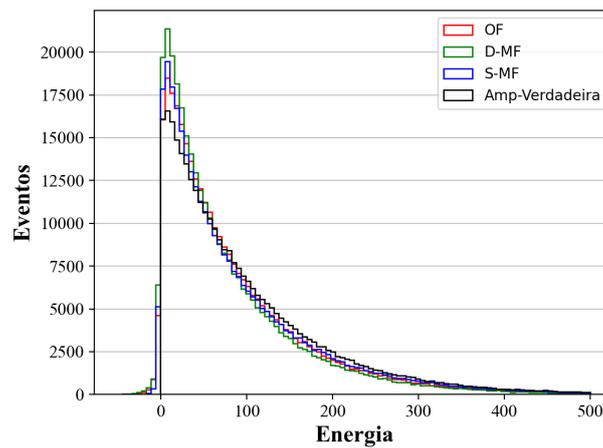
Legenda: Comparação entre amplitudes (em ADC) da célula  $E4$  para  $\langle \mu \rangle \approx 90$  considerando diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.

Fonte: O autor, 2021.

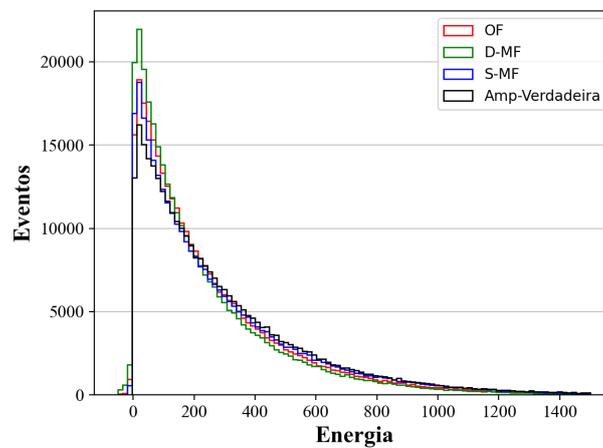
Figura 30 - Comparação entre amplitudes para a célula A9.



(a)



(b)

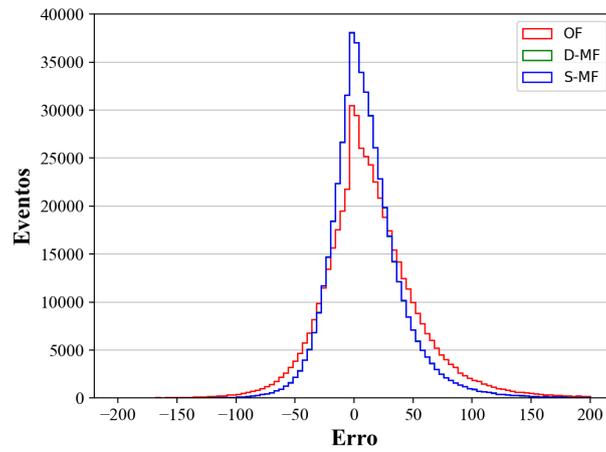


(c)

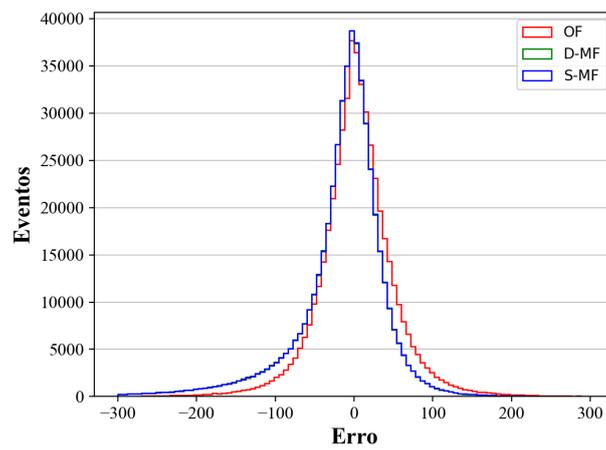
Legenda: Comparação entre amplitudes (em ADC) da célula A9 para  $\langle \mu \rangle \approx 30$  considerando diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.

Fonte: O autor, 2021.

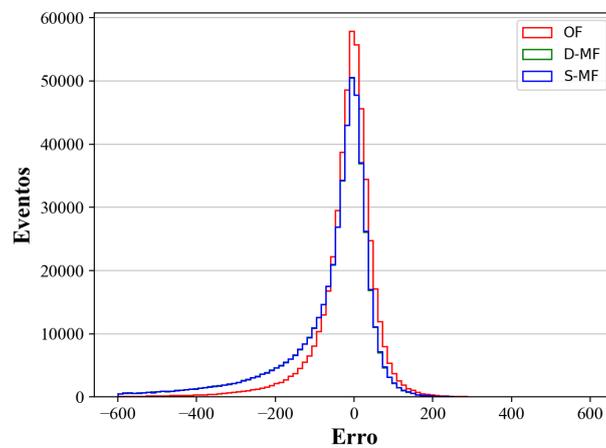
Figura 31 - Comparação entre erros de estimação para a célula  $E4$ .



(a)



(b)

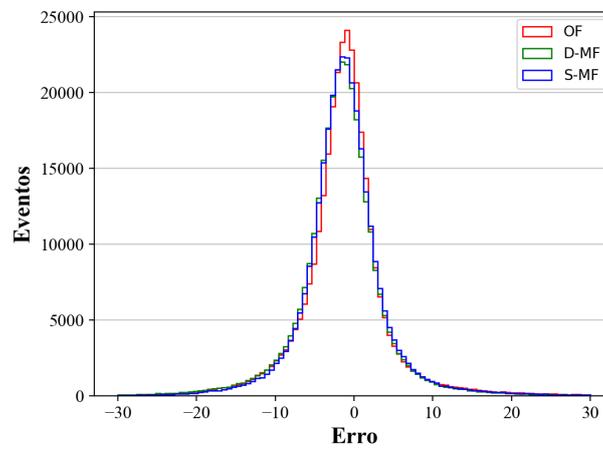


(c)

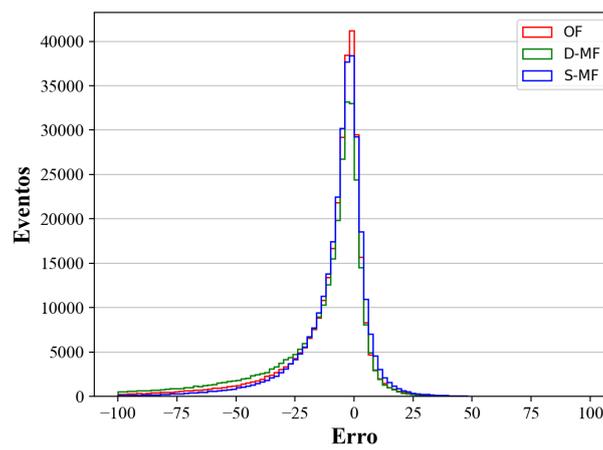
Legenda: Comparação entre erros de estimação (em ADC) da célula  $E4$  para  $\langle \mu \rangle \approx 90$  considerando diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.

Fonte: O autor, 2021.

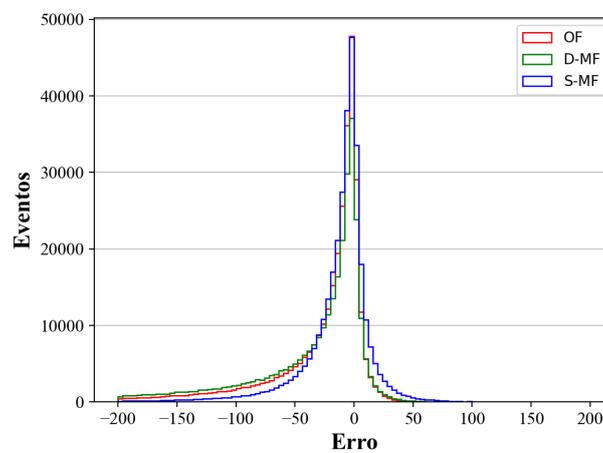
Figura 32 - Comparação entre erros de estimação para a célula A9.



(a)



(b)



(c)

Legenda: Comparação entre erros de estimação (em ADC) da célula A9 para  $\langle \mu \rangle \approx 30$  considerando diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.

Fonte: O autor, 2021.

Tabela 6 - médias e RMSs de cada método para  $E4 <\mu> \approx 90$  e suas respectivas amplitudes.

