

SISTEMA ONLINE DE FILTRAGEM EM UM AMBIENTE COM ALTA TAXA  
DE EVENTOS

André Rabello dos Anjos

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS  
EM ENGENHARIA ELÉTRICA.

Aprovada por:

---

Prof. José Manoel de Seixas, D.Sc.

---

Prof. Fernando Mendes de Azevedo, D.Sc.

---

Prof<sup>a</sup>. Márcia Begalli, Ph.D.

---

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

---

Prof. Marcello Luiz Rodrigues de Campos, Ph.D.

---

Dr. Denis Oliveira Damazio, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2006

ANJOS, ANDRÉ RABELLO DOS

Sistema Online de Filtragem em um Ambiente com Alta Taxa de Eventos [Rio de Janeiro] 2006

XXIV, 315 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia Elétrica, 2006)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Redes Neurais 2. Física de Altas Energias 3. Calorimetria 4. Processamento Rápido 5. Processamento Distribuído

I. COPPE/UFRJ II. Título (série)

Ao **grande** amor de minha vida, Aninha: você  
é a minha luz.

### Agradecimentos:

Em especial, a meu amor, Aninha, que suportou a minha ausência nestes longos meses que passamos afastados por causa deste trabalho. À minha super-mãe Léa, que nunca pára de me apoiar, te amo. Aos meus irmãos Guga e Vá, pela amizade eterna que nutrimos uns pelos outros. Ao meu pai, por seu bom humor de sempre e nossos almoços aos sábados. À Ira, nossa sempre-amiga e mãe. Aos meus pais postigos, Dedé e Sidney, pela filha e pelo carinho de sempre. A meu amigo e mestre Seixas, pelas incontáveis discussões e por me guiar, sempre no melhor dos caminhos. Aos colegas do LPS, pelas contribuições a este trabalho.

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

## SISTEMA ONLINE DE FILTRAGEM EM UM AMBIENTE COM ALTA TAXA DE EVENTOS

André Rabello dos Anjos

Dezembro/2006

Orientador: José Manoel de Seixas

Programa: Engenharia Elétrica

O experimento ATLAS no CERN, Suíça, contará com um Sistema de Filtragem que deverá separar a Física ordinária dos eventos que possam representar decaimentos do raro bóson de Higgs. O Segundo Nível deste Sistema de Filtragem será constituído de cerca de 1.000 computadores ligados em rede, processando cada evento aprovado pelo Primeiro Nível em não mais que 10 milissegundos. Neste nível, operará um conjunto de algoritmos descritos em *software* que executará a seleção de eventos. Dentre estes, algoritmos de detecção de elétrons têm papel fundamental na eficiência da aquisição de dados, uma vez que a ocorrência destas partículas pode representar a Física de interesse. Neste trabalho, apresentamos algoritmos de discriminação mais eficientes baseados em redes neurais artificiais e um sistema de compactação de dados que se beneficia do perfil de deposição energético destas partículas em calorímetros, alcançando uma eficiência de classificação de 97,6% em elétrons para apenas 3,2% de falso-alarmede em jatos. Este algoritmo de detecção é implementado dentro da complexa infraestrutura de *software* do experimento, podendo ser executado em apenas 125 microssegundos.

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

## ONLINE TRIGGER SYSTEM FOR A HIGH EVENT RATE ENVIRONMENT

André Rabello dos Anjos

December/2006

Advisor: José Manoel de Seixas

Department: Electrical Engineering

The ATLAS experiment at CERN, Switzerland, will count on a triggering system that separates the ordinary physics from the one representing decays of the rare Higgs boson. The Second Level of such a Trigger System will be composed 1,000 computers connected by commodity networks, processing each event approved the First Level Trigger in no more than 10 milliseconds. A set of algorithms described via software will operate in this filtering level. Among them, electron detection systems play a fundamental role to the data acquisition since the existence of these particles can represent interesting physics. In this work, we present more efficient discrimination algorithms based on artificial neural networks and a compaction system which benefits from the energy deposit profiles of these particles in calorimeters, reaching a classification efficiency of 97.6% for electrons for a false-alarm of only 3.2% in jets. This detection algorithm is implemented as part of the experiment's complex software infrastructure and can be executed in only 125 microseconds.

# Conteúdo

<b>Conteúdo</b>	<b>vii</b>
<b>Lista de Figuras</b>	<b>x</b>
<b>Lista de Tabelas</b>	<b>xx</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	2
1.2 O experimento ATLAS e o bóson de Higgs . . . . .	3
1.3 Sobre as origens deste trabalho . . . . .	5
1.4 As contribuições deste trabalho . . . . .	10
1.5 Organização do Texto . . . . .	13
<b>2 A Física de Altas Energias</b>	<b>15</b>
2.1 Um Pouco de História . . . . .	15
2.2 Física de Partículas Moderna . . . . .	18
2.2.1 Aceleradores e Detetores . . . . .	19
2.2.2 A Aceleração das Partículas . . . . .	19
2.2.3 Detecção: “Vendo” o que ocorreu após a colisão . . . . .	20
2.3 Calorimetria Moderna . . . . .	23
2.3.1 Tipos de Calorímetros . . . . .	25
2.3.2 A Física da Calorimetria . . . . .	26
2.4 Sistemas de Filtragem . . . . .	29
2.4.1 Sistemas de filtragem em experimentos modernos . . . . .	30
<b>3 O Experimento ATLAS no CERN</b>	<b>34</b>
3.1 O Laboratório CERN . . . . .	34

3.2	O LHC . . . . .	36
3.3	O Experimento ATLAS . . . . .	38
3.4	O Detetor ATLAS . . . . .	40
3.4.1	Os Detetores de Traços . . . . .	41
3.4.2	Os Calorímetros . . . . .	43
3.4.3	O Detetor de Múons . . . . .	50
<b>4</b>	<b>O Sistema de filtragem e aquisição de dados do ATLAS</b>	<b>54</b>
4.1	O primeiro nível de filtragem . . . . .	55
4.2	Os altos níveis de filtragem, aquisição de dados e controle . . . . .	58
4.2.1	O sistema de controle do detetor . . . . .	59
4.2.2	O sistema <i>online</i> . . . . .	61
4.2.3	O sistema de fluxo de dados . . . . .	63
4.2.4	Fluxo de dados e a arquitetura no LVL2 . . . . .	67
4.2.5	Análise funcional do fluxo de dados no LVL2 . . . . .	73
4.2.6	Altos níveis de filtragem e seus algoritmos . . . . .	82
<b>5</b>	<b>Filtragem baseada em calorimetria no experimento ATLAS</b>	<b>89</b>
5.1	Objetos de interesse e RoI's . . . . .	91
5.2	Análise do sistema de filtragem com relação à Física . . . . .	93
5.3	Deteção de elétrons baseada em calorimetria . . . . .	96
5.4	Deteção de elétrons no LVL2 . . . . .	100
5.5	Deteção de elétrons no LVL2 do ATLAS . . . . .	102
5.5.1	Ajuste fino para cada canal . . . . .	107
5.6	Caracterização do T2Calo e do EGammaHypo . . . . .	107
5.6.1	Deteção de elétrons com o EGammaHypo e sua otimização . . . . .	110
<b>6</b>	<b>Análise neural baseada em calorimetria para a deteção elétron/jato</b>	<b>125</b>
6.1	Revisão da literatura . . . . .	125
6.1.1	Métodos neurais para análise <i>offline</i> . . . . .	126
6.1.2	Redes neurais em sistemas de filtragem . . . . .	128
6.1.3	Redes neurais e calorimetria em sistemas de filtragem . . . . .	130
6.1.4	Análise neural resumida . . . . .	132
6.1.5	Discriminador de Fisher na separação elétron/jato . . . . .	134



6.1.6	Resultados para um detetor linear aplicado à separação elétron-jato . . . . .	137
6.2	Métodos neurais de discriminação . . . . .	147
6.2.1	Discriminação neural aplicada às saídas do T2Calo . . . . .	152
6.2.2	Adicionando mais variáveis . . . . .	162
6.2.3	Relevância das 14 características do T2Calo . . . . .	169
6.2.4	Análise de Componentes Principais . . . . .	178
6.3	Mapeamento topológico . . . . .	191
6.3.1	Anelamento dos dados de uma RoI . . . . .	192
6.3.2	Discriminação baseada em anelamento . . . . .	201
6.3.3	Compactação baseada na relevância . . . . .	211
<b>7</b>	<b>Desempenho computacional do sistema de discriminação</b>	<b>231</b>
7.1	Migração para o Sistema de Filtragem do ATLAS . . . . .	238
7.1.1	<i>NeuralRinger</i> e o Athena . . . . .	240
7.1.2	<i>NeuralRinger</i> e o sistema de aquisição . . . . .	249
7.2	Implementação em um DSP . . . . .	254
<b>8</b>	<b>Conclusões</b>	<b>259</b>
8.1	Linhas de Pesquisa . . . . .	263
	<b>Referências Bibliográficas</b>	<b>266</b>
<b>A</b>	<b>As Coordenadas do ATLAS</b>	<b>277</b>
<b>B</b>	<b>Introdução ao processamento neural</b>	<b>279</b>
B.1	Processamento Neural . . . . .	281
B.1.1	Treinamento por retro-propagação de erros simples . . . . .	283
B.1.2	Treinamento por retro-propagação de erros resiliente . . . . .	287
B.1.3	Funções de ativação . . . . .	289
<b>C</b>	<b><i>NeuralRinger</i>: Projeto e implementação</b>	<b>290</b>
<b>D</b>	<b>Publicações</b>	<b>303</b>

# Lista de Figuras

1.1	Um sistema genérico de deteção em um problema de Física de Altas Energias. . . . .	5
2.1	O <i>setup</i> de Rutherford. . . . .	17
2.2	A aceleração de partículas. . . . .	20
2.3	O diagrama de um detetor cilíndrico moderno. Extraído de [3]. . . . .	22
2.4	A interação de partículas com os detetores modernos. Extraído de [3].	23
2.5	Diagrama da seção eletromagnética do calorímetro do experimento CMS. . . . .	27
2.6	O tamanho do evento <i>versus</i> a taxa de produção em experimentos atuais. Extraído de [17]. . . . .	32
3.1	Diagrama esquemático dos aceleradores do CERN e suas conexões. Extraído de [18]. . . . .	36
3.2	Visão esquemática da fronteira franco-suíça mostrando o CERN e os pontos de acesso aos poços dos 4 experimentos do LHC. Extraído de [18]. . . . .	39
3.3	O detetor ATLAS. Extraído de [18]. . . . .	41
3.4	Corte tridimensional do Detetor Interno do ATLAS. . . . .	42
3.5	Diagrama esquemático dos calorímetros no detetor ATLAS. . . . .	44
3.6	O Calorímetro e.m. do ATLAS em um corte transversal. . . . .	46
3.7	Diagrama de um dos gomos do barril da seção e.m. do ATLAS. . . . .	47
3.8	O calorímetro de telhas do ATLAS (ao centro), em sua posição final, envolvido pelo detetor de múons. Extraído de [18]. . . . .	49
3.9	Configuração de leitura das células da seção do barril do Calorímetro de Telhas. . . . .	51

3.10	Visão tridimensional dos detetores de múons do ATLAS. . . . .	52
4.1	Visão funcional simplificada do sistema de filtragem do experimento ATLAS. . . . .	55
4.2	Diagrama simplificado dos sinais de entrada e saída do LVL1. . . . .	57
4.3	Diagrama em blocos indicando as principais funções do LVL1. . . . .	58
4.4	A máquina de estados finitos utilizada por todos os componentes do TDAQ e detetores do ATLAS. . . . .	62
4.5	Os componentes do DF e suas interconexões. . . . .	64
4.6	A ordem das mensagens trocadas pelas aplicações do LVL2. . . . .	68
4.7	A arquitetura da L2PU. . . . .	71
4.8	Desempenho da coleta de dados de RoI para várias combinações de tamanhos de RoI (em bytes) e tarefas processadoras. Extraída de [38].	76
4.9	A taxa inversa em um L2SV para uma única L2PU funcionando com diferentes números de tarefas processadoras. Extraída de [38]. . . . .	78
4.10	A taxa de eventos média para cada L2PU em função do número de L2PU's no sistema para diferentes números de ROB's por RoI. Cada ROB contribui com um montante igual dos dados da RoI. Extraída de [38]. . . . .	80
4.11	A taxa de eventos média para cada L2PU em função do número de L2PU's no sistema. O tempo de processamento emulado para cada evento é de 1 ms. Extraída de [38]. . . . .	81
4.12	A taxa sustentada por L2SV com o aumento de L2PU's no sistema. Extraída de [38]. . . . .	82
4.13	Relação entre os macrossistemas do HLT. . . . .	84
4.14	A seqüência de interações do L2SC com o L2PU e o <i>software</i> de seleção de eventos. Extraída de [39]. . . . .	86
4.15	As principais contribuições para o tempo de processamento mostradas como integrais para a fatia completa do sistema de seleção para RoI's de múons. Extraída de [16]. . . . .	88
5.1	Diagrama esquemático do Sistema de Filtragem do Experimento ATLAS segundo suas funções de filtragem. . . . .	94

