



REPRESENTAÇÃO DA INFORMAÇÃO DE CALORIMETRIA DE ALTAS
ENERGIAS UTILIZANDO AUTO-CODIFICADORES

Carlos Eduardo Covas Costa

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

Orientador: José Manoel de Seixas

Rio de Janeiro
Dezembro de 2019

REPRESENTAÇÃO DA INFORMAÇÃO DE CALORIMETRIA DE ALTAS
ENERGIAS UTILIZANDO AUTO-CODIFICADORES

Carlos Eduardo Covas Costa

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

Orientador: José Manoel de Seixas

Aprovada por: Prof. José Gabriel Rodriguez Carneiro Gomes
Prof. Gilson Antônio Giraldi
Prof. Wagner de Paula Carvalho
Prof. Eduardo Pestana de Aguiar
Dr. Natanael Nunes de Moura Júnior

RIO DE JANEIRO, RJ - BRASIL
DEZEMBRO DE 2019

Eduardo Covas Costa, Carlos

Representação da informação de Calorimetria de Altas Energias utilizando Auto-codificadores/Carlos Eduardo Covas Costa. – Rio de Janeiro: UFRJ/COPPE, 2019.

XIX, 154 p.: il.; 29, 7cm.

Orientador: José Manoel de Seixas

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2019.

Referências Bibliográficas: p. 118 – 136.

1. Redes Neurais.
 2. Redes Auto-codificadoras.
 3. Classificação.
 4. Aprendizado de Representação.
 5. Aprendizado de Máquina.
 6. Ensemble Learning.
 7. Física de Altas Energias.
- I. Seixas, José Manoel de. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*Em memória de minha avó
Maria de Lourdes Coutinho
Covas.
Com a infinita saudade e a terna
lembrança de seu amor e
altruísmo*

Agradecimentos

O processo de produção de uma dissertação de mestrado exige uma forte dedicação para dar conta de uma extensa demanda desde os experimentos e análises até a elaboração da produção escrita. A construção de cada etapa compõe um trabalho exaustivo e solitário, no entanto, a elaboração teórica é resultado de uma construção coletiva. Dessa forma, expresso minha enorme gratidão a todos que, de uma forma ou de outra, com seu conhecimento, paciência, companheirismo, afeto e bom senso me ajudaram a concretizá-la.

As palavras aqui descritas são mera formalidade e não expressam a magnitude da gratidão que tenho pela minha família, pelo afeto e pela compreensão das privações do convívio. Quero agradecer, em especial, aos meus avós, aos meus pais e à minha irmã, pelo incondicional apoio e pelo permanente incentivo. Ao meu pai e à minha mãe, em particular, pelo carinho e incentivo sem os quais essa caminhada não seria possível ao longo dos anos.

Quero prestar um agradecimento especial ao meu orientador, professor Dr. José Manuel de Seixas, pelo estímulo intelectual, pela seriedade, pela paciência e pela confiança, além dos professores Marcia Begalli, Luiz Caloba, Alexandre Evsukoff, Alexandre de la Vega e Tadeu Ferreira, cujas indicações e sugestões foram muito úteis para a conclusão desta pesquisa.

Agradeço também aos meus colegas do laboratório, que foram essenciais nesta caminhada: Micael e Phil, pelo companheirismo sempre presente; Júlio, pelas discussões que proporcionou, que muito me ajudaram; João Victor e Gabriel, por todo apoio com o ambiente de desenvolvimento e com a infraestrutura computacional; Werner, pela orientação e suporte em uma área de aplicação desconhecida para mim; Natanael, pela ajuda na compreensão de diversos tópicos em inteligência computacional; além de Fernando, Breno, João Guedes, Pedro, Vinicius e Luiza pelas discussões e conversas. Essas relações muito contribuíram para que, pouco a pouco, a construção dos experimentos e resultados aqui apresentados fossem efetivados. Gostaria de agradecer, também, à Paraskevi Nousi e Anastasios Tefas pela contribuição técnica.

Agradeço aos colegas de trabalho da Huawei e demais projetos, pelo apoio e compreensão. Gostaria de agradecer à Natalie, à Lorraine, ao Antonio, ao Edu-

ardo, ao Wang Heyang e à Marcela, que trabalharam diretamente comigo, por me darem muito suporte e por me suportarem durante este período intenso e complexo. Agradeço também ao Henrique, pelo incentivo ao mestrado, mencionado nas reuniões de trabalho.

Também agradeço aos amigos de longas jornadas e diversos espaços, em especial, aos amigos e professores do Colégio Pedro II, do CEFET e da UFF, companheiros sempre presentes. Agradeço também aos meus padrinhos (e também amigos) pelo estímulo, confiança e afeto sempre demonstrados.

Por último, agradeço ao Mister e demais colegas de elenco, por terem tornado tardes e noites de escrita e desenvolvimento deste trabalho mais alegres.

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

REPRESENTAÇÃO DA INFORMAÇÃO DE CALORIMETRIA DE ALTAS ENERGIAS UTILIZANDO AUTO-CODIFICADORES

Carlos Eduardo Covas Costa

Dezembro/2019

Orientador: José Manoel de Seixas

Programa: Engenharia Elétrica

Em diversas aplicações de engenharia, encontram-se ambientes que apresentam elevada taxa de amostragem, envolvendo, eventualmente, uma grande quantidade de canais de leitura com eventos raros imersos em ruído de fundo. Na física de altas energias, tais desafios estão presentes no experimento ATLAS, localizado no LHC (Large Hadron Collider), que conta com um sistema de filtragem *online* para eliminar o elevado ruído de fundo que mascara os processos físicos de interesse. O algoritmo *NeuralRinger*, que consiste em um *ensemble* de redes neurais alimentado com informação topológica de calorimetria (medição de energia), entrou em operação *online* em 2017, com o objetivo de detectar elétrons com alta eficiência. O presente trabalho estuda a aplicação de técnicas lineares e não lineares de codificação, sobre os anéis de calorimetria, com o objetivo de encontrar uma representação da informação de calorimetria para a aplicação *offline* do *NeuralRinger*, onde recursos adicionais de processamento se encontram disponíveis. Adicionalmente, avalia-se a possibilidade de se utilizar uma codificação compacta. A análise contempla a qualidade da reconstrução e a consequente capacidade discriminante. Obteve-se que a dimensão do espaço de entrada (100 anéis) pode ser reduzida para 25 componentes sem prejuízo na capacidade discriminativa e para apenas 2 dimensões, apesar de se dobrar a taxa de falso alarme (de 3,82% para 7,91%), quando codificadores não supervisionados são utilizados. Redes auto-codificadoras discriminantes, por outro lado, se mostraram capazes de representar a informação com apenas um componente, mantendo a capacidade de discriminação elétron-jato. Os resultados foram obtidos utilizando uma base de dados de simulação de Monte Carlo para uma energia de 13 TeV no centro de massa da colisão de prótons e abre perspectiva de atuação do *NeuralRinger* em diferentes análises da física do Modelo Padrão e suas possíveis extensões.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

INFORMATION REPRESENTATION IN HIGH-ENERGY CALORIMETRY USING AUTOENCODERS

Carlos Eduardo Covas Costa

December/2019

Advisor: José Manoel de Seixas

Department: Electrical Engineering

Engineering applications usually face scenarios with high sampling rates, possibly involving a large number of readout channels with rare events immersed in background noise. In high energy physics, such challenges are present in the ATLAS experiment, located in the Large Hadron Collider (LHC), which has an online filtering system to eliminate the high background noise that masks the physical processes of interest. The *NeuralRinger* algorithm, which consists of an ensemble of neural networks which receives topological calorimetry (energy measurement) information as input, came into operation in 2017 to detect electrons with high efficiency. The present work studies the application of linear and nonlinear encoding techniques on the calorimetry rings, aiming to find a representation of the calorimetry information for the offline application of *NeuralRinger*, where additional processing resources are available. In addition, the possibility of using a compact encoding is evaluated. The analysis contemplates the quality of the reconstruction and the consequent discriminating capacity. It has been found that the size of the input space (100 rings) can be reduced to 25 components without any impairment on discriminative capacity and to only 2 dimensions, despite doubling the false alarm rate (from 3.82 % to 7, 91 %), when unsupervised encoders are used. Discriminative autoencoders, on the other hand, were able to represent information with only one component, maintaining the ability to discriminate electron-jet. The results were obtained using a Monte Carlo simulation database for a 13 TeV energy at the proton collision center of mass and opens the perspective of *NeuralRinger*'s performance in different analyzes of Standard Model physics and its possible extensions.

Sumário

Sumário	ix
Lista de Figuras	xiii
Lista de Tabelas	xviii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivo	3
1.3 Resultados Obtidos	4
1.4 Organização da Dissertação	4
2 Física de Partículas	6
2.1 Unidades de Energia	7
2.2 Partículas Elementares	7
2.3 Interações Fundamentais	10
2.4 Modelo Padrão	13
2.5 O Bóson de Higgs	14
2.6 Física Além do Modelo Padrão	14
3 Aceleradores e Detectores	18
3.1 Princípios de aceleração de partículas	19
3.2 Princípios de detecção de partículas	21
3.2.1 Detecção baseada na trajetória das partículas	21
3.2.2 Detecção baseada em calorimetria	22
3.2.2.1 Interação de Partículas com o Calorímetro	23
3.2.2.1.1 Interação eletromagnética	23
3.2.2.1.2 Interação Hadrônica	26
3.2.2.2 Tipos de calorímetros	28
4 O LHC e o experimento ATLAS	30
4.1 O Grande Colisor de Hádrons	30

4.2	O Experimento ATLAS	34
4.2.1	O Sistema de Coordenadas	35
4.2.2	O Detector Interno	36
4.2.2.1	O Detector de Pixel	37
4.2.2.2	O Detector de Rastreamento por Semicondutores (SCT)	38
4.2.2.3	Detector de Rastreamento por Radiação de Transição (TRT)	39
4.2.3	O Sistema de Calorimetria	39
4.2.3.1	Calorímetro Eletromagnético (ECAL)	40
4.2.3.2	Calorímetro Hadrônico (HCAL)	43
4.2.3.3	Calorímetro Dianteiro (FCAL)	44
4.2.4	O Espectrômetro de Múons	44
4.3	Sistema de Filtragem do ATLAS	45
4.3.1	L1	46
4.3.2	HLT	46
4.3.3	Sistema de Filtragem <i>Offline</i>	48
4.3.3.1	Grandezas físicas utilizadas na seleção de elétrons . .	49
4.3.3.2	Likelihood	51
4.3.4	O Algoritmo NeuralRinger	51
5	Aprendizado de Representação Através de Auto-codificadores e Trabalhos Relacionados	54
5.1	Aprendizado de Representação	54
5.1.1	PCA e NLPCA	55
5.1.2	Redes neurais auto-codificadoras	56
5.1.3	Redes neurais auto-codificadoras com regularização	58
5.1.4	Redes neurais auto-codificadoras profundas	59
5.1.5	Redes neurais auto-codificadoras discriminantes	60
5.1.6	Codificação com Redes Neurais Recorrentes	62
5.2	Uso de Aprendizado Profundo na Física de Altas Energias	64
5.2.1	Reconstrução de Eventos	64
5.2.2	Identificação e classificação de Jatos	66
5.2.3	Seleção de eventos e análise física	68
5.2.4	Informação de calorimetria representada pela soma de células em anéis concêntricos	69
6	Método Utilizado	70
6.1	Seleção de eventos	70
6.1.1	Método <i>Tag and Probe</i>	72

6.1.2	Conjunto de dados utilizado	73
6.2	Validação cruzada	74
6.3	Normalização	75
6.4	Análise dos Modelos	75
6.4.1	Construção do <i>Ensemble</i>	76
6.4.2	Análise de Reconstrução	76
6.4.2.1	Método de análise	77
6.4.2.2	Especificação da Topologia e Treinamento	77
6.4.2.3	Figuras de Mérito	78
6.4.2.3.1	Erro de Reconstrução	78
6.4.2.3.2	Informação Mútua	79
6.4.2.3.3	Divergência Kullback-Leibler	79
6.4.3	Análise de classificação	80
6.4.3.1	Método de análise	80
6.4.3.2	Especificação da Topologia e Treinamento	80
6.4.3.3	Figuras de mérito	81
6.4.3.4	Relevância de Variáveis	82
7	Resultados	83
7.1	Dados utilizados	83
7.2	Análise de Reconstrução	85
7.2.1	Reconstrução para diferente número de dimensões codificadas	86
7.2.2	Reconstrução da Energia Total da ROI	89
7.2.3	Reconstrução dos anéis	94
7.2.4	Reconstrução das camadas longitudinais	104
7.3	Análise de Discriminação	106
7.3.1	Relevância dos anéis	106
7.3.2	Codificação não-supervisionada	106
7.3.3	Codificação através da rede auto-codificadora discriminante . .	110
8	Conclusão e Trabalhos Futuros	116
8.1	Trabalhos Futuros	117
	Referências Bibliográficas	118
A	Detalhes de Treinamento	137
B	Gráficos de Espalhamento	144

C	Produção Científica	149
C.1	Artigos Publicados em Anais de Congresso	149
C.2	Resumos Publicados em Anais de Congresso	152

Lista de Figuras

2.1	Modelo Padrão de Partículas	14
3.1	Diagrama em árvore mostrando a relação entre tipos e classificação dos detectores de partículas mais comuns	22
3.2	Visão longitudinal do desenvolvimento dos chuveiros no detector ATLAS com os componentes importantes mostrados e como eles respondem a vários tipos de partículas.	28
4.1	Complexo de aceleração do CERN	32
4.2	Diagrama ATLAS	34
4.3	Sistema cartesiano de coordenadas do ATLAS.	36
4.4	O sistema de coordenadas do ATLAS.	36
4.5	Detector Interno	37
4.6	Detector Interno	38
4.7	O sistema de calorimetria do ATLAS.	40
4.8	Corte longitudinal do ECAL.	42
4.9	Granularidade do ECAL para $\eta = 0$	43
4.10	Espectrômetro de múons do ATLAS	45
4.11	Espectrômetro de múons do ATLAS	47
4.12	Fluxograma empregado para seleção de elétrons no HLT do ATLAS.	48
4.13	Construção dos anéis pelo algoritmo NeuralRinger através das seções do calorímetro.	52
5.1	Modelo neural para análise de componentes principais não-linear (NLPCA).	56
5.2	Diagrama ilustrativo da arquitetura de uma rede auto-codificadora.	57
5.3	Diagrama ilustrativo da arquitetura e processo de treinamento de redes auto-codificadoras empilhadas (Stacked Autoencoders).	61
6.1	Conjunto de dados para todo os espaço de fase de E_t e η utilizando simulação por Monte Carlo.	71

6.2	Etapas desde a colisão ou geração das partículas até a análise física, considerando dados de colisão (fluxo da esquerda) e simulação por Monte Carlo (fluxo da direita).	72
6.3	Diagrama ilustrativo do processo de validação cruzada por jackknife.	74
7.1	Comparação entre os perfis médios, considerando o desvio padrão, ao longo dos anéis de energia formados nas camadas do sistema de calorimetria do ATLAS para sinal e o ruído de fundo na região de $20\text{GeV} < E_t < 30\text{GeV}$, $0.0 < \eta < 0.6$	84
7.2	Matriz de Correlação de Pearson e de Informação Mútua calculadas entre os anéis para verificação de presença de correlação linear e não linear, considerando todo o conjunto de dados.	85
7.3	Matriz de Correlação e de Informação Mútua calculada entre os anéis para para as classes sinal e ruído.	85
7.4	Informação mútua entre entrada e reconstrução da rede auto-codificadora para as segmentações longitudinais do Sistema de Calorimetria em função do número de neurônios empregados. No eixo da direita, o valor de SPmax medidos. A dimensão da codificação foi variada de 100 até 1.	86
7.5	Análise da qualidade da reconstrução proporcionada por diferentes modelos ao longo de diferentes dimensões codificadas. Os gráficos da esquerda (Erro médio quadrático) e do centro (informação mútua) comparam o treinamento das redes auto-codificadoras com relação à etapa de ajuste fino.	88
7.6	Análise da qualidade da reconstrução proporcionada por diferentes modelos ao longo de diferentes dimensões codificadas. o gráfico compara estas redes com outros modelos de codificação.	88
7.7	Histogramas da energia total da RoI considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificadora com 3 camadas.	90
7.8	Histogramas da energia total da seção eletromagnética, considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificadora com 3 camadas.	91
7.9	Histogramas da energia total da seção hadrônica, considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificadora com 3 camadas.	92

7.10	Histogramas da energia total da seção eletromagnética, considerando os dados de entrada e os dados reconstruídos pela rede auto-codificadora com 3 camadas, para a classe sinal <i>a</i> e para a classe ruído <i>b</i> ; e da energia total da seção hadrônica, considerando entrada e reconstrução pelo mesmo codificador, para classe sinal <i>c</i> e ruído <i>d</i>	93
7.11	Gráfico Scree e Curva de Carga das Componente Principais.	94
7.12	Comparação de diferentes dimensões de codificação com respeito à sua capacidade de reconstruir as informações originais, para cada um dos 100 anéis, quando são usadas: PCA, NLPCA, rede auto-codificadora simples e rede auto-codificadora com 3 camadas.	95
7.13	Comparação da capacidade de reconstrução da informação original considerando os 4 modelos para codificação em dimensão 2, para cada um dos 100 anéis.	96
7.14	Análise da relação entre o perfil dos anéis (vermelho), perfil reconstruído (azul) e o erro de reconstrução (RMSE) (verde) para os codificadores <i>a</i> PCA e <i>b</i> rede auto-codificada com 3 camadas para codificação em 2 dimensões.	98
7.15	Comparação da capacidade de reconstrução da informação original para cada classe, considerando os modelos <i>a</i> PCA e <i>b</i> rede auto-codificadora com 3 camadas para codificação em dimensão 2, para cada um dos 100 anéis	99
7.16	Histogramas da energia do anel 9, considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificada com 3 camadas.	100
7.17	Histogramas da energia do anel 10, considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificada com 3 camadas.	101
7.18	Histogramas da energia do anel 73, considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificada com 3 camadas.	102
7.19	Histogramas da energia do anel 74, considerando os dados de entrada e os dados reconstruídos pelos codificadores <i>a</i> PCA, <i>b</i> NLPCA, <i>c</i> rede auto-codificadora simples e <i>d</i> rede auto-codificada com 3 camadas.	103
7.20	Relevância de variáveis considerando a contribuição individual de cada anel para o MSE do classificador.	107
7.21	Relevância de variáveis considerando a contribuição individual de cada anel para o índice SP do classificador.	108

7.22	Valor do índice soma produto para diferentes redes auto-codificadoras com múltiplas camadas com e sem o uso de ajuste fino após o treinamento camada a camada. A dimensão da codificação variou de 30 a 1. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$	108
7.23	Impacto da redução de dimensionalidade ao longo de diferentes dimensões codificadas, considerando os modelos testados.	109
7.24	Valor do índice SP para diferentes números de iterações de transformações de dados para a rede auto-codificadora discriminante, considerando apenas uma dimensão na codificação. $30\text{GeV} < E_t < 40\text{GeV}$, $0.0 < \eta < 0.6$	111
7.25	Projeção da informação anelada de calorimetria em duas componentes obtidas através da análise de componentes principais (PCA).	111
7.26	Projeção da informação anelada de calorimetria em duas componentes, obtidas através da rede auto-codificadora discriminativa com três camadas (50-25-2), para diferentes iterações.	112
7.27	Valor do índice soma produto para diferentes modelos de compactação, a dimensão da codificação variando de 30 a 1. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$	113
7.28	Valor dos parâmetros associados a cada anel pela codificação discriminante. Valores positivos (verde) são atribuídos para anéis que apresentam um valor energético apresentado pela classe ruído maior que a classe sinal, enquanto valores negativos (vermelho) são atribuídos para anéis em que a classe sinal deposita mais energia.	114
A.1	Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 1 camada.	138
A.2	Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 3 camadas.	139
A.3	Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 5 camadas.	140
A.4	Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 8 camadas.	141
A.5	Evolução do MSE em função das épocas de treinamento da primeira camada do SAE, quando 100 neurônios são colocados na camada escondida, empregando a otimização por gradiente descendente.	142
A.6	Evolução do MSE em função das épocas de treinamento da primeira camada do SAE, quando 100 neurônios são colocados na camada escondida, empregando a otimização por ADAM.	143

B.1	Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a PCA codificada em 27 dimensões.	145
B.2	Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a rede auto-codificadora profunda com 3 camadas codificada em 27 dimensões.	146
B.3	Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a PCA codificada em 2 dimensões.	147
B.4	Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a rede auto-codificadora profunda com 3 camadas codificada em 2 dimensões.	148

Lista de Tabelas

2.1	Férmions	9
2.2	Bárions e Mésons	9
2.3	As 4 forças fundamentais da natureza.	11
4.1	Parâmetros do LHC durante a Run 2.	33
4.2	Região de cobertura em η , granularidade e número de canais de leitura das camadas dos calorímetros	41
4.3	Definição das grandezas empregadas pelo ATLAS para discriminação de elétrons.	50
6.1	Número de anéis por camada dos calorímetros. Sendo PS o Calorímetro Pré-Amostrador; EM1,2,3 referentes às camadas do Calorímetro Eletromagnético; e HAD1,2,3 referente às camadas do Calorímetro Hadrônico.	76
6.2	Fronteiras para as regiões de η e ET empregado no ajuste dos modelos para o <i>NeuralRinger</i> . Um total de 25 regiões são utilizadas para compor o <i>ensemble</i> de redes neurais no ambiente <i>online</i>	77
7.1	Valores de informação mútua normalizada para a codificação em 27 dimensões, considerando as camadas do calorímetro eletromagnético.	104
7.2	Valores de informação mútua normalizada para a codificação em 27 dimensões, considerando as camadas do calorímetro hadrônico.	104
7.3	Valores de informação mútua normalizada para a codificação em 2 dimensões, considerando as camadas do calorímetro eletromagnético.	105
7.4	Valores de informação mútua normalizada para a codificação em 2 dimensões, considerando as camadas do calorímetro hadrônico.	105
7.5	Valores de eficiência para codificação em 25 dimensões, operando no ponto de máximo SP.	109
7.6	Valores de eficiência para codificação em 2 dimensões, operando no ponto de máximo SP	110
7.7	Valores de eficiência para codificação em 1 dimensões, operando no ponto de máximo SP	115

A.1	Quantidade de passos, MSE mínimo e tempo de execução para 1 treinamento de uma rede neural para classificação elétron-jato (<i>NeuralRinger</i>)	138
A.2	Arquitetura de redes auto-codificadoras para compressão em 2 dimensões.	139

Capítulo 1

Introdução

A área de inteligência computacional tem, atualmente, um papel muito relevante para o desenvolvimento da ciência e da tecnologia. Por ser uma área que possui aplicabilidade em praticamente qualquer domínio do conhecimento, o desenvolvimento e aprimoramento dos modelos bio-inspirados, de meta-heurísticas, de técnicas de análise multivariada e de métodos estatísticos de tomada de decisão proporcionam avanços e soluções para problemas de especialidades diversas do conhecimento científico. A utilização destas técnicas em conjunto com o conhecimento especialista permite que se alcancem níveis de desempenho nunca antes obtidos em problemas de alta complexidade. No caso da engenharia, em diversas áreas, encontram-se aplicações com elevada taxa de amostragem e alta dimensionalidade, havendo ainda a possibilidade do sinal que se deseja ser raro e imerso em elevado volume de ruído de fundo. Estas características podem ser identificadas, por exemplo, em aplicações de telecomunicações, controle, óleo e gás, eletricidade, segurança, processamento de áudio, etc.

1.1 Motivação

A física experimental de altas energias apresenta problemas desafiadores onde características como alta taxa de produção de eventos, enorme quantidade de canais de leitura e eventos extremamente raros de interesse estão presentes. Esta área da ciência busca estudar a natureza das partículas elementares e as interações fundamentais que as governam. O programa experimental desta área gira em torno de dois objetivos principais: testar o Modelo Padrão (Standard Model, SM) de partículas e campos com acuidade crescente e a busca de nova física, além deste Modelo Padrão (Beyond Standard Model, BSM). A fim de obter dados que permitam o estudo aprofundado de ambas as tarefas, aceleradores de partículas são frequentemente utilizados. Tais máquinas impulsionam partículas carregadas a velocidades e energias

elevadas e as colimam em feixes bem definidos, através de seus campos elétrico e magnético. O mais poderoso destes aceleradores, o Large Hadron Collider (LHC), está situado em um túnel de 27 km de circunferência, 175 metros abaixo do solo da fronteira entre a França e a Suíça e, até o momento, é considerada a maior máquina já construída pelo homem. Este colisionador possui 4 pontos de colisão, onde estão situados 7 detectores. O maior deles é o *A Toroidal LHC Apparatus* (ATLAS).

Os dados gerados no experimento ATLAS não estão prontos para análise física, tendo que passar por um sistema de filtragem *online* para identificação dos eventos de interesse (*trigger*). Esta filtragem torna-se fundamental devido a quantidade de dados originados nas colisões, que chega a atingir uma taxa de 70 TB/s. Como o armazenamento desta quantidade de dados não é viável e apenas uma pequena parcela destes eventos constitui interações físicas de interesse, um sistema de filtragem *online* atua com requisitos de baixo tempo de latência.

Como os eventos rejeitados pelo sistema de *trigger* não podem ser recuperados, a eficiência de detecção é aumentada com o custo de um aumento na quantidade de falso alarme. Como consequência, ao final desta etapa de filtragem, ainda há uma quantidade significativa de ruído de fundo em meio aos canais físicos de interesse. Além disso, tais dados armazenados serão destinados a diferentes objetivos de estudo de física, como bóson de Higgs, supersimetria, dimensões extras, etc. A fim de atender a estas diversas necessidades, sistemas de filtragem *offline* são empregados.

Entre os sistemas de detecção do ATLAS, encontram-se o sistema de traço, que mede a trajetória das partículas eletricamente carregadas, localizando espacialmente a ionização resultante, e o sistema de calorimetria, que mede a energia incidente das partículas, bem como o perfil da deposição de energia. O sistema de filtragem *offline* conta com as informações provenientes destes dois sistemas de detecção, no caso da análise de elétrons. Tradicionalmente, variáveis discriminantes de calorimetria, que captam características específicas da física dos processos de interesse, são utilizadas na filtragem dos eventos de interesse, através de cortes bidimensionais nestas variáveis. Mais recentemente, entretanto, os experimentos passaram a utilizar técnicas multivariadas, como o método de verossimilhança, árvores de decisão (Boosted Decision Trees, BDT) e redes neurais, com o objetivo de aumentar seu potencial discriminante.

A fim de obter estatística suficiente para observação dos processos raros, o LHC vem elevando a taxa de colisões próton-próton. Apesar de extremamente positiva para a acumulação de eventos de interesse, essa estratégia leva a uma maior ocupação dos detectores, produzindo o empilhamento de sinais de diferentes colisões. Considerando-se o aumento substancial dos níveis de empilhamento no LHC de alta luminosidade (HL-LHC), que está previsto para iniciar sua operação meados da próxima década, o desafio de filtragem de eventos de interesse se torna ainda mais

significativo. Sendo assim, os métodos de inteligência computacional têm se apresentado como opção conveniente, à medida que os experimentos buscam por soluções que ajudem a explorar eficientemente a estatística disponível.

Um fator importante que afeta o desempenho de métodos de aprendizado de máquina é a escolha da representação de dados. Portanto, muito do esforço da aplicação de algoritmos de aprendizado de máquina se concentra na transformação dos dados e no teste de diferentes técnicas de pré-processamento. Nos últimos anos, a área de aprendizado de representação, principalmente com a utilização de modelos considerados profundos, tem ganhado atenção. Alguns exemplos de sucesso apareceram nas áreas de reconhecimento de fala, reconhecimento de imagens, reconhecimento de objetos, processamento de linguagem natural, contribuindo positivamente.

Na área de física experimental de altas energias, tarefas como reconstrução de traço, seleção e identificação de objetos, análise física, estimação de energia e resiliência ao empilhamento têm sido positivamente afetadas por tais modelos, uma vez que estes permitem aprendizado automático de representação de alto nível de dados multivariados de alta dimensionalidade, devido ao elevado número de parâmetros que os constituem, permitindo melhorar a exploração de todo o potencial do conjunto de dados coletados pela estrutura complexa dos calorímetros e demais detectores.

1.2 Objetivo

Inspirados em [1], desde 1996, a proposta de informação anelada das células do calorímetro [2] vem sendo apresentada no contexto da separação de elétrons e jatos. Hoje, a proposta evoluiu para um ensemble de redes neurais especializadas em regiões do espaço de fase (posição da partícula incidente (η) e sua energia transversa (E_T)) medida pelo sistema de calorimetria do ATLAS. O chamado *NeuralRinger* entrou em operação, em 2017, no segundo nível de seleção de elétrons da filtragem *online*. Entretanto, a adoção do *NeuralRinger* para atuação *online* sem uma contrapartida na análise *offline* inverte a lógica empregada nos experimentos, que, normalmente, degradam um modelo atuante no *offline* para adaptação aos requisitos de latência do *trigger*.

O presente trabalho busca utilizar o ambiente *offline* para estudar diferentes codificações da informação de calorimetria composta pelos anéis utilizados pelo algoritmo *NeuralRinger* para teste de hipótese. Desta forma, é feita a avaliação de diferentes codificadores, comparando os modelos mais simples com codificadores não lineares. O objetivo é testar se codificadores que utilizam transformações não-lineares apresentam boa capacidade de reconstrução da informação de calorimetria, quando esta é comprimida para baixa dimensão, e se a representação obtida por tais codificadores permite a redução do espaço original mantendo alta eficiência de

detecção de sinal sem aumentar o falso alarme ou encontrar uma representação que contribua para um resultado superior de discriminação sinal/ruído. Sendo assim, é possível aprofundar o estudo da proposta de utilização do *NeuralRinger* em ambiente *offline*, ao passo que se possibilita, também, a utilização de informação codificada para alimentar o modelo em operação no ambiente *online*, no caso de se encontrar uma representação que contribua para discriminação elétron-jato.

1.3 Resultados Obtidos

Este trabalho permitiu estudar diferentes modelos de codificação para representação da informação anelada de calorimetria utilizada pelo *NeuralRinger*. O uso de métodos de codificação linear permitiu reduzir a dimensão do espaço de entrada de 100 anéis para 25 componentes de codificação sem prejuízo na capacidade discriminativa. Em um espaço com menos dimensões, por outro lado, a utilização de métodos de codificação que fazem uso de transformações não lineares apresentou vantagens, permitindo a redução do espaço original de 100 para apenas 2 dimensões, com um aumento na taxa de falso alarme de 3,82% para 7,91%. Por outro lado, o modelo de rede neural auto-codificadora discriminante conseguiu manter a capacidade discriminativa com apenas uma dimensão, utilizando apenas uma camada intermediária. Por reduzir o espaço de entrada para apenas uma dimensão, a representação encontrada ofereceu interpretação física, além de possibilitar a customização do classificador para diferentes análises. Sendo assim, os resultados apontam para uma possibilidade de se utilizar esta variável adicionando-a aos modelos empregados atualmente na filtragem *offline*, como a Likelihood, a fim de incorporar a informação anelada de calorimetria ao teste de hipótese empregado pela mesma, potencializando a capacidade de discriminação, com a inserção da informação que ainda não é utilizada no processo atual, sem aumentar significativamente a complexidade da solução. A representação encontrada permite, também, uma codificação dos anéis em baixa dimensão para atuação no ambiente de filtragem *online*.

1.4 Organização da Dissertação

O capítulo 2 apresenta uma breve introdução à física de partículas, descrevendo as unidades utilizadas, as partículas elementares e as interações fundamentais, além de descrever o Modelo Padrão, bem como teorias hipotéticas que vão além do modelo e servem de guia teórico para os experimentos realizados no LHC. O capítulo 3 discute os princípios teóricos de aceleradores e detectores, ferramentas que compõem o aparato complexo necessário para realizar os experimentos de física de altas energias. O capítulo 4 descreve aspectos importantes do acelerador LHC e apresenta

detalhes da composição do sistema de detecção do experimento ATLAS, com foco no sistema de calorimetria e no sistema de filtragem. Neste capítulo o algoritmo *NeuralRinger* será apresentado. O capítulo 5 faz uma revisão dos métodos de aprendizado de representação comuns na literatura e apresenta exemplos de aplicação na área de física de altas energias. No capítulo 6 o método utilizado para avaliação da qualidade de reconstrução e capacidade discriminativa dos experimentos e o conjunto de dados utilizado são descritos. Os resultados são analisados no capítulo 7 e as conclusões são apresentadas no capítulo 8.

Capítulo 2

Física de Partículas

Em Miletus, uma das colônias do império grego, costa oeste do que hoje é conhecido como Turquia, cerca de 650 anos A.C., acreditava-se que o mundo em que vivemos, aparentemente complexo, era intrinsecamente simples e que essa simplicidade poderia ser explicada através de raciocínio lógico [3]. Dois séculos depois, Democritus de Abdera, discípulo de Leucippus, postulava que toda a matéria seria composta de pequenas partículas que não podem ser destruídas ou quebradas, chamando-as de átomo, que significa indivisível [4]. Aristóteles, por outro lado, acreditava que a matéria era contínua, sendo possível dividi-la em porções cada vez menores. Acreditava também que toda a matéria do universo era composta dos quatro elementos básicos: terra, ar, fogo e água [5], e na existência de forças atuantes, como a gravidade, que faria a terra e a água afundarem.

A disputa entre as duas escolas de pensamento perdurou por séculos, sem qualquer comprovação experimental, até o início do século XX, quando Einstein, favorecendo a teoria atomista, publicou a primeira descrição matemática do movimento Browniano ¹ [7]. Esses átomos acabaram sendo divisíveis em elétrons, prótons e nêutrons. Estes dois últimos, por sua vez, apresentando, também uma subestrutura composta de quarks e glúons, que em conjunto com outros bósons e férmions são os constituintes mais fundamentais da matéria que conhecemos, descrita, atualmente, pelo modelo padrão de física de partículas[8].

Embora seja comum atribuir o início da história a Demócrito e aos atomistas gregos, alguns físicos discordam [9], pontuando que suas especulações metafísicas não têm nada em comum com a ciência moderna e atribuindo, então, o nascimento da física de partículas elementares à descoberta do elétron por JJ Thomson em 1897.

¹O estudo do movimento Browniano foi publicado em 1905, ano que ficou conhecido como *Annus mirabilis* de Einstein. Os 4 artigos publicados por ele neste ano, em conjunto com o trabalho de Max Planck na virada do século, transformaram radicalmente nosso entendimento do universo e marcaram o início de uma revolução científica que percorreu o século XX. [6]

De uma forma ou de outra, a busca fascinante das fronteiras internas do Universo, desde o nascimento da filosofia grega, há 25 séculos, até os tempos atuais, culminou nas principais mudanças epistemológicas da História da Ciência. [10] [11]

O objetivo deste capítulo é percorrer a teoria básica de física de partículas, descrevendo as unidades básicas utilizadas no campo, as partículas elementares e as interações fundamentais, bem como fazer uma breve descrição das teorias hipotéticas pesquisadas que vão além do modelo padrão e guiam os experimentos de colisão de partículas, como os do LHC.

2.1 Unidades de Energia

Devido à escala subatômica das partículas elementares, as unidades normalmente utilizadas na mecânica clássica, como gramas e joules, por exemplo, são relativamente grandes. Portanto, a unidade de energia padrão utilizada na física de partículas é o elétron-Volt (eV) e seus derivados ², que representa a energia adquirida por um elétron quando acelerado através de uma diferença de potencial de 1 Volt, sendo $1\text{eV} = 1.6 \times 10^{-19}$ Joules ³. Massa e energia são equivalentes, de acordo com a equação de Einstein

$$E = mc^2 \tag{2.1}$$

Sendo assim, na física de partículas, o momento é medido em MeV/c e massa em MeV/c^2 ⁴.

2.2 Partículas Elementares

Partículas subatômicas são divididas em bósons e férmions, dependendo do seu spin ⁵. Os bósons descobertos possuem spin inteiro igual a 0 ou a 1, mas, teoricamente, os bósons podem ter spin com valor igual a múltiplos inteiros. Férmions

²Ao passo que a física nuclear utiliza keV, a física de partículas utiliza MeV, GeV ou TeV tipicamente [9].

³O maior acelerador de partículas construído até então, o LHC, possui 14TeV no centro de colisão. Se tal valor fosse convertido para Joule, teríamos $22,4 \times 10^{-7}$ Joules

⁴A massa do próton, por exemplo, é $938MeV/c^2 = 1,67 \times 10^{-24}$ g

⁵O Teorema spin-estatística [12] [13] afirma que as partículas com spin que são múltiplos de meio inteiro têm estados quânticos coletivos descritos pela estatística de Fermi-Dirac (férmions), enquanto as partículas com spin inteiros têm estados coletivos descritos pelas estatísticas de Bose-Einstein (bósons). Tal conexão pode ser provada em mecânica quântica relativística, mas é um axioma na teoria não relativística. Bósons e férmions também são diferenciados pelo Postulado da Simetrização [14], que afirma que todos os estados quânticos devem ser simétricos (bósons) ou antisimétricos (férmions), com relação à sua função de onda, sob troca de partículas idênticas. A antissimetria proíbe que férmions idênticos estejam no mesmo estado quântico, que resulta no princípio de exclusão de Pauli [15], enquanto bósons idênticos podem estar no mesmo lugar ao mesmo tempo.

possuem spins múltiplos de meio inteiro (como $\pm 1/2$) e são divididos em quarks e léptons. Os bósons são partículas fundamentais que incluem os bósons de calibre, mediadores das interações fundamentais da natureza, como os fótons, glúons e os bósons W e Z, bem como o recém-descoberto Bóson de Higgs e o bóson hipotético mediador da gravidade quântica, gráviton.

Os léptons⁶ e os quarks⁷ são, também, considerados famílias de partículas fundamentais (ou elementares), uma vez que não são compostos por outras partículas. Cada uma destas famílias contém seis membros, conhecidos como sabores (*flavors*), e podem ser subdivididas em gerações de acordo com suas respectivas massas. As partículas fundamentais possuem, também, suas respectivas antipartículas, que apresentam a mesma massa, mas carga elétrica e momento magnético opostos⁸. No caso de elétron, sua antipartícula é o pósitron⁹, cuja carga elétrica é positiva.

Cada geração é dividida em dois tipos de léptons e dois tipos de quarks, de acordo a carga elétrica. Todos os léptons com carga possuem propriedades similares às do elétron, que possui carga elétrica igual a uma unidade negativa de carga elementar. Os neutrinos, léptons neutros, são produzidos em decaimentos radioativos e possuem carga elétrica igual a zero e massa bem pequena. Os quarks são as partículas elementares que constituem os prótons, nêutron e todas as partículas que chamamos de hádrons. Os quarks possuem uma carga elétrica positiva e igual a $2/3$ da carga elementar ou uma carga negativa equivalente a $1/3$ da carga elementar.

Como pode ser visto na tabela 2.1, cada família tem três gerações, com valores crescentes de massa. A primeira geração é composta pelo elétron, pelo neutrino do

⁶A palavra Lépton deriva do grego e significa fino, pequeno, tendo sido utilizada pelo físico Léon Rosenfeld em 1948 pela primeira vez [16]. A etimologia implica, de forma incorreta, que todos os léptons possuem massa pequena. Quando Rosenfeld as nomeou assim, os únicos léptons conhecidos eram os elétrons e os múons, os quais possuem massa pequena, de fato. Entretanto, o tau, que foi descoberto em meados de 1970, possui um valor de massa que é cerca de 3500 vezes a massa do elétron.

⁷O conceito de quark foi desenvolvido de forma independente, em 1961, por Murray Gell-Mann [17] e Yuval Ne'eman [18]. Tal conceito evoluiu de um esquema de organização de hádrons através da simetria de sabor (SU(3)) que ficou conhecida como Caminho Óctuplo (Eightfold Way), e que estabeleceu a base do modelo de quarks [19]. Ao passo que a nomenclatura Eightfold Way foi escolhida, por Gell-Mann, como alusão a um conjunto de práticas dos ensinamentos budistas, o nome Quark foi encontrado em uma passagem do intraduzível romance *Finnegans Wake* de James Joyce, parte de um poema de 13 linhas indecentes dirigido contra o Rei Mark: "Three quarks for Muster Mark!". Gell-Mann possuía o som desejado, mas sem a ortografia e, na época, a alusão a três quarks parecia perfeita, pois haviam apenas três quarks subatômicos [20]

⁸Além da carga elétrica e momento magnético, as antipartículas possuem números quânticos invertidos, como o número barônico, o número leptônico e a estranheza. [21]

⁹O pósitron é uma das poucas antipartículas que possui seu próprio nome. A maior parte das antipartículas é descrita pela palavra "anti" precedendo o nome da partícula. O pósitron também se diferencia na simbologia, uma vez que as antipartículas normalmente recebem o mesmo símbolo da partícula, mas com um traço acima, como próton p e \bar{p} . O pósitron e^+ foi a primeira antipartícula a ser descoberta, em 1932, pelo físico Carl Anderson, concedendo-lhe o prêmio nobel de física em 1936 [21], e a nomenclatura foi uma sugestão do editor do seu artigo no jornal *Physical Review* [22].

Tabela 2.1: Férmions. Adaptada de [9]

Léptons			Quarks		
Flavor	Massa (GeV/c^2)	Carga Elétrica	Flavor	Massa (GeV/c^2)	Carga Elétrica
ν_e Neutrino do elétron	$(0-2)\times 10^{-9}$	0	u up	0.002	2/3
e Elétron	0.000511	-1	d down	0.005	-1/3
ν_μ Neutrino do múon	$(0.009-2)\times 10^{-9}$	0	c charm	1.3	2/3
μ Múon	0.106	-1	s strange	0.1	-1/3
ν_τ Neutrino do tau	$(0.05-2)\times 10^{-9}$	0	t top	173	2/3
τ Tau	1.777	-1	b bottom	4.2	-1/3

elétron ν_e e pelos quarks up (para cima) e down (para baixo), os quais constituem a matéria ordinária ao nosso redor. A segunda geração é composta pelo múon μ , o neutrino de múon ν_μ e os quarks charm (charmoso) e strange (estranho). Na última geração encontramos o lépton τ , neutrino do tau ν_τ e os quarks top (topo) e bottom (fundo). Os neutrinos são as partículas com menor massa, mas não se sabe, atualmente, o valor exato de suas massas ¹⁰. O sabor, que distingue as partículas entre as gerações dentro de cada família, é conservado para todas as interações, com exceção da interação fraca.

Tabela 2.2: Bárions e Mésons. As informações referentes às demais partículas listadas nesta tabela, bem como uma enorme gama de outras partículas podem ser encontradas em [25].

Hádrons		
	Bárions	Mésons
up, down	p, n, Δ	$\pi, \eta, \rho, \omega, \dots$
strange	$\Lambda, \Sigma, \Xi, \Omega$	K
charm	$\Lambda_c, \Sigma_c, \Xi_c, \Omega_c$	$D, D_s, \eta_c, J/\psi, \chi_c$
bottom	$\Lambda_b, \Sigma_b, \Xi_b, \Omega_b$	$B, B_s, B_c, \Upsilon, \chi_b$
top	-	-

Hádrons são todas as partículas que contém quarks ¹¹. Tais partículas são sensíveis às interações forte e às reações nucleares, enquanto os léptons não. Bárions são compostos por três quarks, enquanto os mésons contém um quark e um anti-quark e ambos juntos formam os hádrons, como pode ser visto na tabela 2.2. A soma das cargas elétricas dos quarks que compõem um hádron é sempre um número inteiro. Os prótons e nêutrons, por exemplo, são os bárions mais leves, pois são compostos

¹⁰Por décadas os físicos acreditavam que os neutrinos não possuíam massa. Em 1998, o observatório Super-Kamiokande, localizado nas profundezas do monte Ikeno, no Japão, descobriu que estas partículas possuem massas diferentes de zero [23]. Não se sabe o valor exato, mas indícios apontam para um valor algumas centenas menor que a massa do elétron [24].

¹¹Não existem hádrons incluindo um quark top por conta de seu curto tempo de vida, já que este quark decai antes que possa formar uma composição com qualquer outro quark.

pelos quarks mais leves, up e down, da primeira geração. Os prótons são formados por dois quarks up e um down, sendo escritos como uud, enquanto os nêutrons, formados por dois quarks down e um up, são escritos como ddu. Repare que ao somar as cargas elétricas dos quarks mencionados, disponíveis na tabela 2.1, obtemos carga elétrica igual a 1 ($2/3 + 2/3 - 1/3$) para o próton e carga elétrica igual a 0 ($-1/3 - 1/3 + 2/3$) para o nêutron. Os píons, que podem ser carregados (π^+ e π^-) ou neutros (π^0) são os mésons mais leves, pois são formados, também, por quarks up e down. O pión carregado positivamente é formado pela combinação $u\bar{d}$, possuindo, então, carga elétrica igual a 1 ($+2/3 + 1/3$), uma vez que as antipartículas possuem a mesma massa, mas cargas físicas (como carga elétrica, carga de cor, ...) opostas. O pión neutro, por outro lado, é constituído por $\frac{u\bar{u} + d\bar{d}}{\sqrt{2}}$, possuindo carga elétrica igual a 0.

Um sumário compreensivo contendo informações e avaliações das medidas das propriedades das partículas elementares conhecidas e um resumo das pesquisas de novas partículas hipotéticas é disponibilizado por uma colaboração internacional conhecida como Particle Data Group (PDG) [25]. Esta colaboração consiste, atualmente, de 227 autores de 159 instituições e 24 países, que atualizam anualmente e publicam a cada dois anos, em um jornal de física de altas energias, o documento conhecido como *Review of Particle Physics*.

2.3 Interações Fundamentais

Existem 4 forças fundamentais na natureza: nuclear forte, eletromagnética, nuclear fraca e gravitacional, as quais são listadas na tabela 2.3 por ordem decrescente de magnitude. Como pode ser visto, essas forças possuem propriedades bastante distintas, como sua magnitude e alcance. A força gravitacional e a força eletromagnética possuem um longo alcance, ao contrário das forças nucleares fraca e forte, que possuem efeito perceptivo apenas em distâncias menores que o tamanho de um próton. Com relação à magnitude, a força forte é superior às demais, seguida da força eletromagnética e da nuclear fraca, sendo a força gravitacional¹² demasiadamente inferior às demais.

Cada uma dessas forças é associada a uma teoria física e mediada pela troca de uma partícula. A força eletromagnética é mediada pelo fóton, a força nuclear forte pelo glúon e a fraca pelos bósons vetoriais intermediários W e Z. Com relação à força gravitacional, a teoria quântica de campos postula a existência de uma partícula denominada gráviton. Caso exista, tal partícula hipotética deve atender a

¹²Percebemos a gravidade, conforme Newton previu, devido a seu alcance teoricamente infinito e à massa de objetos astronômicos, como planetas ou estrelas que interagem por esta força, em escalas cosmológicas.

Tabela 2.3: As 4 forças fundamentais da natureza. Adaptada de [9]

Força	Magnitude	Teoria	Mediador	Alcance (m)
Nuclear forte	10	Cromodinâmica (QCD)	Glúon	10^{-15}
Eletromagnética	10^{-2}	Eletrodinâmica (QED)	Fóton	∞
Nuclear fraca	10^{-13}	Teoria Eletrofraca (EWT)	Bósons W e Z	10^{-18}
Gravitacional	10^{-42}	Relatividade Geral (GR)	gráviton (hipotética)	∞

certas propriedades que concordem com ambas as teorias de gravidade de Newton e Einstein ¹³ [27]. Esses bósons mediadores têm o papel de transmitir a força entre um quark ou lépton e outro. Ao passo que os quarks são afetados pelas 4 interações, os léptons não são afetados pela interação nuclear forte.

A teoria física que descreve a força eletromagnética é conhecida como eletrodinâmica e sua formulação clássica foi feita por Maxwell em 1862, já consistente com a relatividade especial, e aperfeiçoada por Tomonaga, Feynman e Schwinger em 1940. Essa teoria, chamada eletrodinâmica quântica (Quantum Eledrodynamics, QED), portanto, é a mais antiga e a mais simples entre as teorias dinâmicas [9], e postula que todo fenômeno eletromagnético pode ser reduzido ao seguinte processo elementar: uma partícula carregada (um elétron, quark, lépton, mas não um neutrino) *e* entra, emite ou absorve um fóton γ e sai, ou seja, a QED descreve todo fenômeno envolvendo partículas carregadas interagindo por meio da troca de fótons e proporciona um relato da interação de matéria e luz. Tal processo pode ser descrito através de diagramas de Feynman [28].

A cromodinâmica quântica (Quantum Chromodynamics, QCD) descreve as interações, através da força nuclear forte, entre quarks e glúons, partículas fundamentais que compõem hádrons, como prótons, nêutrons e píons [29] [30]. Analogamente à carga elétrica descrita na eletrodinâmica, a cromodinâmica apresenta a carga de cor, que não possui relação alguma com o conceito tradicional de cor. Sendo assim, enquanto a carga elétrica pode ser apenas positiva ou negativa, sendo caracterizada por apenas um número, a carga de cor é caracterizada pela combinação de três cores. A convenção atribui aos quarks 3 tipos de carga, representadas pelas cores vermelho, azul e verde. Suas anti-partículas, por outro lado, apresentam as "anticores" antivermelho, antiazul e antiverde. Já os glúons possuem uma carga composta pela mistura de duas cores¹⁴, sendo uma cor e uma anticor, como vermelho anti-azul,

¹³A fim de manter a coerência com ambas as teorias de Newton e Einstein, a partícula denominada gráviton não deve possuir massa, para que possa manter um alcance infinito; deve possuir um spin par e igual a 2 para que produza uma força atrativa apenas; e deve ser eletricamente neutra. [26]

¹⁴Diferente dos fótons que não possuem carga elétrica e, portanto, não interagem entre si, os glúons possuem carga de cor. Sendo assim, os glúons, além de mediar, participam das interações fortes.

por exemplo. Apesar de termos 9 possibilidades, por razões técnicas, existem apenas 8 tipos de glúons.¹⁵ Devido a um fenômeno da QCD, denominado confinamento, as partículas que apresentam carga de cor não podem existir isoladamente. Desta forma, quarks e glúons não podem ser separados dos hádrons aos quais estão associados sem produzir novos hádrons¹⁶[21]. Quando os quarks são produzidos, então, em experimentos de aceleradores de partículas, os quarks não são vistos individualmente pelos detectores, mas, ao invés, jatos hadrônicos contendo mésons e bárions agrupados são vistos, em um processo chamado de hadronização.

A interação nuclear fraca não era conhecida na física clássica e, portanto, a sua descrição teórica foi feita a partir da formulação quântica relativística, desde o início, pelo físico Enrico Fermi, a fim de explicar o decaimento ou emissão beta, em 1933. A sua teoria descreve o decaimento nuclear radioativo e é chamada de flavordinâmica quântica (Quantum Flavordynamics, QFD). Entretanto, a interação fraca é melhor compreendida nos termos da teoria eletrofraca (EWT) ou teoria de Glashow-Weinberg-Salam (GWS). Este modelo trata as interações fraca e eletromagnética como diferentes manifestações de uma única força eletrofraca e, neste sentido, as 4 forças se reduzem a três. Dentre as suas propriedades, a interação fraca é a única capaz de modificar o sabor (*flavor*) dos quarks, transformando um tipo de quark em outro¹⁷. Enquanto o bóson W media as interações fracas carregadas, o bóson Z media as neutras [9].

Apesar do modelo padrão conter a maior parte das forças fundamentais conhecidas e mencionadas anteriormente (eletromagnetismo, a força nuclear fraca e forte), este modelo não contempla a força gravitacional. A teoria mais madura que descreve tal força é a teoria da relatividade geral, de Einstein, que explica a gravitação pela curvatura do espaço-tempo [32]. Juntas, as teorias de relatividade geral (GR) e a mecânica quântica permitem a explicação de praticamente qualquer fenômeno fundamental observado. Porém, a relatividade geral é baseada no mundo clássico e opera na escala macroscópica, enquanto a mecânica quântica opera na escala subatômica. Desta forma, não é possível utilizar a GR quando se trata de efeitos gravitacionais fortes nas menores escalas de espaço e tempo.¹⁸

¹⁵A QCD é um tipo de teoria quântica de campos chamada de teoria de gauge não abeliana, com o grupo de simetria SU(3). O espaço de matrizes hermitianas 3×3 com traço igual a zero possui 8 dimensões. Glúons se transformam sob essa representação e, portanto, existem 8 glúons.[31]

¹⁶Ao separar dois quarks, uma vez que uma energia suficientemente alta tenha sido introduzida, novos pares de quark se formarão.

¹⁷Todos os quarks e léptons possuem uma "carga fraca". Léptons não possuem cor, então eles não participam na interação forte; neutrinos não possuem carga e, portanto não são afetados pela força eletromagnética, mas todas estas partículas experienciam as interações fracas.

¹⁸Como no centro de um buraco negro, onde um ponto no espaço possuiria uma massa com densidade infinita, levando o espaço-tempo a ter infinitas curvaturas [33] [34] [35], ou no instante do Big Bang [36].

Com o objetivo de unificar as forças fundamentais, considerando o sucesso do modelo padrão em descrever os fenômenos do mundo subatômico para todas as outras interações, acredita-se que a gravidade deve possuir, também, uma natureza quântica. A gravitação quântica é o campo da física teórica que busca desenvolver modelos que permitam descrever a gravidade de acordo com os princípios da mecânica quântica [27]. Entretanto, devido ao fato de a força da gravidade ser extremamente fraca quando considerada na escala subatômica, torna-se praticamente impossível detectar qualquer efeito devido à força gravitacional em algum experimento de física de partículas ¹⁹ Existem, porém, algumas teorias quânticas de gravidade propostas dentre as quais se destacam a teoria das cordas [40], que prediz a existência de grávitons e suas interações bem definidas, e a gravidade quântica em loop [26], que não prevê a existência de tal partícula hipotética.