Método	Média	RMS
$SNR = 0,1$		
OF	14,7927	45,1022
DMF	9,9369	33,0740
SMF	10,0244	33,1045
$SNR = 1$		
OF	3,1279	52,5223
DMF	-17,8174	66,7615
SMF	-17,4486	66,1569
$SNR = 3$		
OF	-22,8733	93,0150
DMF	-79,6094	178,3217
SMF	-76,9586	172,2550

Fonte: O autor, 2021.

Tabela 7 - médias e RMSs de cada método para a célula  $A9 <\mu> \approx 30$  e suas respectivas amplitudes.

Método	Média	RMS
$SNR = 0,1$		
OF	-1,1934	7,1644
DMF	-1,5759	6,7808
SMF	-1,1354	6,7377
$SNR = 1$		
OF	-12,9186	27,8124
DMF	-20,5691	41,1488
SMF	-9,0791	19,8268
$SNR = 3$		
OF	-38,5124	80,3578
DMF	-62,0930	121,1335
SMF	-15,0852	39,3824

Fonte: O autor, 2021.

O erro relativo é dado por

$$\text{erro\_relativo} = (\text{amp\_aproximada} - \text{amp\_verdadeira}) / \text{amp\_verdadeira}. \quad (55)$$

Para  $E4$ , os erros relativos da Figura 33 mostram que a medida que a  $SNR$  diminui, mais concentradas são as aproximações, tendo como consequência clusters longe do eixo das abscissas, que representa erro 0. Este cenário fica ainda mais evidente para  $SNR = 0,1$ , indicando piores aproximações para baixas  $SNRs$ . Para  $SNR = 3$ , percebe-se pequenos clusters perto de  $x = 0$ , mas a medida que a amplitude de referencia aumenta, maior quantidade de aproximações se concentra em torno de zero, indicando que maiores  $SNRs$  irão propiciar melhores aproximações.

Diferentemente de  $E4$ , a Figura 33 mostrou que  $A9$  possui uma quantidade de erro acumulado próximo de  $x = 0$  bem menor. Novamente,  $SNRs$  maiores propiciaram melhores aproximações visto que os erros relativos passam a ser mais próximos de  $y = 0$  a medida que a  $SNR$  aumenta. Por fim, como esperado, o SMF possui a "barra" mais fina em torno de 0, também indicando que suas aproximações possuíram o menor erro relativo.

#### 4.4 Análise da PCA para o SMF

Devido ao uso da PCA no SMF, essa seção abordará brevemente o conceito e aplicabilidade da PCA, seguida da análise da eficiência de diferentes componentes da PCA aplicadas ao SMF.

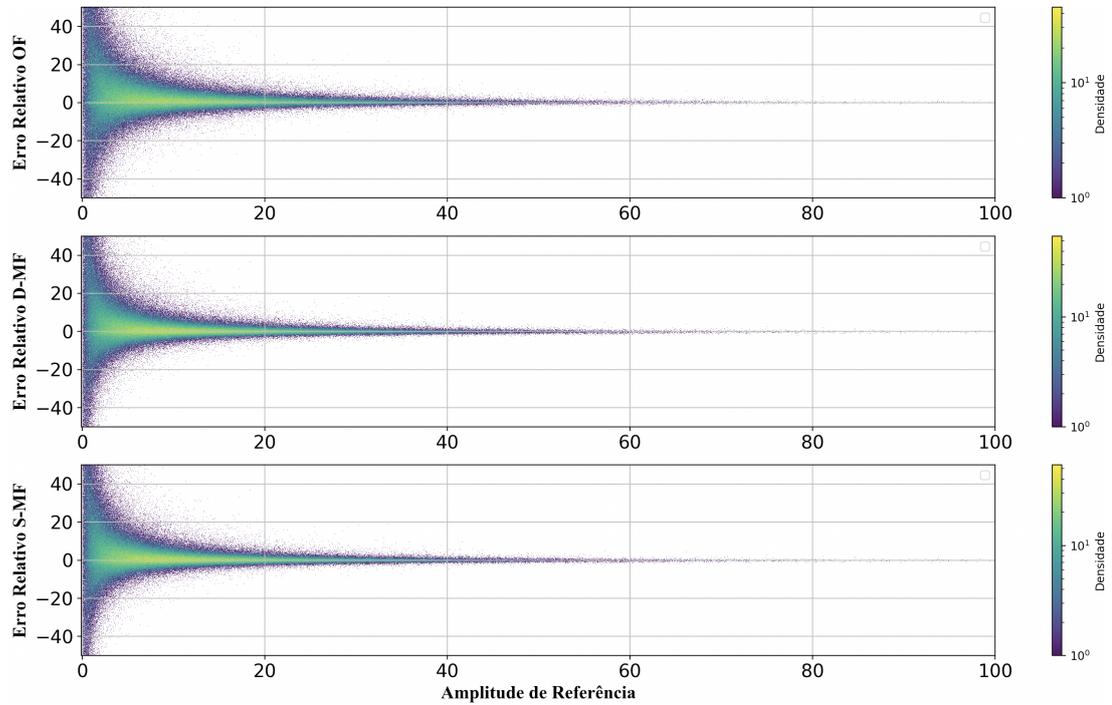
Sucintamente, a PCA é uma técnica de otimização usada em uma base de dados a fim de gerar uma nova base maximizando a decorrelação entre as variáveis aleatórias. Essa nova base possui suas componentes dispostas de forma a representar a maior quantidade de informação utilizando o menor número de variáveis através de seus respectivos auto-vetores e auto-valores.

Seus auto-valores são por sua vez ordenados de forma que a primeira componente represente a variável com maior energia, e a última represente a variável com menor energia. Desta forma, é possível representar um conjunto de dados consideravelmente menor e ótimo utilizando menos componentes que seu conjunto de dados original.

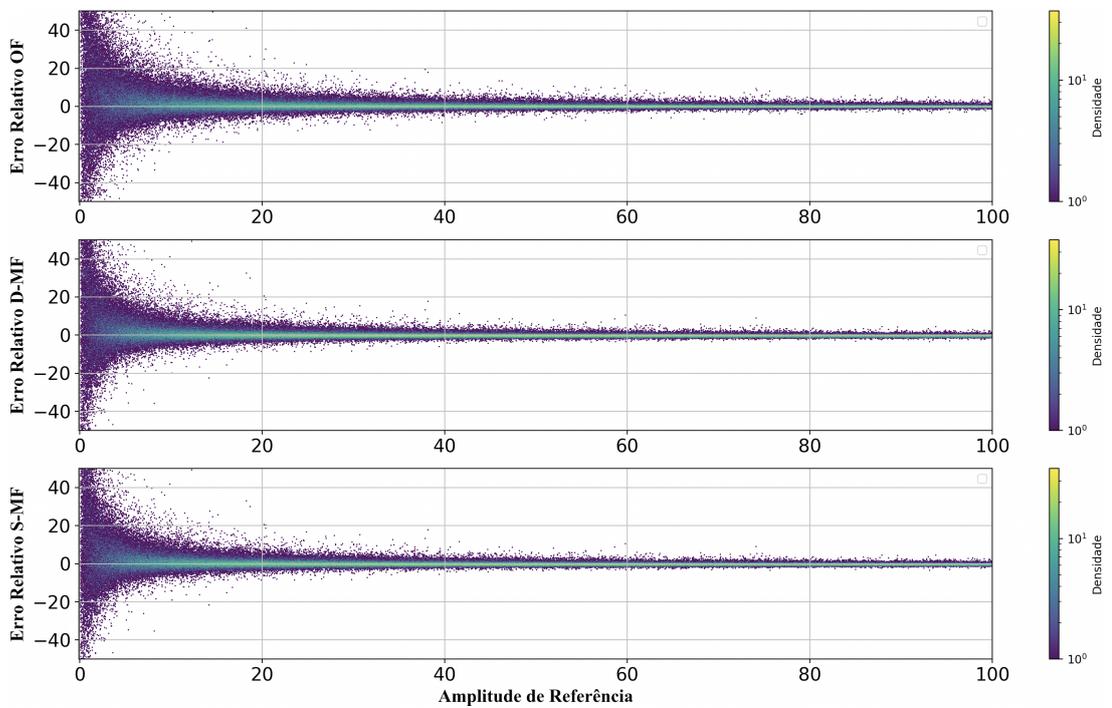
Para o caso do SMF, suas componentes com a energia acumulada estão dispostas na Figura 33. Já a Tabela 8 mostra os valores numéricos referente a cada componente.