5.2	Interação de um elétron com um detector de traços com absorvedores em chumbo. . . . .	97
5.3	Exemplo da relação de deposição energética na seção hadrônica e energia total para elétrons (topo) e jatos (baixo). . . . .	99
5.4	Representação gráfica do algoritmo de detecção de objetos e.m. no LVL1. . . . .	101
5.5	T2Calo, Etapa 1: Cálculo do centro refinado e deposições de energia na segunda camada e.m., para regiões de tamanho $\Delta_\eta = 0,075 \times \Delta_\phi = 0,175$ e $\Delta_\eta = 0,175 \times \Delta_\phi = 0,175$ . . . . .	103
5.6	T2Calo, Etapa 2: Procura dos dois maiores picos na primeira camada e.m.. . . . .	104
5.7	T2Calo, Etapa 3: Cálculo dos somatórios parciais de energia por camada, em uma área equivalente a $\Delta_\eta = 0,075 \times \Delta_\phi = 0,175$ . . . . .	105
5.8	T2Calo, Etapa 4: Valores parciais e somatório das energias, em uma área equivalente a $0,2 \times 0,2$ no plano $\eta \times \phi$ , para as camadas da seção hadrônica. . . . .	106
5.9	A deposição total de energia em uma região $0,4 \times 0,4$ no plano $\eta \times \phi$ para elétrons (em cima) e jatos (em baixo) para a massa de dados disponível para o estudo. . . . .	109
5.10	A deposição total de energia na seção e.m. em uma região $0,4 \times 0,4$ no plano $\eta \times \phi$ para elétrons (em cima) e jatos (em baixo) para a massa de dados disponível para o estudo. As contagens estão normalizadas. . . . .	111
5.11	A deposição total de energia na seção hadrônica em uma região $0,4 \times 0,4$ no plano $\eta \times \phi$ para elétrons (em cima) e jatos (em baixo) para a massa de dados disponível para o estudo. As contagens estão normalizadas. . . . .	112
5.12	Distribuição da variável $R_{e.m.2}^{3 \times 7 / 7 \times 7}$ para elétrons (em cima) e jatos (em baixo). As contagens estão normalizadas. . . . .	113
5.13	Distribuição da variável $R_{e.m.1}^{2-\text{máximos}}$ para elétrons (em cima) e jatos (em baixo). As contagens estão normalizadas. . . . .	114
5.14	Distribuição em $\eta$ dos centros refinados das ROI's para elétrons (em cima) e jatos (em baixo). As contagens estão normalizadas. . . . .	115

5.15	Distribuição em $\phi$ dos centros refinados das RoI's para elétrons (em cima) e jatos (embaixo). As contagens estão normalizadas. . . . .	116
5.16	A curva ROC para 771.579 combinações de valores de corte para o algoritmo EGammaHypo. . . . .	120
5.17	Eficiência de detecção contra falso-alarme por $\eta$ para o conjunto de testes utilizando o algoritmo EGammaHypo. . . . .	122
5.18	Eficiência de detecção contra falso-alarme por $\phi$ para o conjunto de testes utilizando o algoritmo EGammaHypo. . . . .	123
5.19	Eficiência de detecção contra falso-alarme por $E_{T_{3\times 7}}^{e.m.}$ para o conjunto de testes utilizando o algoritmo EGammaHypo. . . . .	124
6.1	Um sistema genérico de detecção em um problema de Física de Altas Energias. . . . .	133
6.2	Diagrama de blocos de um sistema de detecção elétron/jato baseado no Discriminador de Fisher associado ao extrator de características T2Calo. . . . .	135
6.3	O fluxo implementado para o treinamento do discriminador linear. . . . .	136
6.4	Evolução dos valores do EMQ ao longo do treinamento do detetor linear, para o conjunto de treinamento e de teste. . . . .	139
6.5	Evolução dos valores do produto SP ao longo do treinamento do detetor linear, para o conjunto de treinamento e de teste. . . . .	140
6.6	Saída do discriminador linear (após treinamento) para elétrons e jatos. . . . .	141
6.7	Curva R.O.C. para um detetor linear para elétrons e jatos baseado nas 4 variáveis do T2Calo, usadas para discriminação pelo EGammaHypo. . . . .	142
6.8	R.O.Cs comparativas entre a otimização atual para o EGammaHypo e um detetor linear. . . . .	143
6.9	Análise da eficiência de classificação e do produto SP para o classificador linear, para os dados do conjunto de teste ao longo de $\eta$ em (a) e (c) e por $\phi$ , em (b) e (d). . . . .	145
6.10	Análise da eficiência de detecção de elétrons e falso-alarme em jatos (a) e do produto SP (b) por $E_{T_{3\times 7}}^{e.m.}$ para o classificador linear, tendo por base os dados do conjunto de teste. . . . .	146

6.11	Comparação do produto SP para o classificador linear e o EGammaHypo por energia transversa na seção e.m.. . . . .	148
6.12	Gráfico de dispersão mostrando a distribuição das duas variáveis mais discriminantes para o T2Calo, para as classes de elétrons e jatos. . . .	150
6.13	Um discriminador neural para as características definidas pelo T2Calo.	152
6.14	Evolução do EMQ para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do T2Calo. . . . .	154
6.15	Evolução do produto SP para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do T2Calo. . . .	154
6.16	Saída para o conjunto de teste, do discriminador neural baseado nas características extraídas pelo T2Calo. . . . .	156
6.17	R.O.C. para o conjunto de teste, do discriminador neural baseado nas características extraídas pelo T2Calo, comparado com o detetor linear e a técnica de otimização baseado no algoritmo EGammaHypo.	157
6.18	Análise da eficiência de classificação e do produto SP para o discriminador neural baseado nas saídas do T2Calo, para os dados do conjunto de teste ao longo de $\eta$ em (a) e (c) e por $\phi$ , em (b) e (d). . .	158
6.19	Análise da eficiência de detecção de elétrons e falso-alarme em jatos (a) e do produto SP (b) por $E_{T_{3\times 7}}^{e.m.}$ para o discriminador neural baseado nas saídas do T2Calo, tendo por base os dados do conjunto de teste. .	159
6.20	Comparação do produto SP para os 3 detetores abordados até aqui, por faixa de energia e.m. transversa. . . . .	160
6.21	Diagrama de fluxo do sistema de detecção elétron/jato, neural, usando as características do T2Calo como entrada. . . . .	161
6.22	Diagrama em blocos de um sistema de detecção utilizando toda a informação produzida pelo T2Calo. . . . .	163
6.23	Saída para o conjunto de teste, do discriminador neural baseado nas 14 características extraídas pelo T2Calo. . . . .	164
6.24	Evolução do EMQ para os conjuntos de treino e teste de um discriminador elétron/jato neural para as 14 variáveis extraídas pelo T2Calo.	165

6.25	Evolução do produto SP para os conjuntos de treino e teste de um discriminador elétron/jato neural para as 14 variáveis extraídas pelo T2Calo. . . . .	166
6.26	Comparação das ROC's para os 4 detetores abordados até aqui. . . .	167
6.27	Análise da eficiência de classificação e do produto SP para um detetor neural baseado nas 14 saídas do T2Calo, para os dados do conjunto de teste ao longo de $\eta$ em (a) e (c) e por $\phi$ , em (b) e (d). . . . .	168
6.28	Comparação do produto SP máximo por faixa energética para os 2 detetores neurais abordados até aqui e o detetor linear. . . . .	170
6.29	Os valores de relevância para os conjunto de treino e teste, para as 14 variáveis do T2Calo e considerando-se o classificador neural em estudo.	172
6.30	Os valores de relevância de discriminação para os conjunto de treino e teste, para as 14 variáveis do T2Calo e considerando-se o classificador neural em estudo. . . . .	174
6.31	Comparação das ROC's para os 6 detetores abordados até aqui. . . .	176
6.32	Eficiência de deteção e falso-alarmede para o detetor baseado nas 5 mais relevantes (por discriminação) do espaço das 14 variáveis produzidas pelo T2Calo. . . . .	177
6.33	Um discriminador neural para as características definidas pela KLT das saídas do T2Calo. . . . .	178
6.34	Correlação (normalizada) entre as 14 variáveis do T2Calo. . . . .	181
6.35	A curva de carga de energia para as 14 variáveis do novo espaço criado pela KLT. . . . .	182
6.36	Evolução do EMQ ao longo do treinamento para um classificador baseado na KLT das 14 variáveis extraídas pelo T2Calo. . . . .	184
6.37	Evolução do Produto SP ao longo do treinamento para um classificador baseado na KLT das 14 variáveis extraídas pelo T2Calo. . . . .	185
6.38	Curva ROC comparativa entre todos os métodos de deteção avaliados até este momento e incluindo a deteção baseada em PCA. . . . .	186

6.39	Análise da eficiência de classificação e do produto SP para o detetor elétron/jato baseado na transformação linear (via KLT) do espaço das 14 variáveis produzidas pelo T2Calo, para os dados do conjunto de teste ao longo de $\eta$ em (a) e (c) e por $\phi$ , em (b) e (d). . . . .	187
6.40	Análise de relevância de discriminação para detetor baseado em uma KLT das 14 variáveis do T2Calo. . . . .	188
6.41	Comparação das ROC's de diversos detetores baseados da poda (segundo diferentes critérios) de um sistema com 14 variáveis na entrada, resultantes da aplicação da KLT nas 14 variáveis disponíveis na saída do T2Calo. . . . .	190
6.42	Modelo de um objeto e.m. interagindo com um calorímetro. . . . .	192
6.43	Imagem tridimensional mostrando a deposição energética de um elétron simulado na segunda camada e.m. do calorímetro do ATLAS. . . . .	193
6.44	Esquema do processo de anelamento para as diferentes camadas dos calorímetros do ATLAS. . . . .	195
6.45	Imagem tri-dimensional mostrando a deposição energética de um elétron simulado na primeira camada hadrônica do calorímetro do ATLAS. . . . .	197
6.46	Estratégia para obter acesso aos dados tais como processados pelo algoritmo T2Calo. . . . .	200
6.47	Diagrama em blocos de um sistema de detecção utilizando o anelador como extrator de características e um detetor neural. . . . .	201
6.48	Evolução do EMQ para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do anelador. . . . .	203
6.49	Evolução do produto SP para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do anelador. . . . .	204
6.50	Saída para o conjunto de teste, do discriminador neural baseado nas características extraídas pelo anelador. . . . .	205
6.51	Comparação da R.O.C. dos diversos sistemas de filtragem abordados até agora com o sistema neural baseado em anéis de deposição energética. . . . .	207



6.52	Análise da eficiência de classificação e do produto SP para o discriminador neural baseado nas saídas do anelador, para os dados do conjunto de teste ao longo de $\eta$ em (a) e (c) e por $\phi$ , em (b) e (d). . . . .	208
6.53	Análise da eficiência de detecção de elétrons e falso-alarme em jatos (a) e do produto SP (b) por $E_{T_{3 \times 7}}^{e.m.}$ para o discriminador neural baseado nas saídas do anelador, tendo por base os dados do conjunto de teste. . . . .	209
6.54	Gráfico comparativo do produto SP ao longo de $E_T^{e.m.}$ para os discriminador neurais baseados nas saídas do anelador, parcial e completa do T2Calo. . . . .	210
6.55	Os valores de relevância para os conjunto de teste, para as 100 variáveis do anelador e considerando-se o classificador neural em estudo, como indica a Tabela 6.8. . . . .	212
6.56	Os valores de relevância de discriminação para os conjunto de teste, para as 100 variáveis do anelador e considerando-se o classificador neural em estudo, como indica a Tabela 6.8. . . . .	213
6.57	Comparação entre as ROCs de alguns dos sistemas de discriminação abordados até aqui. Em destaque, a ROC do melhor discriminador (linha cheia), determinado a partir da análise de relevância. . . . .	217
6.58	Os valores de relevância de discriminação para os conjunto de treino e teste, para as 62 variáveis que sobreviveram a uma poda baseada na relevância de discriminação. . . . .	218
6.59	Comparação entre as ROCs de alguns dos sistemas de discriminação abordados até aqui. Em destaque, a ROC do melhor discriminador (linha cheia), determinado a partir de uma segunda análise de relevância. . . . .	220
6.60	Os valores de relevância de discriminação para os conjunto de treino e teste, para as 62 variáveis que sobreviveram a uma poda baseada na relevância de discriminação. . . . .	221
6.61	R.O.C. comparativa entre os sistemas de detecção de elétrons e jatos analisados até agora e o novo sistema de detecção gerado a partir da poda dos canais de entrada baseada na relevância. . . . .	223