2.4 Modelo Padrão

Toda essa complexa teoria, que descreve as interações eletromagnéticas, fraca e forte, bem como classifica as partículas elementares, como os bósons, quarks e léptons, como excitações de campo, é conhecida como Modelo Padrão (Standard Model, SM). Desenvolvido na segunda metade do século XX e abrangendo mais de 50 anos de imenso esforço experimental e inspiração teórica de físicos notáveis, o sucesso de sua formulação vem sendo demonstrado através de confirmações experimentais até hoje.

O Modelo Padrão pode ser representado pelo diagrama 2.1, onde estão descritas e nomeadas as partículas fundamentais, ao longo das 3 gerações, bem como suas respectivas massa, carga e spin. Sua formulação matemática é construída através de um arcabouço teórico que combina teoria clássica de campo, relatividade especial e mecânica quântica, sendo conhecida como teoria quântica de campos (Quantum Field Theory, QFT) [42].

¹⁹A principal visão adotada pelos físicos que estudam o problema é que quando chegássemos à escala de Planck, a quantização do espaço-tempo se tornaria inevitável [37] [38], visto que as predições do Modelo Padrão, da teoria quântica de campos e da relatividade geral não são mais conciliáveis, e espera-se que os efeitos quânticos da gravidade dominem. Quanto mais alta a escala de energia do experimento menor a distância que pode ser investigada. Sendo assim, investigar interações na escala de Planck diretamente requer uma máquina que se seja cerca de 15 ordens de grandeza maior que o LHC [39].

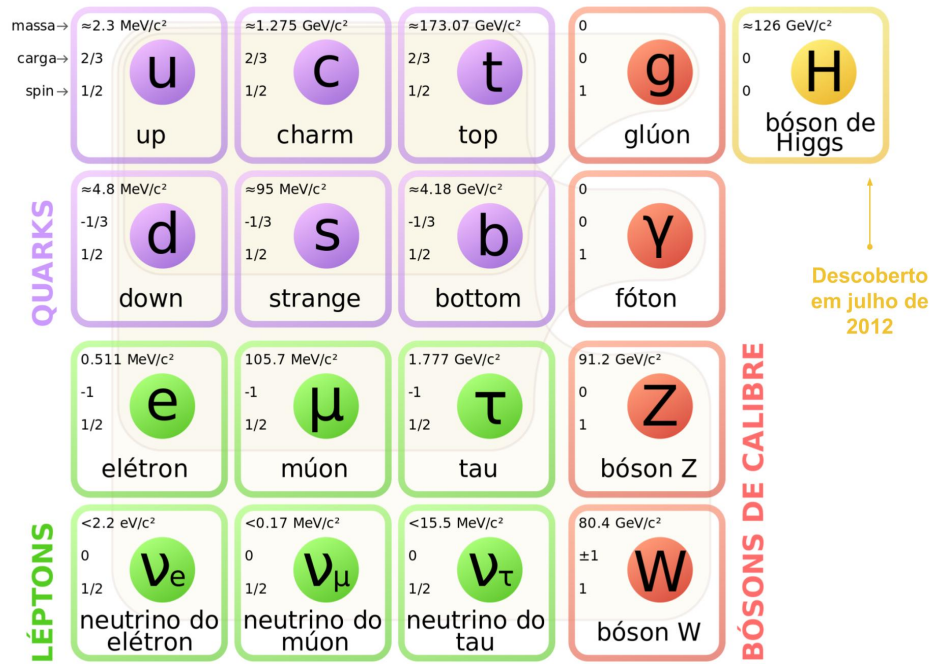


Figura 2.1: Modelo Padrão de Partículas Elementares. Adaptado de [41].

2.5 O Bóson de Higgs

O Modelo Padrão não explica as origens da massa e o motivo de algumas partículas serem massivas e outras não possuírem massa. Em 1964, Peter Higgs em conjunto com outros físicos propuseram um mecanismo que explicaria a origem da massa das partículas. O mecanismo de Higgs seria mediado por um bóson, produzido pela excitação quântica do campo de Higgs. Partículas que interagem intensamente com o campo de Higgs são pesadas, enquanto aquelas com interações fracas são leves. Dentre os objetivos principais da construção do LHC está a possibilidade de observação do Bóson de Higgs, o estudo de suas propriedades e determinar seus decaimentos raros [43]. Em julho de 2012, ambos os experimentos ATLAS e CMS, no LHC, usaram dados coletados para anunciar a descoberta de um bóson com propriedades consistentes com as previstas pelo Modelo Padrão [44], o que rendeu aos físicos Peter Higgs e Francois Englert o prêmio Nobel em física de 2013 [45].

2.6 Física Além do Modelo Padrão

Como mencionado anteriormente, apesar de seu sucesso na previsão dos fenômenos físicos observados, o Modelo Padrão ainda deixa lacunas. Além de não incluir a gravidade e de não permitir entendermos o porquê de tal força possuir

uma magnitude tão inferior às outras, o modelo também não explica, entre outros fenômenos, a oscilação do neutrino, a origem da diferença entre matéria e antimatéria e a natureza da matéria escura e energia escura [46]. Vale notar, com relação a estes últimos dois fenômenos, que a matéria que conhecemos hoje, que compõe as galáxias e estrelas, representa apenas 5% da massa e energia do universo. Estima-se que cerca de 27% seja matéria escura e 68% energia escura [47]. Os desenvolvimentos teóricos que tentam explicar tais deficiências do Modelo Padrão são conhecidos como Física além do Modelo Padrão (Beyond the Standard Model, BSM) e alguns exemplos deste estudo serão brevemente descritos abaixo.

- Anti-matéria

Como mencionado anteriormente, as partículas fundamentais possuem, também, suas respectivas antipartículas. Tal concepção é uma interpretação da equação desenvolvida por Paul Dirac em 1928 ²⁰, que possuiria duas soluções, sendo uma para um elétron com carga positiva e outro com carga negativa. Esta interpretação permitiria galáxias e até universos inteiros feitos de anti-matéria. Quando matéria e antimatéria entram em contatos elas se aniquilam e desaparecem e, de acordo com o que se conhece, o Big Bang deveria ter criado quantidades iguais de ambos. Os físicos buscam, então, estudar o motivo da existência inesperada de uma quantidade bem maior de matéria no universo, por exemplo, através da geração de antiprótons de baixa energia e do estudo do Quark B pelo experimento LHCb [49].

- Supersimetria

A Supersimetria não é uma teoria, mas um princípio ²¹ que propõe que as equações para força e para matéria são idênticas, baseando-se na ideia de que pode-se trocar férmions e bósons em todas as partes das equações de QFT. Sendo assim, para cada bóson existiria um férmion correspondente, e vice-versa [51]. As novas partículas postuladas se diferenciariam de suas partículas correspondentes por meia unidade de spin e por um novo número quântico chamado paridade-R [52]. Equações que contêm supersimetria poderiam explicar porque a gravidade é tão mais fraca que as outras forças, por que o bóson de Higgs existe, além de providenciar uma possível fonte de energia escura. Entretanto, até o presente momento os experimentos de HEP, incluindo o LHC, ainda não permitiram que tais partículas fossem observadas [51].

²⁰A equação de Dirac, que descreve o comportamento dos férmions e prevê a existência de antimatéria proporcionou-lhe, em 1933, o prêmio Nobel em conjunto com Erwin Schrödinger. [48]

²¹Teorias que se baseiam neste princípio são consideradas teorias supersimétricas. Em 1981, uma proposta [50] acrescentou os princípios simétricos ao Modelo Padrão com o número mínimo absoluto de mudanças necessárias, ficando conhecida como Minimal Supersymmetric Standard Model (MSSM) [51]

- Matéria Escura

A existência de matéria escura é creditada devido ao fato de que as galáxias não se formariam, não se manteriam juntas e não possuiriam um movimento rotacional tão rápido se não fosse por uma quantidade razoável de matéria ainda não observada, visto que o efeito gravitacional provocado pela matéria visível não é forte o suficiente para explicar tais comportamentos. Entretanto, diferentemente da matéria ordinária, a matéria escura não interage com a força eletromagnética e, portanto, não emite, reflete ou absorve radiação eletromagnética, como luz visível, ondas de rádio, infra-vermelho e ultra-violeta. Muitos físicos teóricos acreditam que as partículas de matéria escura ²² poderiam ser produzidas pelo LHC, mas escapariam pelo detector sem serem notadas. Sendo assim, sua existência poderia ser inferida pela quantidade de momento e energia faltantes depois da colisão [47].

- Energia Escura

Desde o momento do Big Bang, o universo está expandindo continuamente. Acreditava-se, a partir dos conhecimentos sobre a gravidade, que esta expansão estaria em desaceleração, uma vez que a gravidade é uma força atrativa. No final do século XX, porém, dois times de astrônomos mediram a luz proveniente de explosões chamadas de supernovas do tipo Ia e perceberam que estas supernovas estavam mais distantes do que se esperaria de um universo em desaceleração [53] [54]. O resultado sugeriu que a expansão do universo, na verdade, estaria em aceleração²³ e que existiria uma energia escura impulsionando tal aceleração e contrabalanceando o efeito de atração gravitacional. [51]. Esta energia escura presente em 68% do universo estaria associada ao vácuo no espaço e distribuída uniformemente ao redor do universo em espaço e tempo [47]. Os principais experimentos envolvidos na busca pela natureza da energia escura envolvem equipes de astrônomos, físicos e engenheiros que utilizam telescópios e satélites a fim de investigar a história da expansão do universo. Alguns exemplos podem ser encontrados em [56] [57] [58] [59].

- Dimensões Extras

Uma das possibilidades de explicação do fato de a força gravitacional ser tão inferior em magnitude em relação às demais é que não percebemos todo o efeito gravitacional, pois a gravidade poderia estar se espalhando para dimensões

²²As partículas hipotéticas que constituiriam a matéria escura, e interagiriam com a gravidade e com forças hipotéticas supostamente mais fracas do que as que conhecemos, são denominadas WIMP (Weakly Interacting Massive Particle), mas não há uma definição clara sobre suas propriedades.

²³Esta descoberta proporcionou o Prêmio Nobel da Física de 2011 para Saul Perlmutter, Brian Schmidt e Adam Riess, astrônomos de ambos os times [55].

extras, além das três dimensões espaciais e a quarta temporal que conhecemos. Caso existam, estas dimensões adicionais poderiam explicar, também, o porquê da expansão do universo ser mais rápida do que o esperado. Teorias que sugerem dimensões extras predizem que haveria versões mais pesadas de partículas conhecidas em outras dimensões. Partículas semelhantes aos bósons Z e W com uma massa cerca de 100 vezes maior, por exemplo, poderiam sugerir a existência de dimensões extras e só poderiam ser reveladas em altas energias pelo LHC e detectadas em experimentos como ATLAS e CMS [60].

Para que seja possível, então, testar o Modelo Padrão com acuidade e explorar suas deficiências na busca pela nova física, os experimentos buscam dados para dar mais evidências às teorias já propostas e, se algo novo for levantado, os teóricos se esforçam para dar sentido. A fim de gerar estes dados, as condições que viabilizam a manifestação de fenômenos de interesse são criadas em altas energias, através do uso de aceleradores de partículas e, com a ajuda de detectores, as propriedades destes fenômenos são reconstruídas. O próximo capítulo descreve aspectos teóricos de aceleradores e detectores de partículas.

Capítulo 3

Aceleradores e Detectores

Na busca por uma melhor compreensão dos constituintes básicos da matéria que compõem o universo, a humanidade tem feito uso de um aparato experimental complexo que permite recriar as condições existentes no universo instantes após o Big Bang. O ramo da física que contempla esses experimentos já passou por diversas nomenclaturas ao longo do tempo. No início chamava-se física de raios cósmicos, mas com o advento dos aceleradores de partículas, a disciplina passou a ser conhecida como física de partículas elementares ou física de altas energias. Como a maior parte dos aceleradores utilizados nos experimentos operam na escala energética de GeV ou TeV¹, necessária para investigação do que ocorre nas menores escalas, o termo Física de Altas Energias parece ser o mais apropriado, embora existam questões importantes sobre a estrutura fundamental da matéria que são melhor abordadas em experimentos em baixas energias² [62].

Dentre os principais instrumentos que compõem essa infraestrutura experimental complexa encontram-se os aceleradores de partículas e os detectores. Ao passo que os aceleradores são capazes de produzir uma enorme gama de partículas através da aceleração de prótons e elétrons, entre outras, os detectores permitem medir a energia das partículas produzidas, bem como outros atributos, como momentum, spin, carga. Para efetuar estas medições, os experimentos são, normalmente, compostos por um sistema de traço (*tracking*), que permite reconstruir a trajetória das partículas eletricamente carregadas, e um sistema de calorimetria, que mede a energia das partículas incidentes. Apesar de estas medições serem feitas nos subprodutos finais das colisões, um processo de reconstrução permite que os físicos encontrem as propriedades dos fenômenos originais de interesse. O objetivo deste capítulo é fazer uma breve descrição sobre os princípios destes instrumentos.

¹Experimentos com acelerados também trabalham em problemas baseados, experimentalmente, em baixas energias, como os realizados no desacelerador Anti-próton (CERN) e nos experimentos de feixes de neutrons ultrafrios [61]

²Conservação do número bariônico, simetria matéria-antimatéria e a massa dos neutrinos são alguns exemplos [62].

3.1 Princípios de aceleração de partículas

Os aceleradores são fundamentais para a pesquisa na área e permitem-nos observar fenômenos de interesse nas menores escalas já atingidas. Existem outras maneiras de investigar essas escalas, como os raios cósmicos, mas elas possuem desvantagens. Os raios cósmicos são partículas de energia elevada, tipicamente entre 10^8 e 10^{19} elétron-Volts, vindas do espaço, que ao atingir átomos na atmosfera superior produzem chuviros de partículas secundárias menos energéticas [63]. Apesar de serem constantes e não necessitarem de custo algum para sua produção, a taxa com que conseguem ser detectados não é tão alta, além de não serem controláveis³. Com aceleradores de partículas, por outro lado, conhecemos o que está sendo colidido e podemos posicionar detectores ao redor da região de colisão.

Para que seja possível estudar o que ocorre nessas pequenas escalas, é necessário que os aceleradores sejam capazes de produzir colisões em altíssima energia no centro de massa. Tal relação entre a escala dos fenômenos observados e a energia de colisão decorre das propriedades ondulatórias das partículas, que possuem comprimento de onda inversamente proporcional ao seu momento, de acordo com a equação de Broglie ⁴

$$\lambda = \frac{h}{p} \quad (3.1)$$

onde λ é o comprimento de onda e h a constante de Planck [67]. Sendo assim, de forma similar ao que ocorre com os microscópios (como os de elétrons), quanto maior a energia da partícula, menor o comprimento de onda, e quando menor o comprimento de onda, melhor a resolução que se consegue obter para observação de pequenas estruturas⁵. Colisões em altas energias permitem, também, a produção de partículas pesadas, as quais representam objetos físicos de interesse. Isso ocorre devido à equivalência entre massa e energia, como foi mencionado na Seção 2.1.

Existe uma variedade de tipos de aceleradores de partículas em uso atualmente, dentre os quais se destacam os síncrotrons, cíclotrons e aceleradores lineares. Os aceleradores lineares, também conhecidos como linacs, como o nome sugere, aceleram partículas eletricamente carregadas seguindo uma trajetória retilínea dentro

³Acredita-se que os raios cósmicos são formados majoritariamente por prótons e outros núcleos atômicos [64], mas não se sabe exatamente qual a partícula inicial. Além disso, as colisões não acontecem no vácuo, o que torna o chuviro de raios cósmicos complexo e aumenta as incertezas nas medições [65].

⁴De Broglie recebeu o Nobel da física em 1929 [66], após a confirmação experimental de sua hipótese, publicada em 1924, de que toda matéria tem propriedades ondulatórias, sendo um marco histórico no desenvolvimento da mecânica quântica.

⁵Utilizando a equação mencionada, uma energia de 1TeV, por exemplo, já nos fornece uma resolução de 10^{-18} m. Com colisões altamente energéticas, o LHC permitiu que quarks fossem observados como partículas pontuais na escala de 5×10^{-20} m [68].

de uma câmara de vácuo em forma de tubo, submetendo as partículas a uma série de potenciais elétricos [69]. Os cíclotrons e síncrotrons, por outro lado, possuem formato circular e aceleram as partículas eletricamente, enquanto promovem a trajetória curvada das partículas pela ação dos campos magnéticos⁶[70]. Enquanto os cíclotrons e linacs são utilizados normalmente para pesquisa na área médica, os síncrotrons são os mais apropriados para produzir colisões altamente energéticas e, portanto, adequados para o estudo dos constituintes básicos da matéria [71].

Outro parâmetro fundamental que deve ser considerado no projeto de aceleradores é a partícula que será colidida. Normalmente, elétrons ou prótons são usados nas colisões. Ao passo que os elétrons são partículas elementares que não possuem estrutura interna, os prótons contêm 3 quarks, além dos glúons, que, porventura, podem se converter em pares de quarks e antiquarks temporariamente. Isso faz com que as colisões entre prótons, na verdade, envolvam um par de partículas e, consequentemente, uma quantidade de energia selecionada aleatoriamente dentre as diversas possibilidades. As colisões envolvendo elétrons, por outro lado, possuem uma energia de colisão conhecida, uma vez que sabemos exatamente a energia do elétron. Sendo assim, aceleradores de prótons normalmente são tidos como "máquinas de descobertas", devido às possibilidades de eventos que podem ser gerados, enquanto os aceleradores de elétrons são usados para medições precisas⁷[51]. Vale notar que partículas eletricamente carregadas, quando movidas em círculos, emitem uma radiação que as faz perder energia. Partículas massivas, como prótons perdem energia de forma mais lenta que partículas leves como elétrons e, portanto, é muito mais difícil para um acelerador de elétrons circular alcançar altas energias que um acelerador de prótons.

⁶A principal diferença entre eles é que um cíclotron acelera as partículas em uma espiral, pois mantém a intensidade do campo magnético e a frequência do campo eletromagnético constantes, enquanto o síncrotron varia a intensidade do campo magnético e a frequência do campo eletromagnético no tempo (a frequência diminui e a intensidade aumenta), enquanto a partícula é acelerada, podendo, então, manter as partículas em uma órbita circular [70].

⁷O Brasil inaugurou em 2018 o projeto Sirius, que abriga 3 aceleradores de elétrons e consiste na maior e mais complexa estrutura de pesquisa do país, além de ser uma das primeiras fontes de luz síncrotron de quarta geração do mundo. Localizado no Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), em Campinas, o laboratório com raio X de alto brilho permitirá experimentos com resolução nanométrica, possibilitando investigar a estrutura interna de materiais com mais precisão e, assim, avanços em áreas como saúde, energia e agricultura [72]. Através da microtomografia por raios X à base de luz síncrotron, por exemplo, é possível obter imagens em resolução inédita da arquitetura neuronal do cérebro [73], formas de extrair mais óleo das rochas do pré-sal [74] [75], revelar a localização da água-fresca do bagaço da cana-de-açúcar [76], entre outras aplicações inovadoras [77] [78] [72].

3.2 Princípios de detecção de partículas

De forma similar ao modo como caçadores identificam animais através de suas pegadas e a forma como investigadores de bombas entendem a explosão estudando seus efeitos nos arredores (como as marcas de queimaduras e a profundidade com que os estilhaços penetram em materiais bem compreendidos), as partículas subatômicas são identificadas através do traço que deixam nos detectores, da energia depositada e profundidade com que a deposição de energia penetra em materiais conhecidos. Podemos definir um detector de partículas como um instrumento de medição para detecção e identificação de partículas, como as produzidas por reações em um acelerador de partículas. Os detectores permitem a realização da medição da energia da partícula e outros atributos como momento, spin, carga e, com isso, o tipo de partícula. A maior parte dos mecanismos de detecção se baseiam no fato de que partículas carregadas, de energia elevada, quando passam por material, ionizam os átomos pelo caminho.

Os detectores modernos consistem em camadas de subdetectores, cada uma com o objetivo de identificar propriedades específicas das partículas. Os detectores de traço revelam o caminho de uma partícula eletricamente carregadas à medida que passam e interagem com substâncias adequadas, enquanto os calorímetros absorvem e medem a energia das partículas. A figura 3.1 apresenta um diagrama mostrando a relação entre tipos e classificação dos detectores de partículas mais comuns. Uma breve descrição dos detectores baseados em traço e em calorimetria será feita abaixo, com uma ênfase maior nos detectores baseados em calorimetria, uma vez que estes são a base deste trabalho. Uma descrição mais aprofundada pode ser encontrada em [79] [80] [62].

3.2.1 Detecção baseada na trajetória das partículas

A detecção baseada na trajetória é o processo de reconstruir a trajetória de partículas eletricamente carregadas no qual as partículas que entram nesse detector (*tracking*) deixam um registro preciso de sua passagem pelo dispositivo. Estes detectores possuem um campo magnético que permite determinar o momento local da partícula a partir da curvatura de sua trajetória reconstruída. Isso ocorre pois as partículas produzidas nas colisões normalmente percorrem uma trajetória retilínea, mas na presença de um campo magnético, seus caminhos se tornam curvos. Os eletroímãs ao redor destes detectores geram campos magnéticos para explorar esse efeito. Sendo assim, é possível calcular o momento de uma partícula a partir da curvatura de seu caminho. Partículas com alto momento viajam em linhas quase retas, enquanto aquelas com momento muito baixo avançam em espirais apertadas dentro do detector [82].

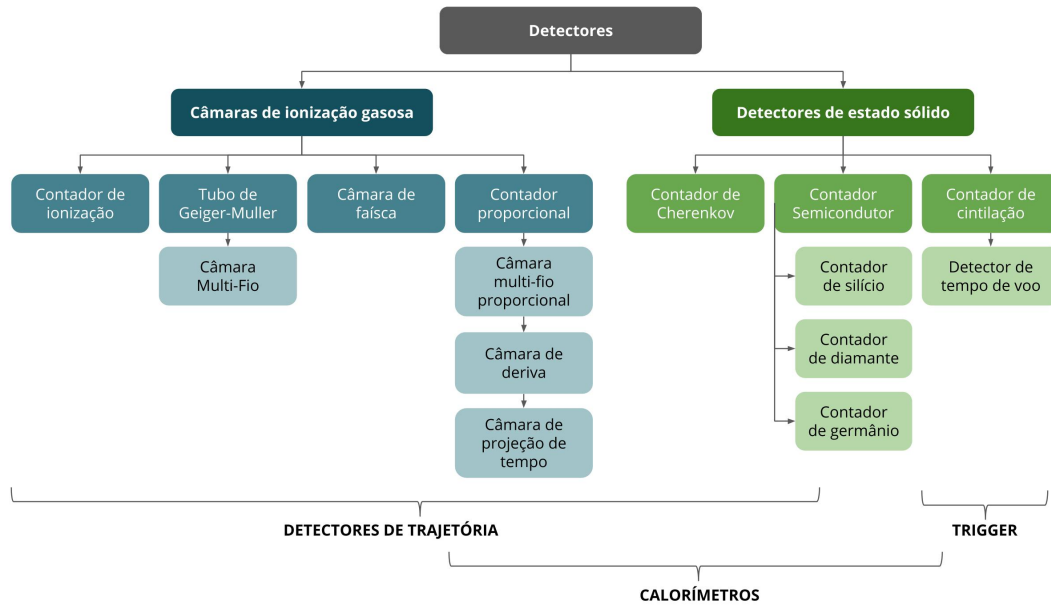


Figura 3.1: Diagrama em árvore mostrando a relação entre tipos e classificação dos detectores de partículas mais comuns. Adaptado de [81].

A reconstrução da trajetória nestes detectores é dividida em duas etapas. Primeiramente ocorre um processo de agrupamento dos *hits* no detector que originem (ou pareçam originar) da mesma trajetória. Na sequência, é realizado um ajuste de uma curva aos *hits* encontrados, com a qual o momento será obtido [83].

Entre os dispositivos mais utilizados para detecção baseada na trajetória das partículas podemos citar as câmaras de nuvens (1920-1950), placas de emulsão nuclear (1937), câmaras de bolhas (1952), câmaras de faísca (1954), câmaras proporcionais com vários fios (1968) e câmaras de deriva (1971) [84]. Com o advento dos semicondutores e da fotolitografia moderna, os detectores de estado sólido (detectores de silício) (1980)[85], passaram a ser utilizados em experimentos que exigem detecção baseada em rastreamento da trajetória de forma compacta, com alta precisão e leitura rápida, como por exemplo, próximo ao ponto de interação principal do LHC [86].

3.2.2 Detecção baseada em calorimetria

O termo calorimetria tem sua origem na termodinâmica, onde é normalmente utilizado em experimentos com o objetivo de determinar o calor específico de alguma substância. Versões modernas destes instrumentos podem ser encontradas em laboratórios de armamentos nucleares, onde são utilizados para ensaio de material fissionável, podendo medir a quantidade precisa de plutônio em uma amostra, de maneira não invasiva, por exemplo [87]. Já na física nuclear e de partículas, calori-

metria se refere ao processo de detecção de partículas e a medição de suas propriedades, através da absorção total em um bloco de matéria denominado calorímetro. Uma característica comum entre os diferentes tipos de calorímetros existentes é que o processo de medição é destrutivo, uma vez que as partículas não estarão mais disponíveis para inspeção após passarem por tal dispositivo ⁸.

Nesta Seção será feita uma breve descrição de aspectos teóricos importantes sobre os calorímetros, que incluem os mecanismos de detecção, algumas formas de interação de partículas com o calorímetro, bem como características de diferentes tipos de calorímetros.

3.2.2.1 Interação de Partículas com o Calorímetro

Uma forma comum de observar objetos é através da detecção da onda eletromagnética ou sonora refletida pelo objeto. No caso da física experimental de altas energias, devido à escala das partículas que queremos observar, a maneira de se detectar estes objetos é através da radiação emitida quando interagem com a matéria. Tais processos de interação dependem da energia e da natureza da partícula e são o resultado das forças eletromagnética, força forte e, mais raramente, força fraca que atuam entre a partícula e os constituintes do meio.

Neste processo de interação, estas partículas formam uma cascata de partículas secundárias dando origem a um chuveiro. Como a energia é conservada, toda vez que novas partículas são produzidas elas têm menos energia. Por exemplo, no caso de uma partícula de entrada e duas de saída, as duas saídas possuem em média 50% da energia da primeira partícula[51]. O processo continua até que uma enorme quantidade de partículas de baixa energia sejam produzidas e, então, absorvidas pelo meio [88]. Os dois tipos básicos de chuveiros são os chuveiros eletromagnéticos que são produzidos normalmente por um fóton ou elétron, os quais interagem exclusivamente através da força eletromagnética, e os chuveiros hadrônicos, produzidos por hádrons e que interagem principalmente através da força nuclear forte.

3.2.2.1.1 Interação eletromagnética O mecanismo de perda de energia mais conhecido que contribui para o processo de absorção é a interação eletromagnética experienciada pelas partículas carregadas que atravessam a matéria. Neste caso, a partícula ioniza o meio caso sua energia seja suficiente para liberar os elétrons do campo de Coulomb gerados pelo núcleo atômico. Este processo forma o princípio no qual muitos detectores de partículas são baseados, uma vez que os elétrons liberados podem ser coletados por meio de um campo elétrico e gerar um sinal elétrico [62].

⁸Os múons são exceção, já que conseguem penetrar a quantidade relativamente grande do material que compõe os calorímetros. Por conta disso, os experimentos, normalmente, contam com uma camada de detecção extra para identificação de múons

As interações eletromagnéticas podem ocorrer de diversas formas, como:

- através da ionização e espalhamento múltiplo de coulomb, entre partículas carregadas e elétrons dos átomos do meio.
- através do espalhamento elástico e inelástico e de *bremstrahlung*, entre partículas carregadas e núcleo atômico.
- através do efeito fotoelétrico e do espalhamento de Compton, entre fótons e elétrons dos átomos do meio.
- através da produção de pares, entre fótons e núcleo atômico.

Com base nessas diferentes formas de interação, diversos mecanismos podem ser utilizados para gerar sinais no calorímetros, sendo os mais importantes descritos abaixo. Uma descrição mais detalhada dos mecanismos de detecção e da interação de partículas carregadas, como elétrons e pósitrons, e partículas sem carga, como fótons, com o meio sobre o qual incidem pode ser encontrada em [62].

- Ionização

Partículas carregadas, ao atravessar um bloco de matéria, podem ionizar os átomos que compõem tal bloco. Desta forma, um ou mais elétrons são liberados deixando para trás um átomo ionizado (íon positivo). Muitos detectores de partículas utilizam, como forma de produção de sinais, um conjunto destes elétrons liberados. Nas câmaras de ionização baseadas em meios líquidos uma diferença de potencial é aplicada sobre o espaço que contém o líquido separando, assim, os elétrons dos íons, sendo os elétrons coletados no ânodo e os íons no cátodo. O sucesso deste método depende da existência de um caminho livre dos elétrons no líquido. Portanto, normalmente líquidos nobres como argônio, criptônio e xenônio são os meios de escolha destes detectores, uma vez que as camadas eletrônicas de seus átomos estão preenchidas. Calorímetros baseados em gases nobres como meio ativo têm sido utilizados em física experimental de partículas desde os anos 1970, sendo o argônio líquido (LAr) o pioneiro. Entre os maiores calorímetros que utilizam argônio líquido encontra-se o calorímetro presente no experimento ATLAS. Previamente o experimento D0 no Tevatron do Fermilab e o experimento H1 do HERA também utilizavam grandes calorímetros compostos de argônio líquido.

- Cintilação

Quando as partículas atravessam a matéria, elas perdem energia devido à interação eletromagnética com os campos de Coulomb dos elétrons. Esta

energia pode ser usada tanto para ionizar átomos ou moléculas pelas quais o meio atravessado pelas partículas é composto, como para trazer estes átomos para um estado excitado. A cintilação é o fenômeno associado com o processo de trazer os átomos para este estado excitado. Como o estado de excitação é instável, o átomo excitado rapidamente retorna ao estado fundamental, liberando a energia de excitação na forma de um ou mais fótons que podem ser transformados em sinal elétrico. A cintilação foi o primeiro processo físico utilizado para geração de sinais de calorimetria [62]. O meio utilizado por detectores baseados em cintilação é um material onde as trajetórias de ionização produzem luz fluorescente. Alguns exemplos são BGO, CsI e cristais PbWO₄.

- Radiação de Cherenkov

O fenômeno da radiação de Cherenkov consiste na emissão de luz quando uma partícula carregada atravessa um meio a uma velocidade maior do que a velocidade da luz neste meio. A luz emitida possui coloração azulada e a radiação é emitida em um ângulo característico, conhecido como ângulo de Cherenkov, em relação à direção da partícula.

Por ser sensível à velocidade das partículas, o mecanismo de Cherenkov pode ser usado para determinar a massa de partículas as quais o momento foi determinado por meio de deflexão em um campo magnético. Uma variedade de dispositivos foi desenvolvida para separar partículas explorando esse efeito. Vale ressaltar que um importante aspecto da luz de Cherenkov é seu caráter instantâneo. Sendo assim, não existe nenhum fator de atraso que afete as características temporais de detectores baseados neste fenômeno. Portanto, os detectores de Cherenkov, incluindo calorímetros, são instrumentos de escolhas para experimentos em que a velocidade do sinal é um requisito de extrema importância (detecção de jatos muito próximo ao tubo de feixe nos experimentos do LHC, por exemplo). Os calorímetros baseados em radiação de Cherenkov possuem como meio um material transparente no qual chuva produz fótons de Cherenkov e o sinal é, portanto, coletado na forma de luz, como os calorímetros de vidro com chumbo, por exemplo.

- Espalhamento de Compton

Neste processo, um fóton é espalhado por um elétron atômico, provocando uma transferência de momento e de energia para o elétron atingido suficiente para colocar este elétron em um estado livre.

- *Bremsstrahlung*

Este fenômeno se refere à radiação eletromagnética produzida pela desaceleração de uma partícula carregada quando desviada por outra partícula

carregada, que normalmente se trata de um elétron desviado por um núcleo atômico. Desta forma, a partícula em movimento perde energia cinética e emite um fóton, satisfazendo assim a lei da conservação de energia.

Em energias acima de 100MeV, e em diversos materiais, mesmo em energias abaixo desta, a principal forma de perda de energia por elétrons e pósitrons é por *bremstrahlung*. Sendo assim, um chuveiro eletromagnético normalmente é causado por uma sequência de interações as quais se iniciam com um elétron na escala GeV irradiando milhares de fótons por *bremstrahlung*. Como o efeito fotoelétrico, no qual o fóton é absorvido e, então, desaparece, tem participação apenas em baixas energias, muitos fótons na escala MeV de energia são absorvidos em uma sequência de processos de espalhamento de Compton, no qual a energia do fóton é reduzida em diversos passos até o ponto onde a absorção final por efeito fotoelétrico acontece. Alguns fótons com energia acima de 10MeV criam pares elétrons-pósitrons, que por sua vez também perdem energia liberando mais fótons, os quais produzem novos pares elétrons-pósitrons. O resultado é um chuveiro que consiste de milhares de partículas diferentes, sendo elas elétrons, pósitrons e fótons [62]. Conforme o chuveiro se desenvolve, a energia média das partículas do chuveiro diminui até o ponto onde não há mais multiplicação de partículas. A profundidade onde isso ocorre é referenciada como ponto máximo do chuveiro.

3.2.2.1.2 Interação Hadrônica No caso das interações hadrônicas, a interação forte entre as partículas do chuveiro e o núcleo atômico do meio também tem um papel importante. Devido à natureza da interação forte, os chuveiros hadrônicos são muito mais complexos que os eletromagnéticos, uma vez que a variedade de processos que ocorrem, tanto a nível da partícula incidente, como os que envolvem o núcleo atingido é muito maior. Uma consequência da interação forte é o fenômeno da energia invisível. Enquanto nos chuveiros eletromagnéticos toda a energia carregada é eventualmente absorvida pelo meio, nos chuveiros hadrônicos, uma fração da energia é indetectável [62].

Como visto na 3.2.2.1.1, elétrons perdem energia numa cadeia contínua de eventos, nos quais átomos do meio atravessado são ionizados e emitem fótons por *bremstrahlung*. Fótons, por outro lado, conseguem penetrar uma quantidade considerável de matéria sem perder energia até interagirem de maneira que suas identidades sejam alteradas, como no caso da produção de pares elétrons-pósitrons. Quando um hádron de alta energia penetra um bloco de matéria, uma combinação destes fenômenos pode ocorrer. Sendo assim, a energia carregada por um hádron é tipicamente depositada na forma de uma componente eletromagnética e uma componente não eletromagnética. A componente eletromagnética é resultado dos π s e η s produzidos nas reações nucleares. Já a componente não eletromagnética consiste de hádrons

carregados e fragmentos nucleares.

Quando o hádron possui carga, ele ioniza os átomos do meio atravessado em um fluxo contínuo de eventos e, em geral, em alguma profundidade, o hádron encontra um núcleo atômico com o qual interage através da força forte. Nesta reação, o hádron pode mudar sua identidade, podendo, por exemplo, se tornar 15 novos hádrons [62]. Hádron neutros não ionizam o meio que atravessam e, portanto, tem como única opção a reação nuclear. Os neutrons, produzidos de forma abundantemente no desenvolvimento do chuveiro hadrônico, possuem essa característica. As partículas produzidas nas primeiras reações nucleares podem perder energia ao ionizar o meio ou induzir novas reações nucleares, causando o desenvolvimento do chuveiro hadrônico.

Conceitualmente, os chuveiros hadrônicos são similares aos eletromagnéticos no que diz respeito ao aumento do número de partículas com a multiplicação dos processo de interação, ao passo que se diminui a energia das partículas, o que ocorre até o ponto máximo do chuveiro. Entretanto, ambos os chuveiros possuem diferenças com relação à escala. O chuveiro eletromagnético é composto por partículas carregadas que ionizam, como os elétrons, e fótons, que não produzem ionização no meio, mas se transformam em pares elétrons-pósitrons. Estes chuveiros se desenvolvem mais rapidamente e penetram pouco. Os hádrons, por outro lado, penetram mais profundamente. Sendo assim, em todos os experimentos que utilizam calorímetros para identificação de elétrons, as diferenças entre o perfil longitudinal do chuveiros eletromagnético e hadrônicos é explorada, sendo esta uma das principais razões para a divisão do sistema de calorimetria em seção eletromagnética e seção hadrônica.

A figura 3.2 mostra uma visão dos diferentes componentes do sistema de detecção do experimento ATLAS. Com base no que foi discutido até então é possível notar os diferentes perfis deixados pelas partículas procuradas e como é possível efetuar uma distinção entre elas através de seus rastros nos diferentes detectores. Os elétrons deixam rastro no detector de traço e depositam energia no calorímetro eletromagnético, através de um chuveiro fino e curto. Os fótons possuem um chuveiro fino e curto, como os dos elétrons devido a seus decaimentos, mas não deixam rastro no detector de traço. As partículas hadrônicas como prótons e neutrons, por outro lado apresentam um chuveiro mais longo e largo, depositando energia no calorímetro hadrônico, porém, enquanto o próton (hádron com carga) deixa rastro no detector de traço, o neutron (hádron neutro) não deixa. Os múons não interagem com nenhum dos calorímetro⁹ e, portanto, precisam de uma camada extra para que possam ser identificados, que no caso do ATLAS é o espectrômetro de múons. Por último, os neutrinos passam despercebidos por tal sistema de detecção e só podem

⁹Rigorosamente, os múons interagem, mas depositando uma quantidade muito pequena de energia nos calorímetros.

ser inferidos através da energia faltando.

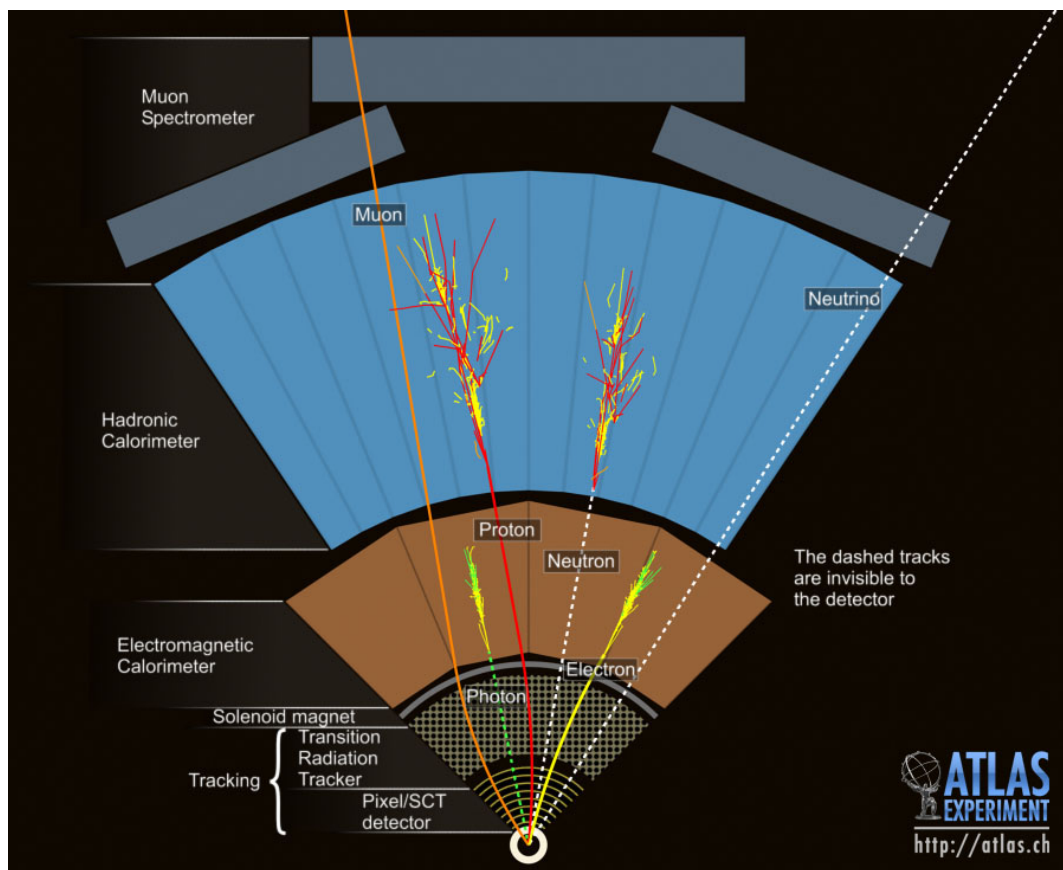


Figura 3.2: Visão longitudinal do desenvolvimento dos chuveiros no detector ATLAS com os componentes importantes mostrados e como eles respondem a vários tipos de partículas. Extraída de [89]

3.2.2.2 Tipos de calorímetros

Calorímetros podem se dividir em dois tipos em relação à sua composição material, são eles os calorímetros homogêneos e os calorímetros de amostragem. Nos calorímetros homogêneos, todo o volume do detector é sensível às partículas e pode contribuir com a geração de sinal. Nos calorímetros de amostragem, por outro lado, as funções de absorção da partícula e geração de sinal são feitas por diferentes materiais, denominados material passivo e ativo, respectivamente. Normalmente, no caso de calorímetros de amostragem, o material passivo é composto por materiais de alta densidade, como o ferro, cobre, tungstênio, chumbo e urânio, enquanto o material ativo é composto por material cintilante sólido ou líquido ou alguma forma de contador proporcional gasoso, de forma que possa gerar luz ou carga que forme a base para gerar sinais neste calorímetro. No caso de calorímetros homogêneos, cristais como NaI, BGO e CsI ou materiais compostos como vidro de chumbo são usados [90].

Por utilizarem apenas um único material, os calorímetros homogêneos possuem uma melhor resolução de energia que os calorímetros de amostragem, uma vez que parte da energia da interação da partícula com o material passivo do calorímetro de amostragem é perdida [62]. Por outro lado, calorímetros de amostragem apresentam um custo menor. Além disso, a resolução de energia é influenciada pela energia das partículas de forma que quanto maior a energia das partículas melhor a resolução. Isso torna os calorímetros de amostragem uma opção interessante e viável para grandes sistemas que operam em altas energias, como o ATLAS.

O próximo capítulo faz uma breve descrição do complexo de aceleração do grande colisor de hádrons (LHC) e do sistema de detecção do experimento ATLAS, contexto na qual este trabalho se insere.

Capítulo 4

O LHC e o experimento ATLAS

Devido à complexidade dos experimentos e das análises, a física de altas energias tem sido conduzida por grandes colaborações internacionais, contando com milhares de membros. Idealizado no pós-guerra e criado em 1954, próximo de Genebra, na Suíça, o CERN (*Conseil européen pour la recherche nucléaire*) se tornou o coração dessa comunidade global de física de partículas, fomentando a pesquisa puramente científica ¹ ao mesmo tempo que desenvolve, como consequência, tecnologias para o benefício da humanidade, como a *World Wide Web* [91], terapias para tratamento de câncer [92], ferramentas de análise de dados [93], etc. No CERN, está situado o mais poderoso acelerador de partículas até o presente momento, o LHC, onde se encontram detectores, com diferentes propósitos, em seus pontos de colisão, entre eles o ATLAS, detector de uso geral.

O objetivo deste capítulo é descrever brevemente alguns aspectos importantes do acelerador LHC, bem como apresentar a composição da estrutura de detecção do experimento ATLAS, com um maior foco no sistema de calorimetria. O sistema de filtragem do ATLAS, foco principal deste trabalho e o algoritmo *NeuralRinger* também serão apresentados.

4.1 O Grande Colisor de Hádrons

O Large Hadron Collider (LHC), está situado em um túnel de 27Km de circunferência, 175 metros abaixo do solo da fronteira entre a França e a Suíça e, até o momento, é considerada o maior colisor de partículas [94]. Sua construção, no CERN, foi fruto de uma colaboração envolvendo cerca de 10.000 cientistas, centenas de universidades e laboratórios e mais de 100 países ao longo de uma década (1998 - 2008), tendo começado a ser idealizado no início da década de 1980.

¹Entre suas principais conquistas encontram-se a descoberta da corrente fraca neutra (1973), dos bósons Z e W (1983), do bóson de Higgs (2012), dos pentaquarks (2015)

O túnel em que o LHC está situado foi construído, originalmente, para abrigar o LEP [95], que proporcionou medições precisas da interação eletrofraca, além de provar a existência das três gerações de partículas. O LHC, por outro lado, foi construído para produzir alta taxa de eventos, o que aumenta as chances de detecção de eventos raros, e atingir valores inéditos de energia no centro de colisão, o que permite a produção de partículas com maiores valores de massa. Tais características permitem ao LHC realizar uma investigação mais profunda das propriedades do bóson de Higgs ², sendo este um dos objetivos principais de sua construção, bem como tentar responder algumas questões ainda abertas na física e testar teorias hipotéticas além do modelo padrão, como visto na Seção 2.6.

O complexo de aceleradores do CERN, do qual o LHC faz parte, consiste em uma cadeia de máquinas que aceleram um feixe de partículas contribuindo para o aumento de sua energia antes de injetá-lo na próxima máquina [94]. A figura 4.1 mostra uma visão esquemática deste complexo.

Os prótons acelerados são obtidos a partir da ionização de átomos de hidrogênio por um campo elétrico. O primeiro elo da cadeia, o LINAC2, acelera os prótons até uma energia de 50 MeV. O feixe é então injetado no Proton Synchrotron Booster, chamado de PS Booster, o qual acelera os prótons até 1,4 GeV e os acumula. A partir de então, eles são injetados no Proton Synchrotron (PS), que carrega o feixe até 25 GeV. Os prótons são então enviados para o Super Proton Synchrotron (SPS), onde são acelerados para 450 GeV de energia. Finalmente, os prótons são enviados para os dois tubos de feixe do LHC, onde um feixe viaja no sentido horário, enquanto o outro no sentido anti-horário, cada um com 6,5 TeV de energia, colidindo a 13 TeV.

Cada um dos cinco aceleradores possui campos elétricos formados por cavidades de radiofrequência[51]. No caso específico do LHC, cada feixe de prótons é acelerado por 8 cavidades de RF, cada uma incrementando a energia dos prótons em 2 GeV. Além disso, empregam-se dipolos e quadripolos que, resfriados a 1,9 K, utilizam supercondutividade para gerar campos magnéticos de 8,3 Tesla, o que é cerca de 170 mil vezes mais intenso que o campo magnético da Terra [51]. Os feixes, que contêm aproximadamente 10^{11} prótons cada, se movem quase à velocidade da luz ao longo do vácuo de 10^{-13} atm, menos denso que o vazio interstelar [100]. Além de colisões próton-próton $p-p$, o LHC também colide, durante parte do ano, íons de chumbo, como $p-Pb$ (próton-chumbo) e $Pb-Pb$ (chumbo-chumbo) no intuito de estudar o fenômeno de plasma quark-glúon[43]. A tabela 4.1 quantifica alguns dos números que caracterizam a grandiosidade do colisor.

²Após o anúncio em 4 de Julho de 2012 da descoberta de um bóson com propriedades consistentes com as previstas pelo modelo padrão, pelos experimentos ATLAS e CMS, os experimentos do LHC começaram a estudar de forma mais detalhada as propriedades do Bóson de Higgs e seus decaimentos raros [96],[97],[98].

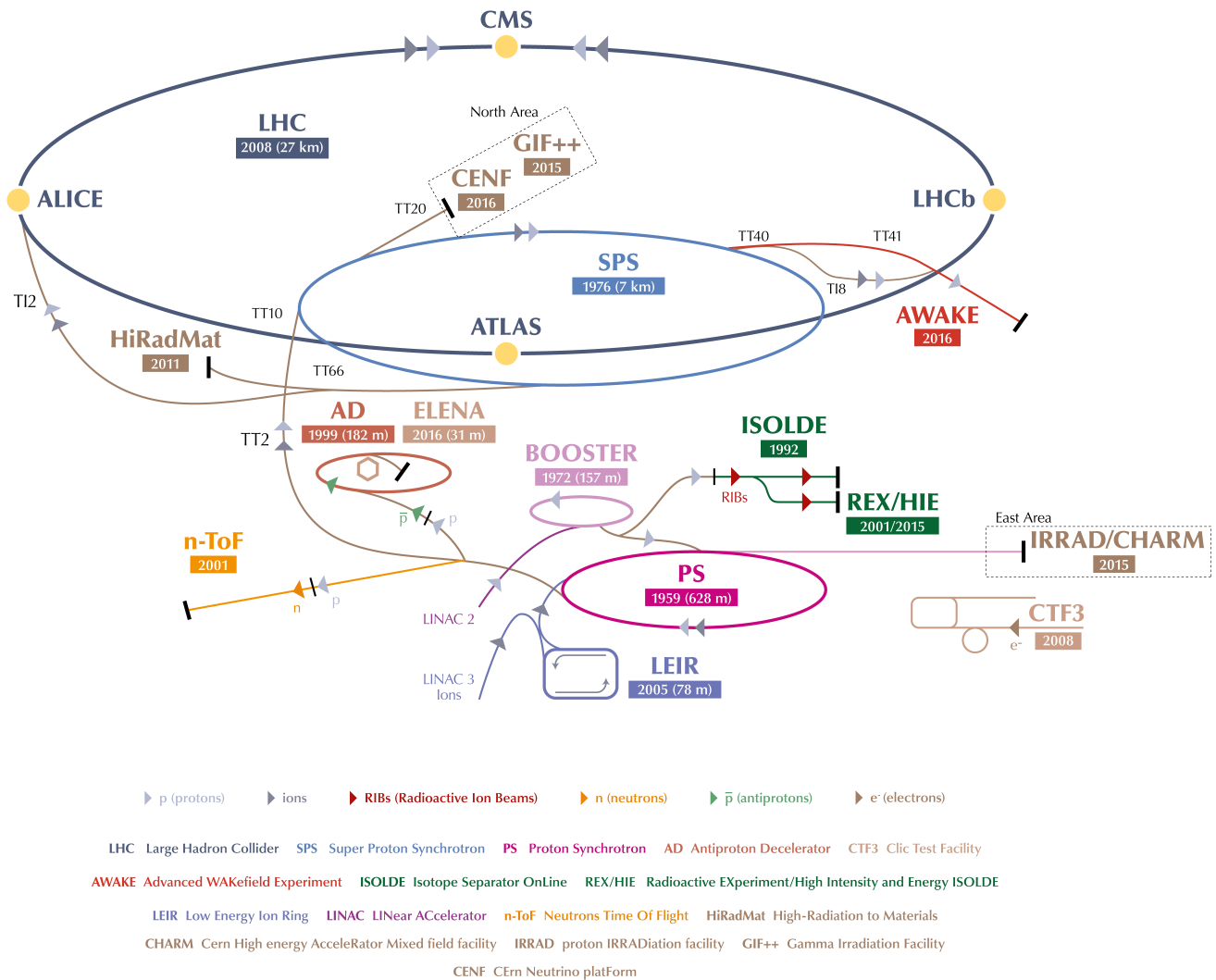


Figura 4.1: Visão esquemática do complexo de aceleração do CERN, que inclui o LHC e os pré-aceleradores, bem como os 4 experimentos principais: ATLAS, CMS, ALICE e LHCb. A figura mostra também sistemas utilizados em outros experimentos. [99].

Tabela 4.1: Energia do LHC durante a Run 2. Adaptado de [43]

Quantidade	Número
Circunferência	26 659 m
Temperatura de funcionamento do dipolo	1,9 K (-271.3°C)
Número de ímãs	9593
Número de dipolos principais	1232
Número de quadrupolos principais	392
Número de cavidades de RF	8 por feixe
Energia nominal, prótons	6.5 TeV
Energia nominal, íons	2.56 TeV/u (energia por nucleon)
Energia nominal, colisões de prótons	13 TeV
No. de pacotes por feixe de prótons	2808
No. de prótons por grupo (no início)	1.2 x 10 ¹¹
Número de voltas por segundo	11245
Número de colisões por segundo	1 bilhão

Este complexo de aceleradores possui 4 pontos de colisão, onde estão instalados 7 detectores: ALICE, ATLAS, CMS, LHCb, LHCf [101], TOTEM [102] e MoEDAL [103], sendo os 4 primeiros os principais. O experimento ALICE (A Large Ion Collider Experiment(ALICE)) [104] é voltado para a física de íons pesados, com a finalidade de estudar um novo estado de matéria, chamado plasma quark-glúon, no qual a matéria é aquecida a uma temperatura significativamente mais quente que o centro do sol. Acredita-se que estas eram as condições no universo instantes após o Big Bang.

O LHCb (LHC Beauty) [49] explora o mistério acerca da assimetria matéria-antimatéria, que foi brevemente descrita na Seção 2.6. O principal objetivo é entender o porquê de o universo ser feito majoritariamente por matéria quando nossa compreensão atual indica que o Big Bang deveria ter criado matéria e antimatéria em quantidades iguais [105]. O *beauty* do LHCb vem do nome antigo de um quark pesado e instável, que agora é chamado de quark bottom. Acredita-se que estudar essa partícula seja a maneira mais promissora de esclarecer essa assimetria entre matéria e antimatéria [106].

Os experimentos ATLAS (A Toroidal LHC ApparatuS) [107] e CMS (Compact Muon Solenoid) [108] são mais semelhante entre si em seus objetivos fundamentais, sendo experimentos com um espectro mais amplo em toda a física. Esses detectores são projetados para detectar colisões entre feixes de prótons em uma energia de colisão de 14 TeV, enquanto buscam por novas partículas ou física desconhecida que as altas energias do LHC possam nos permitir observar pela primeira vez. Este trabalho se apresenta no contexto de uma colaboração com o experimento ATLAS

e, portanto, uma breve descrição de seus componentes será feita na próxima seção.