Neste caso, percebe-se que a primeira componente descreve aproximadamente apenas 50% do processo aleatório. Já a terceira e quinta seriam capazes de entregar aproximadamente 97% e 99,9% respectivamente. Devido a este fato, as componentes selecionadas

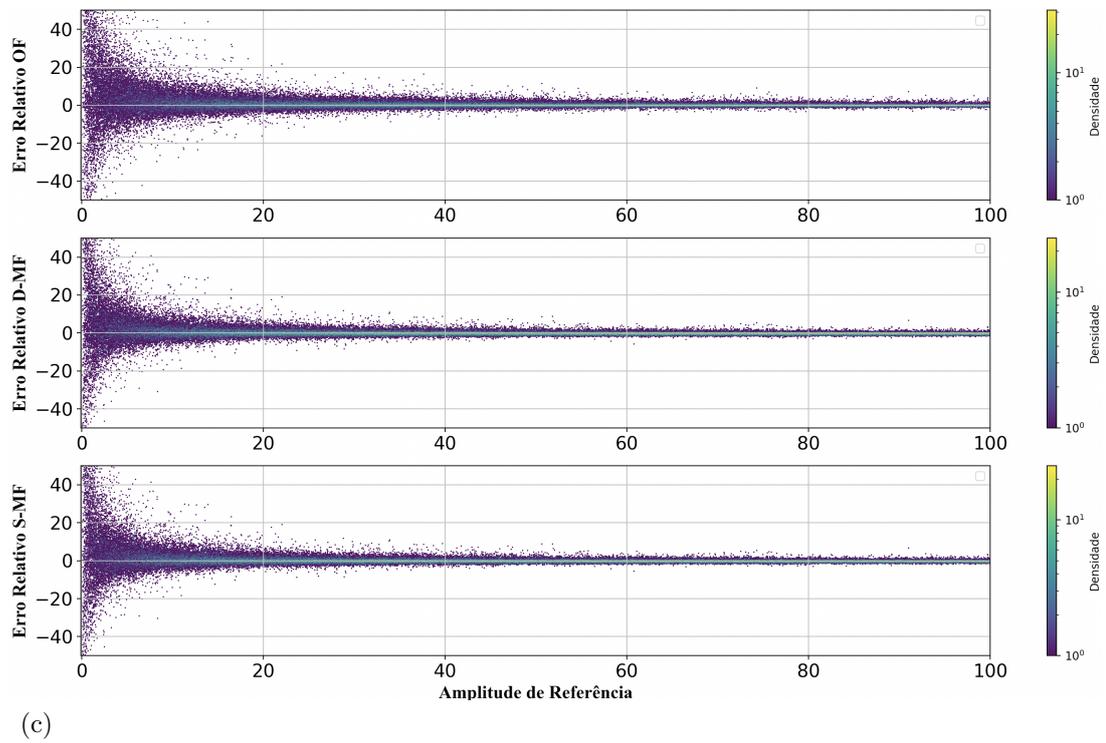
Figura 33 - Comparação entre erros relativos de estimação para  $E4 \langle \mu \rangle \approx 90$ .



(a)



(b)



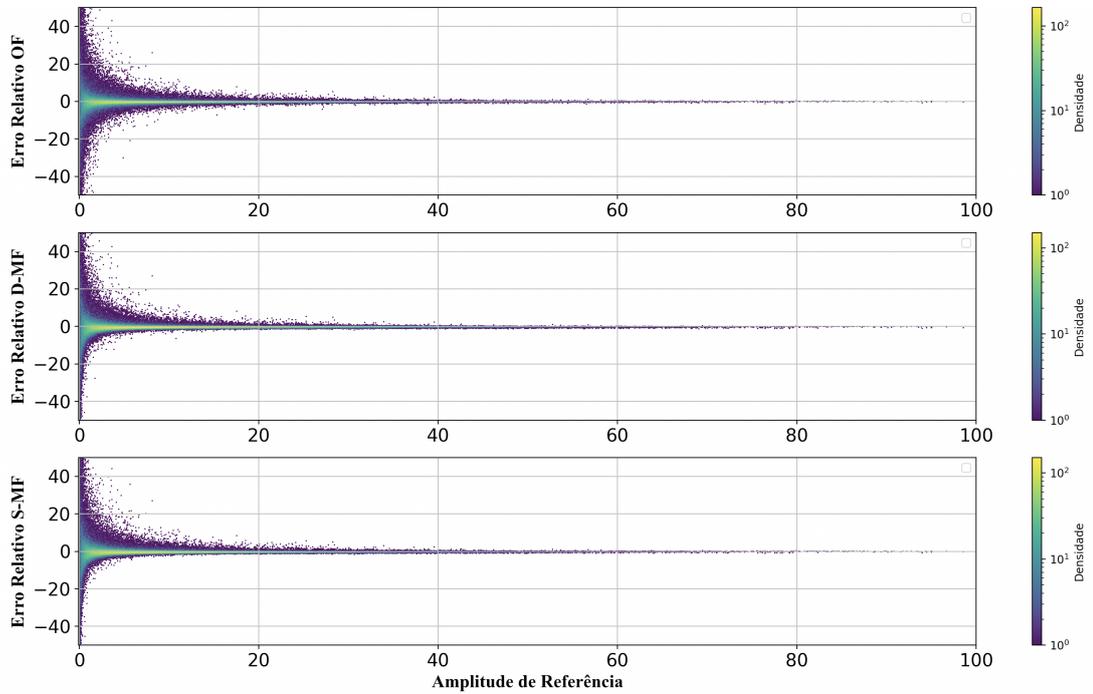
Legenda: Comparação do erro relativo da célula  $E4$  para diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.  
 Fonte: O autor, 2021.

Tabela 8 - Valores numéricos das componentes da PCA para o SMF.

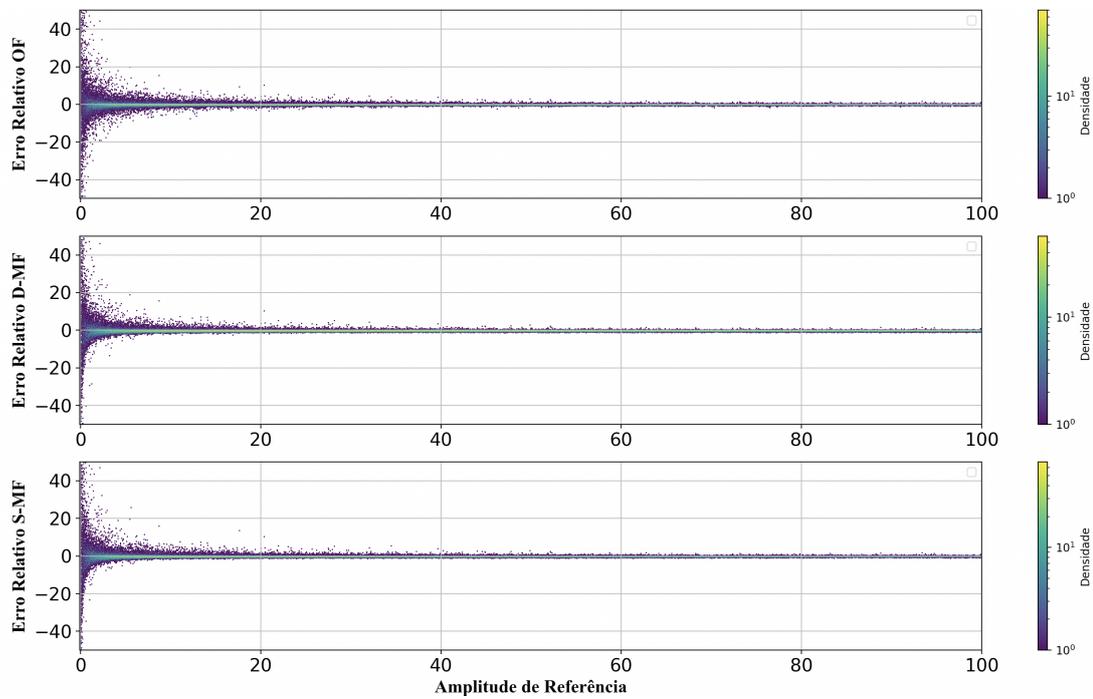
Componente	Energia (%)
1	50,2988
2	37,4277
3	9,2746
4	1,9877
5	0,9487
6	0,0369
7	0,0258

Fonte: O autor, 2021.