6.62	Relevâncias de discriminação para o detetor neural baseado em 34 (dos 100) anéis resultantes da poda por relevância. . . . .	225
6.63	R.O.C. comparativa entre alguns dos sistemas de detecção de elétrons e jatos analisados até agora e o novo sistema de detecção baseado na poda da relevância, com apenas 14 (dos 100) anéis. . . . .	227
6.64	Relevâncias de discriminação para o detetor neural baseado em 14 (dos 100) anéis resultantes da poda por relevância. . . . .	228
6.65	Relevâncias de discriminação para o detetor neural baseado em 13 (dos 14) anéis menos relevantes do último detetor. . . . .	230
7.1	Tempos totais de execução do <i>NeuralRinger</i> em três plataformas distintas, utilizando 100 anéis. . . . .	235
7.2	Tempos totais de execução do <i>NeuralRinger</i> em três plataformas distintas, utilizando 34 anéis. . . . .	236
7.3	Tempos totais de execução do <i>NeuralRinger</i> em três plataformas distintas, utilizando apenas 14 anéis. . . . .	237
7.4	Tempos individuais para cada fase de execução do <i>NeuralRinger</i> , levando-se em consideração a extração e discriminação baseada em 100 anéis. O teste foi executado na plataforma Opteron/250 (pré-serie). . . . .	239
7.5	Exemplo de um cenário de seleção simples para o <i>Steering</i> operando no LVL2, baseado em uma RoI tipo E.M.. . . . .	242
7.6	Erros entre os processos de extração de características e hipótese neural entre o Athena e o <i>NeuralRinger</i> rodando em modo desacoplado. . . . .	245
7.7	Gráfico de correlação entre a saída final do Athena e de uma versão do <i>NeuralRinger</i> rodando em modo desacoplado. . . . .	246
7.8	Tempos de execução do <i>NeuralRinger</i> funcionando de forma integrada ao ambiente Athena. . . . .	248
7.9	Relação entre o posicionamento da RoI em $\eta$ e os tempos de processamento do <i>NeuralRinger</i> em diversas das fases da extração de características no ambiente Athena. . . . .	250

7.10	Relação entre o posicionamento da RoI em $\phi$ e os tempos de processamento do <i>NeuralRinger</i> em diversas das fases da extração de características no ambiente Athena. . . . .	251
7.11	Esquema de bancada de testes para a verificação do funcionamento do NeuralRinger dentro do ambiente de aquisição de dados do ATLAS.	253
7.12	Captura de tela mostrando a operação do sistema de extração de características baseado em anéis rodando dentro de uma banca de testes do sistema de filtragem e aquisição do ATLAS. . . . .	254
7.13	Visão geral da estrutura interna de um DSP. . . . .	255
7.14	Função de distribuição cumulativa para os tempos de processamento, subdivididos por fase de processamento, para implementação do <i>NeuralRinger</i> em um DSP SHARC 21160 da <i>Analog Devices</i> . . . . .	257
7.15	Tempo total de execução para o anelador e detetor neural para diferentes DSP's da família SHARC, da <i>Analog Devices</i> . . . . .	258
A.1	O sistema de coordenadas do ATLAS. . . . .	278
B.1	O diagrama de fluxo do discriminador LMS que será empregado na discriminação elétron-jato. . . . .	282
B.2	Grafo de fluxo de sinal de um neurônio artificial. . . . .	282
B.3	Modelagem de uma rede MLP, totalmente conectada e sem retropropagação de sinal. . . . .	284
C.1	Diagrama de blocos mostrando a relação entre os pacotes do <i>NeuralRinger</i> . . . . .	291
C.2	Diagrama UML mostrando as relações dos componentes do pacote <code>sys</code> .	292
C.3	Diagrama UML mostrando as relações dos componentes do pacote <code>data</code> . . . . .	293
C.4	Diagrama UML mostrando as relações dos componentes do pacote <code>network</code> . . . . .	296
C.5	Exemplo de um gráfico de fluxo produzido pelo método <code>dot()</code> para uma rede LMS, tal qual utilizada para os testes na Seção 6.1.5. . . . .	298
C.6	Exemplo de uma gráfico de fluxo produzido pelo método <code>dot()</code> para uma rede MLP. . . . .	299

# Lista de Tabelas

3.1	A segmentação, camada a camada, dos calorímetros e.m. do ATLAS.	48
3.2	A segmentação, camada a camada, dos calorímetros hadrônicos do ATLAS. . . . .	50
4.1	Taxas de processamento em um sistema com dois núcleos tipo AMD-Athlon 1,533 GHz. . . . .	87
5.1	Taxas de saídas para as assinaturas reconhecidas pelo LVL1, para a luminosidade de pico inicial de $2 \times 10^{33} cm^{-2} s^{-1}$ . Dados baseados em uma simulação do LVL1. . . . .	92
5.2	Valores parciais de detecção e falso-alarme para o detetor baseado no algoritmo EGammaHypo. . . . .	120
6.1	Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado na saída do T2Calo. . .	153
6.2	Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado em todas as variáveis produzidas pelo T2Calo. . . . .	165
6.3	Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado na aplicação da KLT (PCA) nas 14 variáveis originais do T2Calo. . . . .	183
6.4	Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado na aplicação da KLT (PCA) nas 14 variáveis originais do T2Calo. . . . .	191
6.5	Configuração para os anéis formados a partir de uma RoI de tamanho $\eta = 0,4 \times \phi = 0,4$ . . . . .	198

6.6	Algoritmo para a “Normalização Seqüencial” . . . . .	199
6.7	Análise dos parâmetros de treinamento para um detetor neural cuja entrada são os 100 anéis resultantes do processamento de uma RoI. . . . .	202
6.8	Posicionamento relativo dos anéis produzidos pelo anelador no vetor de entrada ao sistema de detecção neural. . . . .	214
6.9	Anéis sobreviventes de uma poda baseada na relevância de discriminação. . . . .	215
6.10	Anéis sobreviventes de uma segunda poda baseada na relevância de discriminação. . . . .	216
6.11	Posicionamento relativo dos anéis produzidos pelo sistema de anelamento podado para ter apenas 34 (dos 100) anéis. . . . .	224
6.12	Posicionamento relativo dos anéis produzidos pelo sistema de anelamento podado usando a análise de relevância discriminante, para ter apenas 14 (dos 100) anéis. . . . .	228
7.1	Tempo de processamento médio (em milissegundos) das diversas fases do algoritmo T2Calo em uma máquina com processadores Intel Xeon de 2,4 GHz e 1 Gb de memória RAM. . . . .	232
7.2	Configurações das máquinas utilizadas para o teste de desempenho do <i>NeuralRinger</i> . . . . .	233

# Glossário

**ATLAS** *A Toroidal LHC ApparatuS* (detetor, experimento)

**CMS** *Compact Muon Solenoid* (detetor, experimento)

**DAQ** Sistema de Aquisição de Dados (*Data Acquisition System*)

**DCS** Sistema de Controle do Detetor (*Detector Control System*)

**DF** Sistema de Fluxo de Dados (*Data Flow System*)

**DFM** Gestor do Sistema de Fluxo de Dados (*Data Flow Manager*)

**DM** Gerente de Dados (*Data Manager*)

**DSS** Sistema de Segurança do Detetor (*Detector Safety System*)

**EB** Construção de Eventos (*Event Building*)

**EDM** Modelo para Dados do Evento (*Event Data Model*)

**EF** Filtro de Eventos (*Event Filter*)

**EFD** Gestor do Fluxo de Dados para o Filtro de Eventos (*Event Filter Dataflow Manager*)

**EM ou e.m.** Eletromagnético

**EMQ** Erro Médio Quadrático (*Mean Square Error, MSE*)

**FEx** Extração de Características (*Feature Extraction*)

**FPGA** *Field-Programmable Gate Array*

**HLT** Filtro(s) de Alto Nível (*High-Level Trigger*)

**ID** Detetor Interno (*Inner Detector*)

**KLT** Transformação de Kharhunen-Loève (*Kharhunen-Loève Transform*)

**L1R** Resultado do Primeiro Nível de Filtragem (*Level 1 Trigger Result*)

**L2PU** Unidade de Processamento do Segundo Nível de Filtragem (*Second Level Trigger Processing Unit*)

**L2R** Resultado do Segundo Nível de Filtragem (*Level 2 Trigger Result*)

**L2SC** ou **PSC** Controlador do Gerente de Operações do Segundo Nível de Filtragem (*LVL2 (PESA) Steering Controller*)

**L2SV** Supervisor do Segundo Nível de Filtragem (*Second Level Trigger Supervisor*)

**LAr** Argônio Líquido (*Liquid Argon*)

**LHC** Grande Colisionador de Hádrons (*Large Hadron Collider*)

**LMS** Mínimo Médio Quadrático (*Least Mean Square*)

**LVL1** Primeiro Nível de Filtragem (*First Level Trigger*)

**LVL2** Segundo Nível de Filtragem (*Second Level Trigger*)

**MLP** Percéptron de Múltiplas Camadas (*Multi-layer Perceptron*)

**PCA** Análise de Componentes Principais (*Principal Component Analysis*)

**pROS** pseudo Sistema de Leitura (*pseudo ReadOut System*)

**PT** Tarefa de Processamento do Filtro de Eventos (*Processing Task*)

**RNA** Redes Neurais Artificiais

**ROB** *Buffer* de Leitura do Detetor (*ReadOut Buffer*)

**ROC** Característica de Operação do Detetor (*Receiver Operating Characteristics*)

**ROD** *Drive de Leitura do Detetor (ReadOut Drive)*

**RoI** *Região de Interesse (Region of Interest)*

**ROS** *Sistema de Leitura do Detetor (ReadOut System)*

**RPROP** *Retropropagação de Erros Resiliente (Resilient Error Backpropagation)*

**SCT** *Semi Conductor Tracker (detetor)*

**SFI** *Entrada de Subfarm (Subfarm Input)*

**SFO** *Saída de Subfarm (Subfarm Output)*

**SMP** *Sistema de Processamento Simétrico (Symmetric Multi-Processor)*

**TDAQ** *Sistema Filtragem e Aquisição de Dados (Trigger and Data Acquisition System)*

**TRT** *Transition Radition Tracker (detetor)*

**TT** *Torre de Filtragem (Trigger Tower)*

**TTC** *Tempo, Filtragem e Controles (Timing, Trigger and Controls)*



# Capítulo 1

## Introdução

Sistemas eletrônicos de aquisição de dados são comumente empregados em muitos campos da engenharia. Exemplos podem ser encontrados na captura de áudio, vídeo, sinais de satélite, de rádio, biológicos e cartográficos, entre outros. Os sinais de fenômenos observados na natureza são registrados por sensores, submetidos a uma conversão analógico-digital e processados por *hardware* especializado, programas rodando em computadores genéricos, sistemas dedicados contendo DSP's, FPGA's ou misturas de todas essas tecnologias.

O número de sensores utilizados em cada nova situação varia de acordo com o problema abordado, podendo passar de alguns poucos (por exemplo, um microfone) a uma quantidade gigantesca (o número de píxeis em câmeras fotográficas digitais). A informação pode se encontrar segmentada ou contida em um único sinal que varia no tempo (e na frequência). Em problemas extremamente segmentados, é típica a utilização de pré-processamento distribuído, habitualmente acoplado ao sensor. Unidades centralizadoras comunicam-se com a eletrônica de aquisição de forma a obter os dados pré-processados e combiná-los para gerar o resultado final esperado por aquele sistema.

Especificamente, em atividades relacionadas ao reconhecimento de padrões captados por múltiplos sensores, a fase de pré-processamento é de extrema importância. A informação dos sensores é recombinaada em um espaço adaptado onde a informação que se deseja discriminar torna-se mais evidente, preferencialmente livre das correlações de ordem elevada que ocorrem frequentemente na natureza. Este processo de compactação ou extração de características inteligente é normalmente desenvolvido

tendo por base algum conhecimento empírico sobre o sinal que será tratado. Exemplos clássicos encontram-se no reconhecimento de imagens e voz, onde aplicam-se transformações ao sinal original para aumentar a qualidade de discriminação obtida nesses ambientes.

Em algumas instâncias, a detecção de padrões deve ser feita *online*, em janelas de tempo bastante limitadas. Como forma de aumentar a velocidade de atuação destes detetores, é possível distribuir as tarefas de pré-processamento e detecção em múltiplos processadores que trabalham paralelamente para executar a tarefa de reconhecimento. Este cenário pode ainda ser piorado se a informação estiver inserida em uma massa de dados muito grande, que deva ser analisada para a detecção da informação procurada.