4.2 O Experimento ATLAS

O ATLAS (A Toroidal LHC ApparatuS) [107] é fisicamente o maior experimento de partículas já construído. O experimento possui 44m de comprimento, 25m de diâmetro e pesa aproximadamente 7000 toneladas distribuídas ao longo de seu formato cilíndrico, como pode ser visto na Figura 4.2. A colaboração envolve cerca de 3000 físicos de 175 instituições e 38 países, incluindo universidades brasileiras como a UFRJ, que participa através da COPPE/UFRJ, da Escola Politécnica e do Instituto de Física.

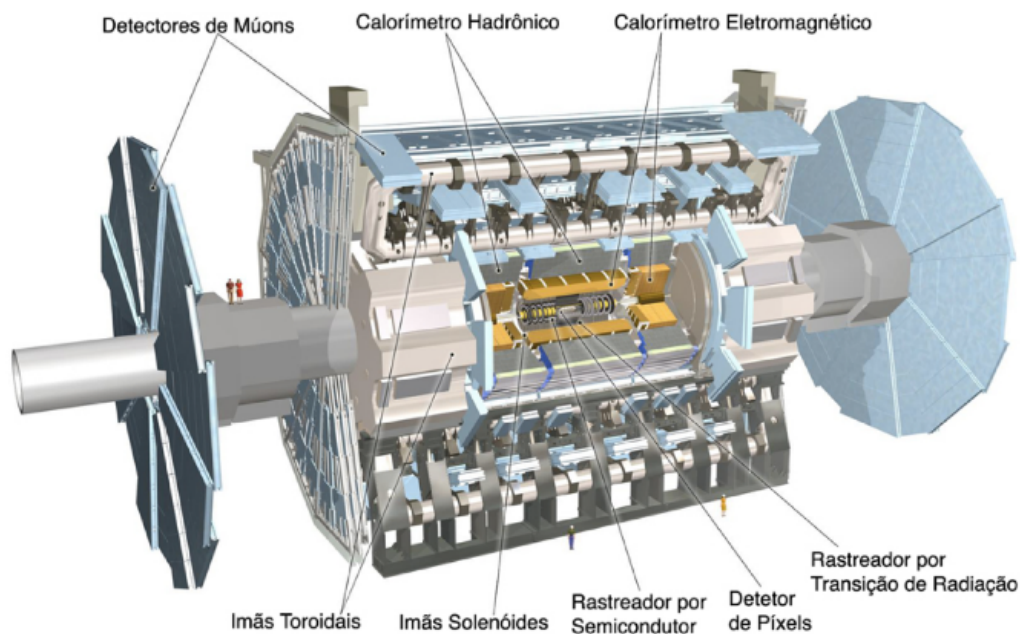


Figura 4.2: Diagrama do detector ATLAS. Extraída de [107].

A infraestrutura do ATLAS é composta por diferentes sub-sistemas de detecção, os quais são envolvidos concentricamente em camadas ao redor do ponto de colisão. Estes detectores registram a trajetória, o momento e a energia depositada por partículas eletricamente carregadas e neutras, usando as características específicas de cada uma delas quando interagem com a matéria, permitindo que elas sejam individualmente identificadas. Os quatro componentes principais do experimento ATLAS são o detector interno (Inner Detector, ID) [86], os calorímetros eletromagnético e hadrônico [109] e o espectrômetro de múons [110]. Encontram-se integrados a estes componentes o sistema de trigger e de aquisição de dados (Data Acquisition System, DAC) [111], que são sistemas especializados multi-nível responsáveis por

selecionar eventos físicos de interesse inseridos em uma enorme quantidade de ruído; e o sistema de computação [112], que desenvolve software com o objetivo de armazenar, processar e analisar de forma ótima a grande quantidade de dados advindos da colisões, em cerca de 130 centros de computação ao redor do mundo [113].

4.2.1 O Sistema de Coordenadas

O ATLAS possui um sistema de coordenadas próprio para descrição das regiões do detector e das interações de partículas. Neste sistema, conhecido como *right-handed coordinate system* por ser um sistema de coordenadas cartesianas que segue a regra da mão direita, a origem é o ponto de interação (IP), onde ocorrem as colisões. Como pode ser visto na Figura 4.3, o eixo-z segue a direção do feixe de partículas, com o z positivo definido no sentido na direção anti-horária do LHC, e o plano x-y é transverso à linha do feixe, com o x positivo apontando horizontalmente para o centro do anel do LHC e o y positivo apontando para a superfície. Entretanto, o sistema de coordenadas cartesiano não é geralmente o mais utilizado. Normalmente, para a análise das colisões, utiliza-se um sistema de coordenadas cilíndricas que acompanha a direção das partículas advindas do ponto de colisão, onde o ângulo azimutal ϕ e o raio ρ são definidos ao redor da linha do feixe no plano xy em relação ao eixo-x positivo e ângulo polar θ é definido a partir da linha do feixe e em relação ao eixo-z positivo, como pode ser visto nas equações 4.1 e 4.2 abaixo:

$$\phi = \arctan\left(\frac{x}{y}\right) \quad (4.1)$$

$$\theta = \arctan\left(\frac{x}{z}\right) \quad (4.2)$$

A direção de projeção das partículas após a colisão é conhecida como pseudorapidez (η) [114] e é definida em função do ângulo polar, conforme equação abaixo:

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (4.3)$$

A Figura 4.4 detalha o sistema de coordenadas utilizado nas análises. A parte superior da Figura mostra o comportamento das coordenadas em relação às coordenadas cartesianas x, y e z, enquanto a parte inferior apresenta um exemplo genérico de um detector segmentado, acompanhando as coordenadas η e ϕ . Pela definição descrita na equação 4.3, é possível verificar que ao alcançar o eixo-z tem-se o valor da pseudorapidez (*eta*) igual a ∞ . Nota-se, portanto, que valores grande de η representam colisões em que as partículas não modificaram significativamente a direção, não produzindo, conseqüentemente, eventos relevantes. Portanto, é comum que se utilizem detectores com baixa resolução quando $\eta > 3$ [114].

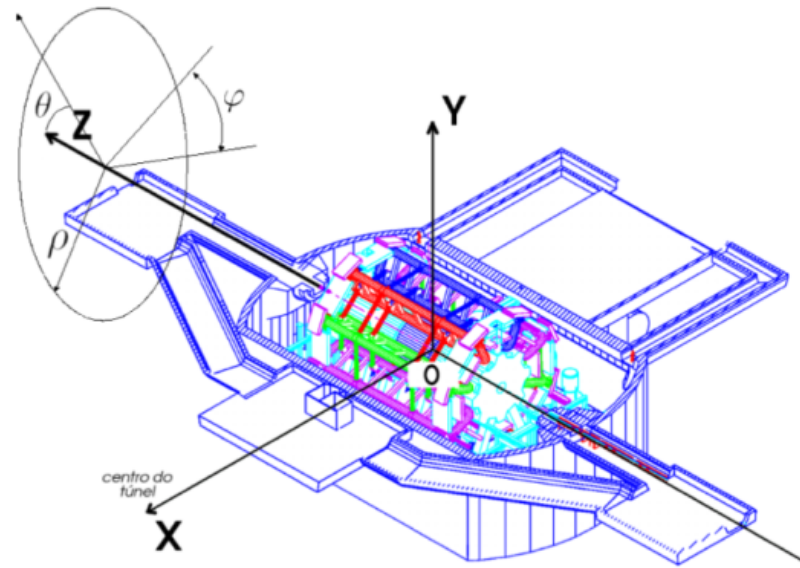


Figura 4.3: Sistema cartesiano de coordenadas do ATLAS. Fonte: [115]

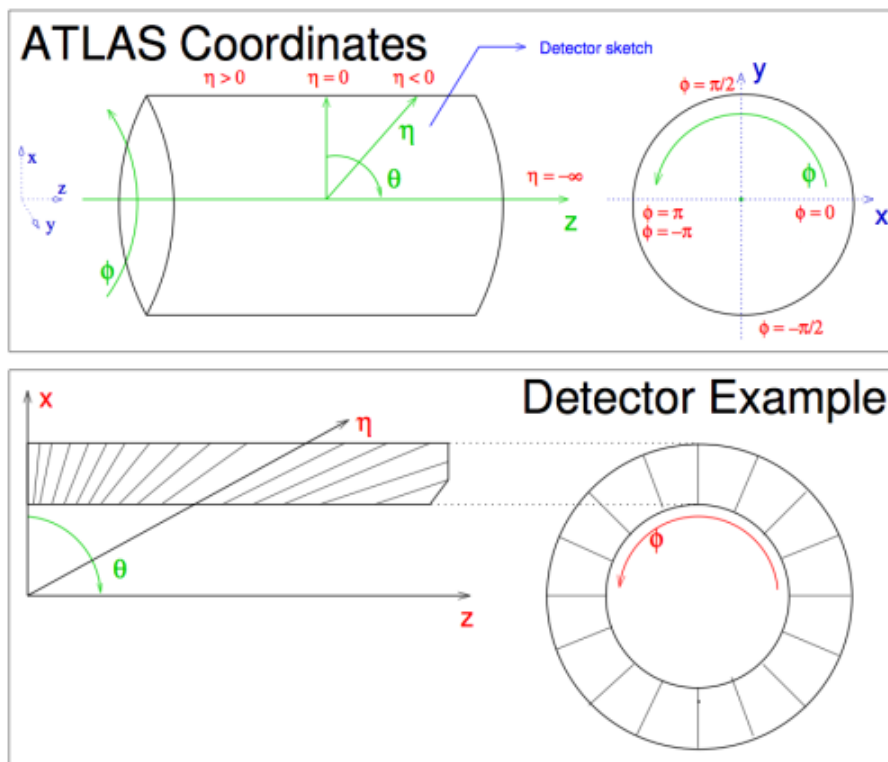


Figura 4.4: O sistema de coordenadas do ATLAS.

4.2.2 O Detector Interno

O detector interno (ID) [116] [117] do ATLAS é responsável por medir a trajetória das partículas oriundas das colisões do LHC. Por estar situado mais internamente,

como o próprio nome sugere, o ID é o primeiro componente do ATLAS a interagir com os produtos dos decaimentos das partículas geradas nas colisões, sendo, portanto, bem compacto e altamente sensível. Ele é circundado por um solenóide que fornece um campo magnético axial de 2 T que altera a trajetória das partículas carregadas. O ID possui uma combinação de detectores baseados em silício e gás que fornecem alta precisão na medição do momento, vértice primário (origem da colisão $p-p$), vértices secundários, parâmetro de impacto direção e carga das partículas carregadas com alta eficiência. Entretanto, apenas partículas eletricamente carregadas têm seu traço registrado por este detector. O ID cobre uma região de $|\eta| < 2,5$ e consiste de três diferentes sistemas de sensores imersos em um campo magnético paralelo ao eixo do feixe: o Detector de Pixel, o Detector de Rastreamento por Semicondutores (SCT) e o Detector de Rastreamento por Radiação de Transição (TRT), que podem ser vistos nas figuras 4.5 e 4.6.

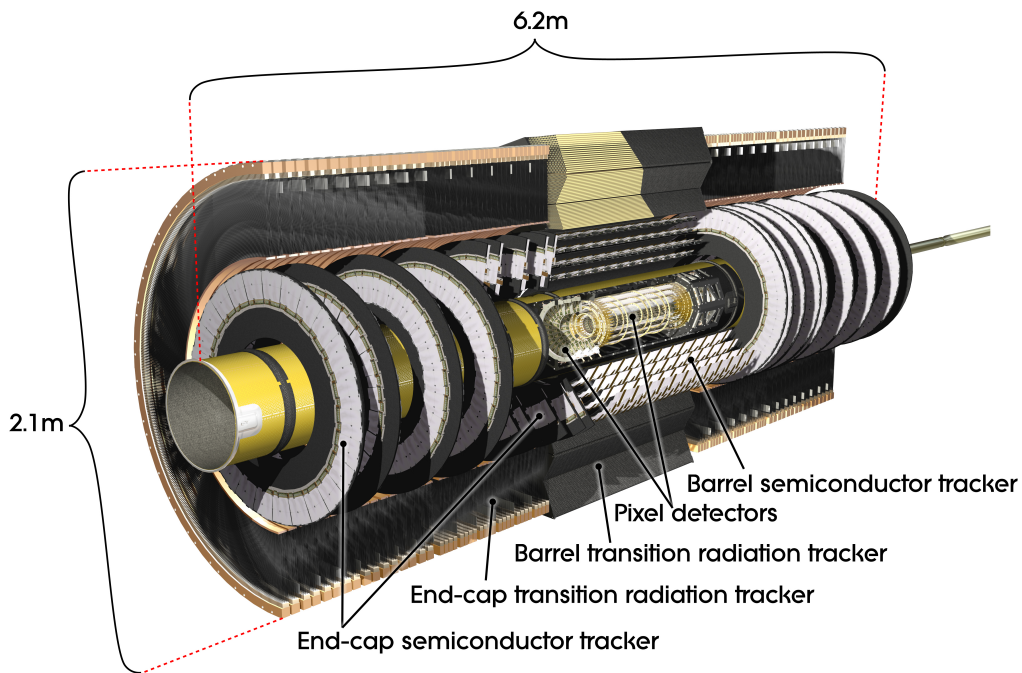


Figura 4.5: Componentes do Detector Interno do ATLAS (barril) [107].

4.2.2.1 O Detector de Pixel

O Detector de Pixel é composto por camadas de sensores de pixel, possuindo 1744 destes sensores, os quais apresentam uma área de 10cm^2 , com os pixels de dimensão $50 \times 400\mu\text{m}^2$ cada, proporcionando um total de 80 milhões de canais e uma resolução de $14 \times 115\mu\text{m}^2$. Inicialmente o Detector de Pixel era composto por 3 camadas dispostas em um barril e duas tampas. o Insertable B-Layer [118] foi adicionado,

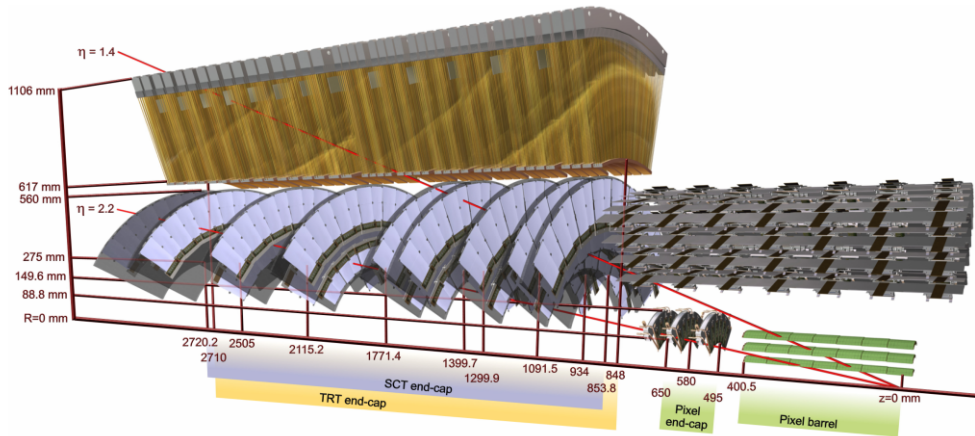


Figura 4.6: Componentes do Detector Interno do ATLAS (barril) [107].

como uma quarta camada de sensores de pixel, durante o desligamento entre os períodos de coleta de dados de 2012 e 2015. Posicionada em um raio de cerca de 33 mm do ponto de interação, comparado a 50 mm da camada mais próxima original. Essa camada permite um rastreamento de maior precisão, incluindo uma maior capacidade de detectar vértices deslocados de jatos b. Este detector foi desenvolvido para fornecer uma alta precisão e granularidade o mais próximo possível do ponto de colisão e, portanto, possui a maior resolução entre os componentes do detector interno, chegando a $10^{-6}m$, e proporciona ao ID a capacidade de distinção entre jatos originados por partículas como hádrons-B e taus.

4.2.2.2 O Detector de Rastreamento por Semicondutores (SCT)

O Detector de Rastreamento por Semicondutores é um detector de tiras de silício, sendo um dos principais dispositivos de rastreamento de precisão no detector interno. O SCT forma a camada intermediária do ID entre o detector de pixels e o TRT. O sistema que compõe o SCT consiste em um rastreador de microfibras de silício composto por 4.088 módulos de dois lados e mais de 6 milhões de canais de leitura através de tiras de leitura implantadas; $60 m^2$ de silício distribuídos em 4 camadas cilíndricas e 18 discos planos de extremidade; e tiras de leitura a cada $80 \mu m$ no silício [86]. Estas tiras permitem que as posições das partículas carregadas sejam registradas com uma precisão de $17 \mu m$ por camada. O SCT permite a distinção de trajetórias de partículas se separadas por uma distância maior que aproximadamente $200 \mu m$.

4.2.2.3 Detector de Rastreamento por Radiação de Transição (TRT)

O Detector de Rastreamento por Radiação de Transição (TRT) é o subsistema mais externo do detector interno e fica localizado em um raio de aproximadamente 550 a 1100 mm, cobrindo uma faixa de $\eta < 2,0$. Este detector fornece a capacidade de identificação de elétrons através da detecção de fótons de raios X de radiação de transição, enquanto está sujeito às mais altas taxas de partículas incidentes já experimentadas por um sistema de rastreamento gasoso de larga escala [119]. Este detector é composto por 350.000 canais de leitura dispostos em um volume de $12m^3$, possuindo cerca de 500.000 tubos cilíndricos de deriva com um diâmetro de 4 mm e um fio de tungstênio banhado a ouro de 0,03 mm de diâmetro no centro. Estes tubos são preenchidos com uma mistura de gases, principalmente Xenônio, que, ao ser atravessado por uma partícula eletricamente carregada, geram uma corrente elétrica (por ionização) que é coletada por um fio no interior de cada tubo. Esta leitura fornece informações adicionais sobre o tipo de partícula que penetrou o sistema de detecção.

4.2.3 O Sistema de Calorimetria

O experimento ATLAS possui um sistema de calorimetria [120] [121] composto por diferentes tecnologias e com o objetivo de detectar e medir a energia para uma variedade de partículas ao longo de uma larga faixa de energia. Este sistema é dividido em duas partes: calorímetro eletromagnético e calorímetro hadrônico.

Como pode ser visto na Figura 4.7, o sistema de calorimetria do ATLAS é composto por calorímetros de amostragem que possuem camadas intercaladas de um material absorvedor e um material ativo. Todas as células dos calorímetros têm geometria aproximadamente retangular no plano $\eta \times \phi$. O calorímetro eletromagnético se situa na camada mais interna e é responsável pela detecção, principalmente, de elétrons e fótons, através da interação eletromagnética, como visto na Seção 3.2.2.1.1. Este calorímetro possui fina granularidade e está dividido em quatro camadas, cada uma com granularidade distinta. O calorímetro hadrônico cobre a camada mais externa e é responsável por detectar partículas hadrônicas, como prótons, píons, káons, que interagem através da interação eletromagnética e força forte, como visto na Seção 3.2.2.1.2. Possui três camadas, também, com granularidades distintas.

A Tabela 4.2 mostra as diferentes regiões de cobertura em η , com granularidade e canais de leitura³ das camadas dos calorímetros. A primeira camada, denominada Pré-Amostrador (PS), não possui meio passivo, sendo constituída unicamente por uma fina camada composta por LAr na região de $|\eta| < 1,8$, sendo o barril abaixo

³Um total de aproximadamente 200.000 canais de leitura dos quais aproximadamente 170.000 do calorímetro eletromagnético

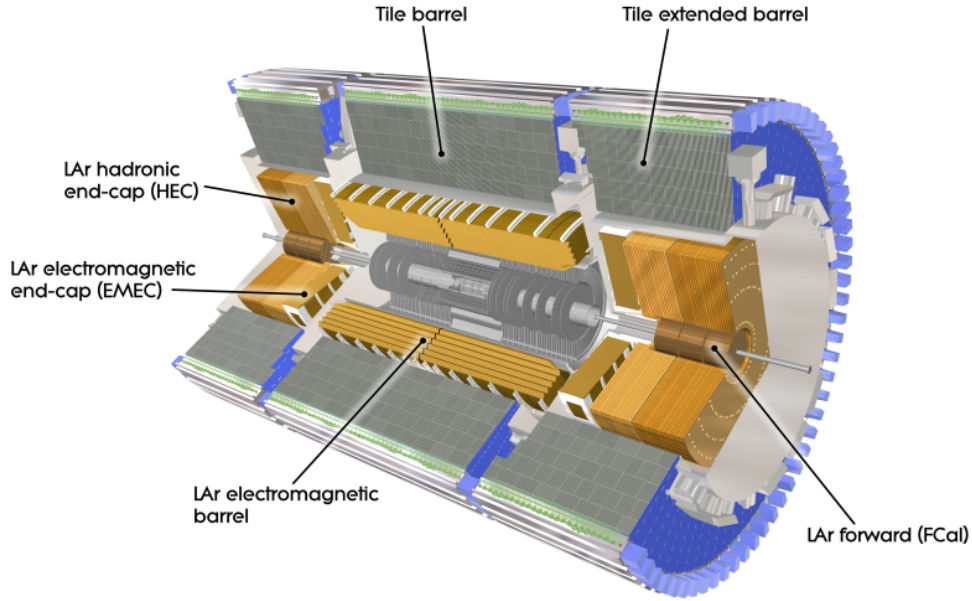


Figura 4.7: O sistema de calorimetria do ATLAS. Extraída de [122]

de $|\eta| = 1,5$ e a tampa acima deste valor. Nesta camada encontram-se eletrodos, com tamanhos diferentes e dispostos de forma distinta no barril e na tampa (de forma perpendicular e paralela respectivamente), que têm como função absorver as partículas advindas de chuveiros que são formados prematuramente antes dos principais calorímetros do ATLAS. O Calorímetro Eletromagnético (ECAL) cobre uma região $|\eta| < 3,2$, na qual o barril se estende na parte central ($|\eta| < 1,475$) e a tampa cobre a região restante ($1,375 < |\eta| < 3,2$). O Calorímetro Hadrônico (HCAL) envolve o ECAL, com mesma cobertura, com o barril, composto pelo *TileCal*, contendo uma parte central ($|\eta| < 1,0$) e uma extensão ($0,8 < |\eta| < 1,7$). A cobertura final ($1,5 < |\eta| < 3,2$), denominada HEC, da região de precisão é realizada por tampas, contendo argônio líquido, devido à maior incidência de radiação nessa região. Por último, é utilizado o Calorímetro Dianteiro (FCAL), que possui menor precisão e tem como objetivo cobrir a região mais próxima do tubo de feixe ($3,1 < |\eta| < 4,9$), e fornecer hermeticidade. Em $|\eta| < 2,47$ a granularidade é mais fina (região de precisão). Devido a maior largura dos chuveiros hadrônicos, a granularidade do HCAL é mais grossa que a do ECAL. Observa-se, também, um decaimento da granularidade com o aumento das camadas de segmentação longitudinal, devido à expansão da espessura lateral do chuveiro à medida que este se propaga pelo calorímetro.

4.2.3.1 Calorímetro Eletromagnético (ECAL)

O Calorímetro Eletromagnético, tanto no barril (EMB) quanto na tampa (EMEC), utiliza meio ativo composto de argônio líquido (LAr) com eletrodos de

Pré-amostrador (PS)	Barril	Tampa	
Cobertura	$ \eta < 1,52$	$1,5 < \eta < 1,8$	
Segmentação Longitudinal	1 camada	1 camada	
Granularidade ($\Delta\eta \times \Delta\phi$)	$0,025 \times 0,1$	$0,025 \times 0,1$	
Canais de Leitura	7808	1536 (ambos os lados)	
Eletrornagnético	Barril	Tampa (EMEC)	
Cobertura	$ \eta < 1,475$	$1,375 < \eta < 3,2$	
Segmentação Longitudinal	3 camadas	3 camadas	$1,5 < pseudorrapidez(\eta) < 2,5$
		2 camadas	$1,375 < \eta < 1,5$
		2 camadas	$2,5 < \eta < 3,2$
Granularidade ($\Delta\eta \times \Delta\phi$)			
Camada 1	$0,003 \times 0,1$	$0,025 \times 0,1$	$1,375 < \eta < 1,5$
		$0,003 \times 0,1$	$1,5 < \eta < 1,8$
		$0,004 \times 0,1$	$1,8 < \eta < 2,0$
		$0,006 \times 0,1$	$2,0 < \eta < 2,5$
		$0,1 \times 0,1$	$2,5 < \eta < 3,2$
Camada 2	$0,025 \times 0,025$	$0,025 \times 0,025$	$1,375 < \eta < 2,5$
		$0,1 \times 0,1$	$2,5 < \eta < 3,2$
Camada 3	$0,050 \times 0,025$	$0,050 \times 0,025$	$1,5 < \eta < 2,5$
Canais de Leitura	101760	62208 (ambos os lados)	
Had. Telhas Cintilantes (TileCal)	Barril	Barril estendido	
Cobertura	$ \eta < 1,0$	$0,8 < \eta < 1,7$	
Segmentação Longitudinal	3 camadas	3 camadas	
Granularidade ($\Delta\eta \times \Delta\phi$)			
Camadas 1, e 2	$0,1 \times 0,1$	$0,1 \times 0,1$	
Camada 3	$0,2 \times 0,1$	$0,2 \times 0,1$	
Canais de Leitura	5760	4092 (ambos os lados)	
Had. Argônio Líquido (HEC)		Tampa	
Cobertura		$1,5 < \eta < 3,2$	
Segmentação Longitudinal		4 camadas	
Granularidade ($\Delta\eta \times \Delta\phi$)		$0,1 \times 0,1$	$1,5 < \eta < 2,5$
		$0,2 \times 0,2$	$2,5 < \eta < 3,2$
Canais de Leitura		5632 (ambos os lados)	
Calorímetro Dianteiro (FCal)		Dianteiro	
Cobertura		$3,1 < \eta < 4,9$	
Segmentação Longitudinal		3 camadas	
Granularidade ($\Delta\eta \times \Delta\phi$)		$\sim 0,2 \times 0,2$	
Canais de Leitura		1762 (ambos os lados)	

Tabela 4.2: Região de cobertura em η , granularidade e número de canais de leitura das camadas dos calorímetros. Adaptado de [123]

cobre como meio absorvedor e meio passivo composto por placas de chumbo. Um corte longitudinal do ECAL está disposto na Figura 4.8. Este calorímetro foi projetado para operar na faixa de 50MeV a 3TeV e possui resolução de energia tanto no barril como na região da tampa, conforme abaixo:

$$\frac{\sigma E}{E} \approx \frac{10\%}{\sqrt{E}} \oplus 0.7\% \quad (4.4)$$

o que significa uma incerteza de 10% na medida para uma energia de 1GeV depositada no calorímetro.

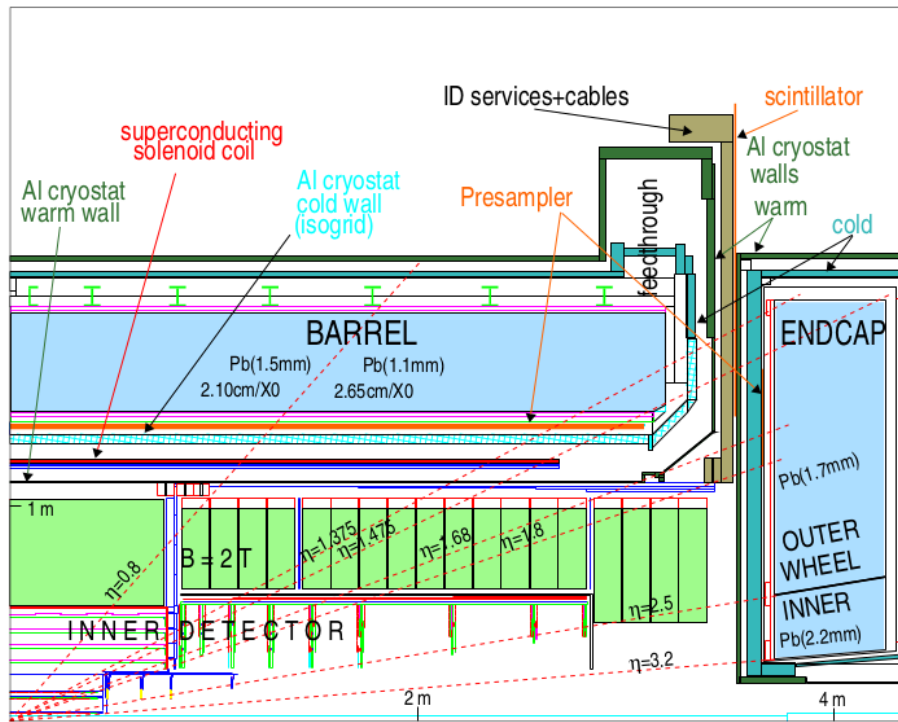


Figura 4.8: Corte longitudinal do ECAL. Extraída de [120]

A absorção de partículas do chuveiro ionizam o argônio líquido permitindo a geração do sinal do ECAL. Existem três regiões de fissura (*crack*) que degradam a resposta do detector, devido a presença de materiais que não contribuem para medida (material morto), as quais estão localizadas em $\eta < 0,02$, $1,37 < \eta < 1,54$ e $2,47 < \eta < 2,5$. O ECAL possui uma estrutura em formato de acordeão, o que possibilita a segmentação do calorímetro. O ECAL possui três camadas, as quais estão ilustradas na Figura 4.9.

A primeira camada denominada EM1 é composta por tiras finas e altamente segmentada em η , com o objetivo de obter uma leitura precisa para análise da dispersão lateral nesta coordenada. Essa camada é importante para detecção de fótons, por não serem medidos no detector de traços, e para casamento em chuveiro e o traço das partículas que possuem carga. A segunda camada EM2 é a camada mais profunda

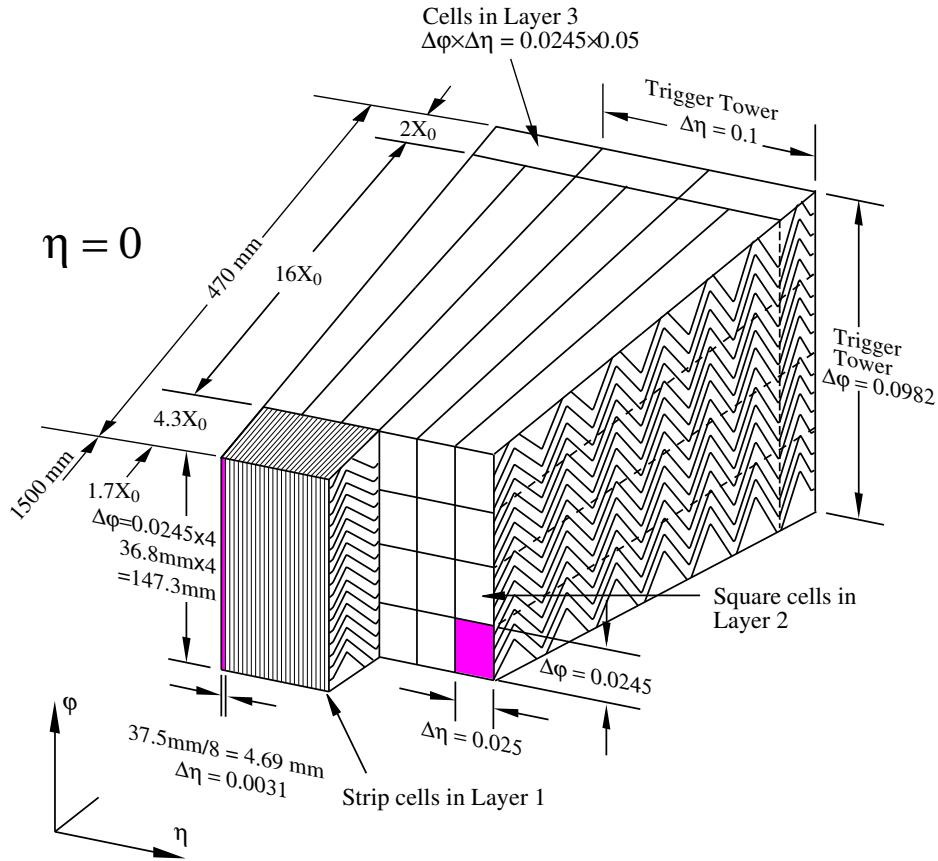


Figura 4.9: Granularidade do ECAL para $\eta = 0$. Extraída de [122]

e a que mais absorve energia das partículas incidentes. A terceira camada EM3 proporciona retenção adicional para partículas eletromagnéticas muito energéticas. A camada denominada Pré-Amostrador se posiciona, longitudinalmente, antes da EM1 e tem como objetivo compensar a existência de material morto e absorver as partículas advindas de chuveiros que são formados antes das 3 camadas mencionadas.

4.2.3.2 Calorímetro Hadrônico (HCAL)

O Calorímetro Hadrônico é composto pelo Calorímetro Hadrônico de Telhas (*TileCal*), Calorímetro Intermediário de Telhas Cintilantes (ITC) e Tampas do Calorímetro Hadrônico (HEC).

O *TileCal* consiste em um barril central e mais dois barris para cada lado e fornece três amostragens longitudinais. É formado por telhas de cintiladores de plástico como material ativo e aço como material passivo. Os cintiladores de plástico, quando excitados pelas partículas, emitem fótons que são capturados por fibras e, então, têm seus sinais amplificados por meio de Tubos fotomultiplicadores (PMT) [124].

O HEC utiliza placas de cobre como material absorvedor e argônio líquido como meio ativo. Sendo assim, a carga de ionização que é deixada pelo chuveiro no LAr é coletada através de uma tensão aplicada entre os eletrodos, o que produz um pulso de corrente para o sinal de leitura [125].

O *TileCal* e o HEC possuem resolução de energia conforme abaixo:

$$\frac{\sigma E}{E} \approx \frac{50\%}{\sqrt{E}} \oplus 3\% \quad (4.5)$$

4.2.3.3 Calorímetro Dianteiro (FCAL)

O calorímetro dianteiro (FCAL) [126] tem como objetivo medir ou melhorar a medida de energia região próxima ao tubo de feixe e aumentar a hermeticidade do detector como um todo. Para isso, deve suportar a elevada taxa de radiação nesta região. Sua primeira camada possui uma matriz de cobre absorvedor contendo buracos destruídos uniformemente, onde são colocadas hastes coaxiais e tubos, também de cobre, separados por fibra de plástico resistente a radiação. A matriz e o tubo são o material absorvedor enquanto o restante do espaço entre o tubo e a haste é preenchido com Lar. Esta estrutura foi inventada especificamente para aplicações em ambientes de alta taxa. A colocação do FCAL adjacente aos outros calorímetros relativamente próximos do ponto de interação fornece vantagens, incluindo calorimetria quase perfeita e blindagem natural para o sistema de múons. A hermeticidade deste calorímetro permite uma boa reconstrução de jatos e determinação da energia transversa faltante. A resolução deste detector é dada por:

$$\frac{\sigma E}{E} \approx \frac{100\%}{\sqrt{E}} \oplus 10\% \quad (4.6)$$

4.2.4 O Espectrômetro de Múons

O espectrômetro de múons é responsável por identificar e medir a trajetória dos múons a fim de determinar a direção, carga e momento destas partículas. Os múons são partículas que normalmente passam pelo detector interno e pelo sistema de calorimetria sem serem detectados. O espectrômetro, portanto, localiza-se ao redor dos calorímetros do ATLAS, sendo composto por 4.000 câmaras individuais de múons, as quais usam quatro tecnologias diferentes [127] [128]. As subseções do seu sistema são a *Thin-gap chamber*, utilizada para filtragem (triggering) e para medição da segunda coordenada nas extremidades do detector. As câmaras de placas resistivas, também utilizadas para filtragem e medição da segunda coordenada, mas na região central. Os tubos de arrasto monitorados, que medem a curvatura da trajetória; e as câmaras de tira catódica, que medem as coordenadas de precisão nas extremidades do detector. Um diagrama do espectrômetro de múons pode ser visto

na Figura 4.10.

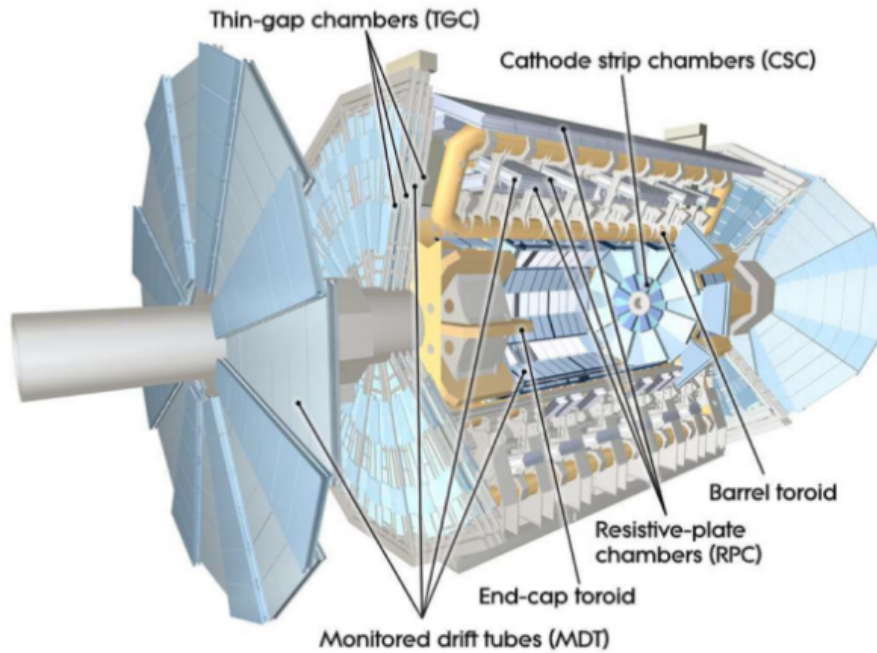


Figura 4.10: Espectrômetro de múons do ATLAS. Extraída de [114]

4.3 Sistema de Filtragem do ATLAS

Os eventos de interesse da física no experimento ATLAS, e nos experimentos do LHC, em geral, são extremamente raros. Um dos principais objetivos do ATLAS, por exemplo, é o estudo do bóson de Higgs. Enquanto o número de interações físicas anuais de qualquer tipo que espera-se que ocorra é de $1,7 \times 10^{16}$, o número de eventos interessantes exibindo o Bóson de Higgs é de aproximadamente 17.000. A taxa de geração de eventos é de 40Mhz e cada evento gerado leva aproximadamente 1,5 MB para ser armazenado. Sendo assim, o resultado da multiplicação resulta em 70 TB de informações geradas a cada segundo. Como não é possível e nem interessante armazenar todos esses dados, um sistema de filtragem online torna-se essencial. No ATLAS, esse sistema é denominado *Trigger*. O sistema de *Trigger* do ATLAS é um sistema multinível que deve identificar os raros padrões de decaimento, que representam a física de interesse, para poder isolá-los da massa de eventos majoritariamente advinda de ruído de cromodinâmica quântica (Quantum Chromodynamics, QCD). O sistema é dividido em 2 etapas:

1. *Trigger* de primeiro nível L1: O *Trigger* de nível 1 é implementado em hardware e utiliza equipamentos eletrônicos personalizados para determinar as regiões de interesse (RoIs) no detector, tomando como entrada informações sobre o calorímetro

de granularidade grossa e o detector de múons. O *Trigger* de nível 1 reduz a taxa de eventos da taxa de *bunchcrossing* do LHC de aproximadamente 30 MHz a 100 kHz. O tempo de decisão para uma aceitação de nível 1 é de 2,5 μ s.

2. *Trigger* de alto nível (HLT): As RoIs formadas no nível 1 são enviadas ao HLT (*High Level Trigger*), no qual algoritmos sofisticados de seleção são executados usando toda a granularidade do detector. O HLT reduz a taxa de saída de nível 1 de 100 kHz para aproximadamente 1 kHz, em média, dentro de um tempo de processamento de cerca de 200 ms. Para tanto, um cluster computacional é utilizado para processamento de reconstrução e filtragem de eventos, sendo tais processos implementados em C++ e Python.

4.3.1 L1

O primeiro nível de filtragem (L1) realiza a seleção inicial, com base nas informações obtidas com granularidade reduzida de um subconjunto dos detectores. Devido ao alto número de canais dos detectores de traço e ao alto custo computacional dos algoritmos desses detectores, decidiu-se usar apenas as informações dos calorímetros e dos detectores rápidos de múons para compor as informações do primeiro nível do Trigger. A granularidade nesse nível não é completa, pois o tempo para a tomada de decisões nesse nível é da ordem de microssegundos. Outra característica desse nível é que tudo é implementado em hardware programável como FPGA, que garante maior flexibilidade aos projetos e permite a implementação de algoritmos mais complexos. Reduzir a quantidade de informações nesse nível é crucial devido aos seus requisitos de latência. Assim, o L1 utiliza janelas deslizantes para separação das regiões de interesse no sistema de calorimetria e busca pelo local de máxima energia calculada na região central EM, como pode ser visto na Figura 4.11. Quando a RoI é obtida, um corte em E_T , dependente da região de η é aplicado. Uma vez que a RoI é aprovada, esta é enviada para o HLT, onde algoritmos serão utilizados para extração de característica e filtragem.

4.3.2 HLT

O filtro feito pelo L1, devido aos requisitos de latência, é baseado em cortes simples, utilizando informações obtidas apenas a partir da leitura das células do sistema de calorimetria. Considerando a informação das RoIs que foram aprovadas pelo L1, o HLT utiliza algoritmos mais rebuscados e que levam mais tempo para tomada de decisão, como cortes lineares e métodos multivariados para seleção de candidatos a elétrons. Além das células do calorímetro usadas no *Trigger* L1, as informações de traço também são usadas nos testes de hipóteses no HLT, a fim de tomar a decisão de aceitação ou rejeição. O *Trigger* de alto nível é subdividido em

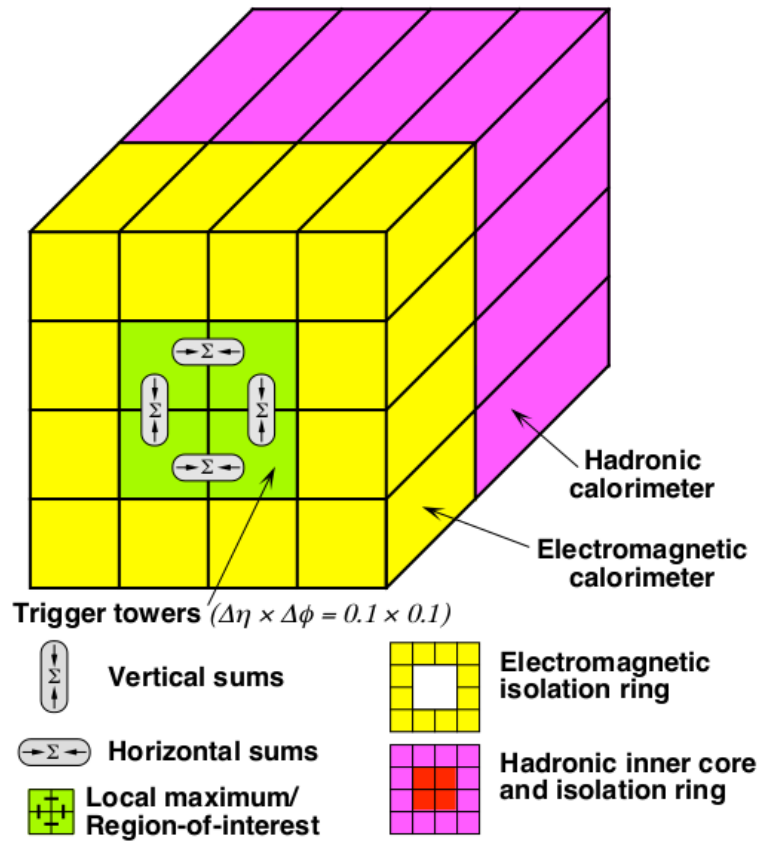


Figura 4.11: Janela deslizante no L1 para seleção de elétrons e fótons. Extraída de [129]

dois subsistemas, sendo eles o *Trigger* rápido e os sistemas de alta precisão. O sistema de *Trigger* rápido é composto por algoritmos de reconstrução de baixa latência e realiza a reconstrução e extração das características do calorímetro com granularidade aumentada, mas não com granularidade total, pois essa etapa permite reduzir ainda mais a taxa de eventos possibilitando que o *Trigger de alta precisão* possa executar um algoritmo ainda mais complexo com informações completas do evento. Na última etapa de trigger, toda a informação do evento é usada em uma versão simplificada do algoritmo *offline*, agora sendo implementada como uma abordagem de chamada de Likelihood (descrita na Seção 4.3.3.2). Após a decisão final, os eventos selecionados prosseguem para o armazenamento. A Figura 4.12 mostra um fluxograma para seleção de elétrons no HLT do ATLAS e inclui todas as etapas presentes na seleção de elétrons neste experimento. Nota-se a presença dos dois algoritmos de seleção na etapa de seleção rápida baseada em calorimetria, com destaque para o *NeuralRinger* acima de 15 GeV. Após esta seleção, os traços são reconstruídos e são realizados cortes finalizando assim a etapa rápida. Na sequência, é realizada a etapa precisa, onde a energia é calibrada e submetida a um novo corte. Por fim,

o candidato é submetido à seleção multivariada baseada no método de Likelihood, que utiliza informação de calorimetria e de traço.

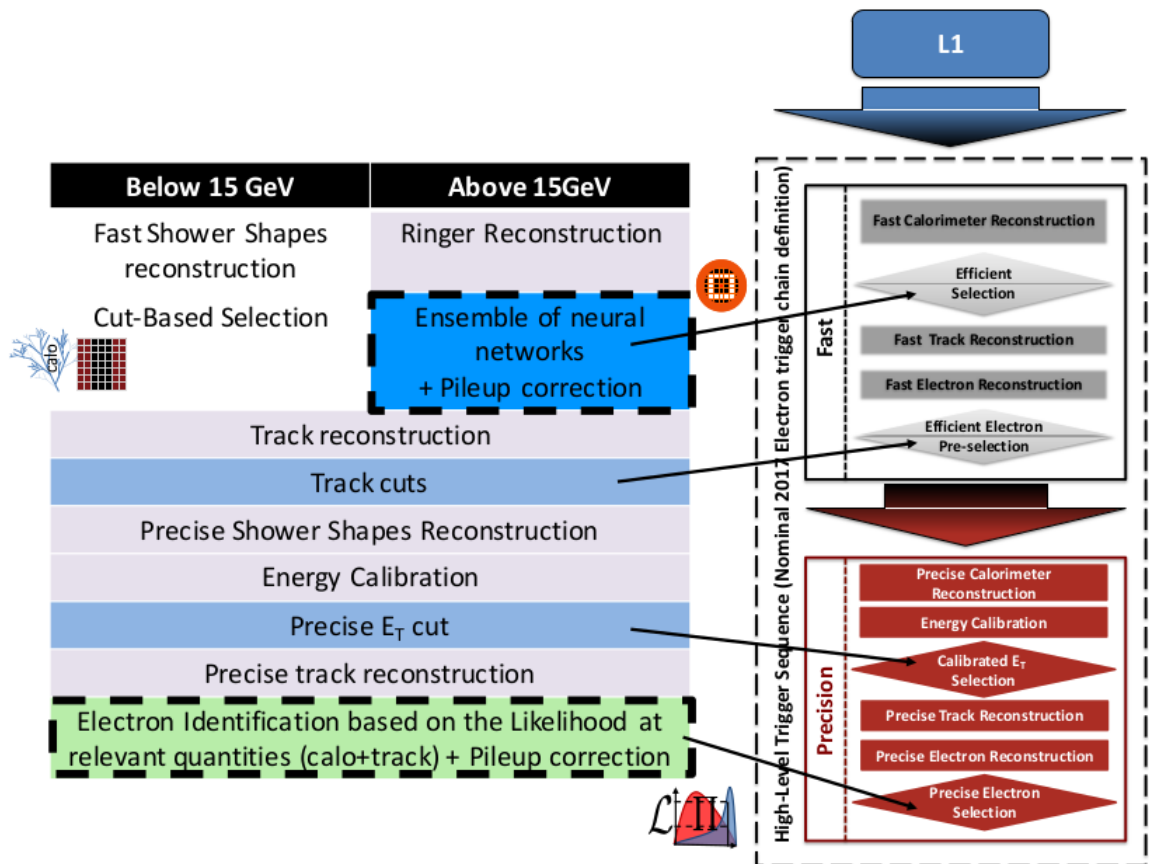


Figura 4.12: Fluxograma empregado para seleção de elétrons no HLT do ATLAS. Fonte: [130]

4.3.3 Sistema de Filtragem *Offline*

Quando um evento é rejeitado pelo ambiente *online*, não é possível recuperá-lo. Para evitar a perda de eventos que possam conter objetos físicos de interesse, a eficiência de detecção é aumentada, gerando, também, um aumento no falso alarme. Desta forma, após todas as etapas de seleção descritas na Seção 4.3, ainda são armazenados, junto com os canais de interesse, eventos provenientes de ruído de fundo. Além disso, existe também a necessidade de separação dos eventos de acordo com os objetivos de estudo pelos diferentes grupos que compõem a colaboração. Alguns grupos, por exemplo, estão interessados em eventos que permitam aprofundar o estudo do Bóson de Higgs, enquanto outros estão procurando Supersimetria.

Com o objetivo de atender aos diferentes interesses de pesquisa, o sistema de filtragem *offline* é utilizado. Neste ambiente, o tempo de filtragem não é mais um

fator crítico e, portanto, algoritmos mais complexos e eficientes podem ser utilizados para analisar os eventos de acordo com os requisitos do estudo a ser realizado. Além disso, as grandezas tanto de calorimetria como de traço estão disponíveis para uso neste sistema. Atualmente, para seleção de elétrons no ambiente *offline*, um método de verossimilhança (Likelihood) [131] é utilizado.

4.3.3.1 Grandezas físicas utilizadas na seleção de elétrons

A seleção de elétrons no ATLAS faz uso de grandezas criadas a partir de conhecimento especialista acumulado, tendo, tais grandezas, alta capacidade de discriminação para separação elétrons-jato. A Tabela 4.3, apresenta essas grandezas, entre as quais encontram-se sete que são relacionadas ao formato e largura do chuveiro de partículas, formadas por razões relacionadas com a energia proveniente do calorímetro. Outras seis são geradas a partir do conhecimento da informação de traço e do casamento dessas duas fontes de informação. Essas 13 variáveis são utilizadas pela Likelihood [131].

Tabela 4.3: Definição das grandezas empregadas pelo ATLAS para discriminação de elétrons. Adaptado de [132].

Tipo	Descrição	Nome
Vazamento Hadrônico	Razão em E_T da primeira camada do calorímetro hadrônico para a E_T do <i>cluster</i> EM (empregado para a região $ \eta < 0,8$ ou $ \eta > 1,37$)	R_{had1}
	Razão em E_T no calorímetro hadrônico para a E_T a E_T do <i>cluster</i> EM (empregado para a região $0,8 < \eta < 1,37$)	R_{had}
Região Traseira do Calorímetro EM	Razão de energia da camada traseira para a energia total do calorímetro de acordeão EM. Essa variável é somente empregada abaixo de 100 GeV pois se sabe que ela é ineficiente em altas energias.	f_3
Camada Central do Calorímetro EM	Largura lateral do chuveiro, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, onde E_i é a energia e η_i a pseudorapidez da célula i e a soma é calcula dentro de uma janela com 3×5 células.	$w_{\eta 2}$
	Razão de energia em 3×3 células para a energia em 3×7 células centradas na posição do <i>cluster</i> do elétron	R_ϕ
	Razão de energia em 3×7 células para a energia em 7×7 células centradas na posição do <i>cluster</i> do elétron	R_η
Camada de Tiras do Calorímetro EM	Largura do chuveiro, $\sqrt{(\sum E_i (i - i_{max})^2)/(\sum E_i)}$, onde i percorre todas as tiras na janela $\Delta\eta \times \Delta\phi \approx 0,0625 \times 0,2$, correspondendo tipicamente a 20 tiras em η . i_{max} é o índice da tira com a maior deposição de energia.	w_{stot}
	Razão da diferença entre os depósitos de maior e a de segunda maior energia no <i>cluster</i> para a soma dessas energias	E_{ratio}
	Razão da energia na camada de tiras para o total de energia no acordeão EM do calorímetro.	f_1
Condições do Traço	Número de pontos na camada mais interna de pixel; discriminante contra conversões de fótons.	n_{blayer}
	Número de pontos no detector de pixel.	n_{pixel}
	Número total de pontos nos detectores de pixel e Detector de Rastreamento por Semicondutores (SCT).	n_{Si}
	Parâmetro transverso do ponto de impacto em relação ao ponto de colisão	d_0
	Significância do parâmetro transverso do ponto de impacto definido como a razão de d_0 e sua incerteza	d_0/σ_{d_0}
	Perda de momento do traço entre o ponto originário e final de medida dividido pelo momento original	$\Delta p/p$
TRT	Probabilidade de verossimilhança baseada na radiação de transição no Detector de Rastreamento por Transição de Radiação (TRT)	eProbabilityHT
	Transformação em eProbabilityHT: $TRT_PID = -\frac{\log((1.0/eProbabilityHT)-1.0)}{15}$	TRT_PID
Casamento <i>cluster</i> -traço	$\Delta\eta$ entre a posição do <i>cluster</i> na camada de tiras e o traço extrapolado	$\Delta\eta_1$
	$\Delta\phi$ entre a posição do <i>cluster</i> na camada central e o traço extrapolado do ponto originário	$\Delta\phi_2$
	Definido como $\Delta\phi_2$, mas o momento do traço é reescalado para a energia do <i>cluster</i> antes de extrapolar o traço do ponto originário até a camada central do calorímetro	$\Delta\phi_{res}$
	Razão da energia do <i>cluster</i> para o momento do traço	E/p

4.3.3.2 Likelihood

O método de verossimilhança faz uso de funções de densidade de probabilidade (PDFs), de sinal e de ruído, de variáveis discriminantes de calorimetria e traço. Com base nestas PDFs, uma probabilidade geral é calculada para o evento ou objeto ser sinal ou ruído. Tais probabilidades são, então, combinadas em um discriminante no qual um corte é aplicado, como pode ser visto na equação abaixo:

$$d_L = \frac{L_S}{L_S + L_B}, \quad L_{S(B)}(\vec{x}) = \prod_{i=1}^n P_{S(B),i}(x_i) \quad (4.7)$$

onde \vec{x} é representado pelas variáveis discriminantes na identificação de elétrons e $P_{S,i}(x_i)$ é o valor da PDF referente ao sinal (elétron) da i -ésima variável dada por x_i . Da mesma forma, $P_{B,i}(x_i)$ refere-se ao valor da PDF para o ruído [131].