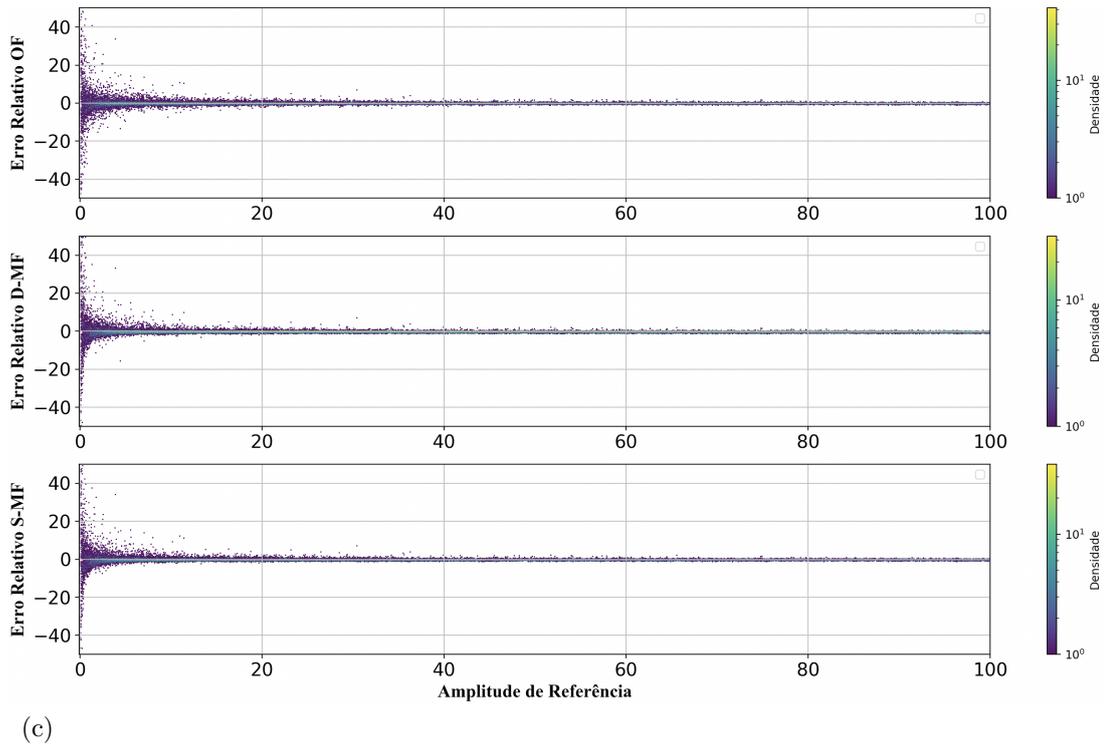
Figura 33 - Comparação entre erros relativos de estimação para  $A9 \langle \mu \rangle \approx 30$ .



(a)

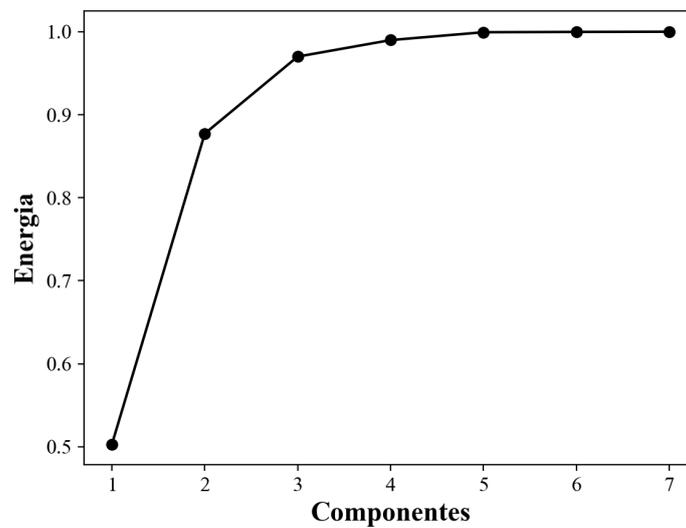


(b)



Legenda: Comparação do erro relativo da célula  $E4$  para diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.  
Fonte: O autor, 2021.

Figura 33 - Componentes da PCA para o SMF.



Fonte: O autor, 2021.

para as análises do SMF são: 1, 3, 5 e 7, onde esta última foi utilizada como base em todos os resultados anteriores deste capítulo.

Portanto, o estudo da PCA para as componentes  $PCA_{comps} = 1, 3, 5, 7$  será aplicado a seguir nos ambientes Simulado e Híbrido seguindo as mesmas estratégias usadas anteriormente.

#### 4.4.1 Ambiente Simulado

Esta seção abordará o estudo da eficiência do SMF para  $SNR = 1$  quando aplicados às quatro componentes da PCA considerando diferentes cenários de ocupação. A Figura 34 exibirá os erros de aproximação do SMF para as ocupações 0%, 50% e 100%. Já a Figura 35 irá mostrar as médias seguidas de seus respectivos RMSs em função de cada ocupação para todas  $PCA_{comps}$ . Por fim, a Tabela 9 destacará as aproximações numéricas para as ocupações 0%, 50% e 100%.

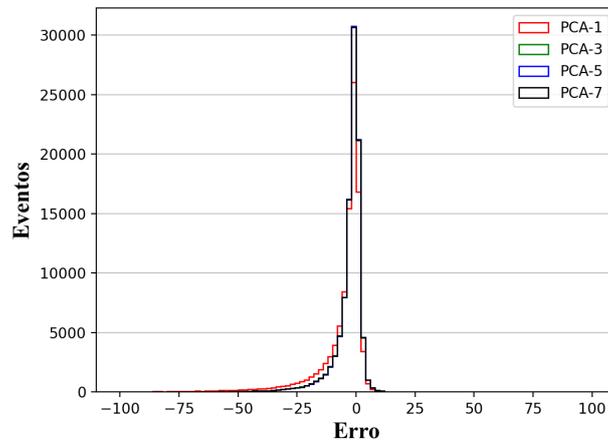
Tabela 9 - médias e RMSs para cada  $PCA_{comp}$  em função das ocupações.

$PCA_{comp}$	Média	Desvio Padrão
Ocupação = 0%		
1	-6.7052	14.1736
3	-3.4817	8.1115
5	-3.4285	8.0124
7	-3.4657	8.0814
Ocupação = 50%		
1	-6.8198	168.9243
3	27.4419	106.2489
5	28.1009	106.1600
7	28.6145	106.1552
Ocupação = 100%		
1	61.8366	179.4365
3	111.8549	126.0592
5	113.1698	126.0002
7	114.2537	125.9966

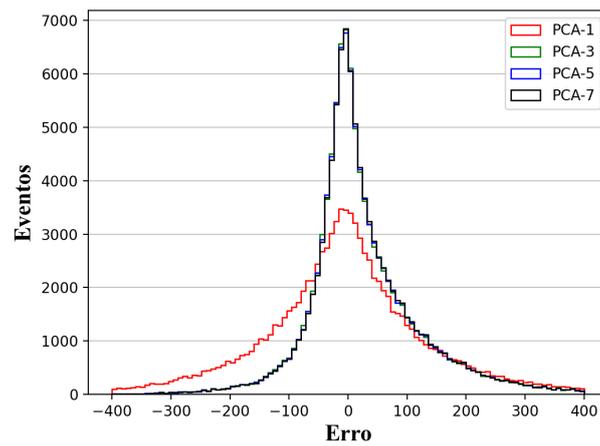
Fonte: O autor, 2021.

Para os dados simulados apresentados pelas Figuras 34 e 35, percebe-se que independentemente das ocupações utilizadas, as  $PCA_{comps} = 3, 5, 7$  obtiveram aproximações muito próximas. Isso mostra que a terceira componente, responsável por representar aproximadamente 97% dos dados pode ser considerada a melhor escolha em termos de

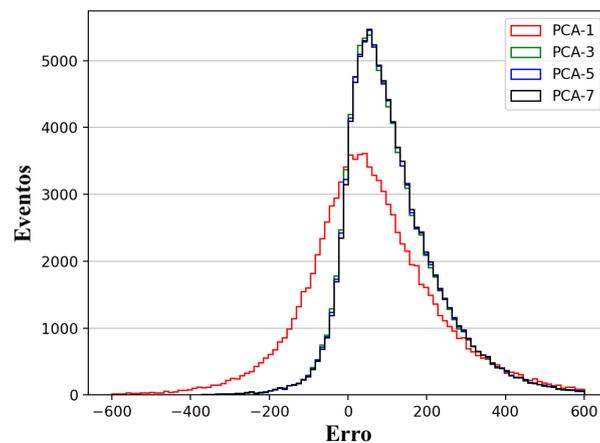
Figura 34 - Comparação entre  $PCA_{comps}$  considerando diferentes ocupações para  $SNR = 1$ .