No caso de eventos raros serem o alvo do detetor, grande acurácia é exigida no sistema de detecção, uma vez que a perda da informação de interesse pode comprometer severamente a funcionalidade do sistema. Este é o caso, por exemplo, de sistemas de apoio ao diagnóstico de doenças (como o câncer ou a tuberculose) a partir de informações do paciente. O evento (doença) é raro, mas deve ser detetado com acurácia.

## 1.1 Motivação

Experimentos em Física de Altas Energias procuram por confirmações experimentais dos modelos propostos em estudos teóricos. Laboratórios deste domínio da Física freqüentemente contam com um sistema de colisão que provoca o aparecimento da física de interesse, associado a complexos sistemas de detecção que registram a evolução no tempo de cada evento produzido. Naturalmente, envolvidos no processo de detecção encontram-se sistemas eletrônicos que automatizam a busca, o registro e a análise dos resultados obtidos.

Dada a natureza complexa e rara dos fenômenos estudados em muitos desses experimentos, a física de interesse está normalmente submersa em uma gigantesca massa de interações que representam ora ruído, provocado pelo mau funcionamento dos sistemas de detecção e colisão, ora efeitos ordinários, já bastante estudados no passado. Em específico, em experimentos que buscam a confirmação de canais físicos

em patamares energéticos elevados, de alguns gigaelétron-volts para cima, a taxa de eventos que representa canais desinteressantes contra a de eventos que possam interessar pode estar na proporção de  $10^9$  para 1. Desta forma, as interações de interesse aparecem escondidas no meio de bilhões de outras reações ordinárias, ou que representam apenas ruído. Ademais, para que se consiga tais interações, milhões de eventos são gerados por base de tempo para que se colete estatística suficiente para a comprovação do canal estudado. O volume de dados associados a cada evento vem aumentando, junto com a ambição dos experimentos. Novos sistemas de detecção exigem alguns *megabytes* para cada evento registrado, o que representa uma dificuldade extra na realização desses experimentos.

Para resolver esse problema, introduzem-se sistemas eletrônicos de filtragem que podem separar *online* os eventos de interesse dos eventos que representam ruído ou física ordinária. Sua arquitetura é projetada para resolver os desafios impingidos pelas duras condições de operação nestes ambientes, seguindo os padrões de complexidade do experimento. Soluções atuais empregam forte paralelização e técnicas modernas de processamento de sinais para responder às demandas de colisões mais e mais energéticas.

Redes Neurais Artificiais vêm sendo empregadas nos sistemas de filtragem em várias instâncias nessa área da Física. Dada a forte segmentação dos dados nos detetores, sistemas neurais conseguem compactar e extrair as informações vitais para a discriminação dos dados, mantendo não só a alta qualidade de classificação como também excelentes níveis de velocidade de processamento.

## 1.2 O experimento ATLAS e o bóson de Higgs

A detecção do bóson de Higgs é uma das grandes “vedetes” da Física de Altas Energias atual. Esta partícula, se existir, possui uma massa bastante elevada (centenas de gigaelétron-volts) e se apresenta como um canal bastante raro e de difícil reprodução laboratorial. A descoberta dessa partícula confirmará mais uma vez o Modelo Padrão, já bastante testado, inicialmente proposto em 1954.

O CERN, na Suíça, é o local onde está sendo desenvolvido o experimento ATLAS, que pretende investigar a rara física do bóson de Higgs. O experimento utilizará

colisões próton-próton, numa taxa de 40 milhões por segundo para conseguir obter algumas dezenas destes bósons por dia de operação. As colisões serão providas pelo Grande Colisionador de Hádrons (do inglês *Large Hadron Collider*, LHC), que será, quando estiver operacional, em 2007, o mais potente no mundo, podendo colidir prótons com 14 TeV no centro de massa. Uma vez que cada evento no ATLAS consumirá cerca de 1,5 *megabytes* de espaço em memória, um dos problemas do projeto e construção do experimento está na articulação de um sistema de filtragem de eventos que seja capaz de fazer uma seleção *online* dos eventos que representem a física de interesse.

O Sistema de Filtragem do ATLAS foi inicialmente projetado para operar em três níveis conectados em cascata com complexidade, qualidade de detecção e tempo de operação por evento crescentes. O Segundo Nível de Filtragem (LVL2), especificamente, será constituído de cerca de 1.000 unidades de processamento ligadas em rede, processando cada candidato aprovado pelo Primeiro Nível (LVL1). No LVL2 cada evento-candidato terá, em média, aproximadamente 10 milissegundos para ser processado (latência).

O LVL2 coordena um conjunto de algoritmos descritos em *software* que executa a seleção de eventos. Dentre estes, algoritmos de detecção de elétrons têm papel fundamental na eficiência da aquisição de dados, uma vez que a ocorrência de elétrons pode representar a Física de interesse (em cerca de 60% dos casos). Assim sendo, elétrons representam o sinal de interesse a ser detetado. Estima-se que, a cada 25.000 candidatos a elétrons definidos pelo LVL1, apenas 1 será verdadeiramente um elétron. Cabe ao LVL2 a redução deste ruído de fundo significativo para a detecção de elétrons.

A informação para a detecção pode ser obtida do sistema de leitura do detetor usando-se as primitivas do complexo *software* de base do sistema de filtragem, disponível em cada unidade de processamento dentro do LVL2. Os dados para cada candidato a elétron são formados por células de detecção de um calorímetro<sup>1</sup>, segmentadas tanto na direção de penetração da partícula, quanto no plano de interseção. Esse sistema forma uma malha de elementos que contém amostras da energia de

---

<sup>1</sup>Calorímetros são detetores especializados em amostrar a energia das partículas que interagem com seu volume. Esses detetores serão especificados e discutidos mais à frente no texto.

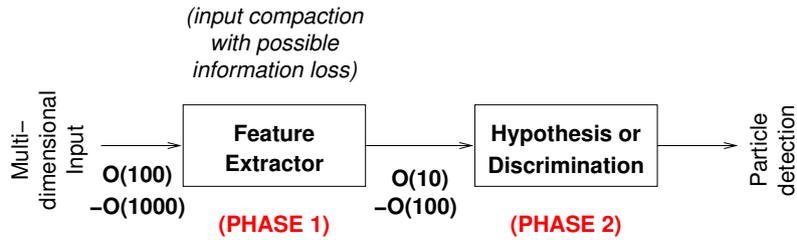


Figura 1.1: Um sistema genérico de detecção em um problema de Física de Altas Energias.

interação da partícula sendo discriminada com o sistema de detecção ao longo de sua trajetória. Em média, cada partícula-candidata ou “objeto” a ser avaliado pelo LVL2 disporá de cerca de 1.300 células de detecção. A dimensão flutuante de entrada (i.e., as diferenças no número total de células analisadas para cada objeto) é devida a variações de granularidade do detetor ao longo de seus eixos e outras facetas geométricas do experimento. Outras causas que podem atingir a dimensão de entrada do sistema de discriminação incluem defeitos na eletrônica do sistema de leitura do detetor ou transporte dos dados até as unidades de processamento.

O processo de discriminação de cada objeto em si é dividido em duas etapas bem definidas: a extração de características ou pré-processamento (fase 1), e a detecção propriamente dita (fase 2), como mostra a Figura 1.1. A extração de características tem por objetivo a compactação do sinal a ser detetado de forma que se realcem as propriedades discriminantes e, ao mesmo tempo, reduza-se a dimensionalidade do espaço de entrada. A detecção propriamente dita que se segue, será simples e robusta de forma que o canal de interesse se exprima claramente mesmo que imerso no ruído de fundo.

### 1.3 Sobre as origens deste trabalho

O experimento ATLAS foi iniciado em 1994 com a publicação de um relatório técnico indicando o objetivo de sua existência (detecção do bóson de Higgs) e sua complexa arquitetura em linhas gerais de operação. Desde então, as arquiteturas de todos os sub-sistemas deste experimento foram modificadas adaptando-se às novas tecnologias e aos desafios que surgiram com sua manufatura, planejamento de

transporte e construção. Algumas áreas sofreram mudanças bastante radicais, como o sistema de filtragem, onde decidiu-se pela utilização pioneira de computadores (PC's) e redes (ethernet) de tipo doméstico atuando como unidades de processamento e interconexão desta parte do experimento. Outras áreas foram afetadas de forma menos drástica, ainda que bastante significativas, como no caso dos detetores, que tiveram suas geometrias modificadas para aumentar a resolução na detecção das partículas de interesse ou para remediar problemas relacionados à instalação.

Uma vez que os dados realísticos provenientes do ATLAS dependem de sua instalação, áreas como o sistema de filtragem tiveram que ser desenvolvidos baseados em simulações das reações que ocorrerão durante a operação do experimento. De fato, este é um problema recorrente em experimentos em Física de Altas Energias. Sistemas complexos de simulação existem e podem simular, com detalhes impressionantes, reações sub-atômicas e a interação destas com detetores formados de múltiplos materiais e módulos. As simulações da Física de interesse e do ruído de fundo também foram amadurecendo com o passar dos anos. Só atualmente é possível contar, por exemplo, com efeitos colaterais que estarão presentes no experimento como ruído eletrônico ou o empilhamento de informação (do inglês, *pile-up*), causado pela sucessiva ocorrência de colisões no interior do sistema de detecção. A geometria dos detetores na simulação foi atualizada para casar com a descrição do sistema (que já se encontra instalado!), assim como os parâmetros de luminosidade que serão esperados quando o LHC entrar em operação.

O trabalho discutido neste documento tem suas origens em 1995 com o Projeto Final de graduação “Sistema de classificação baseado em uma máquina com sistema distribuído”, por este autor, apresentado em setembro de 1997 [1]. Os sistemas de detecção ainda estavam sendo desenvolvidos e poucas áreas tinham atingido um bom nível de maturidade. A arquitetura do Sistema de Filtragem se encontrava, igualmente, em vias de desenvolvimento.

Com base nas escolhas disponíveis na época, desenvolveu-se um sistema de filtragem para o LVL2 baseado em um processamento distribuído, comandado por um nó central de processamento, chamado Supervisor. As unidades de processamento ou nós-escravos eram pré-carregadas com um sistema de detecção neural. Estes nós analisavam os dados de cada evento atribuído pelo Supervisor e distinguiam, *online*,

se o evento era interessante e deveria ser aprovado por este nível de filtragem. No conjunto de dados disponível na época, as características (após compactação) de cada objeto analisado já haviam sido calculadas e, portanto, o sistema desenvolvido limitava-se ao papel de *deteção propriamente dita*, como colocado anteriormente.

No caso em questão, o Supervisor alocava, em *round-robin*, os eventos a cada uma das unidades de processamento que utilizavam o sistema neural para avaliar as características de um objeto, determinando se o evento deveria ser aprovado ou rejeitado. A resposta era enviada de volta ao Supervisor, que simplesmente a registrava em um arquivo de saída. No final do processamento, a saída registrada pelo Supervisor era comparada aos valores obtidos com um programa *offline* para que se determinasse se o sistema estava operando corretamente.

Este modelo foi implementado em um máquina com 16 DSP's tipo T-9000@40 MHz, da companhia INMOS. Cada nó de processamento continha, além das unidades lógicas tradicionais, uma interface de rede especializada, com 4 canais independentes de comunicação. O formato da interconexão entre os nós era configurável e foi adaptado ao problema da distribuição dos dados entre o Supervisor e as Unidades de Processamento. Para executar o *boot* desta máquina, a configuração de interconexão era carregada através de um sistema hospedeiro, seguido de um *micro-kernel* e, por fim, um programa a ser executado. Este programa podia ser codificado em C e compilado em um PC de forma cruzada, para que fosse executado nos *transputers*.

Para essa implementação, o menor tempo de processamento para cada evento foi 390 microssegundos. O sistema de distribuição de eventos em *round-robin* garantia que 9 dos 15 nós de processamento estariam executando uma tarefa dada em um instante tempo, o que era apenas sub-ótimo tendo em vista a capacidade da máquina.

Com o passar dos anos, os diversos sistemas de deteção do experimento começaram a se definir de forma mais concreta, e os programas de simulação de eventos físicos a ganhar detalhes e se aproximar, cada vez mais, de sua forma atual. De posse de dados simulados de calorimetria relativos ao processamento no LVL2, desenvolveu-se, de 1999 a 2001, também por este autor, o trabalho que culminou na dissertação de mestrado intitulada “Sistema neuronal rápido de decisão baseado em calorimetria de altas energias” [2].

Os dados disponíveis nesta época representavam cerca de 270 elétrons e 3600

jatos simulados através de interações simples com o detetor. O conjunto de dados foi obtido a partir de uma simulação primitiva do sistema de detecção do ATLAS e não continha diversos elementos que estarão presentes no sistema de detecção final, tais como efeitos de ruído, empilhamento, variação de granularidade e variação de energia.

