

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Thiago Campos Acácio Paschoalin

**Implementação Online de Representação Esparsa para Reconstrução de
Energia no Calorímetro Hadrônico do ATLAS**

Juiz de Fora

2026

Thiago Campos Acácio Paschoalin

**Implementação Online de Representação Esparsa para Reconstrução de
Energia no Calorímetro Hadrônico do ATLAS**

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

Orientador: Prof. Dr. Luciano Manhães de Andrade Filho

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Paschoalin, Thiago Campos Acácio.

Implementação Online de Representação Esparsa para Reconstrução de Energia no Calorímetro Hadrônico do ATLAS / Thiago Campos Acácio Paschoalin. – 2026.

101 f. : il.

Orientador: Luciano Manhães de Andrade Filho

Tese (Doutorado) – Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2026.

1. Calorimetria. 2. Física de altas energias. 3. Estimacão de Energia. 4. Representação Esparsa. 5. Rede Neural. I. Andrade Filho, Luciano Manhães de, orient. II. Implementação Online de Representação Esparsa para Reconstrução de Energia no Calorímetro Hadrônico do ATLAS.

Thiago Campos Acácio Paschoalin

Implementação Online de Representação Esparsa para Reconstrução de Energia no Calorímetro Hadrônico do ATLAS

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Sistemas Eletrônicos

Aprovada em 06 de março de 2026.

BANCA EXAMINADORA

Prof. Dr. Luciano Manhães de Andrade Filho - Orientador

Universidade Federal de Juiz de Fora

Prof. Dr. Rafael Antunes Nóbrega

Universidade Federal de Juiz de Fora

Prof. Dr. Leandro Rodrigues Manso Silva

Universidade Federal de Juiz de Fora

Prof. Dr. Jorge Luís Machado do Amaral

Universidade do Estado do Rio de Janeiro

Prof. Dr. Michel Pompeu Tcheou

Universidade do Estado do Rio de Janeiro

Juiz de Fora, 13/02/2026.



Documento assinado eletronicamente por **Leandro Rodrigues Manso Silva, Professor(a)**, em 06/03/2026, às 16:43, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Michel Pompeu Tcheou, Usuário Externo**, em 06/03/2026, às 16:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafael Antunes Nobrega, Professor(a)**, em 06/03/2026, às 16:47, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luciano Manhaes de Andrade Filho, Professor(a)**, em 09/03/2026, às 14:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jorge Luís Machado do Amaral, Usuário Externo**, em 09/03/2026, às 14:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2875804** e o código CRC **D82E3DEC**.

AGRADECIMENTOS

Agradeço à minha esposa, Luciana, por estar ao meu lado durante grande parte da minha vida, sempre me incentivando, ouvindo minhas lamentações, comemorando minhas vitórias, compreendendo os prazos necessários e cuidando com tanto carinho do nosso filho. Sem o seu suporte, este trabalho não seria possível.

Ao meu filho, Lucas, agradeço por existir em minha vida. Tudo o que faço tem você como motivação, e a oportunidade de desenvolver este trabalho em casa, compartilhando os dias com a sua presença, foi um dos maiores presentes deste período.

Agradeço aos meus pais, Heloisa e Domingos, por me proporcionarem uma educação de qualidade desde o início da minha infância, permitindo que eu pudesse me dedicar integralmente aos estudos durante boa parte da minha trajetória. Sei o quanto este trabalho lhes traz orgulho.

Agradeço ao meu irmão, Diogo, por celebrar minhas conquistas e contagiar com seu alto astral. Agradeço também aos meus amigos, que compartilharam momentos de lazer e companheirismo ao longo de toda essa caminhada.

Registro uma menção especial ao meu orientador, Luciano, que se tornou um amigo e companheiro de trabalho ao longo desses anos, e que me proporcionou uma experiência de intercâmbio internacional que jamais imaginei vivenciar. Agradeço também ao meu companheiro de profissão, Tiago, que conheci por meio do doutorado, e que contribuiu de forma decisiva em todo o processo, estabelecendo uma colaboração científica fundamental que certamente se estenderá por muitos anos.

À instituição CEFET-MG, onde atuo como professor, agradeço pela oportunidade de desenvolver este trabalho por meio de uma licença remunerada durante os últimos três anos do curso, bem como pelo apoio institucional e pelos recursos destinados à participação em eventos científicos. Agradeço ao meu departamento pela compreensão das minhas demandas e pela autorização do afastamento.

Agradeço ao CERN por proporcionar um ambiente de pesquisa extremamente rico, no qual pude desenvolver novas técnicas, interagir com pesquisadores de diferentes áreas, abrir novas frentes de pesquisa e conhecer outros países.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

À RENAFEA, agradeço pelo suporte à área de Física de Altas Energias no país. Por fim, agradeço ao povo brasileiro, que, de forma indireta, financia as pesquisas e o desenvolvimento científico e tecnológico nacional.

RESUMO

Nos experimentos de Física de Altas Energias, a reconstrução digital de sinais é uma etapa fundamental para a estimação precisa da energia das partículas produzidas em colisões. Esse problema pode ser formulado, do ponto de vista da Engenharia Elétrica, como uma tarefa de processamento de sinais discretos, envolvendo filtragem, deconvolução e estimação de parâmetros em ambientes fortemente ruidosos. No experimento ATLAS, do Grande Colisor de Hádrons (LHC), o calorímetro hadrônico TileCal utiliza eletrônica dedicada e técnicas de filtragem digital no sistema de *trigger* para realizar a seleção inicial de eventos em tempo real. Com a atualização do acelerador para o HL-LHC (*High Luminosity Large Hadron Collider*), o aumento significativo da taxa de eventos simultâneos intensifica o efeito de empilhamento (*pile-up*), degradando o desempenho dos métodos clássicos de reconstrução baseados em filtragem linear. Esse cenário motiva o estudo de novas abordagens de processamento de sinais capazes de operar sob restrições rigorosas de latência e complexidade computacional. A representação esparsa de sinais tem se mostrado promissora nesse contexto, pois explora a estrutura intrínseca dos sinais para melhorar a estimação em ambientes de alta interferência. Entretanto, sua aplicação prática em sistemas de tempo real é limitada pelo uso de algoritmos iterativos. Este trabalho propõe uma adaptação não iterativa desses métodos, baseada em um filtro de deconvolução seguido por uma rede neural *feed-forward*, responsável por aplicar uma função de *shrinkage* com limiar adaptativo definido amostra a amostra. O método proposto foi comparado com técnicas descritas na literatura e apresentou desempenho superior tanto em cenários de baixa quanto de alta ocupação, alcançando reduções de até 54% no erro de reconstrução. A abordagem mantém baixa complexidade computacional e apresenta uma estrutura compatível com processamento em tempo real, tornando-se adequada para implementação em sistemas digitais embarcados, como o sistema de *trigger* do TileCal.

Palavras-chave: calorimetria; física de altas energias; estimação de energia; representação esparsa; rede neural.

ABSTRACT

In High Energy Physics experiments, digital signal reconstruction is a fundamental step for the accurate estimation of the energy of particles produced in collisions. From an Electrical Engineering perspective, this problem can be formulated as a digital signal processing task involving filtering, deconvolution, and parameter estimation in highly noisy environments. In the ATLAS experiment at the Large Hadron Collider (LHC), the hadronic calorimeter TileCal employs dedicated electronics and digital filtering techniques in the trigger system to perform real-time event selection. With the accelerator upgrade to the HL-LHC (High Luminosity Large Hadron Collider), the significant increase in the rate of simultaneous events intensifies the pile-up effect, degrading the performance of classical reconstruction methods based on linear filtering. This scenario motivates the investigation of new signal processing approaches capable of operating under strict latency and computational complexity constraints. Sparse signal representation has shown promising results in this context, as it exploits the intrinsic structure of the signals to improve estimation performance in high-interference environments. However, its practical application in real-time systems is limited by the use of iterative algorithms. This work proposes a non-iterative adaptation of these methods, based on a deconvolution filter followed by a feed-forward neural network, which applies a shrinkage function with an adaptive threshold defined on a sample-by-sample basis. The proposed method was compared with techniques reported in the literature and demonstrated superior performance in both low- and high-occupancy scenarios, achieving reductions of up to 54% in reconstruction error. The approach maintains low computational complexity and features a structure compatible with real-time processing, making it suitable for implementation in embedded digital systems, such as the TileCal trigger system.

Keywords: calorimetry; high-energy physics; energy estimation; sparse representation; neural network.

LISTA DE ILUSTRAÇÕES

Figura 2.1 – Modelo Padrão de partículas fundamentais	20
Figura 2.2 – LHC, CERN e os detectores	21
Figura 2.3 – Plano de atualizações para o LHC	22
Figura 2.4 – Experimento ATLAS	24
Figura 2.5 – Experimento ATLAS	25
Figura 2.6 – Módulo do Tilecal	26
Figura 2.7 – Pulso padrão com 7 amostras	27
Figura 2.8 – Pulso padrão do Tilecal comparado com dados reais	28
Figura 2.9 – Cadeia do sistema de leitura do Tilecal para fase I	29
Figura 2.10–Cadeia do sistema de leitura do Tilecal para fase II	30
Figura 2.11–Diagrama de Blocos do sistema <i>trigger</i> e aquisição de dados da fase I	33
Figura 2.12–Fluxograma do sistema de <i>trigger</i> da fase I	34
Figura 2.13–Fluxograma do sistema de <i>trigger</i> da fase II	36
Figura 2.14–Gráfico do pulso do TileCal demonstrando o efeito de empilhamento do sinal	38
Figura 2.15–Distribuição simulada do ruído nas células do TileCal em função do número médio de interações por cruzamento de feixes	39
Figura 2.16–Esquema da simulação completa do ATLAS	40
Figura 2.17–Fluxograma da simulação TMC	44
Figura 3.1 – Fluxograma da simulação TMC	45
Figura 3.2 – Representação do modelo de um canal do TileCal e estrutura do filtro de estimação de energia.	47
Figura 3.3 – Estrutura do canal modelado como sistema LTI seguido de filtro estimador.	50
Figura 3.4 – Arquiteturas de filtragem: (a) Filtro FIR tradicional; (b) Filtro neural MLP com linha de atraso.	54
Figura 3.5 – Busca exaustiva da configuração ótima da RNA.	55
Figura 3.6 – Rede Neural com Resposta Finita ao Impulso (FIRNN).	55
Figura 3.7 – Comparação entre os coeficientes do filtro FIR e os de um neurônio da RNA.	56
Figura 3.8 – Mapeamento entrada-saída dos neurônios.	57
Figura 4.1 – Representação gráfica da função <i>Shrinkage</i> modificada para o caso de física de altas energias	63
Figura 4.2 – Representação gráfica da função <i>Shrinkage</i> com $\lambda = 2$ (linha contínua), em comparação com uma função identidade (linha pontilhada).	64
Figura 5.1 – Fluxograma detalhando o processo de montagem do banco de dados para treinamento da rede neural	72
Figura 5.2 – Diagrama de blocos resumindo o modelo final do método proposto	73

Figura 6.1 – Parcela de amostras dos dados simulados para uma ocupação de 20% .	75
Figura 6.2 – Relação entre RMSE e número de pesos do filtro FIR para ocupações de 20% e 80%	77
Figura 6.3 – Amostras estimadas pelo filtro FIR para uma ocupação de 20%	78
Figura 6.4 – Função de ativação ReLU	79
Figura 6.5 – Comparação entre o valor previsto e o valor desejado para uma ocupação de 20%	80
Figura 6.6 – Comparação entre o valor previsto e o valor desejado para uma ocupação de 80%	81
Figura 6.7 – RMS obtido por diferentes topologias de rede neural para uma ocupação de 20%	82
Figura 6.8 – RMS obtido por diferentes topologias de rede neural para uma ocupação de 80%	83
Figura 6.9 – Resultado da estimação do método proposto para uma ocupação de 20%	85
Figura 6.10–Resultado da estimação do LS para uma ocupação de 20%	86
Figura 6.11–Resultado da estimação SFRM-FT para uma ocupação de 20%	86
Figura 6.12–Resultado da estimação RNE para uma ocupação de 20%	87
Figura 6.13–Resultado da estimação do método proposto para uma ocupação de 80%	88
Figura 6.14–Comparação de performance dos métodos para diferentes ocupações . .	89
Figura 6.15–Comparação de performance entre o método RNE e o método proposto para todos os cenários	90
Figura 6.16–Comparação de performance entre o modelo genérico e o modelo especialista para o Método Proposto	91

LISTA DE TABELAS

Tabela 2.1 – Operações em cada rodada	24
Tabela 3.1 – Influência sináptica absoluta de cada neurônio.	56
Tabela 6.1 – Multiplicadores necessários em cada topologia de rede neural.	83
Tabela 6.2 – Recursos computacionais necessários para cada um dos métodos.	92

LISTA DE ABREVIATURAS E SIGLAS

ADC	Conversores Analógico-Digitais (<i>Analog-to-Digital Converters</i>)
LHC	Large Hadron Collider
CERN	Conseil Européen pour la Recherche Nucléaire
ATLAS	A Toroidal LHC ApparatuS
L1A	Level-1 Accept
HLT	High-Level Trigger
FIR	Finite Impulse Response (Resposta Finita ao Impulso)
MLP	Multi Layer Perceptron
TMC	Toy Monte Carlo
ALICE	A Large Ion Collider Experiment
LHCb	Large Hadron Collider beauty experiment
CMS	Compact Muon Solenoid
LS2	Long Shutdown 2
LS3	Long Shutdown 3
HL-LHC	High-Luminosity Large Hadron Collider
PMT	Photomultiplier Tube
WLS	Wavelength-Shifting
BC	Bunch Crossing
FE	Front-End
BE	Back-End
ROD	Read-Out Driver
TTC	Timing, Trigger and Control
FPGA	Field Programmable Gate Array
PU	Processing Unit
DSP	Digital Signal Processor
TilePPr	Tile PreProcessor
sROD	Super Read-Out Driver
TDAQi	Trigger and Data Acquisition Interface
FELIX	Front-End Link eXchange
TDAQ	Trigger and Data Acquisition
BCID	Bunch Crossing Identifier
ROB	Readout Buffer
FTK	Fast TrackK
RoI	Region of Interest
nMCM	New Multi-Chip Module
CTP	Central Trigger Processor
FTK++	Full Event Tracking
RoIE	Region of Interest Engine
MC	Monte Carlo
TileCal	Tile Calorimeter

OF	Optimal Filter
BLUE	Best Linear Unbiased Estimator
LTI	Linear Time-Invariant
IIR	Infinite Impulse Response (Resposta Infinita ao Impulso)
LMS	Least Mean Squares
RNA	Redes Neurais Artificiais
FTLNF	Focused Time Lagged Feedforward Network
FIRNN	Finite Impulse Response Neural Network
MP	Matching Pursuit
OMP	Orthogonal Matching Pursuit
RIP	Restricted Isometry Property
ISTA	Iterative Soft Thresholding Algorithm
LS	Least Squares (Mínimos Quadrados)
SFRM-DT	Sparse Free-Running Method with Dynamic Threshold
RNE	Rede Neural Estimadora
RMSE	Root Mean Square Error
ReLU	Rectified Linear Unit
MSE	Mean Square Error

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação	14
1.2	Objetivos	16
1.3	Metodologia e Atividades Realizadas	17
1.4	Organização do Texto	18
2	EXPERIMENTOS DE FÍSICA DE ALTAS ENERGIAS	20
2.1	O LHC	20
2.2	PLANO DE ATUALIZAÇÕES DO LHC	22
2.3	O EXPERIMENTO ATLAS	24
2.3.1	Calorímetro Hadrônico do ATLAS	25
2.3.1.1	Sistema de Leitura do Tilecal na fase I	28
2.3.1.2	Sistema de Leitura do Tilecal na fase II	30
2.3.2	Sistema de <i>Trigger</i> do Tilecal	31
2.3.2.1	Sistema do TDAQ na fase I	32
2.3.2.2	Sistema do TDAQ na fase II	35
2.4	EFEITO <i>PILE-UP</i>	37
2.4.1	<i>Pile-up</i> como fonte de ruído	37
2.4.2	Simulação do <i>Pile-up</i>	39
2.5	CONCLUSÕES PARCIAIS DO CAPÍTULO	44
3	ESTIMAÇÃO DE ENERGIA NO TILECAL	45
3.1	FILTROS DE RECONSTRUÇÃO DE ENERGIA NO SISTEMA DE LEITURA	46
3.1.1	Filtro Ótimo	47
3.1.2	Filtro de Wiener	49
3.2	FILTROS DE DECONVOLUÇÃO NO PRIMEIRO NÍVEL DE TRIGGER	50
3.2.1	Estimativa Inversa do Sistema de Leitura	51
3.2.1.1	Inversão no Domínio- z	51
3.2.1.2	Aproximação via Pseudo-Inversa de Moore-Penrose	52
3.2.2	Filtro Baseado em Mínimos Quadrados	52
3.2.3	Filtro Baseado em Redes Neurais Artificiais	53
3.2.3.1	Interpretação Alternativa: Filtros FIR em Redes Neurais	54
3.3	CONSIDERAÇÕES PARCIAIS	56
4	TEORIA DE ESPARSIDADE APLICADA À PROCESSAMENTO DE SINAIS	58
4.1	REPRESENTAÇÕES ESPARSAS	58
4.2	ESTIMATIVA DE ENERGIA VIA ESPARSIDADE	59
4.2.1	Método Iterativo	61

4.2.2	Método adaptado para processamento <i>online</i> dos sinais	62
4.3	CONCLUSÕES PARCIAIS DO CAPÍTULO	65
5	MÉTODO PROPOSTO	66
5.1	EXPANDINDO O PROCESSO ITERATIVO ISTA	66
5.2	O FILTRO FIR COMO VALOR INICIAL DO PROCESSO	69
5.3	REDE NEURAL PARA DETERMINAÇÃO DO LIMAR DA FUNÇÃO <i>SHRINKAGE</i>	70
5.3.1	Processo de treinamento da Rede Neural	71
5.4	VISÃO GERAL DO MÉTODO PROPOSTO	73
5.5	COMPARAÇÃO COM OUTROS MODELOS DE ESTIMAÇÃO	73
5.6	CONCLUSÕES PARCIAIS DO CAPÍTULO	74
6	RESULTADOS	75
6.1	AVALIAÇÃO DO FILTRO FIR DE DECONVOLUÇÃO	76
6.2	ANÁLISE DA PREDIÇÃO DO VALOR DE LIMAR DA SHRINKAGE	78
6.3	COMPARAÇÃO DE TOPOLOGIAS DA REDE NEURAL NOS RESULTA- DOS	81
6.4	RESULTADOS DE ESTIMAÇÃO DO MÉTODO PROPOSTO	84
6.5	ANÁLISE DE UM MODELO GERAL PARA TODAS AS OCUPAÇÕES	89
6.6	COMPARAÇÃO DE RECURSOS COMPUTACIONAIS NECESSÁRIOS	91
6.7	CONCLUSÕES PARCIAIS DO CAPÍTULO	92
7	CONCLUSÃO	93
7.1	TRABALHOS FUTUROS	94
	REFERÊNCIAS	95
	APÊNDICE A – Produção Bibliográfica	100

1 INTRODUÇÃO

O avanço em sistemas de aquisição de dados e análise de sinais têm sido impulsionado pela crescente demanda por medições confiáveis de variáveis físicas em diversas áreas, como controle industrial, aplicações médicas, gestão energética, domótica e pesquisa científica (Iakovidis et al., 2021). Em tais sistemas, variáveis como temperatura, pressão, tensão elétrica, vazão ou luminosidade precisam ser convertidas em dados digitais, para posterior interpretação e tomada de decisão.

Essa cadeia de medição, nos sistemas de instrumentação modernos, segue um fluxo composto por três etapas principais: a transdução da grandeza física em sinal elétrico, o condicionamento desse sinal — ajustando amplitude e faixa de frequência — e, finalmente, sua digitalização por conversores analógico-digitais (ADC). Após essas etapas, os dados digitalizados são processados para extrair os parâmetros que representam o fenômeno observado.

Embora esse fluxo seja comum em diversas aplicações, desafios mais complexos surgem em cenários onde a quantidade de canais de aquisição é elevada e a taxa de amostragem é extremamente alta. Nestes casos, além do volume de dados ser substancial, pode ocorrer a presença de sinais sobrepostos, exigindo estratégias sofisticadas para seleção e processamento de amostras.

1.1 Motivação

Entre os ambientes mais exigentes para sistemas de aquisição estão os experimentos de física de altas energias, que operam sob condições extremas de taxa de eventos e número de sensores. Esses cenários impõem a necessidade de soluções robustas, rápidas e escaláveis para o tratamento e filtragem dos dados adquiridos.

Um exemplo notável dessa realidade é o Grande Colisor de Hádrons (LHC, do inglês *Large Hadron Collider*), instalado no CERN (*Conseil Européen pour la Recherche Nucléaire*), cujo objetivo é investigar os constituintes fundamentais da matéria. Ao promover colisões entre feixes de prótons a uma taxa de 40 MHz — ou seja, 40 milhões de interações por segundo — o LHC gera uma quantidade massiva de dados experimentais que requerem sistemas extremamente eficientes para seleção e análise.

O experimento ATLAS (*A Toroidal LHC ApparatuS*) é um dos detectores instalados nas regiões de colisão do LHC. Seu sistema de calorimetria desempenha papel essencial na identificação de partículas subatômicas, uma vez que registra a distribuição espacial da energia depositada por essas partículas ao atravessarem diferentes camadas do detector. A análise dessa deposição contribui na identificação de partículas e no seu comportamento ao longo do processo de decaimento.

Os sinais elétricos provenientes do calorímetro são moldados por circuitos de condicionamento para assumir uma forma padrão — conhecida como pulso característico — cujo pico de amplitude está diretamente relacionado à energia depositada. Esse pulso tem uma duração que se estende por várias colisões subjacentes. No sistema em questão, adota-se uma janela de sete amostras sincronizadas ao clock do LHC (período de 25 ns), totalizando 150 ns por pulso.

Essa escolha de janela visa capturar adequadamente o formato do pulso e maximizar a precisão na reconstrução da energia. Porém, o aumento na taxa de incidência de pulsos também amplia a possibilidade de sobreposição de sinais de eventos distintos — fenômeno intensificado com o aumento da luminosidade do colisor. Esse efeito, conhecido como *pile-up*, compromete os pressupostos dos métodos clássicos de estimação, que assumem pulsos isolados imersos em ruído gaussiano.

Com a elevação da luminosidade, torna-se frequente a sobreposição de múltiplos pulsos dentro de uma mesma janela de aquisição. Nessas circunstâncias, as amostras vizinhas ao pico do pulso deixam de apresentar apenas ruído branco, passando a carregar informações correlacionadas com sinais laterais — o que introduz correlações de ordem superior. Métodos tradicionais, de natureza linear, tornam-se ineficientes nessas condições.

Para uma seleção cada vez mais precisa, o sistema de aquisição do ATLAS emprega um esquema de seleção baseado em múltiplos níveis de *trigger*. O primeiro nível — *Level-1 Trigger* — utiliza padrões de energia para selecionar colisões promissoras, emitindo um sinal L1A (*Level-1 Acceptance*) que habilita a coleta de janelas de amostras em cada canal. Em seguida, o *High-Level Trigger* (HLT) realiza uma filtragem mais refinada nesta janela, decidindo quais eventos devem ser armazenados para análise posterior.

Uma vez ativado o L1A, uma janela de sete amostras consecutivas é extraída em cada canal, com o objetivo de estimar a amplitude do pulso central, localizado na quarta amostra. Porém, na presença de *pile-up*, essa tarefa se torna desafiadora, pois a premissa de um único pulso sobre ruído aditivo deixa de ser válida.

Estudos referentes à métodos que processam melhor sinais contendo *pile-up* estão sendo conduzidos (Duarte, 2019; Quirino, 2022). No que diz respeito ao primeiro nível, essas técnicas devem considerar o processamento de sinais em tempo real, acrescentando um desafio à tarefa exercida.

Desenvolvimento de métodos baseados na teoria da esparsidade está avançando nos últimos anos, em especial na área de processamento de imagem. Uma técnica moderna na obtenção da representação esparsa está relacionada com o uso da função limiar, denominada de função *shrinkage*. Essa função não-linear auxilia o processo de reconstrução do sinal na filtragem de ruídos que prejudicam a performance dos estimadores.

Em sinais provenientes de calorímetros, essa teoria pode ser aplicada, em especial em processamento *offline*. Mas modificações das técnicas já desenvolvidas podem ser feitas

para que seja possível a sua utilização em sistemas *online* de processamento, sendo esse o objetivo principal deste trabalho.

Duarte (2019) propôs uma abordagem baseada na teoria de sinais esparsos para aplicação em ambientes *free-running*, nos quais o processamento dos sinais deve ser realizado de forma rápida e em altas frequências de operação. Entretanto, uma limitação desse método está associada ao uso da função *shrinkage* com um limiar fixo, aplicado indistintamente a todas as situações de operação. Por sua vez, Quirino (2022) explora a teoria da esparsidade em conjunto com o teorema de Bussgang para realizar o treinamento *online* de redes neurais artificiais.

O presente trabalho segue a linha de investigação proposta em Duarte (2019), buscando avançar na definição de limiares adaptativos associados a cada valor estimado de energia, com o objetivo de melhorar o processo de reconstrução do sinal. A abordagem proposta baseia-se em um procedimento iterativo capaz de ajustar um modelo específico, adequado à implementação em sistemas de processamento em tempo real das amostras.

1.2 Objetivos

Diante do cenário de aumento de luminosidade e elevado nível de *pile-up*, propõe-se o objetivo principal do trabalho:

- Desenvolvimento de um método baseado em representações esparsas que seja capaz de operar no sistema de filtragem *online* das colisões

Como objetivos específicos, é possível definir:

- Estudo aprofundado da teoria de esparsidade aplicada a processamento de sinais - O aprofundamento no tema permite identificar pontos importantes que levarão ao desenvolvimento do método desejado.
- Análise do papel da função *shrinkage* aplicada em representações esparsas - esse tipo de função é de grande importância para os métodos já desenvolvidos, e merece uma atenção especial nos estudos.
- Desenvolvimento de um método baseado em um método iterativo para utilização em processamento em tempo real - O método iterativo não é possível de ser aplicado em um ambiente *online*. Dessa forma, a expansão deste método permite uma adaptação onde o processo iterativo é contornado.
- Comparação do método proposto com técnicas atuais de estimação de energia baseadas em teoria da esparsidade em ambientes *online*, rede neural e método linear - A comparação com ferramentas que já se mostraram compatíveis com o ambiente citado é essencial para determinar a relevância do método proposto.

- Testes no cenário de altas luminosidades do LHC através de simulação - Como se propõe um método a ser aplicado em ambientes de altas taxas de colisões, é importante verificar sua performance nesse cenário. A simulação se torna importante pois não existem dados reais com essas características.
- Encontrar um modelo genérico aplicável a todas as células - considerando que o calorímetro do ATLAS possui milhares de canais, o treinamento específico para cada célula não é uma tarefa trivial. Portanto, a definição de um modelo com um conjunto fixo de pesos capaz de atender a todas as células é um aspecto fundamental do ponto de vista operacional.

1.3 Metodologia e Atividades Realizadas

O método proposto baseia-se numa estimativa inicial obtida através de um filtro FIR (*Finite Impulse Response*) de deconvolução e na aplicação da função *shrinkage* no resultado deste filtro. Através do processo iterativo no qual a técnica foi baseada, percebeu-se que o valor de limiar da função *shrinkage* era diferente dependendo das amostras vizinhas da leitura do calorímetro e estimação inicial da energia, ou seja, possui um valor dinâmico. Essa relação entre o limiar da referida função, a leitura do calorímetro e a estimação inicial da energia pode ser mapeada através de uma rede neural de múltiplas camadas (MLP, do inglês, *Multi Layer Perceptron*), que prevê o valor a ser utilizado na limiarização da função *shrinkage* de acordo com as amostras vizinhas. Dessa forma, a estimativa inicial do filtro FIR é corrigida por essa importante função utilizada na teoria de esparsidade.

Inicialmente, foram estudados métodos baseados na teoria da esparsidade, tanto para aplicação em processamento *online* e *offline*. Os modelos *offline* apresentavam respostas muito interessantes, mas que não eram possíveis em um ambiente tão desafiador quanto o sistema de trigger L1A. Já o modelo *online* sim era aplicável ao sistema em questão, mas possuía algumas limitações que prejudicavam o processo de estimação. Esse estudo está melhor conduzido no Capítulo 4.

O objetivo principal então foi modificar o modelo aplicável a sistemas *offline* a fim de torná-lo factível no sistema de *trigger*, aprimorando a estimação da energia. O processo iterativo foi avançado em passos posteriores, no qual foi possível realizar algumas trocas de variáveis, descobrindo que a função de limiarização da teoria da esparsidade apresenta valores dinâmicos dependentes de amostras vizinhas do calorímetro. Por mais que essa relação seja uma função não linear que não existe uma fórmula fechada, uma rede neural pode prever o valor a ser utilizado pela função *shrinkage*.

O próximo passo foi desenvolver a estrutura da rede neural que faria essa tarefa descrita. Com isso, foram testadas várias topologias que poderiam ser adequadas para esta situação. Todas estavam associadas ao treinamento do valor de limiar. Esse valor,

por sua natureza e dependendo da situação, pode apresentar infinitos valores para uma mesma amostra analisada. Portanto, foi necessário desenvolver uma metodologia para montar o banco de dados para o treinamento da rede neural.

Com a metodologia da montagem do banco de dados definida, utilizou-se da ferramenta de simulação *Toy Monte Carlo* (TMC) para gerar os dados para esta rede neural. O banco de dados foi construído a partir da análise da estimativa inicial da amplitude e do valor real desejado, em conjunto com a aplicação da função *shrinkage*. Modificou-se a quantidade de neurônios das camadas envolvidas, as funções de ativação até encontrar uma que realizasse a tarefa com o menor uso de recursos computacionais possível.

Por fim, criou-se o modelo final do método proposto que combinou uma estimativa inicial dada por um filtro FIR de deconvolução e uma função *shrinkage* com valor de limiar dado por uma rede neural MLP. Os testes de desempenho foram levantados para este modelo final e comparados com outros métodos, providenciando resultados interessantes. O desempenho observado é resultado do método de treinamento adotado, que se apoia na interpretação da rede neural como um mecanismo de controle de limiar fundamentado nos conceitos da teoria da esparsidade.

1.4 Organização do Texto

O capítulo 2 apresenta alguns conceitos de experimentação física de altas energias importantes e disserta sobre o LHC. Nele é apresentado o calorímetro hadrônico TileCal, bem como conceitos de instrumentação nuclear presentes neste detector, em especial, o sistema de *trigger*. É comentada a atualização que o LHC passará e os desafios gerados por esse novo cenário, apresentando o problema do *pile-up*.

No capítulo 3 é realizada uma revisão bibliográfica das técnicas de estimação de energia já desenvolvidas, em especial sobre as relacionadas ao tema e que faziam sentido serem comparadas com o método proposto.

A teoria da esparsidade aplicada a processamento de sinais é apresentada no capítulo 4. São discutidos conceitos básicos, além de técnicas que podem ser aplicadas tanto em um ambiente de processamento *online* quanto *offline*. São assuntos importantes para a compreensão do método proposto.

No capítulo 5 é desenvolvida a técnica proposta por esse trabalho. É desenvolvida a matemática por trás do método e a criação de um modelo derivado da formulação final.

Os resultados são compartilhados no capítulo 6, onde se discute as etapas parciais do método proposto e o seu resultado final de estimação.

As conclusões e os trabalhos futuros são apresentados no Capítulo 7, no qual são discutidos os resultados obtidos pelo método proposto, bem como destacadas observações

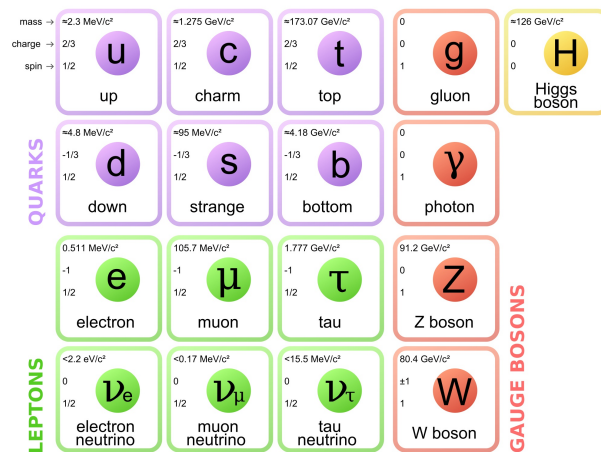
relevantes que podem ser extraídas a partir desses resultados.

No Apêndice A são expostas produções acadêmicas relevantes no período do doutorado trabalhado.

2 EXPERIMENTOS DE FÍSICA DE ALTAS ENERGIAS

Os cientistas buscam sempre entender melhor as leis que regem o universo. Isso ajuda a entender a origem e prever eventos futuros. Neste cenário, entender como a matéria é formada é um dos passos importantes em direção ao objetivo. Inicialmente, propôs-se o átomo como uma partícula fundamental, mas que posteriormente se descobriu que poderia ser dividido em prótons, elétrons e nêutrons. Com o avanço dos estudos, novas subdivisões puderam ser propostas, onde se incluem os quarks, léptons e os bósons (Wigmans, 2017). O modelo padrão de partículas define as partículas fundamentais e está apresentado pela Figura 2.1.