(a)



(b)

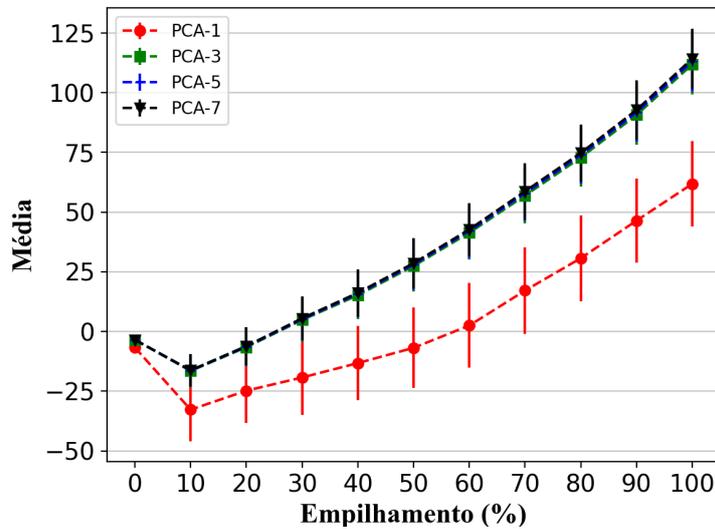


(c)

Legenda: Comparação dos erros gerados (em ADC) em cada  $PCA_{comp}$  para diferentes ocupações. (a) Considera ocupação de 0%, (b) 50% e (c) 100%.

Fonte: O autor, 2021.

Figura 35 - Comparação entre erros de estimação de cada  $PCA_{comp}$  em função de todas as ocupações.



Legenda: As barras de erro foram divididas por um fator de 10 para melhor visualização.

Fonte: O autor, 2021.

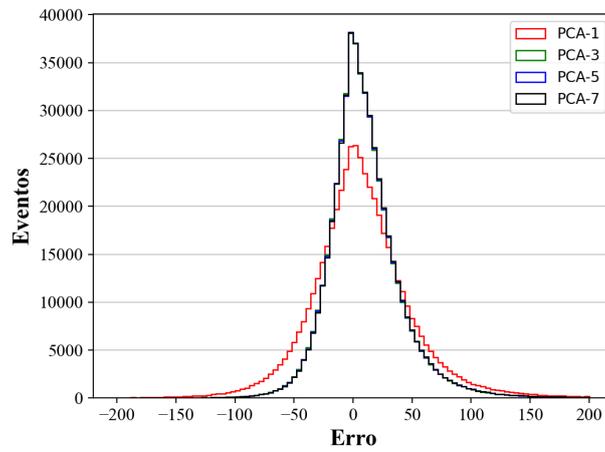
qualidade e velocidade de execução, visto que menos componentes implicará em menos operações matriciais. Já a  $PCA_{comp} = 1$ , responsável por representar aproximadamente apenas 50% teve as piores aproximações, como esperado devido sua baixa porcentagem de energia.

#### 4.4.2 Ambiente Híbrido

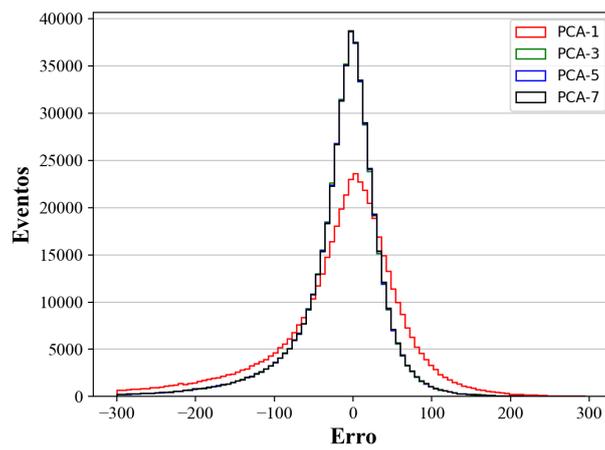
Esta seção abordará o estudo da eficiência do SMF para as células  $E4$  e  $A9$  considerando diferentes  $SNRs$  para às quatro componentes da PCA. As Figuras 36 e 37 exibirão os erros de aproximação do SMF para as  $PCA_{comps} = 1, 3, 5, 7$  considerando as  $SNRs = 0, 1, 1$  e  $3$ . Já a Figura 38 irá mostrar as médias seguidas de seus respectivos RMSs e por fim, a Tabela 10 destacará as aproximações numéricas.

Assim como ocorreu para a abordagem simulada, a variação entre as  $PCA_{comps} = 3, 5, 7$  foram mínimas e satisfatórias, destacando-se mais uma vez a  $PCA_{comp} = 3$  devido sua melhor relação qualidade e velocidade de execução. Igualmente ao ocorrido para a abordagem simulada, a  $PCA_{comp} = 1$  também obteve o pior resultado, podendo ser descartada como uma alternativa de aproximação.

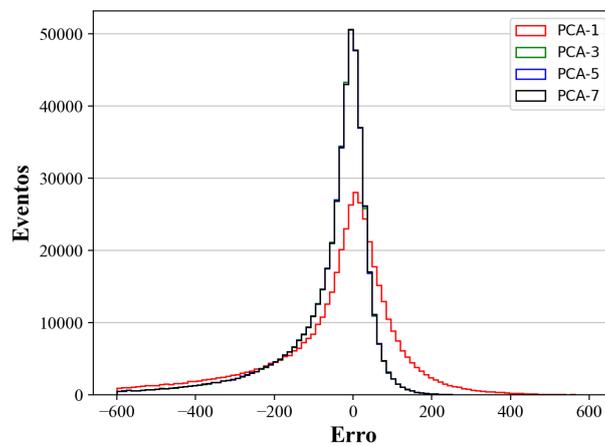
Figura 36 - Comparação entre erros de estimação de cada  $PCA_{comp}$  para a célula  $E4$ .



(a)



(b)

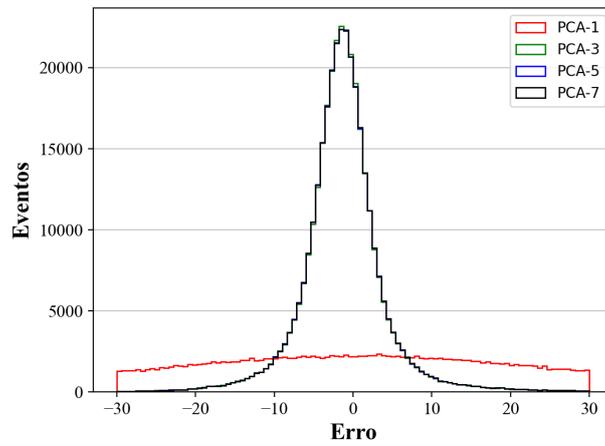


(c)

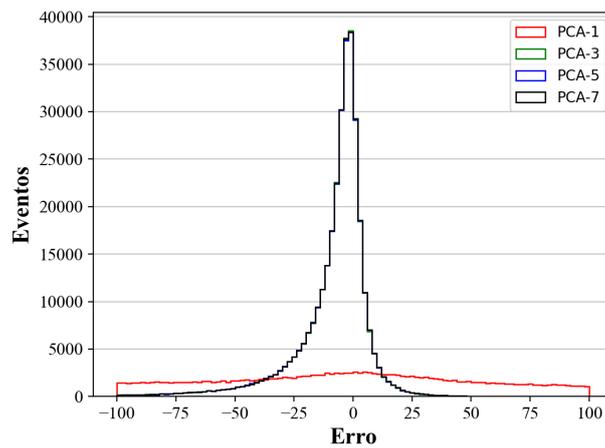
Legenda: Comparação dos erros gerados (em ADC) em cada  $PCA_{comps}$  da célula  $E4$  para  $\langle \mu \rangle \approx 90$  considerando diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.

Fonte: O autor, 2021.

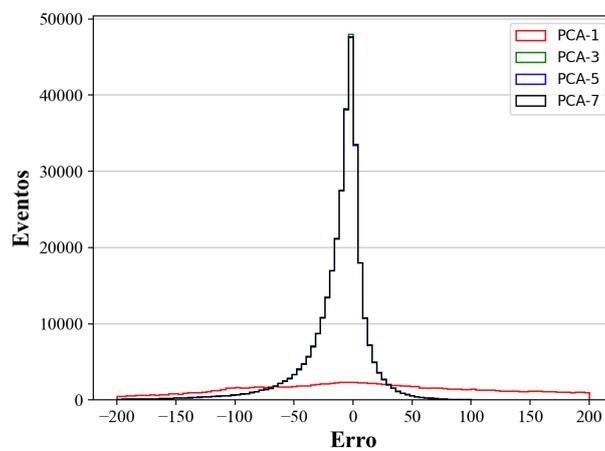
Figura 37 - Comparação entre erros de estimação de cada  $PCA_{comp}$  para a célula A9.