A escolha do valor de corte (ponto de operação) determina a eficiência do sinal ou rejeição do ruído de fundo atingida pelo método. Este método apresenta vantagens, quando comparado aos métodos tradicionais de cortes rígidos em variáveis físicas, com relação ao seu potencial de discriminação, uma vez que a construção da probabilidade permite que as caudas de variáveis discriminantes sejam paricalmente recuperadas, resultando em uma maior eficiência com a mesma rejeição ao ruído [131].

4.3.4 O Algoritmo NeuralRinger

Os chuveiros de partículas depositados nos calorímetros podem ser descritos por seus perfis lateral e longitudinal, como mencionado na Seção 3.2.2.1. Esta informação é crucial para a caracterização e discriminação de eventos de interesse. Os chuveiros gerados por elétrons possuem um formato cônico, estreito e com pouca penetração, enquanto as partículas provenientes de jatos (ruído) apresentam chuveiros mais largos e profundos [62]. Com o objetivo de preservar a informação referente à geometria do chuveiro, e, ao mesmo tempo, reduzir a dimensionalidade da informação de energia das células do sistema de calorimetria, a representação da energia através de anéis concêntricos foi proposta [133].

O sistema de calorimetria do ATLAS conta com 7 camadas longitudinais de amostragem, sendo uma de pré-amostragem (PS, Pre-Sampler), três camadas eletromagnéticas (EM1, EM2 e EM3) e três camadas hadrônicas (HAD1, HAD2 e HAD3). Os anéis de energia utilizados pelo *NeuralRinger* são formados ao longo de todas estas sete camadas, totalizando 100 anéis. A formação destes anéis se inicia na célula mais energética de cada camada, cuja energia é considerada o primeiro anel da camada, e continua através da soma da energia (E_T) das células externas em relação ao anel anterior, até que a Região de interesse (RoI) seja coberta ou um

número máximo de anéis (100) seja atingido, como pode ser visto na Figura 4.13.

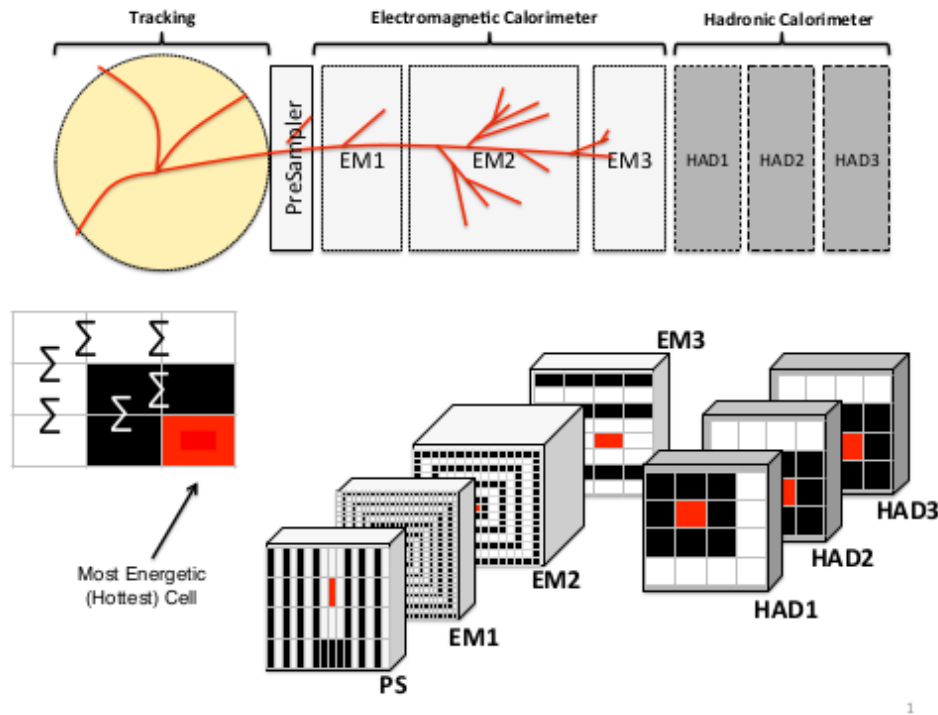


Figura 4.13: Construção dos anéis pelo algoritmo NeuralRinger através das seções do calorímetro. Fonte: [134]

Após o processo de criação dos anéis, o teste de hipótese é realizado através de um *ensemble* de classificadores compostos por redes neurais do tipo Perceptrons de Múltiplas Camadas (Multilayer Perceptron, MLP) [135], que são projetadas para regiões do plano formado pela energia transversa e posição da partícula, para discriminação elétron-hádron. Em 2017, o *NeuralRinger* entrou em operação em uma das etapas do sistema de filtragem online (Trigger) de elétrons do ATLAS, em sua camada de alto nível (High Level Trigger, HLT).

Tradicionalmente, os métodos utilizados na filtragem *online* consistem em degradações dos métodos atuantes na análise *offline*, de maneira que estes possam atender aos requisitos de latência da primeira. No caso do *NeuralRinger*, sua estréia na filtragem *online* ocorreu sem que um método similar estivesse atuando na filtragem *offline*. Desta forma, a fim de manter a consonância entre os métodos de decisão em ambas as etapas, é importante o estudo deste algoritmo em ambiente *Offline*. Este ambiente, por outro lado, apresenta a vantagem de não possuir requisitos restritos de latência, permitindo, assim, o estudo de modelo mais complexos. Neste caso, existem opções que incluem modelos de redes neurais profundas e modelos de codificação, como redes neurais auto-codificadoras. Estes modelos são treinados para codificar e reconstruir as variáveis em que foram treinados e são utilizados, normal-

mente, para redução não linear de dimensionalidade, detecção de anomalia, redução ou remoção de ruído em imagens e até como modelos generativos. No contexto da física experimental de altas energias, especificamente, as redes auto-codificadoras têm sido empregados com o objetivo de codificar o sinal conhecido para, então, avaliar se, ao decodificar a codificação encontrada, pode-se detectar um sinal anômalo, o qual pode ser um candidato à física nova, além do Modelo Padrão [136].

Este trabalho analisa diferentes arquiteturas de redes auto-codificadoras visando analisar a capacidade de reconstrução de eventos que utilizam a informação anelada de calorimetria, bem como busca encontrar uma representação que possa contribuir positivamente para a tarefa de discriminação elétron-jato, seja com redução de dimensionalidade sem perda de eficiência ou com a obtenção de uma representação latente que melhore a capacidade discriminativa do *NeuralRinger*. No caso de sucesso na aplicação destes modelos em ambiente *Offline*, é possível, mais facilmente, propor sua utilização em ambiente *Online*, como normalmente é feito na colaboração ATLAS. O próximo capítulo faz uma revisão de modelos de codificação, descreve as redes auto-codificadoras e apresenta uma revisão de possíveis aplicações tanto na área de física de altas energias, como em outras áreas.

Capítulo 5

Aprendizado de Representação Através de Auto-codificadores e Trabalhos Relacionados

Neste capítulo, será feito um levantamento dos métodos de representação mais comuns na literatura e sua aplicação na área de física de altas energias. Especificamente, serão revisados métodos utilizados para representar a informação de calorimetria, bem como as variáveis provenientes de detectores de traço, tanto na etapa de reconstrução de eventos, como na análise física envolvendo identificação de jatos, de processos físicos de interesse e, inclusive, a busca de nova física.

5.1 Aprendizado de Representação

Um fator importante que afeta o desempenho de métodos de aprendizado de máquina é a escolha de representação de dados em que tais métodos são aplicados. Portanto, muito do esforço da aplicação de algoritmos de aprendizado de máquina se concentra na transformação dos dados e no teste de diferentes técnicas de pré-processamento. Segundo [137], tal esforço destaca a fraqueza dos atuais algoritmos de aprendizado em extrair e organizar as informações discriminativas dos dados. Em [137], Bengio et al fazem uma revisão de técnicas de aprendizado de representação com foco em métodos de aprendizado profundo, formados pela composição de múltiplas transformações não-lineares. Nos últimos anos, a área de aprendizado de representação, principalmente com a utilização de modelos considerados profundos, tem experienciado um aumento de atividade científica de pesquisa ao mesmo tempo que tanto a academia como a indústria tem apresentado sucessos empíricos. Alguns exemplos deste sucesso apareceram nas áreas de reconhecimento de fala, reconhecimento de imagens, reconhecimento de objetos, processamento de lingua-

gem natural, contribuindo positivamente, inclusive, para a área de físicas de altas energias. Uma perspectiva histórica dos trabalhos mais relevantes na área pode ser encontrada em [138].

5.1.1 PCA e NLPCA

A análise de componentes principais (PCA) projeta dados N -dimensionais em M direções ortogonais nas quais os dados possuem mais variância. Essas M direções principais formam um sub-espço de dimensão menor, sendo possível representar amostras N -dimensionais por suas projeções nas M direções principais. A PCA é uma análise multivariada por autovetores (Vetores Próprios). Com frequência, sua operação pode ser tomada como sendo reveladora da estrutura interna dos dados, de uma forma que melhor explica a variância nos dados. Considerando uma rede auto-codificadora (ou autoencoder) com neurônios lineares nas camadas escondidas e de saída e com M menor que N neurônios na camada escondida, as unidades escondidas irão aprender funções lineares dos dados e minimizar o erro de reconstrução, reproduzindo o que é feito pela PCA e, portanto, o erro de reconstrução da rede auto-codificadora será o mesmo de PCA. Entretanto, apesar de as M unidades escondidas gerarem o mesmo espaço que as M componentes principais, elas podem ser uma rotação ou inclinação desses eixos. Os vetores de pesos obtidos pela rede auto-codificadora não serão ortogonais e tendem a ter variâncias iguais. Além disso, o gradiente descendente estocástico (SGD) será menos eficiente que o algoritmo usado para PCA, a não ser que se trate de uma grande quantidade de dados. Uma descrição completa da forma convexa da função custo, que possui um único ponto mínimo, quando o erro médio quadrático é usado no aprendizado de redes neurais com ativações lineares é apresentado por Baldi et Hornik [139], mostrando que o resultado obtido com o algoritmo de retropropagação do erro (Backpropagation) seria uma projeção no mesmo sub-espço gerado pelos primeiros componentes principais da matriz de covariância associada às variáveis de treino.

Bourlard et al [140] critica o uso de redes neurais auto-associativas, colocando que as ‘não-linearidades na camada escondida são inúteis’ e que valores ótimos para os parâmetros podem ser derivados diretamente por técnicas lineares puras utilizando decomposição em valores singulares e aproximação de matrix por baixo posto, de forma similar à transformada de Karhunen-Loève, além de que estas proporcionam uma clara interpretação para o papel de diferentes parâmetros. Ao explicar a motivação do uso de três camadas escondidas na sua proposta de redes neurais auto-associativas para análise de componentes principais não linear (NLPCA), Kramer 1991 [141] afirma que, no caso da utilização de apenas uma camada escondida (retirando as camadas de mapeamento e de-mapeamento propostas), se as ativações

da camada de gargalo fossem lineares, os resultados seriam similares à PCA, como mostrado por Sanger 1989 [142], e no caso de ativações sigmoidais, as formas funcionais da codificação e decodificação ainda seriam severamente restringidas, sendo possível representar apenas uma combinação linear das entradas comprimida pela sigmoide no intervalo $(-1,1)$. Kramer sustenta que para alcançar extração não-linear de características três camadas escondidas são necessárias, sendo uma camada de mapeamento, uma de gargalo e uma de desmapeamento, como pode ser visto na Figura 5.1. Os resultados de Cybenko [143], de que uma camada escondida é suficiente para modelar qualquer função contínua não linear em subconjuntos compactos e com um número suficiente de unidades escondidas, aplica-se para as funções G (mapeamento) e H (de-mapeamento) individualmente. Kramer também apresentou uma rede NLPCA sequencial, que possui uma série de redes auto-associativas, em que cada rede treina apenas um nó de gargalo e que cada nó de gargalo, supostamente, modela um fator diferente. Sendo assim, é possível evitar a tendência de um ou mais nós de gargalos alinharem com o primeiro fator.

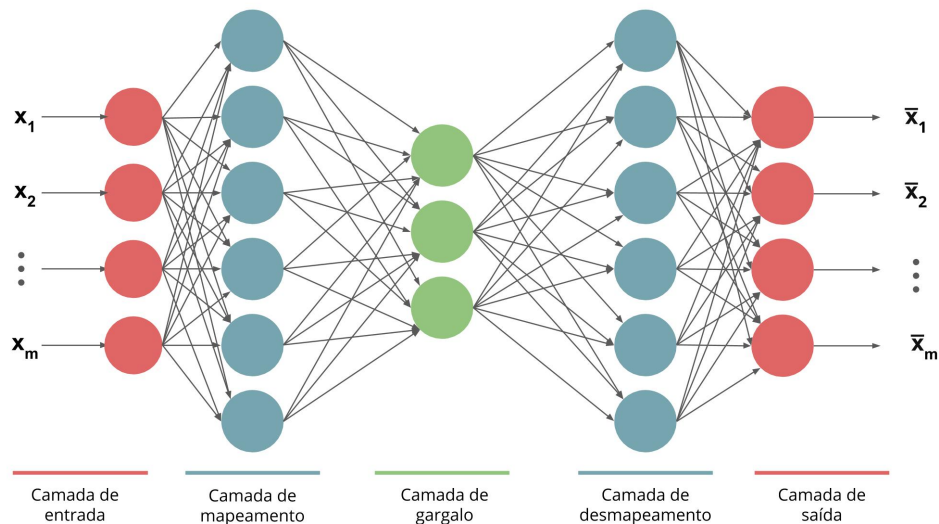


Figura 5.1: Modelo neural para análise de componentes principais não-linear (NLPCA).

5.1.2 Redes neurais auto-codificadoras

Uma rede auto-codificadora é uma Rede Neural Feed-Forward treinada para reproduzir as entradas na camada de saída, combinando uma função de codificador, que converte os dados de entrada em uma representação diferente, e uma função de

decodificador, que converte a nova representação de volta ao formato original, como pode ser visto na Figura 5.2. As redes auto-codificadoras são treinadas para preservar o máximo de informações possível quando uma entrada é executada através do codificador e, em seguida, do decodificador, sendo um algoritmo com perdas treinado de forma não supervisionada afim de obter de códigos eficientes [144]. Entre seus principais objetivos de uso estão a redução de dimensionalidade não-linear, onde este tipo de rede pode comprimir dados semelhantes aos dados em que foi treinado, mas obtendo resultados ruins em dados que não apareceram no treinamento. Desta maneira uma rede auto-codificadora não apresenta características interessantes como outros algoritmos de compressão, mas, em contrapartida, pode atuar como detector de anomalias, baseando-se no erro de reconstrução de classes em que este não foi treinado. Outra habilidade está na possibilidade de extrair características significativas dos dados, podendo separar os fatores de variação que explicam os dados observados, à medida que diferentes restrições são postas na função custo que norteia a reconstrução. Redes auto-codificadoras também são utilizadas em aprendizado semi-supervisionado ou pré-treino, afim de tirar proveito da vasta disponibilidade de dados não rotulados, podendo, então, melhorar o poder discriminativo de classificadores que possuem apenas uma pequena quantidade de eventos. Diferentes tipos de redes auto-codificadoras alcançam diferentes propriedades e, recentemente, as conexões teóricas entre estas redes e os modelos de variáveis latentes trouxeram estes para a linha de frente da modelagem generativa.

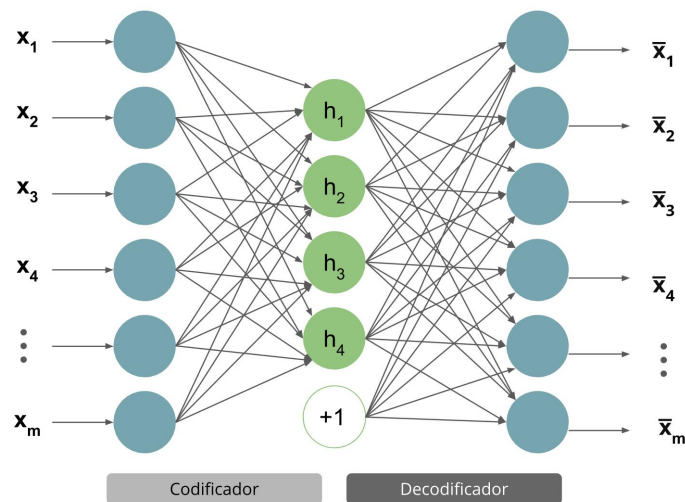


Figura 5.2: Diagrama ilustrativo da arquitetura de uma rede auto-codificadora.

Redes auto-codificadoras profundas foram pensadas na década de 1980, mas

não era possível treiná-las de maneira a conseguir resultados significativamente melhores que a PCA, principalmente devido ao problema no treinamento relacionado ao desaparecimento do gradiente. As primeiras aparições do uso de redes neurais auto-associativas, com gargalo para concentrar informação, na literatura remetem a Hinton et al 1985 [145], Lecun et al 1987 [146], [147]. Ballard 1987 [148] utilizou múltiplas redes de uma camada, com aprendizado auto-associativo, para criar codificações complexas, utilizando acoplamento hierárquico das camadas treinadas afim de obter uma melhor performance do algoritmo de treinamento com relação à escalabilidade.

5.1.3 Redes neurais auto-codificadoras com regularização

Uma rede auto-codificadora regularizada, com a penalidade da norma Frobenius quadrática da matriz jacobiana de derivadas parciais associadas a função do codificador é apresentado em [149]. Essa abordagem, chamada de Contractive Autoencoder, tem conexões teóricas com os autoencoders de aprendizado múltiplo e modelagem probabilística e permite encontrar uma representação que captura melhor as direções locais de variação ditadas pelos dados, correspondendo a uma variedade não-linear de menor dimensão, sendo mais invariável para a grande maioria das direções ortogonais à variedade. Tais direções encontradas supostamente se aproximam dos planos tangentes à variedade. Experimentos utilizando os datasets de CIFAR e MNIST mostraram que o CAE aplicado a imagens deve aprender vetores tangentes que mostram como a imagem muda a medida que os objetos na imagem mudam gradualmente.

Alain e Bengio [150] mostraram que no limite de ruído gaussiano na entrada, o erro de reconstrução do Denoising Autoencoder é equivalente a uma penalidade de contração na função de reconstrução. O DAE faz com que a função de reconstrução resista a perturbações pequenas, mas de tamanho finito, da entrada, e apenas indiretamente encoraja resiliência ao ruído, pois o requerimento de invariância é compartilhado entre encoder e decoder, enquanto o CAE faz com que a função de extração de características resista a perturbações infinitesimais da entrada, o que é melhor em aprendizado de representação para tarefas de discriminação, pois é o encoder que será usado para classificação.

No caso de um autoencoder linear, a norma Frobenius quadrática da jacobiana corresponde a um weight decay, e, quando a penalidade do CAE é aplicada a unidades sigmoidais, uma maneira de encolher a jacobiana é fazer com que as unidades saturam para 0 ou 1, o que encoraja o CAE a codificar valores extremos, como um código binário, tornando-o parecido com o autoencoder esparso [150].

5.1.4 Redes neurais auto-codificadoras profundas

O problema associado ao treinamento de redes profundas, devido ao processo de otimização ser difícil, ao problema do desaparecimento do gradiente e à demora ou inviabilidade do treinamento foi superado em 2006 com a idéia do pré-treino camada a camada e empilhamento de autoencoders, com uma etapa não-supervisionada e uma etapa de ajuste fino dos pesos, de acordo com a tarefa final, apresentada por Hinton e Salakhutdinov [151]. A partir de então, a área de redes neurais profundas apresentou um aumento da atividade de pesquisa em aprendizado profundo e aprendizado de representação com diversas camadas, bem como modelos generativos [152], [153].

Bengio et al [154] faz um estudo utilizando Denoising Autoencoders e Restricted Boltzman Machines empilhadas e mostra que o ganho com o ajuste fino supervisionado, após o treinamento de autoencoders profundos para aprendizado de representação, é bem expressivo. A hipótese que esta pesquisa sugere é que, atualmente, os algoritmos de aprendizagem não supervisionada para arquiteturas profundas ainda não são capazes de separar suficientemente bem os fatores explicativos subjacentes de variação.

Em [151], um modelo composto por uma pilha de RBMs foi treinado em imagens de faces humanas e depois seus pesos foram usados para inicializar um autoencoder profundo com camadas ocultas gradualmente menores, culminando em um gargalo de 30 unidades. O código resultante produziu menos erros de reconstrução do que o PCA em 30 dimensões, e a representação aprendida foi qualitativamente mais fácil de interpretar e relacionar com as categorias subjacentes, com essas categorias se manifestando como clusters bem separados. No mesmo trabalho, no contexto de recuperação da informação, 10 componentes extraídas com autoencoder profundo obtiveram melhores resultados que 50 componentes extraídas com um método linear Latent Semantic Analysis. Quando PCA foi utilizada para reduzir a dimensionalidade de LSA para 2-D afim de comparar com a visualização de um autoencoder profundo em 2D, a separação das classes ficou, também, visualmente mais clara com a representação obtida pelo Autoencoder.

A Figura 5.3 mostra a arquitetura e o processo de treinamento de redes-autocodificadoras profundas. Primeiramente a rede é treinada camada a camada (Figura *a*). De forma opcional, na etapa de ajuste fino todas as camadas que foram treinadas separadamente são empilhadas para um novo treinamento (Figura *b*). Quando a tarefa em questão trata-se de classificação, as camadas de codificação, que foram treinadas separadamente, podem ser empilhadas e treinadas em conjunto com um classificador (Figura *c*).

Goodfellow e at [144] pontuam que técnicas de aprendizado profundo baseadas

em aprendizado supervisionado, regularizadas com Dropout [155] ou Batch Normalization [156], são capazes de atingir desempenho de nível humano em muitas tarefas, mas apenas com conjuntos de dados rotulados extremamente grandes. Essas mesmas técnicas superam o pré-treinamento não supervisionado em conjuntos de dados de tamanho médio. Em conjuntos de dados extremamente pequenos, os métodos Bayesianos superam os métodos baseados em pré-treinamento não supervisionado. Por estas razões, a popularidade de pré-treinamento não supervisionado diminuiu. No entanto, o pré-treinamento não supervisionado continua a influenciar as abordagens contemporâneas a medida que novas aplicações e estratégias são apresentadas para explorar uma codificação latente.

5.1.5 Redes neurais auto-codificadoras discriminantes

As redes neurais auto-codificadoras discriminantes propostas em [157] tentam aumentar a compactação intra-classe e a separação inter-classe, ao mesmo tempo que apresentam uma codificação latente em dimensão reduzida. Para atingir este objetivo, a rede auto-codificadora reconstrói, no lugar da própria entrada, uma transformação da entrada mais discriminante, de forma que as amostras estejam mais próximas de outras amostras da mesma classe e mais distantes das amostras das outras classes, através de um processo iterativo de mudança de alvo. Desta forma, este modelo minimiza a função:

$$l = \|\hat{x} - \tilde{x}^{(t)}\|_2^2 \quad (5.1)$$

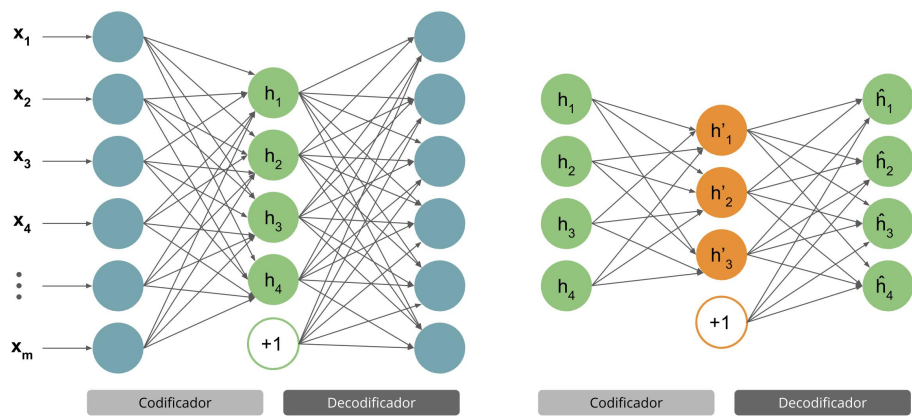
onde \hat{x} representa o vetor de entradas e $\tilde{x}^{(t)}$ o vetor de entradas após t iterações de transformação da entrada. No lugar das entradas, o novo alvo é transformado de maneira que as amostras se movam na direção dos centros de classe, por um valor pequeno de α :

$$\tilde{x}_i^{(t+1)} = (1 - \alpha)\tilde{x}_i^{(t)} + \alpha\left(\frac{1}{|C_i|} \sum_{\tilde{x}_j^{(t)} \in C_i} \tilde{x}_j^{(t)}\right) \quad (5.2)$$

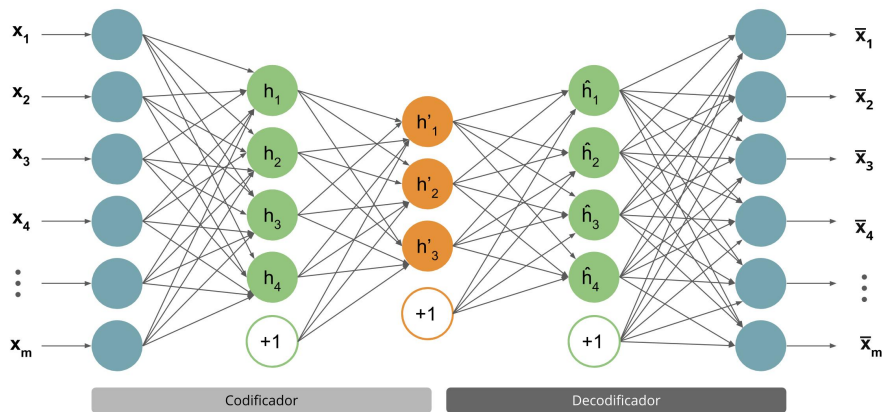
onde α é um parâmetro que permite ajustar a magnitude do deslocamento e C_i o conjunto de amostras pertencentes a class i . De forma similar, as distâncias entre as amostras e centros de outras classes podem ser intensificadas afastando cada amostra da média das amostras pertencentes a outras classes:

$$\tilde{x}_i^{(t+1)} = (1 + \alpha)\tilde{x}_i^{(t)} - \alpha\left(\frac{1}{N - |C_i|} \sum_{\tilde{x}_j^{(t)} \notin C_i} \tilde{x}_j^{(t)}\right) \quad (5.3)$$

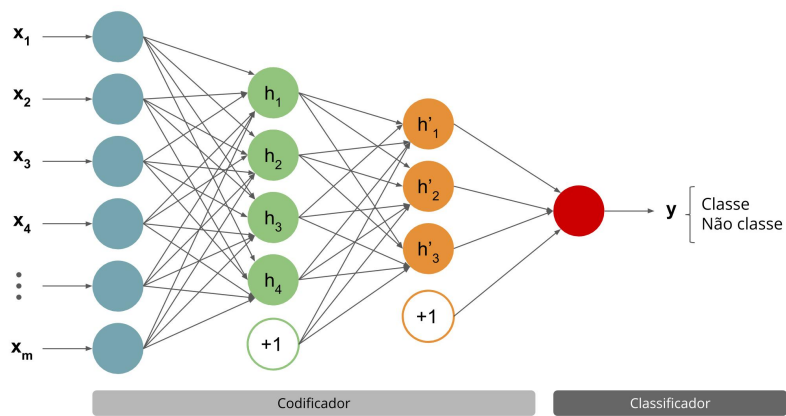
onde N representa o total de amostras.



(a)



(b)



(c)

Figura 5.3: Diagrama ilustrativo da arquitetura e processo de treinamento de redes auto-codificadoras empilhadas (Stacked Autoencoders).

5.1.6 Codificação com Redes Neurais Recorrentes

Sutskever et al em 2014 [158] apresentou um modelo, chamado de Seq2Seq, baseado em redes Long Short-Term Memory (LSTM), para resolver problemas relacionados ao mapeamento de sequências para sequências com tamanho de vetores de entrada e saída indeterminados. A rede LSTM recebe uma sequência de entrada e aprende uma representação codificada de dimensão fixa para, então, outra rede LSTM extrair a sequência de saída desta representação. Os autores perceberam que ao reverter a ordem das sequências de entrada, mas não da decodificação na saída, pode-se obter uma melhoria na eficiência do modelo. Eles atribuíram tal ganho ao fato de que o modelo poderia encontrar uma codificação com um maior número de dependências de curto prazo, já que o valor da sequência decodificada estará mais próximo do codificado, facilitando o processo de otimização. O método foi aplicado em uma tarefa de tradução de inglês para francês, obtendo resultados próximos ao melhor método para esta tarefa, mesmo sem ter investido numa otimização dos hiperparâmetros da rede. Outro fato interessante foi a capacidade que a rede apresentou em traduzir frases bem longas.

Em [159], Srivastava et al aplicou e estendeu o framework Seq2Seq para o aprendizado de representações de sequências de vídeos. Uma rede LSTM, então, codifica a sequência de entrada e uma ou múltiplas redes LSTM decodificam a representação afim de reconstruir a sequência de entrada ou prever a sequência futura. Uma tentativa de visualização e interpretação das características aprendidas foi feita e o modelo que obteve os melhores resultados utilizou um autoencoder combinado com um previsor de quadros futuros na tarefa de reconhecimento de ação.

Em [160], foi feito um trabalho no contexto de sumarização abstrativa de textos utilizando modelos Seq2Seq. Neste trabalho, desenvolveu-se a arquitetura Pointer-Generator Networks, que se baseia em um sumarizador abstrativo por redes recorrentes, porém, com características de modelos extrativos para superar a fraqueza de modelos tradicionais para sumarização abstrativas com relação a palavras raras ou fora do vocabulário, que normalmente levam a erros factuais, além do problema das repetições constantes das partes mais importantes do documento recorrente quando modelos de Redes Neurais Recorrentes (RNN) são utilizados. Os autores adicionaram duas alterações ao modelo de RNN para contornar as dificuldades apresentadas na sumarização abstrativa: um esquema de ponteiros a fim de melhorar a performance de palavras raras ou palavras fora do vocabulário OOV e um mecanismo de Cobertura que reduz o risco do modelo repetir partes do texto.

Em outro trabalho, [161] propõe uma maneira alternativa de calcular as distâncias entre duas sequências de fonemas usando métodos de aprendizagem profunda não supervisionados (LSTM Autoencoders). O modelo é treinado para prever

cada pronúncia nos dados usando a própria pronúncia como o único preditor. Como a representação interna do autoencoder é limitada, ele é forçado a aprender representações compactas de palavras que são úteis para a reconstrução da entrada. Estas representações resultantes são vetores multidimensionais em um espaço onde pronúncias similares encontram-se próximas umas das outras. O modelo é diferente do modelo em [158], pois o modelo Seq2Seq fornece a representação oculta da sequência de entrada para prever o primeiro símbolo em uma sequência de destino e, em seguida, usa o símbolo de destino previsto como uma entrada para a camada LSTM para prever o símbolo atual. Neste artigo as sequências de entrada e de saída são idênticas, possuindo o mesmo tamanho.

No contexto de reconhecimento de áudio, [162] efetuou o treinamento de redes RNN Autoencoders em dados provenientes de sequências acústicas utilizando espectrogramas. A sequência de entrada é alimentada a um codificador composto de RNNs de múltiplas camadas. O estado oculto final do codificador RNN é então passado através de uma camada totalmente conectada, cuja saída é usada para inicializar o estado oculto do decodificador de múltiplas camadas RNN, que, por sua vez, efetua a reconstrução da sequência de entrada. A rede é treinada para minimizar o erro quadrático médio da raiz quadrada entre a sequência de entrada e a reconstrução e quando o treinamento é concluído, as ativações da camada totalmente conectada são usadas como a representação de uma sequência de entrada, permitindo, assim, que a dinâmica temporal dos audios seja aprendida e codificada. Em seguida, as representações são utilizadas na tarefa de classificação de áudio, obtendo resultados competitivos com o estado da arte nesta tarefa.

Uma avaliação abrangente do estado da arte em detecção de novidades acústicas é feita em [163], que apresenta uma breve descrição dos principais métodos utilizados, com abordagens estatísticas ou baseadas em redes neurais. O artigo foca em abordagens utilizando RNN baseadas em Autoencoders. Os métodos avaliados são Gaussian Mixture Models, Hidden Markov Models, One-Class Support Vector Machines, MLP, LSTM e BLSTM, Autoencoder básico, comprimido, Denoising e Non-linear Predictive Autoencoder. As características espectrais auditivas (ASF) são calculadas aplicando a Transformada de Fourier de Curto Prazo (STFT) utilizando um tamanho de quadro de 30 ms e uma etapa de quadro de 10 ms. Cada STFT produz o espectrograma de potência que é convertido para a escala Mel. As características são extraídas do sinal de entrada e o erro de reconstrução entre as entradas e as características reconstruídas é então processado por um bloco de limiar que detecta se o evento é uma novidade ou não.

Um método de treinamento baseado no erro de reconstrução (RE) no contexto de reconhecimento contínuo de emoção em fala, utilizando BLSTMs é proposto em [164]. Presume-se que, ao treinar o modelo como um autoencoder, o erro de

reconstrução indicaria para quais entradas há maior dificuldade no treinamento. As características referentes à habilidade do modelo de codificar a informação de emoções são acrescidas à entrada do modelo para que este aumente a atenção nas regiões de sensibilidade ao erro, abordagem inspirada pelo mecanismo de atenção em NLP. Após experimentos de previsão de emoção (excitação e valência) com o modelo proposto, os autores notaram que as características do erro de reconstrução têm relação com a melhoria de desempenho e que o segundo modelo testado BLSTM-RNN presta mais atenção às regiões de entrada que possuem altos valores de erro de reconstrução. Sendo assim, o modelo possui capacidade de aprender com suas fragilidades.

5.2 Uso de Aprendizado Profundo na Física de Altas Energias

Os dados coletados por experimentos de física de altas energia (High Energy Physics, HEP) são complexos e de alta dimensionalidade. Técnicas tradicionais de análise de dados em HEP usam uma sequência de decisões seguidas de análise estatística nos dados selecionados. Durante as últimas décadas, os físicos de partículas têm procurado melhorar o poder de suas análises empregando técnicas multivariadas (MVA), que podem ser consideradas exemplos de aprendizado de máquina. Para identificação de elétrons em análise offline, por exemplo, originalmente, os experimentos empregavam cortes sequenciais em grandezas físicas de calorimetria para obtenção da decisão final. Entretanto, devido à crescente dificuldade de seleção de eventos, imersos em grande quantidade de ruído, os físicos passaram a se apoiar em abordagens multivariável Preferencialmente, técnicas de árvores de decisão (BDT), implementadas no pacote de software TMVA, e redes neurais têm sido utilizadas. Essas ferramentas proporcionaram um impulso importante para muitas tarefas de análise de dados, mas suas capacidades foram entendidas como limitadas, especialmente quando a dimensionalidade dos dados aumentava [165]. Após o surgimento da aprendizagem profundo, houve uma grande quantidade de trabalhos publicados em físicas de altas energias, tanto do ponto de vista de discriminação como de aprendizado de representação utilizando tais técnicas, especialmente nas áreas de reconstrução de traço, seleção e identificação de objetos e análise física.

5.2.1 Reconstrução de Eventos

O sistema de reconstrução converte os sinais digitais advindos do detector nas propriedades físicas das partículas e, para tanto, conta com as informações provenientes do sistema de calorimetria e do detector de traços, por exemplo. Enquanto

os calorímetros medem a energia incidente das partículas, bem como o perfil da deposição de energia, detectores de traço medem a trajetória de partículas eletricamente carregadas localizando espacialmente a ionização. Detectores de traço com boa precisão fornecem resolução espacial suficiente para permitir a localização da criação de partículas e o ponto de decaimento.

O algoritmo de reconstrução normalmente envolve uma sequência de passos para transformar as medições cruas em objetos de alto nível, como: extração de características, sendo uma exemplo a caracterização de um sinal de uma célula do calorímetro acima de um nível de ruído; reconhecimento de padrões (associação da passagem de um sinal ou grupo de partículas quando um padrão é reconhecido em elementos adjacentes do detector, conhecido como agrupamento em calorimetria); caracterização do objeto (medição das propriedades do objeto, após o ajuste de um padrão de hits no detector de traço e a extração da energia, localização e outras propriedades do agrupamento no calorímetro); e a reconstrução combinada, onde os objetos em ambos os detectores (ou em diferentes detectores) são associados para criar um candidato a partícula [166]. Técnicas de aprendizado de máquina, supostamente, podem ser aplicadas em qualquer um destes passos, como por exemplo o treinamento de algoritmos em variáveis de reconstrução combinada para identificação de partículas.

Técnicas de aprendizado de máquina têm sido úteis em diversos aspectos no que se refere à reconstrução de traços. Por exemplo, no caso em que múltiplos traços passam pelo mesmo agrupamento de pixels, o ATLAS utiliza redes neurais para identificar agrupamentos mesclados, auxiliando o solucionador de ambiguidade e minimizando a perda de eficiência devido às limitações no número de agrupamentos compartilhados por traço, obtendo assim, uma medida para cada traço ao invés de atribuir cada traço ao centro do cluster [167], [168]. O LHCb também utiliza diversas redes neurais para reconstrução de traço para solucionar o problema do aumento do ruído por conta de hits conectados de forma errada, causados pela longa distância entre os sensores no detector de traço do LHCb [169]. Sendo assim, a etapa de reconstrução de traços é eficiente e traços reconstruídos de forma errada são raros. Vale ressaltar, entretanto, que o reconhecimento de padrões para o detector de traço é o passo mais computacionalmente custoso entre todas as tarefas de reconstrução de baixo nível. Na verdade, este passo se torna computacionalmente intratável para o HL-LHC. A esperança é que técnicas de aprendizado de máquina ajudem a superar tal desafio, proporcionando soluções que escalam linearmente com a densidade de colisão do LHC. Um esforço em andamento é o HEP.TrkX que investiga algoritmos de aprendizado profundo, como redes LSTM para reconhecimento de padrão de traço em ambientes online e offline [170].

5.2.2 Identificação e classificação de Jatos

Aprendizado de máquina tem sido aplicado a uma ampla gama de problemas de classificação de jatos, para identificação de jatos de quarks pesados, como quarks c, b e t , ou de quarks leves, glúons e bósons W, Z e H . Esses problemas de classificação, normalmente, são agrupados, mas não limitados a isso, em identificação de sabor (flavor tagging), que discrimina entre b, c e quarks leves; análise das propriedades, incluindo a subestrutura do jato (jet substructure tagging), que discrimina entre jatos de bóson W, Z , quark t e bóson H ; e identificação de quarks e glúons (quark-gluon tagging). [165].

Na identificação de sabor, a informação discriminante é espacial e se baseia nos traços reconstruídos por sensores de alta granularidade próximos ao ponto de interação e nos vértices correspondentes. A análise da subestrutura do jato e a identificação quark-gluon dependem da informação criada em um local espacial durante o decaimento da partícula original e, tradicionalmente, se baseiam em variáveis discriminativas projetadas por físicos a fim de explorar a subestrutura de jatos [171]. A utilização de técnicas de aprendizado de máquina para identificação de jatos ocorre pelo menos desde a década de 1990, quando a biblioteca JETNET foi implementada proporcionando modelos como redes neurais artificiais (MLP) e Mapas auto-organizáveis (SOM) [172], [173].

Em [174] (2015), a deposição de energia de partículas, proveniente da torre de calorímetros, presentes em quase todos os modernos detectores em físicas de altas energias, foi reconhecida como similar aos pixels de uma imagem e, portanto, uma abordagem inspirada em técnicas de visão computacional, como reconhecimento facial em imagens, foi introduzida como método de pré-processamento para identificação de classificação de jatos a partir de sua codificação como imagem. Um discriminante para classificar as imagens dos jatos, derivado do discriminador de Fisher, foi utilizado em dados provenientes de simulação de Monte Carlo, no contexto da identificação de decaimentos do bóson W com relação ao ruído de jatos de quark e gluon. Esta proposta de representação dos dados permitiu que os físicos aproveitassem os avanços em classificação de imagens como redes neurais convolucionais (CNN), que vêm apresentando resultados impressionantes em diversas áreas. De fato, esta nova representação apresentou, em alguns casos testados, resultados melhores ou iguais aos de redes neurais utilizando as informações construídas com conhecimento especialista. Em [175], algoritmos de aprendizado profundo, como as redes ConvNet e MaxOut, foram treinados em tais imagens de jatos e foram desenvolvidas técnicas para visualização e quantificação de como as redes aprendiam as características a partir de imagens e quais informações adicionais contribuíam para a melhoria da performance na identificação dos decaimentos do bóson W imersos

em uma grande quantidade de ruído proveniente dos processos da cromodinâmica quântica (QCD), além de conectar tais visualizações com as propriedades físicas dos processos, de maneira, inclusive, a contribuir para elaboração futura de outras grandezas físicas.

Em um estudo utilizando dados gerados com MADGRAPH5 [176] para simular colisões e decaimentos imediatos, PYTHIA [177] para simular o desenvolvimento do chuveiro e o processo de hadronização e DELPHES [178] para simular a resposta dos detectores, a abordagem utilizando o depósito de energia nos calorímetros como pixels de uma imagem como entrada para uma rede neural profunda, com camadas totalmente conectadas empilhadas no topo de camadas localmente conectadas, foi comparada com a abordagem utilizando BDT treinada em 6 variáveis físicas tradicionais projetadas por físicos. Os resultados mostraram um ganho modesto na tarefa de identificação de jatos, com a utilização das redes profundas, bem como uma certa resiliência, de tais redes, ao empilhamento de sinais.

Apesar do sucesso da abordagem baseada em imagens, a geometria do detector não é perfeitamente regular e, portanto, técnicas de pré-processamento são necessárias para que se possa obter uma representação razoável dos jatos como imagem. Além disso, as imagens são, normalmente, bem esparsas [179]. Neste sentido, algumas tentativas de utilizar Redes Neurais Recorrentes para identificação de jatos foram publicadas recentemente na literatura [180],[181], [182], [183]. Por exemplo, Redes Neurais Recorrente foram propostas no experimento ATLAS com o objetivo de identificar jatos originados de Quarks b, sendo a identificação destes jatos (b-tagging) importante tanto para as medições precisas do Modelo Padrão, incluindo o setor de Higgs, como para explorar novos cenários físicos nas colisões próton-próton. Normalmente, tal identificação, no ATLAS, baseia-se em vários algoritmos de b-tagging de baixo nível que podem ser agrupados em algoritmos baseados em parâmetros de impacto (IP) e baseados em vértices secundários. Como ambas as classes de algoritmos oferecem informações complementares, as saídas destes algoritmos são combinadas em Boosted decision Trees, algoritmo padrão de alto-nível utilizado para análises físicas no ATLAS. Os algoritmos de IP atribuem probabilidades por traço de que o traço se originou de um jato de determinado ‘sabor’ e combinam essas probabilidades como um produto de probabilidade, ignorando quaisquer interdependências entre parâmetros de traço de diferentes traços em um dado jato. Essa simplificação é impulsionada por limitações práticas no algoritmo de verossimilhança. [184] introduz um novo algoritmo de b-tagging de baixo nível (RNNIP), baseado em modelos RNN com uma sequência de traços como variáveis de entrada, com o objetivo de superar esses desafios.

5.2.3 Seleção de eventos e análise física

Nos últimos anos, alguns estudos demonstraram que redes profundas baseadas em entradas com alta dimensionalidade e que recebem menos pré-processamento apresentam, em alguns casos, melhores resultados em seleção de eventos de interesse (discriminação sinal x background) que redes neurais baseadas em variáveis inspiradas na física (variáveis de alto nível). Em [185], o desempenho de redes neurais ‘rasas’ e redes profundas é comparado na tentativa de distinguir decaimentos exóticos do bósons de Higgs do ruído de fundo dominante. Este estudo utilizou um conjunto de dados em que um grande conjunto de variáveis de baixo nível foi reduzido a um conjunto menor de variáveis de alto nível inspirados na física. Isto deu origem a comparações reveladoras sobre o conteúdo relativo de informação das variáveis de baixo e alto nível e o poder dos classificadores para extraí-lo, a medida que as redes profundas que utilizam os dados de nível mais baixo superaram significativamente as redes ‘shallow’ que dependiam de características inspiradas na física, como massas invariantes reconstruídas. Tais conclusões, entretanto, não são universais. Utilizando a mesma abordagem, os autores [185] consideraram uma segunda tarefa de classificação com o objetivo de identificar um processo onde partículas supersimétricas eram produzidas. Para análise de tais partículas, entretanto, houve um esforço significativo na criação de variáveis discriminativas por parte dos físicos e, neste caso, a utilização de redes profundas não obteve vantagens em relação a abordagem, considerando as redes ‘shallow’ com variáveis inspiradas na física. Em [136], aprendizado não supervisionado, através da utilização de autoencoders, é apresentado como um método de busca de nova física. A idéia é utilizar a capacidade do autoencoder na tarefa de detecção de anomalias no contexto de discriminação entre sinal e ruído. Neste caso, diferentes autoencoders foram testados e, à medida que foram treinados para reconstruir apenas as amostras de ruído, compostas de jatos de QCD, as amostras de sinal, compostas de gluino (uma partículas supersimétrica hipotética do gluon), apresentaram um erro de reconstrução relativamente alto. Um autoencoder linear, que obtém resultados parecidos com a PCA, um autoencoder não linear e um autoencoder convolutivo (CNN Autoencoder) são testados e a correlação entre a massa dos jatos com o erro de reconstrução dos autoencoders é examinada. Desta forma, foi mostrado que a distribuição da massa dos jatos se mantém estável a medida que são feitos cortes no erro de reconstrução, com melhor performance apresentada pelo autoencoder convolutivo. Sendo assim, o estudo mostra que seria possível facilitar a descoberta de gluinos de 400GeV utilizando autoencoders, abrindo, então, uma área interessante para exploração de novas físicas sem expectativas prévias ou atreladas à teoria. Outros estudos que seguiram nesta direção apresentaram, também, resultados interessantes [186], [187], [188], [189], [190].

5.2.4 Informação de calorimetria representada pela soma de células em anéis concêntricos

Inspirados em [1], desde 1996, a proposta de informação anelada das células do calorímetro [2] vem sendo apresentada no contexto da separação de elétrons e jatos, através de um ensemble de redes neurais especializadas em regiões do espaço de fase do calorímetro do ATLAS. Tal proposta, então, entrou em operação, em 2017, no segundo nível de seleção de elétrons no detector do ATLAS. A partir de tal representação, diversas tentativas foram feitas na direção de obter uma codificação mais discriminante e que apresente uma redução de dimensionalidade das 100 variáveis referentes aos anéis.

O trabalho [191] utilizou análise de componentes independentes (PCA) ao explorar a segmentação do sistema de calorimetria com o objetivo de reduzir a dimensionalidade da informação de calorimetria no segundo nível do Trigger do ATLAS. Foram propostas duas abordagens de soma de células para montagem dos dados, sendo uma no formato de anéis e outra no formato de serpentina, com a representação em formato de anéis tendo obtido resultados superiores. Em [192], abordagens de pré-processamento como análise de componentes principais (PCA), análise de componentes principais não linear (NLPCA) e análise de componente principais de discriminação (PCD) são aplicados em conjunto com análise de componentes independentes (ICA) nos dados de calorimetria provenientes do segundo nível do trigger do ATLAS, afim de estimar os fatores independentes da informação de calorimetria representada pelos anéis. Tal aplicação foi feita para cada camada do calorímetro e depois tais componentes são concatenadas em um único vetor de características que é dado como entrada para uma rede MLP, obtendo, então, resultados interessantes na discriminação de elétrons imersos em ruído de jatos QCD. Neste estudo, foi mostrado que, para os dados em questão, a maior compactação foi obtida com a utilização de PCA, com 23 componentes, em seguida com PCD, obtendo-se 28 componentes e por último, com NLPCA, obtendo-se 35 componentes, mas os melhores resultados do ponto de vista da discriminação foram obtidos com ICA segmentada em conjunto com PCD. De forma similar, no estudo [193], análise de componentes independentes não linear (NLICA) foi aplicada no mesmo contexto. Outros trabalhos utilizaram a informação anelada de calorimetria e exploraram outras formas de codificação da mesma, tanto em diferentes dados de simulação como de colisão em ambiente online e offline [194], [195], [123], [196], [197], [198].

O presente trabalho se insere neste contexto, sendo uma tentativa de obter uma codificação dos anéis com informação de calorimetria, através da utilização de diferentes variações de Autoencoders, no ambiente offline com o objetivo de discriminar elétrons de jatos.

Capítulo 6

Método Utilizado

Neste capítulo será descrito todo o processo utilizado para o estudo dos modelos de codificação aplicados à informação anelada de calorimetria, na busca por codificadores que consigam apresentar boa qualidade de reconstrução e que possibilitem redução de dimensionalidade ou extração de características latentes para alimentar o algoritmo *NeuralRinger* sem que este apresente perda de capacidade discriminante ou até melhore a capacidade de separação elétron-jato. A análise feita tem como referência o *NeuralRinger* no espaço original de características para $E_t > 15\text{GeV}$ utilizado no ambiente *offline*.

6.1 Seleção de eventos

O experimento ATLAS possui interesses de estudo diversos. O sistema de filtragem *offline* valida a decisão do *trigger* e permite uma filtragem de eventos de acordo com os diferentes objetivos de estudo da física. Normalmente, a colaboração fornece um conjunto de menus¹ que seguem diferentes critérios de aquisição dos dados. No caso da identificação de elétrons, os menus são descritos pelos critérios *Tight*, *Medium*, *Loose* e *Very Loose*, que representam os pontos de operação de todos os testes de hipóteses utilizados na cadeia. Esses critérios se diferenciam pela relação entre a pureza da amostra e quantidade de amostras disponíveis do processo de interesse. Ao passo que o critério *Tight* prioriza a pureza da amostra e apresenta um limiar de corte que proporciona uma probabilidade de detecção do objeto físico de interesse menor, em relação aos outros critérios, o *Very Loose* obtém uma grande quantidade de amostras, com maior contaminação, à medida que opera em um limiar que proporciona alta probabilidade de detecção. Este último critério é utilizado, por

¹Uma assinatura de *Trigger*, também conhecida como cadeia (ou *chain*), é composta por um esquema de algoritmos de extração de características e testes de hipóteses configurados para detectar um objeto de interesse que obedece a certos critérios, como elétrons acima de 28GeV , por exemplo. O conjunto de assinaturas utilizado durante uma rodada de dados é conhecido como menu.

exemplo, na busca de processos que contemplem a física nova. O critério *Medium* apresenta equilíbrio entre a pureza e a estatística do processo.

Os dados utilizados pela colaboração ATLAS são divididos em diferentes regiões do espaço de fase de E_t e η , devido às diferentes características de deposição de energia ao longo do calorímetro. Como pode ser visto na Figura 6.1, os dados do ambiente *Offline*, especificamente, são divididos em 36 regiões ao longo destes espaços.

Sinal / Ruído

η				
8 : 1.97->2.47	9337 63563	42474 36248	81747 8358	107265 5412
7 : 2.01->2.37	37590 251280	178982 143096	367895 32368	520972 20722
6 : 1.81->2.01	21364 138384	106335 78453	231614 18071	354117 11617
5 : 1.52->1.81	31221 253168	162018 139843	376346 30843	610423 19329
4 : 1.37->1.52	14654 160998	83285 88381	209418 20371	351689 12380
3 : 1.15->1.37	23309 206436	136614 104942	354996 23291	576554 15326
2 : 0.80->1.15	37562 290415	229949 162514	659446 37726	1006221 24500
1 : 0.60->0.80	23517 163739	135327 93012	414629 21880	618203 14453
0 : 0.00->0.60	77050 474093	498182 268598	1245005 63981	1949812 42058
	0 : 15->20 [GeV]	1 : 20->30 [GeV]	2 : 30->40 [GeV]	3 : 40->50000 [GeV]

E_T

Figura 6.1: Conjunto de dados para todo o espaço de fase de E_t e η utilizando simulação por Monte Carlo.

As análises da física e o desenvolvimento de modelos para *trigger* podem ser feitos tanto com dados de colisão como em dados de simulação por Monte Carlo [199]. Quando se utiliza dados de colisão, as grandezas físicas medidas pelo detector são afetadas por suas imperfeições, como não-linearidades do detector ou regiões com baixa resolução. Quando os dados utilizados são provenientes de simulação, é possível, então, ter acesso ao evento verdadeiro (*truth*), gerado de acordo com a física presente no Modelo Padrão ou ao evento reconstruído pela eletrônica após interação das partículas com os detectores (*reconstructed*). Um fluxograma com as etapas referentes aos dados de colisão e aos gerados por simulação Monte Carlo encontram-se na Figura 6.2. Quando os dados são gerados por Monte Carlo, normalmente a colaboração utiliza Pythia [200] ou Herwig [201] para geração de eventos de colisão, e Geant4 [202] para simulação da passagem das partículas pela matéria com uso de métodos de Monte Carlo. Em ambos os casos, o gerenciamento de praticamente todos os fluxos de produção do ATLAS, como geração de eventos, simulação e reconstrução de eventos é feito pelo *framework Athena* [203], principal infraestrutura em software do ATLAS. Por último, é comum a utilização da ferramenta ROOT [204] para análise física e modelagem de métodos estatísticos pela colaboração².