A partir deste conjunto de dados, desenvolveu-se um sistema de discriminação neural que pudesse operar dentro do Segundo Nível de Filtragem do ATLAS. Este discriminador utilizava um pré-processamento formando “anéis” de deposição energética ao redor do ponto de impacto do objeto em estudo, se aproveitando do padrão de interação das partículas com as células de calorímetros. Ao invés de compactar o sinal da interação de elétrons com o detetor formando apenas 4 variáveis, como sugeria o sistema empregado no CERN naquela época, 58 variáveis (anéis) eram produzidas. Um sistema de normalização desta saída de dimensão 58, simplificado, baseado na energia total do objeto, foi empregado de forma que o processo de detecção se tornasse independente da energia da partícula sendo analisada.

O resultado desta *extração de anéis* baseada nas células dos calorímetros, normalizado, era utilizado como entrada para um detetor neural com 58 neurônios de entrada, 5 neurônios na camada escondida e apenas uma saída que indicava se o objeto na entrada era ou não um elétron. A rede neural era treinada de forma supervisionada (usando-se o método de retro-propagação clássica), utilizando-se metade do conjunto de dados disponível. Neste caso, obteve-se 95% de eficiência na detecção de jatos contra 97% para a detecção de elétrons. Este resultado foi comparado com uma análise simplificada do sistema de detecção desenvolvido no CERN, que apontava uma capacidade de discriminação de 95% para elétrons para um falso-alarme de 11,6% para jatos.

Uma análise baseada no impacto de cada um dos 58 anéis na saída da rede neural (análise de relevância) revelou que era possível utilizar apenas 20 dos 58 anéis iniciais para uma detecção marginalmente inferior apenas àquela utilizando todos os anéis, mas ainda muito superior à análise proposta no CERN. Uma implementação simplificada do algoritmo foi executada, indicando a possibilidade de ser utilizado no LVL2.

Atualmente, dispõe-se de um sistema muito mais elaborado para a simulação do



detetor ATLAS, e das colisões que acontecerão quando o acelerador for finalmente ligado, em meados de 2007. Dentre os novos elementos presentes, encontram-se:

- Dimensão flutuante para os dados utilizados na detecção. No caso do estudo para a dissertação de mestrado, limitou-se o escopo da análise a dados com granularidade constante. Uma simulação muito mais realística do sistema de detecção do ATLAS está presente na nova massa de dados, onde a granularidade da região de interesse varia a cada camada e ao longo dos eixos do detetor, trazendo um novo desafio a sistemas de detecção candidatos ao LVL2;
- Introdução de ruído relativo ao sistema de detecção. Este efeito não estava presente nos dados obtidos anteriormente, mas estará certamente presente quando o experimento estiver operacional. A adição de ruído aos sistemas de detecção diminui a acuidade das medidas, introduzindo uma perturbação não desprezível aos algoritmos de detecção;
- Um volume maior de dados está disponível para este trabalho, o que possibilita uma análise muito mais detalhada e acurada do problema de detecção elétron/jato. Enquanto que para os trabalhos anteriores dispunha-se de uma massa muito limitada de dados, para este trabalho dispõe-se de aproximadamente 30.000 objetos (cerca de 22.500 elétrons e 7.500 jatos) para análise, o que diminui a margem para possíveis erros estatísticos introduzidos por uma análise sobre um conjunto de dados limitado;
- A infraestrutura de suporte para o sistema de filtragem está disponível. Todos os resultados, tanto de desempenho físico quanto de velocidade de processamento devem ser mencionados com relação a uma implementação compatível e estável dentro desse ambiente. Dessa forma, é possível fazer o sistema de detecção operar tal qual um algoritmo de detecção operaria no sistema de filtragem e comparar seu desempenho tendo em vista todos os parâmetros funcionais deste complexo ambiente;
- O algoritmo empregado atualmente no CERN está também disponível em um formato mais evoluído, em que seja possível uma comparação direta em termos de eficiência e desempenho com um sistema alternativo. De posse dos

resultados de cada um dos algoritmos é possível conduzir uma análise mais detalhada das diferenças entre os dois sistemas.

## 1.4 As contribuições deste trabalho

Este trabalho desenvolve e emprega métodos de análise neural na detecção de elétrons no Segundo Nível de Filtragem do experimento ATLAS. Isto é feito baseando-se no modelo apresentado na Figura 1.1, onde os métodos de extração de características (fase 1) são analisados tendo por base sistemas de discriminação (fase 2) baseados em cortes uni-dimensionais sucessivos, detetores lineares baseados na Análise de Componentes Principais e Redes Neurais Artificiais. Os diversos sistemas são apresentados seguindo a modelagem do processamento dentro do LVL2, como indica a figura.

Utiliza-se como base um novo conjunto de simulações mais avançadas do detetor e dos eventos que ocorrerão no ATLAS. Esta nova base de dados contém uma estrutura de eventos com granularidade variante e com muitos dos detalhes operacionais do sistema de detecção final para o experimento. Estudos detalhados de tendências estatísticas, para as diversas combinações de compactadores e discriminadores, foram realizados, levando-se em consideração a energia das partículas ou sua localização no detetor (plano  $\eta \times \phi$ ).

Inicialmente, levando-se em conta a massa de dados disponível, determinou-se a eficiência de discriminação, que poderia ser obtida pelo sistema atualmente empregado no CERN (fase 1 + fase 2). Este sistema baseia-se na determinação de 4 variáveis altamente discriminantes (algoritmo T2Calo, fase 1), seguido de um processo de discriminação utilizando cortes unidimensionais (algoritmo EGammaHypo, fase 2). Determina-se sua eficiência em 91,85% para a detecção de elétrons contra um falso-alarme de 10,19% em jatos.

Com base nestes “valores canônicos de detecção”, observados com a análise deste sistema de discriminação elétron/jato, desenvolve-se um detetor linear para as 4 variáveis especialistas, substituindo-se o sistema de discriminação *ad-hoc* empregado no CERN. Este novo detetor tem um resultado equivalente ao sistema anterior, indicando uma inseparabilidade linear, inerente às variáveis no espaço compactado

provido pelo algoritmo T2Calo. Desenvolvemos então um detetor neural que atenda aos critérios da filtragem de elétrons combinado a esse algoritmo de extração de características. Esse novo sistema de discriminação apresenta uma eficiência de classificação superior e também maior robustez na separação elétron/jato, atingindo 92,38% na detecção de elétrons contra 9,05% de falso-alarme.

Seguindo as tendências atuais de processamento neural aplicadas à Física de Altas Energias, consideramos variáveis normalmente desprezadas para fins de separação elétron/jato, mas produzidas durante o processo de compactação do T2Calo, na criação de um novo sistema de discriminação mais eficiente. Este novo sistema, utilizando 14 variáveis, obtém 94,79% de eficiência na detecção de elétrons contra uma taxa de falso-alarme de apenas 7,93%, apresentando-se ainda mais robusto que o sistema T2Calo (4 variáveis) + detetor neural. Nesse momento, conduzimos uma análise de relevância das variáveis que revela que utilizando apenas 5 das 14 variáveis iniciais é possível desenvolver um detetor tão eficiente quanto aquele usando as 14 variáveis. Observando-se as correlações existentes entre as 14 variáveis, conduzimos uma Análise de Componentes Principais no objetivo de remover correlações no espaço de entrada do detetor neural, aumentando, dessa forma, o nível de compactação que poderia ser obtido.

Conduz-se então uma análise baseada no pré-processamento em anéis. Para esta análise, um novo sistema de mapeamento das informações do detetor em anéis foi elaborado, levando-se em consideração vários aspectos que ainda não haviam sido abordados: a problemática dos dados faltantes, a introdução de ruído proveniente da eletrônica, a variação de granularidade dentro da camada e camada a camada. Com base neste pré-processamento, 100 variáveis correspondentes aos anéis são extraídas de cada objeto a ser analisado. Um detetor neural é elaborado para a fase de detecção propriamente dita, levando-se em consideração uma otimização do espaço de parâmetros para seu treinamento. Este novo sistema obtém uma eficiência de classificação de elétrons de 96,55% contra apenas 3,12% de falso-alarme. Uma análise de relevância baseada das variáveis de entrada (anéis) ao detetor neural é conduzida, tendo em vista seu poder de discriminação. Finalmente, um detetor utilizando apenas 14 dos 100 anéis iniciais é desenvolvido, baseando-se na relevância discriminativa dos anéis de deposição energética. Este sistema atinge uma eficiência

de 97,59% na detecção de elétrons contra um falso alarme de 3,22%.

Atualmente, algoritmos candidatos ao Sistema de Filtragem devem atender todas as restrições, não somente de eficiência, mas também de tempo de execução, integrando-se perfeitamente à pilha de *software* do ATLAS. Uma vez tendo desenvolvido o sistema de detecção, ele foi completamente integrado a esta infraestrutura de processamento, de forma a verificar sua portabilidade, tempo de processamento e acuidade, enquanto embebido no inóspito ambiente de detecção do ATLAS. Testes de velocidade e exatidão são conduzidos em várias plataformas, donde se conclui que o sistema baseado em análise neural é viável para implementação no segundo nível de filtragem do experimento. Enquanto o sistema atual empregado no CERN consome cerca de 3,64 milissegundos para ser executado, o sistema baseado na extração em anéis e detecção neural, mais que 3 vezes mais eficiente na rejeição de falsos positivos, é executado em 3,8 milissegundos dentro da mesma infraestrutura de base, apenas cerca de 150 microssegundos de diferença, em média. Finalmente, uma implementação baseada em DSP's é derivada, indicando uma nova plataforma de desenvolvimento e a portabilidade desejada ao sistema.

Este trabalho foi desenvolvido com o objetivo de avaliar novas técnicas de detecção para o experimento ATLAS, no contexto da colaboração internacional entre o CERN e a UFRJ. As contribuições deste trabalho também incluem o desenvolvimento e a construção de boa parte do Segundo Nível de Filtragem e sua infraestrutura *online*, que serão narradas como um contínuo no decorrer do texto. Dentre as mais significativas, podemos destacar:

1. Desenvolvimento e aplicação da Unidade Central de Processamento no LVL2, a L2PU;
2. Desenvolvimento e aplicação da biblioteca global de formatação de dados do ATLAS (*eformat*);
3. Desenvolvimento e aplicação do *Pseudo* Sistema de Leitura (PROS);
4. Desenvolvimento e aplicação de uma biblioteca para rápido acesso e processamento de dados de calorimetria;
5. Desenvolvimento e aplicação da camada de *software* que combina a L2PU e a suíte de processamento *offline* Athena;

6. Medidas de desempenho do sistema de filtragem para diferentes cenários, com ou sem a utilização de algoritmos de discriminação.

## 1.5 Organização do Texto

O Capítulo 2 traz uma visão geral da Física de Partículas atual. Este capítulo inicia-se resumindo fatos históricos que culminaram na formulação do Modelo Padrão para a descrição das interações sub-atômicas. Em seguida, é feita uma revisão das técnicas de aceleração e detecção de partículas mais comuns em experimentos nessa área. Na Seção 2.3 descrevem-se detetores conhecidos como *calorímetros* em maiores detalhes. Aqui, discutem-se as partículas de interesse para o trabalho e a forma como interagem com este tipo de detetor. O capítulo termina introduzindo noções gerais sobre os sistemas de filtragem que equipam os experimentos modernos.

O Capítulo 3 narra especificidades do experimento ATLAS, que investigará a Física do bóson de Higgs. O capítulo é encabeçado por um breve histórico do CERN, adentrando por detalhes operacionais do experimento e do colisionador LHC, que gerará as interações próton-próton necessárias. A Seção 3.4.2 traz um resumo das características operacionais dos Calorímetros do ATLAS e um conjunto de referências para maiores detalhes.

O Sistema de Filtragem do ATLAS é introduzido no Capítulo 4. Este capítulo descreve todos os sub-sistemas que compõem o Sistema de Filtragem do experimento, desde o Primeiro Nível de Filtragem até chegar à gravação de um evento selecionado em mídia permanente. É neste ambiente que são executados os algoritmos de filtragem abordados neste trabalho. O Segundo Nível de Filtragem e algumas das contribuições deste trabalho são detalhados a partir da Seção 4.2.4. Resultados obtidos durante simulações do comportamento desta parte do sistema são apresentados na Seção 4.2.5. O final deste capítulo é dedicado ao ambiente de processamento de dados Athena e sua interação com o Sistema de Filtragem do ATLAS. Neste capítulo descreve-se em detalhes os componentes de infraestrutura produzidos no contexto da colaboração entre o CERN e a UFRJ.

O Capítulo 5 apresenta os aspectos da filtragem elétron/jato conduzida no experimento ATLAS, partindo do LVL1 até que os eventos estejam disponíveis no LVL2.