(a)



(b)

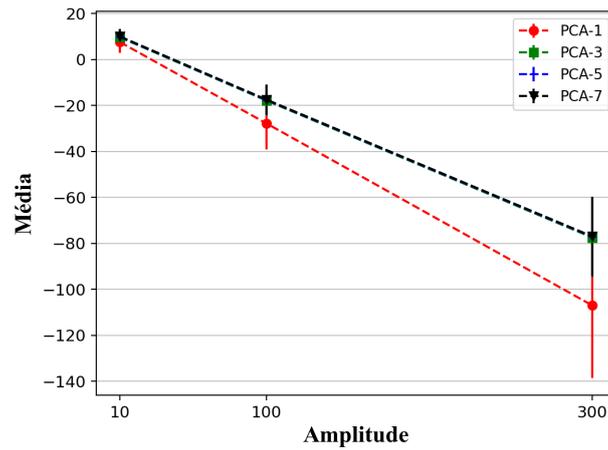


(c)

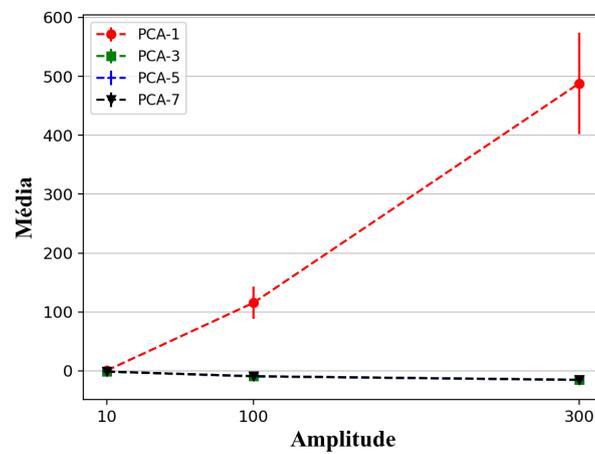
Legenda: Comparação dos erros gerados (em ADC) em cada  $PCA_{comps}$  da célula A9 para  $\langle \mu \rangle \approx 30$  considerando diferentes  $SNRs$ .  $SNRs$ : (a) 0,1, (b) 1 e (c) 3.

Fonte: O autor, 2021.

Figura 38 - Comparação entre as médias e RMSs para cada  $PCA_{comp}$  em função das amplitudes.



(a)



(b)

Legenda: médias das células: (a) E4, (b) A9. As barras de erro foram divididas por um fator de 10 para melhor visualização.

Fonte: O autor, 2021.

Tabela 10 - médias e RMSs para cada  $PCA_{comp}$  em função das amplitudes.

$PCA_{comp}$	média $E4$	média $A9$
$SNR = 0,1$		
1	$7,5873 \pm 46,6563$	$0,7398 \pm 49,2427$
3	$9,6490 \pm 33,1364$	$-1,1550 \pm 6,7491$
5	$9,8655 \pm 33,1056$	$-1,1410 \pm 6,7384$
7	$10,0244 \pm 33,1045$	$-1,1354 \pm 6,7377$
$SNR = 1$		
1	$-27,8006 \pm 113,2587$	$115,6170 \pm 272,3430$
3	$-17,8660 \pm 66,3004$	$-9,1844 \pm 20,0783$
5	$-17,6050 \pm 66,1581$	$-9,1575 \pm 19,9070$
7	$-17,4486 \pm 66,1569$	$-9,0791 \pm 19,8268$
$SNR = 3$		
1	$-106,9646 \pm 315,6145$	$488,0812 \pm 863,5177$
3	$-77,4491 \pm 172,7041$	$-15,2738 \pm 39,9635$
5	$-77,1155 \pm 172,2648$	$-15,2875 \pm 39,5406$
7	$-76,9586 \pm 172,2550$	$-15,0852 \pm 39,3824$

Fonte: O autor, 2021.

## CONCLUSÃO

Neste trabalho, o experimento ATLAS do LHC e seu calorímetro de telhas, TileCal, foram apresentados em detalhes. Nos próximos anos de operação, uma série de atualizações do LHC é prevista com o objetivo de aumentar consideravelmente a luminosidade do feixe, produzindo mais sinais a cada colisão de seus feixes de partículas. O aumento da luminosidade introduz novos desafios à atual estrutura do ATLAS e do TileCal como, por exemplo, o aumento do efeito de empilhamento de sinais que degrada a reconstrução da energia depositada nos canais de leitura do TileCal. Neste contexto, três métodos tipicamente utilizados em calorímetros modernos para estimar a energia foram descritos: o Filtro Ótimo (OF), Filtro Casado Determinístico (DMF) e Filtro Casado Estocástico (SMF).

Os métodos testados foram baseados em dois conjuntos de dados. O conjunto de dados simulados teve o objetivo de reproduzir sinais provenientes das colisão ocorridas no TileCal, simulando também diferentes taxas de ocupação que produzem diferentes condições de empilhamento de sinais. Já a segunda base de dados, um conjunto de dados híbrido teve o objetivo de reproduzir condições mais próximas possível das condições reais, por isso, foi composto de diferentes ruídos reais acrescidos de sinais com amplitudes. Células com diferentes níveis de ocupação foram escolhidas para a análise de eficiência.

Os resultados simulados mostraram que todos os métodos conseguem uma aproximação da amplitude com boa eficiência. No entanto, o método OF tende a ter caudas negativas indesejadas à medida que o efeito de empilhamento de sinais aumenta, visto que características não gaussianas são adicionadas. Vale destacar que o método SMF apresentou uma melhor capacidade de lidar com os efeitos de defasagem do sinal recebido, visto que este efeito não é considerado nos projetos dos métodos DMF e OF.

O DMF e SMF foram similares em suas aproximações, e sua média cresceu linearmente a medida que a ocupação aumentou, o que pode indicar que uma correção linear nestes métodos pode melhorar suas aproximações. O SMF se destacou em suas estimativas de energia considerando baixas ocupações, visto que, independente da  $SNR$  usada, seu erro médio e RMS foram melhores que os demais métodos.

Já para os resultados híbridos, notou-se a dificuldade dos métodos conseguirem estimar a amplitude em condições severas de empilhamento. Entretanto, à medida que a  $SNR$  aumenta, melhores foram as estimações. Entre os métodos, notou-se novamente a similaridade das aproximações feitas pelo DMF e SMF, isso provavelmente se dá ao fato de ambos usarem a informação da matriz de covariância do ruído em seus projetos. Conforme observado para os dados simulados, para baixas ocupações, o SMF obteve um desempenho superior aos outros métodos considerados.

Em condições severas de empilhamento, o OF demonstrou ser capaz de realizar

aproximações melhores, principalmente para alta  $SNR$ . No entanto, conforme constatado nos dados simulados, o SMF apresentou o melhor desempenho para baixas condições de empilhamento de sinais. Nestas condições, comparando com os métodos OF e DMF, o melhor resultado foi alcançado na condição de  $SNR = 3$ , onde o SMF foi 60,8% superior ao OF e 75,7% superior ao DMF, em termos da dispersão do erro de estimação.

Também foi apresentado um estudo de compactação de dados utilizando o método PCA. Foi constatado que o método SMF pode ser projetado utilizando-se, por exemplo, três componentes principais, reduzindo o esforço computacional. A representação dos dados utilizando três e cinco componentes computacionais mostraram que, respectivamente, 97% e 99,9% da variância do conjunto de dados é mantida. Desta forma o projeto do método SMF pode ser consideravelmente simplificado sem perda de eficiência.

## Trabalhos Futuros

Como trabalhos futuros, sugere-se a análise de correção linear aplicada às aproximações do Filtro Casado Determinístico e do Filtro Casado Estocástico. Adicionalmente, sugere-se uma análise do comportamento das estimações de cada método em função do desvio de fase do sinal de resposta a fim de se estudar a sensibilidade de cada método ao desvio de fase.

O SMF se mostrou mais eficiente em células com menor empilhamento, por isso análises envolvendo células mais distantes do ponto de colisão como, por exemplo as células  $D$  da Figura 10, podem indicar bons resultados da aplicação deste método. Além disso, técnicas não-lineares como redes neurais artificiais podem ser elaboradas e aplicadas no contexto *offline*, a fim de lidar com os efeitos indesejáveis e não-lineares do empilhamento de sinais e desvio de fase.