²No caso deste trabalho, tal modelagem e análise de dados foi feita utilizando bibliotecas da

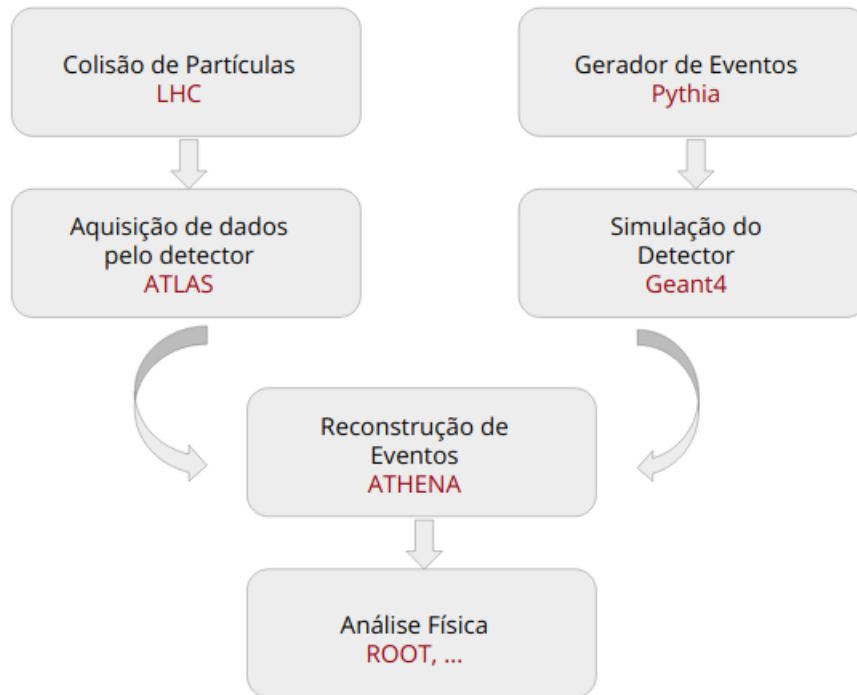


Figura 6.2: Etapas desde a colisão ou geração das partículas até a análise física, considerando dados de colisão (fluxo da esquerda) e simulação por Monte Carlo (fluxo da direita).

6.1.1 Método *Tag and Probe*

O método *Tag and Probe* [132] permite uma seleção de amostras de interesse a partir de conhecimento especialista e é útil para aplicações em dados reais, onde a natureza das partículas observadas é desconhecida. Isto é feito considerando-se processos físicos bem conhecidos que geram os objetos físicos de interesse, como elétrons no caso deste trabalho. A partir do conhecimento destes processos, uma série de critérios são aplicados. Como as análises físicas realizam comparações entre simulações e dados de colisão, as medições de eficiência e ajuste dos modelos em simulações empregam o mesmo procedimento para garantir uniformidade entre ambos.

No caso das análises cujo objeto físico de interesse são elétrons, os decaimentos empregados pela colaboração ATLAS são $Z \rightarrow ee$, $Z \rightarrow ee\gamma$ e $J/\psi \rightarrow ee$. Sendo assim, exige-se que o evento de cruzamento de pacotes contenha ao menos dois candidatos a elétrons reconstruídos pelo *Offline* e procura-se no evento por pares de elétrons que atendam a um critério no qual um dos candidatos (*tag*) seja aceito como elétron, considerando todo o conhecimento da análise *offline*, enquanto o outro candidato *probe* precisa formar, em conjunto com o *tag*, as propriedades conhecidas

linguagem Python, especialmente Keras, Numpy e Scikit-learn.

das partículas. Com o uso deste método, os elétrons utilizados nas análises serão obtidos de forma não tendenciosa e com grande pureza para faixas de energia acima de 30 GeV. Os critérios utilizados pelo *Tag and Probe* neste trabalho, considerando decaimento de Z em dois elétrons e que a massa do bóson Z é de aproximadamente 90 GeV, são descritos abaixo:

- exigência de ao menos dois candidatos a elétrons não oriundos de conversão de fótons;
- o evento contém ao menos um elétron de menor energia sem a aplicação de pré-escala³ obedecendo a uma das restrições abaixo para os diferentes critérios:
 - $E_T > 26$ GeV, quando o critério é *tight*
 - $E_T > 60$ GeV, quando o critério é *medium*
 - $E_T > 140$ GeV, quando o critério é *loose*
 - $E_T > 300$ GeV, quando não há critério de seleção.
- o *tag* não se encontra na região do *crack*⁴;
- o *tag* deve possuir $E_T > 25$ GeV;
- exige-se que o *tag* e o *probe* sejam elétrons de acordo com a verdade de Monte Carlo;
- a massa obtida com o *tag* e o *probe* deve estar dentro da região [80, 100] GeV.

6.1.2 Conjunto de dados utilizado

O conjunto de dados utilizado foi obtido através de simulação, por Monte Carlo, dedicada a representar as condições de colisão $p - p$ a uma energia de 13 TeV no centro de massa referentes ao ano de 2016 (*mc16a*). A seleção de eventos aplica o método *Tag and Probe* (descrito na Seção 6.1.1) para a coleta de elétrons, além da verdade de simulação (*truth*), enquanto somente verdade de simulação foi empregada para ruído. Para seleção de eventos de interesse, optou-se por utilizar dados como amostra de sinal de elétrons que tenham sido aprovados pelo critério *medium* e como ruído de fundo foi exigido que estes tenham sido reprovados no critério *Very Loose*.

³O objetivo da pré-escala é ponderar o processamento em determinadas cadeias de seleção de objetos físicos de interesse. Sendo assim, constitui o número de vezes que uma determinada cadeia precisa ser acionada para que a mesma seja processada. Se uma cadeia possuir pré-escala igual a 0, toda vez que o critério exigido por essa for obedecido esta cadeia será processada. Entretanto, quando o número da pré-escala é alto, o critério deve ser observado esta quantidade de vezes para ser processado.

⁴Região de fissura, onde há uma degradação da resposta do detector, conforme visto na Seção 4.2.3.1. Esta região é definida entre $1,37 \leq -\eta < 1,54$

Neste trabalho são utilizados apenas os decaimentos de Z não radioativos: $Z \rightarrow ee$.

6.2 Validação cruzada

A utilização de métodos estatísticos como os empregados neste trabalho torna necessária a avaliação de como os resultados da análise irão generalizar em um conjunto de dados independentes [205]. Com o objetivo de levar em consideração a incerteza inerente aos dados utilizados, foi empregado um método de validação cruzada. Normalmente, na validação cruzada o conjunto de dados é dividido em conjunto de treinamento, que é utilizado para seleção dos parâmetros do modelo e outro de teste (validação) utilizado para análise da capacidade de generalização. Dentre as diferentes formas de se realizar o particionamento dos subconjuntos, neste trabalho, o método utilizado foi o jackknife [206], onde a partir do conjunto de dados são geradas N partições, onde N-1 são usadas no treinamento e 1 como validação/teste. Cada modelo avaliado é, então, treinado para todas as N diferentes combinações de amostras em treino e teste. No caso deste trabalho, foram utilizadas 10 partições e as diferentes combinações foram feitas conforme a Figura 6.3.

TREINO 1	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 2	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 3	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 4	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 5	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 6	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 7	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 8	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 9	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10
TREINO 10	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5	Parte 6	Parte 7	Parte 8	Parte 9	Parte 10

Legenda: Teste Treino

Figura 6.3: Diagrama ilustrativo do processo de validação cruzada por jackknife.

6.3 Normalização

Como forma de preservar o perfil energético do chuveiro em frações da energia total, foi proposto em [2] uma normalização absoluta usando a energia da RoI total. Sendo assim, cada um dos 100 anéis de energia é dividido pela soma total, conforme a equação 6.1:

$$r'_i = \frac{r_i}{|\sum_{j=1}^N r_j|} \quad \forall i = 1, 2, 3 \dots 100 \quad (6.1)$$

onde N é o número de anéis (100) e r_i é o valor de energia do anel i .

O termo absoluto é usado para evitar refletir os valores ao longo do eixo, devido ao acúmulo negativo de ruído, um comportamento que impactaria a representação física dos valores normalizados e exigiria um limite de decisão mais complexo [195]. Um estudo com dados de simulação mostrou que essa estratégia tinha eficiência compatível com outros esquemas de normalização e seu uso foi preferido devido à sua simplicidade, uma vez que trata-se de uma abordagem não paramétrica, e por permitir fácil interpretação do perfil do chuveiro. Entretanto, esta normalização está sujeita a diminuir as contribuições de sinal em caso de contaminação rigorosa por empilhamento. Outro ponto sensível é que as contribuições de energia dos anéis hadrônicos externos podem dominar e deteriorar o perfil do sinal.

6.4 Análise dos Modelos

A utilização de codificadores no contexto do sistema de filtragem do experimento ATLAS tem como objetivo responder duas perguntas. Os testes referentes à primeira pergunta tentam responder se um codificador que faz uso de transformações não lineares consegue reconstruir bem os dados utilizados no experimento. A segunda diz respeito à capacidade de um codificador de reduzir a dimensão do espaço original mantendo alta eficiência de detecção de sinal, sem aumentar o falso alarme ou de encontrar uma representação que contribua para um resultado superior de discriminação sinal-ruído. Como a qualidade da reconstrução de um codificador não implica, necessariamente, em uma melhoria no desempenho da classificação, os procedimentos referentes à análise de reconstrução e da contribuição da codificação para a tarefa de discriminação serão feitos de formas distintas. Nesta seção será descrito o processo de construção dos *ensembles*, o procedimento de análise de reconstrução, com especificação de figuras de mérito, topologia e treinamento e, por último, o mesmo será feito para a análise na tarefa de classificação.

6.4.1 Construção do *Ensemble*

O algoritmo *NeuralRinger* acima de 15GeV opera, desde 2017, em uma das etapas do sistema de filtragem online (Trigger) de elétrons do ATLAS, em sua camada de alto nível (High Level Trigger, HLT). Este algoritmo utiliza um conjunto de classificadores neurais, onde cada rede deste conjunto é treinada para uma região específica do espaço de fase, o qual é definido em função da energia transversa e da posição da partícula.

Para treinamento destes classificadores, são utilizados os anéis de energia apresentados na Seção 4.3.4. No total são 100 anéis, formados ao longo das 7 camadas dos calorímetros eletromagnético (PS, EM1, EM2, EM3) e hadrônico (HAD1, HAD2 e HAD3). A Tabela 6.1 apresenta a quantidade de anéis definidos para cada uma dessas camadas. No ambiente *online*, o *NeuralRinger* opera em um espaço de fase dividido em 25 regiões do plano $\eta \times E_T$. A Tabela 6.2 apresenta as divisões deste plano para o *ensemble* em operação no ambiente *online*, onde E_T representa a energia transversa e η representa a pseudorapidez [114], que identifica a posição da partícula definida em função do ângulo polar utilizado nas coordenadas cilíndricas do experimento, conforme discutido na Seção 4.2.1. No caso do ambiente *offline*, onde este trabalho se insere, o espaço de fase é dividido em 36 bins (como mencionado na Seção 6.1) utilizados para ajuste do algoritmo Likelihood em operação. Por conta das diferentes análises e do caráter exploratório deste trabalho, apenas alguns bins serão utilizados.

Tabela 6.1: Número de anéis por camada dos calorímetros. Sendo PS o Calorímetro Pré-Amostrador; EM1,2,3 referentes às camadas do Calorímetro Eletromagnético; e HAD1,2,3 referente às camadas do Calorímetro Hadrônico.

PS	EM1	EM2	EM3	HAD1	HAD2	HAD3	Total
8	64	8	8	4	4	4	100

6.4.2 Análise de Reconstrução

A análise das redes neurais auto-codificadoras será feita em comparação com codificadores lineares e não-lineares conhecidos, chamados aqui de baseline. Nesta etapa, a análise será feita para os codificadores PCA e NLPCA, como baseline, e rede auto-codificadora simples (1 camada) e rede auto-codificadora com 3 camadas. Abaixo serão descritos o método de análise, a especificação da topologia treinamento e, por último, as figuras de mérito.

Tabela 6.2: Fronteiras para as regiões de η e ET empregado no ajuste dos modelos para o *NeuralRinger*. Um total de 25 regiões são utilizadas para compor o *ensemble* de redes neurais no ambiente *online*.

Ajuste do Modelo					
Fronteira das Regiões em E_T [GeV]					
15	20	30	40	50	∞
Fronteiras das Regiões em $ \eta $					
$0 \leq \eta < 0.8$					
$0.8 \leq \eta < 1.37$					
$1.37 \leq \eta < 1.54$					
$1.54 \leq \eta < 2.5$					
$ \eta \geq 2.5$					

6.4.2.1 Método de análise

Para que fosse possível estudar o comportamento dos codificadores para compressões cada vez mais intensas, os modelos serão avaliados para diferentes números de dimensões codificadas. Considerando que o espaço original possui 100 dimensões, variou-se a dimensão codificada em intervalo de 10 unidades. Para codificações abaixo de 30, variou-se a dimensão codificada em intervalos de 1 unidade. Com o objetivo de levar em consideração o conhecimento especialista de calorimetria, os codificadores serão analisados em relação à sua capacidade de reconstruir a energia do evento apresentada pelos anéis, considerando tanto a energia total, como a de ambas as seções eletromagnética e hadrônica. Também foram analisadas a capacidade de reconstrução dos anéis individualmente e da energia das camadas longitudinais pelos codificadores, conforme abaixo.

1. Reconstrução para diferentes números de dimensões codificadas.
2. Reconstrução da Energia Total e da Energia dos calorímetros Eletromagnético e Hadrônico.
3. Reconstrução dos anéis.
4. Reconstrução das camadas longitudinais.

6.4.2.2 Especificação da Topologia e Treinamento

O Erro Médio Quadrático (MSE) foi usado como função custo no treinamento das redes auto-codificadores e do modelo NLPCA. Nestes treinamentos, o algoritmo Adam [207] foi utilizado para otimização com taxa de aprendizagem adaptativa, com 1024 amostras como tamanho da batelada. As redes auto-codificadoras foram treinadas com função de ativação tangente hiperbólica na camada intermediária e linear na camada de saída. Um número máximo de 5000 épocas foi utilizado

em todos os treinamentos, com parada antecipada baseada no erro no conjunto de validação (MSE).

Os modelos baseados em redes neurais possuem uma camada de gargalo, cuja saída representa a codificação aprendida e, portanto, a quantidade de neurônios é igual a dimensão da codificação. No caso das redes auto-codificadoras simples (AE) e as redes contrativas (CAE), que possuem apenas uma camada escondida, o número de neurônios desta camada é igual ao número de dimensões codificadas, uma vez que esta é a camada de gargalo. As redes auto-codificadoras profundas e o modelo NLPCA, por outro lado, possuem outras camadas além da camada de gargalo e foram treinadas com um número de neurônios incremental ao longo destas camadas.

A escolha do número de neurônios foi feita através de somas incrementais pequenas, conforme se aumenta o número de camadas, em relação à quantidade de neurônios na camada de gargalo (dimensão de codificação desejada). Essa estratégia foi levada em consideração devido à complexidade de avaliar todas as combinações de neurônios possíveis ao longo das arquiteturas com diferentes números de camadas. Sendo assim, o modelo NLPCA utilizou 30 neurônios a mais nas camadas de mapeamento e de-mapeamento do que na camada de gargalo para todas as codificações, estrutura similar à utilizada nas redes auto-codificadoras profundas com duas camadas⁵. A diferença no treinamento de ambas, portanto, consistiu no fato de que as redes auto-codificadoras foram treinadas camada a camada, com a possibilidade de ajuste fino, como visto na Seção 5.1.4, enquanto o modelo NLPCA foi treinado de uma só vez. As redes auto-codificadoras profundas com 3 camadas apresentaram acréscimo de 20 e 30 neurônios com relação à camada de gargalo.

O apêndice A apresenta alguns detalhes da escolha de hiperparâmetros, bem como as arquiteturas utilizadas pelos diferentes codificadores e suas curvas de treinamento.

6.4.2.3 Figuras de Mérito

As principais medidas utilizadas na avaliação da qualidade de reconstrução dos modelos de codificação estão descritas abaixo.

6.4.2.3.1 Erro de Reconstrução O erro de reconstrução utilizado na avaliação dos modelos consiste no erro médio quadrático, conforme abaixo

$$R_{erro} = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X})^2 \quad (6.2)$$

⁵As redes auto-codificadoras com 2 camadas não serão apresentadas nesta parte análise, mas serão apresentadas quando o impacto da codificação for avaliado.

onde N é o número de anéis (100), X_i é o vetor de entradas e \hat{X} o vetor de saída da rede auto-codificadora, representando as entradas reconstruídas.

6.4.2.3.2 Informação Mútua A informação mútua [208] é uma medida da dependência mútua entre duas variáveis aleatórias e apresenta a incerteza de uma variável a partir do conhecimento de outra variável. Esta medida foi utilizada para a comparação das distribuições das variáveis de entrada com as distribuições da reconstrução das entradas, permitindo avaliar a capacidade de reconstrução dos diferentes codificadores. Considerando x e y como duas variáveis aleatórias, então, a informação mútua entre elas é dada por

$$I(x, y) = H(x) + H(y) - H(x, y) \quad (6.3)$$

onde $H(\cdot)$ é a função entropia e p é uma função densidade de probabilidade.

A informação mútua é uma medida que possui apenas limite inferior, o que dificulta a análise da qualidade da reconstrução entre a distribuição de entrada e a distribuição reconstruída dos anéis. Para fornecer um limite superior para essa medida e facilitar as análises, foi utilizada a informação mútua normalizada [209]. Sendo assim, aplica-se

$$\rho_{x,y}^2 = 1 - e^{-2I(x,y)} \quad (6.4)$$

o que limita o valor da informação mútua na região $[0,1]$.

6.4.2.3.3 Divergência Kullback-Leibler A Divergência de Kullback-Leibler (ou divergência KL) [210] [211] é uma medida do quanto uma distribuição de probabilidade é diferente de uma outra. A Divergência KL não é uma distância por não ser simétrica, mas é sempre não negativa e zero se e somente se as duas distribuições forem iguais. Esta medida também foi utilizada na avaliação da capacidade de reconstrução dos codificadores através da comparação das distribuições de entrada e entrada reconstruída. A Divergência KL entre duas distribuições $p(x)$ e $q(x)$ pode ser definida conforme abaixo

$$D_{KL}(p||q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \quad (6.5)$$

onde $p(x)$ e $q(x)$ são distribuições de probabilidade. No caso deste trabalho, a distribuição $p(x)$ representa a distribuição de energia de um anel, enquanto a distribuição $q(x)$ representa a distribuição da reconstrução da energia deste anel por um dado codificador.

Considerando estas medidas, vale ressaltar que, a normalização utilizada, conforme descrita em 6.3, quando somada para todas as variáveis apresentará valor 1.

Sendo assim, não é possível utilizar as medidas descritas para os valores normalizados para avaliação da energia total, por exemplo. Sendo assim, a avaliação dos modelos foi feita considerando a distribuição dos anéis originais sem a normalização e a distribuição dos anéis reconstruídos denormalizados.

6.4.3 Análise de classificação

A qualidade da reconstrução de um codificador não implica, necessariamente, em uma melhora no desempenho da classificação, mas, em alguns casos, o processo de codificação pode evitar que variáveis que possuem correlação linear ou não-linear sejam utilizadas no processo de treinamento do classificador, evitando redundância e contribuindo para uma melhor performance na tarefa em questão. Em outros casos, o processo de representação pode aprender uma codificação latente, na qual as variáveis encontradas são os fatores que explicam os dados e que, quando utilizados, podem também apresentar uma melhora para o classificador. Abaixo serão descritos o método de análise, a especificação da topologia do treinamento e, por último, as figuras de mérito.

6.4.3.1 Método de análise

Para avaliar o impacto da codificação na capacidade discriminativa, o espaço de entrada será codificado e fornecido ao classificador (rede MLP). Inicialmente, a análise da relevância das variáveis será feita com o objetivo de avaliar os anéis mais relevantes para classificação, a fim de verificar se anéis importantes para classificação, quando mal reconstruídos, podem reduzir a capacidade discriminante. Na sequência, os modelos serão avaliados para diferentes codificações, com o objetivo de verificar como a redução de dimensionalidade afeta a classificação. Por último, o índice SP será apresentado em conjunto com a probabilidade de detecção e o falso alarme para comparação do impacto dos diferentes codificadores.

1. Análise de relevância de variáveis para o Modelo *NeuralRinger* sem codificação.
2. Reconstrução para diferentes números de dimensões codificada.
3. Comparação do índice SP (será definido na Seção), probabilidade de detecção (Pd) e probabilidade de falso alarme (Pf) para uma dimensão fixa.

6.4.3.2 Especificação da Topologia e Treinamento

O classificador utilizado foi uma rede neural MLP, como descrito na Seção 4.3.4. A arquitetura do classificador consistiu em duas camadas com 5 neurônios na camada escondida e 1 neurônio na camada de saída. Tanto a camada escondida como a camada de saída tiveram como função de ativação a tangente hiperbólica.

O treinamento desse modelo foi feito utilizando o Erro Médio Quadrático (MSE) como função custo. Nestes treinamentos, o algoritmo Adam também foi utilizado para otimização com taxa de aprendizagem adaptativa, com 2048 amostras como tamanho da batelada. Um número máximo de 10000 épocas foi utilizado em todos os treinamentos, com parada antecipada baseada no erro no conjunto de validação (MSE).

As redes auto-codificadoras utilizadas nesta etapa foram as mesmas treinadas para a análise de reconstrução, conforme descrito na Seção 6.4.2. Sendo assim, a escolha do número de neurônios foi feita através de somas incrementais pequenas, conforme se aumenta o número de camadas, em relação à quantidade de neurônios na camada de gargalo (dimensão de codificação desejada). Além das redes já mencionadas, as redes profundas com 5 camadas apresentaram acréscimos de 10, 20, 35 e 60 e as redes profundas com 8 camadas apresentaram acréscimos de 5, 25, 35, 45, 55, 65. Após o treinamento destas redes, a projeção na saída da camada de gargalo foi utilizada para obtenção da representação.

No caso das rede auto-codificadora discriminante, definida na Seção 5.1.5, como seu treinamento consiste em reconstruir uma transformação da entrada, em que as amostras se encontram mais próximas das amostras da mesma classe e distantes das amostras da outra classe, estes modelos foram treinados variando o número de iterações em que os dados eram transformados de 1 até 20 iterações. O parâmetro α , que pondera a intensidade da transformação, como descrito na Seção 5.1.5, foi escolhido igual a 0.2 em todas as transformações. Essas redes foram treinadas, inicialmente, com 3 camadas com arquitetura 50,25,1.

O apêndice A apresenta alguns detalhes da escolha de hiperparâmetros, bem como as arquiteturas utilizadas pelos diferentes codificadores e suas curvas de treinamento.

6.4.3.3 Figuras de mérito

O índice Soma-Produto (SP) [212] foi utilizado como figura de mérito para comparação dos modelos. Este índice permite uma avaliação equilibrada entre a probabilidade de detecção (P_D) e o falso alarme (P_F) dos classificadores, conforme equação:

$$SP = \sqrt{\sqrt{P_D \cdot (1 - P_F)} \cdot \frac{P_D + (1 - P_F)}{2}} \quad (6.6)$$

onde P_D é a probabilidade de detecção de elétrons e P_F é a probabilidade de classificar como elétron um jato hadrônico (ruído).

6.4.3.4 Relevância de Variáveis

Uma maneira de se entender a importância das variáveis utilizadas por um classificador é observar como cada uma contribui, de forma isolada, para a composição da saída deste. Desta forma, é possível efetuar descarte de componentes irrelevantes e entender quais variáveis oferecem maior capacidade discriminativa. Isso pode ser feito através de um método de análise de relevância [195]. A relevância R_i da i -ésima entrada fornecida a um classificador é dada por

$$R_i^{MSE} = \frac{1}{N} \sum_{j=1}^N [O(x_j) - O(x_j|x_j = x_{medio})]^2 \quad (6.7)$$

onde N é o número de eventos aplicados ao classificador e $O(x)$ a saída produzida pelo classificador para um determinado vetor de entradas x . Sendo assim, a relevância analisa o erro médio quadrático produzido quando a média da i -ésima entrada é inserida em seu lugar.

Outra maneira de avaliar a relevância de uma dada variável é através do índice SP mencionado acima. Quando este é utilizado, a relevância da i -ésima entrada será dada por

$$R_i^{SP} = SP_{Max}(x) - SP_{Max}(x|x_i = x_{imedio}) \quad (6.8)$$

Estas análises se diferenciam, pois enquanto a equação 6.7 analisa o desvio na saída do classificador, a equação 6.8 observa o impacto da tarefa de classificação.

Capítulo 7

Resultados

Neste capítulo serão analisados os resultados obtidos ao longo do desenvolvimento deste trabalho. Os modelos de codificação foram testados a fim de avaliar se os codificadores não-lineares apresentam vantagens na capacidade de reconstrução de eventos, quando treinados com a informação dos 100 anéis de energia como entrada, e se permitem redução de dimensionalidade sem impactar negativamente na capacidade de discriminação do *NeuralRinger* ou aprender uma representação que contribua para um melhor desempenho na separação sinal-ruído. Como mencionado na Seção 6.4, o processo de avaliação dos modelos de codificação e do impacto da codificação no desempenho do classificador serão apresentados de maneira separada. Sendo assim, inicialmente as características dos dados utilizados serão apresentadas, posteriormente uma análise da capacidade de reconstrução dos codificadores avaliados em relação à referência será feita e, por último, o impacto da codificação na capacidade de discriminação do classificador *NeuralRinger*, quando este utiliza, como entrada, a representação encontrada por um dado codificador, será analisada.

7.1 Dados utilizados

O conjunto de dados analisado foi adquirido através de simulação de Monte Carlo (ano de 2016) referente ao ambiente *Offline*. Conforme mencionado na Seção 6.1.2, os dados utilizados pela colaboração são divididos em diferentes regiões do espaço de fase de E_t e η . Para efeito de análise, o conjunto de dados utilizado se situa na região com energia na faixa de 30 a 40 GeV e de η entre 0 e 0.6, por ser uma região de energia que favorece eventos de interesse e que se situa na região do barril, próximo à posição central do detector. Este conjunto contém um total de 1308986 amostras, sendo 1245005 de sinal e 63981 de ruído de fundo.

A Figura 7.1 compara o perfil médio, considerando o desvio padrão, ao longo dos anéis de energia formados nas camadas do sistema de calorimetria do ATLAS para sinal e o ruído de fundo na região dos dados utilizados. Nota-se que a maior

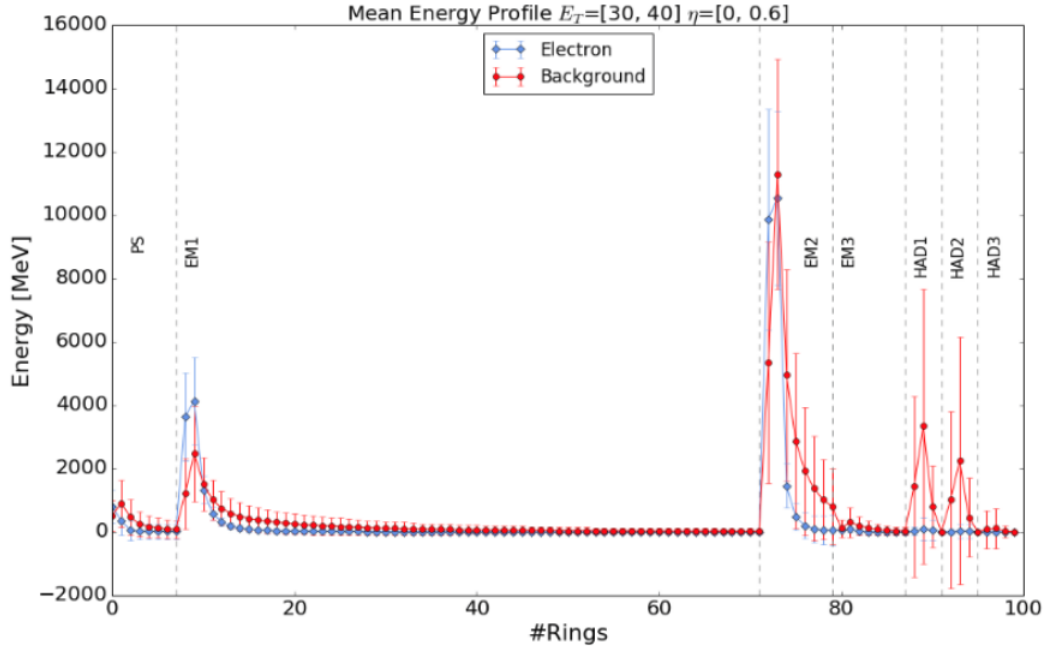


Figura 7.1: Comparação entre os perfis médios, considerando o desvio padrão, ao longo dos anéis de energia formados nas camadas do sistema de calorimetria do ATLAS para sinal e o ruído de fundo na região de $30\text{GeV} < E_t < 40\text{GeV}$, $0.0 < \eta < 0.6$.

quantidade de energia é depositada na camada EM2, especialmente nos anéis 73 e 74 (anéis 2 e 3 desta camada). Há também uma quantidade de energia significativa depositada nos anéis 9 e 10, majoritariamente para a classe sinal, e nos anéis 89, 90, 91, 93, 94, 95, das camadas HAD1 e HAD2 respectivamente, sendo nestes casos o depósito de energia feito para a classe ruído.

Como mencionado na Seção 4.3.4, a representação da informação de calorimetria através de anéis concêntricos tem como objetivo preservar a estrutura geométrica do chuveiro produzido pela interação de partículas com a calorimetria do ATLAS. Como o processo de desenvolvimento do chuveiro possui correlações entre o perfil de deposição de energia ao longo das células, espera-se que tais correlações sejam refletidas nos anéis. A Figura 7.2 apresenta a matriz de correlação entre os anéis, calculada através do coeficiente de correlação de Pearson. Verifica-se a presença de correlação entre anéis de uma mesma camada. Este índice consegue medir apenas a correlação linear. Para analisar a presença de correlações não lineares, a informação mútua normalizada foi utilizada, conforme discutido na Seção 6.4.2.3.2, como pode ser visto na mesma figura. A Figura 7.3 mostra novamente a correlação entre os anéis, porém separada por classe. No caso da classe sinal, existe uma correlação mais intensa entre os primeiros anéis da camada EM1 e também destes com os primeiras da camada EM2. A presença significativa de correlação não linear justifica

a utilização de métodos que busquem encontrar uma representação latente através do uso transformações não lineares.

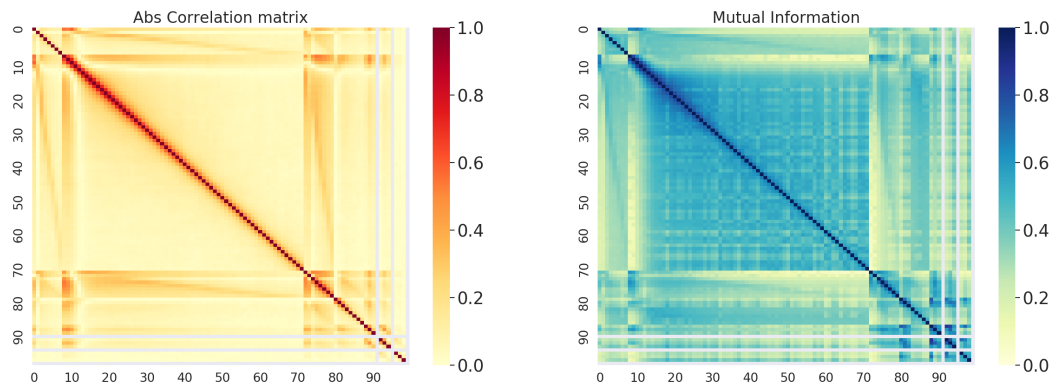


Figura 7.2: Matriz de Correlação de Pearson e de Informação Mútua calculadas entre os anéis para verificação de presença de correlação linear e não linear, considerando todo o conjunto de dados.

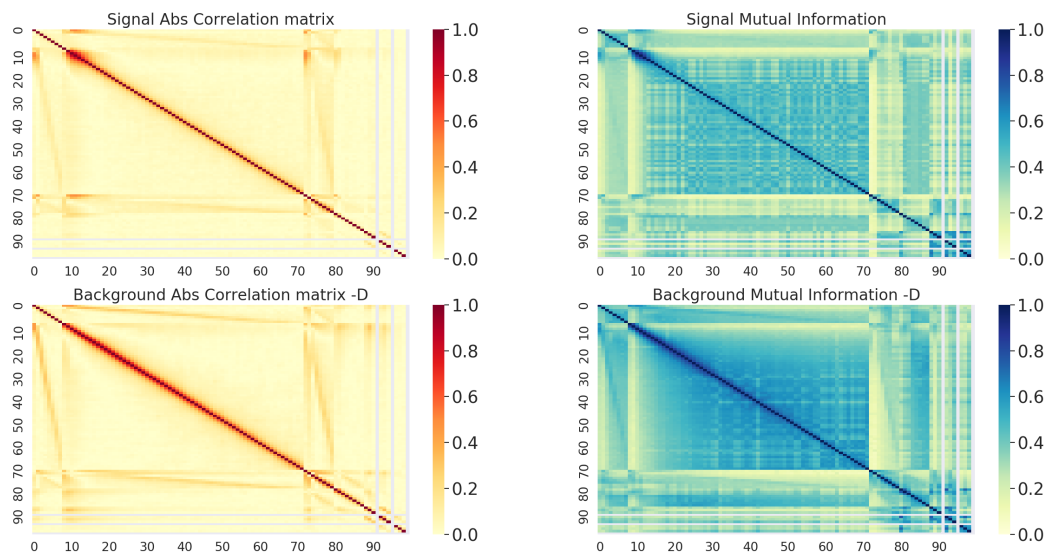


Figura 7.3: Matriz de Correlação e de Informação Mútua calculada entre os anéis para para as classes sinal e ruído.

7.2 Análise de Reconstrução

Como mencionado na Seção 6.4.2, a análise das redes neurais auto-codificadoras será feita em comparação com os codificadores de baseline: PCA e NLPCA. Inici-

almente, o comportamento dos diferentes modelos será analisado conforme se reduz a dimensão codificada. Com o objetivo de levar em consideração o conhecimento especialista de calorimetria, os codificadores serão analisados em relação à sua capacidade de reconstruir a energia do evento apresentada pelos anéis, considerando tanto a energia total, como a de ambas as seções eletromagnética e hadrônica. Por último, a reconstrução dos anéis será avaliada individualmente.

7.2.1 Reconstrução para diferente número de dimensões codificadas

Como a representação da informação de cada dimensão é dada por energia, é possível realizar a soma das mesmas para verificar como as segmentações longitudinais do sistema de calorimetria estão sendo reconstruídas pela rede auto-codificadora. Nesta análise, empregou-se a rede auto-codificadora simples (com apenas uma camada) e, como figura de mérito, para verificar a reconstrução, empregou-se a informação mútua normalizada, conforme discutido em 6.4.2.3.2, onde as distribuições da soma da energia de cada segmentação longitudinal e da soma da energia reconstruída destas segmentações são comparadas por esta medida.

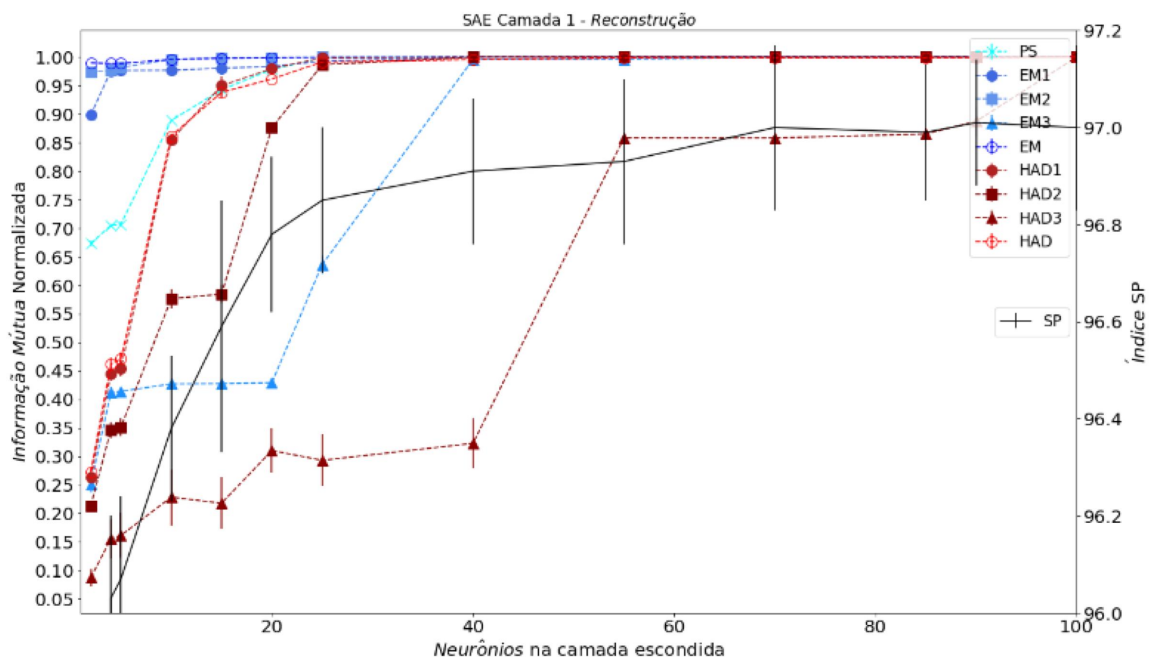


Figura 7.4: Informação mútua entre entrada e reconstrução da rede auto-codificadora para as segmentações longitudinais do Sistema de Calorimetria em função do número de neurônios empregados. No eixo da direita, o valor de SP-max medidos. A dimensão da codificação foi variada de 100 até 1.

Observa-se na figura 7.4 que a seção HAD3 é a primeira a perder capacidade

de reconstrução, conforme a redução do número de neurônios no codificador, o que é um comportamento esperado já que essa camada captura apenas ruído na dada faixa de energia ao se posicionar de, forma longitudinal, no final do calorímetro. Com uma configuração empregando menos de 40 neurônios na camada escondida, a rede auto-codificadora começa a degradar a informação da EM3, última camada da seção eletromagnética e que fornece informação similar à HAD1 neste faixa de energia. Em torno de 25 neurônios, há degradação da informação da seção HAD. Como a rede auto-codificadora é treinada para minimizar o erro médio quadrático, é natural que a seção hadrônica sofra degradação primeiro, uma vez que constitui de uma menor perda na função objetivo que a informação na seção EM, dada a maior concentração de captura da energia na última. A camada PS, que captura apenas eventuais desenvolvimentos de chuveiro prematuros ao calorímetro, também começa a sofrer degradação quando a dimensão codificada é reduzida para este valor. Por conter maior parcela da energia dos eventos, a EM2 é a última camada a sofrer degradação.

Na Figura 7.5, diferentes arquiteturas de redes auto-codificadoras são comparadas em relação à sua capacidade de reconstrução da informação original. Os gráficos da esquerda e direita mostram, respectivamente, o erro médio quadrático e a informação mútua entre os valores de entrada e os valores reconstruídos pelos modelos, treinados camada a camada (*layer-wise pretraining*), com e sem a etapa de ajuste fino. Observando ambas as medidas, ao longo de diferentes dimensões de codificação, nota-se a importância desta etapa para diminuir a perda de informação decorrente da codificação. Sendo assim, a partir desta análise, os modelos de redes auto-codificadoras considerados foram treinados, também, com a etapa de ajuste fino. A Figura 7.6 apresenta a comparação das arquiteturas de auto-codificadores profundos com os demais modelos testados. Verifica-se que os modelos obtêm performance de reconstrução parecidas quando a dimensão codificada é igual ou superior a 3. Abaixo deste valor de dimensão codificada, os modelos que utilizam estatística de ordem superior aparentam ter obtido uma reconstrução melhor.

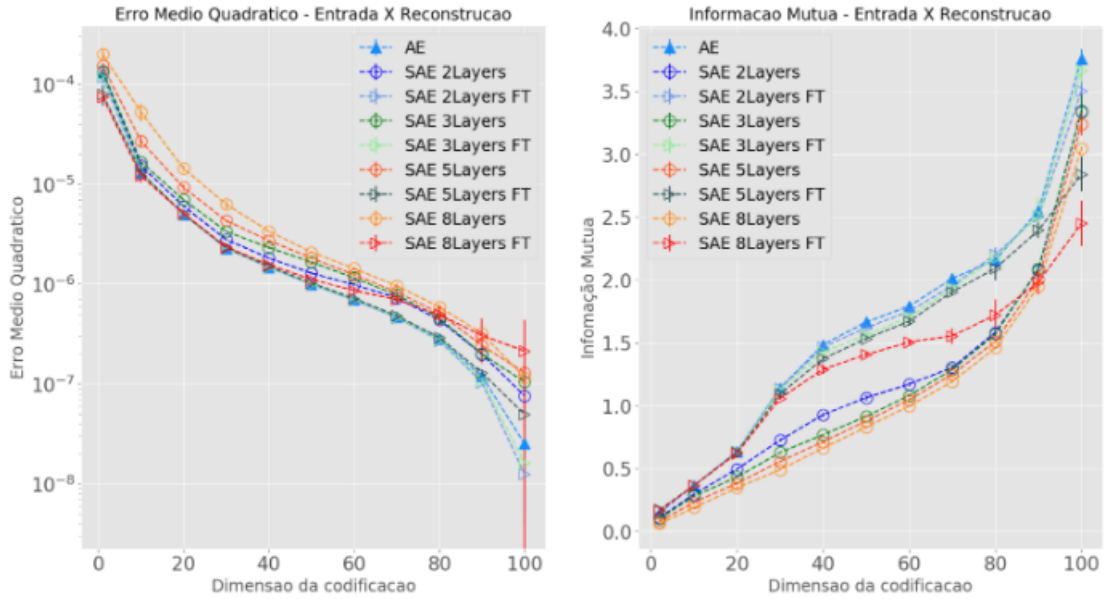


Figura 7.5: Análise da qualidade da reconstrução proporcionada por diferentes modelos ao longo de diferentes dimensões codificadas. Os gráficos da esquerda (Erro médio quadrático) e do centro (informação mútua) comparam o treinamento das redes auto-codificadoras com relação à etapa de ajuste fino.

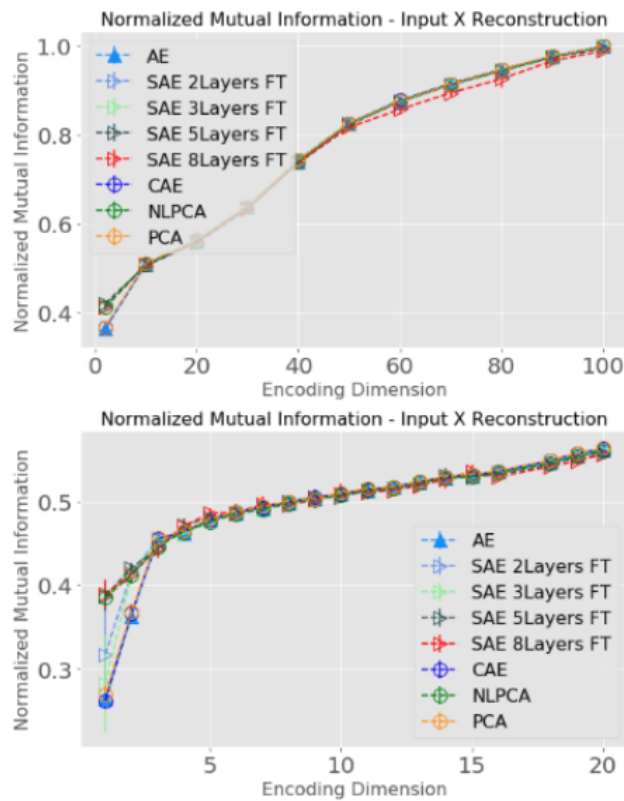
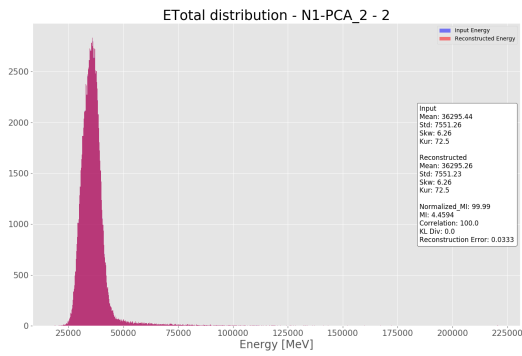


Figura 7.6: Análise da qualidade da reconstrução proporcionada por diferentes modelos ao longo de diferentes dimensões codificadas. o gráfico compara estas redes com outros modelos de codificação.

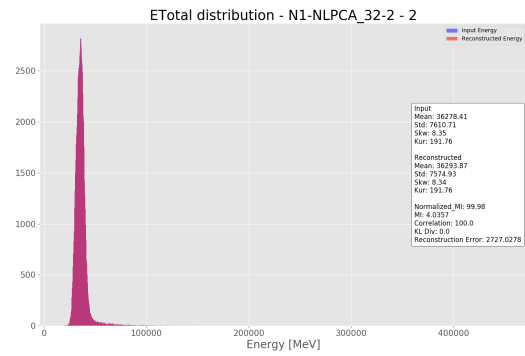
7.2.2 Reconstrução da Energia Total da RoI

Considerando que a diferença entre os modelos testados ao longo de diferentes dimensões codificadas é notada somente quando a dimensionalidade do espaço original é reduzida abaixo de 3, a análise de reconstrução da energia total foi feita para codificação em duas dimensões a fim de verificar o motivo de tal diferença. Sendo assim, a Figura 7.7 mostra os histogramas da energia total da RoI, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a)* PCA, *b)* NLPCA, *c)* rede auto-codificadora simples e *d)* rede auto-codificadora com 3 camadas. A mesma análise é feita para a energia total da seção eletromagnética (Figura 7.8) e seção hadrônica (Figura 7.9). No caso da reconstrução da energia total da RoI, todos os modelos obtiveram cerca de 0.999 para informação mútua normalizada. Entretanto, a PCA obteve um erro de reconstrução menor, o que é um possível efeito do fato desta codificação olhar apenas para os dois primeiros momentos estatísticos (média e variância). No caso da energia total da seção eletromagnética, os modelos NLPCA e rede auto-codificadora com 3 camadas, obtiveram valores de informação mútua parecidos, mas não significativamente maiores que o valor obtido pelas codificações por PCA e pela rede auto-codificadora simples. Entretanto, o erro de reconstrução dos primeiros foi consideravelmente inferior. No caso da seção hadrônica, tanto o erro de reconstrução como a informação mútua normalizada apresentaram valores significativamente melhores para os modelos que acessam estatística de ordem superior.

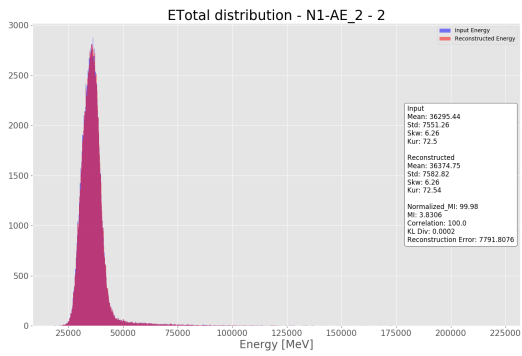
Na Figura 7.10 o histograma da energia total das seções eletromagnética e hadrônica são analisadas para as diferentes classes, considerando apenas a rede auto-codificadora com 3 camadas. No caso da seção eletromagnética, nota-se que o erro de reconstrução da classe sinal foi consideravelmente inferior, enquanto a informação mútua normalizada foi ligeiramente superior. Quando a seção hadrônica é analisada, por outro lado, é possível notar que o histograma da classe sinal tem qualidade de reconstrução muito inferior à classe ruído, o que é indicado também pelas outras medidas analisadas, como a divergência KL. A ocorrência de tal fato era esperada, uma vez que os elétrons, que compõem a classe sinal, praticamente não depositam energia na seção hadrônica do sistema de calorimetria do ATLAS, como pode ser visto no perfil energético das classes mostrado na Seção 7.1.



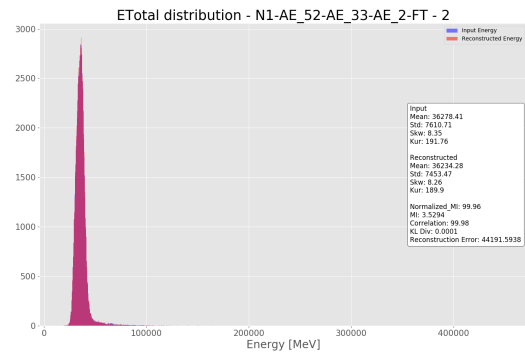
(a)



(b)

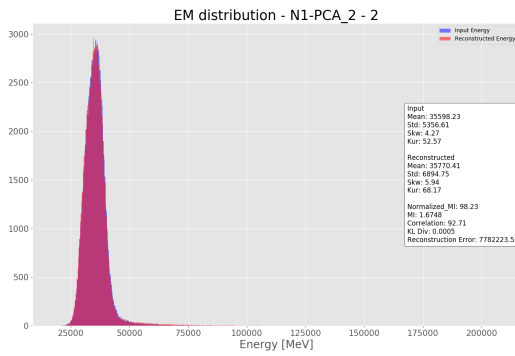


(c)

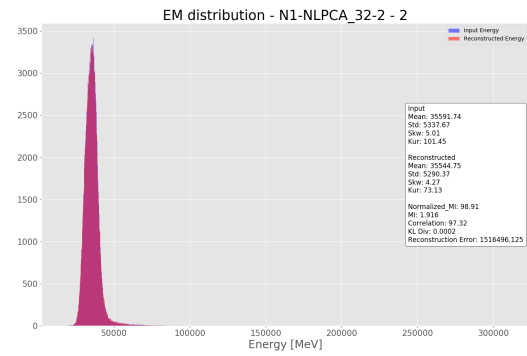


(d)

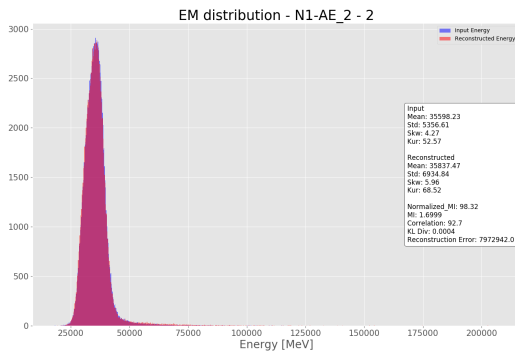
Figura 7.7: Histogramas da energia total da ROI considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.



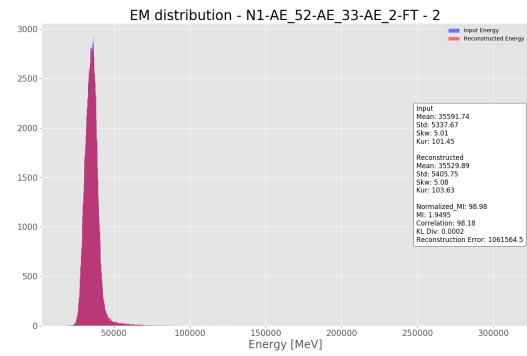
(a)



(b)



(c)



(d)

Figura 7.8: Histogramas da energia total da seção eletromagnética, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.

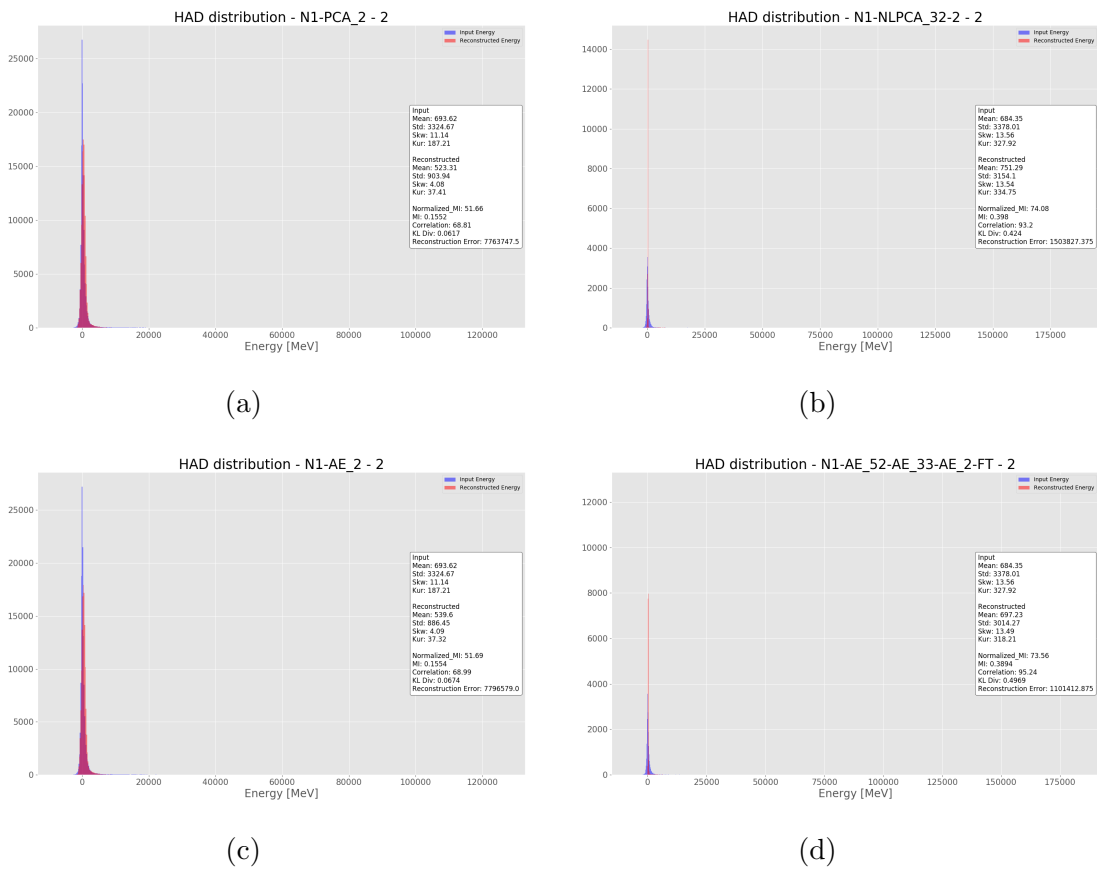


Figura 7.9: Histogramas da energia total da seção hadrônica, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.