Figura 2.1 – Modelo Padrão de partículas fundamentais



Fonte: (Gagnon, 2016)

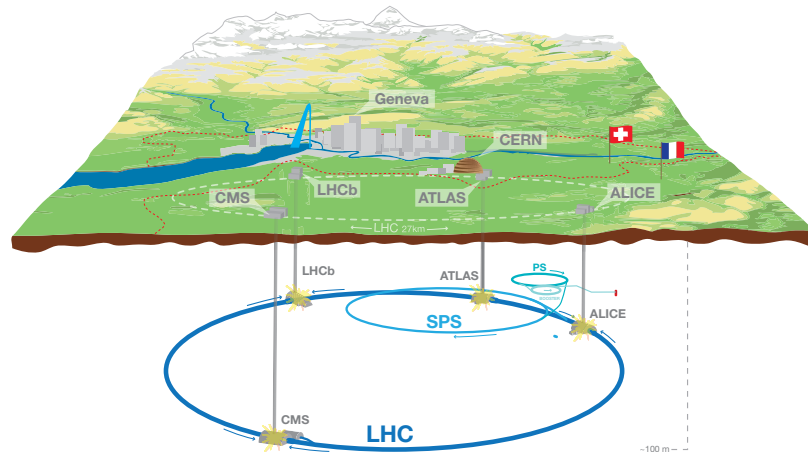
Um importante passo é a validação dos modelos teóricos através de experimentos físicos e os colisionadores de partículas cumprem um papel fundamental para os estudos do modelo padrão. Partículas estáveis (como elétrons ou prótons) são aceleradas em sentidos opostos até se colidirem, formando novas partículas que serão detectadas e medidas através de sistemas de instrumentação eletrônica (Andrade Filho, 2009). O LHC é atualmente o principal colisionador de partículas, e um marco interessante foi a observação do Bóson de Higgs (CERN, 2013). Essa partícula teorizada em 1964 é responsável por gerar a massa de outras partículas.

2.1 O LHC

O LHC é um acelerador de partículas construído entre as fronteiras da França e da Suíça. Este é considerado o maior colisionador em termos de dimensões físicas e o mais energético do mundo. O LHC acelera prótons que são forçados a se colidirem, gerando outras partículas fundamentais. O seu nome é inspirado no tipo de partícula dos prótons, que são considerados hádrons. Em 2008, este colisionador foi inaugurado no CERN, que é

uma organização internacional com colaboradores espalhados por todo o planeta e que tem como principal objetivo o estudo da matéria. O LHC, a sede do CERN e os detectores associados estão representados na Figura 2.2.

Figura 2.2 – LHC, CERN e os detectores



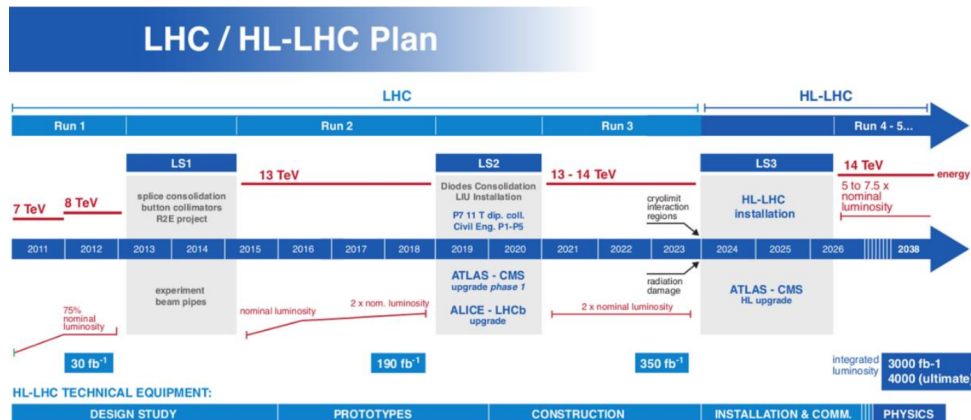
Fonte: (Service graphique, 2014)

O LHC é formado por um anel circular com 27 km de circunferência onde conjuntos de prótons formam feixes que são acelerados em sentidos opostos, cada um atingindo energias de $6,5 \text{ TeV}$. O encontro entre os dois feixes se dá em posições determinadas em momentos específicos, gerando uma colisão com energia total de 13 TeV , valor propositalmente alto para aumentar a probabilidade de gerar as partículas desejadas (Ferreira, 2009; ATLAS Collaboration, 2017a).

Os pontos de colisão mencionados são os detectores que estão construídos ao longo do anel circular do LHC. Ao todo, o colisionador é formado por quatro experimentos, cada um com sua função determinada. Esses são: ALICE (*A Large Ion Collider Experiment*) - construído para o estudo de íons pesados; o LHCb (*Large Hadron Collider beauty*) - desenvolvido para analisar o desbalanceamento entre matéria e antimatéria; CMS (*Compact Muon Solenoid*) e ATLAS (*A Toroidal LHC ApparatuS*) - detectores de propósito geral (Iiyama, 2017).

A estrutura circular do LHC é constituída de eletroímãs que proporcionam uma trajetória circular aos feixes de prótons e também aumentam a densidade dessas partículas no feixe, considerando sua área de seção transversal. Com o aumento da densidade, a probabilidade de ocorrer eventos de interesse durante a colisão entre os feixes é maior.

Figura 2.3 – Plano de atualizações para o LHC



Fonte: (ATLAS Collaboration, 2017a)

Essa probabilidade pode ser definida como luminosidade (\mathcal{L}), e é um importante fator para analisar o desempenho de colisionadores de partículas (Aberle et al., 2020). A luminosidade é determinada por:

$$\mathcal{L} \approx \frac{n_{bunch}^2}{\Delta t_c \cdot S_c} \quad (2.1)$$

onde n_{bunch} é a quantidade de prótons nos dois feixes opostos, Δt_c é o tempo da colisão e S_c é a área da seção transversal da colisão (Wiedemann, 2015).

Outro parâmetro importante relacionado à luminosidade é a taxa média de colisões por cruzamento (μ), que define a média probabilística da taxa de ocorrência de colisões. A Equação (2.2) demonstra a relação entre esse parâmetro e a luminosidade, sendo S_{inel} a área da seção transversal não elástica próton-próton e f_r a frequência de colisão do LHC. Esses dois últimos fatores possuem os valores de 71,5 mb e 40 MHz, respectivamente (Novakova, 2014).

$$\mu = \frac{\mathcal{L} \cdot S_{inel}}{n_{bunch} \cdot f_r} \quad (2.2)$$

O LHC possui um plano de atualizações, descrito na próxima seção, onde mudanças substanciais acontecerão em seu funcionamento, elevando os valores de luminosidade suportados nos experimentos. Como é de se esperar, uma série de componentes e instrumentos do colisionador deverá ser revista para atingir os valores desejados de energia.

2.2 PLANO DE ATUALIZAÇÕES DO LHC

Esta seção abordará as atualizações previstas para o LHC, em especial temas que envolvem o cenário que este trabalho está inserido. Um diagrama com um planejamento geral das atualizações dos experimentos é apresentado na Figura 2.3.

Nos primeiros testes com o LHC, realizados em 2011 e 2012, o equipamento ainda operava abaixo de seu potencial máximo, tanto em energia quanto em luminosidade. Durante a fase inicial de experimentação, conhecida como *Run 1*, as colisões atingiram energias de 7 a 8 TeV , com apenas 75% da luminosidade instantânea ideal sendo utilizada. Em números, isso representa uma luminosidade de cerca de $0,75 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, o que gerou, por cruzamento de feixes, uma média de até 12 colisões (Pastore, 2016).

Entre 2013 e 2015, o LHC passou por sua primeira parada técnica extensa (*LS1 - Long Shutdown*), período em que foram realizadas atualizações importantes nos sistemas do acelerador. Após esse intervalo, teve início uma nova fase experimental, conhecida como *Run 2*, que se estendeu de meados de 2015 até o fim de 2018. Nessa etapa, as colisões de partículas foram conduzidas com energia de 13 TeV . Ao longo desse período, a intensidade dos feixes foi sendo elevada gradualmente, até superar a luminosidade originalmente planejada, atingindo aproximadamente o dobro do valor de projeto, em torno de $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Com isso, a frequência de colisões por cruzamento de feixes variou, em média, entre 25 e 50 (Pastore, 2016).

Com foco na modernização dos principais detectores — ALICE, LHCb, CMS e ATLAS —, uma nova fase do LHC teve início em 2019, quando foi realizada a segunda parada técnica prolongada (*Long Shutdown 2 - LS2*). Essa etapa marca o início da Fase I do projeto. Já os experimentos da terceira rodada, chamados *Run 3*, estão ocorrendo desde o ano de 2022, utilizando feixes com energia de até 13,6 TeV . Mesmo mantendo a luminosidade instantânea em $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, o desempenho aprimorado do acelerador permite que a taxa média de colisões por cruzamento se situe entre 55 e 80 eventos — um avanço significativo em relação às fases anteriores (Pastore, 2016).

A partir de 2026, está programada uma nova fase de reestruturação no LHC, conhecida como *LS3 (Long Shutdown 3)*. Diferente das paradas anteriores, o foco agora é preparar o acelerador para operar em um novo patamar de desempenho, inaugurando o que se convencionou chamar de Fase II. Essa decisão estratégica do CERN busca explorar ao máximo as capacidades do colisor, elevando significativamente sua luminosidade. Para isso, será implementada uma série de melhorias que transformarão o atual sistema no HL-LHC (*High Luminosity Large Hadron Collider*) — uma versão de alta luminosidade capaz de produzir um volume muito maior de interações por segundo. Estima-se que, nesse novo estágio, a luminosidade atinja valores próximos de $7,5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, o que poderá gerar entre 140 e 200 colisões por cruzamento de feixes em diferentes momentos da operação. Esse aumento substancial no número de eventos tem como principal objetivo ampliar as chances de observar fenômenos raros, ainda pouco acessíveis nas rodadas anteriores (Brüning and Zerlauth, 2024).

De modo a resumir a operação do LHC nas rodadas citadas, a tabela 2.1 demonstra os importantes parâmetros em cada uma delas.

Tabela 2.1 – Operações em cada rodada

Rodada	\sqrt{s}	Luminosidade Máxima	Colisões por Cruzamento
Run 1	8 TeV	$0,75 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$	0 – 12
Run 2	13 TeV	$2 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$	25 – 50
Run 3	13–14 TeV	$2,5 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$	55 – 80
Run 4+	14 TeV	$5 \cdot 10^{34} - 7,5 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$	140 – 200

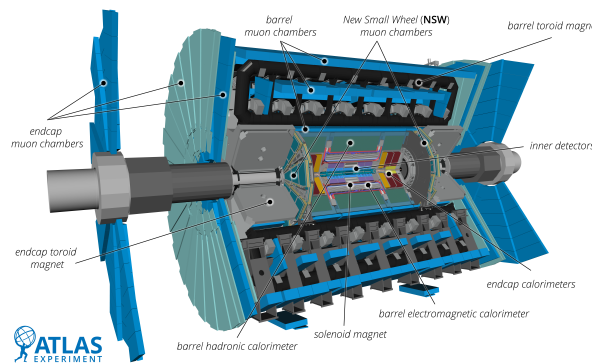
Fonte: (ATLAS Collaboration, 2017a)

2.3 O EXPERIMENTO ATLAS

Este trabalho utilizará os dados processados pelo experimento ATLAS. Desta forma, serão apresentados mais detalhadamente os sistemas que estão associados à aquisição e processamento dos sinais. Problemas futuros provenientes da nova operação do ATLAS serão abordados.

Com 44 metros de comprimento e 25 metros de diâmetro, o ATLAS é o detector de maiores proporções entre os quatro principais do LHC, conforme representado na Figura 2.4. Seu projeto foi concebido para abranger uma ampla gama de estudos físicos relacionados às colisões de partículas. Entre seus objetivos, destacam-se a análise detalhada dos produtos das colisões e seus respectivos decaimentos, permitindo investigar propriedades como massa, spin, paridade e interações do bóson de Higgs. Além disso, o ATLAS busca evidências de partículas supersimétricas, possíveis constituintes da matéria escura, fenômenos associados à escala eletrofraca, cenários alternativos com assinaturas genéricas, manifestações de física além do Modelo Padrão, bem como a existência de novos bósons (Aberle et al., 2020).

Figura 2.4 – Experimento ATLAS

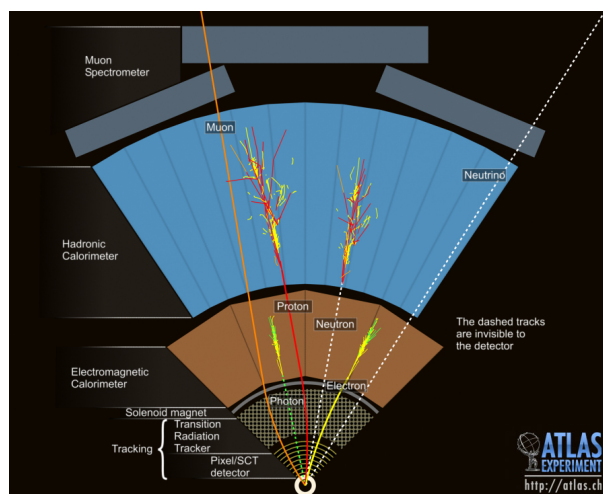


Fonte: (Bianchi and Collaboration, 2022)

A arquitetura do detector ATLAS é composta por uma série de sistemas distribuídos em camadas cilíndricas concêntricas, como evidenciado na vista em corte da Figura 2.5. Esses sistemas incluem: um módulo de rastreamento interno de partículas, chamado *Inner Detector*; o calorímetro eletromagnético, que utiliza argônio líquido como meio ativo (LAr - *LiquiArgon*); o calorímetro hadrônico, conhecido como Tilecal (*Tile Calorimeter*),

baseado em módulos de absorção segmentados; e o espectrômetro de múons, responsável por identificar e medir essas partículas de alta penetração. Cada camada desempenha um papel específico na reconstrução dos eventos gerados pelas colisões (ATLAS Collaboration, 2008).

Figura 2.5 – Experimento ATLAS



Fonte: (Pequenao, 2008)

Entre os diversos módulos que compõem o ATLAS, este estudo direciona sua atenção ao Tilecal, o calorímetro hadrônico, com ênfase no modo como seus dados são processados. Assim, aprofundaremos apenas neste subsistema.

2.3.1 Calorímetro Hadrônico do ATLAS

Em experimentos de altas energias, como os realizados no LHC, dispositivos chamados calorímetros são essenciais para investigar as partículas geradas. Eles funcionam capturando essas partículas e medindo a energia que transportam. Devido às energias extremamente elevadas envolvidas — chegando a 14 TeV —, os calorímetros operam com excelente precisão, tornando-se ferramentas indispensáveis para analisar os produtos das colisões (GALLART, 1998).

Calorímetros utilizam blocos de materiais densos, cuja espessura é projetada para garantir que a energia de uma partícula incidente seja completamente absorvida. À medida que a partícula interage com o material, ocorre a formação de um chuveiro de partículas secundárias, processo ilustrado na Figura 2.5. Um dos diferenciais desses detectores é a possibilidade de dividi-los em múltiplas camadas, o que permite mapear com precisão o padrão espacial da energia depositada. Essa segmentação fornece informações importantes para a identificação da partícula responsável pela interação (Ferbel, 1987).

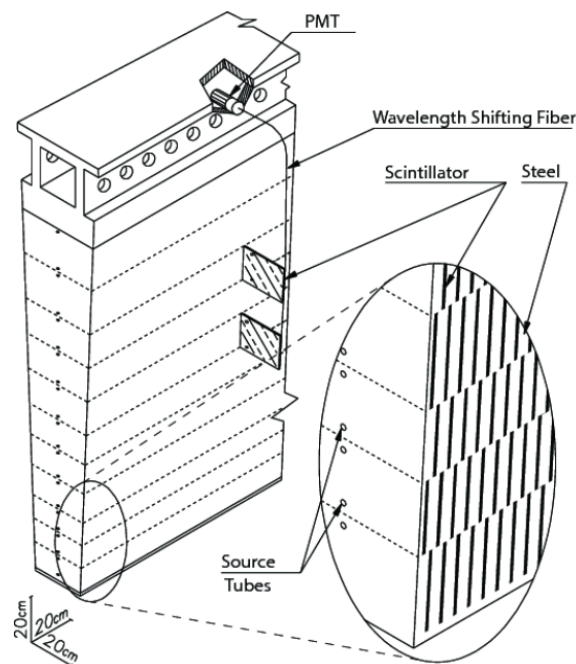
No ATLAS, os calorímetros utilizam uma configuração em camadas conhecida como sistema de amostragem. Nessa abordagem, alternam-se materiais densos, responsáveis por

absorver a maior parte da energia das partículas, com camadas sensíveis encarregadas de registrar os sinais gerados durante essa interação. Embora a maior parte da energia se perca como calor no material absorvedor, a porção captada pelo componente ativo é suficiente para estimar com precisão a energia da partícula que deu origem ao evento (GALLART, 1998).

O Tilecal é o subsistema do ATLAS projetado para detectar hádrons gerados nas colisões. Sua operação baseia-se em uma técnica de amostragem óptica, onde a energia das partículas é convertida em luz. Para isso, o detector é construído com camadas alternadas de aço — que atuam como material absorvedor — e blocos de plástico cintilante, que funcionam como meio ativo. A luz produzida nesses cintiladores é então captada por tubos fotomultiplicadores (PMTs - *Photomultipliers*), responsáveis por converter os sinais luminosos em pulsos elétricos mensuráveis (GALLART, 1998).

Graças ao seu desenho modular e repetitivo, o Tilecal foi concebido a partir da montagem de unidades menores, o que permitiu a construção de um detector de grande porte, como ilustrado na Figura 2.6 (GALLART, 1998). Cada módulo incorpora diversos cintiladores, cuja luz gerada é coletada por fibras ópticas especiais — conhecidas como WLS (*wavelength-shifting*) — que redirecionam o sinal luminoso para células de leitura. Essas células são formadas por grupos de cintiladores conectados a um mesmo canal de detecção. O número total de células foi dimensionado não apenas para garantir eficiência na captação da luz, mas também para oferecer redundância operacional, essencial frente à longa duração esperada do experimento ATLAS (ATLAS Collaboration, 2017a).

Figura 2.6 – Módulo do Tilecal

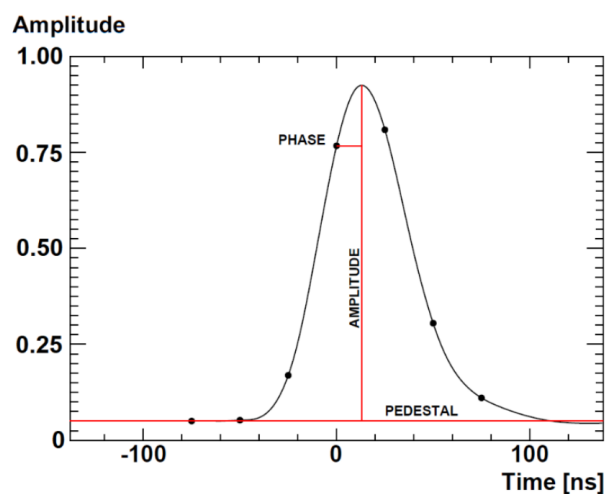


Fonte: (ATLAS Collaboration, 2017a)

O sistema de leitura óptica do Tilecal conta com cerca de 10.000 canais independentes, responsáveis por captar a luz emitida nos cintiladores e convertê-la em sinais elétricos (Mlynarikova, 2017). Cada um desses canais é equipado com um circuito específico para realizar essa conversão de forma precisa.

A parte eletrônica envolvida nesse processo será explorada com mais detalhes adiante. Esses circuitos não apenas detectam e amplificam os sinais luminosos, como também os digitalizam, gerando pulsos cuja amplitude está diretamente relacionada à energia depositada no detector — como exemplificado na Figura 2.7 (Paschoalin, 2016).

Figura 2.7 – Pulso padrão com 7 amostras



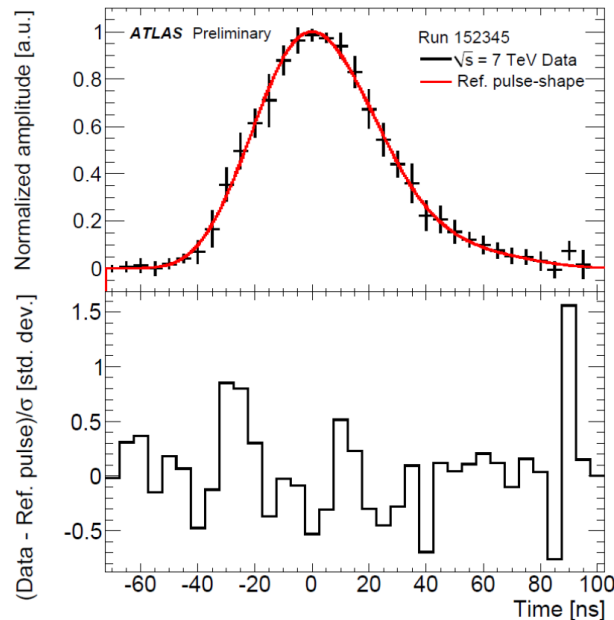
Fonte: (Sánchez, 2010)

O Tilecal possui um circuito conformador que gera um pulso elétrico contínuo, como apresentado pela linha preta da Figura 2.7. O sinal antes do circuito citado é um sinal impulsivo com amplitude proporcional à luz depositada (representada em vermelho), onde o papel do detector é reconstruir da forma mais fidedigna possível essa energia original. O processamento do pulso é realizado através de sistemas digitais, que necessitam que o mesmo seja amostrado a cada 25 ns , formando uma janela de leitura com 7 amostras (150 ns). Importante frisar que o sistema de amostragem é síncrono às colisões (BC - *Bunch Crossings*), de forma que a posição temporal do pulso dentro da janela de amostragem pode variar. Como consequência, o pico do pulso pode não coincidir exatamente com uma das amostras digitalizadas, exigindo o uso de algoritmos de reconstrução capazes de estimar corretamente a amplitude do sinal a partir do conjunto discreto de amostras disponíveis.

Para garantir a confiabilidade na medição dos sinais, o TileCal emprega diferentes estratégias de calibração e verificação contínua dos seus canais ópticos. Esse processo assegura que as formas de pulso geradas em cada canal se mantenham praticamente idênticas ao longo do tempo, sem variações relevantes entre elas. A Figura 2.8 ilustra essa

uniformidade, mostrando no topo a comparação entre um pulso padrão e a resposta de um canal real, enquanto a parte inferior apresenta a distribuição do erro relativo entre ambos (Peralva, 2015).

Figura 2.8 – Pulso padrão do Tilecal comparado com dados reais



Fonte: (Peralva, 2015)

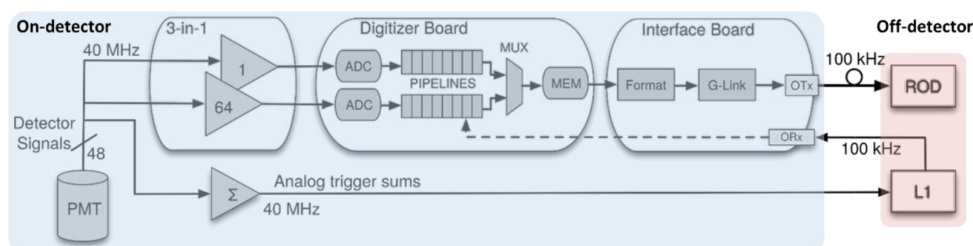
Entender o funcionamento do circuito de leitura atual e suas atualizações planejadas para a próxima fase de operação é importante para o desenvolvimento de novas técnicas de reconstrução do sinal que se comportem melhor em cenários existentes no Tilecal.

2.3.1.1 Sistema de Leitura do Tilecal na fase I

A infraestrutura eletrônica do TileCal é segmentada em dois níveis distintos: o *front-end* (FE), diretamente acoplado ao detector, e o *back-end* (BE), posicionado externamente. No projeto atual (ver Figura 2.9), o *front-end* é encarregado de processar os sinais provenientes das colisões com uma taxa extremamente alta, operando a 40 MHz — a mesma frequência dos cruzamentos de feixes no LHC. Por outro lado, o sistema de *back-end* atua em uma etapa posterior, tratando apenas os eventos selecionados. Esse processamento ocorre a uma taxa muito mais baixa, de aproximadamente 100 kHz , valor que reflete a taxa média de aceitação do primeiro nível de disparo, ativado por meio do comando L1A (Peralva et al., 2010).

O ponto inicial da cadeia de leitura na eletrônica de FE do TileCal está nos fotomultiplicadores (PMTs), cuja função é converter a luz oriunda dos cintiladores em sinais elétricos. Esses sinais seguem para um componente fundamental: a placa conhecida como *3-in-1*. Ela desempenha múltiplas tarefas, como moldar a forma do pulso para se

Figura 2.9 – Cadeia do sistema de leitura do Tilecal para fase I



Fonte: (Yue, 2019)

alinhar a um padrão de referência (vide Figura 2.7), aplicar diferentes níveis de amplificação — através de dois canais com ganhos distintos — e ainda combinar os sinais de forma analógica. A saída desse processo gera sinais compostos, que são organizados em estruturas chamadas torres de trigger. Esses sinais são então direcionados ao primeiro estágio do sistema de *trigger*, que será descrito mais à frente. A lógica por trás dessa arquitetura garante que os dados estejam prontos, em tempo real, para avaliação e eventual seleção dos eventos mais relevantes (Yue, 2019).

Após a amplificação, os sinais com ganho alto (*high-gain*) são os principais utilizados para leitura, enquanto os de ganho baixo (*low-gain*) funcionam como uma alternativa de segurança em casos de saturação. Ambos os sinais são direcionados às placas de digitalização, que utilizam conversores analógico-digitais (ADCs) para amostrar os pulsos a cada 25 ns. Quando o sinal de alta sensibilidade ultrapassa o limite do ADC, o sistema automaticamente recorre ao canal de baixo ganho para preservar a integridade da informação. Os dados digitalizados são então armazenados temporariamente em memórias do tipo *pipeline*. Essas memórias mantêm os pulsos até que um comando de disparo (L1A) selecione os intervalos de interesse, que são então encaminhados para os próximos estágios do processamento (Peralva et al., 2010).

Após a seleção dos eventos pelo sistema de disparo, os segmentos de sinal escolhidos são enviados para a Placa de Interface. Nessa etapa, os dados passam por compressão e reestruturação antes de serem encaminhados à eletrônica de BE, especificamente para os módulos ROD (*Read-Out Drivers*). Os RODs exercem um papel central no processo de aquisição de dados: eles atuam como ponte entre os sinais provenientes do detector e o sistema global de leitura. Entre suas funções estão a organização e multiplexação dos dados originados na eletrônica de FE, além da distribuição do sinal TTC (*Timing, Trigger and Control*), responsável por alinhar temporalmente os diversos subsistemas com os instantes exatos de colisão dos feixes (Faria, 2017).

Outro papel fundamental desempenhado pelas placas RODs é a reconstrução da energia depositada nos pulsos, feita em tempo real por meio de algoritmos executados diretamente na eletrônica. Para isso, os dados são direcionados a duas FPGAs (*Field*

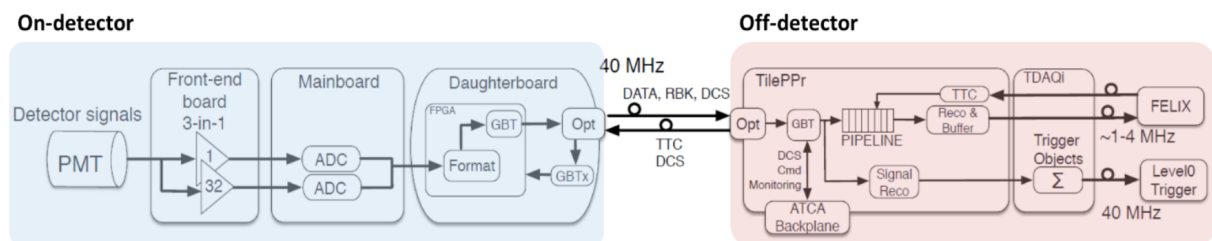
Programmable Gate Array) dedicadas, que os encaminham às chamadas Unidades de Processamento (PUs - *Processing Units*). Cada PU é equipada com dois processadores digitais de sinais (DSPs - *Digital Signal Processors*), responsáveis por aplicar os algoritmos de estimativa da amplitude dos pulsos — processo essencial para a determinação da energia associada a cada evento (Valero et al., 2007).

Essa eletrônica de leitura do Tilecal foi projetada há aproximadamente 30 anos. A instrumentação eletrônica tem avançado nesse período, e as limitações tecnológicas envolvidas no sistema atual são uma realidade, justificando a atualização do sistema descrito (Yue, 2019).

2.3.1.2 Sistema de Leitura do Tilecal na fase II

Para a fase II do LHC, está previsto um amplo processo de modernização nos sistemas de leitura do TileCal. O objetivo é substituir os circuitos atuais por versões mais robustas frente à radiação, com maior precisão na coleta de dados, conectores aprimorados e fontes de alimentação compatíveis com os novos requisitos energéticos (Cerqueira, 2015). A Figura 2.10 ilustra a nova arquitetura eletrônica proposta para essa etapa de atualização.

Figura 2.10 – Cadeia do sistema de leitura do Tilecal para fase II



Fonte: (Yue, 2019)

Assim como na arquitetura atual, a nova configuração eletrônica do Tilecal continua utilizando o sinal elétrico proveniente da PMT como ponto de partida. Esse sinal é encaminhado para a placa de *front-end 3-em-1*, cuja base tecnológica permanece semelhante à da geração anterior. No entanto, foram realizadas modificações pontuais, como ajustes nos níveis de ganho e a eliminação da necessidade de geração dos sinais de torres de *trigger*, otimizando o desempenho e simplificando o circuito (ATLAS Collaboration, 2017b).

Na sequência do fluxo de processamento, os sinais provenientes de até 12 placas 3-em-1 são digitalizados pela placa-mãe, conhecida como *mainboard*. Essa unidade também é responsável por direcionar os dados digitalizados para a *daughterboard* (placa-filha). A *daughterboard*, por sua vez, realiza o empacotamento dos dados e os transmite para a eletrônica de *back-end* por meio de conexões ópticas redundantes, garantindo maior confiabilidade na comunicação (Yue, 2019).

Na nova arquitetura, uma das mudanças centrais é a transmissão ininterrupta (ou *streaming*) dos dados digitalizados diretamente para o sistema de leitura *back-end*, operando agora em uma taxa significativamente maior — saltando de 100 *kHz* para 40 *MHz*. Esse aumento de desempenho demanda a substituição da atual placa ROD por um novo componente, o TilePPr (*Tile PreProcessor*), anteriormente conhecido em versões passadas do Tilecal e de outros subdetectores como super ROD (sROD) (Carrió et al., 2014).

A nova placa TilePPr foi projetada para substituir a antiga ROD, assumindo suas principais funções, agora implementadas em FPGAs modernos. Entre essas funções, destaca-se a reconstrução *online* da energia dos pulsos — atividade que continua sendo central neste trabalho. Enquanto essa reconstrução é realizada, os dados provenientes do sistema de *front-end* são mantidos temporariamente em uma memória do tipo *pipeline*, aguardando a emissão do sinal L1A, responsável por indicar se o evento deve ser preservado para análises futuras.

Paralelamente, a TilePPr também cuida do pré-processamento dos sinais de *trigger*, os quais são posteriormente enviados ao módulo TDAQi (*Trigger and Data Acquisition Interface*). Essa interface adapta os dados para garantir a compatibilidade entre os diferentes elementos da cadeia eletrônica (Zakareishvili, 2020).

Com isso, o sistema deixa de apenas reconstruir a energia de pulsos já selecionados e passa também a fornecer uma estimativa inicial da energia depositada em cada interação de partículas — realizada a cada intervalo de 25 *ns*. Essa informação adicional é usada pelo primeiro nível do sistema de seleção para auxiliar na decisão sobre a emissão do L1A. Essa funcionalidade, inédita em versões anteriores, constitui a base exploratória deste estudo, sendo discutida com mais profundidade nos capítulos seguintes.

Após a validação pelo sistema de disparo, os eventos considerados de interesse não seguem diretamente para o armazenamento. Em vez disso, eles são encaminhados à infraestrutura de transmissão de dados da fase II, denominada FELIX (*Front-End Link eXchange*). Esse sistema foi concebido para atuar como uma ponte eficiente entre os elementos eletrônicos de leitura e os servidores de agregação e arquivamento. O módulo TDAQi participa ativamente desse processo, sendo responsável por intermediar a comunicação e garantir que os dados cheguem ao destino final com o formato e a sincronização exigidos pela arquitetura do sistema de aquisição (Yue, 2019).