Os algoritmos de detecção de elétrons nesses dois níveis de filtragem são avaliados em detalhes. O capítulo termina com uma análise completa do algoritmo do LVL2, tendo por base os dados utilizados durante este trabalho de tese.

O Capítulo 6 começa trazendo uma revisão das técnicas de análise neural que são atualmente empregadas em experimentos em Física de Altas Energias, especialmente focado no emprego de redes neurais em calorimetria. O texto continua desenvolvendo os diversos sistemas de discriminação propostos. Na primeira parte, com sistemas baseados na reutilização do algoritmo de extração de características T2Calo e, numa segunda, utilizando um processamento baseado em anéis de deposição energética remodelado e melhorado.

O Capítulo 7 adentra pelo emprego do sistema desenvolvido no capítulo anterior, no Sistema de Filtragem do ATLAS. Demonstra-se que é possível construir e executar sistemas de detecção neural dentro desse complexo ambiente. Testes de acuidade e velocidade são conduzidos e analisados em diversas plataformas. Este capítulo termina com a descrição de uma implementação do sistema de anelamento e detecção neural em DSP's.

O Capítulo 8 traz uma conclusão dos estudos realizados e apresenta extensões de estudo para este trabalho.

Para a melhor compreensão dos termos utilizados na descrição da geometria dos objetos estudados neste trabalho, o Apêndice A introduz o sistema de coordenadas do ATLAS. O Apêndice B contém a descrição da base matemática de Redes Neurais Artificiais que foram utilizadas neste trabalho. O Apêndice C sumariza os detalhes de implementação do pacote de *software* `Neuralringer`, utilizado para as medidas de eficiência e relevância, que foi transplantado para operação dentro do Sistema de Filtragem do experimento. Finalmente, o Apêndice D traz um resumo da produção científica durante o período.

## Capítulo 2

# A Física de Altas Energias

Física de Partículas ou Física de Altas Energias (do inglês, *High Energy Physics*, comumente abreviado por *HEP*) denomina o ramo da física que estuda a matéria considerando-se sua característica sub-atômica, formada por partículas sub-atômicas como elétrons, quarks e prótons.

Neste capítulo, introduzimos seus conceitos fundamentais, iniciando com um breve histórico desta área. O material desta seção encontra-se detalhado em [3].

### 2.1 Um Pouco de História

Desde a Antiguidade, as pessoas têm buscado entender o comportamento da matéria: - *Por que objetos soltos no espaço caem no chão?*, - *Por que materiais diferentes têm propriedades diferentes?*, e assim por diante. Outros mistérios incluem a característica do próprio universo onde habitamos e o comportamento dos corpos celestes. Muitas teorias foram propostas desde então, a maior parte errada, embora este seja um efeito colateral da empreitada científica. As teorias físicas na Antiguidade eram baseadas somente em termos filosóficos e raramente verificadas por testes experimentais sistemáticos.

Já no início do século XVII, Kepler formulou um modelo do sistema solar baseado em cinco sólidos platônicos, em uma tentativa de explicar por que as órbitas dos planetas tinham os tamanhos relativos observados. Seu acesso às observações acuradas realizadas por outro cientista da época o capacitou a determinar que seu modelo estava inconsistente com as órbitas observadas. Depois de um esforço he-

róico de 7 anos, Kepler concluiu que os planetas não descrevem uma órbita circular, mas elíptica, tendo o Sol em um de seus focos. Kepler também sugeriu que uma “força” emanada do Sol afastaria os planetas de suas órbitas naturais, levando-os a perseguir uma órbita elíptica.

Durante este mesmo século, Galileu fez uso, pela primeira vez, de um experimento tendo por base o comportamento sub-atômico dos materiais, para validar suas teorias. O uso do experimento por Galileu e sua insistência de que os resultados observados terão sempre que anteceder os resultados teóricos eliminaram a aceitação de “dogmas” e fizeram nascer uma nova era, onde as idéias científicas eram abertamente discutidas e rigorosamente testadas.

Dentre outros grandes momentos, em 1687, Isaac Newton publicou o *Principia Mathematica*, detalhando dois compreensivos e bem sucedidos modelos físicos: a lei do movimento, de onde parte a mecânica clássica; e a lei da gravidade, que descreve esta força fundamental. Ambas as teorias concordavam bastante bem com experimentos executados na época. A mecânica clássica foi exaustivamente estendida por vários outros cientistas, que produziram novas formulações, princípios e resultados.

O início do século XX iniciou uma revolução na Física. As aclamadas teorias newtonianas mostraram-se incorretas em várias circunstâncias. Não somente a Mecânica Quântica mostrou que as leis do movimento não se confirmavam em pequenas escalas como, o que foi ainda mais inquietante, a relatividade geral mostrou que a relação fixa do espaço-tempo, do qual a mecânica newtoniana e a relatividade especial dependiam, não poderia existir.

Em 1904, Thomson propôs o primeiro modelo atômico, conhecido como “Pudim de Ameixas”. Já em 1905, apenas um ano depois, Einstein formulou a “Teoria de Relatividade Especial”, unificando espaço e tempo em uma entidade singular, chamada de espaço-tempo. Em 1915, o próprio estendeu sua teoria de relatividade especial com a “Teoria de Relatividade Geral”, substituindo a lei newtoniana sobre a gravidade. As teorias de Newton e Einstein concordam para regimes de pouca massa e energia.

Ainda em 1911, Rutherford descobriu a existência de um núcleo atômico compacto, consistindo de cargas positivas chamadas de prótons. Para tal, utilizou-se de um feixe de partículas alfa, emitido por uma fonte radioativa, uma folha de ouro



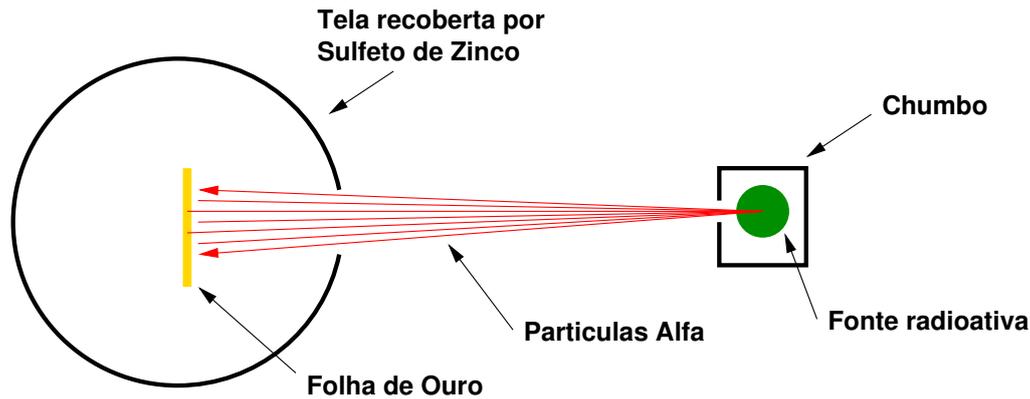


Figura 2.1: A configuração do experimento de Rutherford para a constatação de que o núcleo atômico era denso e positivo. Extraída de [3].

e um simples sistema de detetores feito de Sulfeto de Zinco para testar sua teoria [4]. As partículas alfa, ao baterem no detetor, marcavam-no (veja a configuração do experimento de Rutherford na Figura 2.1). Embora não pudesse ver o que ocorria no mundo sub-atômico, Rutherford podia teorizar, testar sua hipótese e então analisar os dados experimentais para verificar se seus cálculos estavam corretos. A única teoria cabível, após seus experimentos, é que o átomo teria que ser composto por um núcleo compacto e positivo e uma periferia negativamente carregada de partículas mais leves e menores que aquelas no núcleo.

Os nêutrons, os constituintes nucleares de carga neutra, foram descobertos em 1932 por Chadwick. Planck e Bohr entre outros físicos desenvolveram as teorias quânticas para explicar várias anomalias experimentais, introduzindo o conceito de níveis (ou *quanta*) de energia. Em 1925, Heisenberg e Schrödinger formularam a Mecânica Quântica, que sintetizou o formalismo matemático que substituiu a dinâmica de Newton. Nesta área, os resultados de medidas físicas são de natureza inerentemente probabilística. A teoria modela o cálculo destas probabilidades e, de forma bem-sucedida, descreve o comportamento da matéria em pequenas escalas.

A mecânica quântica também fornece ferramentas teóricas para a física de matéria condensada, que estuda o comportamento físico de sólidos e líquidos, incluindo fenômenos de cristalização, semi-condutividade e super-condutividade.

A Teoria Quântica dos Campos foi formulada para estender a mecânica quântica fazendo-a consistente com a relatividade especial, atingindo sua forma moderna no final dos anos 40 com o trabalho de Feynman, Schwinger, Tomonaga e Dyson. Eles

formularam a Eletrodinâmica Quântica, que descreve a interação eletromagnética.

A teoria quântica dos campos fornece o ferramental de fundo para a física de partículas moderna, que estuda primordialmente as forças elementares entre as partículas. Em 1954, Yang e Mills desenvolveram uma classe de teorias de Gauge, introduzindo o Modelo Atômico Padrão ou simplesmente Modelo Padrão, que foi completado nos anos 70. O Modelo Padrão descreve de forma bem-sucedida quase todas as partículas elementares observadas até hoje e abrange as forças fundamentais “forte”, “fraca” e “eletromagnética” assim como as partículas de que se compõe toda a matéria conhecida. O Modelo Padrão não é, no entanto, uma teoria fundamental de todas as interações existentes na natureza, primariamente por não descrever a gravidade.

## 2.2 Física de Partículas Moderna

Embora de grande sucesso, modelando resultados encontrados experimentalmente, o Modelo Padrão nunca foi aceito como uma teoria completa de física fundamental. Isto acontece por causa de dois aspectos:

1. O modelo contém 19 parâmetros livres, como a massa das partículas, que precisam ser determinados experimentalmente (e mais outros 10 para massas de neutrinos). Estes parâmetros não podem ser calculados independentemente;
2. O modelo não descreve a interação gravitacional;

Desde a finalização do Modelo Padrão, muitos esforços foram travados para resolver ambos os pontos. Um deles é conhecido como a Grande Unificação, do qual muitos aspectos estão bastante longe de serem observados em laboratório. O bóson de Higgs, que é previsto pelo modelo padrão, ainda não foi observado até os dias de hoje. Alguns experimentos, dentre os quais o experimento ATLAS (do inglês, *A Toroidal LHC Apparatus*) no CERN, Suíça, estão sendo construídos para o estudo desta física.

## 2.2.1 Aceleradores e Detetores

De forma similar ao que fez Rutherford, os físicos atuais usam um feixe de partículas aceleradas para recriar em laboratório condições que permitam o estudo das interações sub-atômicas que se deseja averiguar. Estes feixes podem colidir com um alvo fixo ou com um outro feixe de partículas, que é acelerado em direção contrária ao feixe primário. Para visualizar eletronicamente os sub-produtos de tais interações físicas, utilizam-se múltiplos detetores.

A aceleração das partículas resolve dois problemas que os físicos de hoje encontram em seus experimentos:

1. Comprimento de Onda - O comprimento de onda determina a acurácia do que é possível observar [3].

Uma vez que as partículas também apresentam características de onda, não é possível obter uma medida acurada usando-se partículas comuns, como um elétron, na observação de partículas muito pequenas. Um elétron não serve nem mesmo para observar outro elétron. A aceleração da partícula, no entanto, aumenta seu momento, diminuindo<sup>1</sup> seu comprimento de onda e permitindo que medidas acuradas possam ser tomadas, usando-se partículas maiores, como léptons.

2. Energia Cinética - Deseja-se, nos experimentos modernos, que o impacto seja o mais aniquilador possível. Isto é interessante, pois ao se aniquilar matéria, liberando muita energia, partículas mais massivas e menos estáveis são geradas. Ao se acelerar uma partícula, aumenta-se sua energia cinética, tornando a colisão com o alvo mais eficiente.

## 2.2.2 A Aceleração das Partículas

A aceleração de partículas é um processo bastante simples: inicialmente devem-se escolher partículas eletricamente carregadas para um experimento - elétrons, prótons ou íons são utilizados normalmente. As partículas eletricamente carregadas são posicionadas no interior de um túnel e aceleradas por pulsos eletromagnéticos

---

<sup>1</sup>O comprimento de onda ( $\lambda$ ) e o momento ( $p$ ) de um corpo são inversamente proporcionais -  $p = h/\lambda$ .