## REFERÊNCIAS

- AAD, Georges et al. Evidence for Higgs boson decays to a low-mass dilepton system and a photon in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Lett. B*, v. 819, p. 136412. 39 p, Mar 2021. Disponível em: <<http://cds.cern.ch/record/2757278>>.
- AAMODT, K. et al. The ALICE experiment at the CERN LHC. *JINST*, v. 3, p. S08002, 2008.
- ADRAGNA, P. et al. The atlas hadronic tile calorimeter: from construction toward physics. *IEEE Transactions on Nuclear Science*, v. 53, n. 3, p. 1275–1281, 2006.
- ALVES JR., A. Augusto et al. The LHCb Detector at the LHC. *JINST*, v. 3, p. S08005, 2008.
- ANDERSON, Brian D. O.; MOORE, John B. Book. *Optimal filtering / Brian D. O. Anderson, John B. Moore*. [S.l.]: Prentice-Hall Englewood Cliffs, N.J, 1979. x, 357 p. ; p. ISBN 0136381227.
- BARBOSA, Davis Pereira. *Estimação de energia para calorimetria em física de altas energias baseada em representação esparsa*. 2017. 102 f. Doutorado em Engenharia Elétrica — Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, 2017.
- BERTUCCIO, Giuseppe et al. Sampling and optimum data processing of detector signals. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Elsevier, v. 322, n. 2, p. 271–279, 1992.
- CALDWELL, A et al. Design and implementation of a high precision readout system for the ZEUS calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Elsevier, v. 321, n. 1-2, p. 356–364, 1992.
- CAVALCANTI, Elmano Pontes. *REVOLUÇÃO DA INFORMAÇÃO: ALGUMAS REFLEXÕES*. São Paulo: [s.n.], 1995. Disponível em: <<http://www.ancibe.com.br/artigos%20de%20si/artigo%20-%20Revoluç~ao%20da%20informaç~ao%20-%20algumas%20reflex~oes.pdf>>.
- CERN. *World-Wide Web: invented at CERN*. Geneva: CERN, 1998. Disponível em: <<https://cds.cern.ch/record/2634986>>. Acesso em: 12 fev. 2020.
- \_\_\_\_\_. The first capacitive touch screens at CERN. *CERN Courier*, v. 50, n. 3, p. 13–16, Apr 2010. Disponível em: <<https://cds.cern.ch/record/1734498>>. Acesso em: 05 fev. 2020.
- \_\_\_\_\_. *Tile Calorimeter Public Plots*. 2013. Disponível em: <<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTile>>. Acesso em: 24 mar. 2021.
- \_\_\_\_\_. *LHC Season 2. Facts figures*. 2018. Apresenta Fatos e importantes contribuições do LHC ao longo dos anos. Disponível em: <[https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en\\_0.pdf](https://home.cern/sites/home.web.cern.ch/files/2018-07/factsandfigures-en_0.pdf)>. Acesso em: 15 jun. 2021.

\_\_\_\_\_. *CERN Accelerating science*. 2020. Disponível em: <<https://home.cern/science/accelerators/large-hadron-collider>>. Acesso em: 11 nov. 2020.

\_\_\_\_\_. \_\_\_\_\_. 2020. Disponível em: <<https://home.cern/science/engineering/vacuum-empty-interstellar-space>>. Acesso em: 11 nov. 2020.

\_\_\_\_\_. *História do CERN*. 2020. Apresenta eventos históricos relacionados ao CERN. Disponível em: <<https://home.cern/about/who-we-are/our-history>>. Acesso em: 15 jun. 2021.

\_\_\_\_\_. *CERN Accelerating science*. 2021. Disponível em: <[https://cds.cern.ch/record/2741140/files/DJI\\_0007b.jpg](https://cds.cern.ch/record/2741140/files/DJI_0007b.jpg)>. Acesso em: 11 jul. 2021.

\_\_\_\_\_. \_\_\_\_\_. 2021. Disponível em: <<https://cds.cern.ch/record/40525>>. Acesso em: 11 jul. 2021.

CERQUEIRA, A.S. ATLAS Tile Calorimeter Readout Electronics Upgrade Program for the High Luminosity LHC. In: *Proceedings, International School on High Energy Physics : Workshop on High Energy Physics in the near Future. (LISHEP 2013): Rio de Janeiro, Brazil, March 17-24, 2013*. [S.l.: s.n.], 2013.

CHAPMAN, John. ATLAS simulation computing performance and pile-up simulation in ATLAS. In: *LPCC Detector Simulation Workshop*. [S.l.: s.n.], 2011.

CIODARO, T.; SEIXAS, J. M. de; CERQUEIRA, A. Use of hadronic calorimetry information in the atlas level-1 muon trigger. *IEEE Transactions on Nuclear Science*, v. 61, n. 2, p. 1047–1055, 2014.

CLELAND, WE; STERN, EG. Signal processing considerations for liquid ionization calorimeters in a high rate environment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Elsevier, v. 338, n. 2-3, p. 467–497, 1994.

COLLABORATION, ATLAS. *ATLAS tile calorimeter: Technical Design Report*. Geneva: CERN, 1996. (Technical design report. ATLAS). Disponível em: <<https://cds.cern.ch/record/331062>>.

\_\_\_\_\_. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, v. 3, p. S08003. 437 p, 2008. Also published by CERN Geneva in 2010. Disponível em: <<https://cds.cern.ch/record/1129811>>. Acesso em: 12 fev. 2021.

\_\_\_\_\_. Readiness of the ATLAS Tile Calorimeter for LHC collisions. *Eur. Phys. J. C*, v. 70, p. 1193–1236, 2010.

\_\_\_\_\_. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, Elsevier, v. 716, n. 1, p. 1–29, 2012.

\_\_\_\_\_. *Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System*. [S.l.], 2013. Final version presented to December 2013 LHCC. Disponível em: <<https://cds.cern.ch/record/1602235>>. Acesso em: 30 jun. 2021.

COLLABORATION, CMS. Time reconstruction and performance of the CMS electromagnetic calorimeter. *Journal of Instrumentation*, IOP Publishing, v. 5, n. 03, p. T03011–T03011, mar 2010. Disponível em: <<https://doi.org/10.1088/1748-0221/5/03/T03011>>. Acesso em: 04 mar. 2021.

D0, Collaboration. The upgraded D0 detector. *arXiv preprint physics/0507191*, 2005.

DAS, Ashok; FARBEL, Thomas. *Introduction to Nuclear and Particle Physics*. 2. ed. Singapore: World Scientific Publishing Company, 2003. 416 p.

DELMASTRO, Marco. A stand-alone signal reconstruction and calibration algorithm for the ATLAS electromagnetic calorimeter. In: IEEE. *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515)*. [S.l.], 2003. v. 2, p. 1110–1114.

EDWARDS, Donald A; SYPHERS, Michael J. *An introduction to the physics of high energy accelerators*. [S.l.]: John Wiley & Sons, 2008.

ERIKSSON, Daniel et al. A prototype for the upgraded readout electronics of TileCal. *Journal of Instrumentation*, IOP Publishing, v. 7, n. 02, p. C02006, 2012.

LHC Machine. *JINST*, v. 3, p. S08001, 2008.

FRANCAVILLA, Paolo. The ATLAS tile hadronic calorimeter performance at the LHC. *Journal of Physics: Conference Series*, IOP Publishing, v. 404, p. 012007, dec 2012. Disponível em: <<https://doi.org/10.1088/1742-6596/404/1/012007>>.

FULLANA, Esteban et al. Digital signal reconstruction in the ATLAS hadronic tile calorimeter. *IEEE transactions on nuclear science*, IEEE, v. 53, n. 4, p. 2139–2143, 2006.

GONÇALVES, Guilherme Inácio. *Estimação de energia para calorímetros finamente segmentados operando em condições severas de empilhamento de sinais*. 2020. 95 f. Dissertação (Mestrado em Modelagem Computacional) — Universidade do Estado do Rio de Janeiro (UERJ), Nova Friburgo, 2020.

GRIFFITHS, David. *Introduction to elementary particles*. [S.l.]: John Wiley & Sons, 2008.

JENNI, Peter et al. *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Geneva: CERN, 2003. (Technical design report. ATLAS). Disponível em: <<https://cds.cern.ch/record/616089>>.

JOLLIFFE, I. T.; CADIMA, Jorge. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 374, 2016.

KAY, Steven M. *Fundamentals of statistical signal processing*. Upper Saddle River, NJ: Prentice Hall PTR, 1993. (Prentice Hall signal processing series). Disponível em: <<https://cds.cern.ch/record/2012069>>. Acesso em: 12 fev. 2021.

KESSY, Agnan; LEWIN, Alex; STRIMMER, Korbinian. Optimal whitening and decorrelation. *The American Statistician*, American Statistical Association, v. 72, n. 4, p. 309–314, out. 2018. ISSN 1537-2731.