2.3.2 Sistema de *Trigger* do Tilecal

Devido à altíssima frequência de colisões no LHC e à grande quantidade de canais de leitura presentes no detector ATLAS, a geração de dados é massiva, tornando inviável o armazenamento contínuo de todas as amostras capturadas. Para lidar com essa limitação prática, implementa-se um mecanismo de filtragem em tempo real, conhecido como

sistema de *trigger*, que atua na identificação dos eventos com potencial interesse físico. Esse sistema opera de forma integrada ao processo de aquisição de dados, formando a arquitetura denominada TDAQ (*Trigger and Data Acquisition*), cuja função é selecionar e preservar apenas as informações consideradas relevantes para análise posterior (Peralva et al., 2010).

O sistema TDAQ é estruturado em múltiplos níveis hierárquicos, com o propósito de filtrar, em tempo real, apenas uma fração das milhões de amostras geradas a cada segundo, possibilitando o armazenamento de milhares delas. Esse processo envolve uma triagem instantânea dos dados provenientes dos *bunch-crossings*, selecionando os eventos que apresentam maior potencial de conter informações significativas sobre partículas geradas nas colisões. Aos eventos identificados como promissores é atribuído um código único, conhecido como *Bunch Crossing ID* (BCID), que os marca para a posterior análise (Pastore, 2016).

O sistema TDAQ do experimento ATLAS está passando por um processo de modernização alinhado às atualizações do LHC. As fases I e II desse aprimoramento são de especial interesse, pois visam adaptar o sistema às novas demandas experimentais impostas pelo aumento expressivo da luminosidade. Esse crescimento implica em uma elevação significativa no volume de dados a ser tratado — chegando a ser cem vezes maior — o que requer a implementação de soluções tecnológicas mais avançadas e eficientes para garantir a continuidade e a qualidade do processamento em tempo real (Solans Sánchez, 2019).

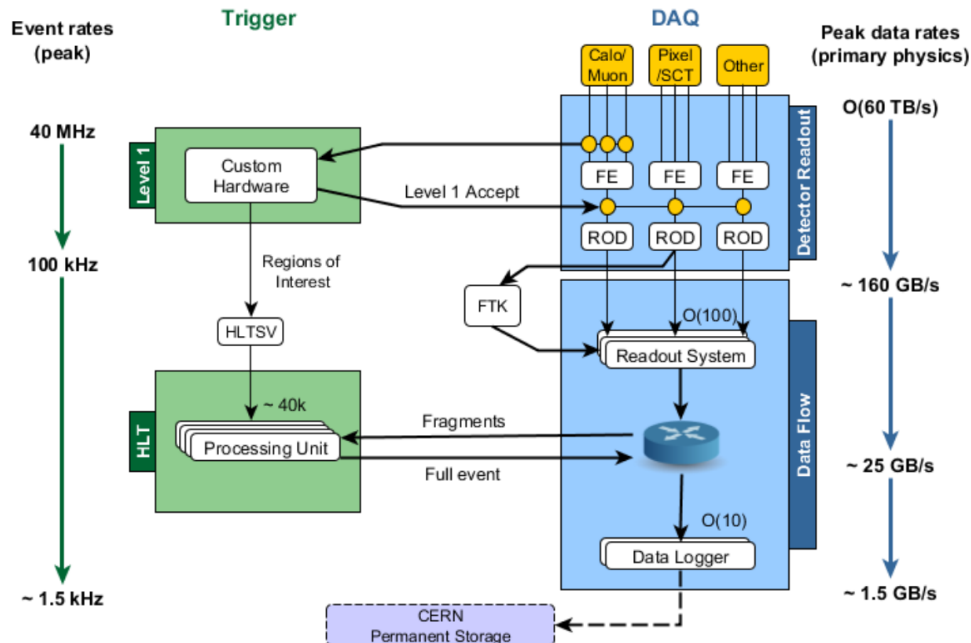
A Figura 2.11 apresenta uma representação da estrutura funcional do sistema TDAQ conforme utilizado na *Run 2*, cuja configuração permanece inalterada durante a fase I do LHC. Nessa configuração, o mecanismo de seleção de eventos se baseia na análise conjunta de dados vindos dos módulos de leitura dos calorímetros (tanto o eletromagnético quanto o hadrônico) e do espectrômetro de múons. Essa integração permite ao sistema de *trigger* identificar, em tempo real, quais eventos possuem maior potencial de conter informações físicas relevantes, promovendo um descarte eficiente de dados menos significativos.

2.3.2.1 Sistema do TDAQ na fase I

A partir de 2019, com o início da segunda grande parada do LHC (LS2), teve início a fase I do processo de modernização do sistema TDAQ do experimento ATLAS. Essa etapa foi concebida para preparar o sistema para operar em condições de luminosidade instantânea próximas de $\mathcal{L} \approx 2 - 3 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. Durante a *Run 3*, espera-se que o número médio de colisões por cruzamento de feixes (μ) alcance valores em torno de 80, demandando maior capacidade de processamento e seleção eficiente de eventos (Quirino, 2022).

No ambiente do ATLAS, o sistema TDAQ é concebido em duas etapas complementares, tanto para os sistemas de *trigger* quanto para a aquisição de dados. O estágio

Figura 2.11 – Diagrama de Blocos do sistema *trigger* e aquisição de dados da fase I



Fonte: (Vazquez and on behalf of the ATLAS Collaboration, 2017)

inicial é baseado em componentes de *hardware* de alta velocidade, responsáveis por uma filtragem imediata dos eventos. Já o segundo estágio atua com soluções de *software* mais sofisticadas, encarregadas de processar com profundidade os dados que passaram pela primeira triagem (Quirino, 2022).

As medições no ATLAS ocorrem a uma taxa de 40 MHz. No entanto, com a atuação do primeiro nível de *trigger* (L1A - *Level 1 Accept*), essa frequência é filtrada e reduzida para cerca de 100 kHz. Essa etapa inicial de seleção é essencial, pois permite uma queda drástica na taxa de dados armazenados, de aproximadamente 60 TB/s para cerca de 160 GB/s. Apesar das mudanças, o tempo de latência permanece fixado em 2,5 μs — valor mantido das versões anteriores para evitar alterações mais profundas na arquitetura do sistema de aquisição (Faria, 2017).

O filtro de eventos, também conhecido como *trigger* de alto nível (*High Level Trigger* – HLT), atua como a segunda etapa no sistema de seleção. Sua principal função é reduzir significativamente a taxa de eventos aceitos, passando de 100 kHz para aproximadamente 1 kHz. Esse processo ocorre com uma latência estimada em torno de 550 ms. Como resultado, o volume de dados a ser armazenado é drasticamente diminuído — caindo de cerca de 160 GB/s para aproximadamente 1,5 GB/s (Quirino, 2022).

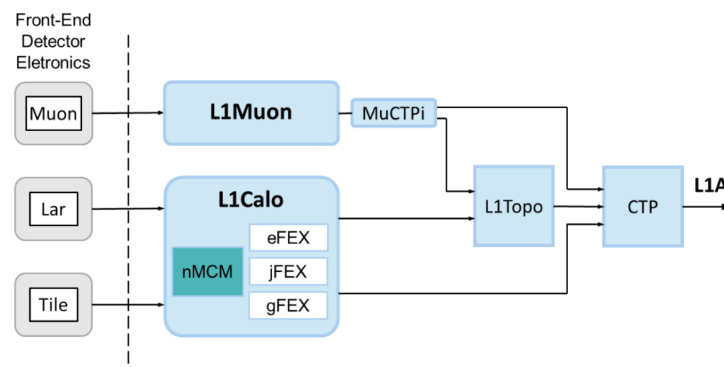
Quando ocorre o sinal de aceitação de *trigger*, L1A, os dados da eletrônica de FE são transmitidos por fibra óptica para as RODs, que desempenham a função de transferir os dados para o sistema de leitura. Este sistema é composto pelos *Readout Buffers* (ROBs), que armazenam temporariamente os fragmentos dos eventos. Paralelamente, o módulo

FTK (*Fast Track*) é responsável por realizar uma reconstrução acelerada das trajetórias, apoiando o processo de montagem de eventos no nível HLT e facilitando a reconstrução global dos traços identificados (Quirino, 2022).

O HLT é um sistema que implementa algoritmos de processamento rápido, com o objetivo de realizar uma análise mais aprofundada dos dados dentro de uma janela de tempo aproximada de 200 *ms*. Ele se baseia nas regiões de interesse (RoI - *Regions of Interest*) definidas pelo *trigger* de nível 1, a partir dos dados processados pelos ROB. Com essas informações, o HLT é capaz de identificar e selecionar eventos relevantes para a análise das partículas, armazenando esses eventos de forma permanente para análise posterior (Carrió et al., 2014).

A importância do sinal de L1A é destacada neste trabalho, uma vez que ele será utilizado como um dado especializado para o algoritmo de reconstrução de energia. Para ilustrar esse processo, o fluxograma da Figura 2.12 apresenta a formação do sinal, que ocorre dentro do circuito L1Calo. Este circuito envolve componentes eletrônicos específicos que processam informações dos calorímetros (L1Calo) e do detector de múons (L1Muon). Esses componentes têm a tarefa de ajustar os sinais detectados e realizar uma comparação entre os níveis de energia registrados e os valores pré-estabelecidos para diferentes partículas. Com base nessa análise, o sistema é capaz de identificar eventos que possam estar relacionados a partículas, através de uma avaliação de quão bem os dados correspondem aos modelos de partículas esperados (Quirino, 2022).

Figura 2.12 – Fluxograma do sistema de *trigger* da fase I



Fonte: (Chen, 2017), modificado

O L1Calo é formado por diversos sistemas especializados, como o gFEX, que extrai características globais de energia, o jFEX, que foca na detecção de jatos, e o eFEX, dedicado à análise de elétrons. Além disso, há um pré-processador avançado, o nMCM (*new Multi-Chip Module*), que utiliza uma FPGA (*Field Programmable Gate Array*) para melhorar a resolução dos sinais recebidos da eletrônica de FE. A associação geométrica entre jatos, elétrons e múons é feita pelo módulo L1Topo, que avalia padrões espaciais específicos para selecionar eventos de interesse (Palka, 2016). Por fim, o processador de

trigger central (CTP - *Central Trigger Processor*) é responsável por combinar os dados provenientes dos detectores processados pelo L1Topo, determinando quais eventos têm maior probabilidade de representar fenômenos significativos e gerando o sinal L1A (ATLAS Collaboration, 2017a).

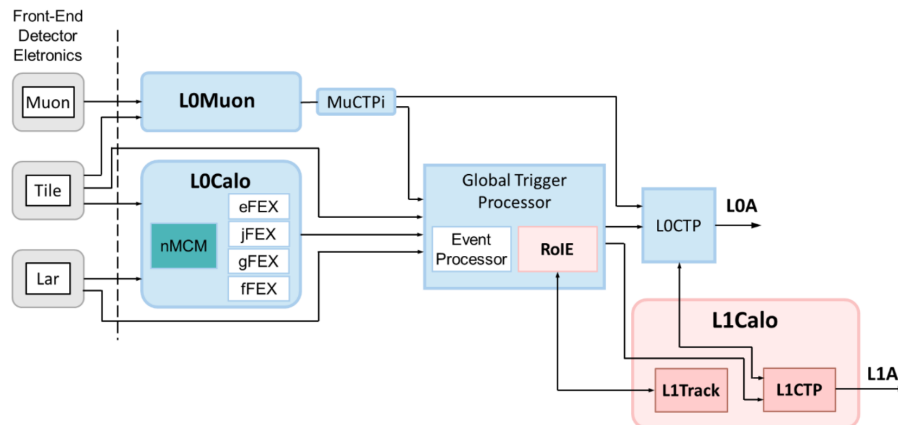
2.3.2.2 Sistema do TDAQ na fase II

A fase II de atualização do sistema foca em uma maior integração com o projeto de fase I, ao mesmo tempo em que incorpora interfaces de sistemas comerciais. A estrutura principal, conforme ilustrado na Figura 2.11, permanece essencialmente a mesma. Entretanto, com o aumento da luminosidade esperada, torna-se necessário um ajuste na taxa de armazenamento e leitura, além de um aumento na latência do *trigger*. Como resultado, está prevista a ampliação da taxa de leitura na eletrônica de BE para 40 MHz, com a substituição dos RODs pelos sRODs, que permitirão um processamento mais detalhado dos dados na cadeia de DAQ. O sistema FELIX será o intermediário entre os sRODs e a rede comercial de alta capacidade, enviando as RoI para o HLT, atendendo às exigências da nova arquitetura (Solans Sánchez, 2019; Goff, 2020).

Na atualização da fase II, a latência do primeiro nível de *trigger* é aumentada em quatro vezes, chegando a 10 μs , o que resulta em uma taxa final de armazenamento de dados de 10 kHz. Esse aumento exige uma modernização no processador FTK, que passa a ser denominado FTK++ (*Full Event Tracking*). A nova versão do FTK++ é projetada para identificar trajetórias com momentos transversais menores, otimizando a precisão e a eficiência no processamento dos dados de colisão (Camplani, 2020).

As mudanças no circuito de formação do sinal de aceitação de *trigger* para a fase II consideram duas abordagens, que são visualmente diferenciadas pelas cores no fluxograma mostrado na Figura 2.13. O primeiro cenário, representado em azul, mantém a estrutura de um único nível de *trigger* (L0), similar ao primeiro nível utilizado na fase I. O segundo cenário, destacado em rosa, introduz a combinação de dois níveis de *trigger* (L0/L1), com a adição de novos componentes que permitem a formação do nível 1 (L1), expandindo a capacidade do sistema (Quirino, 2022).

Os sistemas de extração de características que eram parte do L0Calo foram atualizados com base nos componentes do L1Calo da fase I, com a introdução de um novo subsistema projetado para aprimorar a identificação de elétrons. Esse novo subsistema, denominado fFEX, foca na extração de características eletromagnéticas mais avançadas, especialmente para jatos. O pré-processador nMCM, responsável pela melhoria da resolução dos sinais detectados, permanece em uso, com a função de otimizar a precisão dos dados. Os sinais, após processados, são direcionados para os sistemas de extração de características das partículas. Além disso, para otimizar a precisão da medição de energia, algoritmos especializados de deconvolução são implementados na FPGA do nMCM, facilitando a

Figura 2.13 – Fluxograma do sistema de *trigger* da fase II

Fonte: (ATLAS Collaboration, 2013), modificado

geração de sinais de *trigger* mais eficazes (Quirino, 2022).

Outro aspecto relevante é a mudança no processador L1Topo, que será substituído pelo *Global Trigger Processor*. Esse componente será integrado a todo o sistema de *trigger*, independentemente de qual abordagem de *trigger* for aplicada, garantindo maior versatilidade no tratamento de eventos (Quirino, 2022).

Quando o sistema de *trigger* L0 for utilizado isoladamente, o sinal L0A será gerado de maneira similar à implementação da fase I, mas agora pelo L0CTP (Processador Central de Trigger do Nível 0). A diferença principal está na taxa de operação, que será elevada para 1MHz , permitindo o fluxo contínuo de dados para o processamento no HLT e, subsequentemente, o armazenamento de eventos significativos (Quirino, 2022).

No segundo cenário, que envolve a combinação dos níveis de *trigger* L0 e L1, a taxa do sinal L0A é ajustada para 4MHz , sendo responsável por controlar o fluxo de dados apenas para o detector de trajetórias do ATLAS. A latência permanece em $10\ \mu\text{s}$. Neste contexto, o módulo RoIE (*Regions of Interest Engine*) é integrado ao *Global Trigger Processor*. Sua função é calcular as RoIs após o L0 com a máxima granularidade (Collaboration, 2015), levando em consideração a necessidade de processar todos os dados das células que, ao serem analisadas, apresentam uma discriminação clara entre o ruído e a energia, conforme o critério $|E| > 2\sigma_{\text{ruído}}$ (ATLAS Collaboration, 2017a).

No contexto do nível 1, a adição do processador L1Track permite a reconstrução das trajetórias nas RoIs, a partir dos dados fornecidos pelo RoIE. Após esse passo, o L1CTP, como processador central de *trigger*, é responsável por combinar e organizar as informações recebidas tanto do L0CTP quanto de outros processadores centrais. O sinal L1A gerado a partir desse processo controla a distribuição de dados de todos os detectores — calorímetros, sistema de múons e detector de trajetórias — com uma taxa de transmissão de 800kHz e uma latência de $35\ \mu\text{s}$, conforme especificado para este cenário (Izzo, 2020).

Essas propostas estão sendo avaliadas para determinar qual cenário oferece o armazenamento de dados mais eficiente, considerando o aumento previsto no número de eventos do LHC, especialmente a partir da fase I (Quirino, 2022).

2.4 EFEITO *PILE-UP*

O principal desafio da atualização para a fase II é a sobreposição de medições em um mesmo canal (*pile-up*), que adiciona significativamente ruído nos canais dos calorímetros. Esse *pile-up* impacta a reconstrução de energia nas células dos calorímetros, prejudicando a definição dos padrões de deposição de energia de cada partícula.

O efeito de empilhamento dos sinais na calorimetria pode ocorrer de duas maneiras: *in-time pile-up* e *out-of-time pile-up*. O *in-time pile-up* surge quando a medição das energias depositadas por partículas de diferentes interações no mesmo cruzamento de feixes ocorre simultaneamente. Nesse caso, uma deposição de energia é relevante (referente ao cruzamento analisado), enquanto as outras são indesejadas (referentes aos cruzamentos vizinhos). Já o *out-of-time pile-up* acontece devido à sobreposição de leituras de energia de cruzamentos de feixes consecutivos e, neste caso, ambas as leituras são de interesse.

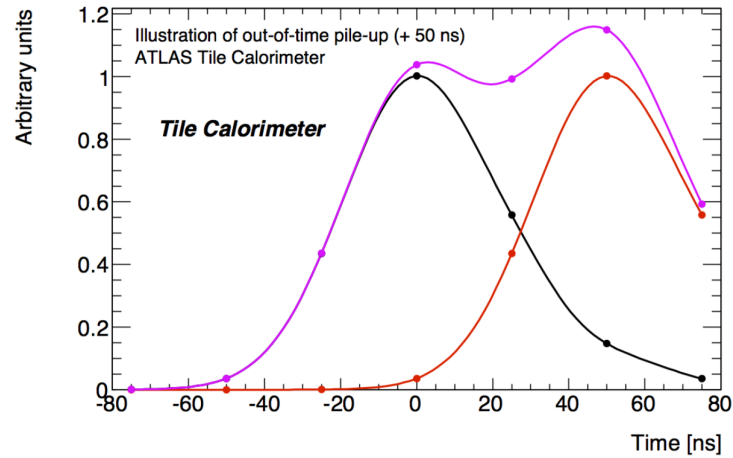
Os efeitos do *in-time pile-up* e do *out-of-time pile-up* têm impacto direto no ruído total nos detectores, com base na luminosidade instantânea (\mathcal{L}) e na taxa de colisões por cruzamento de feixe (μ). Esses efeitos estão intimamente ligados ao número de interações que ocorrem durante cada cruzamento de feixe de prótons. Como a luminosidade instantânea está projetada para valores elevados no HL-LHC, o *pile-up* será proporcional a esses valores, aumentando o impacto nas medições de energia e comprometendo a reconstrução dos sinais.

O pulso característico dos canais no TileCal tem uma duração de 150 ns (Anderson et al., 1998), enquanto o intervalo entre os cruzamentos de feixes no LHC é de apenas 25 ns. A decisão de não reduzir a duração do pulso, conforme discutido por Cerqueira (2015), visa evitar uma maior exigência de resposta em frequência dos componentes do sistema de leitura. No entanto, essa escolha resulta em uma maior probabilidade de sobreposição entre colisões, aumentando o efeito do *out-of-time pile-up*, conforme ilustrado na Figura 2.14.

2.4.1 *Pile-up* como fonte de ruído

Com o aumento da luminosidade, o ruído total nos canais de leitura passa a ser fortemente influenciado pela presença de *pile-up*. Nesse cenário, o ruído introduzido pelas múltiplas interações por cruzamento de feixes se sobrepõe ao ruído eletrônico, tornando-se o componente predominante por uma margem significativa — frequentemente superando-o em mais de dez vezes (ATLAS Collaboration, 2017b).

Figura 2.14 – Gráfico do pulso do TileCal demonstrando o efeito de empilhamento do sinal



Fonte: (ATLAS Collaboration, 2017b)

O ruído proveniente do *pile-up* pode ser tratado como uma das componentes do ruído total, sendo somado em quadratura ao ruído eletrônico, conforme apresentado na Equação (2.3). Essa abordagem se justifica pela suposição de que ambos os ruídos — o eletrônico e o causado pelo *pile-up* — não apresentam correlação entre si, permitindo que suas contribuições sejam combinadas dessa forma (Novakova, 2014).

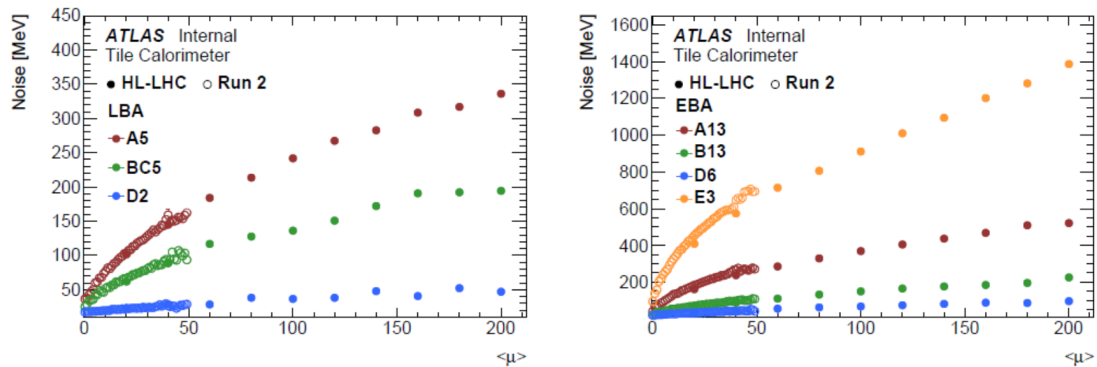
$$\sigma = \sqrt{\sigma_{pileup}^2 + \sigma_{eletr\tilde{on}ico}^2} \quad (2.3)$$

O ruído eletrônico pode ser atribuído a instabilidades de natureza sistemática ou aleatória nos sinais medidos. Entre os fatores sistemáticos, destaca-se a variação no processo de conversão óptico-eletrônica, responsável pelo acúmulo de carga. Já as variações aleatórias podem surgir de inconsistências na impedância do circuito ou de deficiências no aterramento das fontes de alimentação, impactando diretamente a qualidade do sinal (Anderson et al., 1998).

O ruído eletrônico apresenta comportamento independente da amplitude da medição, ou seja, à medida que a energia do sinal aumenta, sua influência relativa diminui, melhorando assim a razão sinal-ruído em altos níveis de energia. Em contrapartida, o ruído decorrente do *pile-up* tende a afetar as leituras de forma semelhante em toda a faixa energética, sendo diretamente influenciado pela elevação da luminosidade instantânea. A Figura 2.15 demonstra como o crescimento da taxa de colisões por cruzamento de feixes intensifica gradualmente o ruído registrado em múltiplas células do TileCal (representadas por sua respectiva letra e número), evidenciando a predominância do ruído de *pile-up* nesse cenário (Gonçalves et al., 2021).

Um modelo aproximado pode ser utilizado para descrever a relação entre a taxa de colisões por cruzamento de feixes (μ) e o ruído de *pile-up*. Essa relação é expressa

Figura 2.15 – Distribuição simulada do ruído nas células do TileCal em função do número médio de interações por cruzamento de feixes



Fonte: (ATLAS Collaboration, 2017b)

na Equação (2.4), onde o termo σ_{pileup}^{norm} representa uma constante associada ao *pile-up*. Essa constante garante que o modelo se ajuste adequadamente à quantificação do ruído, especialmente em cenários com alta frequência de interações e com deposições de energia ocorrendo em intervalos curtos (Novakova, 2014).

$$\sigma_{pileup} = \sigma_{pileup}^{norm} \cdot \sqrt{\mu} \quad (2.4)$$

Atualmente, simular o impacto do *pile-up* representa um desafio, já que ainda não há observações experimentais suficientes sobre o comportamento dos calorímetros sob as condições extremas de luminosidade previstas para os futuros ciclos dos aceleradores (Fabjan and Gianotti, 2003). Por isso, é necessário construir modelos preditivos capazes de estimar como o *pile-up* se manifestará nesses contextos de operação intensificada.

2.4.2 Simulação do *Pile-up*

Devido à natureza aleatória dos fenômenos analisados no experimento ATLAS, torna-se essencial o uso de abordagens estatísticas para a realização de simulações. O objetivo dessas simulações é reproduzir tanto a energia quanto a trajetória das partículas geradas nas colisões entre prótons. Nesse contexto, os métodos de Monte Carlo (MC) mostram-se eficazes, uma vez que permitem a geração de variáveis aleatórias de acordo com distribuições probabilísticas que refletem o comportamento esperado dos eventos físicos simulados (Guatelli et al., 2011).

A utilização de simulações desempenha um papel fundamental no processamento e interpretação dos dados, pois permite a construção de modelos geométricos representativos dos chuveiros de partículas. Esses modelos, por sua vez, são essenciais para a calibração precisa dos calorímetros (Budagov et al., 2006).

No ATLAS, a simulação dos eventos físicos conta com uma variedade de algoritmos

representar todos os aspectos físicos envolvidos. Nesse tipo de abordagem, as propriedades estatísticas dos sinais, como distribuição de energia e o ruído associado, podem ser ajustadas com precisão, permitindo um controle refinado sobre os parâmetros mais relevantes. Como o foco está em replicar as distribuições probabilísticas associadas às medições, elementos considerados secundários ao processamento podem ser intencionalmente omitidos, sem comprometer a validade dos resultados (Rimoldi, 2011).

Quando se trata de simular os canais de medição dos calorímetros utilizando simuladores TMC, é fundamental controlar diversos parâmetros que impactam diretamente a precisão e a eficiência da simulação. Entre esses parâmetros estão: o ruído eletrônico, o ruído causado pelo *pile-up*, a amplitude dos sinais e o desvio de fase. Esses elementos são cruciais para garantir que a simulação reflita adequadamente as condições reais do experimento.

No contexto do simulador TMC, a taxa de ocupação (γ_{ocup}) é usada como medida para quantificar o *pile-up*. Essa taxa é calculada a partir da janela de medição, sem depender diretamente dos processos físicos. A sincronização entre os intervalos de eventos do LHC (f_r) e a taxa de amostragem de 40 MHz (f_s) facilita esse cálculo. Especificamente, a taxa de ocupação é a razão entre o número de amostras que apresentam um depósito de energia (N_{BCID}) e o número total de amostras de medição (N). Esse valor é expresso dentro do intervalo de 0 a 1, conforme indicado pela Equação (2.5) (Quirino, 2022).

$$\gamma_{ocup} = \frac{N_{BCID}}{N} \quad (2.5)$$

A taxa de amostras ocupadas (v) e a frequência de amostragem (f_s), definidas pelas equações (2.6) e (2.7), podem ser utilizadas para encontrar a ocupação em cada canal. Dessa forma, é possível manipular as equações e encontrar a ocupação como expresso na Equação (2.8)

$$v = \frac{\Delta N_{BCID}}{\Delta t} \quad (2.6)$$

$$\gamma = \frac{\Delta N}{\Delta t} \quad (2.7)$$

$$\gamma_{ocup} = \frac{v}{f_s} \quad (2.8)$$

A ocupação pode ser interpretada como uma medida de probabilidade ($P[ocp]$), representando a chance de uma amostra estar ocupada, considerando várias amostras. Ela pode ser calculada pelo produto da taxa de amostras ocupadas (v) e o tempo de duração de cada amostra (τ), gerando a Equação (2.9). Essa expressão é consistente

com a Equação (2.8), sendo reescrita em termos do período de amostragem ($\Delta t = f_s^{-1} = \tau$) (Wielers, 2015).

$$P[ocp] = v \cdot \tau = \gamma_{ocup} \quad (2.9)$$

Sabendo a probabilidade de ocupação, a probabilidade de vacância (evento sem colisão) representada por $P[vac]$ é dada pela Equação (2.10), já que a soma dessas duas probabilidades deve totalizar um.

$$P[vac] = 1 - v \cdot \tau = 1 - \gamma_{ocup} \quad (2.10)$$

A média de colisões por cruzamento (μ) em uma distribuição geométrica (α), em conjunto com a probabilidade de vacância, pode determinar a taxa de amostras ocupadas (v) como mostrado na Equação (2.11) (Wielers, 2015). A distribuição geométrica é um valor fixo para cada canal e é encontrada através do sistema de coordenadas do ATLAS. Dessa forma, esse termo é constante para a análise de um canal específico (Quirino, 2022).

$$v = \alpha \cdot \mu \cdot P[vac] \quad (2.11)$$

Definindo o coeficiente normalizado de ocupação para cada canal como $\gamma_{ocup}^{norm} = \alpha\tau$, é possível determinar a taxa de ocupação do canal e a taxa média de colisões por cruzamento através das equações (2.12) e (2.13). As equações 2.8, 2.10 e 2.11 são utilizadas para chegar a este resultado.

$$\gamma_{ocup} = \frac{\mu}{\gamma_{ocup}^{norm} + \mu} \quad (2.12)$$

$$\mu = \gamma_{ocup}^{norm} \frac{\gamma_{ocup}}{1 - \gamma_{ocup}} \quad (2.13)$$

Controlar a ocupação em um canal permite, de maneira indireta, reproduzir os efeitos da taxa média de interações por cruzamento de feixes. Essa relação torna possível, ao se empregar a Equação (2.13) junto à Equação (2.4), estabelecer um vínculo direto entre a ocupação e o ruído associado ao *pile-up*. Assim, a intensidade desse tipo de ruído também pode ser ajustada por meio do parâmetro de ocupação, como descrito na Equação (2.14) (Quirino, 2022).

$$\sigma_{pileup} = \sigma_{pileup}^{norm} \sqrt{\gamma_{ocup}^{norm} \frac{\gamma_{ocup}}{1 - \gamma_{ocup}}} \quad (2.14)$$

Os resultados obtidos a partir da modelagem matemática do *pile-up* servem como base para o desenvolvimento do modelo TMC aplicado aos canais do TileCal. Essa

abordagem é viável graças à alta qualidade dos sinais obtidos pela eletrônica de leitura, assegurada pelos sistemas de calibração. Como ilustrado na Figura 2.8, um dos principais propósitos da calibração é garantir que a forma do pulso se mantenha constante, promovendo uma resposta padronizada entre os milhares de canais que integram o sistema calorimétrico (Peralva et al., 2010).

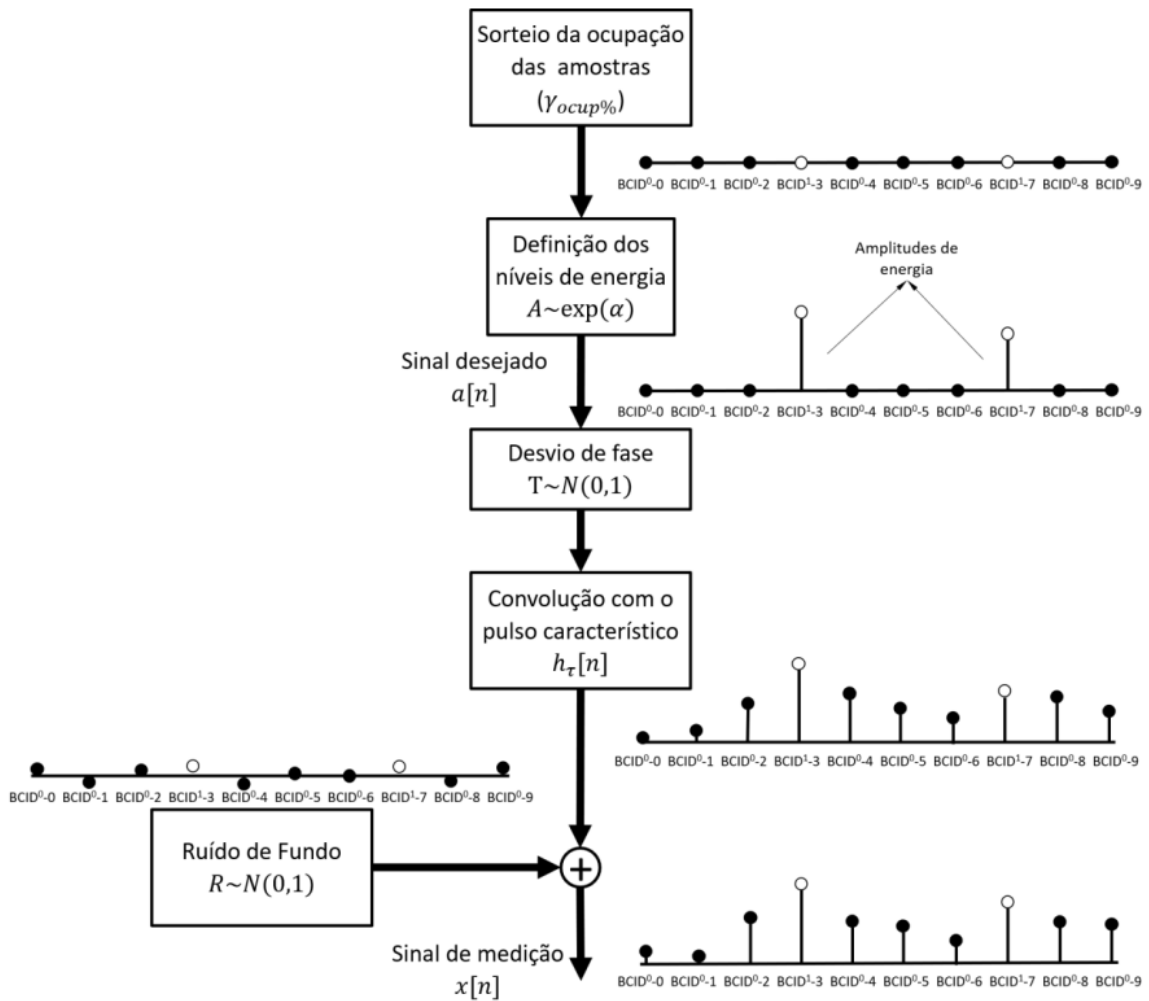
A Figura 2.17 ilustra a lógica por trás da simulação no modelo TMC, destacando um exemplo prático da geração de sinais. Nessa representação, os círculos sem preenchimento indicam as amostras escolhidas aleatoriamente para conter depósitos energéticos simulados, conforme os critérios definidos pelo modelo.