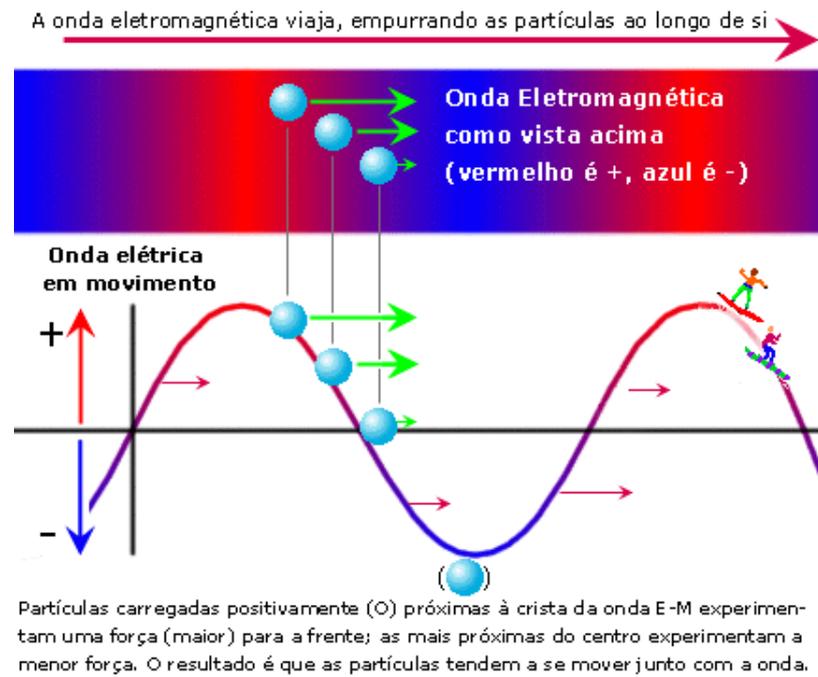


Figura 2.2: A aceleração de partículas eletricamente carregadas usando-se de pulsos e.m.. Extraído de [3].

(e.m). A Figura 2.2 exemplifica como elétrons possam ser acelerados: as ondas e.m. aceleram as partículas, pois elementos eletricamente carregados adquirem força (aceleração) quando envoltos por um campo eletromagnético.

Os aceleradores podem ter dois formatos: linear e circular. Num acelerador linear, as partículas são injetadas em uma extremidade e percorrem uma reta até que colidam com outras partículas ou com um alvo fixo. A outra possibilidade é ter um acelerador circular ou síncrotron. Num acelerador circular, as partículas são injetadas em um ponto do anel de aceleração e lá permanecem até que tenham adquirido velocidade suficiente ao experimento a que se destinam. A aceleração circular exige que ímãs poderosos curvem a trajetória das partículas injetadas. Aceleradores circulares também permitem que vários experimentos sejam conduzidos em pontos de sua circunferência, de forma simultânea.

### 2.2.3 Detecção: “Vendo” o que ocorreu após a colisão

Depois de um acelerador ter fornecido energia suficiente para suas partículas, estas são colocadas em rota de colisão com um alvo fixo ou, então, com as partículas

de um outro feixe acelerado. Cada uma dessas colisões forma um *evento* físico. O objetivo dos físicos é isolar cada evento, coletar dados a seu respeito e verificar se o processo do qual a partícula participou está de acordo com a teoria que está sendo testada no experimento em questão.

A análise de cada evento pode ser bastante complexa, já que muitas subpartículas podem ser produzidas. A maioria dessas partículas têm tempo de vida tão curto que viajam por distâncias extremamente curtas, antes de decaírem em outras partículas, sem deixarem pistas detetáveis.

Para procurar esses vários objetos e os produtos de seus decaimentos, os físicos projetam detetores com multi-componentes, que testam diferentes aspectos de um evento [5, 6, 7, 8]. Cada componente de um detetor moderno é usado para medir vários parâmetros das partículas provenientes de um evento, e/ou distinguir os diferentes tipos de objetos gerados. Quando todos esses componentes funcionam juntos para detetar um evento, partículas individuais podem ser distinguidas da multidão para efeito de análise e o evento original pode ser reconstruído.

Seguindo cada evento, os sistemas de processamento coletam e interpretam a vasta quantidade de dados dos detetores e apresentam os resultados extrapolados aos físicos.

Os físicos interessam-se pelos eventos que ocorrem durante ou mesmo depois da colisão das partículas. Por essa razão, colocam detetores em regiões nas quais os objetos resultantes daquela interação passarão. Os detetores são construídos de diferentes maneiras, dependendo do tipo de colisão analisada:

**Alvo fixo** Num experimento envolvendo um alvo fixo, as partículas produzidas geralmente projetam-se para a frente; por isso, os detetores são na forma de cones e são colocados ao longo da direção do feixe;

**Feixes de colisão** Durante um experimento envolvendo feixes em colisão, as partículas são espalhadas em todas as direções; assim, o detetor mais adequado é esférico ou, mais comumente, cilíndrico.

## Composição dos Detetores

Os detetores modernos são feitos de peças distintas, que testam diferentes aspectos de um evento. Esses vários componentes são organizados de tal maneira que os

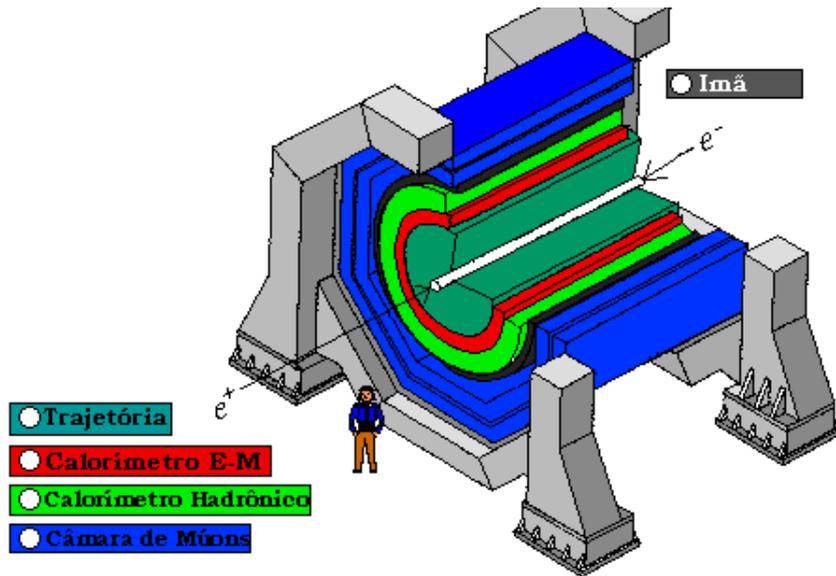


Figura 2.3: O diagrama de um detector cilíndrico moderno. Extraído de [3].

físicos possam obter o máximo de informação sobre as partículas geradas durante um evento. A Figura 2.3 mostra um diagrama esquemático de um pequeno detector cilíndrico moderno. Uma figura humana é mostrada em escala, indicando as grandes proporções dos detectores típicos em Física de Altas Energias.

Na Figura 2.3, é possível observar que um detector moderno é composto, basicamente, de 4 sub-detetores principais. O primeiro detector, de dentro para fora (tomando o feixe de partículas como origem), tem a função de determinar a rota e curvatura de partículas, por isto é chamado de detector de trajetória, arrasto ou, mais comumente, traços. A curvatura das partículas sobre o campo magnético a que são expostas no detector dá informações sobre a carga da partícula e seu momento. O segundo detector é chamado calorímetro eletromagnético (e.m.), e tem a função de determinar a energia total de elétrons, pósitrons e fótons (raios  $\gamma$ ). Estas partículas interagem com este detector, originando chuviscos compostos de outros elétrons, pósitrons e fótons. O terceiro detector, o calorímetro hadrônico, mede a energia total de chuviscos originados por hádrons (prótons, nêutrons ou mésons). No caso de partículas hadrônicas, este chuvisco é mais largo e profundo que o equivalente para partículas como elétrons e fótons. O quarto detector é um detector de múons. Somente estas partículas e neutrinos escapam dos outros sistemas de detecção. Neutrinos, infelizmente, nem por este último são detetados. Seu momento pode ser

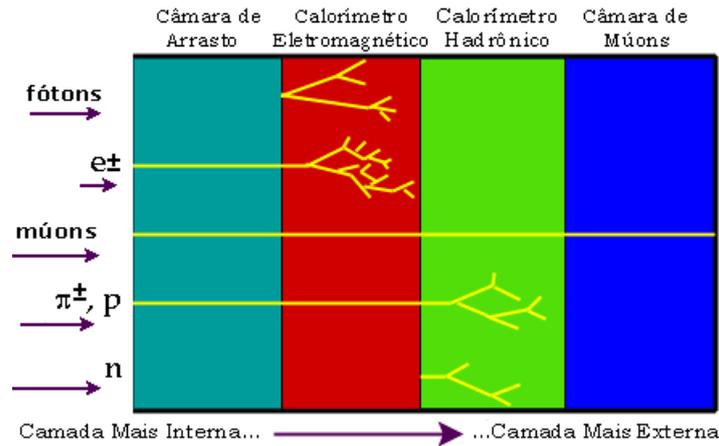


Figura 2.4: A interação de partículas com os detectores modernos. Extraído de [3].

estimado, no entanto, através da energia faltante no evento [7]. A Figura 2.4 mostra como alguns tipos de partículas interagem com estes detectores. Fótons, por exemplo, têm carga nula e não são detetados pelos detectores de traços, mas interagem com os calorímetros e.m.. Elétrons e pósitrons interagem tanto com os detectores de traços quanto com o calorímetro e.m., desenvolvendo cascata e.m. neste último. Prótons são detetáveis pelos detectores de traços, mas desenvolvem cascata somente nos calorímetros hadrônicos. Nêutrons somente desenvolvem cascata nos detectores hadrônicos.

## 2.3 Calorimetria Moderna

Calorímetros são detectores compostos utilizando a absorção total de partículas para medir a energia e a posição de partículas e jatos. Durante o processo de absorção, chuveiros são gerados por cascatas de interações. Durante a formação do chuveiro, eventualmente, a maior parte da energia da partícula será convertida em radiação ionizante, o que explica o nome deste tipo de detector. É claro, nenhuma temperatura é medida, mas as características de interação com a matéria (e.g. excitação atômica, ionização) são utilizadas para gerar um efeito detetável, através de partículas carregadas. A calorimetria é também uma forma de detetar e medir partículas neutras entre as secundárias produzidas em uma colisão de altas energias [9].

Estes detectores são normalmente construídos tendo por base diferentes sub-

detetores e têm seu desempenho otimizado para as diferentes partículas que incidirão sobre seu corpo. Cada sub-detetor de um calorímetro é por sua vez formado por múltiplas células de detecção, que medem a energia das partículas que interagem por seu volume. As células são alinhadas para formar torres tipicamente na direção de incidência. A análise da energia depositada nas células e torres permite medir o perfil lateral e longitudinal das cascatas, e daí sua geometria ser otimizada para este propósito.

Tipicamente, partículas eletromagnéticas incidentes (i.e. elétrons e fótons) são completamente absorvidos em calorímetros eletromagnéticos, a primeira das camadas de calorímetros compostos. Sua construção aproveita-se da geometria de chuveiros eletromagnéticos, que são compactos e curtos, para medir a energia e posição com precisão ótima para as partículas, que incluem  $\pi^0$ 's que decaem eletromagneticamente.

Hádrons incidentes (prótons, nêutrons, etc.), diferentemente, podem começar sua interação no calorímetro eletromagnético, mas quase sempre só serão completamente absorvidos em camadas posteriores, i.e. em calorímetros hadrônicos, construídos exatamente para conterem estes chuveiros. Chuveiros hadrônicos têm formato bastante diversificado.

A qualidade da discriminação entre chuveiros eletromagnéticos e hadrônicos é um critério importante para um calorímetro. É importante conter chuveiros eletromagnéticos em um espaço mais curto, sem que se inicie um grande número de chuveiros hadrônicos.

Calorímetros também provêem assinaturas para partículas que não são absorvidas: múons e neutrinos. Múons não formam chuveiro na matéria, mas sua carga deixa um sinal de ionização, que pode ser identificado em um calorímetro se a partícula está suficientemente isolada (e a eletrônica associada permitir), e ser associado a um traço detetado nos detetores de traço e/ou no detetor de múons. Neutrinos, por outro lado, não deixam sinais em um calorímetro, mas sua existência pode ser, algumas vezes, inferida através do princípio da conservação de energia: num calorímetro hermeticamente fechado, ao menos um neutrino suficientemente energético ou grupo de neutrinos pode ser observado tomando-se a soma vetorial de todos os outros momentos medidos, e a soma da energia observada em cada célula do ca-



lorímetro. A precisão de tais métodos, usualmente limitada à direção transversal, requer um vazamento mínimo de energia em todas as direções e daí o desafio para o projeto de calorímetros na prática.