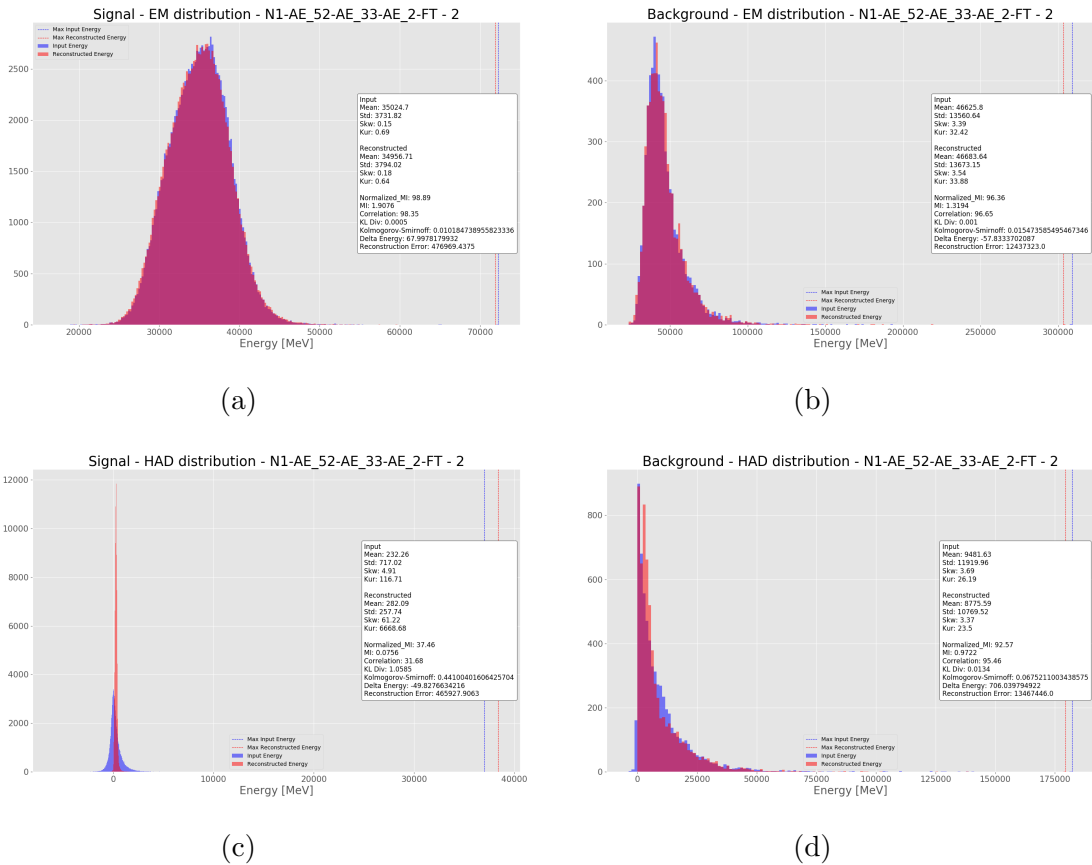


Figura 7.10: Histogramas da energia total da seção eletromagnética, considerando os dados de entrada e os dados reconstruídos pela rede auto-codificadora com 3 camadas, para a classe sinal *a* e para a classe ruído *b*; e da energia total da seção hadrônica, considerando entrada e reconstrução pelo mesmo codificador, para classe sinal *c* e ruído *d*.

7.2.3 Reconstrução dos anéis

Em seguida, uma análise individual dos anéis foi feita a fim de observar quais anéis apresentam melhor qualidade de reconstrução e quais codificadores apresentam melhores resultados nesta tarefa. A Figura 7.11 apresenta a curva de carga e o gráfico Scree da análise de componentes principais para o conjunto de dados utilizado. A partir deste gráfico quatro cortes foram feitos com base variância retida pelos componentes da PCA. Para um corte de 99%, 27 componentes são necessárias. Quando o corte é de 97%, 17 componentes são necessárias. Com 90% no valor de corte, apenas 5 componentes são utilizadas e, por último, para manter 70% de energia são necessários apenas 2 componentes. Vale ressaltar, entretanto, que tais cortes são feitos com base apenas no segundo momento estatístico (variância), o que, considerando a característica não gaussiana das distribuições de energia dos anéis, não torna tais cortes suficientes para avaliação e comparação dos modelos. De qualquer forma, tal análise é utilizada apenas como referência para comparação da qualidade de reconstrução dos anéis, ao longo de diferentes dimensões codificadas.

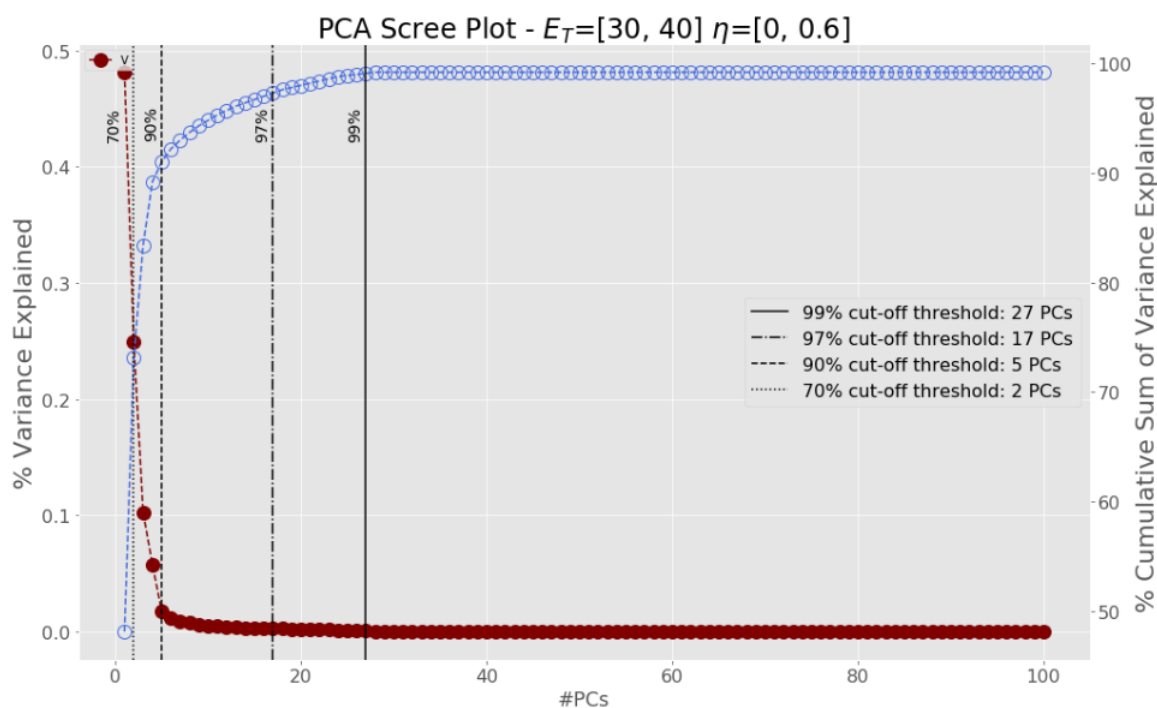


Figura 7.11: Gráfico Scree e Curva de Carga das Componente Principais.

A Figura 7.12 apresenta uma comparação da qualidade de reconstrução dos anéis, com base na informação mútua normalizada, com respeito à capacidade de reconstruir a informação original, para diferentes dimensões ao longo dos 100 anéis, quando os modelos PCA, NLPCA, rede auto-codificadora simples e uma rede auto-codificadoras profunda com 3 camadas são utilizados. Em geral, nota-se que quando

a dimensão de codificação é reduzida para 27, podemos observar uma diminuição significativa na qualidade da reconstrução nos anéis de energia mais baixa: os da camada EM1 acima do anel 20 (40%) e os da EM3 (70%) e HAD3 (60%). Quando reduzida para 17 dimensões, há uma diminuição nos anéis da PS (45%) e da HAD2 (35%). Quando reduzido para 5, a diminuição ocorre nos primeiros anéis do PS (50%) e no último da EM2 (30%). Quando reduzido para 3, há uma ligeira diminuição nos anéis 9, 10 (18%), 75 (22%) e 90 (27%), que são altamente energéticos. Quando reduzido para duas dimensões, há uma diminuição significativa na qualidade de reconstrução dos anéis 9 (45%) e 10 (27%). Analisando individualmente a rede auto-codificadora com 3 camadas, quando a dimensão de codificação é reduzida para 2 dimensões, não há uma diminuição significativa na qualidade de reconstrução dos anéis 9 e 10, o que mostra que a qualidade da reconstrução dos codificadores varia de forma similar ao longo das dimensões codificadas.

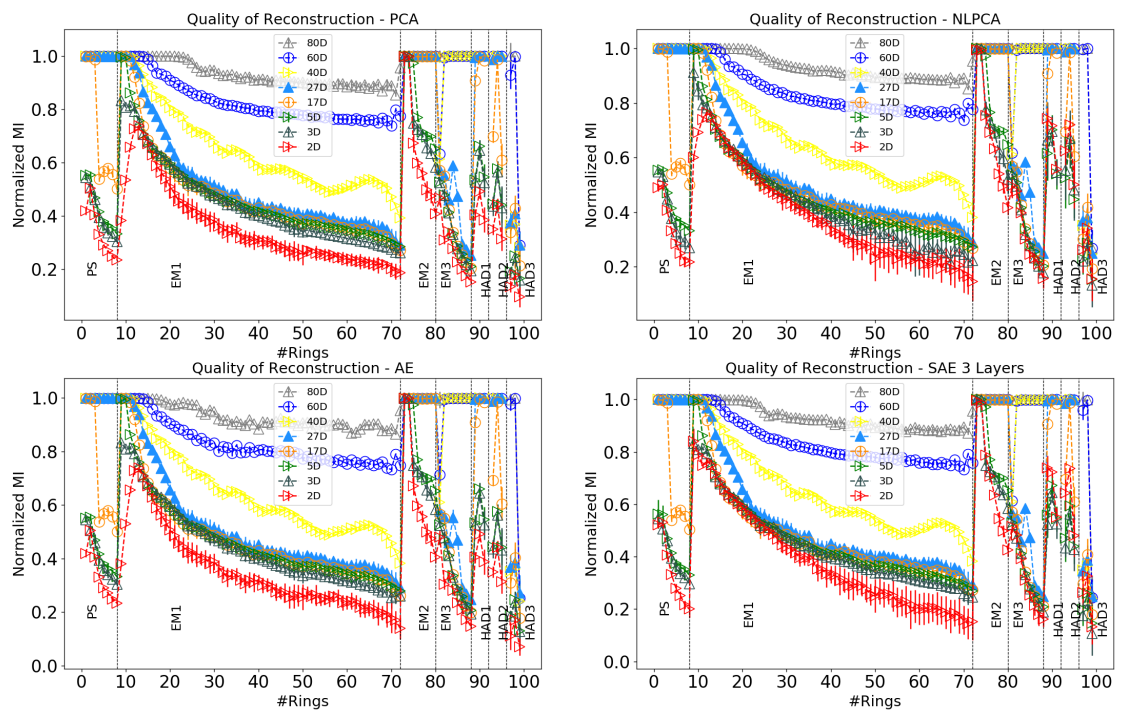


Figura 7.12: Comparação de diferentes dimensões de codificação com respeito à sua capacidade de reconstruir as informações originais, para cada um dos 100 anéis, quando são usadas: PCA, NLPCA, rede auto-codificadora simples e rede auto-codificadora com 3 camadas.

A Figura 7.13 mostra uma comparação entre a qualidade de reconstrução dos dois modelos lineares, a rede auto-codificadora com uma camada e a PCA, e dois modelos não lineares, a NLPCA e a rede auto-codificadora profunda com 3 camadas,

quando a dimensão é reduzida para 2 de forma que possibilite ressaltar as diferenças na reconstrução individual dos anéis para estes codificadores. É possível observar que os modelos de codificação linear têm o mesmo comportamento ao longo dos 100 anéis, enquanto os modelos de codificação não linear também têm comportamento muito semelhante, exceto que o SAE apresenta uma melhor reconstrução para os anéis 9 e 10. Quando comparados, os modelos não lineares apresentam desempenho superior em todas as camadas da seção hadrônica, em alguns anéis da EM2, nos anéis mais energéticos da EM1 e nos primeiros anéis do PS.

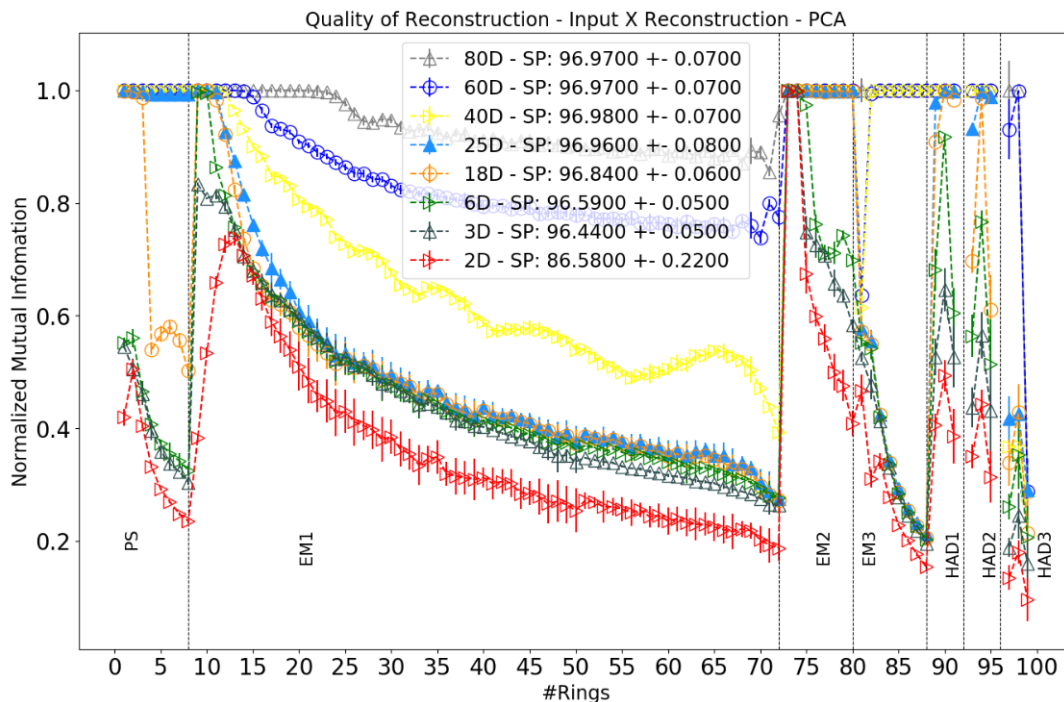


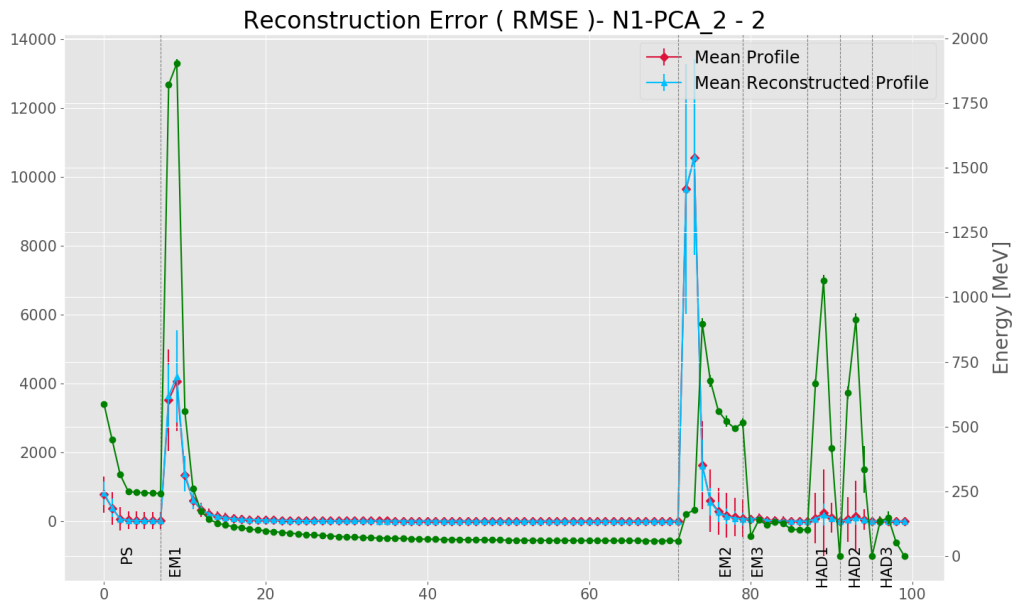
Figura 7.13: Comparação da capacidade de reconstrução da informação original considerando os 4 modelos para codificação em dimensão 2, para cada um dos 100 anéis.

A Figura 7.14 permite uma análise da relação entre o perfil dos anéis (original e reconstruído) e o erro de reconstrução (RMSE), para os codificadores *a* PCA e *b* rede auto-codificadora com 3 camadas, com codificação em 2 dimensões. A PCA possui um erro de reconstrução superior nas camadas HAD1 e HAD2 da seção hadrônica, no PS e no anéis 9 e 10, pertencentes à seção EM1. Por outro lado, os anéis mais energéticos, anéis 73 e 74, pertencentes à EM2, apresentam erro de reconstrução muito inferior para a PCA quando comparada à rede auto-codificadora. Isso mostra que a PCA, por se basear na variância, tende a dar ênfase nestes anéis com maior valor de energia e, neste caso, maior variância. Como consequência, acaba perdendo informação proveniente de outros anéis de outras camadas, que também possuem valores de energia relativamente altos. A rede auto-codificadora profunda apresenta

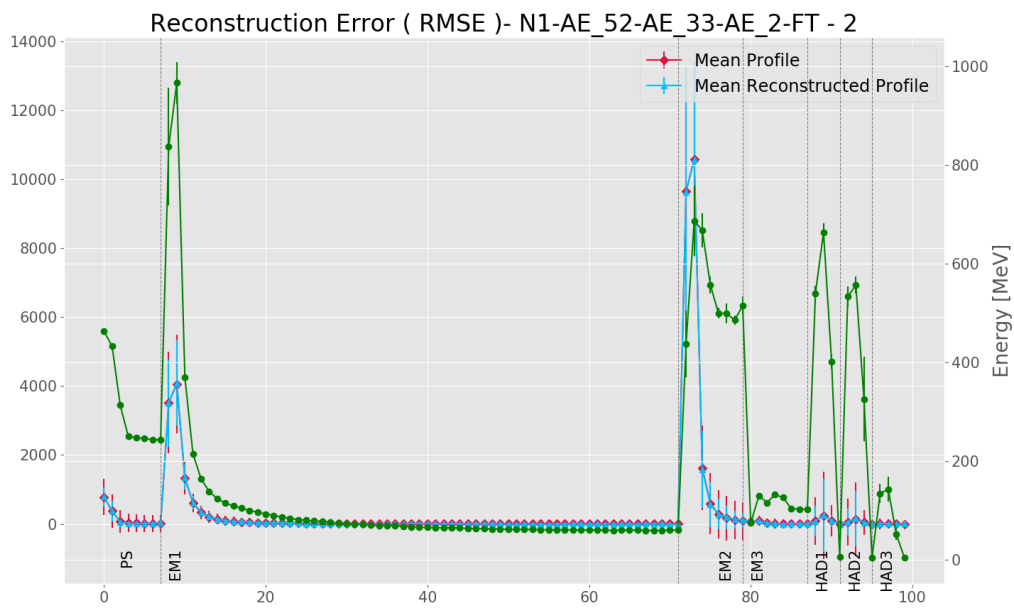
maior erro de reconstrução nos anéis com mais energia, mas não da ênfase especial aos 2 anéis mais energéticos. A comparação da capacidade de reconstrução da informação original, por classe, utilizando a informação mútua normalizada é feita na Figura 7.15, onde nota-se o mesmo comportamento. É possível perceber neste caso, porém, que ambos os modelos apresentam reconstrução melhor para a classe ruído em quase todos os anéis, com exceção dos primeiros da EM1 e dos anéis 73 e 74, mostrando que a classe que deposita, em média, mais energia em um dado anel é melhor reconstruída. Para estes últimos, no caso da PCA, a reconstrução é igual para ambas as classes, enquanto a rede auto-codificadora com 3 camadas apresenta uma reconstrução melhor para sinal.

As Figuras 7.16, 7.17, 7.18, 7.19, mostram os histogramas da energia dos anéis 9, 10, 73 e 74, respectivamente, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a)* PCA, *b)* NLPCA, *c)* rede auto-codificadora simples e *d)* rede auto-codificadora com 3 camadas para codificação em duas dimensões. A ênfase em tais anéis é dada por serem os que apresentam o maior valor de energia depositada nas células que os compõe e por terem apresentado qualidade de reconstrução diferente dependendo do codificador utilizado. Nota-se em todos os casos a superioridade dos modelos que utilizam transformação não-linear, especialmente a rede auto-codificadora com 3 camadas na reconstrução da energia destes anéis.

Considerando que 27 componentes fornecem 99% na variância retida pela PCA e que os métodos de codificação não linear apresentam vantagens apenas para codificações com 2 componentes, o gráfico de espalhamento dos anéis, considerando entrada e reconstrução, obtida pela PCA e pela rede auto-codificadora com 3 camadas, é apresentado para ambas as codificações em 27 e 2 dimensões no Apêndice B.

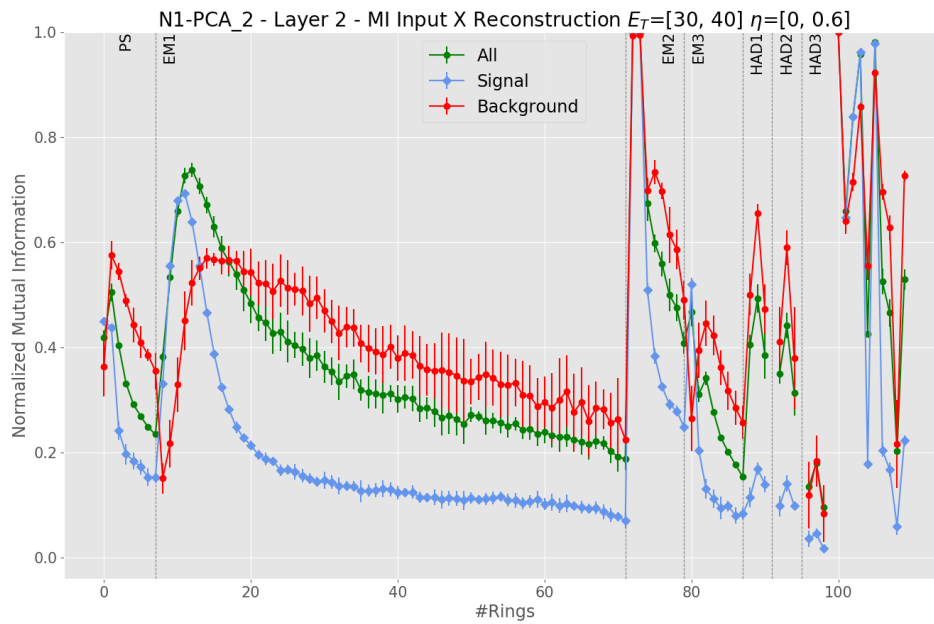


(a)

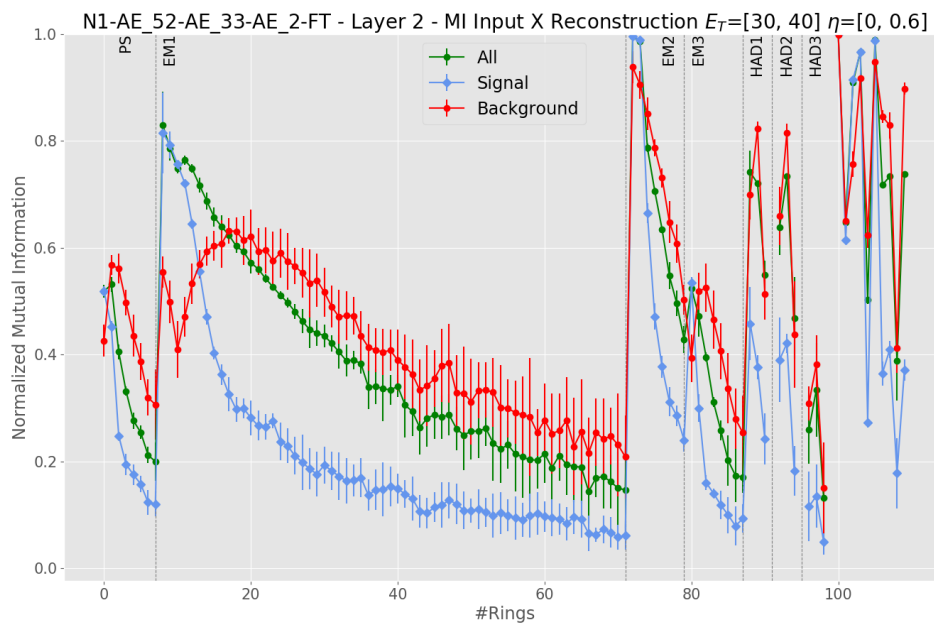


(b)

Figura 7.14: Análise da relação entre o perfil dos anéis (vermelho), perfil reconstruído (azul) e o erro de reconstrução (RMSE) (verde) para os codificadores *a* PCA e *b* rede auto-codificada com 3 camadas para codificação em 2 dimensões.

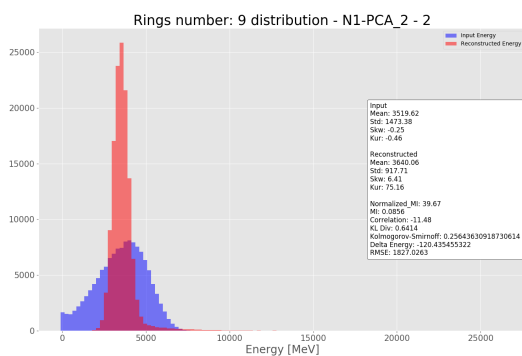


(a)

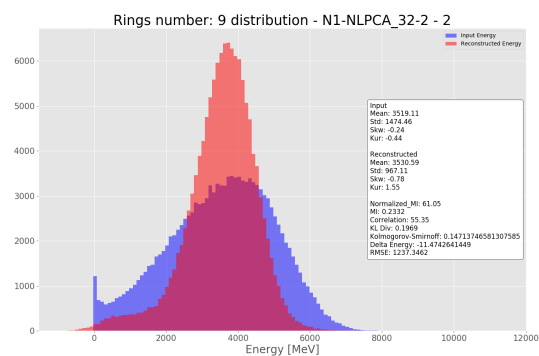


(b)

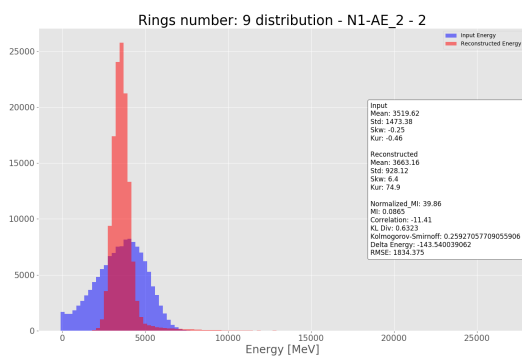
Figura 7.15: Comparação da capacidade de reconstrução da informação original para cada classe, considerando os modelos *a* PCA e *b* rede auto-codificadora com 3 camadas para codificação em dimensão 2, para cada um dos 100 anos



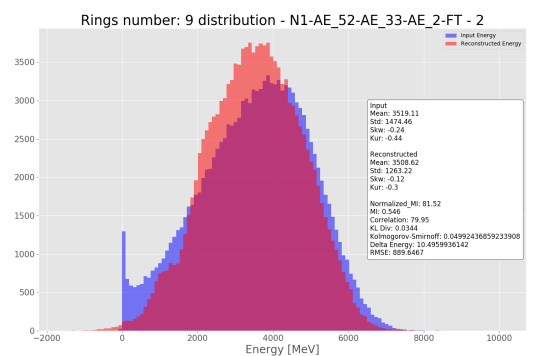
(a)



(b)

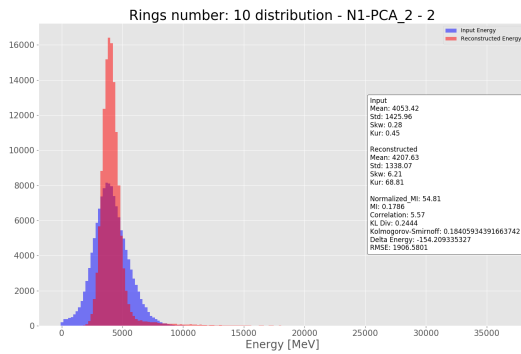


(c)

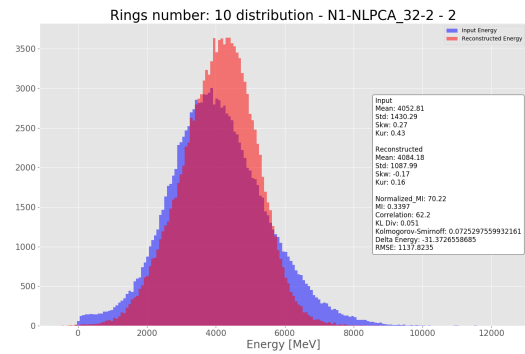


(d)

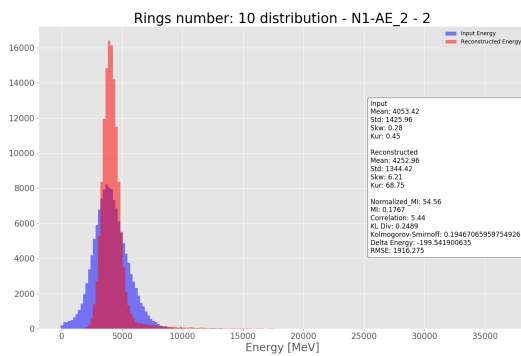
Figura 7.16: Histogramas da energia do anel 9, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.



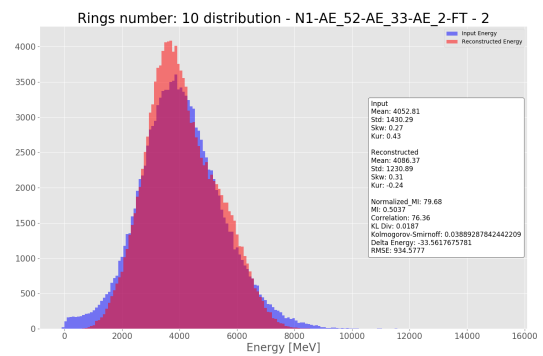
(a)



(b)



(c)



(d)

Figura 7.17: Histogramas da energia do anel 10, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.

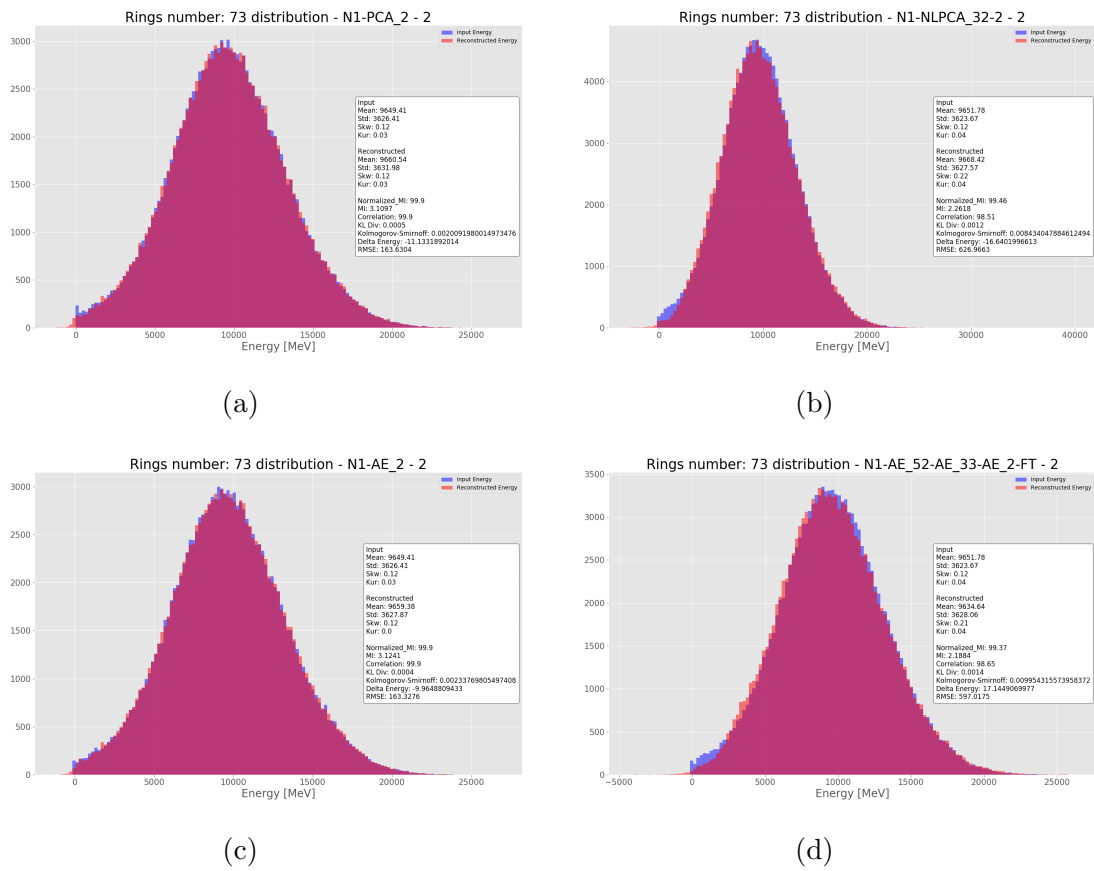


Figura 7.18: Histogramas da energia do anel 73, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.

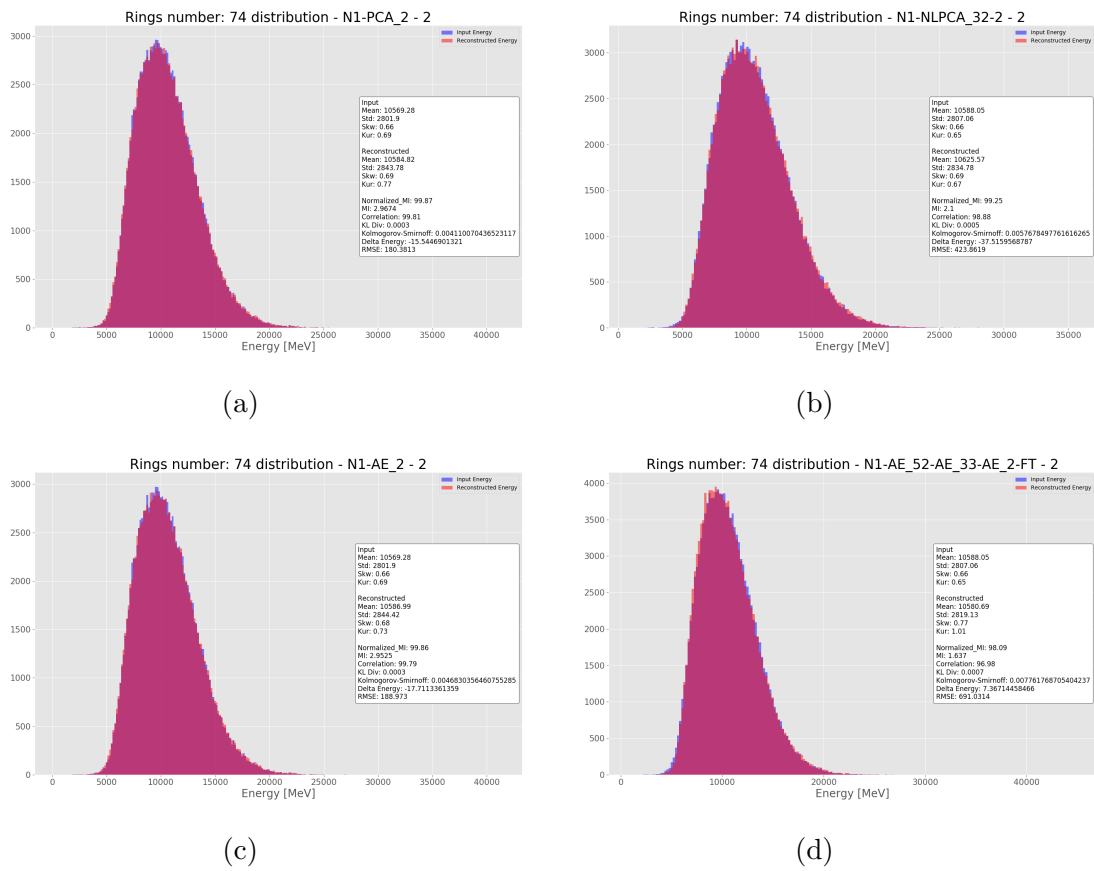


Figura 7.19: Histogramas da energia do anel 74, considerando os dados de entrada e os dados reconstruídos pelos codificadores *a* PCA, *b* NLPCA, *c* rede auto-codificadora simples e *d* rede auto-codificadora com 3 camadas.

7.2.4 Reconstrução das camadas longitudinais

As Tabelas 7.1 e 7.2 apresentam o valor de reconstrução das seções eletromagnética e hadrônica, respectivamente, para 27 dimensões codificadas, considerando todos os codificadores utilizados. Nota-se que todos os modelos apresentaram valores similares e encontram-se dentro da barra de erro. As Tabelas 7.3 e 7.4 apresentam o valor de reconstrução das seções eletromagnética e hadrônica, respectivamente, para 2 dimensões codificadas. Neste caso, os modelos não lineares apresentam melhor qualidade de reconstrução. É possível notar tal superioridade, especificamente, nas camadas EM1 e EM3 da seção eletromagnética e em todas as camadas do calorímetro hadrônico.

Tabela 7.1: Valores de informação mútua normalizada para a codificação em 27 dimensões, considerando as camadas do calorímetro eletromagnético.

Codificação	<i>PS</i>	<i>EM1</i>	<i>EM2</i>	<i>EM3</i>
<i>AE</i>	0.9981±0.0001	0.9934±0.0000	0.9998±0.0000	0.7213±0.0131
<i>AE2CamadasFT</i>	0.9980±0.0001	0.9933±0.0001	0.9998±0.0000	0.7325±0.0056
<i>AE3CamadasFT</i>	0.9979±0.0001	0.9933±0.0001	0.9998±0.0000	0.7336±0.0078
<i>AE5CamadasFT</i>	0.9979±0.0000	0.9933±0.0001	0.9997±0.0000	0.7340±0.0074
<i>AE8CamadasFT</i>	0.9976±0.0005	0.9934±0.0003	0.9997±0.0000	0.7462±0.0183
<i>CAE</i>	0.9980±0.0000	0.9933±0.0001	0.9997±0.0000	0.7346±0.0063
<i>NLPCA</i>	0.9980±0.0000	0.9933±0.0001	0.9998±0.0000	0.7337±0.0046
<i>PCA</i>	0.9980±0.0000	0.9933±0.0001	0.9998±0.0000	0.7358±0.0067

Tabela 7.2: Valores de informação mútua normalizada para a codificação em 27 dimensões, considerando as camadas do calorímetro hadrônico.

Codificação	<i>HAD1</i>	<i>HAD2</i>	<i>HAD3</i>
<i>AE</i>	0.9994±0.0000	0.9981±0.0001	0.4288±0.0268
<i>AE2CamadasFT</i>	0.9994±0.0000	0.9979±0.0001	0.4270±0.0356
<i>AE3CamadasFT</i>	0.9994±0.0000	0.9978±0.0002	0.4233±0.0265
<i>AE5CamadasFT</i>	0.9993±0.0001	0.9976±0.0002	0.4256±0.0426
<i>AE8CamadasFT</i>	0.9991±0.0002	0.9966±0.0008	0.4206±0.0238
<i>CAE</i>	0.9994±0.0000	0.9979±0.0000	0.4466±0.0313
<i>NLPCA</i>	0.9994±0.0000	0.9979±0.0001	0.4272±0.0367
<i>PCA</i>	0.9994±0.0000	0.9979±0.0000	0.4427±0.0274

Tabela 7.3: Valores de informação mútua normalizada para a codificação em 2 dimensões, considerando as camadas do calorímetro eletromagnético.

Codificação	<i>PS</i>	<i>EM1</i>	<i>EM2</i>	<i>EM3</i>
<i>AE</i>	0.6604±0.0095	0.8373±0.0052	0.9578±0.0008	0.4269±0.0040
<i>AE2CamadasFT</i>	0.6495±0.0119	0.9092±0.0047	0.9670±0.0010	0.4987±0.0077
<i>AE3CamadasFT</i>	0.6496±0.0111	0.9094±0.0030	0.9677±0.0016	0.5042±0.0077
<i>AE5CamadasFT</i>	0.6563±0.0126	0.9159±0.0057	0.9693±0.0013	0.5114±0.0091
<i>AE8CamadasFT</i>	0.6481±0.0096	0.9243±0.0076	0.9681±0.0027	0.5097±0.0075
<i>CAE</i>	0.6589±0.0086	0.8385±0.0030	0.9584±0.0008	0.4283±0.0050
<i>NLPCA</i>	0.6550±0.0105	0.9078±0.0032	0.9681±0.0014	0.4988±0.0103
<i>PCA</i>	0.6596±0.0090	0.8390±0.0032	0.9584±0.0008	0.4262±0.0023

Tabela 7.4: Valores de informação mútua normalizada para a codificação em 2 dimensões, considerando as camadas do calorímetro hadrônico.

Codificação	<i>HAD1</i>	<i>HAD2</i>	<i>HAD3</i>
<i>AE</i>	0.5246±0.0248	0.4660±0.0267	0.1964±0.0226
<i>AE2CamadasFT</i>	0.7139±0.0053	0.7346±0.0079	0.3935±0.0633
<i>AE3CamadasFT</i>	0.7200±0.0057	0.7338±0.0065	0.3760±0.0803
<i>AE5CamadasFT</i>	0.7352±0.0117	0.7469±0.0085	0.4023±0.0655
<i>AE8CamadasFT</i>	0.7352±0.0066	0.7267±0.0404	0.3548±0.0577
<i>CAE</i>	0.5236±0.0227	0.4617±0.0255	0.2065±0.0190
<i>NLPCA</i>	0.7136±0.0067	0.7323±0.0091	0.3977±0.0680
<i>PCA</i>	0.5254±0.0247	0.4660±0.0268	0.2029±0.0228

A partir das análises feitas, é possível concluir que os modelos que fazem uso de transformações não lineares apresentam vantagens na qualidade de reconstrução em relação aos codificadores lineares. Entretanto, tais vantagens aparecem apenas quando a dimensão codificada é inferior a 3. Sendo assim, em casos onde não se necessita uma codificação tão intensa da informação do espaço de entrada, métodos de codificação lineares, como a PCA, podem ser utilizados por serem mais simples. O estudo feito apresentou, também, os pontos de codificação onde a reconstrução da energia total de camadas específicas dos calorímetros eletromagnético e hadrônico começa a sofrer deterioração, bem como em quais codificações a reconstrução da distribuição de anéis específicos começa a ser degradada. Como a função custo utilizada pelas redes auto-codificadoras é o erro médio quadrático, os anéis mais energéticos foram os últimos a terem a qualidade de reconstrução degradada. De forma similar, a PCA também tem preferência para reconstrução dos anéis mais energéticos, por estes serem, normalmente, os que apresentam maior variância. Entretanto, devido

à característica não gaussiana da distribuição dos anéis, a rede auto-codificadora profunda e o modelo NLPCA obtiveram melhores resultados na reconstrução destes anéis. Na próxima seção, a avaliação do impacto das codificações realizadas na capacidade discriminativa do classificador será analisada.

7.3 Análise de Discriminação

O impacto das codificações, efetuadas pelos diferentes modelos, na tarefa de discriminação elétron-hádron será analisada em duas partes. Primeiramente, será feita uma análise utilizando apenas pré-processamento por codificação não-supervisionada. Em um segundo momento, a rede auto-codificadora discriminante, que utiliza informação das classes para efetuar a representação da informação, será analisada e, então, comparada com os demais modelos.

7.3.1 Relevância dos anéis

Inicialmente é feita a análise da relevância de variáveis, considerando a contribuição relativa de cada anel para o MSE e para o índice SP do classificador, conforme discutido na Seção 6.4.3.4. Para esta análise, o algoritmo *NeuralRinger* é utilizado sem nenhuma codificação do espaço de entrada. A Figura 7.20 mostra o impacto de cada anel no MSE do classificador, enquanto a Figura 7.21 apresenta o impacto no índice SP, discutido na Seção 6.4.3.3. Esta análise aponta os anéis 9 e 10, pertencentes à EM1, e o anel 90, pertencente à HAD1, como os que mais contribuem para o índice SP, sendo consideradas, portanto, variáveis altamente discriminantes. Os primeiro anéis da PS, os primeiros da EM1, os anéis da EM2 e os anéis mais energéticos da HAD1 e HAD2 também apresentaram impacto considerável.

7.3.2 Codificação não-supervisionada

Na Figura 7.22, nota-se a importância da etapa de ajuste fino para melhoria da capacidade de discriminação elétron-hádron em codificações com poucas dimensões. Os modelos sem ajuste fino apresentaram maior instabilidade com uma barra de erro relativamente larga, considerando os diferentes resultados de validação cruzada.

Na sequência, o impacto da redução de dimensionalidade ao longo de diferentes dimensões codificadas, considerando os modelos testados, é analisada. Na Figura 7.23, nota-se que a capacidade discriminativa das diferentes codificações se mantém similar e constante a medida que se reduz a dimensionalidade até uma representação com 25 dimensões. Verifica-se, também, reduções na capacidade discriminativa das codificações quando a dimensão é reduzida abaixo de 18 e, em seguida, de 6 dimensões, acarretando em decréscimo de 0.2% no índice SP em todos os casos. O

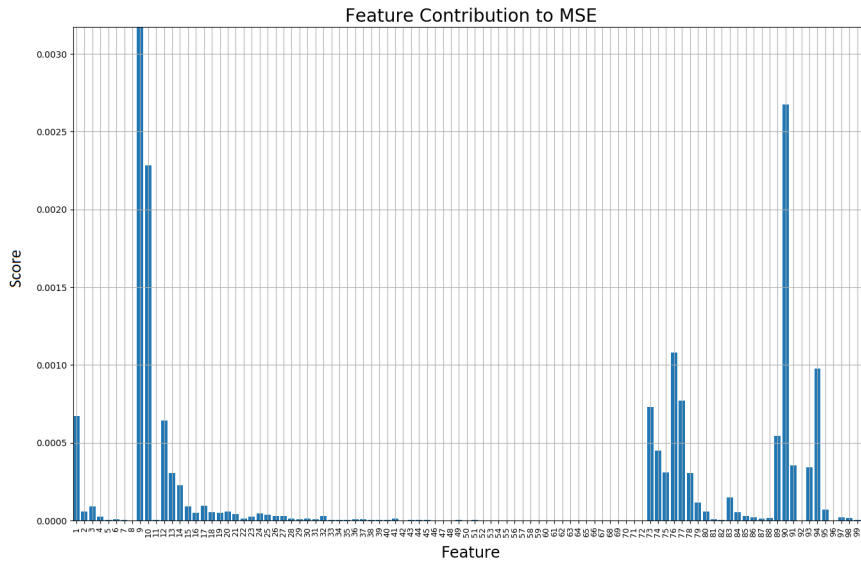


Figura 7.20: Relevância de variáveis considerando a contribuição individual de cada anel para o MSE do classificador.

modelo auto-codificador profundo com 8 camadas apresentou instabilidade em codificações acima de 70 dimensões, o que pode ter ocorrido devido à grande quantidade de parâmetros empregada por este modelo, acarretando em sobreajuste. Quando se observa uma codificação de apenas duas dimensões, entretanto, os modelos não-lineares de múltiplas camadas (SAE e NLPCA) apresentam vantagem significativa em relação aos demais modelos, especialmente os modelos com 3 e 5 camadas, o que mostra um ganho na utilização de codificações que acessam a estatística de ordem superior quando chegamos a esta dimensão. Isso pode ser um indício que a qualidade superior da reconstrução dos anéis 9 e 10 por estes modelos, para codificações em duas dimensões, quando comparada aos modelos lineares, pode impactar positivamente na capacidade discriminativa do classificador, dada a relevância destes anéis para a classificação.

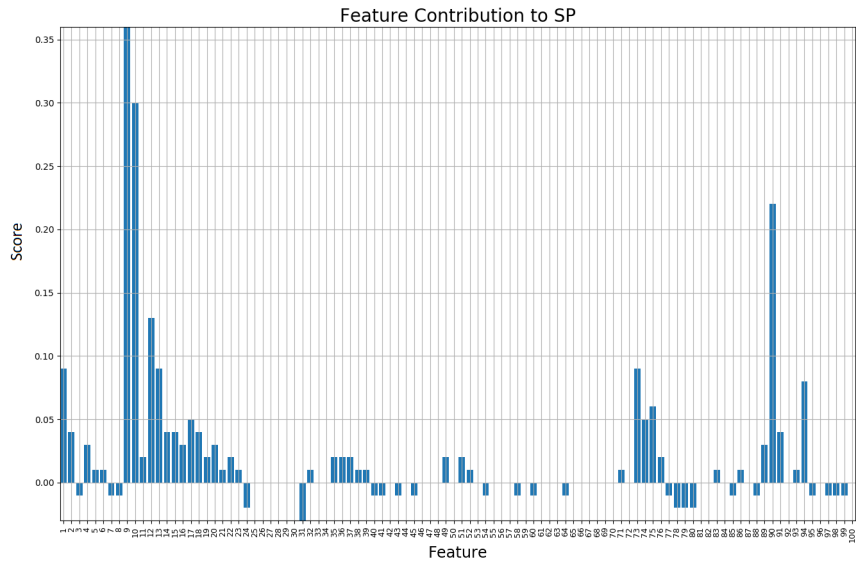


Figura 7.21: Relevância de variáveis considerando a contribuição individual de cadaanel para o índice SP do classificador.

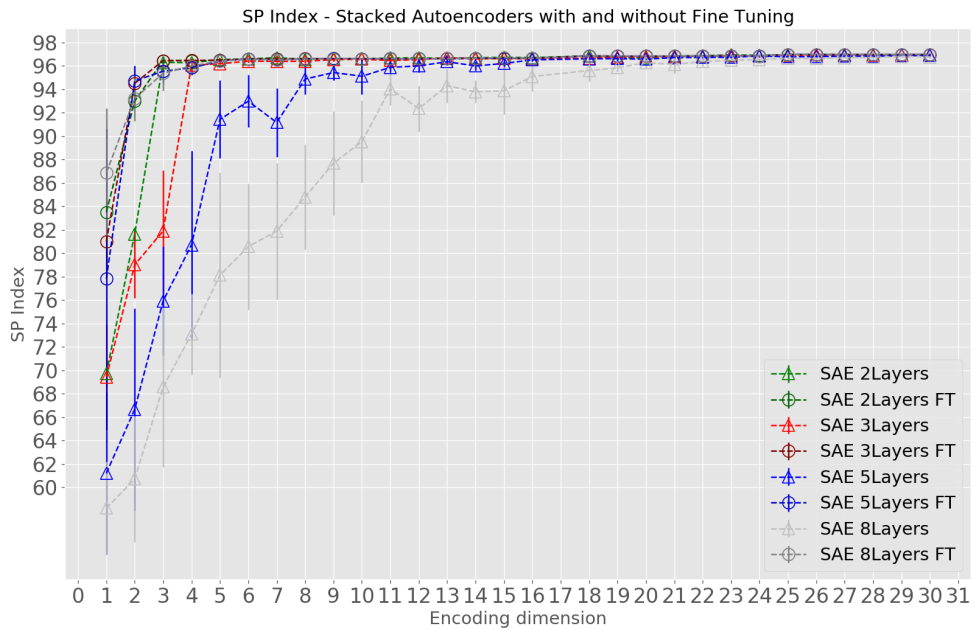


Figura 7.22: Valor do índice soma produto para diferentes redes auto-codificadoras com múltiplas camadas com e sem o uso de ajuste fino após o treinamento camada a camada. A dimensão da codificação variou de 30 a 1. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$

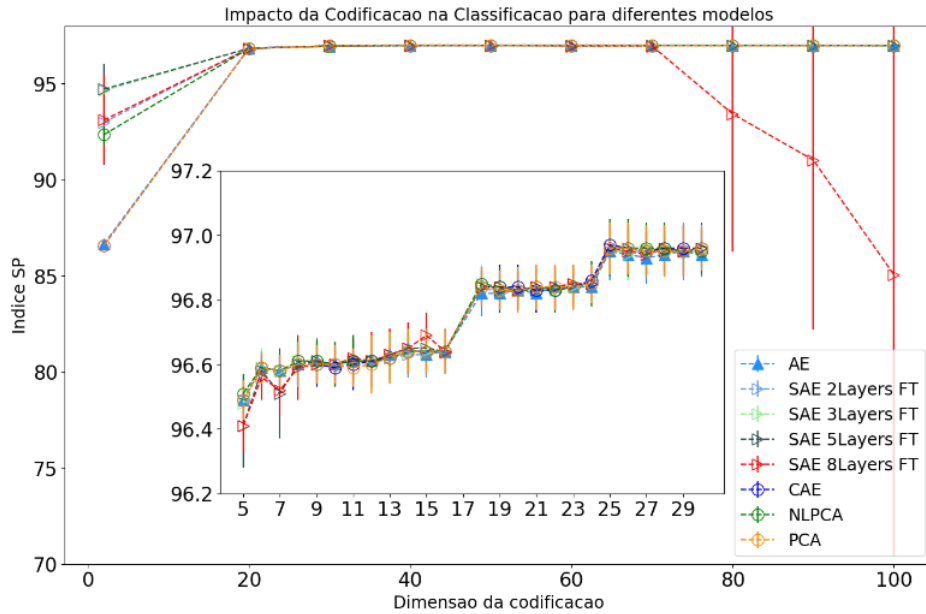


Figura 7.23: Impacto da redução de dimensionalidade ao longo de diferentes dimensões codificadas, considerando os modelos testados.