O processo de simulação pode ser resumido na seguinte ordem:

1. Para a geração do sinal simulado, estabelece-se previamente um valor para a taxa de ocupação (γ_{ocup}), o qual define a proporção de amostras que irão conter depósitos de energia (BCID¹) em relação às que permanecerão vazias (BCID⁰). Esse processo segue uma distribuição de Poisson, adequada para modelar a ocorrência de eventos discretos — neste caso, o número de depósitos energéticos — ao longo de um intervalo fixo de medição (Quirino, 2022).
2. Os valores de energia associados aos BCID¹ são gerados de forma aleatória, com base em uma distribuição exponencial, cuja média é de 30 unidades ADC. Esse parâmetro é determinado a partir de informações extraídas do banco de dados do LHC, garantindo a fidelidade estatística em relação às medições reais realizadas no experimento (Barbosa, 2017).
3. Os desvios de fase podem provocar um deslocamento no pulso característico, o qual é modelado por uma distribuição de probabilidade aproximada à gaussiana, com média de $0ns$ e desvio padrão de $1ns$. Este efeito reflete as variações no tempo de voo das partículas geradas durante cada colisão (Faria, 2017).
4. Para representar a dispersão nos canais dos calorímetros, cada amostra de energia é ajustada por meio de uma convolução com um pulso característico, normalizado em 7 BC, conforme o modelo de conformação implementado na placa *3-in-1* (Peralva et al., 2010).
5. Por fim, o sinal é complementado com amostras de ruído gaussiano, a fim de simular o ruído eletrônico presente nos canais dos calorímetros (Faria, 2017).

Dessa forma, o simulador TMC é capaz de se aproximar de ambientes com presença de *pile-up*, auxiliando a sugestão e estudo de novas técnicas de reconstrução de energia, comparando-se a outros modelos já propostos.

Figura 2.17 – Fluxograma da simulação TMC



Fonte: (Quirino, 2022)

2.5 CONCLUSÕES PARCIAIS DO CAPÍTULO

Neste capítulo, foi apresentado o contexto do trabalho, com uma visão geral dos experimentos realizados no LHC e suas atualizações previstas para o HL-LHC, na chamada fase II. O foco foi o experimento ATLAS, particularmente o sistema de calorimetria, com ênfase na seleção de dados online em cenários de alta luminosidade, onde o efeito *pile-up* se torna relevante.

Foi realizada uma análise detalhada do *pile-up*, seguida de uma discussão sobre a simulação dos sinais, o que destacou a importância de novos algoritmos para o processamento desses sinais, com o objetivo de melhorar a reconstrução de energia no HL-LHC. Com isso, foi feita uma avaliação do estado atual das técnicas utilizadas na reconstrução de energia do TileCal.

3 ESTIMAÇÃO DE ENERGIA NO TILECAL

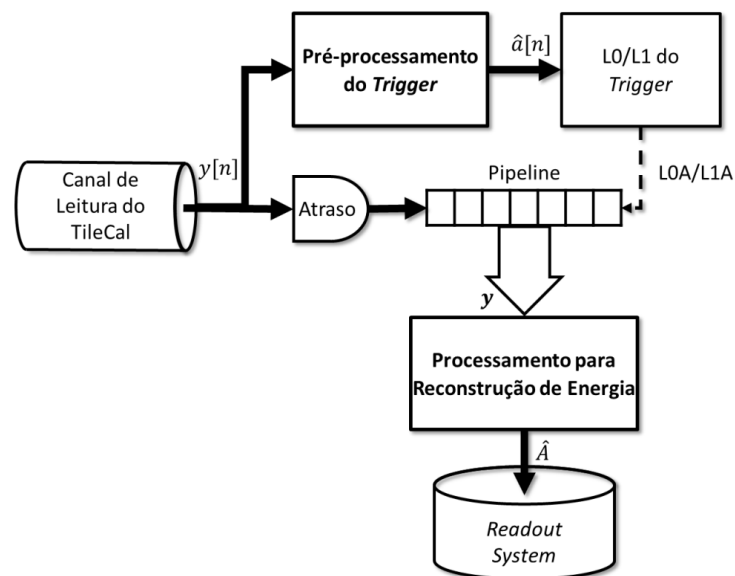
Os experimentos realizados no LHC representam um marco no avanço da física de partículas, e esse progresso tende a se intensificar com as atualizações previstas para o HL-LHC, a Fase II do colisor. Com a elevação significativa da luminosidade, espera-se uma densidade maior de partículas nos feixes, aumentando a taxa de interações nas colisões. Isso permite a obtenção de um volume mais robusto de dados experimentais, ampliando a sensibilidade à detecção de processos raros.

Durante as fases iniciais do ATLAS, o sistema de calorimetria foi projetado sem considerar os efeitos da luminosidade sobre as medições, sendo o principal fator de distorção o ruído eletrônico. Nessa etapa, algoritmos de processamento de sinais atuavam para selecionar apenas as amostras relevantes à análise física, eliminando informações redundantes.

Entretanto, com o aumento da luminosidade, emerge o fenômeno conhecido como *pile-up*, caracterizado pela sobreposição de múltiplos eventos em uma mesma leitura. Esse efeito compromete a precisão das medições e não pode ser tratado eficientemente pelos métodos tradicionais utilizados anteriormente (Peralva et al., 2010).

Diante desse desafio, propõem-se estratégias avançadas de processamento de sinais em tempo real, aplicadas aos canais do TileCal. O objetivo é mitigar simultaneamente os impactos do *pile-up* e do ruído eletrônico. Esses algoritmos devem ser integrados tanto no pré-processamento do primeiro nível de *trigger*, quanto na etapa de reconstrução de energia, conforme ilustrado na Figura 3.1.

Figura 3.1 – Fluxograma da simulação TMC



Fonte: (Quirino, 2022)

A determinação da energia depositada nos canais do TileCal é realizada através de um sistema digital implementado em FPGA, denominado TilePPr. Esse sistema opera utilizando janelas de sete amostras consecutivas, armazenadas em um *pipeline*, que formam o vetor de entrada \mathbf{y} para o processamento. Essa janela é configurada previamente de modo que a amostra central esteja alinhada com o pico do sinal, representando a amplitude depositada (A). A reconstrução da energia estimada (\hat{A}) ocorre somente quando um evento relevante é identificado pelo sistema de *trigger*, momento em que as amostras são selecionadas para armazenamento.

O estágio inicial do sistema de *trigger* (níveis L0/L1) possui um módulo de pré-processamento responsável por melhorar a eficiência na emissão do sinal de aceitação (L0A ou L1A). Essa etapa, também baseada em FPGA, é realizada pelo componente nMCM (representado na Figura 2.12 e Figura 2.13). Nessa fase, o algoritmo executa uma estimativa contínua e em tempo real do sinal amostrado $y[n]$, aplicando uma técnica de deconvolução para calcular $\hat{a}[n]$, que representa a energia depositada em cada intervalo de leitura.

Devido à introdução de latência pelos blocos de processamento do *trigger*, é necessário sincronizar o fluxo de dados com o sistema de leitura, o que é feito ajustando o tempo de retenção das amostras no *buffer*. Embora ambos os sistemas — *trigger* e leitura — processem os sinais de maneira simultânea, não há compartilhamento direto dos dados entre eles, pois cada um tem objetivos distintos.

Enquanto o sistema de *trigger* precisa realizar estimativas rápidas e contínuas para identificar qual *bunch crossing* contém sinais significativos e fornecer uma estimativa aproximada de energia para cada instante ($\hat{a}[n]$), o sistema de leitura realiza uma reconstrução mais precisa. Este último foca na estimativa exata da amplitude do sinal associado ao pico, que ocorre tipicamente na quarta posição da janela de sete amostras.

Apesar de funcionarem de maneira separada, existe uma conexão funcional entre os dois sistemas: a decisão do *trigger* (L1A) sobre quais eventos devem ser registrados influencia diretamente quais amostras são efetivamente armazenadas. Portanto, mesmo que o processamento seja independente, o sistema de leitura depende da atuação prévia do *trigger* para definir o conteúdo que será preservado.

3.1 FILTROS DE RECONSTRUÇÃO DE ENERGIA NO SISTEMA DE LEITURA

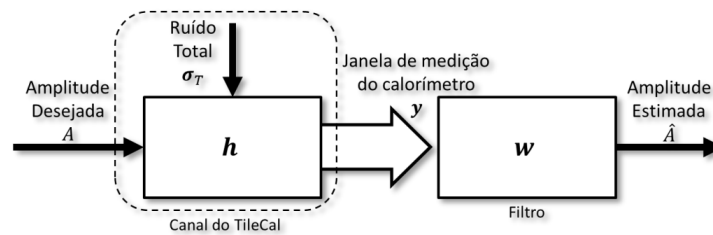
Os algoritmos de reconstrução de energia dos canais do TileCal são fundamentados em um modelo matemático que descreve a resposta do sistema, conforme ilustrado na Figura 3.2 e formalizado na Equação (3.1). Nesse modelo, uma janela de medição é extraída da resposta do canal do TileCal, permitindo que um filtro atue como estimador da amplitude do sinal. Nesse contexto, o efeito de *pile-up* é tratado como um componente de ruído, uma vez que o foco do processamento está na energia depositada que corresponde à

amostra central da janela de leitura. As demais contribuições, resultantes de sobreposições temporais de outros sinais, são incorporadas ao vetor de ruído.

Além disso, o ruído eletrônico — intrínseco ao sistema de leitura — também influencia a medida final. Ambos os tipos de ruído são combinados por soma em quadratura, conforme apresentado anteriormente na Equação (2.3), compondo o ruído total σ_{Total} .

$$\mathbf{y} = \mathbf{h} \cdot A + \sigma_{\text{Total}} \quad (3.1)$$

Figura 3.2 – Representação do modelo de um canal do TileCal e estrutura do filtro de estimação de energia.



Fonte: Adaptado de Haykin (2014)

A Equação (3.1) representa o modelo matemático para o vetor de observação \mathbf{y} . Neste modelo adotado, o vetor \mathbf{h} representa o pulso de referência do canal com amplitude normalizada, composto por sete amostras alinhadas temporalmente: $\mathbf{h} = [h_{-3}, h_{-2}, \dots, h_3]$. A cada evento identificado pelo sistema de *trigger* (L0/L1A), uma nova janela de observação \mathbf{y} é formada. Essa janela contém a superposição do sinal escalado por sua amplitude real (A) e os ruídos mencionados.

A estimativa da energia depositada, denotada por \hat{A} , é realizada de forma online através da aplicação de filtros digitais às amostras da janela. O objetivo é minimizar o impacto do ruído total sobre o valor estimado. Os coeficientes desses filtros podem ser determinados com base no conhecimento prévio da forma do pulso de resposta do canal ou por métodos que analisam diretamente a relação entre sinais de entrada e saída do sistema. Os principais procedimentos para esse tipo de reconstrução de energia são discutidos a seguir.

3.1.1 Filtro Ótimo

Desde 2014, o método empregado, tanto na leitura de dados quanto no primeiro nível de *trigger* do TileCal, é o chamado Filtro Ótimo (*Optimal Filter* - OF), cuja formulação original foi estendida com variações discutidas por Peralva et al. (2010). Inicialmente adotado em calorímetros de ionização líquida, como o LAr do próprio detector ATLAS, o OF demonstrou desempenho robusto em ambientes de alta luminosidade (Duarte, 2019), o que incentivou sua aplicação também no TileCal.

A técnica de parametrização do OF é fundamentada no conceito de estimador linear não tendencioso de mínima variância, conhecido como BLUE (*Best Linear Unbiased Estimator*). A estratégia consiste em minimizar a variância do estimador de amplitude a partir da função de autocorrelação do ruído, o que permite estimar simultaneamente a amplitude da energia depositada e sua posição temporal, atenuando tanto o ruído eletrônico quanto o efeito de *pile-up* sobre a amostra central (Fullana et al., 2005).

A expressão geral para o cálculo da amplitude estimada \hat{A} é dada pela Equação (3.2).

$$\hat{A} = \sum_{k=0}^{M-1} w_k y[k] \quad (3.2)$$

Nesse contexto, \mathbf{y} representa a janela de observação com $M = 7$ amostras, e os coeficientes w_k são os pesos fixos aplicados a cada uma dessas amostras (Gonçalves et al., 2018). Para que o estimador seja não tendencioso, o seu valor esperado deve coincidir com o valor real da amplitude:

$$\mathbb{E}\{\hat{A}\} = \sum_{k=0}^{M-1} w_k \mathbb{E}\{y[k]\} = A \quad (3.3)$$

A variância associada à estimativa é determinada por:

$$\text{var}\{\hat{A}\} = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (3.4)$$

onde \mathbf{C} representa a matriz de covariância do ruído. A minimização dessa variância com o modelo matemático dado pela Equação (3.1) fornece os coeficientes ótimos:

$$\mathbf{w} = \frac{\mathbf{h}^T \mathbf{C}^{-1}}{\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}} \quad (3.5)$$

Assim, a forma final do estimador da amplitude é expressa como:

$$\hat{A} = \frac{\mathbf{h}^T \mathbf{C}^{-1} \mathbf{y}}{\mathbf{h}^T \mathbf{C}^{-1} \mathbf{h}} \quad (3.6)$$

Embora o OF apresente bom desempenho quando o ruído predominante é eletrônico — geralmente com distribuição gaussiana e média nula — sua eficácia é comprometida em cenários onde o *pile-up* se torna dominante. Nesse caso, o ruído apresenta distribuição não gaussiana e média deslocada, reduzindo a precisão do estimador (Paschoalin, 2016).

Outro fator limitante está na hipótese de que todos os canais do TileCal compartilham uma mesma forma de pulso característica e invariável. Essa suposição, ao ser aplicada uniformemente aos mais de 10.000 canais, pode não refletir as variações reais do sistema. A dependência excessiva da forma do pulso e as simplificações feitas na modelagem do ruído motivam a investigação de abordagens mais robustas e adaptativas para o processamento de sinais Gonçalves et al. (2018).

3.1.2 Filtro de Wiener

Considerando-se que as características estatísticas do sistema — incluindo a relação entre entrada e saída dos canais — são conhecidas, torna-se viável aplicar técnicas de equalização de filtros lineares sem a necessidade de se recorrer diretamente à resposta impulsiva do canal. Nesse cenário, o filtro de Wiener surge como uma solução ótima, fundamentada no princípio da ortogonalidade (Haykin, 2014), sendo essa a base da proposta apresentada por Gonçalves et al. (2018) para aplicação no sistema de leitura do TileCal.

A construção do filtro de Wiener envolve a minimização do erro quadrático médio entre a estimativa \hat{A} e o valor real da amplitude A , como descrito pela função de custo $\mathcal{J}(\mathbf{w})$, cujo mínimo define os coeficientes ótimos \mathbf{w}_o (Haykin, 2014):

$$\mathbf{w}_o = \arg \min_{\mathbf{w}} \mathcal{J}(\mathbf{w}) \quad (3.7)$$

$$\mathcal{J}(\mathbf{w}) = \mathbb{E}\{(A - \hat{A})^2\} = \mathbb{E}\{(A - \mathbf{w}^T \mathbf{y})^2\} \quad (3.8)$$

Para encontrar o vetor de pesos que minimiza essa função, deriva-se \mathcal{J} em relação a \mathbf{w} e iguala-se o gradiente a zero, conforme o método tradicional da descida de gradiente:

$$\nabla_{\mathbf{w}} \mathcal{J} = \mathbb{E}\{\mathbf{y}(A - \mathbf{w}^T \mathbf{y})\} = 0 \quad (3.9)$$

A equação resultante pode ser expressa em termos da matriz de autocorrelação da entrada $\mathbf{R} = \mathbb{E}\{\mathbf{y}\mathbf{y}^T\}$ e do vetor de correlação cruzada $\mathbf{p} = \mathbb{E}\{\mathbf{y}A\}$, levando à equação normal:

$$\mathbf{R}\mathbf{w}_o = \mathbf{p} \quad (3.10)$$

Assumindo que \mathbf{R} seja não singular, os pesos ótimos podem ser obtidos por:

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} \quad (3.11)$$

No caso do TileCal, o ruído de *pile-up* apresenta média distinta de zero. Para lidar com esse viés, introduz-se um termo adicional no vetor de entrada: $\mathbf{y}_b = [\mathbf{y}, 1]$. Consequentemente, o vetor de pesos também é estendido: $\mathbf{w}_b = [\mathbf{w}_o, b]$, onde o termo b atua como um elemento de compensação (bias), ajustando o filtro para que ele permaneça não tendencioso mesmo diante do desvio médio do ruído (Gonçalves et al., 2018).

Embora a amplitude real da energia (A) não seja diretamente observável nos dados experimentais do TileCal, a disponibilidade de modelos estatísticos confiáveis — obtidos por meio de simulações Toy Monte Carlo (TMC) — permite extrair pares (A, \mathbf{y})

representativos do comportamento do canal (Duarte, 2019). Essas amostras simuladas viabilizam a aplicação do método de Wiener em fase de projeto, com a equalização sendo feita de forma offline. Assim, uma vez estimados, os coeficientes são fixados para uso operacional.

Os estudos realizados por Gonçalves et al. (2018) demonstram que o filtro de Wiener oferece vantagens em relação ao Filtro Ótimo tradicional, especialmente em condições onde o ruído *pile-up* compromete a precisão das técnicas anteriores empregadas na reconstrução de energia no TileCal.

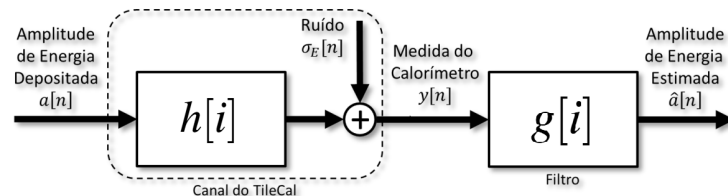
3.2 FILTROS DE DECONVOLUÇÃO NO PRIMEIRO NÍVEL DE TRIGGER

No contexto do primeiro nível de *trigger*, o tratamento do *pile-up* apresenta particularidades em relação ao sistema de leitura, principalmente porque, nesse estágio, ainda não é possível identificar quais pulsos correspondem a interações físicas relevantes. Diante dessa incerteza, torna-se necessário recorrer a métodos que permitam separar, de forma eficiente, o sinal de interesse da resposta do sistema. Um dos caminhos viáveis é o uso de algoritmos de deconvolução, capazes de reverter o efeito da convolução entre a resposta impulsiva do canal e os sinais de entrada (Haykin, 2014).

Modela-se o canal como um sistema digital linear e invariante no tempo (LTI), seguido por um filtro cuja função é recuperar o sinal impulsivo de energia original a partir das observações ruidosas. Esse modelo, representado esquematicamente na Figura 3.3 e descrito pela Equação (3.12), estabelece que o sinal observado $y[n]$ resulta da convolução do vetor de entrada $a[n]$ com a resposta impulsiva $h[i]$, somada ao ruído eletrônico $\sigma_E[n]$:

$$y[n] = \sum_{i=-\infty}^{\infty} h[i] \cdot a[n - i] + \sigma_E[n] \quad (3.12)$$

Figura 3.3 – Estrutura do canal modelado como sistema LTI seguido de filtro estimador.



Fonte: Adaptado de Haykin (1996)

Esse tipo de formulação não é exclusivo da física de partículas: técnicas semelhantes são amplamente utilizadas em outras áreas que envolvem reconstrução de sinais degradados. Em telecomunicações, por exemplo, algoritmos de deconvolução são empregados na equalização de canais com interferência intersimbólica (Peralva et al., 2010). No campo da geofísica, são usados para recuperar estímulos sísmicos a partir de medições em superfície,

mesmo quando o impulso original não é diretamente observável, desde que o modelo do meio esteja previamente caracterizado (Haykin, 2014).

Esse cenário é comparável ao do TileCal: não se conhece a energia real depositada pelas partículas, mas é possível medir a resposta do canal e utilizar um modelo predefinido do sistema. Essas condições tornam aplicáveis os métodos de deconvolução no processamento dos sinais coletados, o que tem motivado esforços dentro da colaboração brasileira do experimento ATLAS para substituir as abordagens baseadas no Filtro Ótimo por estratégias mais robustas e adaptativas.

Assim, nas seções seguintes, são apresentadas as técnicas de filtragem por deconvolução que se alinham com esta proposta, com ênfase em sua aplicação prática no sistema de leitura do TileCal.

3.2.1 Estimativa Inversa do Sistema de Leitura

As abordagens clássicas para a construção de filtros de deconvolução baseiam-se na aproximação da inversa do sistema de leitura, considerando o comportamento no domínio da frequência e a interpretação de sistemas lineares. Parte-se da premissa de que a resposta impulsiva do canal, representada pelo vetor de coeficientes \mathbf{h} , é bem caracterizada. Sendo essa hipótese válida, torna-se possível recuperar a estimativa $\hat{a}[n]$ do sinal de entrada $a[n]$ a partir da saída do sistema $y[n]$, utilizando um filtro com pesos definidos por \mathbf{w} (Mitra, 2006).

3.2.1.1 Inversão no Domínio- z

No domínio- z , a formulação do sistema inverso é direta, conforme a Equação (Mitra, 2006):

$$W(z) = \frac{1}{H(z)} \quad (3.13)$$

onde $H(z)$ representa a Transformada- z da resposta impulsiva do canal, e $W(z)$ corresponde à Transformada- z do filtro de deconvolução. No entanto, esse método geralmente resulta em um filtro do tipo IIR (Resposta Infinita ao Impulso), o qual pode apresentar instabilidade e fase não mínima, propriedades indesejáveis em aplicações instrumentais que exigem robustez (Mitra, 2006).

Para viabilizar sua utilização, o filtro deve ser convertido em uma versão FIR (Resposta Finita ao Impulso), através de aproximações que podem comprometer seu desempenho (Duarte, 2019). Além disso, o filtro deve aceitar um atraso fixo na resposta, quando comparado com o filtro inverso teórico.

3.2.1.2 Aproximação via Pseudo-Inversa de Moore-Penrose

O comportamento do canal e do filtro pode ser representado por sistemas lineares, nos quais a convolução é vista como uma multiplicação entre vetores e matrizes (Barbosa, 2017). Define-se, então, a matriz de convolução \mathbf{H} , composta por deslocamentos da resposta impulsiva \mathbf{h} :

$$\mathbf{H} = \begin{bmatrix} h_{-3} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ h_3 & \cdots & h_{-3} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_3 \end{bmatrix}$$

Com isso, a operação do canal pode ser expressa matricialmente como:

$$\mathbf{y} = \mathbf{H}\mathbf{a} \quad (3.14)$$

A obtenção da pseudo-inversa de Moore-Penrose permite calcular uma aproximação da matriz inversa \mathbf{W} :

$$\hat{\mathbf{a}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} = \mathbf{W}\mathbf{y} \quad (3.15)$$

Esse método resolve o problema dos mínimos quadrados, em que se busca minimizar a função de custo quadrática do erro entre a saída medida e a esperada. O resultado é equivalente ao da pseudo-inversa em sistemas sobredeterminados (Van Loan and Golub, 1996).

Contudo, essa abordagem pode induzir ao espalhamento de energia entre amostras consecutivas, gerando estimativas enviesadas das amplitudes de energia (Barbosa, 2017).

3.2.2 Filtro Baseado em Mínimos Quadrados

Uma alternativa foi desenvolvida por Duarte (2016), que propôs um filtro adaptativo linear com base no método dos mínimos quadrados (LMS - *Least Mean Squares*). A ideia é otimizar iterativamente os coeficientes do filtro de forma a minimizar a função de custo global:

$$\mathcal{J}(\mathbf{w}) = \sum_{k=0}^{N-1} (a[k] - \hat{a}[k])^2 \quad (3.16)$$

Em sua forma matricial, essa função é reescrita como:

$$\mathcal{J}(\mathbf{w}) = (\mathbf{a} - \mathbf{Y}\mathbf{w})^T (\mathbf{a} - \mathbf{Y}\mathbf{w}) \quad (3.17)$$

A matriz \mathbf{Y} é composta por N amostras e possui $M + 1$ colunas, sendo M a ordem do filtro FIR acrescida de uma coluna constante. O vetor de pesos ótimo \mathbf{w} é então obtido por:

$$\mathbf{w} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{a} \quad (3.18)$$

Embora essa estrutura seja simples e eficaz para dados simulados do TileCal, ela apresenta limitações na presença de *pile-up*, por não considerar componentes estatísticos de ordem superior, tornando-se sensível ao viés introduzido pelos dados de entrada (Duarte, 2016).

3.2.3 Filtro Baseado em Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) do tipo *feedforward* operam segundo uma arquitetura em que a informação é propagada unidirecionalmente, da camada de entrada até a saída, sem ciclos. Esse tipo de rede, originalmente classificado como um modelo de *Processamento Paralelo Distribuído*, foi proposto por Rumelhart et al. (1986). Quando se utiliza o modelo de neurônio do perceptron, essa estrutura recebe a designação de Perceptron Multicamada (MLP, do inglês *Multi-Layer Perceptron*).

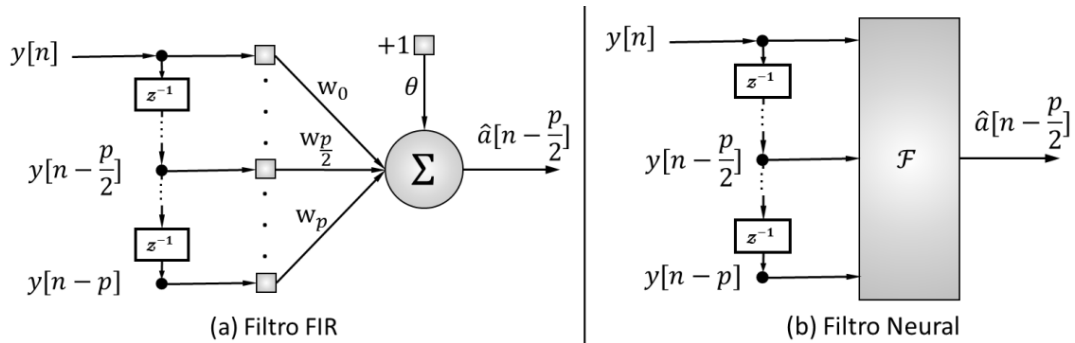
No contexto do TileCal, a aplicação de uma MLP como filtro de deconvolução foi analisada por Faria (2017), com o objetivo de substituir o filtro FIR treinado offline por Duarte (2016). A motivação para tal substituição está na habilidade da MLP de realizar tarefas de processamento de sinais quando se utiliza uma linha de atraso na entrada da rede, conforme proposto por Widrow (1988). Essa abordagem é denominada *Focused Time Lagged Feedforward Network* (FTLFFN), uma rede que incorpora memória temporal por meio de registradores de deslocamento, possibilitando a identificação implícita de padrões ao longo do tempo (Haykin, 2001).

As linhas de atraso, típicas em filtros digitais, determinam a ordem do filtro neural. De forma análoga ao filtro FIR (Figura 3.4a), o filtro neural (Figura 3.4b) emprega pesos associados às amostras anteriores para estimar a energia depositada. Neste contexto, a MLP assume o papel de filtro, em que os pesos substituem diretamente os coeficientes do FIR.

O valor estimado da energia $\hat{a}[n - p/2]$ é obtido a partir de uma combinação linear ponderada das amostras da janela de observação. A centralização na amostra $n - p/2$ deve-se ao formato típico do pulso detectado no canal, o que, por sua vez, impõe uma latência proporcional à metade da ordem do filtro.

Na construção do modelo, Faria (2017) investigou o impacto da variação no número de neurônios e na largura da janela de entrada (linha de atraso). Embora em muitas aplicações o número de entradas da rede não seja tratado como um hiperparâmetro, no caso

Figura 3.4 – Arquiteturas de filtragem: (a) Filtro FIR tradicional; (b) Filtro neural MLP com linha de atraso.



Fonte: Adaptado de Faria (2017)

específico da filtragem neural, o tamanho da janela precisa ser cuidadosamente definido.

A MLP é treinada a partir de pares entrada/saída $\{\mathbf{y}, \mathbf{a}\}$ obtidos via simulações TMC, permitindo à rede capturar a distribuição estatística dos sinais. O treinamento é realizado utilizando o algoritmo de retropropagação do erro, frequentemente com o método de Levenberg-Marquardt como otimizador, que combina características do gradiente descendente e do método de Gauss-Newton:

$$\mathbf{W}_{i+1} = \mathbf{W}_i - (\mathbf{J}^T \mathbf{J} + \mu \mathbf{I})^{-1} \nabla E_i \quad (3.19)$$

Neste contexto, \mathbf{W} representa os pesos da rede, \mathbf{J} o jacobiano do erro, μ um fator de amortecimento e ∇E_i o gradiente do erro na iteração i .

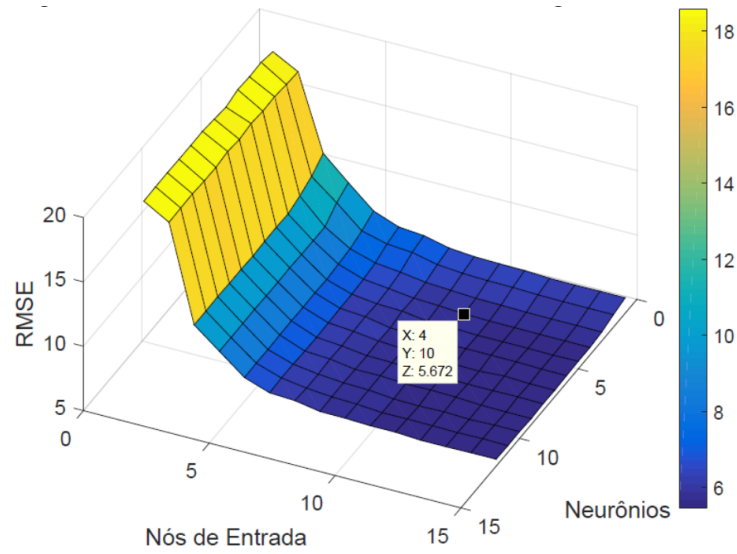
Para garantir a capacidade de generalização da rede, os dados foram divididos em três conjuntos: treino (para ajuste dos pesos), validação (para evitar sobreajuste) e teste (para avaliação final do desempenho). A configuração ideal foi determinada por uma busca exaustiva em diferentes combinações de parâmetros, conforme ilustrado na Figura 3.5. A arquitetura escolhida, 10-4-1, corresponde a 10 entradas, 4 neurônios ocultos e uma saída.

Apesar dos bons resultados obtidos, a principal limitação do modelo está na forte dependência dos dados simulados. Isso impõe um desafio considerável, pois a precisão da estimativa depende diretamente da fidelidade do simulador em representar todos os aspectos relevantes do fenômeno físico.