- KHANDAI, P. K. et al. Hadron spectra in p+p collisions at rhic and lhc energies. *International Journal of Modern Physics A*, World Scientific Pub Co Pte Lt, v. 28, n. 16, p. 1350066, Jun 2013. ISSN 1793-656X. Disponível em: <<http://dx.doi.org/10.1142/S0217751X13500668>>.
- KLIMEK, Pawel. Quality Factors in TileCal and out-of-time Pile-up. Oct 2011. Disponível em: <<http://cds.cern.ch/record/1392389>>. Acesso em: 04 mar. 2021.
- KNOLL, Glen. *Radiation Detection and Measurement (4th ed.)*. Hoboken, NJ: John Wiley, 2010. ISBN 978-0-470-13148-0.
- LEO, William R. *Techniques for nuclear and particle physics experiments: a how-to approach; 2nd ed.* Berlin: Springer, 1994. Disponível em: <<https://cds.cern.ch/record/302344>>.
- LUNDBERG, Johan. Performance of the ATLAS first-level trigger with first LHC data. In: IEEE. *2010 17th IEEE-NPSS Real Time Conference*. [S.l.], 2010. p. 1–5.
- MARJANOVIĆ, Marija. Atlas tile calorimeter calibration and monitoring systems. *IEEE Transactions on Nuclear Science*, v. 66, n. 7, p. 1228–1235, 2019.
- MEYER, Chris. The ATLAS Tile Calorimeter Calibration and Performance. *EPJ Web Conf.*, v. 60, n. arXiv:1310.2945. ATL-TILECAL-PROC-2013-008, p. 20051. 3 p, Oct 2013. Comments: 3 pages, 8 figures. Proceeding to LHCP 2013: First Large Hadron Collider Physics Conference, Barcelona, Spain, 13-18 May 2013. Disponível em: <<https://cds.cern.ch/record/1558952>>. Acesso em: 20 fev. 2021.
- MITRA, Sanjit Kumar; KUO, Yonghong. *Digital signal processing: a computer-based approach*. [S.l.]: McGraw-Hill New York, 2006. v. 2.
- MOBS, Esmá. The CERN accelerator complex. Complexe des accélérateurs du CERN. Jul 2016. General Photo. Disponível em: <<https://cds.cern.ch/record/2684277>>. Acesso em: 07 ago. 2021.
- PAGANINI, Pascal; VULPEN, Ivo Van. *Pulse amplitude reconstruction in the CMS ECAL using the weights method 025*. Geneva, 2004. Disponível em: <<https://cds.cern.ch/record/801816>>. Acesso em: 20 fev. 2021.
- PAPOULIS, Athanasios; PILLAI, S Unnikrishna. *Probability, random variables, and stochastic processes*. [S.l.]: Tata McGraw-Hill Education, 2002.
- PASTORE, F. The atlas trigger system: Past, present and future. *Nuclear and Particle Physics Proceedings*, v. 273-275, p. 1065–1071, 2016. ISSN 2405-6014. 37th International Conference on High Energy Physics (ICHEP). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405601415006562>>.
- PEEBLES, Peyton Z. *Probability, random variables, and random signal principles; 2nd ed.* New York, NY: McGraw Hill, 1987. Disponível em: <<https://cds.cern.ch/record/2705970>>. Acesso em: 11 fev. 2020.
- PEQUENAO, Joao. Computer Generated image of the ATLAS calorimeter. 2008. Disponível em: <<https://cds.cern.ch/record/1095927>>. Acesso em: 11 fev. 2021.

\_\_\_\_\_. Computer generated image of the whole ATLAS detector. 2008. Disponível em: <<https://cds.cern.ch/record/1095924>>. Acesso em: 11 fev. 2021.

PEQUENAO, Joao; SCHAFFNER, Paul. How ATLAS detects particles: diagram of particle paths in the detector. 2013. Disponível em: <<https://cds.cern.ch/record/1505342>>. Acesso em: 11 fev. 2021.

PERALVA, B.S. et al. A matched filter based approach for high-energy estimation in calorimetry. *Journal of Instrumentation*, IOP Publishing, v. 16, n. 02, p. P02016–P02016, feb 2021. Disponível em: <<https://doi.org/10.1088/1748-0221/16/02/p02016>>.

PERALVA, Bernardo S. The tilecal energy reconstruction for collision data using the matched filter. In: *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. [s.n.], 2013. p. 1–6. Disponível em: <<https://cds.cern.ch/record/1614492>>.

PERALVA, Bernardo Sotto-Maior. *Detecção de sinais e estimação de energia para calorimetria de altas energias*. 2012. 79 f. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, 2012.

PERALVA, Bernardo Sotto-Maior. *Reconstrução de energia para calorímetros finamente segmentados*. 2015. 142 f. Mestrado em Engenharia Elétrica — Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, 2015.

PERALVA, B. Sotto-Maior. The TileCal Online Energy Estimation for the Next LHC Operation Period. *J. Phys. Conf. Ser.*, v. 608, n. 1, p. 012043, 2015.

SÁNCHEZ, Carlos Antonio Solans. Implementation of the ROD Crate DAQ Software for the ATLAS Tile Calorimeter and a Search for a MSSM Higgs Boson decaying into Tau pairs. 2010. Disponível em: <<https://cds.cern.ch/record/1309926>>. Acesso em: 19 fev. 2021.

SEIXAS, J.M.; CALOBA, L.P.; PINTO, L.A.F. Fast particle discriminator based on matched filters. In: *Proceedings of the 39th Midwest Symposium on Circuits and Systems*. [S.l.: s.n.], 1996. v. 2, p. 843–846 vol.2.

\_\_\_\_\_. \_\_\_\_\_. In: *Proceedings of the 39th Midwest Symposium on Circuits and Systems*. [S.l.: s.n.], 1996. v. 2, p. 843–846 vol.2.

THEODORIDIS, Sergios. *Machine learning: a Bayesian and optimization perspective*. [S.l.]: Academic Press, 2015.

TOFFLER, Alvin. *Powershift: As Mudanças Do Poder: Um perfil da sociedade do século xxi pela análise das transformações na natureza do poder*. 1. ed. Rio de Janeiro: Record, 1991. 613p p.

TREES, Harry L Van. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. [S.l.]: John Wiley & Sons, 2004.

TYLMAD, Maja. Pulse shapes for signal reconstruction in the ATLAS Tile Calorimeter. In: *IEEE. 2009 16th IEEE-NPSS Real Time Conference*. [S.l.], 2009. p. 543–547.

VANKOV, P. *ATLAS Upgrades Towards the High Luminosity LHC: extending the discovery potential*. Geneva, 2013. Disponível em: <<https://cds.cern.ch/record/1631339>>.

WESTON, Thomas.

*Searching for Supersymmetry using the Higgs Boson*, 2019. Presented 17 Dec 2019. Disponível em: <<https://cds.cern.ch/record/2709365>>.

WIEDEMANN, Helmut. *Particle accelerator physics*. [S.l.]: Springer, 2015.

WIGMANS, R. *Calorimetry: Energy measurement in particle physics*. [S.l.: s.n.], 2000. v. 107.

\_\_\_\_\_. *Calorimetry: Energy Measurement in Particle Physics*. OUP Oxford, 2017. (International Series of Monographs on Physics). ISBN 9780191089695. Disponível em: <<https://books.google.com.br/books?id=vJc4DwAAQBAJ>>. Acesso em: 12 fev. 2020.

## APÊNDICE A – Publicações científicas

### A.1 Trabalhos publicados em anais de congressos e revistas

1. ROCHA G. P., PERALVA B. S., SEIXAS J. M., *Estimação De Energia Baseada Em Filtragem Casada Para Um Calorímetro Operando Em Condições De Empilhamento De Sinais*, Anais do XXIII ENMC – Encontro Nacional de Modelagem Computacional e XI ECTM – Encontro de Ciências e Tecnologia de Materiais, Palmas, TO (2020).

**Resumo:** A descrição e reprodução computacional de fenômenos é um desafio comum na área da Modelagem Computacional. No contexto da calorimetria de altas energias, a energia da partícula proveniente de colisões é absorvida e amostrada em forma de um sinal digitalizado. Em operação de alta taxa de eventos e luminosidade, o empilhamento de sinais pode ocorrer, distorcendo o sinal esperado. Desta forma, este trabalho avalia o desempenho de uma abordagem estocástica do método conhecido como Filtro Casado (FC). Para a análise de eficiência, dados simulados considerando diferentes condições de empilhamento de sinais no Calorímetro de Telhas (TileCal) do ATLAS no LHC foram utilizados. A eficiência do FC estocástico foi comparada com o método atualmente empregado no TileCal, conhecido como Filtro Ótimo (OF), para a estimação da energia. Os resultados mostram que o FC apresenta um desempenho superior ao método OF.

**Palavras-chave:** *Filtro Casado, Filtro Ótimo, Estimação de Sinais, Compactação de Dados, Calorimetria de Altas Energias.*