As tabelas 7.5 e 7.6 apresentam os resultados para o algoritmo *NeuralRinger* sem nenhuma codificação (100 anéis), quando comparado com as demais codificações estudadas. A tabela 7.5 apresenta uma codificação de dimensão 25, enquanto a tabela 7.6 apresenta uma codificação de dimensão 2. Nota-se que é possível reduzir a dimensionalidade em 75% sem perder a capacidade discriminativa ou em 98% com o custo de dobrar a taxa de falso alarme.

Tabela 7.5: Valores de eficiência para codificação em 25 dimensões, operando no ponto de máximo SP.

Codificação	$SP[\%]$	$P_D[\%]$	$F_a[\%]$	$AUC[\%]$
100aneis(Ref)	96.97 ± 0.07	97.77 ± 0.22	3.82 ± 0.28	96.97 ± 0.07
AE	96.95 ± 0.09	97.87 ± 0.11	3.98 ± 0.22	96.95 ± 0.09
AE2CamadasFT	96.95 ± 0.08	97.91 ± 0.19	4.00 ± 0.23	96.96 ± 0.08
AE3CamadasFT	96.96 ± 0.08	97.82 ± 0.31	3.89 ± 0.28	96.96 ± 0.08
AE5CamadasFT	96.96 ± 0.08	97.75 ± 0.31	3.83 ± 0.29	96.96 ± 0.08
AE8CamadasFT	96.96 ± 0.08	97.83 ± 0.21	3.90 ± 0.19	96.96 ± 0.08
CAE	96.97 ± 0.07	97.96 ± 0.18	4.03 ± 0.18	96.97 ± 0.07
NLPCA	96.96 ± 0.09	97.85 ± 0.20	3.93 ± 0.22	96.96 ± 0.09
PCA	96.96 ± 0.08	97.88 ± 0.18	3.95 ± 0.21	96.96 ± 0.08

Tabela 7.6: Valores de eficiência para codificação em 2 dimensões, operando no ponto de máximo SP

Codificação (2D)	$SP[\%]$	$P_D[\%]$	$F_a[\%]$	$AUC[\%]$
100aneis(<i>Ref</i>)	96.97±0.07	97.77±0.22	3.82±0.28	96.97±0.07
<i>AE</i>	86.65±0.19	90.44±0.94	17.05±1.05	86.69±0.18
<i>AE2CamadasFT</i>	92.98±0.54	97.68±0.80	11.61±0.53	93.03±0.54
<i>AE3CamadasFT</i>	94.63±0.45	99.16±0.17	9.79±0.87	94.69±0.44
<i>AE5CamadasFT</i>	94.71±1.32	97.39±1.53	7.91±3.16	94.74±1.29
<i>AE8CamadasFT</i>	93.10±2.36	98.39±0.85	12.01±4.83	93.19±2.27
<i>CAE</i>	86.58±0.23	90.07±0.99	16.82±0.88	86.62±0.23
<i>NLPCA</i>	92.34±0.52	97.16±0.80	12.36±0.62	92.40±0.52
<i>PCA</i>	86.58±0.22	90.05±0.97	16.82±0.97	86.62±0.22

7.3.3 Codificação através da rede auto-codificadora discriminante

A rede auto-codificadora discriminante tenta aumentar a compactação intra-classe e a separação inter-classe, enquanto fornece uma codificação em dimensão reduzida. Para tanto, esta rede reconstrói uma transformação da entrada mais discriminante, tornando as amostras de uma classe mais próximas entre si e mais distantes da classe adversária. Isso é feito através de um processo iterativo de mudança de alvo, como foi descrito na Seção 5.1.5. Nesta análise, o conjunto de dados utilizado corresponde à faixa de energia entre 20 e 30GeV e η entre 0.6 e 0.8, totalizando 228339, sendo 135327 amostras de sinal e 93012 de ruído.

Inicialmente, diferentes números de iterações, foram comparados para esta rede, utilizando os parâmetros mencionados em 6.4.3.2. como pode ser visto na Figura 7.24. Como o índice SP se estabiliza após 9 iterações, o valor de 10 iterações foi escolhido para comparação com os demais modelos.

A Figura 7.25 mostra a projeção dos 100 anéis de energia em duas componentes com uso de análise de componentes principais (PCA). Nota-se que a capacidade discriminativa se torna restrita, a medida que as componentes desta projeção apresentam uma área de sobreposição significativa. A Figura 7.26 mostra a projeção em duas componentes, obtidas através da rede auto-codificadora discriminativa com três camadas (50-25-2), para diferentes iterações de transformação dos dados. As componentes apresentadas são obtidas através da projeção do espaço de entrada na camada intermediária, após um número de 1, 4, 6 e 10 iterações de transformação dos dados de entrada.

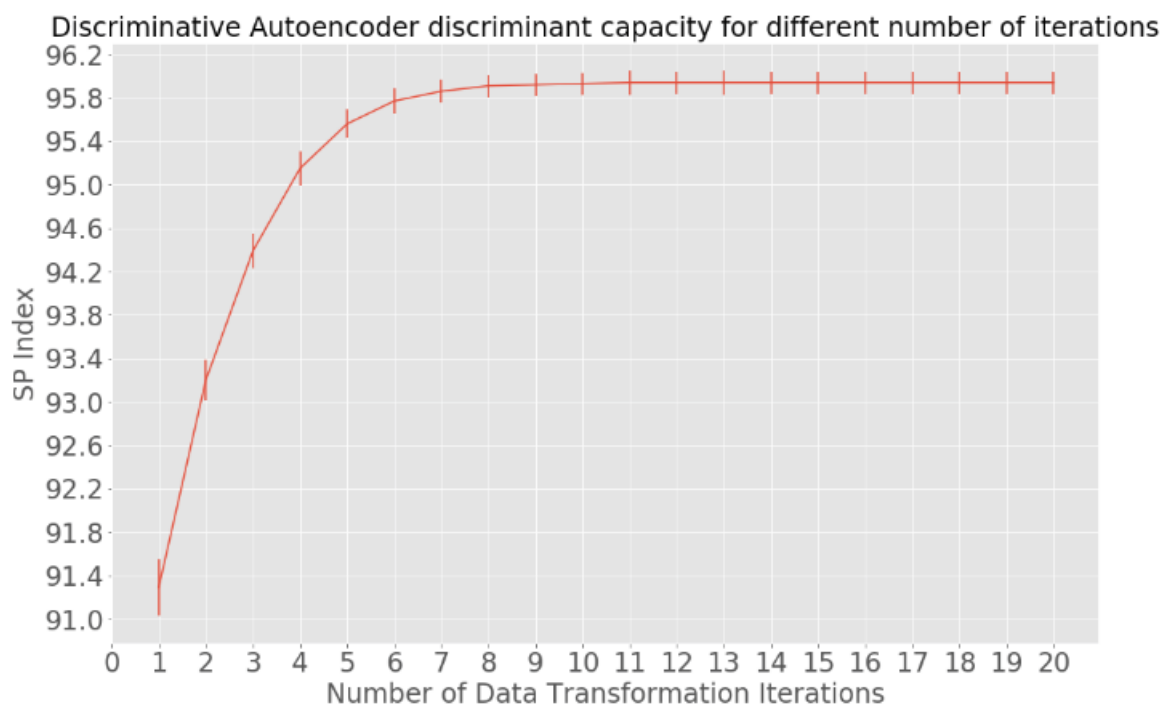


Figura 7.24: Valor do índice SP para diferentes números de iterações de transformações de dados para a rede auto-codificadora discriminante, considerando apenas uma dimensão na codificação. $30\text{GeV} < E_t < 40\text{GeV}$, $0.0 < \eta < 0.6$

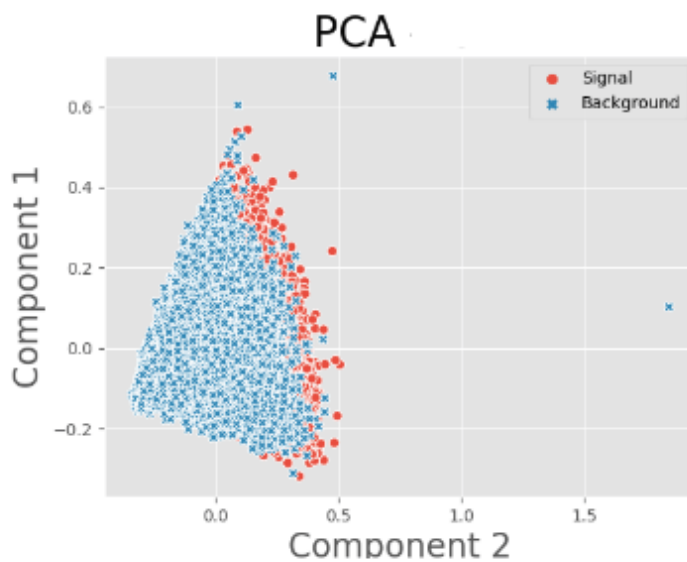


Figura 7.25: Projeção da informação anelada de calorimetria em duas componentes obtidas através da análise de componentes principais (PCA).



Figura 7.26: Projeção da informação anelada de calorimetria em duas componentes, obtidas através da rede auto-codificadora discriminativa com três camadas (50-25-2), para diferentes iterações.

Os diferentes modelos de codificação foram comparados, desta vez com a rede auto-codificadora discriminante, com relação à capacidade de discriminação fornecida pela estrutura latente encontrada. Na figura 7.27 verifica-se que este modelo obtém os melhores resultados, o que era esperado, a medida que este é o único modelo que utiliza informação de classe durante o treinamento, não consistindo, exatamente, de um modelo de caráter não-supervisionado como os outros. Os demais modelos apresentam degradação no índice SP para codificações abaixo de 25 dimensões. Os modelos NLPCA e auto-codificador profundo com 2 camadas apresentaram, entretanto, uma melhor performance, para dimensões abaixo de 3, quando comparados aos demais modelos. Sendo assim, nota-se uma vantagem na utilização de codificações não-lineares que exploram a estatística de ordem superior para baixas dimensões.



Figura 7.27: Valor do índice soma produto para diferentes modelos de compactação, a dimensão da codificação variando de 30 a 1. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$

A rede auto-codificadora discriminante apresentou resultados similares, independente da quantidade de camadas, quando foram testadas 3 arquiteturas: 100-50-25-1, 100-25-1 e 100,1. Sendo assim, verificou-se que, mesmo com apenas uma camada, esta rede consegue manter a capacidade discriminativa com apenas 1 dimensão codificada, o que apresenta a possibilidade de codificar os anéis de forma simples. A Figura 7.28 apresenta uma análise dos parâmetros aprendidos por tal rede. Com base na figura, é possível verificar que a rede fornece valores de peso positivos, em

verde aos anéis que apresentam, em média, valor de energia maior para a classe ruído que para a classe sinal e, por outro lado, valores negativos para os anéis onde a classe sinal deposita mais energia. Nota-se também que a magnitude dos parâmetros tem relação com a diferença média de energia entre ambas as classes. Na parte inferior da Figura 7.28, a multiplicação dos pesos pela média de cada classe é feita e somada. Nota-se que após a aplicação da função tangente hiperbólica, na soma obtida, o valor para a classe sinal é convertido para -0.94 , enquanto para a classe ruído o valor é de 0.99 . Isso mostra que o modelo de rede auto-codificadora discriminativa alcança uma separação significativa das classes, ao mesmo tempo que é interpretável. Sendo assim, é possível que esta parametrização sirva para codificação de uma variável discriminante representando a informação anelada de calorimetria e auxilie na análise *offline*, uma vez que se conhece o processo de tomada de decisão.

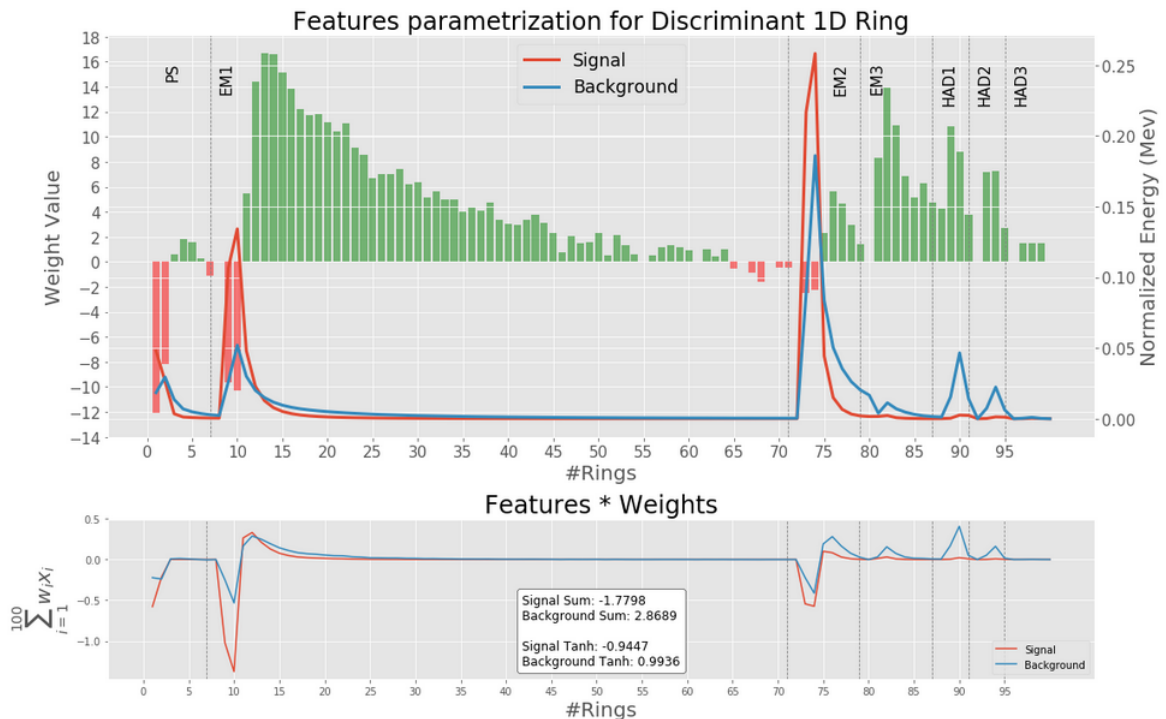


Figura 7.28: Valor dos parâmetros associados a cada anel pela codificação discriminante. Valores positivos (verde) são atribuídos para anéis que apresentam um valor energético apresentado pela classe ruído maior que a classe sinal, enquanto valores negativos (vermelho) são atribuídos para anéis em que a classe sinal deposita mais energia.

A Tabela 7.7 apresenta os resultados para o algoritmo *NeuralRinger* sem nenhuma codificação (100 anéis), comparando com as demais codificações estudadas. Nota-se um bom indicativo que a rede auto-codificadora discriminante possibilita a redução da dimensionalidade em 99% sem perder a capacidade discriminativa, uma vez que esta se encontra dentro da barra de erro quando comparada com a referência

Tabela 7.7: Valores de eficiência para codificação em 1 dimensões, operando no ponto de máximo SP

Codificação	$SP[\%]$	$P_D[\%]$	$P_f[\%]$
100aneis(<i>Ref</i>)	95.98±0.11	97.19±0.34	5.22±0.31
<i>AE</i>	87.70±0.25	91.92±0.97	16.41±1.14
<i>AE, 2Camadas</i>	81.62±0.92	86.15±1.34	22.78±1.10
<i>AE, 2CamadasFT</i>	90.15±1.55	95.48±2.05	15.02±1.27
<i>AE, 3Camadas</i>	82.64±4.36	86.56±4.80	21.19±4.00
<i>AE, 3CamadasFT</i>	88.53±1.68	92.58±2.18	15.42±1.51
<i>CAE</i>	87.84±0.25	91.99±1.17	16.20±1.20
<i>Disc10IterAE</i>	95.93±0.10	96.94±0.39	5.07±0.39
<i>NLPCA</i>	88.42±0.99	92.58±1.87	15.64±1.08
<i>PCA</i>	87.83±0.25	91.61±1.01	15.86±1.07

(*NeuralRinger* com o espaço original de 100 anéis).

Capítulo 8

Conclusão e Trabalhos Futuros

O algoritmo *NeuralRinger* entrou em operação em 2017 em uma das etapas do sistema de filtragem *online* (*Trigger*) de elétrons do ATLAS, em sua camada de alto nível (High Level Trigger, HLT). Entretanto, tradicionalmente os modelos utilizados no ambiente *online* consistem em degradações dos modelos sofisticados que operam em *offline*, devido à importância da consonância entre ambos os ambientes. Por não existir uma contrapartida no *offline*, justifica-se um estudo aprofundado neste ambiente para este algoritmo. O ambiente *offline* não possui critérios rígidos de latência, permitindo, assim, o estudo de modelos complexos. Este trabalho se propôs a estudar métodos de codificação na tentativa de validar a capacidade de reconstrução da informação de calorimetria, provenientes do anéis de energia, obtida por codificadores lineares e não lineares. O estudo também buscou identificar se é possível reduzir a dimensão do espaço de entrada sem perder capacidade discriminante ou encontrar uma representação que contribua para um melhor desempenho de classificação.

As redes auto-codificadoras forneceram um modelo no qual é possível modular o número de componentes fazendo um compromisso entre acuidade na reconstrução, estatística e pureza na discriminação. Os modelos forneceram, também, flexibilidade para acomodar diferentes requisitos da física de interesse, conforme o ambiente *offline* de um experimento complexo como o ATLAS exige. O estudo apresentou a possibilidade de se reduzir a dimensão do espaço de entrada de 100 anéis de informação para 25 componentes sem prejuízo na capacidade discriminativa e para apenas 2 dimensões, com um decréscimo de 2,5% no índice SP ou o dobro na taxa de falso alarme (de 3,82% para 7,91%). Neste último caso, os modelos que fazem uso de transformação não linear apresentaram resultados superiores em relação aos modelos lineares ou com apenas uma camada, o que mostra uma vantagem na utilização de codificação não linear na representação dos dados com poucas dimensões. Este resultado é importante, pois apresenta a possibilidade de manutenção da capacidade de filtragem com redução do processamento computacional, através do uso

de representações lineares (análise de componentes principais), o que pode se tornar uma opção para o ambiente *online*, bem como fornece a opção de redução ainda mais drástica de dimensionalidade ao custo de coletar o dobro de eventos indesejáveis. Por último, verificou-se também a importância da etapa de ajuste fino no treinamento de redes auto-codificadoras com múltiplas camadas, quando o treinamento é feito camada a camada.

A rede auto-codificadora discriminante, que foi também avaliada, manteve a capacidade discriminativa com apenas uma dimensão, ao passo que a representação encontrada por este modelo permitiu um entendimento do processo de tomada de decisão. A análise de tal codificação é interessante, uma vez que esta possui potencial tanto para o ambiente *offline* bem como *online*, além de permitir interpretação física. Destaque-se que é uma variável especialista obtida através do aprendizado de máquina. Como os experimentos em física de altas energias exigem um entendimento do comportamento dos modelos, os resultados apontam para uma possibilidade de se adicionar esta variável aos modelos empregados atualmente na filtragem *offline*, baseados em Likelihood, a fim de incorporar-lhe a informação anelada de calorimetria, potencializando a capacidade de discriminação, sem aumentar significativamente a complexidade da solução. A representação encontrada permite, também, uma codificação extremamente compacta da informação anelada para atuação no ambiente de filtragem *online*.

8.1 Trabalhos Futuros

Cada mudança, tanto no ambiente *online* como no *offline*, precisa ser avaliada com profundidade e, portanto, novas análises ainda são necessárias para verificar, por exemplo, se a compactação encontrada cria tendências. Necessita-se, também, analisar se há ganhos reais, sem perda de informação relevante, ao se adicionar a variável encontrada na decisão do *offline* pela Likelihood e, por último, é necessário analisar o comportamento dos modelos testados ao longo das demais faixas de energia e posição da partícula no detector, considerando-se a solução ensemble (*online* e *offline*). Além disso, o trabalho abre perspectivas tanto para a avaliação de outras alternativas de aprendizado profundo (outras variações de rede auto-codificadoras, redes neurais convolucionais, redes neurais recorrentes) na filtragem e identificação de elétrons, como para a aplicação dos modelos avaliados em outros ambientes do ATLAS, como a filtragem e identificação de fótons.

Referências Bibliográficas

- [1] ACOSTA, D., AVELLINO, P., BUONTEMPO, S., et al. “The performance of a lead/scintillating-fiber calorimeter at lhc/ssc compatible gate widths”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 314, n. 3, pp. 431–449, 1992. 3, 69
- [2] SEIXAS, J., CALOBA, L., SOUZA, M., et al. “Neural second-level trigger system based on calorimetry”, *Computer Physics Communications*, v. 95, n. 2-3, pp. 143–157, 1996. 3, 69, 75
- [3] LEDERMAN, L. M., TERESI, D. *The God particle: if the universe is the answer, what is the question?* Houghton Mifflin Harcourt, 2006. 6
- [4] NE’EMAN, Y., KIRSH, Y., OTHERS. *The particle hunters*. Cambridge University Press, 1996. 6
- [5] BERRYMAN, S. “Ancient atomism”, 2005. 6
- [6] BARACCA, A. “The revolution in physics of the early Nineteenth century revisited in the context of science-and-society interaction”, *arXiv preprint arXiv:1802.02504*, 2018. 6
- [7] HAWKING, S. *A brief history of time: from big bang to black holes*. Random House, 2009. 6
- [8] HODDESON, L., BROWN, L. M., BROWN, L., et al. *The rise of the standard model: A history of particle physics from 1964 to 1979*. Cambridge University Press, 1997. 6
- [9] GRIFFITHS, D. *Introduction to elementary particles*. John Wiley & Sons, 2008. 6, 7, 9, 11, 12
- [10] CARUSO, F., OGURI, V. “A ETERNA BUSCA no INDIVISIVEL: no ATOMO FILOSOFICO AOS QUARKS E LEPTONS”, *Química nova*, v. 20, pp. 3, 1997. 7

- [11] CARUSO, F., OGURI, V. *Física Moderna: origens clássicas e fundamentos quânticos*. Elsevier, 2006. 7
- [12] PAULI, W. “Relativistic Quantized Theory of Particles Obeying Bose-Einstein Statistics”, *Annals de l’Institut Henri Poincare*, v. 6, pp. 137, 1936. 7
- [13] PAULI, W. “The connection between spin and statistics”, *Physical Review*, v. 58, n. 8, pp. 716, 1940. 7
- [14] MESSIAH, A., GREENBERG, O. “Symmetrization postulate and its experimental foundation”, *Physical Review*, v. 136, n. 1B, pp. B248, 1964. 7
- [15] CURCEANU, C., GILLASPY, J., HILBORN, R. C. “Resource letter SS–1: The spin-statistics connection”, *American Journal of Physics*, v. 80, n. 7, pp. 561–577, 2012. 7
- [16] ROSENFELD, L. “Nuclear forces”, 1948. 8
- [17] GELL-MANN, M. “The Eightfold Way: A Theory of strong interaction symmetry. 1961”, *Cité page*, v. 27. 8
- [18] NE’EMAN, Y. “Derivation of strong interactions from a gauge invariance”, *Nuclear physics*, v. 26, n. 2, pp. 222–229, 1961. 8
- [19] GELL-MANN, M. “A schematic model of baryons and mesons”. In: *Murray Gell-Mann: Selected Papers*, World Scientific, pp. 151–152, 2010. 8
- [20] GELL-MANN, M. *The Quark and the Jaguar: Adventures in the Simple and the Complex*. Macmillan, 1995. 8
- [21] MARTIN, B. R., SHAW, G. *Nuclear and particle physics: an introduction*. Wiley, 2019. 8, 12
- [22] ANDERSON, C. D. “The positive electron”, *Physical Review*, v. 43, n. 6, pp. 491, 1933. 8
- [23] FUKUDA, Y., HAYAKAWA, T., ICHIHARA, E., et al. “Evidence for oscillation of atmospheric neutrinos”, *Physical Review Letters*, v. 81, n. 8, pp. 1562, 1998. 9
- [24] THOMAS, S. A., ABDALLA, F. B., LAHAV, O. “Upper bound of 0.28 eV on neutrino masses from the largest photometric redshift survey”, *Physical Review Letters*, v. 105, n. 3, pp. 031301, 2010. 9

- [25] TANABASHI, M., HAGIWARA, K., HIKASA, K., et al. “Review of Particle Physics”, *Phys. Rev. D*, v. 98, pp. 030001, Aug 2018. doi: 10.1103/PhysRevD.98.030001. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevD.98.030001>>. 9, 10
- [26] ROVELLI, C. “Loop quantum gravity”, *Living reviews in relativity*, v. 11, n. 1, pp. 5, 2008. 11, 13
- [27] ROVELLI, C. “Notes for a brief history of quantum gravity”, *arXiv preprint gr-qc/0006061*, 2000. 11, 13
- [28] KAISER, D. “Physics and Feynman’s Diagrams: In the hands of a postwar generation, a tool intended to lead quantum electrodynamics out of a decades-long morass helped transform physics”, *American Scientist*, v. 93, n. 2, pp. 156–165, 2005. 11
- [29] MARCIANO, W., PAGELS, H. “Quantum chromodynamics”, *Physics Reports*, v. 36, n. 3, pp. 137–276, 1978. 11
- [30] CAMPBELL, J., HUSTON, J., KRAUSS, F. *The black book of quantum chromodynamics: a primer for the LHC era*. Oxford University Press, 2018. 11
- [31] BOTTOMLEY, J., BAEZ, J. “Why are there eight gluons and not nine?” 1996. 12
- [32] ANDERSON, E. “Geometrodynamics: spacetime or space?” *arXiv preprint gr-qc/0409123*, 2004. 12
- [33] ROMERO, G. E. “Philosophical issues of black holes”, *arXiv preprint arXiv:1409.3318*, 2014. 12
- [34] SFETCU, N. “Ontology of Gravitational Singularities”, . 12
- [35] CURIEL, E., BOKULICH., P. “Singularities and Black Holes.” *The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University*, 2018. Disponível em: <<https://plato.stanford.edu/archives/sum2018/entries/spacetime-singularities/>>. 12
- [36] HAWKING, S. “The beginning of time”, *Stephen Hawking: The Official Website*, 1996. 12
- [37] DIÓSI, L. “Planck length challenges non-relativistic quantum mechanics of large masses”, *arXiv preprint arXiv:1903.04852*, 2019. 13

- [38] GORELIK, G. “First steps of quantum gravity and the Planck values”, *Studies in the history of general relativity*, v. 3, pp. 364–79, 1992. 13
- [39] FOLGER, T. “A Gravidade Quântica chega ao Laboratório”, *Scientific American Brasil*, Junho 2019. 13
- [40] BECKER, K., BECKER, M., SCHWARZ, J. H. *String theory and M-theory: A modern introduction*. Cambridge University Press, 2006. 13
- [41] WIKIPEDIA, THE FREE ENCYCLOPEDIA. “Modelo Padrão de Partículas Elementares”. 2008. Disponível em: <https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg>. [Online; Acessado em 24 de junho, 2019]. 14
- [42] PESKIN, M. E. *An introduction to quantum field theory*. CRC Press, 2018. 13
- [43] “Facts and figures about the LHC”. Disponível em: <<https://home.cern/resources/faqs/facts-and-figures-about-lhc>>. Acesso em: 16 de julho de 2019. 14, 31, 33
- [44] OFF, C. P. “CERN experiments observe particle consistent with long-sought Higgs boson”. 2012. 14
- [45] “The Nobel Prize in Physics 2013”. . Disponível em: <<https://www.nobelprize.org/prizes/physics/2013/summary/>>. Acesso em: 16 de julho de 2019. 14
- [46] ELLIS, J. “Outstanding questions: physics beyond the Standard Model”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 370, n. 1961, pp. 818–830, 2012. 15
- [47] “Dark Matter”. Disponível em: <<https://home.cern/science/physics/dark-matter>>. Acesso em: 14 de julho de 2019. 15, 16
- [48] “The Nobel Prize in Physics 1933”. . Disponível em: <<https://www.nobelprize.org/prizes/physics/1933/summary/>>. Acesso em: 15 de julho de 2019. 15
- [49] ALVES JR, A. A., ANDRADE FILHO, L., BARBOSA, A., et al. “The LHCb detector at the LHC”, *Journal of instrumentation*, v. 3, n. 08, pp. S08005, 2008. 15, 33
- [50] DIMOPOULOS, S., GEORGI, H. “Softly broken supersymmetry and SU (5)”, *Nuclear Physics B*, v. 193, n. 1, pp. 150–162, 1981. 15

- [51] LINCOLN, D. *The quantum frontier: The large Hadron collider*. JHU Press, 2009. 15, 16, 20, 23, 31
- [52] FAYET, P. “The Supersymmetric Standard Model”. In: *The Standard Theory of Particle Physics: Essays to Celebrate 's 60th Anniversary*, World Scientific, pp. 397–454, 2016. 15
- [53] RIESS, A. G., FILIPPENKO, A. V., CHALLIS, P., et al. “Observational evidence from supernovae for an accelerating universe and a cosmological constant”, *The Astronomical Journal*, v. 116, n. 3, pp. 1009, 1998. 16
- [54] PERLMUTTER, S., ALDERING, G., GOLDHABER, G., et al. “Measurements of Ω and Λ from 42 high-redshift supernovae”, *The Astrophysical Journal*, v. 517, n. 2, pp. 565, 1999. 16
- [55] “The Nobel Prize in Physics 2011”. . Disponível em: <<https://www.nobelprize.org/prizes/physics/2011/summary/>>. Acesso em: 15 de julho de 2019. 16
- [56] HILL, G. J., GEBHARDT, K., KOMATSU, E., et al. “The Hobby-Eberly Telescope Dark Energy Experiment (HETDEX): Description and Early Pilot Survey Results”, *arXiv preprint arXiv:0806.0183*, 2008. 16
- [57] ABBOTT, T., ABDALLA, F., ALLAM, S., et al. “The Dark Energy Survey: Data Release 1”, *The Astrophysical Journal Supplement Series*, v. 239, n. 2, pp. 18, 2018. 16
- [58] DAWSON, K. S., SCHLEGEL, D. J., AHN, C. P., et al. “The Baryon oscillation spectroscopic survey of SDSS-III”, *The Astronomical Journal*, v. 145, n. 1, pp. 10, 2012. 16
- [59] BASSETT, B. A., NICHOL, B., EISENSTEIN, D. J. “Sounding the dark cosmos”, *Astronomy & Geophysics*, v. 46, n. 5, pp. 5–26, 2005. 16
- [60] “Extra dimensions, gravitons, and tiny black holes”. Disponível em: <<https://atlas.cern/tags/extra-dimensions>>. Acesso em: 16 de julho de 2019. 17
- [61] POKOTILOVSKI, Y. N. “Experiments with ultracold neutrons—first 50 years”, *arXiv preprint arXiv:1805.05292*, 2018. 18
- [62] WIGMANS, R., WIGMANS, R. *Calorimetry: Energy measurement in particle physics*, v. 107. Oxford University Press, 2000. 18, 21, 23, 24, 25, 26, 27, 29, 51

- [63] ILLINGWORTH, V., CLARK, J. O. E. *The Facts on File dictionary of astronomy*, v. 229. Facts on file New York, 1979. 19
- [64] “Cosmic Rays”. Disponível em: <https://web.archive.org/web/20121028154200/http://imagine.gsfc.nasa.gov/docs/science/known_11/cosmic_rays.html>. Acesso em: 17 de novembro de 2019. 19
- [65] SOMMERS, P. “Experimental aspects of cosmic rays”, 2006. 19
- [66] “The Nobel Prize in Physics 1929”. . Disponível em: <<https://www.nobelprize.org/prizes/physics/1929/broglie/biographical/>>. Acesso em: 25 de julho de 2019. 19
- [67] FEYNMAN, R. P. *QED: The strange theory of light and matter*. Princeton University Press, 2006. 19
- [68] BY LUCIO ROSSI. “Seeing is believing”, Feb 2015. Disponível em: <<https://cds.cern.ch/record/2000237>>. 19
- [69] “What are the Differences Between Linear Accelerators, Cyclotrons, and Synchrotrons?” Disponível em: <<https://www.machinedesign.com/whats-difference-between/what-are-differences-between-linear-accelerators-cyclotrons-and-synchrotrons>>. Acesso em: 17 de novembro de 2019. 20
- [70] “Introduction to Accelerators: Evolution of Accelerators and Modern Day Applications”. Disponível em: <https://uspas.fnal.gov/materials/110DU/Lecture1a_Intro_to_Accelerators.pdf>. Acesso em: 28 de julho de 2019. 20
- [71] “A primer on particle accelerators”. Disponível em: <<https://www.symmetrymagazine.org/article/a-primer-on-particle-accelerators>>. Acesso em: 28 de julho de 2019. 20
- [72] “Sirius”. Disponível em: <<https://www.lnls.cnpem.br/sirius/>>. Acesso em: 28 de julho de 2019. 20
- [73] DE CASTRO FONSECA, M., ARAUJO, B. H. S., DIAS, C. S. B., et al. “High-resolution synchrotron-based X-ray microtomography as a tool to unveil the three-dimensional neuronal architecture of the brain”, *Scientific reports*, v. 8, n. 1, pp. 12074, 2018. 20

- [74] PAK, T., ARCHILHA, N. L., MANTOVANI, I. F., et al. “An X-ray computed micro-tomography dataset for oil removal from carbonate porous media”, *Scientific data*, v. 6, pp. 190004, 2019. 20
- [75] ARCHILHA, N. L., MISSAGIA, R. M., HOLLIS, C., et al. “Permeability and acoustic velocity controlling factors determined from x-ray tomography images of carbonate rocks”, *AAPG Bulletin*, v. 100, n. 8, pp. 1289–1309, 2016. 20
- [76] DRIEMEIER, C. E., LING, L. Y., YANCY-CABALLERO, D., et al. “Location of water in fresh sugarcane bagasse observed by synchrotron X-ray microtomography”, *PloS one*, v. 13, n. 12, pp. e0208219, 2018. 20
- [77] WESTFAHL, H., TOLENTINO, H. C., MENEAU, F., et al. “X-Ray Microscopy at Sirius, the New Brazilian Synchrotron Light Source”, *Microscopy and Microanalysis*, v. 24, n. S2, pp. 176–179, 2018. 20
- [78] COSTA, G., LOPES ARCHILHA, N., VASCONCELOS, G., et al. “Automation Solutions and Prototypes for the X-Ray Tomography Beamline of Sirius, the New Brazilian Synchrotron Light Source”, 2018. 20
- [79] GRUPEN, C., SHWARTZ, B. *Particle detectors*, v. 26. Cambridge university press, 2008. 21
- [80] KNOLL, G. F. *Radiation detection and measurement*. John Wiley & Sons, 2010. 21
- [81] WIKIPEDIA, THE FREE ENCYCLOPEDIA. “Tree diagram showing the relationship between types and classification of most common particle detectors”. 2009. Disponível em: <https://commons.wikimedia.org/wiki/File:Detectors_summary_3.png>. [Online; Acessado em 10 de junho, 2019]. 22
- [82] “How a detector works”. Disponível em: <<https://home.cern/science/experiments/how-detector-works>>. Acesso em: 10 de novembro de 2019. 21
- [83] STRANDLIE, A., FRÜHWIRTH, R. “Track and vertex reconstruction: From classical to adaptive methods”, *Reviews of Modern Physics*, v. 82, n. 2, pp. 1419, 2010. 22
- [84] BLUM, W., RIEGLER, W., ROLANDI, L. *Particle detection with drift chambers*. Springer Science & Business Media, 2008. 22

- [85] TURALA, M. “Silicon tracking detectors—historical overview”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 541, n. 1-2, pp. 1–14, 2005. 22
- [86] “The Inner Detector”. Disponível em: <<https://atlas.cern/discover/detector/inner-detector>>. Acesso em: 4 de agosto de 2019. 22, 34, 38
- [87] KOUZES, R. T., GEELHOOD, B. D. *Composite signatures of nuclear and non-nuclear technologies for weapons material and component measurement*. Relatório técnico, Pacific Northwest National Lab., Richland, WA (US), 2002. 22
- [88] KÖHN, C., EBERT, U. “The structure of ionization showers in air generated by electrons with 1 MeV energy or less”, *Plasma Sources Science and Technology*, v. 23, n. 4, pp. 045001, 2014. 23
- [89] “ATLAS”. . Disponível em: <<http://atlas.ch/>>. Acesso em: 17 de novembro 2019. 28
- [90] FRASS, W., WLACZAK, R. “Particle detectors”. 2009. 28
- [91] “Where the web was born”. Disponível em: <<https://home.cern/science/computing/where-web-was-born>>. Acesso em: 16 de novembro de 2019. 30
- [92] “Using CERN magnet technology in innovative cancer treatment”. Disponível em: <<https://home.cern/news/news/knowledge-sharing/using-cern-magnet-technology-innovative-cancer-treatment>>. Acesso em: 16 de novembro de 2019. 30
- [93] “Root - Data Analysis Framework”. Disponível em: <<https://root.cern.ch/>>. Acesso em: 16 de novembro de 2019. 30
- [94] EVANS, L., BRYANT, P. “LHC machine”, *Journal of instrumentation*, v. 3, n. 08, pp. S08001, 2008. 30, 31
- [95] *LEP design report*. Geneva, CERN, 1984. Disponível em: <<https://cds.cern.ch/record/102083>>. Copies shelved as reports in LEP, PS and SPS libraries. 31
- [96] COLLABORATION, C., OTHERS. “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, *arXiv preprint arXiv:1407.0558*, 2014. 31

- [97] CHATRCHYAN, S., KHACHATRYAN, V., SIRUNYAN, A., et al. “Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states”, *Journal of High Energy Physics*, v. 2014, n. 1, pp. 96, 2014. 31
- [98] AAD, G., ABBOTT, B., ABDALLAH, J., et al. “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”, *Journal of high energy physics*, v. 2015, n. 4, pp. 117, 2015. 31
- [99] MOBS, E. “The CERN accelerator complex. Complexe des accélérateurs du CERN”, Jul 2016. Disponível em: <<https://cds.cern.ch/record/2197559>>. General Photo. 32
- [100] “A vacuum as empty as interstellar space”. Disponível em: <<https://home.cern/science/engineering/vacuum-empty-interstellar-space>>. Acesso em: 29 de julho de 2019. 31
- [101] ADRIANI, O., BONECHI, L., BONGI, M., et al. “The LHCf detector at the CERN large hadron collider”, *Journal of Instrumentation*, v. 3, n. 08, pp. S08006, 2008. 33
- [102] ANELLI, G., ANTCHEV, G., ASPELL, P., et al. “The totem experiment at the cern large hadron collider”, *Journal of Instrumentation*, v. 3, n. 08, pp. S08007, 2008. 33
- [103] PINFOLD, J. L. “The MoEDAL experiment at the LHC”. In: *Astroparticle, Particle, Space Physics and Detectors for Physics Applications*, World Scientific, pp. 515–524, 2012. 33
- [104] AAMODT, K., QUINTANA, A. A., ACHENBACH, R., et al. “The ALICE experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. 08, pp. S08002, 2008. 33
- [105] “The matter-antimatter asymmetry problem”. Disponível em: <<https://home.cern/science/physics/matter-antimatter-asymmetry-problem>>. Acesso em: 29 de julho de 2019. 33
- [106] “LHCb”. Disponível em: <<https://home.cern/science/experiments/lhcb>>. Acesso em: 29 de julho de 2019. 33
- [107] COLLABORATION, A., OTHERS. “The ATLAS experiment at the CERN large hadron collider”. 2008. 33, 34, 37, 38

- [108] COLLABORATION, C., OTHERS. “The CMS experiment at the CERN LHC”. 2008. 33
- [109] “Calorimeter”. Disponível em: <<https://atlas.cern/discover/detector/calorimeter>>. Acesso em: 4 de agosto de 2019. 34
- [110] “Muon Spectrometer”. . Disponível em: <<https://atlas.cern/discover/detector/muon-spectrometer>>. Acesso em: 4 de agosto de 2019. 34
- [111] “Trigger and Data Acquisition System”. Disponível em: <<https://atlas.cern/discover/detector/trigger-daq>>. Acesso em: 4 de agosto de 2019. 34
- [112] “Software and Computing”. . Disponível em: <<https://atlas.cern/discover/detector/software-computing>>. Acesso em: 4 de agosto de 2019. 35
- [113] “Software and Computing”. Disponível em: <<https://atlas.cern/discover/detector/software-computing>>. Acesso em: 17 de novembro 2019. 35
- [114] AAD, G., BENTVELSEN, S., BOBBINK, G. J., et al. “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST*, v. 3, pp. S08003. 437 p, 2008. Disponível em: <[http://cds.cern.ch/record/1129811',note=>](http://cds.cern.ch/record/1129811)>. 35, 45, 76
- [115] AAD, G., BUTTERWORTH, J., THION, J., et al. “The ATLAS experiment at the CERN large hadron collider”, *Jinst*, v. 3, pp. S08003, 2008. 36
- [116] *ATLAS inner detector: Technical Design Report, 1*. Relatório técnico, Geneva, 1997. Disponível em: <<http://cds.cern.ch/record/331063>>. 36
- [117] HAYWOOD, S., ROSSI, L., NICKERSON, R., et al. *ATLAS inner detector: Technical Design Report, 2*. Relatório técnico, Geneva, 1997. Disponível em: <<http://cds.cern.ch/record/331064>>. 36
- [118] CAPEANS, M., FLICK, T., VUILLERMET, R., et al. *ATLAS insertable B-layer technical design report*. Relatório técnico, ATLAS-TDR-019, 2010. 37
- [119] VOGEL, A. *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*. Relatório Técnico ATL-INDET-PROC-2013-005, CERN, Geneva, Apr 2013. Disponível em: <<https://cds.cern.ch/record/1537991>>. 39

- [120] *ATLAS calorimeter performance: Technical Design Report*. Relatório técnico, Geneva, 1996. Disponível em: <<http://cds.cern.ch/record/331059>>. 39, 42
- [121] *ATLAS liquid-argon calorimeter: Technical Design Report*. Relatório técnico, Geneva, 1996. Disponível em: <<http://cds.cern.ch/record/331061>>. 39
- [122] PEQUENAO, J. *Computer Generated image of the ATLAS calorimeter*. Relatório técnico, 2008. 40, 43
- [123] SIMAS, E. *Análise não-linear de componentes independentes para uma filtragem online baseada em calorimetria de alta energia e com fina segmentação*. Tese de Doutorado, PhD thesis, PEE/COPPE/UFRJ, 2010. 41, 69
- [124] AAD, G., ABBOTT, B., ABDALLAH, J., et al. “Readiness of the ATLAS Tile Calorimeter for LHC collisions”, *The European Physical Journal C*, v. 70, n. 4, pp. 1193–1236, 2010. 43
- [125] GINGRICH, D. “Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter”, *Journal of Instrumentation*, v. 2, n. 05, pp. P05005, 2007. 44
- [126] ARTAMONOV, A., BAILEY, D., BELANGER, G., et al. “The ATLAS forward calorimeter”, *Journal of Instrumentation*, v. 3, n. 02, pp. P02010, 2008. 44
- [127] “Muon Spectrometer”. . Disponível em: <<https://atlas.cern/discover/detector/muon-spectrometer>>. Acesso em: 3 de novembro de 2019. 44
- [128] *ATLAS muon spectrometer: Technical Design Report*. Relatório técnico, Geneva, 1997. Disponível em: <<http://cds.cern.ch/record/331068>>. 44
- [129] ACHENBACH, R., ADRAGNA, P., ANDREI, V., et al. “The ATLAS level-1 calorimeter trigger”, *Journal of Instrumentation*, v. 3, n. 03, pp. P03001, 2008. 47
- [130] DA FONSECA PINTO, J. V. “Ring-shaped Calorimetry Information for a Neural Egamma Identification with ATLAS Detector”, *J. Phys. Conf. Ser.*, v. 762, n. 1, pp. 012049, 2016. doi: 10.1088/1742-6596/762/1/012049. 48

- [131] ALISON, J., BRENDLINGER, K., HEIM, S., et al. “Description and Performance of the Electron Likelihood Tool at ATLAS using 2012 LHC Data”, *ATLAS Internal Communication ATL-COM-PHYS-2013-378*, 2013. 49, 51
- [132] *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*. Relatório Técnico ATLAS-CONF-2016-024, CERN, Geneva, Jun 2016. Disponível em: <<https://cds.cern.ch/record/2157687>>. 50, 72
- [133] SEIXAS, J., CALOBA, L., SOUZA, M., et al. “Neural second-level trigger system based on calorimetry”, *Computer Physics Communications*, v. 95, n. 2-3, pp. 143–157, 1996. 51
- [134] SPOLIDORO FREUND, W., OTHERS. *An Ensemble of Neural Networks for Online Electron Filtering at the ATLAS Experiment*. Relatório técnico, ATL-COM-DAQ-2019-067, 2019. 52
- [135] HAYKIN, S. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994. 52
- [136] FARINA, M., NAKAI, Y., SHIH, D. “Searching for New Physics with Deep Autoencoders”, *arXiv preprint arXiv:1808.08992*, 2018. 53, 68
- [137] BENGIO, Y., COURVILLE, A., VINCENT, P. “Representation learning: A review and new perspectives”, *IEEE transactions on pattern analysis and machine intelligence*, v. 35, n. 8, pp. 1798–1828, 2013. 54
- [138] SCHMIDHUBER, J. “Deep learning in neural networks: An overview”, *Neural networks*, v. 61, pp. 85–117, 2015. 55
- [139] BALDI, P., HORNIK, K. “Neural networks and principal component analysis: Learning from examples without local minima”, *Neural networks*, v. 2, n. 1, pp. 53–58, 1989. 55
- [140] BOURLARD, H., KAMP, Y. “Auto-association by multilayer perceptrons and singular value decomposition”, *Biological cybernetics*, v. 59, n. 4-5, pp. 291–294, 1988. 55
- [141] KRAMER, M. A. “Nonlinear principal component analysis using autoassociative neural networks”, *AIChE journal*, v. 37, n. 2, pp. 233–243, 1991. 55

- [142] SANGER, T. D. “Optimal unsupervised learning in a single-layer linear feed-forward neural network”, *Neural networks*, v. 2, n. 6, pp. 459–473, 1989. 56
- [143] CYBENKO, G. “Approximation by superpositions of a sigmoidal function”, *Mathematics of control, signals and systems*, v. 2, n. 4, pp. 303–314, 1989. 56
- [144] GOODFELLOW, I., BENGIO, Y., COURVILLE, A., et al. *Deep learning*, v. 1. MIT press Cambridge, 2016. 57, 59
- [145] ACKLEY, D. H., HINTON, G. E., SEJNOWSKI, T. J. “A learning algorithm for Boltzmann machines”, *Cognitive science*, v. 9, n. 1, pp. 147–169, 1985. 58
- [146] YANN, L. *Modeles connexionnistes de l'apprentissage*. Tese de Doutorado, PhD thesis, These de Doctorat, Universite Paris 6, 1987. 58
- [147] HINTON, G. E., ZEMEL, R. S. “Autoencoders, minimum description length and Helmholtz free energy”. In: *Advances in neural information processing systems*, pp. 3–10, 1994. 58
- [148] BALLARD, D. H. “Modular Learning in Neural Networks.” In: *AAAI*, pp. 279–284, 1987. 58
- [149] RIFAI, S., VINCENT, P., MULLER, X., et al. “Contractive auto-encoders: Explicit invariance during feature extraction”. In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 833–840. Omnipress, 2011. 58
- [150] ALAIN, G., BENGIO, Y. “What regularized auto-encoders learn from the data-generating distribution”, *The Journal of Machine Learning Research*, v. 15, n. 1, pp. 3563–3593, 2014. 58
- [151] HINTON, G. E., SALAKHUTDINOV, R. R. “Reducing the dimensionality of data with neural networks”, *science*, v. 313, n. 5786, pp. 504–507, 2006. 59
- [152] BENGIO, Y., LAMBLIN, P., POPOVICI, D., et al. “Greedy layer-wise training of deep networks”. In: *Advances in neural information processing systems*, pp. 153–160, 2007. 59
- [153] POULTNEY, C., CHOPRA, S., CUN, Y. L., et al. “Efficient learning of sparse representations with an energy-based model”. In: *Advances in neural information processing systems*, pp. 1137–1144, 2007. 59

- [154] LAMBLIN, P., BENGIO, Y. “Important gains from supervised fine-tuning of deep architectures on large labeled sets”. In: *NIPS* 2010 Deep Learning and Unsupervised Feature Learning Workshop*, pp. 1–8, 2010. 59
- [155] SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., et al. “Dropout: a simple way to prevent neural networks from overfitting”, *The Journal of Machine Learning Research*, v. 15, n. 1, pp. 1929–1958, 2014. 60
- [156] IOFFE, S., SZEGEDY, C. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, *arXiv preprint arXiv:1502.03167*, 2015. 60
- [157] NOUSI, P., TEFAS, A. “Deep learning algorithms for discriminant autoencoding”, *Neurocomputing*, v. 266, pp. 325–335, 2017. 60
- [158] SUTSKEVER, I., VINYALS, O., LE, Q. V. “Sequence to sequence learning with neural networks”. In: *Advances in neural information processing systems*, pp. 3104–3112, 2014. 62, 63
- [159] SRIVASTAVA, N., MANSIMOV, E., SALAKHUDINOV, R. “Unsupervised learning of video representations using lstms”. In: *International conference on machine learning*, pp. 843–852, 2015. 62
- [160] SEE, A., LIU, P. J., MANNING, C. D. “Get to the point: Summarization with pointer-generator networks”, *arXiv preprint arXiv:1704.04368*, 2017. 62
- [161] RAMA, T., ÇÖLTEKIN, Ç. “LSTM autoencoders for dialect analysis”. In: *Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3)*, pp. 25–32, 2016. 62
- [162] FREITAG, M., AMIRIPARIAN, S., PUGACHEVSKIY, S., et al. “audeep: Unsupervised learning of representations from audio with deep recurrent neural networks”, *The Journal of Machine Learning Research*, v. 18, n. 1, pp. 6340–6344, 2017. 63
- [163] MARCHI, E., VESPERINI, F., SQUARTINI, S., et al. “Deep recurrent neural network-based autoencoders for acoustic novelty detection”, *Computational intelligence and neuroscience*, v. 2017, 2017. 63
- [164] HAN, J., ZHANG, Z., RINGEVAL, F., et al. “Reconstruction-error-based learning for continuous emotion recognition in speech”. In: *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 2367–2371. IEEE, 2017. 63

- [165] GUEST, D., CRANMER, K., WHITESON, D. “Deep learning and its application to LHC physics”, *Annual Review of Nuclear and Particle Science*, v. 68, pp. 161–181, 2018. 64, 66
- [166] ALBERTSSON, K., ALTOE, P., ANDERSON, D., et al. “Machine learning in high energy physics community white paper”. In: *Journal of Physics: Conference Series*, v. 1085, p. 022008. IOP Publishing, 2018. 65
- [167] COLLABORATION, A., OTHERS. “Track reconstruction performance of the ATLAS inner detector”, *CERN, Geneva, Tech. Rep. ATLPHYS-PUB-2015-018*, 2015. 65
- [168] COLLABORATION, A., OTHERS. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC run 2,(2017)”, *arXiv preprint arXiv:1704.07983*. 65
- [169] DE CIAN, M., STAHL, S., SEYFERT, P., et al. *Fast neural-net based fake track rejection*. Relatório técnico, 2017. 65
- [170] FARRELL, S., ANDERSON, D., CALAFIURA, P., et al. “The HEP. TrkX Project: deep neural networks for HL-LHC online and offline tracking”. In: *EPJ Web of Conferences*, v. 150, p. 00003. EDP Sciences, 2017. 65
- [171] LARKOSKI, A. J., MOULT, I., NACHMAN, B. “Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning”, *arXiv preprint arXiv:1709.04464*, 2017. 66
- [172] LÖNNBLAD, L., PETERSON, C., RÖGNVALSSON, T. “Pattern recognition in high energy physics with artificial neural networks—JETNET 2.0”, *Computer Physics Communications*, v. 70, n. 1, pp. 167–182, 1992. 66
- [173] PETERSON, C., RÖGNVALDSSON, T., LÖNNBLAD, L. “JETNET 3.0—A versatile artificial neural network package”, *Computer Physics Communications*, v. 81, n. 1-2, pp. 185–220, 1994. 66
- [174] COGAN, J., KAGAN, M., STRAUSS, E., et al. “Jet-images: computer vision inspired techniques for jet tagging”, *Journal of High Energy Physics*, v. 2015, n. 2, pp. 118, 2015. 66
- [175] DE OLIVEIRA, L., KAGAN, M., MACKEY, L., et al. “Jet-images—deep learning edition”, *Journal of High Energy Physics*, v. 2016, n. 7, pp. 69, 2016. 66
- [176] ALWALL, J., HERQUET, M., MALTONI, F., et al. “MadGraph 5: going beyond”, *Journal of High Energy Physics*, v. 2011, n. 6, pp. 128, 2011. 67

- [177] SJÖSTRAND, T., MRENNNA, S., SKANDS, P. “PYTHIA 6.4 physics and manual”, *Journal of High Energy Physics*, v. 2006, n. 05, pp. 026, 2006. 67
- [178] OVYN, S., ROUBY, X., LEMAITRE, V. “DELPHES, a framework for fast simulation of a generic collider experiment”, *arXiv preprint arXiv:0903.2225*, 2009. 67
- [179] CHENG, T. “Recursive neural networks in quark/gluon tagging”, *Computing and Software for Big Science*, v. 2, n. 1, pp. 3, 2018. 67
- [180] COLLABORATION, A., OTHERS. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Relatório técnico, ATL-PHYS-PUB-2017-013, 2017. 67
- [181] COLLABORATION, C., OTHERS. “Heavy flavor identification at CMS with deep neural networks”, *CMS Detector Performance Summary CMS-DP-2017-005*, 2017. 67
- [182] EGAN, S., FEDORKO, W., LISTER, A., et al. “Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC”, *arXiv preprint arXiv:1711.09059*, 2017. 67
- [183] LOUPPE, G., CHO, K., BECOT, C., et al. “QCD-aware recursive neural networks for jet physics”, *arXiv preprint arXiv:1702.00748*, 2017. 67
- [184] COLLABORATION, A., OTHERS. “Identification of jets containing b-hadrons with recurrent neural networks at the ATLAS experiment”, *ATLAS note: ATL-PHYS-PUB-2017-003*, <http://cds.cern.ch/record/2255226>, 2017. 67
- [185] BALDI, P., SADOWSKI, P., WHITESON, D. “Searching for exotic particles in high-energy physics with deep learning”, *Nature communications*, v. 5, pp. 4308, 2014. 68
- [186] HEIMEL, T., KASIECZKA, G., PLEHN, T., et al. “QCD or What?” *arXiv preprint arXiv:1808.08979*, 2018. 68
- [187] HAJER, J., LI, Y.-Y., LIU, T., et al. “Novelty Detection Meets Collider Physics”, *arXiv preprint arXiv:1807.10261*, 2018. 68
- [188] RACAH, E., KO, S., SADOWSKI, P., et al. “Revealing fundamental physics from the Daya Bay neutrino experiment using deep neural networks”. In: *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, pp. 892–897. IEEE, 2016. 68

- [189] NARDELLI, A. “New physics detection in High Energy Physics experiments through unsupervised Deep Learning”, 2018. 68
- [190] “Anomaly detection using Deep Autoencoders for the assessment of the quality of the data acquired by the CMS experiment”, 2018. 68
- [191] LIMA JR, H., SEIXAS, J. “A segmented principal component analysis applied to calorimetry information at ATLAS”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 559, n. 1, pp. 129–133, 2006. 69
- [192] SIMAS FILHO, E. F., SEIXAS, J. M., CALÔBA, L. P. “Optimized calorimeter signal compaction for an independent component based ATLAS electron/jet second-level trigger”, *PoS*, p. 102, 2008. 69
- [193] SEIXAS, J., CALOBA, L., SIMAS, E. “ATLAS Second-Level Electron/Jet Neural Discriminator based on Nonlinear Independent Components”. In: *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, v. 93, p. 052. SISSA Medialab, 2011. 69
- [194] TORRES, R. C., DE LIMA, D. E. F., DE SIMAS FILHO, E. F., et al. “Neural online filtering based on preprocessed calorimeter data”. In: *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, pp. 530–536. IEEE, 2009. 69
- [195] TORRES, R. “Sistema online de filtragem em um ambiente com alta taxa de eventos e fina granularidade”, *Rio de Janeiro, UFRJ/COPPE*, 2010. 69, 75, 82
- [196] CIODARO, T. *Identificação online de sinais baseada em calorimetria de altas energias e fina segmentação*. Tese de Doutorado, Rio de Janeiro Federal U., 2012. 69
- [197] DOS ANJOS, A. *Sistema Online de Filtragem em um Ambiente com Alta Taxa de Eventos*. Tese de Doutorado, PhD thesis, PEE/COPPE/UFRJ, 2006. 69
- [198] FREUND, W. *Identificação de Elétrons Baseada em um Calorímetro de Altas Energias Finamente Segmentado*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2018. 69
- [199] AY, C., BUCKLEY, A., BUTTERWORTH, J., et al. “Monte Carlo generators in ATLAS software”. In: *Journal of Physics: Conference Series*, v. 219, p. 032001. IOP Publishing, 2010. 71

- [200] SJÖSTRAND, T., MRENNNA, S., SKANDS, P. “A brief introduction to PYTHIA 8.1”, *Computer Physics Communications*, v. 178, n. 11, pp. 852–867, 2008. 71
- [201] BÄHR, M., GIESEKE, S., GIGG, M. A., et al. “Herwig++ physics and manual”, *The European Physical Journal C*, v. 58, n. 4, pp. 639–707, 2008. 71
- [202] AGOSTINELLI, S., ALLISON, J., AMAKO, K. A., et al. “GEANT4—a simulation toolkit”, *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 506, n. 3, pp. 250–303, 2003. 71
- [203] CALAFIURA, P., MARINO, M., LEGGETT, C., et al. “The athena control framework in production, new developments and lessons learned”, 2005. 71
- [204] BRUN, R., RADEMAKERS, F., CANAL, P., et al. *ROOT: An Object-Oriented Data Analysis Framework: Users Guide 5.26*. Geneva, CERN, 2009. 71
- [205] KOHAVI, R., OTHERS. “A study of cross-validation and bootstrap for accuracy estimation and model selection”. In: *Ijcai*, v. 14, pp. 1137–1145. Montreal, Canada, 1995. 74
- [206] EFRON, B., GONG, G. “A leisurely look at the bootstrap, the jackknife, and cross-validation”, *The American Statistician*, v. 37, n. 1, pp. 36–48, 1983. 74
- [207] KINGMA, D. P., BA, J. “Adam: A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014. 77
- [208] COVER, T. M., THOMAS, J. A. *Elements of information theory*. John Wiley & Sons, 2012. 79
- [209] JOE, H. “Relative entropy measures of multivariate dependence”, *Journal of the American Statistical Association*, v. 84, n. 405, pp. 157–164, 1989. 79
- [210] KULLBACK, S. “Letter to the editor: The Kullback-Leibler distance”, 1987.
- [211] HYVÄRINEN, A., OJA, E. “Independent component analysis: algorithms and applications”, *Neural networks*, v. 13, n. 4-5, pp. 411–430, 2000. 79

- [212] DOS ANJOS, A., TORRES, R., SEIXAS, J. D., et al. “Neural triggering system operating on high resolution calorimetry information”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 559, n. 1, pp. 134–138, 2006. 81

Apêndice A

Detalhes de Treinamento

Neste apêndice, serão apresentados alguns detalhes do treinamento das redes auto-codificadoras. A infraestrutura utilizada para realização dos treinamentos consistiu em um computador com CPU Intel i7 8700k, GPU GeForce GTX 1080ti e 64GB de RAM.