3.2.3.1 Interpretação Alternativa: Filtros FIR em Redes Neurais

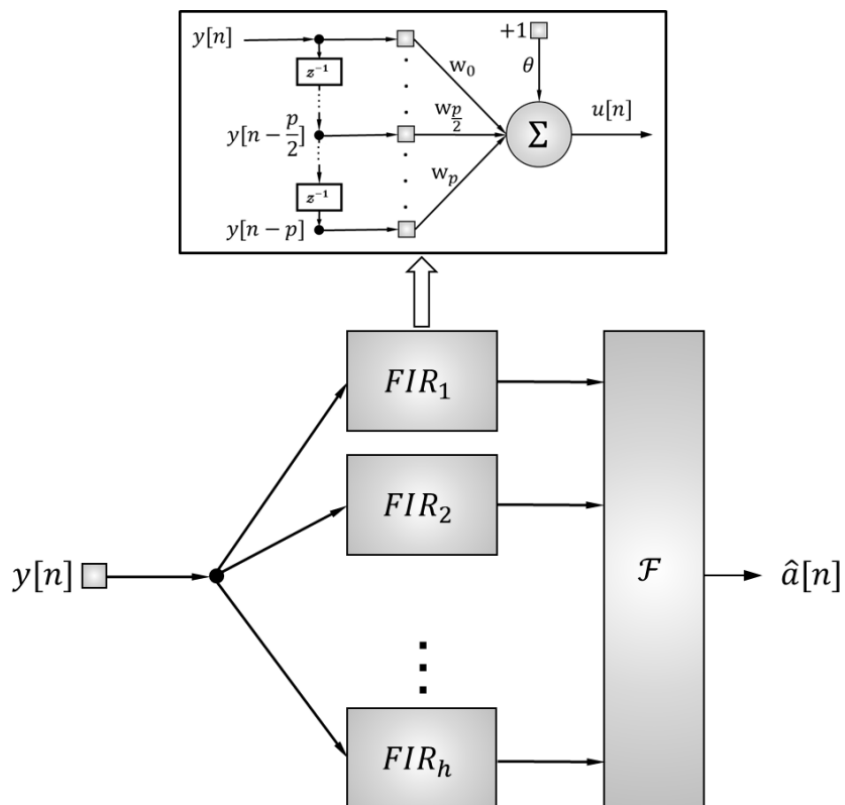
Uma maneira complementar de compreender o funcionamento das MLPs na filtragem de sinais é considerá-las como uma composição de filtros FIR lineares (um por neurônio da camada de entrada), seguidos por uma camada de ativação que introduz a não linearidade (Wan, 1991), conforme ilustrado na Figura 3.6. Essa estrutura é conhecida como *Finite Impulse Response Neural Network* (FIRNN) (Haykin, 2001).

Figura 3.5 – Busca exaustiva da configuração ótima da RNA.



Fonte: (Faria, 2017)

Figura 3.6 – Rede Neural com Resposta Finita ao Impulso (FIRNN).

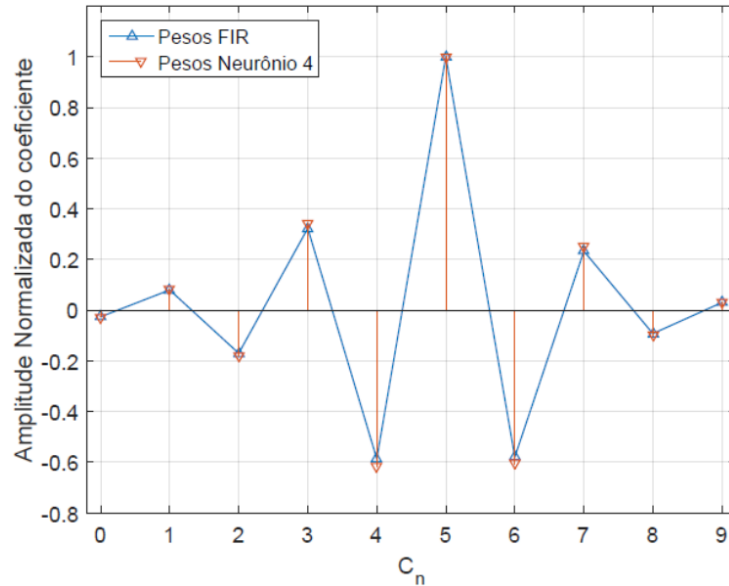


Fonte: Adaptado de Haykin (2001)

Faria (2017) demonstra experimentalmente essa interpretação, verificando que os pesos de um dos neurônios convergem para os coeficientes do filtro FIR treinado com mínimos quadrados, como mostrado na Figura 3.7. Esse neurônio atua majoritariamente

na região linear da resposta, sendo o mais influente na saída da rede (Tabela 3.1).

Figura 3.7 – Comparação entre os coeficientes do filtro FIR e os de um neurônio da RNA.



Fonte: (Faria, 2017)

Tabela 3.1 – Influência sináptica absoluta de cada neurônio.

Neurônio	Peso Absoluto
1	0.09
2	0.03
3	0.03
4	0.85

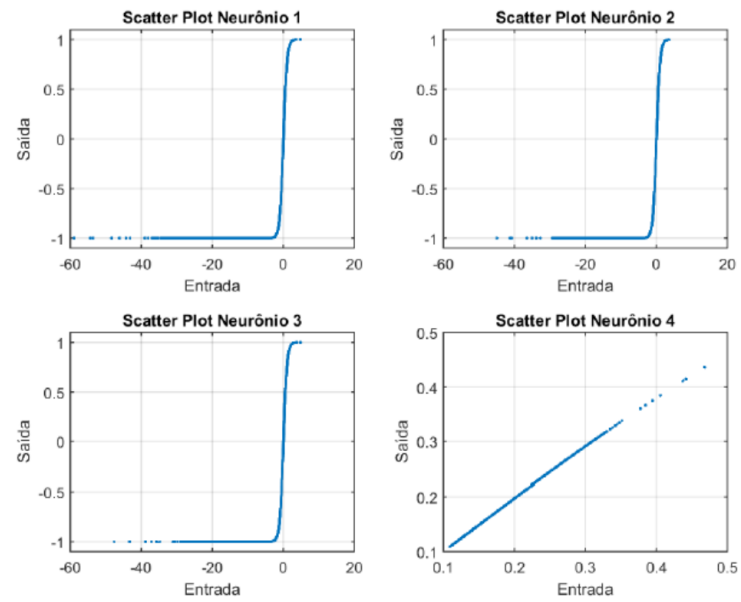
Fonte: (Faria, 2017)

Nota-se, à partir da Figura 3.8, que enquanto um dos neurônios responde de forma linear, os demais ajustam componentes não lineares do sinal, fornecendo à rede a flexibilidade necessária para superar os métodos clássicos na tarefa de estimação de energia.

3.3 CONSIDERAÇÕES PARCIAIS

Os resultados apresentados indicam que, embora os filtros lineares representem soluções robustas e computacionalmente eficientes para mitigação de *pile-up*, métodos baseados em redes neurais demonstram maior capacidade de modelagem para sinais não lineares, como os observados no TileCal. No entanto, a alta dependência de dados simulados impõe uma limitação crítica à aplicabilidade prática. Essa restrição motiva a busca por abordagens híbridas ou novos treinamentos menos sensíveis à acurácia do simulador.

Figura 3.8 – Mapeamento entrada-saída dos neurônios.



Fonte: (Faria, 2017)

4 TEORIA DE ESPARSIDADE APLICADA À PROCESSAMENTO DE SINAIS

A representação esparsa aplicada à processamento de sinais é discutida na literatura nas últimas décadas, com avanços significativos e resultados promissores. Neste capítulo serão discutidos a teoria de sinais esparsos, bem como algumas aplicações no cenário de reconstrução de sinais em física de altas energias.

4.1 REPRESENTAÇÕES ESPARSAS

Nas últimas décadas, o conceito de esparsidade tem se consolidado como uma ferramenta fundamental em modelagem matemática, processamento de sinais e aprendizado de máquina, impulsionado principalmente pelos avanços em algoritmos de otimização e pela crescente demanda por representação eficiente de dados (Elad, 2010). O termo “representação esparsa” refere-se à capacidade de expressar um sinal ou vetor com poucos coeficientes não nulos em relação ao seu comprimento total. Essa premissa está alinhada com a observação de que muitos sinais naturais, embora complexos no domínio temporal ou espacial, tornam-se significativamente mais simples quando representados em uma base apropriada.

Inicialmente, os estudos focaram no desenvolvimento de algoritmos para obtenção de soluções esparsas de sistemas lineares subdeterminado, com destaque para técnicas como *Matching Pursuit* (MP), *Orthogonal Matching Pursuit* (OMP) e métodos baseados em relaxações convexas, como o ℓ_1 -minimization (Tropp, 2004). Posteriormente, a teoria da esparsidade evoluiu para oferecer garantias formais, como o conceito de *Restricted Isometry Property* (RIP), que estabelece condições sobre a matriz de medição para que a reconstrução esparsa seja possível com alta probabilidade (Candes and Tao, 2005).

A importância da esparsidade transcende a álgebra linear e se estende a diversas áreas aplicadas. No processamento de imagens, ela permite a compressão e restauração de dados com mínima perda de informação. Em deconvolução, a esparsidade facilita a reconstrução de sinais originais a partir de observações ruidosas ou degradadas. Também é um dos pilares teóricos da amostragem compressiva (*compressive sensing*), que permite recuperar sinais esparsos a partir de um número muito menor de amostras do que o exigido pelo Teorema de Nyquist (Donoho, 2006).

Formalmente, considere um vetor $\Omega \in \mathbb{R}^n$, definido por $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_n\}$. Este vetor é dito estritamente esparsa se a maioria de seus elementos for nula, ou seja, se o seu suporte $\Lambda(\Omega) = \{i \in \{1, \dots, n\} \mid \Omega_i \neq 0\}$ possuir cardinalidade k tal que $k \ll n$. Na prática, mesmo vetores não estritamente esparsos podem admitir uma aproximação esparsa, em que apenas os maiores coeficientes são mantidos, desprezando-se os demais por serem insignificantes em termos energéticos.

Adicionalmente, muitos sinais que não apresentam esparsidade em seu domínio original podem ser representados de maneira esparsa em um espaço transformado adequado, como nas bases de Fourier, wavelets, coseno discreto (DCT), ou em dicionários aprendidos de forma adaptativa (Starck et al., 2010). Essas representações são cruciais em tarefas como codificação, compressão, filtragem e aprendizado supervisionado ou não supervisionado. O uso de representações esparsas também está fortemente ligado a métodos de aprendizado profundo, onde mecanismos como regularização ℓ_1 , autoencoders esparsos e codificações por dicionário têm-se mostrado eficazes na obtenção de representações interpretáveis e generalizáveis (Chen and Guo, 2023; Rajamanoharan et al., 2024; Xiong et al., 2025).

Portanto, a modelagem esparsa não apenas oferece uma maneira eficiente de representar e processar informações, como também introduz uma estrutura matemática que permite explorar a redundância inerente aos dados naturais. Essa abordagem tem sido uma das bases para diversas inovações em tecnologias de sensoriamento, comunicação e inteligência artificial Zhang et al. (2016).

Para o caso específico de reconstrução da energia da colisão dos aceleradores de partículas, a representação esparsa dos sinais está se mostrando um caminho promissor para lidar com o ruído, recuperando os sinais impulsivos desejados no fim do processo.

4.2 ESTIMATIVA DE ENERGIA VIA ESPARSIDADE

O modelo convolucional do calorímetro, apresentado anteriormente na Equação (4.1), pode ser representado de forma matricial ao negligenciarmos o ruído eletrônico. Nessa representação, a convolução é expressa por:

$$\mathbf{H}\mathbf{a} = \mathbf{y} \tag{4.1}$$

onde \mathbf{H} é uma matriz de convolução com N linhas e M colunas ($N > M$), composta por colunas formadas por versões deslocadas da resposta ao impulso do calorímetro; \mathbf{a} representa o vetor de amplitudes dos pulsos de referência; e \mathbf{y} é o vetor de observações obtido do calorímetro.

Por se tratar de um sistema indeterminado, a equação possui infinitas soluções. O objetivo é encontrar a solução mais esparsa possível, ou seja, aquela em que o vetor \mathbf{a} possua o menor número de coeficientes distintos de zero.

Esse problema pode ser reformulado como uma tarefa de otimização, onde se busca minimizar uma função custo que promova esparsidade. Uma forma direta de fazer isso é utilizando a *norma* ℓ_0 , que conta o número de elementos não nulos em \mathbf{a} . De maneira geral, a norma- p de um vetor é dada por:

$$\|\mathbf{a}\|_p = \left(\sum_{i=1}^M |a_i|^p \right)^{1/p}, \quad \text{para } p > 0 \quad (4.2)$$

enquanto que a norma- ℓ_0 é definida como:

$$\|\mathbf{a}\|_0 = \#\{i : a_i \neq 0\} \quad (4.3)$$

A formulação de otimização com base na norma ℓ_0 leva aos seguintes problemas:

$$(P_0) : \min_{\mathbf{a}} \|\mathbf{a}\|_0 \quad \text{sujeito a } \mathbf{H}\mathbf{a} = \mathbf{y} \quad (4.4)$$

$$(P_{0,\varepsilon}) : \min_{\mathbf{a}} \|\mathbf{a}\|_0 \quad \text{sujeito a } \|\mathbf{H}\mathbf{a} - \mathbf{y}\|_2^2 \leq \varepsilon \quad (4.5)$$

onde ε representa um limite aceitável de erro, permitindo certa tolerância à presença de ruído.

Resolver exatamente o problema com a norma ℓ_0 é extremamente custoso do ponto de vista computacional. Assim, uma estratégia alternativa envolve a busca inicial pelo suporte do vetor esparso (posições não nulas), seguida de uma estimativa dos valores desses coeficientes através de regressão linear. Algoritmos que seguem essa abordagem são conhecidos como algoritmos vorazes (greedy), como Matching Pursuit (MP) e Orthogonal Matching Pursuit (OMP).

Outra estratégia bastante explorada é a substituição da norma ℓ_0 pela norma ℓ_1 , que relaxa o problema, tornando-o convexo e mais tratável numericamente. Esse novo problema é formulado como:

$$(P_{1,\varepsilon}) : \min_{\mathbf{a}} \|\mathbf{a}\|_1 \quad \text{sujeito a } \|\mathbf{H}\mathbf{a} - \mathbf{y}\|_2^2 \leq \varepsilon \quad (4.6)$$

onde $\|\mathbf{a}\|_1 = \sum_{i=1}^M |a_i|$ representa a soma dos valores absolutos dos coeficientes.

A versão sem restrições deste problema pode ser formulada através da introdução de um multiplicador de Lagrange λ , resultando na seguinte função custo:

$$J(\mathbf{a}) = \lambda \|\mathbf{a}\|_1 + \frac{1}{2} \|\mathbf{H}\mathbf{a} - \mathbf{y}\|_2^2 \quad (4.7)$$

O parâmetro λ atua como fator de ponderação entre a esparsidade e a fidelidade da reconstrução ao sinal original. Sua escolha influencia diretamente o equilíbrio entre os dois termos e pode ser ajustada empiricamente.

Existem diversas técnicas para minimizar $J(\mathbf{a})$. Uma abordagem simples e eficaz, especialmente útil para filtragem de ruído, é a técnica de *shrinkage* ou limiarização suave, proposta por Donoho (1995). Essa técnica se baseia na separação da função custo em

componentes individuais para cada entrada de \mathbf{a} , permitindo a aplicação de soluções otimizadas por variável.

Uma solução apresentada por Elad (2010) é a utilização de um algoritmo iterativo desenvolvido a partir da Equação (4.7), onde é utilizada a função *shrinkage* recursivamente para rastrear a solução mais próxima desejada. Por se tratar de um processo iterativo, a velocidade de convergência dificulta sua implementação em sistemas de estimação *online* de energia.

Duarte (2019) demonstra um método também derivado da mesma Equação (4.7), preservando a atuação de limiarização, mas que envolve o uso de um filtro FIR de deconvolução. O método permite o uso das representações esparsas em um ambiente *free-running* de dados.

4.2.1 Método Iterativo

Através da Equação (4.7) de minimização, Elad (2010) considera o caso onde a matriz \mathbf{H} é uma matriz unitária. Neste caso, a transformação mantém a norma euclidiana do sinal, sendo sua inversa dada pela transposta da matriz de convolução ou \mathbf{H}^T . Desta forma, as amostras vizinhas podem ser desacopladas e o resíduo definido pelo termo $\|\mathbf{H}\mathbf{a} - \mathbf{y}\|_2^2$ pode ser simplificado. Considerando a estimação inicial $\mathbf{a}_0 = \mathbf{H}^T \mathbf{y}$, a Equação de minimização é expressa como:

$$J(\mathbf{a}) = \lambda \|\mathbf{a}\|_1 + \frac{1}{2} \|\mathbf{a}_0 - \mathbf{a}\|_2^2 \quad (4.8)$$

A função *shrinkage* é um método eficaz para resolver problemas de minimização como o proposto. Esse tipo de função é obtida derivando a Equação (4.8) em relação a \mathbf{a} e igualando o resultado a zero. Isolando-se o termo em questão, permite-se que os sinais encontrados de \mathbf{a} sejam a solução mais esparsa possível do sinal. A função *shrinkage* é determinada então por:

$$\mathbf{a} = S_\lambda(\mathbf{a}_0) = \begin{cases} a_0[n] - \lambda, & a_0[n] > \lambda \\ 0, & -\lambda \leq a_0[n] \leq \lambda \\ a_0[n] + \lambda, & a_0[n] < -\lambda \end{cases} \quad (4.9)$$

Por considerar a matriz \mathbf{H} como unitária, restringe-se a solução a problemas muito específicos, que não são o caso da estimação de energia em física de altas energias, por exemplo. Desta forma, Elad (2010) propõe um método iterativo conhecido como *Iterative Soft Thresholding Algorithm* (ISTA) que mantém o uso da função *shrinkage* encontrada, simplificando o processo. O uso de uma função de suporte permite encontrar uma solução próxima à da função custo original, levando a um processo eficiente e que converge para a

solução desejada. A Equação (4.10) define o processo iterativo, onde c é uma constante necessária para que a função suporte seja estritamente convexa.

$$\mathbf{a}_{k+1} = S_\lambda \left(\frac{1}{c} \mathbf{H}^T (\mathbf{y} - \mathbf{H} \mathbf{a}_k) + \mathbf{a}_k \right) \quad (4.10)$$

O processo iterativo é realizado até que a diferença entre iterações consecutivas atinja um valor menor que um limite pré-estabelecido. Este método apresenta resultados satisfatórios, mas por se tratar de um processo iterativo, o uso do mesmo em ambientes com processamento *online* dos sinais, onde o tempo de resposta é restrito, não é adequado. Essa limitação não o credencia como uma solução para o sistema de *trigger* apresentado anteriormente.

4.2.2 Método adaptado para processamento *online* dos sinais

Duarte (2019) demonstra um método derivado da Equação (4.7) onde se permite usar a teoria de esparsidade no sistema de *trigger*. Partindo da hipótese de que a matriz \mathbf{H} admite inversa, ou seja, $\mathbf{W}_0 = \mathbf{H}^{-1}$, é possível reescrever o sistema linear original $\mathbf{y} = \mathbf{H} \mathbf{a}$ como $\mathbf{a} = \mathbf{H}^{-1} \mathbf{y}$. A partir disso, pode-se formular uma função de custo do tipo:

$$J_o(\mathbf{a}) = \lambda \|\mathbf{a}\|_1 + \frac{1}{2} \|\mathbf{a} - \mathbf{W}_0 \mathbf{y}\|_2^2, \quad (4.11)$$

ou, de forma equivalente, ao definir $\mathbf{a}_0 = \mathbf{W}_0 \mathbf{y}$, tem-se:

$$J_o(\mathbf{a}) = \lambda \|\mathbf{a}\|_1 + \frac{1}{2} \|\mathbf{a} - \mathbf{a}_0\|_2^2. \quad (4.12)$$

No contexto da calorimetria, onde se esperam somente sinais com valores positivos, pode-se restringir a análise às componentes $a_k \geq 0$, o que permite eliminar o módulo da norma. Derivando a função em relação a \mathbf{a} e igualando o resultado a zero para os casos em que $a_k > 0$, obtém-se:

$$0 = (\mathbf{a} - \mathbf{a}_0) + \lambda \cdot \mathbf{1}, \quad \text{para } a_k > 0. \quad (4.13)$$

Assim, a solução que minimiza a função de custo é dada por:

$$\mathbf{a} = \mathbf{a}_0 - \lambda \cdot \mathbf{1}, \quad \text{para } a_k > 0. \quad (4.14)$$

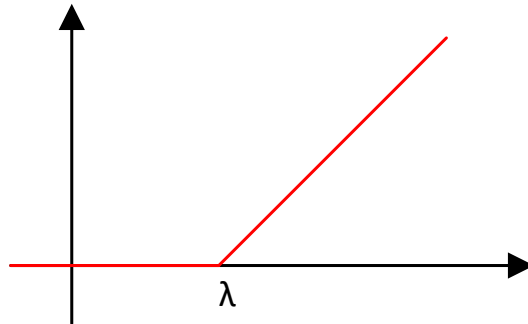
Considerando que \mathbf{w}_{0k} representa a k -ésima linha da matriz \mathbf{W}_0 , pode-se expressar a componente $a[k]$ como:

$$a[k] = \mathbf{w}_{0k} \mathbf{y} - \lambda, \quad \text{se } a[k] > 0. \quad (4.15)$$

onde \mathbf{w}_{0k} é um vetor do tipo linha.

Dessa maneira, o vetor \mathbf{y} pode ser interpretado como um sinal de entrada submetido a um filtro FIR, com w_{0k} sendo os coeficientes do mesmo, ou seja, uma forma linear de realizar a deconvolução do sinal. No entanto, a saída do filtro deve ser ajustada por uma subtração da constante λ , seguida de uma limitação inferior em zero, devido às restrições físicas de amplitude. Esse comportamento pode ser descrito por uma função não linear composta por uma subtração seguida de um operador de limiarização, conforme ilustrado na Figura 4.1.

Figura 4.1 – Representação gráfica da função *Shrinkage* modificada para o caso de física de altas energias



Fonte: Do autor

No entanto, como \mathbf{H} é a matriz de convolução, sua natureza não quadrada impede a inversibilidade direta. Para contornar essa limitação, utiliza-se a pseudo-inversa, aproximando-se \mathbf{W} como:

$$\mathbf{W} \approx \mathbf{H}^+ = (\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}. \quad (4.16)$$

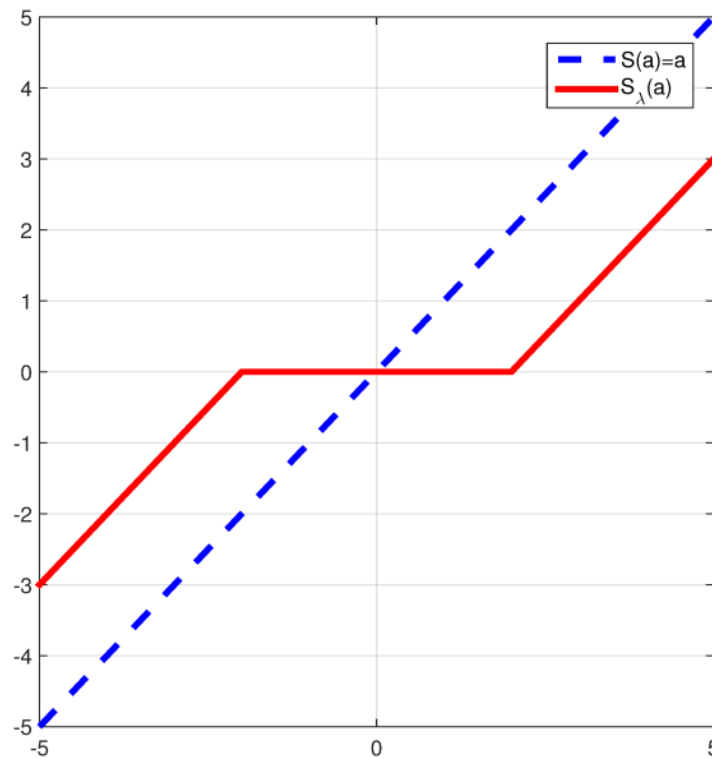
É importante salientar que, quanto maior for a dimensão do vetor de entrada \mathbf{y} , melhor será a aproximação da pseudo-inversa. Isso implica que o filtro FIR associado deve possuir ordem elevada para garantir desempenho adequado na reconstrução por deconvolução.

Generalizando esse procedimento para além do domínio da calorimetria, introduz-se a função *Shrinkage*, denotada como $S_\lambda(a)$, a qual realiza uma compressão seletiva dos coeficientes. Essa função é definida como:

$$S_\lambda(a) = \begin{cases} a - \lambda, & \text{se } a > \lambda, \\ 0, & \text{se } |a| \leq \lambda, \\ a + \lambda, & \text{se } a < -\lambda. \end{cases} \quad (4.17)$$

A Figura 4.2 apresenta a forma da função *Shrinkage* para $\lambda = 2$, evidenciando a diferença em relação a uma função linear simples.

Figura 4.2 – Representação gráfica da função *Shrinkage* com $\lambda = 2$ (linha contínua), em comparação com uma função identidade (linha pontilhada).



Fonte: (Duarte, 2019)

A aplicação desta função após a filtragem permite eliminar componentes de baixa magnitude, que podem ser atribuídas a ruído e a artefatos oriundos do processo de deconvolução linear, ao mesmo tempo em que preserva os elementos significativos do sinal. Esse método, portanto, associa filtragem linear com uma etapa de processamento não linear simples e eficiente, viabilizando sua implementação em sistemas embarcados para reconstrução de energia.

Por considerar um limiar (λ) fixo para todas as amostras, vão existir situações em que o sinal de interesse será filtrado por ser considerado ruído. Vão existir casos também onde o ruído será maior que o limiar, não filtrando corretamente como desejado. Um método onde se considerasse um limiar adaptativo para cada amostra analisada permitiria um processamento mais eficiente, aproximando-se mais da energia original que se deseja

estimar. Este método proposto, que será discutido no próximo capítulo, é o tema central desta tese e combina o método iterativo com algumas ideias apresentadas por Duarte (2019).

4.3 CONCLUSÕES PARCIAIS DO CAPÍTULO

No presente capítulo, foi possível entender como a teoria de esparsidade pode ser aplicada a processamento de sinais. O modelo do calorímetro pode ser comparado ao modelo convolucional demonstrado e a função *shrinkage* é um importante passo para a filtragem de ruído presente no processo.

O método iterativo não possibilita uma utilização em um cenário de processamento em tempo real dos sinais, e o método adaptado pode ser melhorado. No próximo capítulo será demonstrado um novo método a partir do método iterativo que também é possível de ser aplicado em sistemas *online* de estimação de energia.

5 MÉTODO PROPOSTO

Duarte (2019) apresentou um método baseado na representação esparsa de sinais para a estimação de energia em calorímetros, adequado também para aplicações em tempo real. Neste capítulo, discute-se um método que combina o conceito de processamento *online* com o processo iterativo proposto por Elad (2010). Dessa forma, espera-se aprimorar a reconstrução da energia.

5.1 EXPANDINDO O PROCESSO ITERATIVO ISTA

Na Seção 4.2.1, foi discutido o método conhecido como ISTA, fundamentado no conceito de representação esparsa de sinais. Ao expandir o processo iterativo e introduzir algumas considerações adicionais, podem-se obter conclusões relevantes.

Primeiramente, a função *shrinkage*, representada pela Equação (4.9), pode ser modificada a partir da mesma consideração de Duarte (2019). Como a energia das partículas não assume valores negativos, não há justificativa para considerar a última hipótese da referida equação. Assim, a função *shrinkage* pode ser definida como:

$$\mathbf{a} = S_\lambda(\mathbf{a}_0) = \begin{cases} a_0[n] - \lambda, & a_0[n] > \lambda \\ 0, & a_0[n] \leq \lambda \end{cases} \quad (5.1)$$

A partir da Equação (4.10), é possível determinar as primeiras iterações. A primeira consideração consiste em adotar a_0 como valor inicial, cuja escolha será discutida posteriormente. A Equação (5.2) corresponde à primeira iteração do processo, lembrando que $S_\lambda(\cdot)$ representa a função *shrinkage* definida anteriormente, e que λ é um fator fixo relacionado ao ruído envolvido no processo.

$$\mathbf{a}_1 = S_\lambda \left(\frac{1}{c} \mathbf{H}^T (\mathbf{y} - \mathbf{H} \mathbf{a}_0) + \mathbf{a}_0 \right) \quad (5.2)$$

Pode-se realizar uma substituição de variável, conforme a Equação (5.3), em que Δ_0 é um vetor que atualiza os valores de a_0 . Aplicando a função *shrinkage* com essa substituição de variável e analisando a operação em cada amostra, obtém-se a Equação (5.4).

$$\Delta_0 = \frac{1}{c} \mathbf{H}^T (\mathbf{y} - \mathbf{H} \mathbf{a}_0) \quad (5.3)$$

$$a_1[n] = S_\lambda(\Delta_0[n] + a_0[n]) = \begin{cases} \Delta_0[n] + a_0[n] - \lambda, & \Delta_0[n] + a_0[n] > \lambda \\ 0, & \Delta_0[n] + a_0[n] \leq \lambda \end{cases} \quad (5.4)$$

Uma nova substituição de variável pode ser realizada, conforme a Equação (5.5). Dessa forma, a função *shrinkage* aplicada em cada amostra é representada pela Equação (5.6).

$$\lambda_1 = \lambda - \Delta_0 \quad (5.5)$$

$$a_1[n] = S_\lambda(\Delta_0[n] + a_0[n]) = \begin{cases} a_0[n] - \lambda_1[n], & a_0[n] > \lambda_1[n] \\ 0, & a_0[n] \leq \lambda_1[n] \end{cases} \quad (5.6)$$

Uma conclusão importante desse passo é que, nesse contexto, cada amostra apresenta um valor de limiar distinto. Esse limiar depende de \mathbf{y} e de \mathbf{a}_0 . A forma exata dessa função não é relevante neste momento, sendo de interesse apenas as variáveis que a compõem, uma vez que, posteriormente, será apresentada uma alternativa para determinar essa relação.

Com a primeira iteração estabelecida, pode-se analisar o próximo valor de energia gerado pelo processo iterativo. A operação correspondente está descrita na Equação (5.7), em que a mesma substituição de variável pode ser aplicada (Equação (5.8)), permitindo a aplicação da função *shrinkage* conforme a Equação (5.9). A operação amostra a amostra é, então, expressa na Equação (5.10).

$$\mathbf{a}_2 = S_\lambda \left(\frac{1}{c} \mathbf{H}^T (\mathbf{y} - \mathbf{H} \mathbf{a}_1) + \mathbf{a}_1 \right) \quad (5.7)$$

$$\Delta_1 = \frac{1}{c} \mathbf{H}^T (\mathbf{y} - \mathbf{H} \mathbf{a}_1) \quad (5.8)$$

$$\mathbf{a}_2 = S_\lambda(\Delta_1 + \mathbf{a}_1) \quad (5.9)$$

$$a_2[n] = S_\lambda(\Delta_1[n] + a_1[n]) = \begin{cases} \Delta_1[n] + a_1[n] - \lambda, & \Delta_1[n] + a_1[n] > \lambda \\ 0, & \Delta_1[n] + a_1[n] \leq \lambda \end{cases} \quad (5.10)$$

O valor de \mathbf{a}_1 foi explicitado anteriormente. Um aspecto relevante é que, por envolver uma função não linear, existem duas possibilidades para essa variável. Considerando $a_1[n] = a_0[n] - \lambda_1[n]$, a aplicação da função *shrinkage* é representada pela Equação (5.11).

$$a'_2[n] = \begin{cases} \Delta_1[n] + a_0[n] - \lambda_1[n] - \lambda, & \Delta_1[n] + a_0[n] - \lambda_1[n] > \lambda \\ 0, & \Delta_1[n] + a_0[n] - \lambda_1[n] \leq \lambda \end{cases} \quad (5.11)$$

Recorrendo novamente a uma substituição de variável, conforme a Equação (5.12), o valor de $a_2[n]$ para essa primeira possibilidade é obtido pela Equação (5.13).

$$\lambda'_2 = \lambda + \lambda_1 - \mathbf{\Delta}_1 \quad (5.12)$$

$$a'_2[n] = S_\lambda(\Delta_1[n] + a_1[n]) = \begin{cases} a_0[n] - \lambda'_2[n], & a_0[n] > \lambda'_2[n] \\ 0, & a_0[n] \leq \lambda'_2[n] \end{cases} \quad (5.13)$$

Mais uma vez, observa-se que cada amostra possui um limiar distinto, o qual depende apenas de \mathbf{y} e \mathbf{a}_0 . Outra conclusão relevante é que o valor de $a_2[n]$ está relacionado unicamente ao valor inicial \mathbf{a}_0 , tornando a primeira iteração (\mathbf{a}_1) transparente no processo.

Para completar a análise, considera-se $a_1[n] = 0$, que corresponde à segunda possibilidade de valor para $a_1[n]$. Nesse caso, pode-se afirmar que $a_1[n] = a_0[n] - a_0[n]$. Assim, a operação que conduz à segunda possibilidade para $a''_2[n]$ é expressa como:

$$a''_2[n] = \begin{cases} \Delta_1[n] + a_0[n] - a_0[n] - \lambda, & \Delta_1[n] + a_0[n] - a_0[n] > \lambda \\ 0, & \Delta_1[n] + a_0[n] - a_0[n] \leq \lambda \end{cases} \quad (5.14)$$

Nesse caso específico, é possível definir λ''_2 como na Equação (5.15). Assim, a expressão final para essa consideração é apresentada na Equação (5.16).

$$\lambda''_2 = \lambda + \mathbf{a}_0 - \mathbf{\Delta}_1 \quad (5.15)$$

$$a''_2[n] = S_\lambda(\Delta_1[n] + a_1[n]) = \begin{cases} a_0[n] - \lambda''_2[n], & a_0[n] > \lambda''_2[n] \\ 0, & a_0[n] \leq \lambda''_2[n] \end{cases} \quad (5.16)$$

Nessa segunda análise, o limiar difere do obtido anteriormente, mas mantém a mesma premissa de depender apenas de \mathbf{y} e \mathbf{a}_0 . Como não é necessário conhecer a relação exata entre essas variáveis para determinar os valores de limiar, pode-se unificar as duas possibilidades conforme:

$$a_2[n] = S_\lambda(\Delta_1[n] + a_1[n]) = \begin{cases} a_0[n] - \lambda_2[n], & a_0[n] > \lambda_2[n] \\ 0, & a_0[n] \leq \lambda_2[n] \end{cases} \quad (5.17)$$

onde λ_2 é o vetor de limiar associado a cada amostra, sendo uma função não linear com entradas \mathbf{y} e \mathbf{a}_0 .

O processo iterativo pode prosseguir, realizando-se substituições de variáveis como as efetuadas anteriormente. Considerando que a atualização do valor de \mathbf{a} atingiu sua etapa final, essa última iteração pode ser expressa como:

$$a_N[n] = S_\lambda(\Delta_N[n] + a_N[n]) = \begin{cases} a_0[n] - \lambda_N[n], & a_0[n] > \lambda_N[n] \\ 0, & a_0[n] \leq \lambda_N[n] \end{cases} \quad (5.18)$$

Esse resultado é relevante, pois demonstra que é possível determinar o valor final do processo iterativo a partir do valor inicial \mathbf{a}_0 e do vetor de limiar λ_N de cada amostra. O limiar é resultado de uma cascata de funções não lineares (funções *shrinkage*) que dependem de \mathbf{a}_0 e \mathbf{y} . Embora não seja viável obter uma fórmula fechada para essa função, caso tais valores possam ser determinados, é possível alcançar o mesmo resultado do processo iterativo por meio de um processamento direto do tipo *feedforward*, conforme proposto por Duarte (2019).

A definição eficiente do valor inicial \mathbf{a}_0 , bem como a correta determinação do limiar, possibilitam o desenvolvimento de um método baseado em representações esparsas de sinais. Esse método se inspira no processo iterativo proposto por Elad (2010), com vistas à aplicação em sistemas de processamento *online*. Nas seções seguintes, serão apresentadas estratégias para a determinação eficiente desses dois parâmetros.

5.2 O FILTRO FIR COMO VALOR INICIAL DO PROCESSO

O filtro FIR é amplamente empregado em processamento digital de sinais devido à sua fácil implementação, estabilidade e ao baixo custo em termos de recursos computacionais (Oppenheim et al., 2010). Sua utilização como valor inicial no processo descrito anteriormente constitui uma estratégia relevante.

Na Seção 4.2, foi apresentado o modelo convolucional do calorímetro, que relaciona a matriz de convolução, o vetor de energias e a leitura do calorímetro. Como tanto a matriz de convolução quanto a leitura do calorímetro são conhecidas, essas informações podem ser empregadas para estimar o valor inicial do processo descrito.

A matriz \mathbf{H} não é quadrada e, portanto, não admite inversão direta. Nesse contexto, a pseudo-inversa de Moore-Penrose surge como uma solução para obter uma aproximação da matriz de deconvolução, representada por \mathbf{W} neste método. Assim, o valor de \mathbf{a}_0 pode ser calculado como:

$$\mathbf{a}_0 = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H} \mathbf{y} = \mathbf{W} \mathbf{y} \quad (5.19)$$

Utilizando a mesma ideia apresentada por Duarte (2019), ao observar a k -ésima linha da matriz \mathbf{W} representada por \mathbf{w}_k , o termo $a_0[k]$ pode ser determinado como:

$$a_0[k] = \mathbf{w}_k \cdot \mathbf{y} \quad (5.20)$$

A Equação (5.20) estabelece a relação entre os sinais de leitura do calorímetro e um vetor linha contendo os pesos correspondentes a cada amostra. Essa operação equivale exatamente à aplicação de um filtro FIR. Ao analisar uma linha estratégica da matriz de deconvolução \mathbf{W} , observa-se que o filtro FIR se aproxima do sistema inverso da resposta em frequência do canal e, dependendo da dimensão da matriz de convolução, tende a convergir para os valores do filtro de Wiener.

Assim, a proposta do método consiste em obter o valor inicial $a_0[n]$ por meio da saída de um filtro FIR, cujos pesos correspondem à k -ésima linha da pseudo-inversa da matriz de convolução, sendo a entrada do filtro a leitura do calorímetro. A determinação do tamanho do filtro FIR e do *buffer* associado à leitura do calorímetro exige uma análise dos recursos computacionais utilizados, do atraso de processamento e da performance para diferentes configurações. Essa análise será desenvolvida no próximo capítulo.

5.3 REDE NEURAL PARA DETERMINAÇÃO DO LIMIAR DA FUNÇÃO *SHRINKAGE*

Na seção 5.1 foi discutido o termo $\lambda_N[n]$, que corresponde ao limiar da função *shrinkage* para cada amostra de energia a ser estimada. Também foi mencionado que não é possível obter uma fórmula fechada para a determinação desse parâmetro, uma vez que envolve a aplicação sequencial de diversas funções não lineares.

Este trabalho propõe o uso de uma rede neural MLP do tipo *feed-forward* para estimar o valor do limiar desejado. Esse tipo de rede apresenta implementação relativamente simples, pois não requer mecanismos recursivos. Tal característica é fundamental em cenários de processamento em tempo real, nos quais o tempo de resposta é restrito.

Ainda que não se conheça uma formulação fechada para a função que mapeia o limiar, suas variáveis de entrada são conhecidas. Durante a expansão do método iterativo apresentada na seção 5.1, foi destacado que o limiar depende exclusivamente das variáveis \mathbf{a}_0 e \mathbf{y} , que correspondem, respectivamente, à estimativa inicial da energia e às leituras do calorímetro.

Dessa forma, uma rede neural MLP, tendo como entradas os valores iniciais de energia estimada e os sinais registrados pelo calorímetro, pode ser empregada para prever o limiar da função *shrinkage* associado a cada energia a ser estimada. Considerando que o objetivo é a aplicação em um sistema *free-running*, a definição da janela de observação desses sinais torna-se um aspecto crítico. Uma escolha inadequada pode aumentar a complexidade da rede neural sem necessidade.

Outros fatores relevantes no desempenho da rede são o número de camadas ocultas e a quantidade de neurônios em cada uma delas. Uma configuração inadequada pode comprometer a predição dos valores de limiar, resultando em desempenho insatisfatório.

A análise da influência desses parâmetros, bem como a definição da janela de observação mencionada anteriormente, será apresentada no capítulo seguinte.

Por fim, destaca-se que o processo de treinamento é essencial em redes neurais supervisionadas, como a proposta neste trabalho. Dada a sua complexidade e caráter pouco intuitivo neste caso específico, uma seção dedicada, apresentada a seguir, detalhará como esse processo é conduzido.

5.3.1 Processo de treinamento da Rede Neural

Um dos principais desafios da rede neural proposta consiste na construção de um conjunto de treinamento que permita que sua operação atenda às expectativas. Como o resultado da predição é o valor do limiar da função *shrinkage*, existem situações em que o valor a ser estimado não é único, mas pode assumir uma infinidade de possibilidades, desde que respeitadas as condições adequadas.

Ao analisar a função *shrinkage* determinada pela Equação (5.21), observa-se que a estimação pode se enquadrar em duas regiões distintas. Na primeira, quando $a_0[n]$ é superior ao limiar, o valor é bem definido, pois a energia estimada final corresponde à estimativa inicial subtraída desse limiar. Assim, conhecendo-se a amplitude real da energia da partícula e o primeiro valor estimado, o limiar pode ser determinado com precisão. Já nos casos em que $a_0[n]$ não ultrapassa o limiar, a energia estimada é nula. Nessa condição, o limiar não possui um valor único, podendo assumir qualquer valor superior à medida inicial encontrada.

$$a_N[n] = S_\lambda(\Delta_N[n] + a_N[n]) = \begin{cases} a_0[n] - \lambda_N[n], & a_0[n] > \lambda_N[n] \\ 0, & a_0[n] \leq \lambda_N[n] \end{cases} \quad (5.21)$$

Para a construção do banco de dados de treinamento da rede neural, é necessário conhecer previamente um conjunto de amplitudes reais que se deseja estimar, as leituras do calorímetro e a saída do filtro FIR com os valores de $a_0[n]$. Métodos de simulação de Monte Carlo podem ser empregados para gerar os dois primeiros parâmetros, possibilitando o treinamento adequado da rede.

Primeiramente, considera-se o caso em que a amplitude real a ser estimada é positiva. Nesse cenário, a única possibilidade é que o valor inicial do processo seja superior ao limiar da função *shrinkage*. Assim, $\lambda_N[n]$ pode ser facilmente determinado, como expresso na Equação (5.22).

$$\lambda'_N[n] = a_0[n] - a_N[n] \quad (5.22)$$

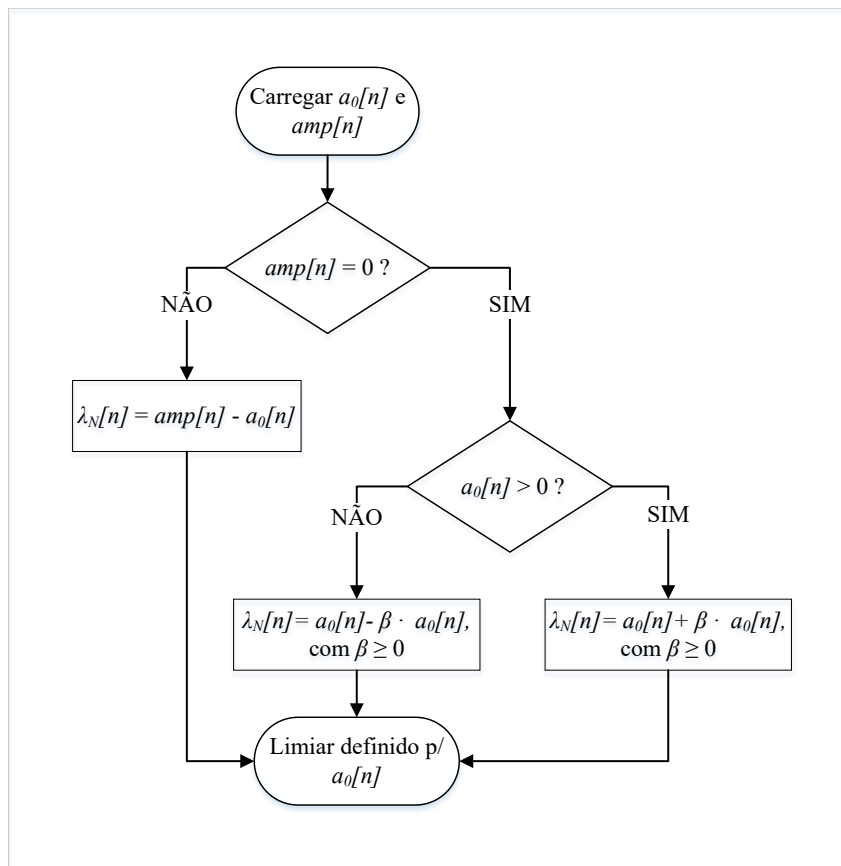
Nos casos em que a energia real seja nula — ou seja, não ocorram colisões naquele instante — surge uma situação com certo grau de liberdade. O valor do limiar deve obriga-

toriamente ser maior ou igual a $a_0[n]$ (Equação (5.23)), mas não existe um limite superior definido. Uma solução adequada consiste em atribuir a $\lambda_N[n]$ um valor ligeiramente superior a $a_0[n]$. Definir o valor exatamente igual a $a_0[n]$ pode tornar o modelo excessivamente sensível, de modo que pequenos erros de predição da rede comprometam a anulação da saída. Por outro lado, valores demasiadamente elevados de limiar podem prejudicar a capacidade da rede neural de identificar padrões nos dados durante o treinamento, levando a predições incorretas.

$$\lambda_N''[n] \geq a_0[n] \quad (5.23)$$

Para facilitar a compreensão do processo de montagem do banco de dados de treinamento da rede neural, a Figura 5.1 apresenta um fluxograma que resume as etapas descritas. Com isso, a implementação de um algoritmo que classifique e determine os valores de limiar desejados torna-se mais direta. Nesse fluxograma, $amp[n]$ representa a amplitude real a ser recuperada, $a_0[n]$ é a estimativa inicial obtida pelo filtro FIR de deconvolução, $\lambda_N[n]$ corresponde ao valor de limiar da amostra, e β é um fator de relaxamento aplicado ao limiar expresso em porcentagem.

Figura 5.1 – Fluxograma detalhando o processo de montagem do banco de dados para treinamento da rede neural

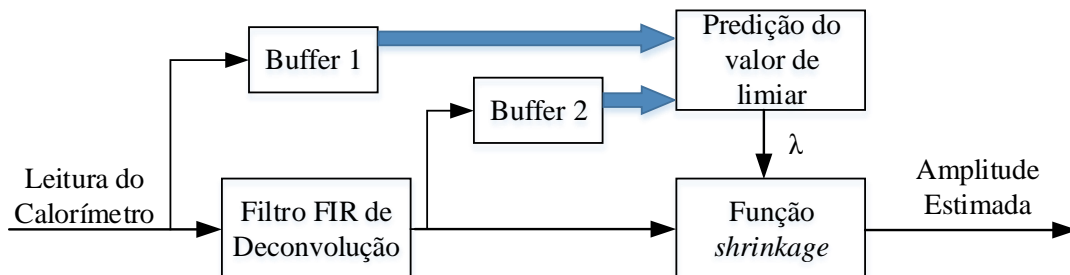


5.4 VISÃO GERAL DO MÉTODO PROPOSTO

Foi apresentada a fundamentação teórica do método proposto. Em seguida, desenvolveram-se os passos intermediários, com a devida explanação acerca da utilização do filtro FIR e da rede neural. Como etapa complementar, torna-se relevante demonstrar como esses elementos se integram e como o processo se desenrola a partir do recebimento dos dados.

A Figura 5.2 apresenta um diagrama de blocos que interliga as etapas que compõem o método proposto. Os sinais recebidos pelo calorímetro são inicialmente processados pelo filtro FIR, configurado com uma janela de tamanho determinado. A saída desse filtro corresponde à estimativa inicial da amostra central dessa janela. Dois *buffers* armazenam uma quantidade pré-definida de amplitudes iniciais estimadas e de sinais provenientes da leitura do calorímetro — sendo que a amostra central do *buffer* permite sincronizar a leitura do calorímetro com sua energia estimada. Em seguida, a rede neural recebe os valores retidos pelos *buffers* e calcula o limiar da função *shrinkage*. Finalmente, o valor estimado pelo filtro FIR é ajustado por meio da função *shrinkage*, obtendo-se assim a estimativa final do método proposto.

Figura 5.2 – Diagrama de blocos resumindo o modelo final do método proposto



Fonte: Do autor

Dessa forma, o processo pode ser implementado em um sistema de processamento de sinais em tempo real, mostrando-se adequado para integração ao sistema de *trigger* do TileCal. O modelo baseia-se em representações esparsas e em técnicas derivadas de processos iterativos para alcançar o objetivo estabelecido.

5.5 COMPARAÇÃO COM OUTROS MODELOS DE ESTIMAÇÃO

Duarte (2019) apresentou um método também fundamentado na teoria da esparsidade para aplicações em tempo real. Por meio da formulação matemática adotada, o autor propôs a aplicação da função *shrinkage* ao resultado do filtro FIR de deconvolução, a partir

de um limiar fixo. Nesse caso, o mesmo valor é utilizado para todas as amostras analisadas, o que torna essencial a determinação de um valor único de λ capaz de maximizar o desempenho na reconstrução da energia.

Ao analisar o processo iterativo desenvolvido por Elad (2010), observa-se que a fixação desse limiar inviabiliza a obtenção de melhores estimativas utilizando a teoria da esparsidade. Nesse sentido, o método proposto buscou determinar o valor de λ de forma dinâmica, de modo a potencializar a performance na tarefa de reconstrução de energia.

O estudo desenvolvido por Faria (2017) apresentou uma rede neural integralmente responsável pelo processo de estimação a partir das leituras do calorímetro. Nesse caso, a rede produz diretamente, em sua saída, a estimativa da amplitude na janela analisada, o que torna a tarefa mais complexa para o modelo.

Por sua vez, a rede neural utilizada no método aqui proposto constitui apenas uma parte do processo geral, atuando no ajuste da saída do filtro FIR. Em outras palavras, ela pode suprimir o sinal (retornando zero) ou realizar um refinamento na estimativa, a depender do valor do neurônio de saída e do resultado do filtro FIR. Dessa forma, um eventual erro de predição causa um impacto menor quando comparado à rede desenvolvida por Faria (2017).

5.6 CONCLUSÕES PARCIAIS DO CAPÍTULO

Foi desenvolvido um método baseado na teoria da esparsidade, derivado de um processo iterativo. A proposta consistiu em eliminar a etapa de atualização sucessiva a cada iteração, estabelecendo uma relação direta entre a leitura do calorímetro, a saída do filtro FIR de deconvolução e o valor de limiar da função *shrinkage*. Tal relação é obtida por meio de uma rede neural MLP, responsável por prever o valor do limiar.

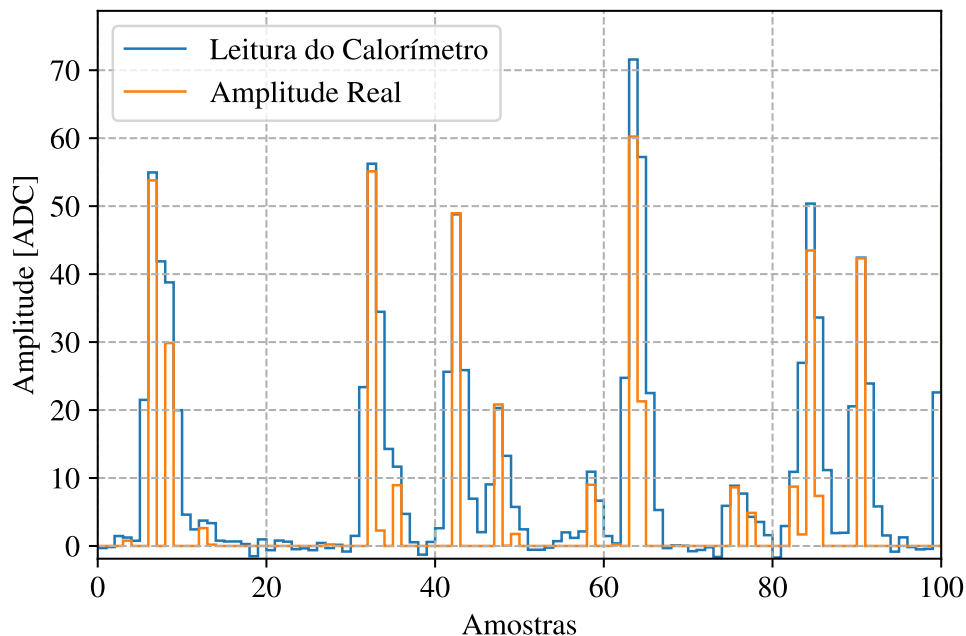
O método proposto apresenta potencial para aplicação em sistemas de processamento *online* de estimação de energia, uma vez que utiliza uma estrutura não recursiva. Os resultados e análises correspondentes serão discutidos no capítulo subsequente.

6 RESULTADOS

Para avaliar a performance do método proposto para a estimação da energia dos sinais do TileCal, são utilizadas simulações baseadas em dados de operação real do calorímetro. Conforme descrito em 2.4.2, o simulador *Toy Monte Carlo* constitui um instrumento amplamente utilizado pelo grupo de pesquisa e será mantido como referência para mensurar e comparar os resultados obtidos com o método proposto.

Os dados simulados contêm 10.000 amostras para cada condição de ocupação analisada, constituídas por pares formados pela saída do calorímetro e pela respectiva energia da partícula detectada. A distribuição de energia das colisões segue uma lei exponencial com média de 30 ADC (*ADC count*), enquanto o ruído é modelado por uma distribuição gaussiana de média zero e desvio padrão de 1 ADC. Dessa forma, é possível avaliar o comportamento do calorímetro em condições próximas às reais, abrangendo diferentes cenários de luminosidade. A Figura 6.1 ilustra um exemplo de subconjunto dos dados utilizados para uma ocupação de 20%, apresentando a energia da partícula absorvida e a saída correspondente do sistema de leitura do calorímetro.

Figura 6.1 – Parcela de amostras dos dados simulados para uma ocupação de 20%



Fonte: Do autor

Inicialmente, serão avaliadas as etapas intermediárias do método proposto, fundamentais para a compreensão e validação do modelo final. Assim, torna-se necessária a análise da performance do filtro FIR de deconvolução e da predição dos valores de limiar da função *shrinkage*.

Na sequência, procede-se à avaliação do modelo final do método proposto, o qual será comparado com técnicas previamente descritas na literatura. Os métodos de referência considerados serão: o dos Mínimos Quadrados (LS), o método desenvolvido por Duarte (2019) — SFDM-FT (*Sparse Free-Running Method with Dynamic Threshold*) — e o proposto por Faria (2017) — RNE (Rede Neural Estimadora). Ressalta-se que todos são aplicáveis em processamento *online* para o primeiro nível de *trigger*.

Como métrica de avaliação será utilizada a raiz do erro quadrático médio (*Root Mean Square Error* — RMSE). O RMSE é definido pela Equação (6.1), onde o erro obtido no processo de estimação ($e[n] = A[n] - \hat{A}[n]$) é empregado no cálculo, refletindo tanto a dispersão quanto a tendência do erro Bishop (2006). Quanto menor o RMSE, mais fidedigna será a reconstrução de energia fornecida pelo método.

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} |e[n]|^2} \quad (6.1)$$

É importante destacar que tanto o método proposto quanto todos os métodos utilizados para fins de comparação foram implementados na linguagem de programação Python, escolhida por se tratar de uma ferramenta de código aberto amplamente utilizada na comunidade científica e adequada ao desenvolvimento de pesquisas.

Também será avaliado um modelo genérico, com pesos fixos para todos os canais do TileCal, utilizando uma rede neural treinada com um conjunto de dados composto por todas as ocupações. Por fim, serão apresentados e discutidos os recursos computacionais necessários para a implementação dos algoritmos avaliados.

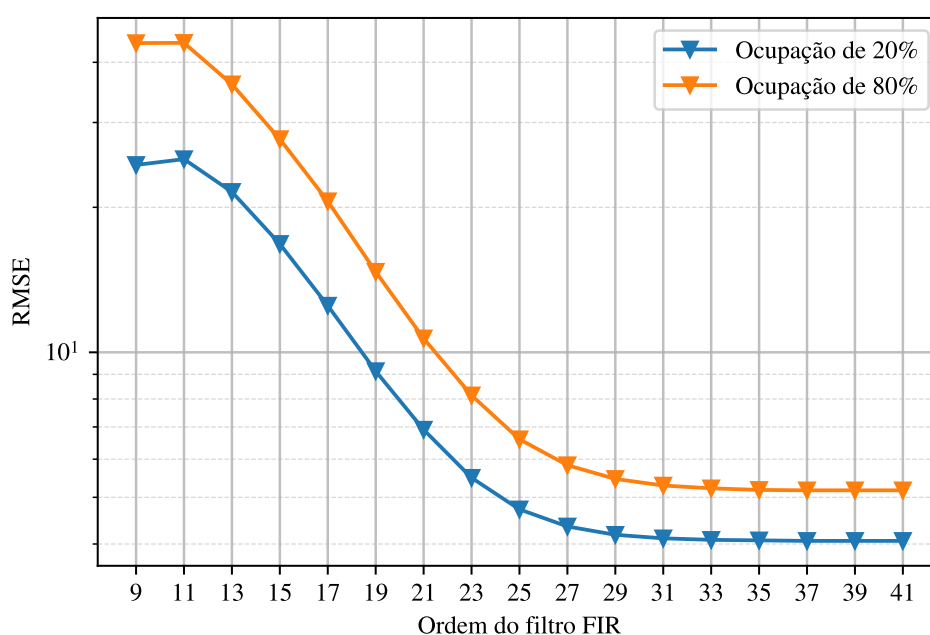
6.1 AVALIAÇÃO DO FILTRO FIR DE DECONVOLUÇÃO

O método proposto utiliza como estimativa preliminar o resultado obtido pelo filtro FIR de deconvolução. Os pesos desse filtro são derivados das linhas da pseudo-inversa da matriz de convolução do sistema, em que o valor máximo localiza-se na coluna central da respectiva linha. Quanto maior a matriz de convolução, maior será o número de pesos do filtro e, conseqüentemente, melhor o processo de deconvolução do sinal. No entanto, essa melhoria apresenta um limite, pois há um ponto a partir do qual apenas valores nulos são adicionados às extremidades da linha, sem promover ganhos significativos de desempenho.

Outro fator limitante para o tamanho da matriz de convolução está relacionado ao objetivo do método proposto. Busca-se sua implementação em um sistema de processamento em tempo real; portanto, prioriza-se o menor número de pesos possível que satisfaça adequadamente os requisitos, reduzindo o consumo de recursos. Além disso, o aumento do número de pesos impacta diretamente no atraso da resposta do filtro, uma vez que filtros de ordem elevada acarretam maior latência, o que pode comprometer determinadas aplicações.

Para uma ocupação de 20%, representando cenários de baixa taxa de colisões, foi analisado o RMSE para diferentes números de pesos do filtro FIR, variando de 9 a 41 com incremento de 2 entre os testes. A mesma análise foi realizada para uma ocupação de 80%, uma vez que o método proposto deve apresentar bom desempenho em condições de alta ocupação. A Figura 6.2 mostra a relação entre o RMSE e o número de pesos do filtro FIR para ambos os casos, utilizando escala logarítmica.

Figura 6.2 – Relação entre RMSE e número de pesos do filtro FIR para ocupações de 20% e 80%



Fonte: Do autor

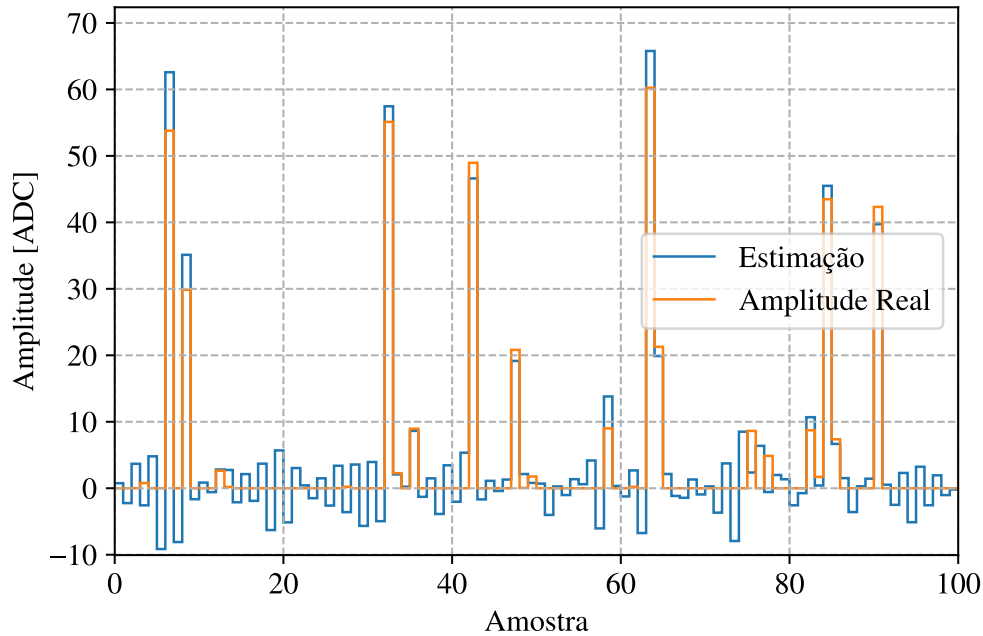
O comportamento das duas curvas é semelhante. Observa-se que, para números reduzidos de pesos, o RMSE é elevado. Como esperado, o aumento da quantidade de pesos melhora o desempenho da estimação, até atingir um limite próximo a 37, a partir do qual o RMSE permanece praticamente constante.

Apesar de o limite ser alcançado em torno de 37 pesos, do ponto de vista de uso de *hardware*, é mais vantajoso empregar a menor quantidade possível que ainda apresente resultados satisfatórios. Nesse caso, a escolha por 31 pesos mostra-se adequada, pois fornece bom desempenho e representa apenas uma estimativa inicial no método proposto.

Assim, o modelo final adotou 31 pesos para o filtro FIR de deconvolução. O resultado da estimação inicial desse filtro para dados com ocupação de 20% está ilustrado na Figura 6.3, na qual se observa a presença de artefatos indesejados entre amostras correspondentes a energias reais de colisão. Espera-se que a função *shrinkage*, fundamentada na teoria da esparsidade, seja capaz de suprimir tais artefatos e aprimorar as estimativas

de energia.

Figura 6.3 – Amostras estimadas pelo filtro FIR para uma ocupação de 20%



Fonte: Do autor.

Encerrada essa análise, a etapa seguinte consiste na avaliação da predição do valor de limiar da função *shrinkage*.

6.2 ANÁLISE DA PREDIÇÃO DO VALOR DE LIMIAR DA SHRINKAGE

Na Seção 5.3 foi discutida a motivação para o uso de uma rede neural na predição dos valores de limiar da função *shrinkage*. Na presente seção, é apresentado um exemplo de topologia escolhida para a rede e os resultados obtidos para a predição dos valores desejados.

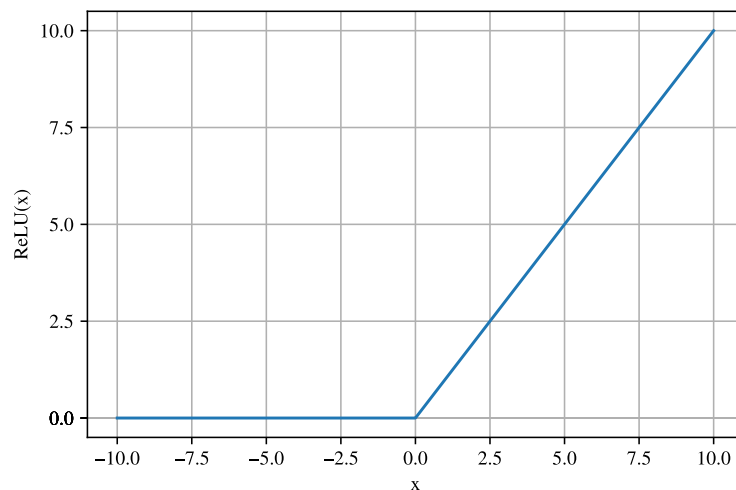
Considerando que o objetivo do método é sua aplicação em um sistema de processamento em tempo real, busca-se empregar a rede mais simples possível que atenda de forma adequada à finalidade proposta. Um pré-requisito estabelecido para o desenvolvimento é a utilização de apenas uma camada oculta. Dessa forma, a rede neural é composta por uma camada de entrada, uma camada oculta e uma camada de saída.

Na camada de entrada, são fornecidos os dados provenientes das leituras do calorímetro e da saída do filtro FIR de deconvolução, ambos com o mesmo tamanho de janela de amostras. Durante os testes, verificou-se que o uso de 5 amostras de cada um dos sinais produziu resultados satisfatórios, totalizando 10 valores de entrada para a rede.

O aumento do tamanho da janela não resultou em melhorias significativas em relação à topologia escolhida. Estes testes serão mais bem detalhados na próxima seção

A camada oculta possui dois parâmetros configuráveis que influenciam diretamente os resultados: a função de ativação e o número de neurônios. Primeiramente, buscou-se uma função de ativação simples de implementar em *hardware* — com baixo consumo de recursos — e que fornecesse saídas consistentes. Por esse motivo, adotou-se a função ReLU (do inglês, *Rectified Linear Unit*), ilustrada na Figura 6.4. Essa função apresenta rápida convergência e pode ser implementada por meio de circuitos condicionais, envolvendo apenas comparadores e multiplexadores.

Figura 6.4 – Função de ativação ReLU



Fonte: Do autor.

Quanto ao número de neurônios na camada oculta, verificou-se que o valor de 10 apresentou resultados satisfatórios. O aumento desse número implica maior quantidade de multiplicações e, conseqüentemente, maior uso de DSPs (do inglês, *Digital Signal Processor*) em implementações em FPGA. Como valores superiores não proporcionaram ganhos expressivos de desempenho, optou-se pela utilização de 10 neurônios na configuração dos testes do modelo.