As Figuras A.1, A.2, A.3 e A.4 apresentam o erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 1 camada, 3, 5 e 8, respectivamente. Como o treinamento utilizou parada antecipada baseada no erro de validação, os treinamentos possuem número de épocas distintos.

A Tabela A.1 apresenta a quantidade de passos, MSE mínimo e tempo de execução para 1 treinamento de uma rede neural para classificação elétron-jato (*NeuralRinger*). O número de passos é definido pela quantidade de amostras dividido pelo tamanho da batelada. Nota-se que, em alguns casos, o tamanho de batelada maior implica em mais passos devido à demora na convergência. Considerando estes resultados, o valor de 1024 foi escolhido para batelada.

A Tabela A.2 mostra um exemplo de arquitetura das redes auto-codificadoras utilizadas. Neste caso, as redes permitem compressão em duas dimensões. A lógica empregada para escolha da quantidade de neurônios nas camadas escondidas foi definida na Seção 6.4.2.2 e, portanto, pode-se utilizar este exemplo para compreender a arquitetura das demais redes utilizadas.

As Figuras A.5 e A.6 apresentam a evolução do MSE em função das épocas de treinamento da primeira camada do SAE, quando 100 neurônios são colocados na camada escondida, empregando a otimização por gradiente descendente e ADAM, respectivamente. Devido à superioridade apresentada pelo ADAM, este passou a ser utilizado.

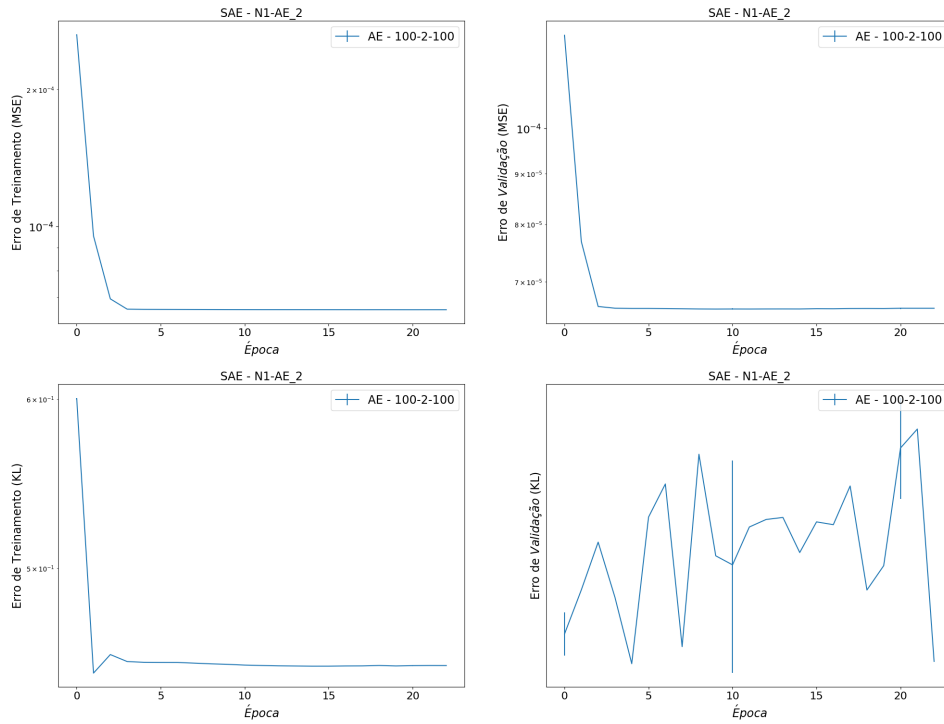


Figura A.1: Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 1 camada.

Tabela A.1: Quantidade de passos, MSE mínimo e tempo de execução para 1 treinamento de uma rede neural para classificação elétron-jato (*NeuralRinger*)

Batch Size	Passos	$MSE(Min)$	Tempo
28	1.644M	0.1237	1m0s
64	2.261M	0.1238	38s
128	4.932M	0.1238	45s
256	3.288M	0.1240	15s
512	5.754M	0.1238	15s
1024	9.248M	0.1238	16s
5000	28.98M	0.1239	27s
20000	87.34M	0.1241	1m17s
83711	323.3M	0.1241	4m25s

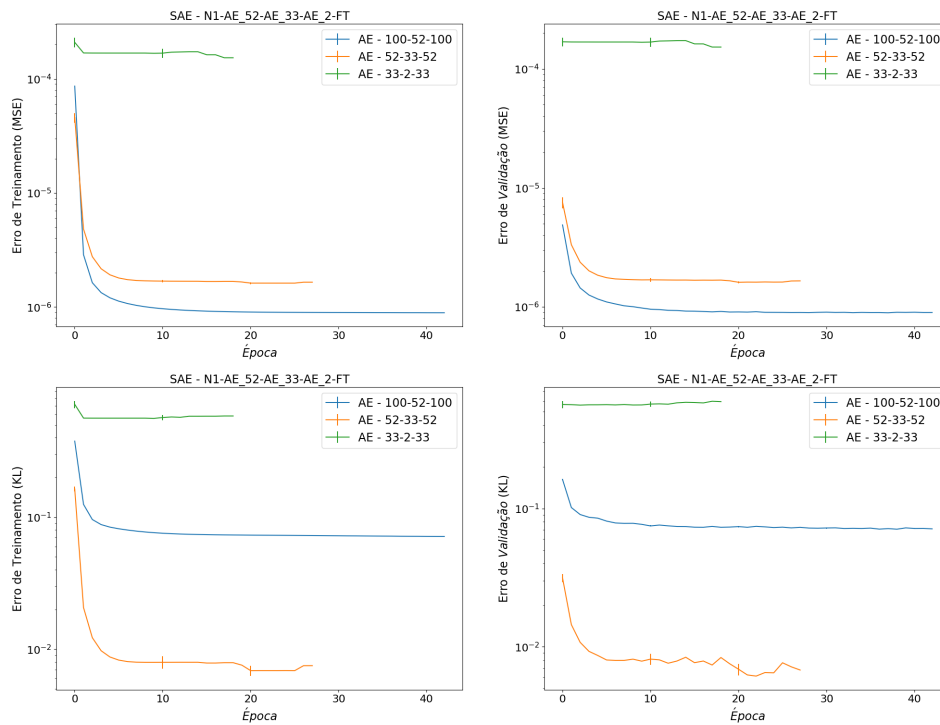


Figura A.2: Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 3 camadas.

Tabela A.2: Arquitetura de redes auto-codificadoras para compressão em 2 dimensões.

Quantidade de camadas	Arquitetura
1 Camada	100-2-100
2 Camadas	100-33-2-33-100
3 Camadas	100-52-33-2-33-52-100
5 Camadas	100-63-37-23-13-2-13-23-37-63-100
8 Camadas	100-67-57-47-37-27-17-7-2-7-17-27-37-47-57-67-100

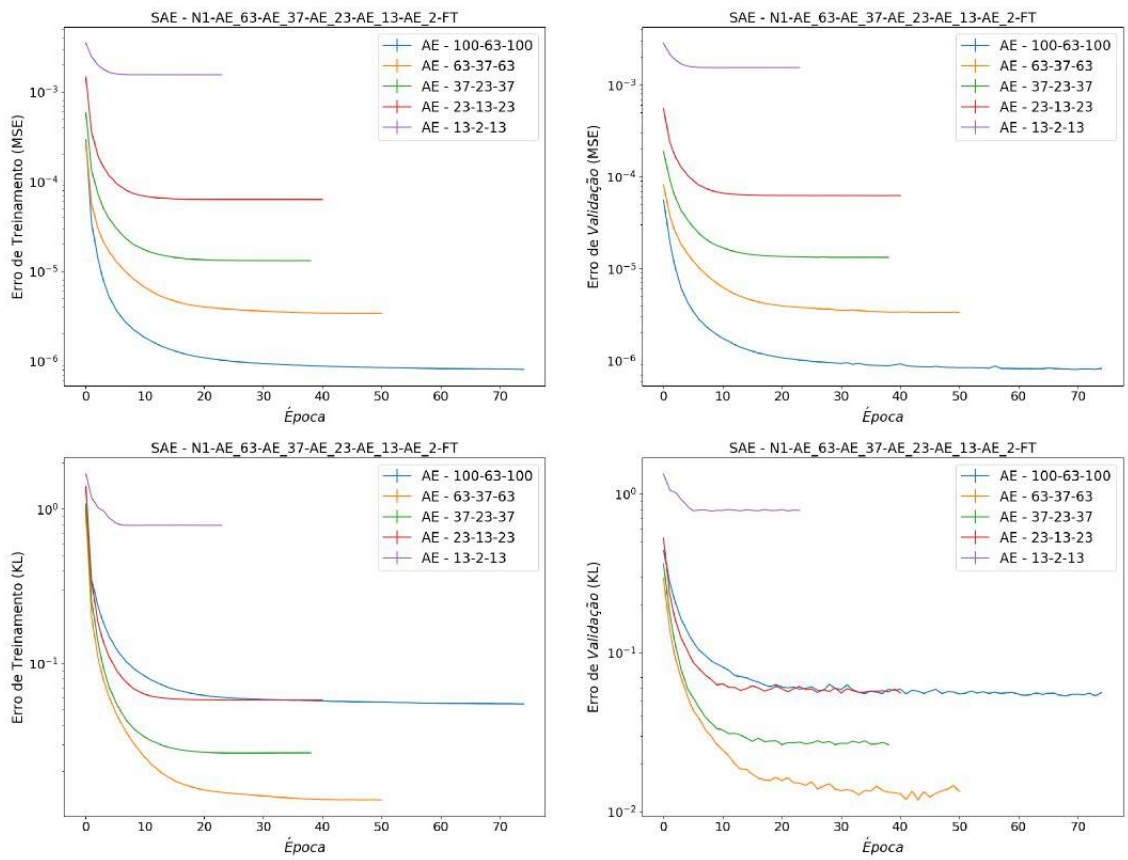


Figura A.3: Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 5 camadas.

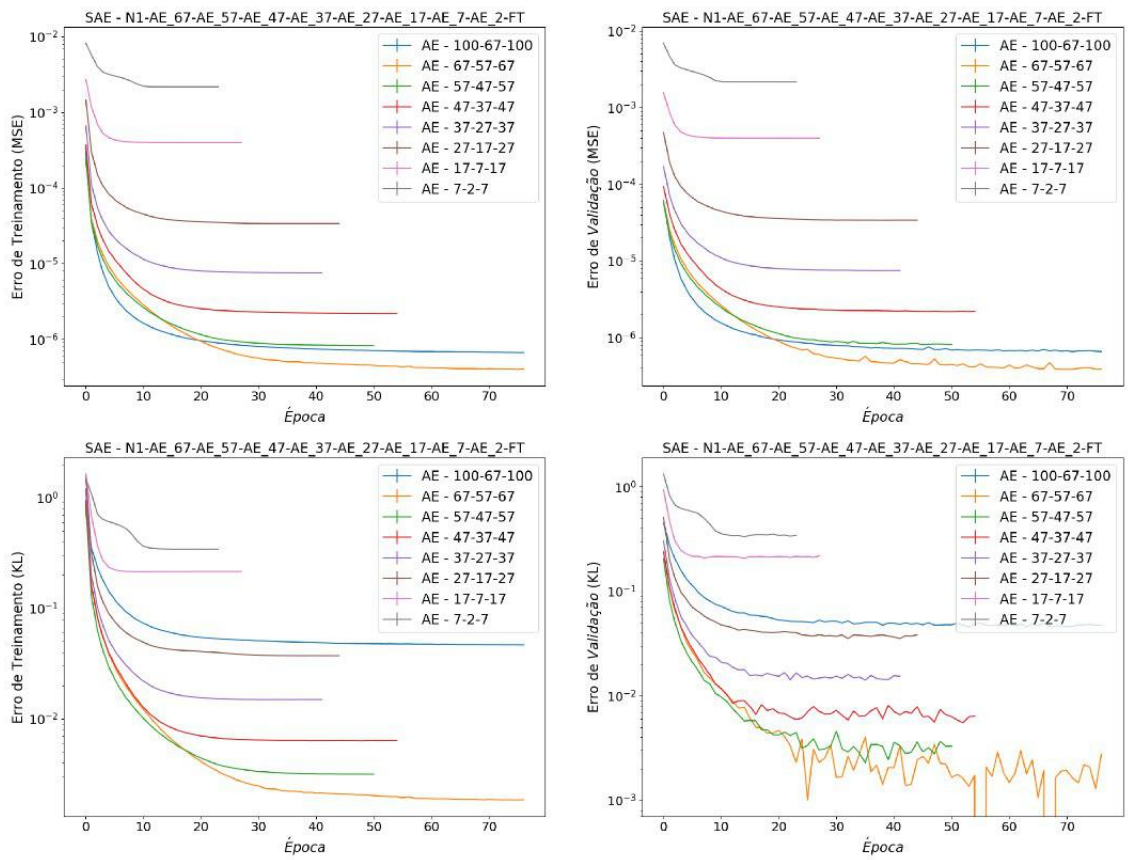


Figura A.4: Erro médio quadrático e divergência KL para o conjunto de treino e teste para a rede auto-codificadora com 8 camadas.

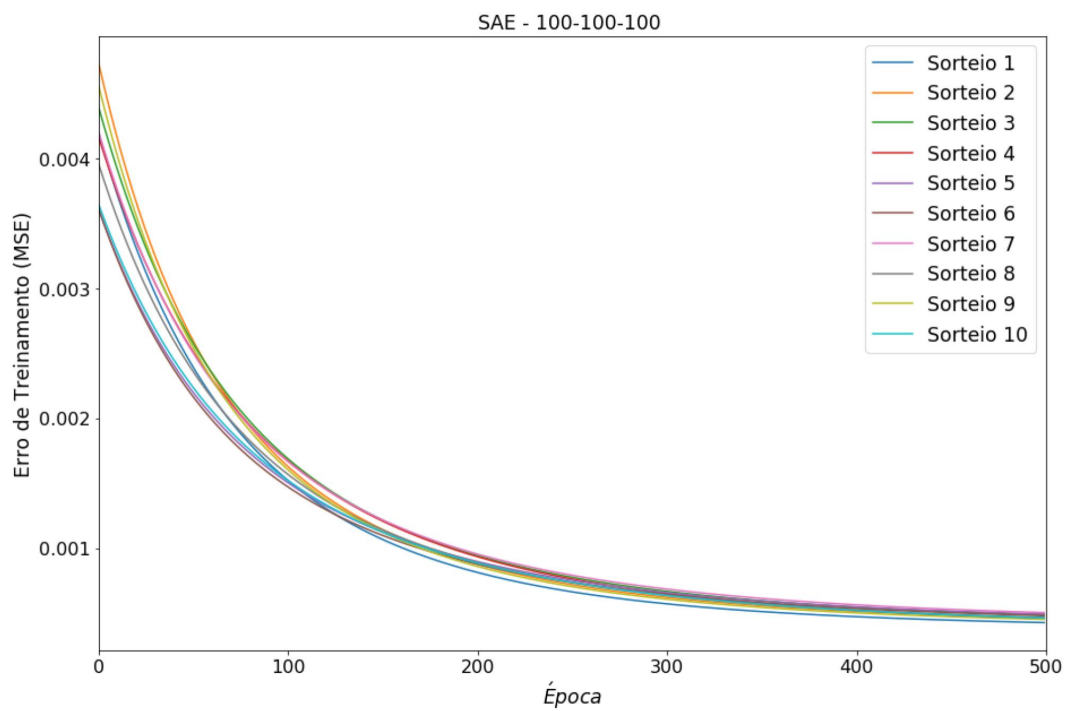


Figura A.5: Evolução do MSE em função das épocas de treinamento da primeira camada do SAE, quando 100 neurônios são colocados na camada escondida, empregando a otimização por gradiente descendente.

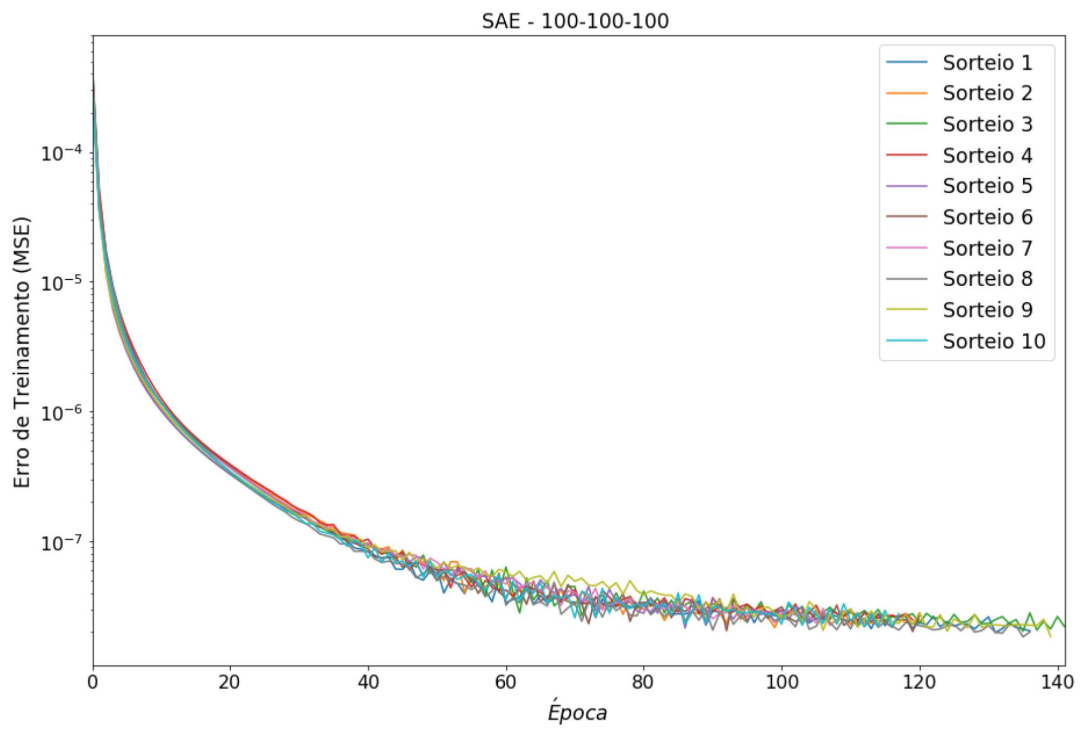


Figura A.6: Evolução do MSE em função das épocas de treinamento da primeira camada do SAE, quando 100 neurônios são colocados na camada escondida, empregando a otimização por ADAM.

Apêndice B

Gráficos de Espalhamento

Neste apêndice, serão apresentados os gráficos de espalhamento para cada anel individualmente, considerando entrada e reconstrução, obtida pela PCA (Figura B.1) e pela rede auto-codificadora (Figura B.2) quando a dimensão codificada é 27 e pela PCA (B.3) e pela rede auto-codificadora com 3 camadas (B.4) quando a dimensão codificada é 2.

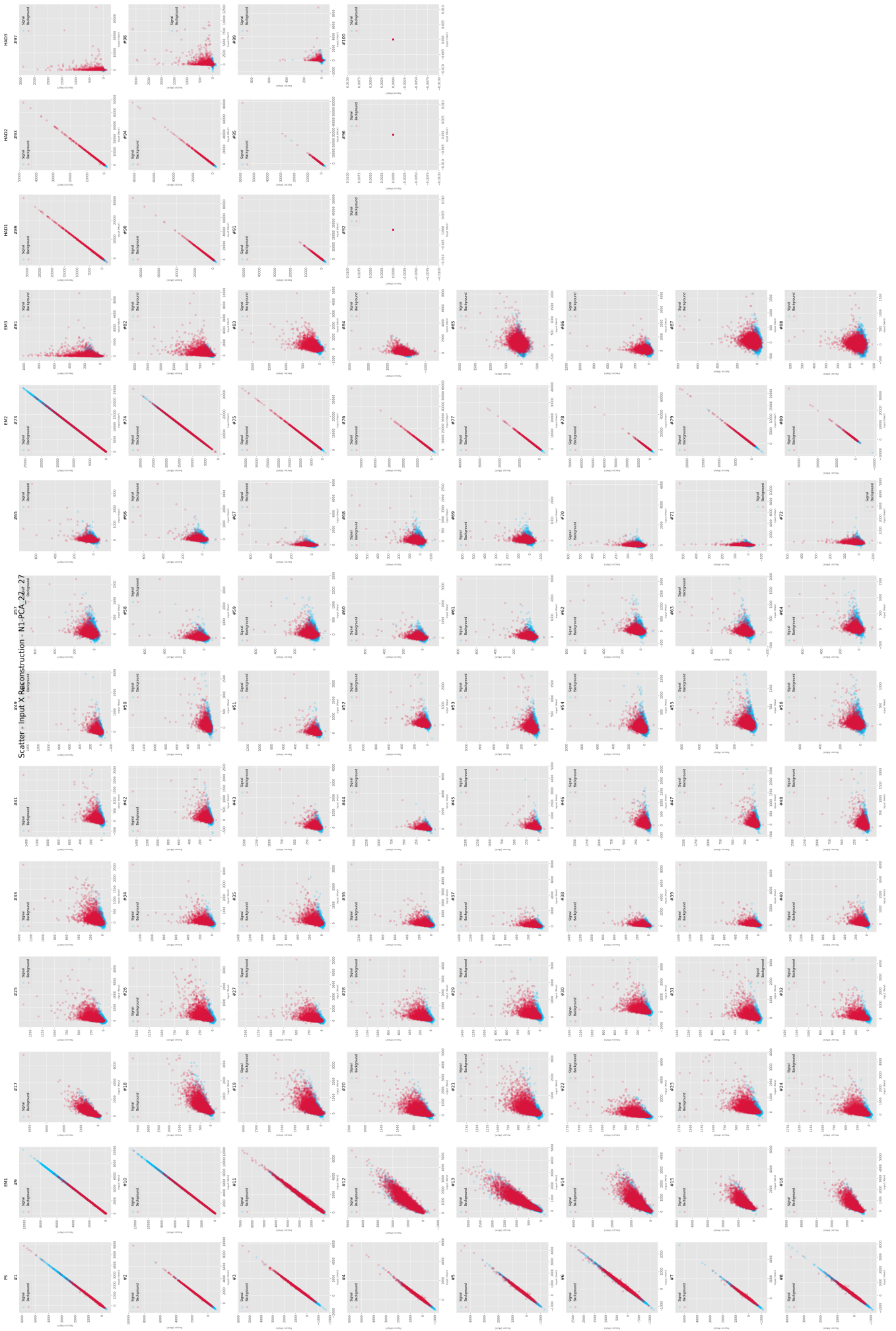


Figura B.1: Gráfico de espalhamento dos eixos de entrada e da reconstrução dos mesmos para a PCA codificada em 27 dimensões.

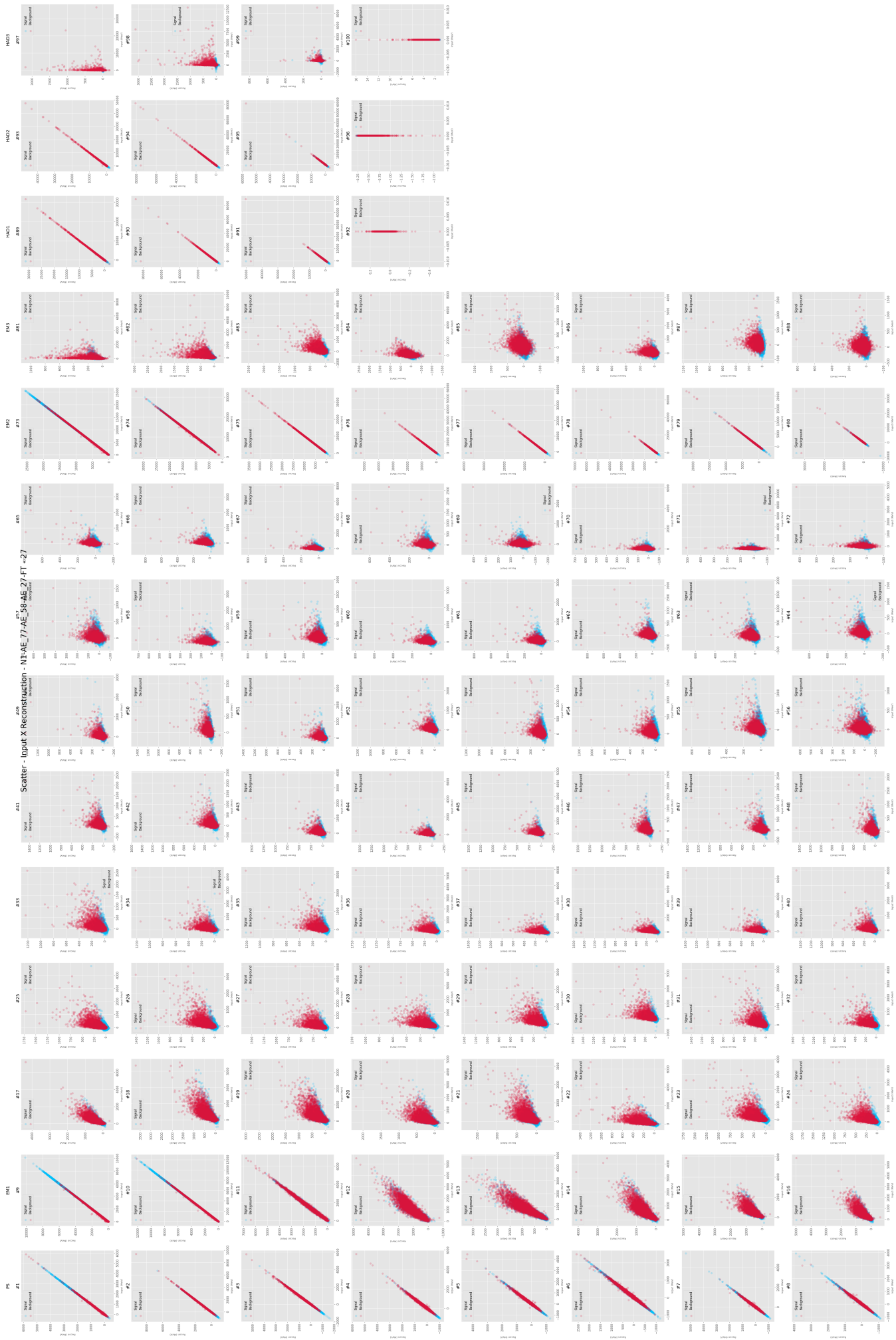


Figura B.2: Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a rede auto-codificadora profunda com 3 camadas codificadas em 27 dimensões.

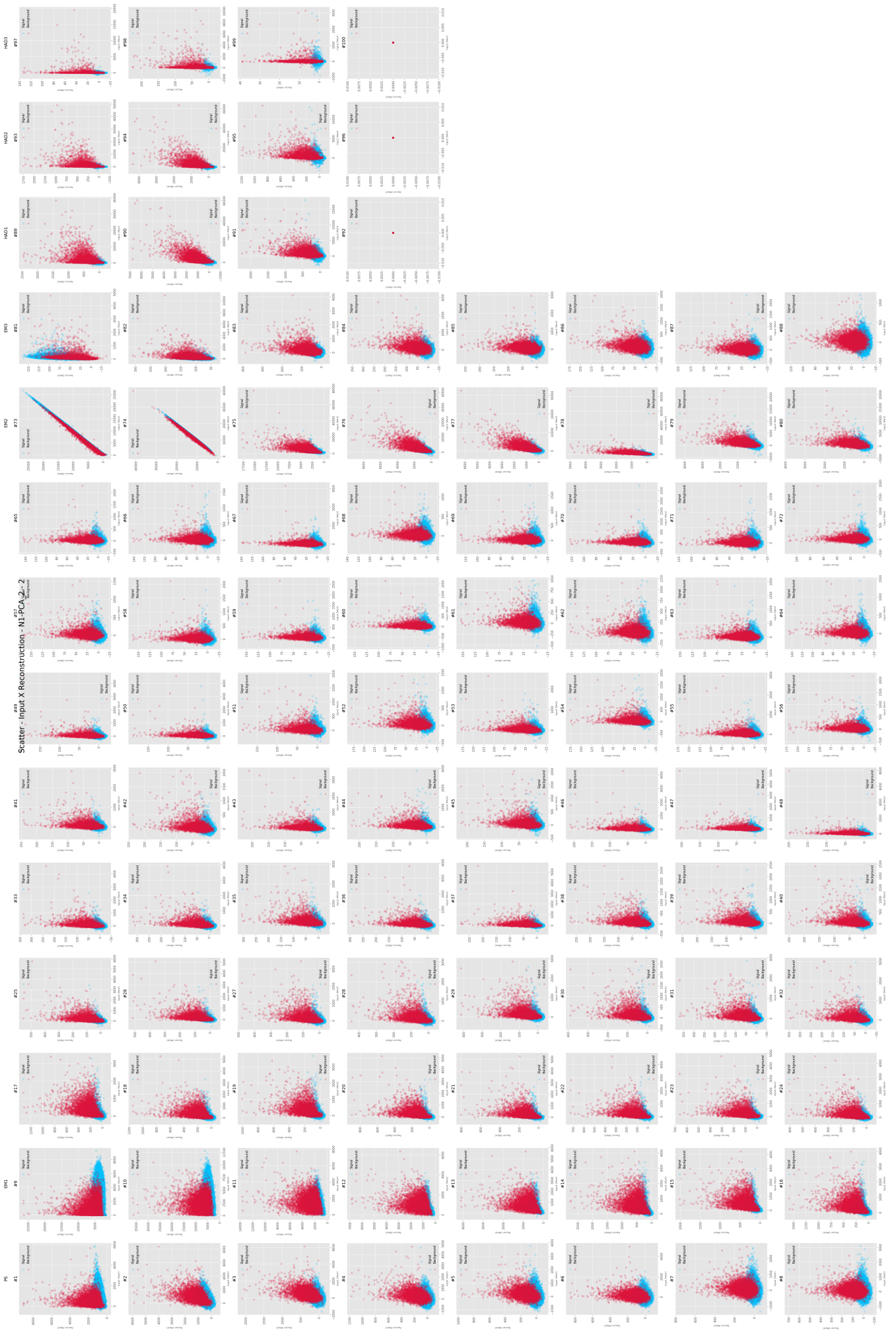


Figura B.3: Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a PCA codificada em 2 dimensões.

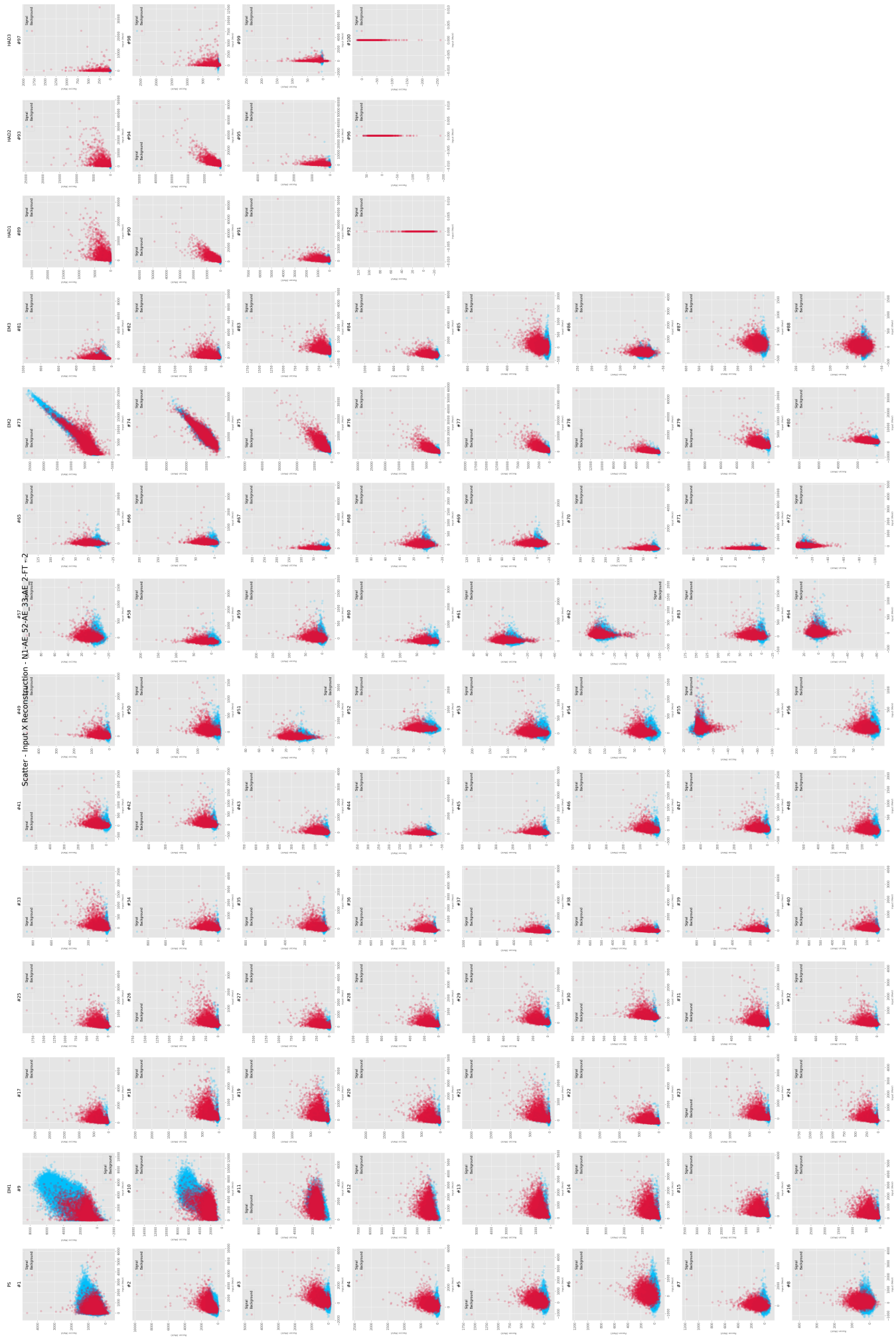


Figura B.4: Gráfico de espalhamento dos anéis de entrada e da reconstrução dos mesmos para a rede auto-codificadora profunda com 3 camadas codificadas em 2 dimensões.

Apêndice C

Produção Científica

Neste apêndice serão listados os resumos e artigos publicados em anais de congressos.

C.1 Artigos Publicados em Anais de Congresso

1. COSTA, C. E. C.; PINTO, J. V. F. ; ARAUJO, M. V. ; FREUND, W. S. ; SEIXAS, J. M. . "REPRESENTAÇÃO DO SINAL DE CALORIMETRIA DE ALTAS ENERGIAS UTILIZANDO AUTO-CODIFICADORES". In: *ENMC & ECTM 2019*, 2019, Juiz de Fora. *XXII Encontro Nacional de Modelagem Computacional e X Encontro de Ciência e Tecnologia de Materiais*, 2019.

Alguns sistemas de aquisição, medição e análise de dados possuem alta taxa de produção de eventos, uma enorme quantidade de canais de leitura e eventos raros de interesse. Tais características acarretam em dificuldade para filtragem e análise de dados devido à alta dimensionalidade no espaço original e ao grande volume gerado. Neste caso, a codificação da informação pode auxiliar com a redução da dimensionalidade e, conseqüentemente, do custo computacional. Neste contexto encontra-se o sistema de filtragem do experimento ATLAS. O algoritmo *NeuralRinger* é utilizado neste sistema para identificação de elétrons, através do uso de anéis concêntricos (total de 100) na representação da energia depositada nas células do sistema de calorimetria (medidores de energia) do detector. Este trabalho busca explorar a estatística de ordem superior para encontrar uma representação da informação anelada de calorimetria com um menor número de componentes. Diferentes arquiteturas de redes auto-codificadoras são avaliadas, com relação à sua capacidade de reconstrução dos eventos e discriminação elétron-hádron. Os resultados são apresentados utilizando-se uma base de dados de simulação de Monte Carlo, que representa as condições de colisões de prótons a uma energia de 13 TeV

no centro de massa.

2. COSTA, C. E. C.; ARAUJO, M. V. ; PINTO, J. V. F. ; FREUND, W. S. ; SEIXAS, J. M. . "Redes auto-codificadoras como pré-processamento em calorimetria de altas energias com fina segmentação." In: *XIV Congresso Brasileiro de Inteligência Computacional*, 2019, Belém. *Congresso Brasileiro de Inteligência Computacional*, 2019. v. XIV.

Ambientes que apresentam elevada taxa de amostragem, uma grande quantidade de canais de leitura e eventos raros imersos em uma grande quantidade de ruído de fundo são comuns em diversas aplicações de engenharia. Nestes ambientes, a filtragem do sinal de interesse se torna desafiadora devido à alta dimensionalidade no espaço de características original e ao grande volume de dados gerados. Na física de partículas, este contexto se faz presente, como no caso do sistema de filtragem do experimento ATLAS, situado no LHC (Large Hadron Collider), que tem como objetivo identificar partículas envolvidas em processos físicos de interesse de estudo. Desde 2017, o algoritmo *NeuralRinger*, que consiste em um conjunto de redes neurais do tipo perceptron de múltiplas camadas, é utilizado no sistema de filtragem *online* do ATLAS com o objetivo de discriminar elétrons (sinal de interesse) de jatos (ruído). O *NeuralRinger* oferece redução de dimensionalidade através da representação da energia depositada nas células do sistema de calorimetria (medidores responsáveis pela estimação da energia associada aos eventos) com o uso de anéis concêntricos (total de 100) que alimentam as redes neurais. Este trabalho avalia a aplicação de técnicas de codificação para encontrar uma representação da informação anelada de calorimetria com um menor número de componentes e capaz de manter a capacidade discriminante. Redes auto-codificadoras discriminantes se mostraram capazes de representar a informação anelada de calorimetria com apenas um componente, mantendo a capacidade de discriminação elétron-jato. Os resultados foram obtidos utilizando uma base de dados de simulação de Monte Carlo, que representa as condições de colisões de prótons a uma energia de 13 TeV no centro de massa.

3. ARAUJO, M. V. ; COSTA, C. E. C. ; PINTO, J. V. F. ; FREUND, W. S. ; SEIXAS, J. M. ; BEGALLI, M. "Avaliação da Contribuição Longitudinal de um Calorímetro Finamente Segmentado para a Filtragem Online Realizada por um Ensemble de Redes Neurais em Ambiente de Alta Taxa de Eventos." In: *XIV Congresso Brasileiro de Inteligência Computacional*, 2019, Belém. *Congresso Brasileiro de Inteligência Computacional*, 2019. v. XIV.

Em grandes experimentos, a tomada de dados pode ser feita por uma grande quantidade de canais de leitura, que são empregados de modo a obter a in-

formação de interesse registrada por uma grande quantidade de sensores. Se houver o interesse por eventos raros, o emprego de técnicas multivariadas e um sistema de filtragem que possa eliminar os eventos não interessantes de modo online pode vir a ser considerada para manter estável a tomada de dados. No experimento LHC (Large Hadron Collider) de colisão de partículas (prótons) de altas energias (13 TeV, no centro de massa), o sistema NeuralRinger utiliza um conjunto de redes neurais do tipo perceptron de múltiplas camadas, para realizar a filtragem online de eventos no ATLAS, o maior experimento do LHC, buscando detectar elétrons, cuja presença é indicador para vários processos de física nova que se deseja observar experimentalmente. O NeuralRinger utiliza-se das informações provenientes do sistema de calorímetros (equipamentos responsáveis por medir a energia das partículas) para alimentar as redes neurais e, assim, tomar a decisão de filtragem. Todavia, com o aumento da quantidade de colisões no LHC, e o fato de que partes do sistema de calorimetria necessitam de maior tempo de leitura durante a tomada de dados, uma estratégia de filtragem que possa contornar tais limitações é mandatória para alguns canais de interesse. Este trabalho avalia uma extensão do NeuralRinger para atuar na detecção de elétrons até 15 GeV e que não utiliza toda a segmentação longitudinal disponível nos calorímetros do ATLAS. Resultados indicam que, apesar de uma pequena perda de eficiência, esta opção pode vir a ser considerada quando for retomada a coleta de dados do ATLAS, em 2021.

4. ARAUJO, M. V. ; COSTA, C. E. C. ; PINTO, J. V. F. ; FREUND, W. S. ; SEIXAS, J. M. ; BEGALLI, M. . "FILTRAGEM ONLINE BASEADA EM CALORIMETRIA DE ALTAS ENERGIAS EM UM AMBIENTE COM ALTA TAXA DE EVENTOS UTILIZANDO UM ENSEMBLE DE REDES NEURASIS." In: *ENMC & ECTM 2019*, 2019, Juiz de Fora. *XXII Encontro Nacional de Modelagem Computacional e X Encontro de Ciência e Tecnologia de Materiais*, 2019.

Durante a aquisição de dados de uma dada aplicação, o emprego de diversos sensores para compor os canais de leitura tem por consequência, em geral, o registro da informação de interesse por meio destes sensores. Devido a isto, e a ao interesse por eventos raros, torna o emprego de uma análise multivariada e um sistema de filtragem que atue de forma online com abordagem mais indicada. Neste contexto está o sistema de filtragem online (Trigger) do ATLAS que utiliza, para selecionar elétrons, desde 2017, um conjunto de redes neurais do tipo perceptron de múltiplas camadas que faz uso das características de deposição de energia dos elétrons nos calorímetros (medidores de energias) para

alimentar as redes neurais, as quais tomam a decisão de filtragem. Entretanto, este enfoque de aprendizado de máquina é empregado apenas na filtragem de elétrons com energia transversa maior do que 15 GeV; abaixo deste patamar são utilizados cortes em variáveis extremamente discriminantes da calorimetria. Este trabalho propõe a extensão do processamento neural para a região abaixo de 15 GeV, na qual se utilizam processos físicos para o treinamento dos modelos neurais diferentes daqueles usados para os elétrons acima de 15 GeV.

C.2 Resumos Publicados em Anais de Congresso

1. COSTA, C. E. C. ; SEIXAS, J. M. . "Representação do Sinal de Calorimetria do ATLAS Utilizando Auto-Codificadores." In: *XLII RTFNB - XL ENFPC 2019*, 2019, Campos do Jordão, SP. *XLII Reunião de Trabalho sobre Física Nuclear no Brasil - XL Encontro Nacional de Física de Partículas e Campos*, 2019.

O NeuralRinger é um algoritmo para identificação de elétrons, que utiliza anéis concêntricos na representação da energia depositada nas células das sete camadas do sistema de calorimetria do ATLAS, e que entrou em operação no sistema de filtragem online (Trigger) de elétrons, em sua camada de alto nível (HLT), em 2017. A formação destes anéis se inicia na célula mais energética de cada camada, cuja energia é considerada o primeiro anel da camada, e continua através da soma das células externas em relação ao anel anterior, até que a Região de interesse (RoI) seja coberta ou um número máximo de anéis (100) seja atingido. Desta forma, se explora a geometria dos chuveiros formados pela interação das partículas com o calorímetro, preservando-se suas informações lateral e longitudinal, enquanto se obtém uma significativa redução de dimensionalidade na RoI. O teste de hipótese é realizado através de um ensemble de classificadores compostos por redes neurais do tipo Pércptrons de Múltiplas Camadas (MLP) que são projetadas para regiões do plano formado pela energia transversa (ET) e posição da partícula (η) para discriminação elétron-hádron. Este trabalho busca explorar a estatística de ordem superior para encontrar uma representação da informação anelada de calorimetria com um menor número de componentes, no contexto do sistema de reconstrução (SR) do experimento ATLAS. Para tanto, são avaliadas diferentes arquiteturas de redes auto-codificadoras, possibilitando compactação não-linear e extração de características latentes, com relação à sua capacidade de reconstrução dos eventos e discriminação elétron-hádron. Os resultados são apresentados utilizando-se uma base de dados de simulação de Monte Carlo que representa as condições de colisões a uma energia de 13TeV no centro de

massa em 2016, contendo decaimentos $Z \rightarrow ee$. É demonstrada a possibilidade de redução de dimensionalidade mantendo-se a capacidade discriminativa na identificação de elétrons no ambiente offline do ATLAS.

2. COSTA, C. E. C. ; FREUND, W. S. ; SEIXAS, J. M. . "Filtragem Offline de Elétrons Baseada em um Ensemble de Redes Neurais Especialistas." In: *Workshop RENAFAE 2018*, 2018, São Paulo. *Anais do Workshop RENAFAE 2018*, 2018.

O gradativo aumento de luminosidade no LHC tem exigido métodos mais eficientes para a manutenção da capacidade de seleção online e offline. Nesse sentido, o ATLAS adotou o algoritmo NeuralRinger, que possibilitou aliviar a demanda por processamento durante a filtragem online a partir de 2017. Essa estratégia, dedicada à seleção de elétrons, emprega outra representação da informação de calorimetria, baseada em 100 anéis concêntricos de energia ao baricentro de energia que é aproximado, na filtragem online, pela célula mais energética. A tomada de decisão emprega uma assembleia de redes neurais específicas por regiões de energia e posição da partícula incidente na região de precisão do experimento, o que possibilita limitar o impacto da resposta do detector no perfil dos anéis. Este trabalho avalia a atuação do NeuralRinger para a seleção offline de elétrons no experimento ATLAS. Desenvolvem-se aprimoramentos a fim de estender o método para atingir sua eficiência máxima nesse ambiente. A necessidade de explorar toda a informação discriminante proveniente do Sistema de Calorimetria e Detector Interno é atacada como um problema de fusão de dados. Nesse âmbito, são utilizadas, adicionalmente aos anéis, as representações tradicionais em grandezas físicas para o treinamento de redes especialistas que, posteriormente, alimentam, através dos neurônios da camada escondida, a entrada de outra rede neural responsável pela fusão da informação e tomada de decisão. É realizada a busca por um espaço latente mais discriminante a partir da representação dos anéis empregando pré-processamento estatístico via Análise de Componentes Independentes (ICA), conhecimento especialista por intermédio de um mapeamento não-linear com dois parâmetros otimizados por busca em grade (Ringer-Rp) e compactação não-linear cega via Stacked AutoEncoder. Como a estatística disponível para o ajuste dos modelos é extremamente volumosa, foi desenvolvida uma estratégia própria para reduzir o processamento necessário no desenvolvimento dos modelos avaliados enquanto mantendo as amostras relevantes para o ajuste. Compararam-se as propostas de extensão do NeuralRinger com a referência em atuação no ambiente offline, baseada em verossimilhança nas grandezas tradicionais, através de amostras de provas obtidas pelo método

T&P para o decaimento $Z \rightarrow ee$, e seu respectivo ruído de fundo, em colisões próton-próton ocorridas em 2016 e simulações equivalentes. Quando apenas aplicando a fusão via redes especialistas, observou-se uma redução na taxa de falsos elétrons de 2,54% para 1,13% em valor central e com incerteza de validação cruzada desprezível para ambos operando com a mesma taxa de detecção de elétrons em dados de simulação.