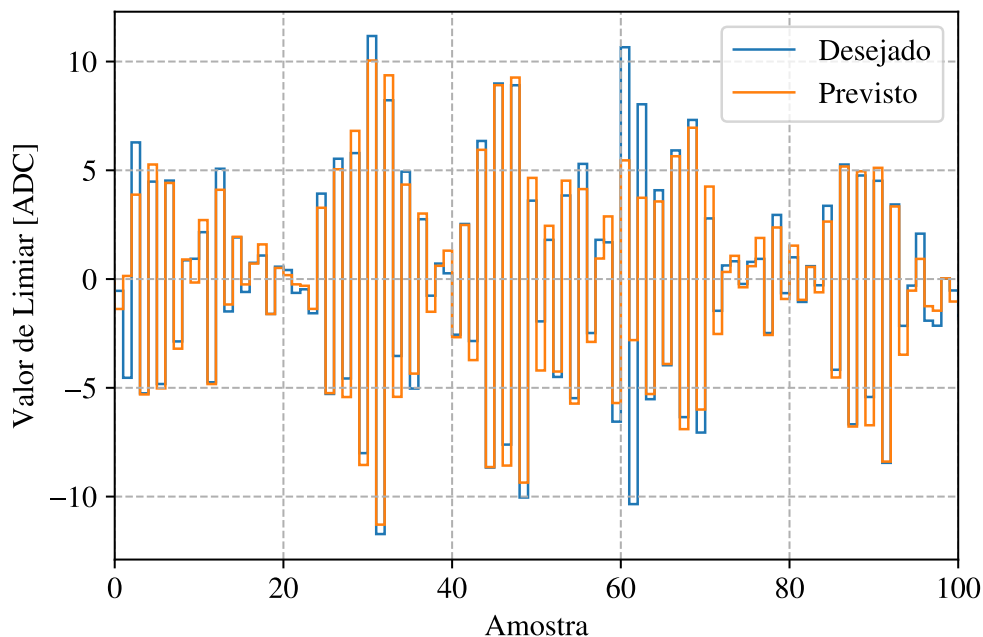
Por fim, é necessário definir a função de ativação da camada de saída, visto que a rede neural produz apenas um valor final: o limiar da função *shrinkage*. Como esse valor pode assumir tanto magnitudes positivas quanto negativas, a função ReLU não é aplicável nesse caso. Uma alternativa simples e eficiente é a função Linear, cuja adoção apresentou resultados promissores, conforme discutido adiante, incluindo na estimação final da energia.

Portanto, a rede neural MLP *feed-forward* utilizada nesta etapa possui a seguinte configuração: camada de entrada com 10 amostras; uma camada oculta com 10 neurônios

e função de ativação ReLU; e uma camada de saída com 1 neurônio e função de ativação Linear. Para o treinamento da rede, foram utilizadas 8.000 amostras, com 500 épocas, paciência (*early stop*) de 30, função de perda MSE (do inglês, *Mean Square Error*) e métricas também baseadas em MSE. As 2.000 amostras restantes foram reservadas para o teste da predição. É importante reiterar que essa topologia pode ser alterada de acordo com o ambiente que será aplicado, que será melhor abordado na próxima seção.

Para ilustrar o desempenho da predição do valor de limiar da função *shrinkage*, inicialmente analisou-se um conjunto de dados com 20% de ocupação. O gráfico da Figura 6.5 apresenta a comparação entre os valores previstos e os valores de referência, evidenciando que a rede obteve uma boa aproximação. O RMSE do processo de predição foi de 1,67.

Figura 6.5 – Comparação entre o valor previsto e o valor desejado para uma ocupação de 20%

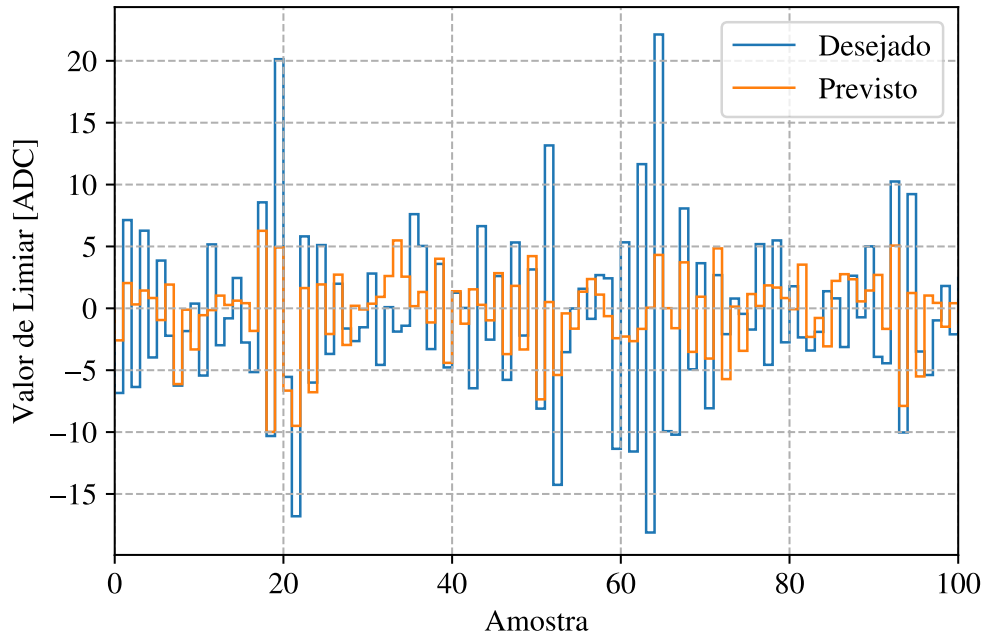


Fonte: Do autor.

Na análise dos dados com 80% de ocupação, cujos resultados estão apresentados na Figura 6.6, observa-se que os valores previstos já não se aproximam tanto dos valores desejados, resultando em um RMSE de 4,03. Para altas ocupações, a predição apresenta menor sensibilidade, uma vez que o número de amostras com energia nula é significativamente menor em comparação com situações de baixa ocupação. Nesses casos, o limiar da função *shrinkage* tende a atuar predominantemente como um corretor da amplitude estimada.

A análise preliminar da performance da rede neural na predição do limiar da

Figura 6.6 – Comparação entre o valor previsto e o valor desejado para uma ocupação de 80%



Fonte: Do autor.

função *shrinkage* é relevante; entretanto, a verificação isolada, sem considerar o processo de limiarização, não reflete integralmente o procedimento final de estimação de energia. Assim, torna-se necessário avaliar o contexto completo, com todos os processos integrados.

6.3 COMPARAÇÃO DE TOPOLOGIAS DA REDE NEURAL NOS RESULTADOS

Na seção anterior, foram apresentados os resultados de predição do valor da função *shrinkage* para uma topologia fixa. Para a definição da topologia mais adequada à aplicação desejada, diversos fatores podem influenciar essa escolha. Em cenários com abundância de recursos computacionais, uma rede neural mais complexa pode ser adotada. Por outro lado, em situações com *hardware* limitado ou com requisitos rigorosos de latência, deve-se priorizar a utilização da rede neural mais simples possível. Dessa forma, nesta seção são apresentadas diferentes topologias, juntamente com seus respectivos resultados, permitindo que o usuário defina quais critérios são prioritários.

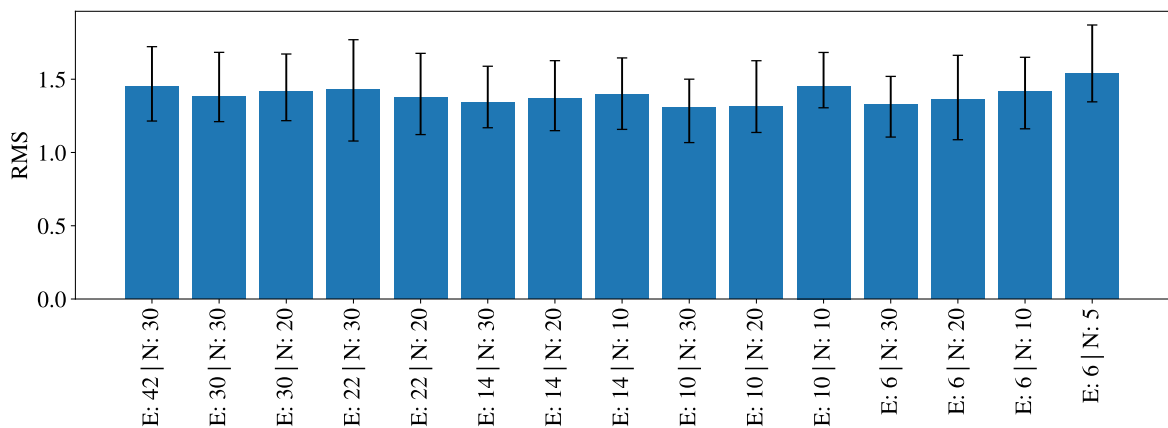
O ambiente de estimação de energia em tempo real para o TileCal se enquadra neste último cenário. Trata-se de milhares de canais operando em paralelo a uma frequência de 40 MHz, o que torna prioritária a minimização do uso de recursos computacionais, mesmo que isso implique a não utilização da topologia com a melhor performance absoluta.

Com o objetivo de comparar as diferentes topologias, será analisado o erro quadrá-

tico médio (*Root Mean Square Error* – RMSE) da diferença entre o resultado do método proposto e o valor de amplitude desejado. Será utilizado o método de validação cruzada denominado *k-Fold*, com $k = 10$. Para cada configuração, serão apresentados os valores médios de RMSE, juntamente com barras de erro correspondentes aos valores máximos e mínimos observados.

Os dois parâmetros variáveis da rede neural são o número de entradas e o número de neurônios na camada oculta. O número de entradas corresponde ao total de amostras consideradas, incluindo tanto as amostras do calorímetro quanto as amostras provenientes da saída do filtro FIR. O primeiro cenário analisado corresponde a uma ocupação de 20%, caracterizando um ambiente com menor incidência de eventos de *pile-up*. A Figura 6.7 apresenta o gráfico com os resultados obtidos para diferentes configurações, onde “E” representa o número de entradas da rede (contabilizando leitura do calorímetro e saída do filtro FIR) e “N” o número de neurônios presentes na camada oculta. Observa-se que o aumento do número de neurônios na camada oculta tende a gerar melhores resultados; entretanto, o ganho de performance é relativamente pequeno para redes mais complexas. Pode-se afirmar que existe uma equivalência estatística entre os modelos.

Figura 6.7 – RMS obtido por diferentes topologias de rede neural para uma ocupação de 20%

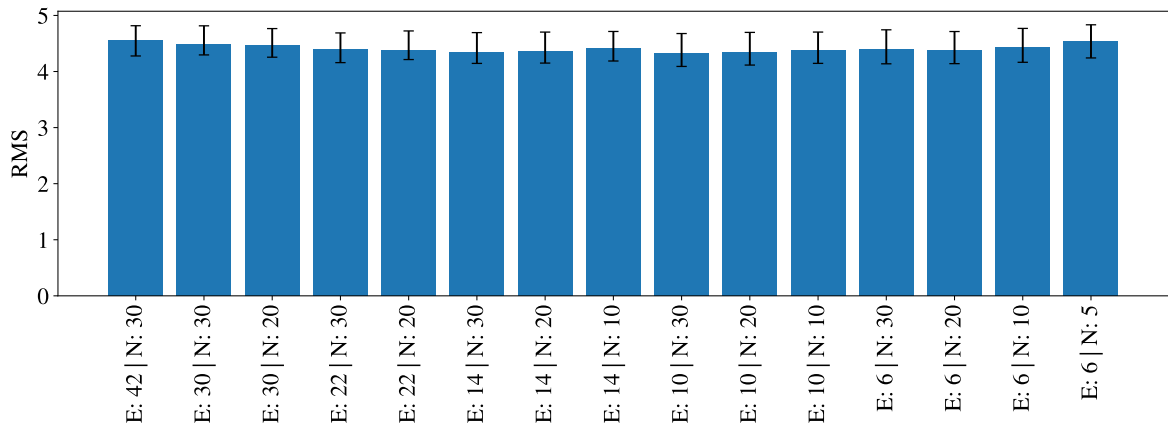


Fonte: Do autor.

Para a avaliação de ambientes com maior presença de *pile-up*, são utilizados dados com ocupação de 80%. O gráfico com os resultados, apresentado na Figura 6.8, evidencia que, nesse cenário, as diferenças de performance entre as topologias são ainda menores, reforçando a indicação de que estruturas de rede mais simples podem ser empregadas no método proposto.

Com o intuito de comparar o custo computacional associado a cada topologia, são contabilizados o número de multiplicadores necessários em cada uma das configurações analisadas. A Tabela 6.1 resume esses resultados, evidenciando um aumento significativo

Figura 6.8 – RMS obtido por diferentes topologias de rede neural para uma ocupação de 80%



Fonte: Do autor.

no número de multiplicadores à medida que a complexidade da rede cresce. Esse comportamento se deve ao fato de que cada entrada está conectada a cada neurônio da camada oculta por meio de um peso.

Tabela 6.1 – Multiplicadores necessários em cada topologia de rede neural.

Número de entradas	Neurônios na camada oculta	Multiplicadores
42	30	1290
30	30	930
30	20	620
30	10	310
22	30	690
22	20	460
22	10	230
14	30	450
14	20	300
14	10	150
10	30	330
10	20	220
10	10	110
6	30	210
6	20	140
6	10	70
6	5	35

Fonte: Do autor

Conforme descrito anteriormente, para o ambiente de testes, o uso do menor número possível de recursos computacionais é um requisito fundamental. Como o ganho

de performance não justifica a adoção de redes mais complexas e não existe uma diferença estatisticamente significativa entre as redes, a topologia contendo 6 entradas e 5 neurônios na camada oculta mostrou-se uma escolha adequada e será utilizada nos testes apresentados na próxima seção.

6.4 RESULTADOS DE ESTIMAÇÃO DO MÉTODO PROPOSTO

Definidas as etapas anteriores necessárias ao processo de estimação, o método proposto pode ser avaliado. Conforme mencionado anteriormente, ele será comparado com os métodos LS, SFRM-FT e RNE.

No caso do método LS, será utilizada a mesma janela empregada no filtro FIR de deconvolução, de tamanho 31. O mesmo critério é adotado para o método SFRM-FT, que também depende da etapa prévia do filtro de deconvolução. Já o método RNE apresenta uma rede neural voltada à estimação da amplitude, configurada da seguinte forma para obter resultado similar ao do método proposto: uma camada de entrada com 13 neurônios; uma camada oculta com 12 neurônios e função de ativação ReLU; e, por fim, uma camada de saída com função de ativação linear.

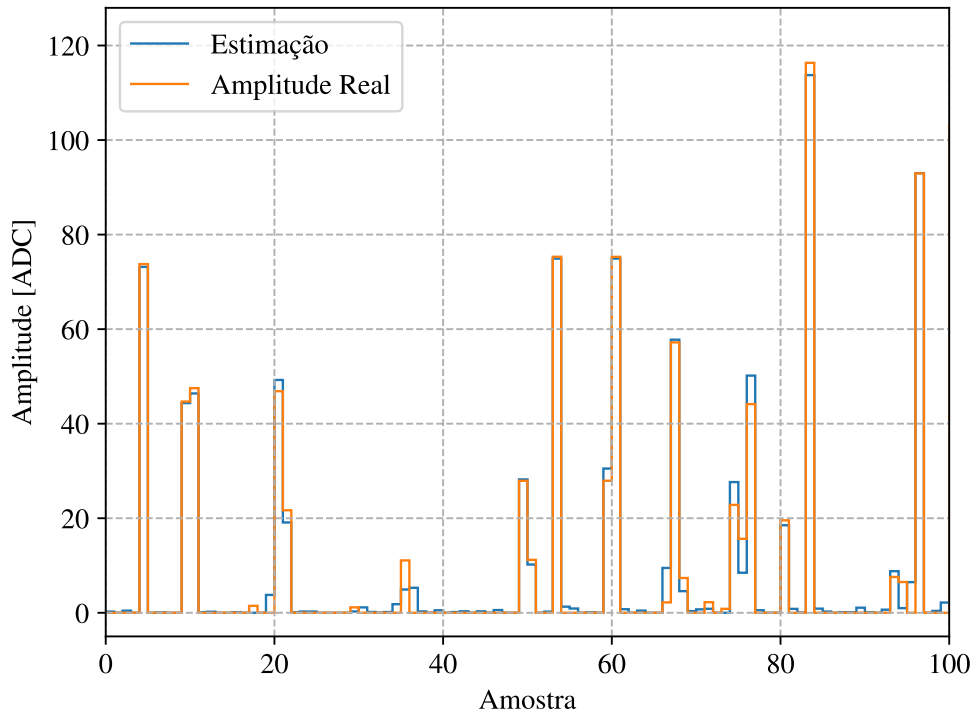
O modelo final do método proposto foi configurado da seguinte maneira: um filtro FIR de deconvolução com 31 coeficientes, responsável por fornecer o valor inicial necessário ao processo de estimação; uma rede neural para predição do limiar da função *shrinkage*, cuja camada de entrada recebe 3 leituras do calorímetro e 3 saídas do filtro FIR, seguida por uma camada oculta com 5 neurônios e função de ativação ReLU, e uma camada de saída composta por um único neurônio com ativação linear. A etapa final consiste em aplicar a função *shrinkage* ao valor inicial obtido pelo filtro e ao respectivo limiar estimado pela rede neural.

Como o método proposto e o RNE envolvem redes neurais, foi empregado o procedimento de validação cruzada, com $k = 10$. Dessa forma, foram obtidos o valor médio do RMSE, bem como os valores máximo e mínimo desse parâmetro.

Para avaliar a estimação em cenários de baixa ocupação, utilizou-se um conjunto de dados correspondente a 20% de ocupação. Na Figura 6.9, apresentam-se as primeiras 100 amostras da estimação do método proposto, em comparação com o valor real da amplitude da partícula. Observa-se que, nas amostras que continham energia real, houve uma filtragem consideravelmente mais eficiente, com remoção dos artefatos. Nesse cenário, o RMSE médio obtido foi de 1,72.

Aplicando o método LS, o RMSE encontrado foi de 3,76, cujas primeiras amostras estão apresentadas na Figura 6.10. Nota-se que o LS não conseguiu eliminar os artefatos com a mesma eficiência, embora tenha superado o desempenho do filtro FIR puro. O método proposto apresentou um resultado 54% superior em relação ao LS.

Figura 6.9 – Resultado da estimação do método proposto para uma ocupação de 20%



Fonte: Do autor.

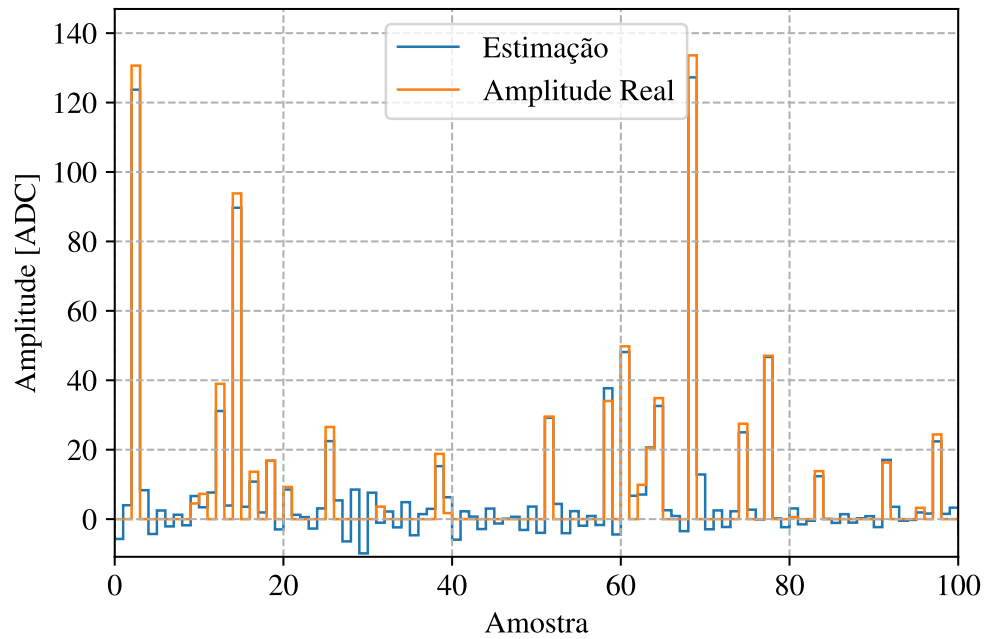
O método SFRM-FT obteve desempenho melhor que o LS na redução de artefatos, mas ainda não alcançou o mesmo nível de qualidade do método proposto, como pode ser observado na Figura 6.11. O RMSE obtido foi de 2,63, valor 34% superior ao alcançado pelo método proposto.

Por fim, a Figura 6.12 apresenta os resultados obtidos com a rede neural RNE. Esse método produziu resultados bastante próximos aos do método proposto, alcançando um RMSE médio de 1,74. Considerando a variação de valores de RMSE resultante do processo de treinamento, pode-se afirmar que ambos os métodos apresentaram desempenhos equivalentes. Entretanto, pela análise gráfica das amostras estimadas, percebe-se que o método proposto proporciona uma filtragem mais eficiente dos artefatos indesejados.

Em condições de baixa ocupação, conclui-se que o método proposto apresenta maior capacidade de filtragem dos artefatos nas amostras com energia, aproximando-se mais consistentemente da teoria da esparsidade. Esse resultado pode ser explicado pela própria natureza dos dados, em que a predominância de elementos nulos favorece a atuação da função *shrinkage*, que elimina (zera) tais elementos indesejados.

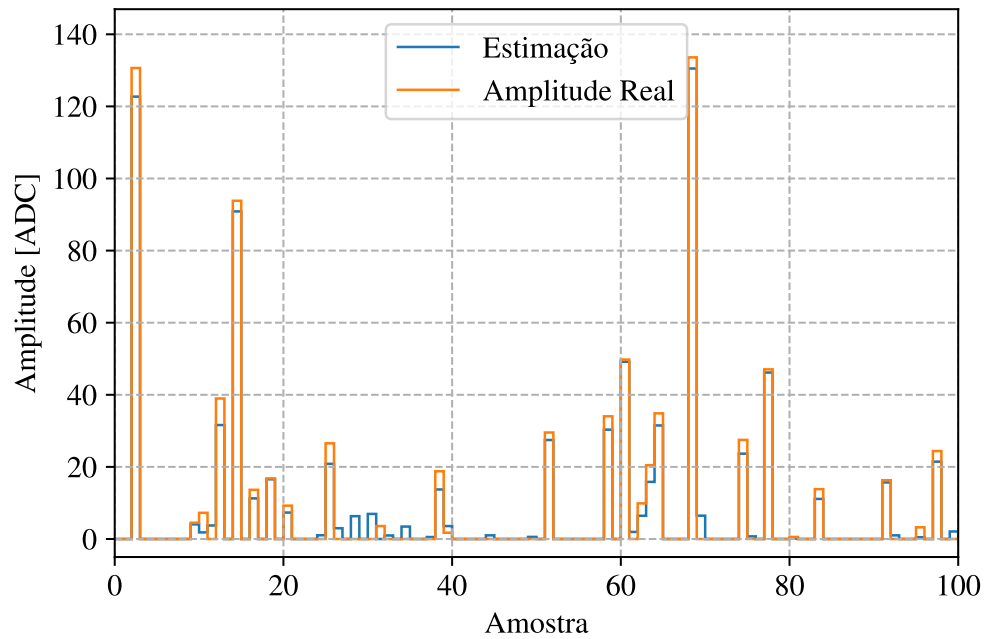
Prosseguindo para a análise em cenários de alta ocupação, considera-se agora o caso de 80%. O método proposto apresentou um RMSE médio de 4,50, cujos resultados iniciais de estimação estão ilustrados na Figura 6.13. Observa-se que, nesse regime, a

Figura 6.10 – Resultado da estimação do LS para uma ocupação de 20%



Fonte: Do autor.

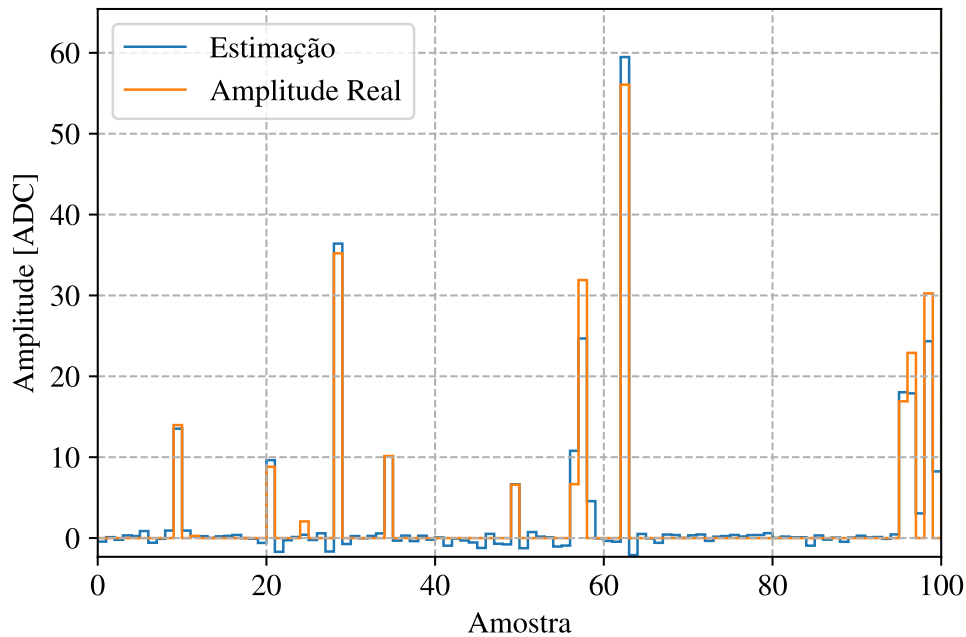
Figura 6.11 – Resultado da estimação SFRM-FT para uma ocupação de 20%



Fonte: Do autor.

quantidade de elementos nulos é substancialmente menor em comparação ao cenário de

Figura 6.12 – Resultado da estimação RNE para uma ocupação de 20%



Fonte: Do autor.

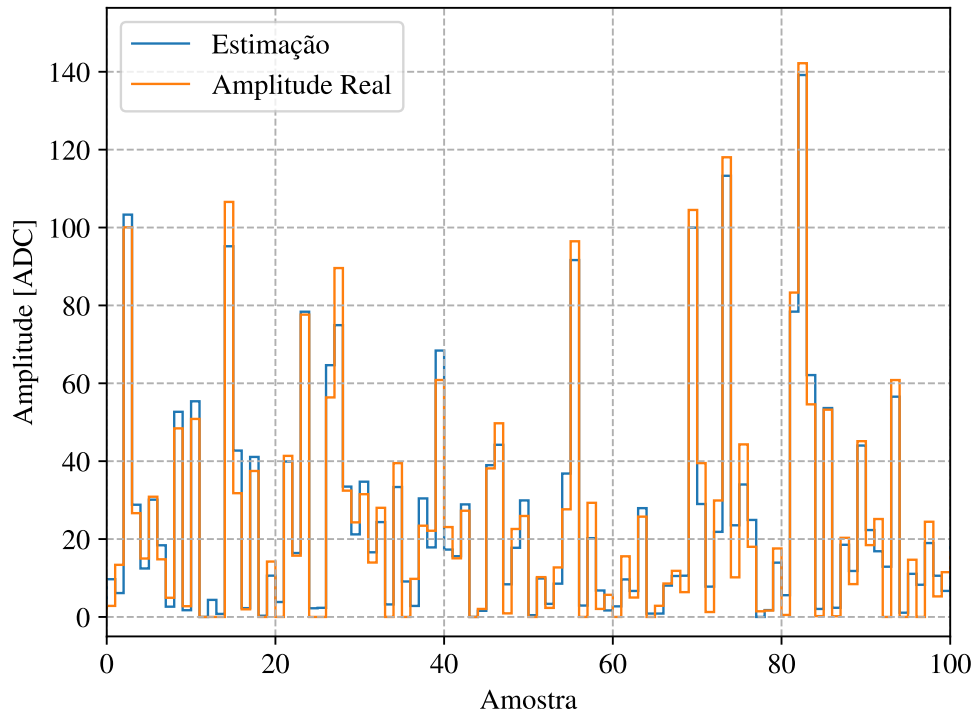
20% de ocupação. Devido à elevada densidade de sinais, a inspeção visual das amostras torna-se pouco informativa, motivo pelo qual não são apresentadas figuras análogas para os demais métodos.

Nos métodos de comparação, o LS obteve RMSE médio de 5,11 e o SFRM-FT de 5,14. Assim, o método proposto apresentou uma redução de aproximadamente 12% no RMSE em relação ao melhor resultado entre esses métodos. Já a rede neural RNE apresentou um RMSE médio de 5,89, posicionando-se entre o desempenho do LS e do método proposto. Ressalta-se, contudo, que este resultado não implica superioridade absoluta do método proposto frente à RNE, visto que a etapa de treinamento pode introduzir variações estatisticamente relevantes no desempenho.

Em ocupações elevadas, o número reduzido de amostras nulas faz com que o método proposto atue de maneira semelhante a um corretor de amplitude: quando a saída inicial do filtro de deconvolução é inferior ao limiar definido pela função *shrinkage*, esta exerce o papel de ajuste fino, refinando a estimação final da energia.

Com o intuito de avaliar o desempenho de forma global, o nível de ocupação foi variado de 10% a 100%, em incrementos de 10%. O RMSE médio foi calculado para cada método, resultando na curva da Figura 6.14. A comparação direta com a RNE é apresentada separadamente. Em todos os cenários, mesmo considerando variações decorrentes da validação cruzada, o método proposto obteve o melhor desempenho. Além

Figura 6.13 – Resultado da estimação do método proposto para uma ocupação de 80%



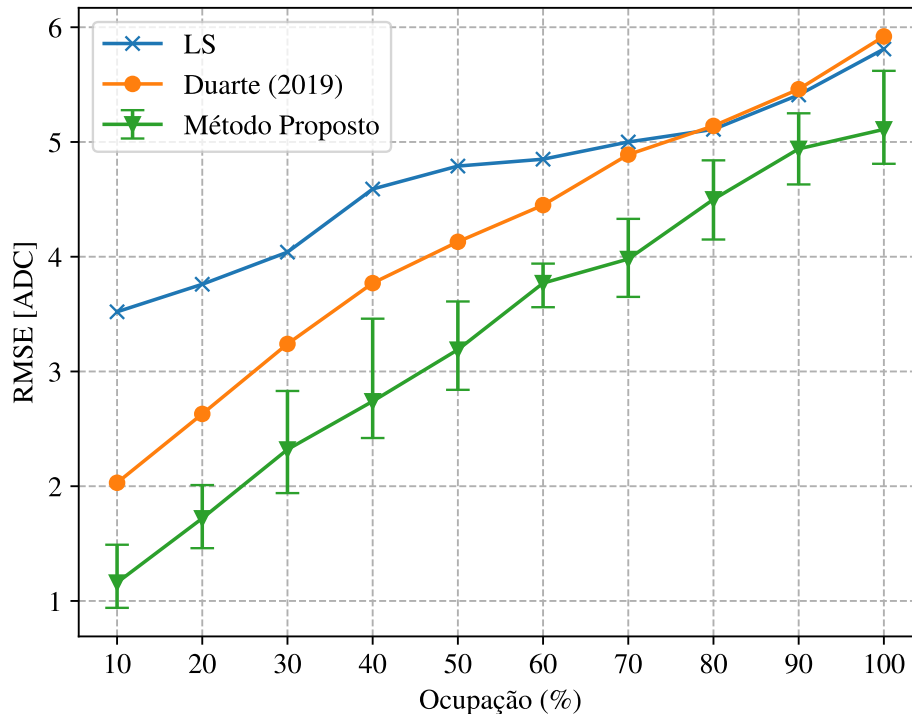
Fonte: Do autor.

disso, o resultado pode ser melhorado ao utilizar uma rede neural mais complexa para predição do valor do limiar da função *shrinkage*.

Ao se confrontar diretamente os resultados do método proposto com a RNE (Figura 6.15), nota-se que ambas as abordagens apresentam desempenhos próximos. Entretanto, para alcançar resultados equivalentes, a rede neural isolada demanda maior complexidade. O ganho de desempenho obtido está relacionado ao método de treinamento empregado, o qual se torna possível ao interpretar a rede neural como um sistema de ajuste de limiar fundamentado na teoria da esparsidade. Além disso, a tarefa da RNE é intrinsecamente mais desafiadora, uma vez que deve estimar diretamente o valor da energia. No modelo proposto, por outro lado, a rede neural assume papel auxiliar, funcionando como mecanismo de ajuste fino sobre a estimação inicial, o que simplifica sua função e potencializa a robustez do processo.

Os resultados apresentados até o momento utilizaram bancos de dados de treinamento compostos exclusivamente por grandezas correspondentes à ocupação específica em análise. Considerando que o TileCal opera com milhares de canais, a realização de um treinamento específico para cada canal não constitui uma tarefa trivial. Dessa forma, torna-se fundamental avaliar o impacto da utilização de um modelo único que, uma vez treinado, seja capaz de atender a diferentes níveis de ocupação sem a necessidade de ajuste

Figura 6.14 – Comparação de performance dos métodos para diferentes ocupações



Fonte: Do autor.

dos pesos. Essa abordagem é apresentada e analisada na seção subsequente.

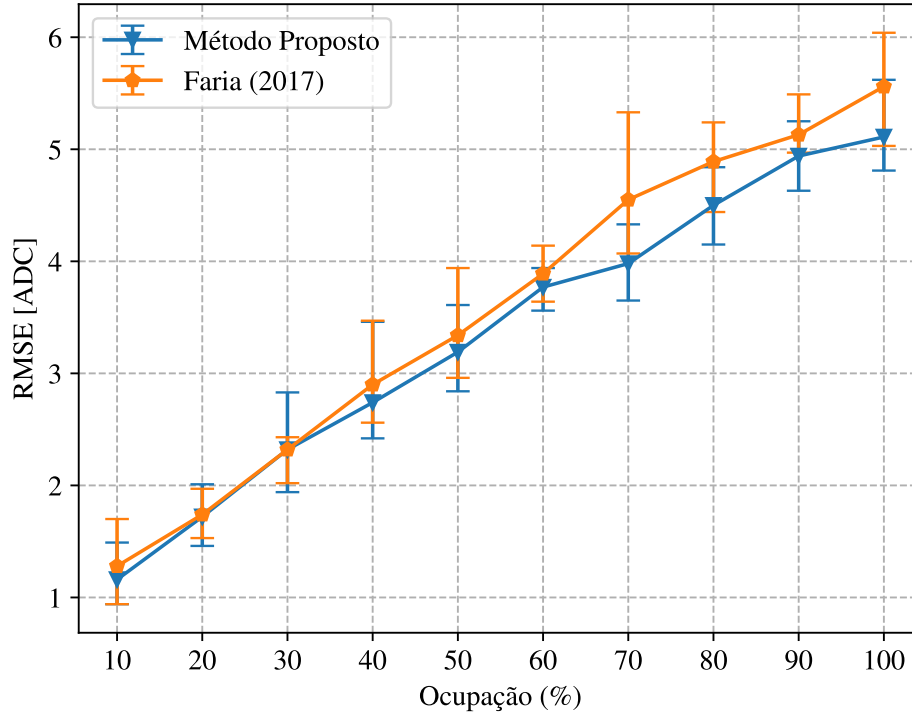
6.5 ANÁLISE DE UM MODELO GERAL PARA TODAS AS OCUPAÇÕES

Com o objetivo de obter um conjunto de pesos capaz de atender a todos os níveis de ocupação, é necessário que o banco de dados utilizado no treinamento da rede neural contemple todos os cenários de interesse. Para esse fim, foram gerados 10.000 eventos para ocupações variando de 10% a 100%, em incrementos de 10%. Assim, o conjunto de treinamento é composto por um total de 100.000 eventos.

Como a proposta consiste em avaliar um único modelo aplicável a todas as ocupações, o método de validação cruzada não foi empregado para a quantificação da performance em cada cenário específico. Em vez disso, foram gerados conjuntos de teste independentes, contendo 10.000 eventos para cada nível de ocupação, resultando em 10 conjuntos distintos de teste. Dessa forma, é possível realizar a estimação de amplitude individualmente para cada ocupação, utilizando sempre o mesmo modelo treinado.

O primeiro modelo avaliado adota a mesma topologia considerada na análise apresentada na seção anterior, sendo composta por um filtro FIR de deconvolução com 31 coeficientes e uma rede neural responsável pela predição do limiar da função *shrinkage*. A

Figura 6.15 – Comparação de performance entre o método RNE e o método proposto para todos os cenários



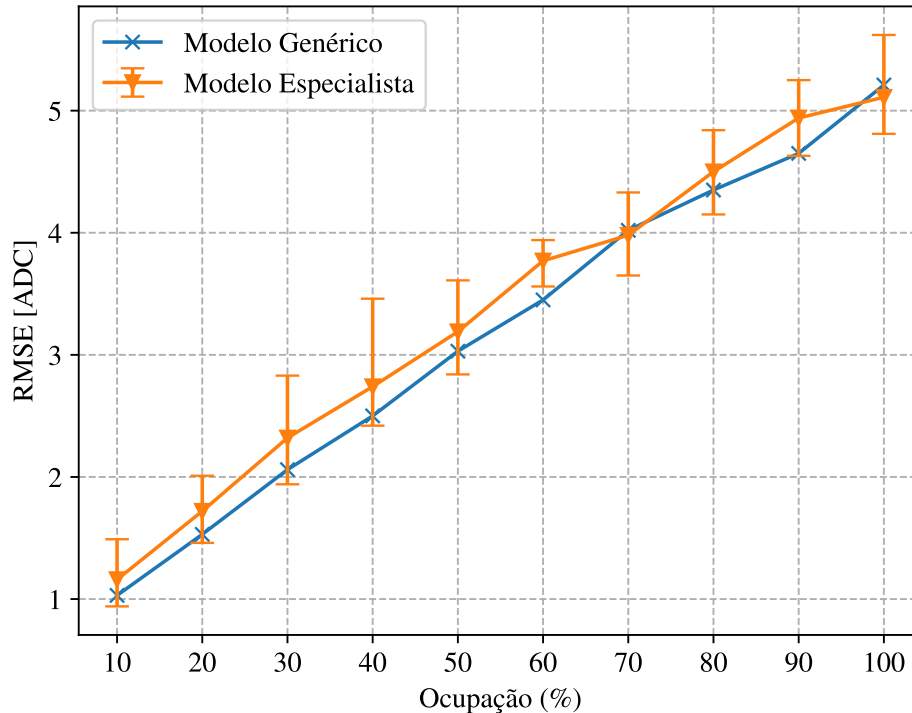
Fonte: Do autor.

camada de entrada da rede neural recebe 3 amostras do calorímetro e 3 valores provenientes da saída do filtro FIR, seguida por uma camada oculta com 5 neurônios e função de ativação ReLU, e por uma camada de saída constituída por um único neurônio com ativação linear.

A comparação de desempenho entre o modelo genérico e os modelos especialistas treinados para ocupações específicas pode ser observada no gráfico apresentado na Figura 6.16. O modelo genérico apresentou uma performance muito semelhante à dos modelos especialistas, mantendo-se dentro da faixa de erro observada nos resultados obtidos por validação cruzada. Esses resultados indicam que é possível empregar um conjunto fixo de pesos capaz de atender adequadamente a todos os cenários de ocupação, sem degradação significativa da performance do algoritmo. Essa característica é particularmente relevante para o ambiente de aplicação considerado.

Por fim, para uma comparação final entre o método proposto e as demais abordagens apresentadas neste trabalho, a análise dos recursos computacionais necessários para a implementação em *hardware* de cada solução é apresentada na próxima seção.

Figura 6.16 – Comparação de performance entre o modelo genérico e o modelo especialista para o Método Proposto



Fonte: Do autor.

6.6 COMPARAÇÃO DE RECURSOS COMPUTACIONAIS NECESSÁRIOS

Para a definição do método a ser adotado, é fundamental conhecer os recursos exigidos para sua implementação em *hardware* dedicado. No ambiente de testes descrito neste trabalho, o uso de FPGA é amplamente difundido para a estimação de energia em tempo real. O número de elementos lógicos consumidos pelos métodos comparados depende diretamente da forma como os algoritmos são descritos em Linguagem de Descrição de *Hardware*, não sendo possível determinar esses valores de maneira precisa antes da efetiva implementação.

Entretanto, existem três tipos de elementos cuja quantidade pode ser estimada previamente ao desenvolvimento do circuito em lógica combinacional: multiplicadores, adicionadores e multiplexadores. Como todos os métodos analisados utilizam pesos para realizar a estimação, o uso de multiplicadores e adicionadores é inerente a todas as abordagens. Adicionalmente, métodos como o apresentado em Duarte (2019), o método proposto e a rede neural utilizam estruturas de decisão (equivalentes à instrução *if-else* em lógica de programação), as quais são implementadas por meio de multiplexadores.

A Tabela 6.2 apresenta um resumo dos recursos necessários para a implementação

em FPGA dos quatro métodos avaliados, considerando as mesmas topologias utilizadas na Seção 6.4. Conforme esperado, o método LS apresenta a menor demanda por recursos computacionais, seguido pelo método proposto em Duarte (2019), que adiciona apenas um multiplexador. O método proposto neste trabalho ocupa a terceira posição, demandando aproximadamente 113% a mais de multiplicadores e 125% a mais de adicionadores em relação ao método LS, além de empregar 5 multiplexadores, correspondentes às funções de ativação dos neurônios na camada oculta. Por sua vez, a rede neural convencional apresenta um crescimento expressivo no uso de recursos computacionais, evidenciando a necessidade da simplificação adotada no método proposto para alcançar desempenho semelhante com menor custo de implementação.

Tabela 6.2 – Recursos computacionais necessários para cada um dos métodos.

Método	Multiplicadores	Adicionadores	Multiplexadores
LS	31	32	0
Duarte (2019)	31	32	1
Faria (2017)	168	181	12
Método Proposto	66	72	5

Fonte: Do autor

Cabe destacar que a utilização da topologia de rede neural mais simples, identificada na Seção 6.3, reduz de forma significativa a quantidade de multiplicadores necessários. A adoção de redes mais complexas resulta em um aumento substancial do uso de recursos, sem que haja um ganho proporcional de desempenho, o que reforça a escolha da arquitetura proposta.

6.7 CONCLUSÕES PARCIAIS DO CAPÍTULO

O método proposto apresentou desempenho superior em todos os cenários avaliados. O filtro FIR de deconvolução mostrou-se eficaz na obtenção de uma estimativa inicial, enquanto a rede neural incorporada ao método foi responsável por determinar o limiar da função *shrinkage*, possibilitando a aplicação de técnicas baseadas em teoria da esparsidade no processamento em tempo real dos sinais.

Para que uma rede neural estimadora isolada atinja resultados comparáveis, faz-se necessária uma arquitetura significativamente mais complexa, o que implica maior custo computacional em sua implementação.

Foi possível obter um modelo genérico, com pesos fixos, capaz de operar adequadamente em todos os canais do TileCal, eliminando a necessidade de treinamentos específicos para cada cenário.

7 CONCLUSÃO

A proposta inicial deste trabalho consistiu no desenvolvimento de um método de reconstrução de energia no contexto da Física de Altas Energias, em especial para o TileCal. O método proposto fundamenta-se na teoria da esparsidade, adaptada para processamento de dados em tempo real, ou seja, compatível com o primeiro nível de *trigger* da nova eletrônica do HL-LHC. Considerando os diferentes cenários de colisão previstos para a próxima fase, buscou-se um método versátil, capaz de apresentar bom desempenho em diferentes níveis de ocupação das células.

Métodos baseados em esparsidade já foram explorados em processamentos *offline* de sinais. Um deles, também fundamentado nesse princípio, foi desenvolvido para aplicação em sistemas de processamento em tempo real, mas apresenta algumas limitações, em especial no que se refere à função *shrinkage*.

No presente trabalho, propôs-se analisar o limiar da função *shrinkage* a partir de métodos iterativos da teoria da esparsidade. Identificou-se que diferentes valores de limiar poderiam ser atribuídos a cada amostra por meio de uma rede neural MLP, eliminando assim a necessidade do processo iterativo, o qual inviabilizaria sua utilização em sistemas de processamento *online*.

Os resultados obtidos demonstraram um método de estimação de energia aplicável em processamento *online*, que apresenta desempenho superior em relação às abordagens analisadas em estudos anteriores.

O método proposto caracteriza-se por sua estrutura simples de ser implementada em *hardware*, como FPGAs. A utilização de um filtro FIR é uma técnica amplamente adotada em aplicações de tempo real, e a inclusão de uma rede neural *feed-forward* com poucos parâmetros mantém a simplicidade do método nesse contexto. Além disso, a aplicação da função *shrinkage* é de fácil implementação, uma vez que depende apenas de comparadores e de um circuito de adição.

Outra característica relevante do método é sua versatilidade. Em situações de baixa ocupação, atua como um filtro capaz de eliminar artefatos indesejados, enquanto em condições de alta ocupação exerce o papel de corretor de amplitude.

Quando comparado a uma rede neural pura, diversas vantagens podem ser ressaltadas. Para alcançar resultados equivalentes aos do método proposto, seria necessária uma arquitetura neural mais complexa, com maior número de parâmetros, o que impactaria diretamente os recursos de *hardware* exigidos para a implementação. Essa superioridade de desempenho decorre da forma de treinamento adotada, viabilizada pelo entendimento da rede neural como um mecanismo de controle do limiar derivado da teoria da esparsidade. Outro aspecto importante é que, em uma rede neural pura, todo o processo de estimação encontra-se encapsulado, enquanto no método proposto a rede é utilizada apenas para a

estimação do valor de limiar da função *shrinkage*. Dessa forma, garante-se maior controle sobre as etapas envolvidas na reconstrução do sinal.

A possibilidade de aplicar o método proposto utilizando uma topologia e um conjunto de pesos fixos para todos os cenários constitui uma característica de grande relevância para a operação do TileCal. O treinamento específico para cada canal torna-se uma tarefa onerosa, dada a elevada quantidade de canais do detector.

7.1 TRABALHOS FUTUROS

Neste trabalho foi apresentado um novo método acompanhado de uma série de análises detalhadas. À partir dos resultados obtidos, é possível identificar novas frentes de estudo e contribuições que podem dar continuidade à linha de pesquisa desenvolvida.

A sequência natural deste trabalho consiste na implementação do método em *hardware*, utilizando FPGA. A avaliação do número de elementos lógicos requeridos, da frequência máxima de operação e dos atrasos de resposta é fundamental para a determinação da viabilidade e da escalabilidade do método em um ambiente real de aplicação.

O método proposto foi desenvolvido considerando um cenário no qual apenas sinais positivos são de interesse. Entretanto, é possível estender a abordagem de forma a permitir o processamento de sinais tanto positivos quanto negativos, viabilizando sua aplicação em contextos distintos da Física de Altas Energias.

Por fim, há espaço para avançar na investigação de novas estratégias para a determinação do limiar da função *shrinkage* sem o emprego de redes neurais, preservando, ao mesmo tempo, a característica de processamento de dados em tempo real exigida pela aplicação.

REFERÊNCIAS

- Aad, G. et al. (2010). The atlas simulation infrastructure. *The European Physical Journal C*, 70(3):823–874. Acessado em 11 de agosto de 2025.
- Aberle, O., Béjar Alonso, I. e., Brüning, O. e., Fessia, P. e., Rossi, L. e., Tavian, L. e., and Zerlauth, M. e. (2020). High-luminosity large hadron collider (hl-lhc): Technical design report. Technical Report CERN-2020-010, CERN. Acessado em 8 de agosto de 2025.
- Anderson, K., Pilcher, J., Sanders, H., Tang, F., Berglund, S., Bohm, C., Holmgren, S.-O., Jon-And, K., Blanchot, G., and Cavalli-Sforza, M. (1998). Front-end electronics for the atlas tile calorimeter. Technical note, University of Chicago.
- Andrade Filho, L. M. d. (2009). *Detecção e Reconstrução de Raios Cósmicos Usando Calorimetria de Altas Energias*. Tese de doutorado, Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro.
- ATLAS Collaboration (2008). Overview of the atlas detector. *JINST*, 3:S08003. Acessado em 8 de agosto de 2025.
- ATLAS Collaboration (2010). The atlas simulation infrastructure. *The European Physical Journal C*, 70:823–874.
- ATLAS Collaboration (2013). Technical design report for the phase-i upgrade of the atlas tdaq system. Technical Report ATLAS TDR 023, CERN-LHCC-2013-018, CERN, Geneva, Switzerland.
- ATLAS Collaboration (2017a). Technical design report for the phase-ii upgrade of the atlas tile calorimeter. ATLAS Technical Design Report ATLAS-TDR-028; CERN-LHCC-2017-019, CERN, Geneva. Acessado em 8 de agosto 2025.
- ATLAS Collaboration (2017b). Technical design report for the phase-ii upgrade of the atlas tile calorimeter. Technical Report ATLAS-TDR-028; CERN-LHCC-2017-019, CERN. Acessado em 8 de agosto de 2025.
- Barbosa, D. P. (2017). *Estimação de Energia para Calorimetria em Física de Altas Energias Baseada em Representação Esparsa*. Tese de doutorado, Universidade Federal de Juiz de Fora.
- Bianchi, R. M. and Collaboration, A. (2022). ATLAS experiment schematic or layout illustration. General Photo.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York, NY, USA.
- Brüning, O. and Zerlauth, M. (2024). Hl-lhc counts down to ls3. CERN Courier. Acessado em 28 de agosto de 2025.
- Budagov, Y. A., Khorauli, G. D., Khubua, J., Khukhunaishvili, A., Kulchitskii, Y. A., and Solodkov, A. (2006). Study of TileCal Sampling Fraction for Improvement of Monte-Carlo Data Reconstruction.

- Camplani, A. (2020). Atlas track trigger: On behalf of the atlas tdaq collaboration. Apresentação no Common ATLAS CMS Electronics Workshop. Disponível em: <https://cds.cern.ch/record/2719419/files/ATL-DAQ-SLIDE-2020-120.pdf>. Acessado em 8 de agosto de 2025.
- Candes, E. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215.
- Carrió, F., Castillo, V., Ferrer, A., Fiorini, L., Hernández, Y., Higón, E., Mellado, B., Moreno, P., Reed, R., Solans, C., Valero, A., and Valls, J. A. (2014). The srod module for the atlas tile calorimeter upgrade demonstrator. In *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3.
- CERN (2013). New results indicate new particle higgs boson. <https://home.cern/news/news/physics/new-results-indicate-new-particle-higgs-boson>. Acessado em 8 de agosto de 2025.
- Cerqueira, A. S. (2015). Tile calorimeter upgrade program for the luminosity increasing at the lhc. Acessado em 8 de agosto de 2025.
- Chen, K. (2017). Bnl readout & trigger/daq road map. CPAD2017, Novo México.
- Chen, S. and Guo, W. (2023). Auto-encoders in deep learning—a review with new perspectives. *Mathematics*, 11(8).
- Collaboration, A. (2015). Atlas phase-ii upgrade scoping document. Technical Report CERN-LHCC-2015-020 / LHCC-G-166, CERN. Disponível em: <https://cds.cern.ch/record/2055248>. Acesso em: 11 ago. 2025.
- Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.
- Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627.
- Duarte, J. B. d. S. (2016). Técnicas de deconvolução aplicadas à estimação de energia online em calorimetria de altas energias em condições de alta taxa de eventos. In *Anais do XXI Congresso Brasileiro de Automática*, pages 2612–2617, [S.l.]. SBA/PPGEE-UFES.
- Duarte, J. P. B. d. S. (2019). *Deconvolução Free-running Aplicada à Estimação Online de Energia na Calorimetria do ATLAS*. Tese de doutorado, Universidade Federal de Juiz de Fora.
- Elad, M. (2010). *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, New York.
- Fabjan, C. W. and Gianotti, F. (2003). Calorimetry for particle physics. *Rev. Mod. Phys.*, 75:1243–1286.
- Faria, M. H. M. d. (2017). Estimação de energia no calorímetro hadrônico do atlas para o primeiro nível de trigger utilizando redes neurais artificiais. Dissertação de mestrado, Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brasil.

Ferbel, T. (1987). *Experimental Techniques in High-Energy Nuclear and Particle Physics*. World Scientific Publishing Co., Singapore.

Ferreira, B. C. (2009). Detecção de raios cósmicos com calorimetria de altas energias.

Fullana, E. et al. (2005). Optimal filtering in the atlas hadronic tile calorimeter. Technical report, CERN, Geneva. Acessado em 11 de agosto 2025.

Gagnon, P. (2016). *Who Cares about Particle Physics? Making Sense of the Higgs Boson, the Large Hadron Collider and CERN*. Oxford University Press, Oxford.

GALLART, S. B. i. (1998). *Calibration and Light Yield measurements of Steel-plastic scintillator calorimeters for the ATLAS detector*. Tese (doutorado), Universitat Autònoma de Barcelona, Departament de Física, Institut de Física d'Altes Energies, Bellaterra: Barcelona. Disponível em: <https://cds.cern.ch/record/683727/files/tilecal-98-167.pdf>. Acessado em 8 ago. 2025.

Goff, F. L. (2020). Daq in atlas. Apresentação em nome do ATLAS TDAQ. Disponível em: <http://cds.cern.ch/record/2719417/files/ATL-DAQ-SLIDE-2020-118.pdf>. Acesso em: 11 de agosto de 2025.

Gonçalves, G. I. et al. (2021). Energy reconstruction techniques in tilecal under high pile-up conditions. Technical note / proceedings, CERN. Acessado em 28 agosto de 2025.

Gonçalves, G. I., Peralva, B. S., Andrade Filho, L. M., Cerqueira, A. S., and Seixas, J. M. (2018). Estimação de energia para o calorímetro de telhas do atlas baseada na filtragem de wiener-hopf. In *Anais do XXI Encontro Nacional de Modelagem Computacional (ENMC)*, Búzios, Brasil.

Guatelli, S., Cutajar, D., Oborn, B., and Rosenfeld, A. B. (2011). Introduction to the geant4 simulation toolkit. *AIP Conference Proceedings*, 1345(1):303–322. Acessado em 11 de agosto de 2025.

Haykin, S. (2001). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ, 2nd edition.

Haykin, S. S. (1996). *Adaptive Filter Theory*. Prentice Hall, Upper Saddle River, NJ, 3rd edition.

Haykin, S. S. (2014). *Adaptive Filter Theory*. Pearson, Upper Saddle River, 5 edition.

Iakovidis, D. K. et al. (2021). Roadmap on signal processing for next generation measurement systems. *Measurement Science and Technology*, 33(1):012002.

Iiyama, Y. (2017). *The LHC and the CMS Experiment*, pages 47–83. Springer International Publishing, Cham.

Izzo, V. (2020). Atlas upgrades. Apresentação na 8th Annual Large Hadron Collider Physics Conference (LHCP 2020). Disponível em: <https://cds.cern.ch/record/2732959/files/ATL-UPGRADE-PROC-2020-001.pdf>. Acessado em 8 de agosto de 2025.

- Mitra, S. K. (2006). *Digital Signal Processing: A Computer-Based Approach*. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill, 3 edition.
- Mlynarikova, M. (2017). Performance of the atlas hadronic tile calorimeter. Acessado em 8 de agosto de 2025.
- Novakova, J. (2014). *Standard Model Measurements with the ATLAS Detector: Monte Carlo Simulations of the Tile Calorimeter and Measurement of the $Z \rightarrow$ Cross Section*. Springer International Publishing.
- Oppenheim, A. V., Willsky, A. S., and Nawab, S. H. (2010). *Signals and Systems*. Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition.
- Palka, M. (2016). Operation of the enhanced atlas first level calorimeter trigger at the start of run-2. *Journal of Instrumentation*, 11(02):C02059. Acessado em 11 de agosto 2025.
- Paschoalin, T. C. (2016). Reconstrução de energia em calorímetros operando em alta luminosidade usando estimadores de máxima verossimilhança. Master's thesis, Universidade Federal de Juiz de Fora.
- Pastore, F. (2016). The atlas trigger system: Past, present and future. In *Proceedings of the 37th International Conference on High Energy Physics (ICHEP)*. Elsevier B.V. Resumo da evolução do sistema de trigger do ATLAS (Run-1: 7/8TeV, luminosidade pico $0,7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$). Acessado em 8 de agosto de 2025.
- Pequenao, J. (2008). Event Cross Section in a computer generated image of the ATLAS detector.
- Peralva, B. S.-M. (2015). *Reconstrução de energia para calorimetria em Física de Altas Energias*. PhD thesis, Universidade Federal de Juiz de Fora. Tese de Doutorado.
- Peralva, B. S.-M., Cerqueira, A. S., Filho, L. M., and Seixas, J. M. (2010). Optimal signal selection for a highly segmented calorimeter. In *Proceedings of INTERNOISE 2010*. Acessado em 8 de agosto de 2025.
- Quirino, T. M. (2022). *Aprendizado online de redes neurais artificiais aplicadas à reconstrução de energia no calorímetro hadrônico do ATLAS*. Tese de doutorado, Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brasil.
- Rajamanoharan, S. et al. (2024). Improving dictionary learning with gated sparse autoencoders. *ArXiv*, abs/2404.16014.
- Rimoldi, A. (2011). Simulation strategies for the atlas experiment at lhc. Disponível em: <https://cds.cern.ch/record/1322150/files/ATL-SOFT-PROC-2011-016.pdf>. ATL-SOFT-PROC-2011-016.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- Service graphique, C. (2014). Overall view of the LHC. General Photo.
- Solans Sánchez, C. A. (2019). The felix detector interface for the atlas tdaq upgrades and its deployment in the itk demonstrator setup. Apresentação no ACAT 2019 (on behalf of the ATLAS Collaboration). Acessado em 8 de agosto de 2025.

- Starck, J.-L., Murtagh, F., and Fadili, J. M. (2010). *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, Cambridge, UK. Discusses sparse representations (e.g., Fourier, wavelet, DCT, dictionary learning) and applications such as compressed sensing. Acesso em: 12 de agosto de 2025.
- Sánchez, C. A. S. (2010). *Implementation of the ROD Crate DAQ Software for the ATLAS Tile Calorimeter and a Search for a MSSM Higgs Boson decaying into Tau pairs*. Doutorado em física, Universitat de València – CSIC, València.
- Tropp, J. A. (2004). Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242.
- Valero, A. et al. (2007). Installation and commissioning of the TileCal Read-Out Drivers. In *2007 IEEE Nuclear Science Symposium and Medical Imaging Conference*, IEEE Nucl.Sci.Symp.Conf.Rec., pages 490–493.
- Van Loan, C. F. and Golub, G. H. (1996). *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 3 edition.
- Vazquez, W. P. and on behalf of the ATLAS Collaboration (2017). The atlas data acquisition system in lhc run 2. Technical Report ATL-DAQ-PROC-2017-007, ATLAS Experiment.
- Wan, E. A. (1991). Temporal backpropagation: An efficient algorithm for finite impulse response neural networks. In Touretzky, D. S., Elman, J. L., Sejnowski, T. J., and Hinton, G. E., editors, *Connectionist Models*, pages 131–137. Morgan Kaufmann. Acessado em 11 de agosto de 2025.
- Widrow, B. (1988). Learning algorithms in neural networks. In *Annual Meeting of the Optical Society of America, Technical Digest Series*, page THX1, Santa Clara, CA. Optica Publishing Group.
- Wiedemann, H. (2015). *Particle Accelerator Physics*. Springer, 4 edition.
- Wielers, M. (2015). Introduction to trigger and data acquisition. <https://indico.cern.ch/event/518474/?print=1>. Disponível em: <https://indico.cern.ch/event/518474/?print=1>. Acessado em 11 de agosto de 2025.
- Wigmans, R. (2017). *Calorimetry: Energy Measurement in Particle Physics*. Oxford University Press, New York, 2 edition.
- Xiong, P., Zhang, K., Shi, Z., et al. (2025). Deep dictionary learning with reconstruction for texture recognition. *Scientific Reports*, 15:31164.
- Yue, X. (2019). Tile TDAQ interface module for the Phase-II Upgrade of the ATLAS Tile Calorimeter. Technical report, CERN, Geneva. Acessado em 8 de agosto de 2025.
- Zakareishvili, T. (2020). Upgrade of the atlas hadronic tile calorimeter for the high luminosity lhc. *Journal of Instrumentation*, 15(09):C09003. Acessado em 8 de agosto de 2025.
- Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D. (2016). A survey of sparse representation: algorithms and applications. *IEEE Access / arXiv*.

APÊNDICE A – Produção Bibliográfica

Destacam-se neste capítulo as contribuições acadêmicas importantes no período de desenvolvimento deste trabalho.

- Artigos publicados em Periódicos

PASCHOALIN, T.C.A.; QUIRINO, T.M.; DUARTE, J.P.B.S; ANDRADE FILHO, L.M.de. **Learned Soft-Thresholding for Streaming Sparse Signal Recovery in Nuclear Instrumentation**. IEEE Transactions on Nuclear Science 2025, Early Access. <https://doi.org/10.1109/TNS.2025.3644239>

RIMES, S.; PERALVA, B.; LIBOTTE, G.; PASCHOALIN, T.; ANDRADE FILHO, L.; SEIXAS, J.M. **A Maximum Likelihood Approach for Signal Estimation and Data Quality Assessment for High-Energy Calorimeter Systems**. IEEE Transactions on Nuclear Science 2025, vol. 72, no. 12, pp. 3801-3808. <https://doi.org/10.1109/TNS.2025.3630987>

QUIRINO, T.M.; PASCHOALIN, T.C.A.; GONÇALVES, G.I.; LISBOA, P.H.B.; DE ANDRADE FILHO, L.M.; PERALVA, B.S.-M. **Online Pulse Compensation for Energy Spectrum Determination: A Pole-Zero Cancellation and Unfolding Approach**. Electronics 2025, 14, 493. <https://doi.org/10.3390/electronics14030493>

- Artigo em Congresso Internacional

PASCHOALIN, T.C.A; DIAS, U.R.F.; AGUIAR, M.S.; SANTOS, D.F.; QUIRINO, T.M.; ANDRADE FILHO, L.M.de. **Uncorrelated Pseudo-Random Generator for FPGA**. Em: Integrated Circuits and Systems Design Symposium (SBCCI), 2025, Manaus. The 38th Integrated Circuits and Systems Design Symposium (SBCCI), 2025.

- Artigos em Congresso Nacional

PASCHOALIN, T.C.A; ALMEIDA, A.; LIMA, M.; ANDRADE FILHO, L.M.de. **Estudo de um Método para Detecção de Múons no Experimento ATLAS Baseado em Rede Neural Convolucional**. Em: Congresso Brasileiro de Automática (CBA), 2024, Rio de Janeiro. Congresso Brasileiro de Automática (CBA), 2024

LIMA, M.; PASCHOALIN, T.C.A; ALMEIDA, A. **Controle com Droop de Tensão Baseado em Impedância Virtual para uma Microrrede CC Isolada com Geração Fotovoltaica e Baterias**. Em: Congresso Brasileiro de Automática (CBA), 2024, Rio de Janeiro. Congresso Brasileiro de Automática (CBA), 2024.

LUNA, F.C.; DIAS, U.F.; LISBOA, P.H.B.; PASCHOALIN, T.C.A.; QUIRINO, T.M.; ANDRADE FILHO, L.M.de. **Real-Time FPGA-Based for the Tile Calorimeter Readout System in the ATLAS Experiment**. Em: Encontro Nacional de Modelagem Computacional (ENMC), 2024, Ilhéus. XXVII Encontro Nacional de Modelagem Computacional (ENMC), 2024.

PASCHOALIN, T.C.A.; QUIRINO, T.M.; DUARTE, J.P.B.S; ANDRADE FILHO, L.M.de. **Método Iterativo Utilizando Teoria de Esparsidade à Estimação de Energia para um Calorímetro Operando com Alta Taxa de Eventos**. Em: Encontro Nacional de Modelagem Computacional (ENMC), 2024, Ilhéus. XXVII Encontro Nacional de Modelagem Computacional (ENMC), 2024.

PASCHOALIN, T.C.A.; QUIRINO, T.M.; ANDRADE FILHO, L.M.de. **Implementação em FPGA do Circuito do Condicionamento de Sinal do Calorímetro Hadrônico do ATLAS**. Em: Simpósio Brasileiro de Automação Inteligente (SBAI), 2025, São João Del Rei. XVII Simpósio Brasileiro de Automação Inteligente (SBAI), 2025.

OLIVEIRA, V.P.D.de; PASCHOALIN, T.C.A.; ANDRADE FILHO, L.M.de. **Simulador de Pulsos para o Upgrade do Calorímetro Hadrônico do Experimento ATLAS no CERN**. Em: Simpósio Brasileiro de Automação Inteligente (SBAI), 2025, São João Del Rei. XVII Simpósio Brasileiro de Automação Inteligente (SBAI), 2025.

OLIVEIRA, V.P.D.de; LISBOA, P.H.B.; DIAS, U.R.F.; PASCHOALIN, T.C.A.; QUIRINO, T.M.; ANDRADE FILHO, L.M.de. **Multi-memory approach for Monte Carlo simulation in FPGA**. Em: Students Forum (SFORUM), 2025, Manaus. 25th Students Forum (SFORUM), 2025.

LUNA, F.C.; PASCHOALIN, T.C.A.; ANDRADE FILHO, L.M.de. **Design and Implementation of a Digital IIR Filter with Floating-Point Representation**. Em: Students Forum (SFORUM), 2025, Manaus. 25th Students Forum (SFORUM), 2025.

PASCHOALIN, T.C.A; OLIVEIRA, V.P.D.; ANDRADE FILHO, L.M.de.; CARDINOT, D.A.; LIBOTTE, G.B. **Simulador De Front-End do Calorímetro Hadrônico do Detector do ATLAS no LHC Usando Fpga e Processador Arm Embarcado**. Em: Encontro Nacional de Modelagem Computacional (ENMC), 2025, Montes Claros. XXVIII Encontro Nacional de Modelagem Computacional (ENMC), 2025.

TERREZO, A.; MORETT, G.B.; ANDRADE FILHO, L.M.de; PERALVA, B.S.-M.; QUIRINO, T.M; PASCHOALIN, T.C.A; LOPES, C.W. **Análise de Variação dos Parâmetros Temporais do Pulso do Calorímetro Cintilante do ATLAS do Calo Paramétrica em Relação a Variação do Circuito de Condicionamento de Sinal**. Em: Encontro Nacional de Modelagem Computacional (ENMC), 2025, Montes Claros. XXVIII Encontro Nacional de Modelagem Computacional (ENMC), 2025.