### 2.3.1 Tipos de Calorímetros

Quanto à construção, é possível se distinguir os seguintes tipos de calorímetro [9, 10]:

- **Homogêneos:** Em calorímetros deste tipo, as funções de absorção e geração e leitura de sinais são combinadas em um único tipo de material. Estes materiais são quase que exclusivamente usados para calorímetros eletromagnéticos, por exemplo cristais (vidros dopados com chumbo), materiais compostos ou gases nobres em estado líquido;
- **Heterogêneos:** Também conhecidos como calorímetros de amostragem. Neste tipo de detector, as funções de absorção de energia e leitura do sinal são separadas. Isto permite uma escolha ótima do material de absorção, e liberdade no tratamento do sinal. Calorímetros heterogêneos são, em sua maior parte, construídos como sanduíches (e.g. chumbo, aço e ferro), alternando estes materiais com camadas de material ativo (e.g. cintiladores líquidos, sólidos ou contadores proporcionais). Somente parte da energia do chuveiro absorvida pelo material ativo é medida. Calorímetros hadrônicos, necessitando de uma profundidade considerável e largura para criar e absorver os chuveiros, são necessariamente deste tipo (veja a Seção 2.3.2).

Ainda que o desempenho não dependa fortemente da orientação, a espessura não pode variar em demasiado para assegurar uma resolução independente da direção e da posição dos chuveiros.

A calorimetria moderna é a arte de escolher entre duas restrições conflitantes; a principal é normalmente formulada em termos da resolução em energia, coordenadas espaciais, capacidade de filtragem, resistência à radiação dos materiais usados, e parâmetros eletrônicos como a faixa de operação e a extração de sinais. Na maior parte dos casos, o custo final de produção, a segunda restrição, é o parâmetro mais

limitante. O número de soluções em calorimetria é maior para calorímetros do que para detectores de traços, e soluções bastante engenhosas foram encontradas nos últimos 15 anos [10].

### 2.3.2 A Física da Calorimetria

A acurácia das medidas de energia em calorímetros aumenta com o aumento da energia de partículas incidentes [10], de acordo com a fórmula empírica:

$$\frac{\sigma_E}{E} \approx \frac{a}{\sqrt{E}} + b \quad (2.1)$$

Onde  $E$  é a energia da partícula incidente e  $\sigma_E$  representa o desvio padrão da medida de energia e  $a$  e  $b$  são constantes que dependem do tipo de detector, e.g. sua espessura e características das camadas ativas e passivas.

#### Calorimetria eletromagnética

Partículas eletromagnéticas perdem energia através de dois processos durante a interação com calorímetros:

- **radiação:** através do fenômeno conhecido como *Bremsstrahlung*, onde a partícula incidente muda de rota, perdendo energia e gerando um fóton. Este fóton seguirá um caminho de interação independente do elétron original, possivelmente através de espalhamento Compton, efeito foto-elétrico ou produção de pares elétron-pósitron, dependendo diretamente da energia com o qual foi gerado. Os elétrons remanescentes deste processo, se tiverem energia, poderão repetir este processo até que as partículas formadas não possam mais irradiar. Esta energia limite é conhecida como Energia Crítica ( $\varepsilon_c$ ) e depende do número atômico ( $Z$ ) do material por onde se desenvolve a cascata eletromagnética. Valores típicos de  $\varepsilon_c$  estão na faixa de dezenas de MeV (exemplo:  $\varepsilon_c(Cu) = 25$  MeV). Quanto maior o número atômico, maiores as chances de perda por radiação[11, 12].
- **ionização:** interagindo com os elétrons do material em que viaja a partícula, ionizando-o e desta forma perdendo energia.

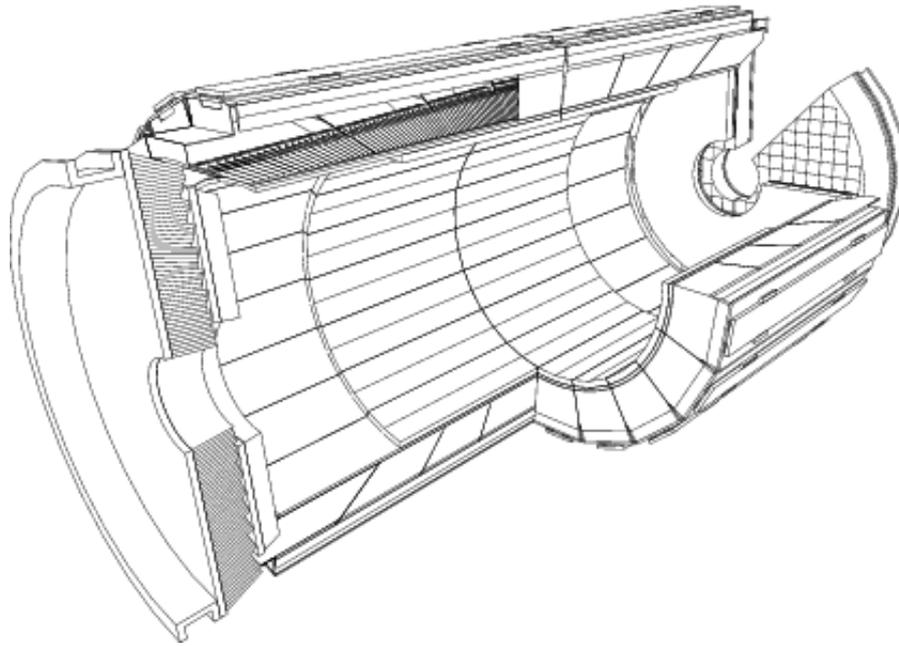


Figura 2.5: Diagrama da seção eletromagnética do calorímetro do experimento CMS.

Por estas razões, chuveiros eletromagnéticos são curtos e bem contidos em calorímetros pouco espessos. A cascata desenvolve-se a partir do eixo de penetração da partícula, de forma aproximadamente isotrópica, num espaço de tempo extremamente curto, na ordem de centenas de picossegundos.

A Figura 2.5 mostra um diagrama esquemático em três dimensões da seção eletromagnética do calorímetro do experimento CMS (do inglês *Compact Muon Solenoid*) no CERN, Suíça. Este calorímetro em formato cilíndrico é feito de cristais compostos. Na cavidade central deste aparato encaixam-se os detectores de traço do experimento. A segmentação do detetor é indispensável para a localização de objetos no espaço.

### Calorimetria hadrônica

O chuveiro hadrônico é dominado por uma sucessão de interações inelásticas deste tipo. Em altas energias, estas são caracterizadas pela produção de partículas múltiplas e pela emissão de partículas originárias de decaimentos nucleares do material excitado. Devido a freqüente geração de  $\pi^0$ 's (diz-se “pi-zeros”, ou seja, píons que não possuem carga elétrica), chuveiros hadrônicos têm também componentes eletromagnéticas.

Quando um hádron altamente energético penetra num bloco de matéria, ele, em algum ponto, interagirá com algum núcleo atômico. Neste processo, mésons são usualmente gerados (píons, káons, etc.). Outra fração da energia inicial da partícula é transferida para o núcleo com o qual o hádron interagiu. Este núcleo excitado liberará esta energia emitindo um certo número de núcleons (prótons ou nêutrons) e num estado posterior,  $\gamma$ 's de baixa energia, perdendo sua energia cinética por ionização. As partículas produzidas nesta reação (mésons, núcleons e  $\gamma$ 's), por sua vez, podem perder sua energia cinética por ionização ou induzir novas reações formando uma cascata ou chuva [10].

Limites intrínsecos na resolução em energia de calorímetros hadrônicos são [10]:

- Uma componente  $\pi^0$ , flutuante dentre os chuveiros secundários, que interaja eletromagneticamente sem qualquer outra interação nuclear ( $\pi^0 \rightarrow \gamma\gamma$ ). A fração de  $\pi^0$ 's é dada por  $\frac{\pi^0}{all} \approx 0,10 \log E$ ,  $E$  em GeV. Chuveiros hadrônicos podem desenvolver uma componente eletromagnética dominante;
- Uma boa parte da energia disponível é convertida em excitação e quebra de núcleos atômicos. Somente uma pequena fração desta energia aparecerá como sinal detetável, havendo largas flutuações evento-a-evento. Estas flutuações podem ser compensadas com a boa escolha dos materiais ativos e absorvedores do detetor;
- Uma fração considerável da energia da partícula incidente é despendida em reações que não resultarão em sinal observável, tais como vazamento de energia em várias formas (evaporação ou quebra nuclear, excitação nuclear, etc.).

De forma contrária a chuveiros eletromagnéticos, que se desenvolvem num curto espaço de tempo, chuveiros hadrônicos são caracterizada por diferentes escalas de tempo. Os fenômenos mais lentos podem durar de centenas de nanossegundos a um microssegundo.

Sumarizando:

1. Calorímetros são sensíveis a partículas neutras e carregadas;

2. Devido a diferenças na forma de deposição de energia das partículas, a identificação de partículas com alta eficiência pode ser atingida usando-se calorímetros;
3. Quanto maior a energia da partícula, mais acurado é o resultado. Isto não acontece com outros tipos de detetores;
4. Para conter o desenvolvimento de cascatas dos objetos a serem medidos, a profundidade dos calorímetros aumenta logaritmicamente com a energia, o que permite o projeto de detetores mais compactos;
5. Não precisam de campos magnéticos (como os detetores de traços);
6. Podem ser segmentados, o que permite acurada medida da energia e a visualização da trajetória das partículas;
7. Resposta rápida (melhor que 50 ns) pode ser atingida, o que é importante num ambiente com alta taxa de eventos;
8. A informação de energia pode ser usada para filtrar eventos interessantes com alta seletividade;

## 2.4 Sistemas de Filtragem

Os sinais detetados por calorimetria são normalmente utilizados como fonte de discriminação rápida de eventos interessantes em experimentos modernos. Um Sistema de Filtragem, ou *Trigger*, é um conjunto de dispositivos, normalmente uma combinação de componentes eletrônicos e computadores, provendo um sinal rápido quando a física interessante acontece [9, 13, 14, 15, 16, 17]. Tipicamente, um sistema de filtragem está associado a experimentos com detetores e seus sinais ativam o sistema de gravação de eventos para registrar parte ou a totalidade dos impulsos captados em mídia permanente. As condições que levam o sistema de filtragem a produzir o sinal de aceitação são comumente chamadas de assinaturas do evento. As condições podem ser tão simples quanto identificar um traço gerado por uma partícula carregada passando por cintiladores durante um período de tempo, ou tão

complexos quanto o critério de massa efetiva entre léptons identificados que devam satisfazer colisões de altas energias.

Em muitos experimentos, a aquisição de dados, através do tempo morto que causa, é um fator crítico determinante, que limita as estatísticas e o potencial físico. Um sistema de filtragem eficaz é, desta forma, o ponto-chave para a transmissão de dados que têm alta probabilidade de conter física de interesse e rejeição, com base nas possibilidades de detecção, de todos ou da maioria dos eventos que representem física ordinária e trivial.

Dependendo do acelerador utilizado, sistemas de filtragem podem ser chaveados (e.g. por pacotes de partículas chegando ao ponto de colisão) ou permanentemente ativados (como para o estudo de raios cósmicos). As implementações podem ser síncronas ou operar em tempo real, ou ainda se comporem de vários circuitos assíncronos operando paralelamente, respeitando uma unidade de controle central (que também tem a função de ressincronizar o sistema como um todo). As implementações nos dias de hoje, vão desde simples portas E/OU até *Field-Programmable Gate Arrays* ou FPGA's. Os tempos de retardo dependem do volume dos dados a serem analisados (e comumente da ocupação de recursos de processamento a que se destinam). Algoritmos de filtragem também podem ter seu desempenho variando no tempo, já que dependem da complexidade dos objetos em análise.

Em grandes experimentos, sistemas de filtragem são implementados em múltiplos níveis, tipicamente consistindo de um primeiro nível síncrono, que identifica candidatos a partir de um subconjunto dos dados colhidos pelos detetores, reduzindo a taxa de eventos por algum fator. Subseqüentemente, os dados são digitalmente transmitidos para bancos de memória (do inglês *buffers*) e para um segundo nível de filtragem, normalmente assíncrono, onde algoritmos mais complexos, baseados em um conjunto de dados mais completo, conseguem reduzir novamente a taxa de eventos. Eventualmente, depois de uma terceira ou quarta iteração, o evento é gravado em mídia permanente.

### 2.4.1 Sistemas de filtragem em experimentos modernos

Um experimento em Física de Altas Energias como o DZERO [5], localizado no FermiLab, próximo a Chicago nos Estados Unidos, é um exemplo de um experimento

com um sistema de filtragem moderno. O cruzamento de pacotes de prótons e anti-prótons que geram as colisões a serem estudadas ocorrem numa taxa de 2,5 MHz (ou seja, 2,5 milhões de colisões por segundo). Cada colisão poderá produzir um evento que possui 0,25 *megabytes* em dados. Se a taxa total fosse escrita em fitas magnéticas<sup>2</sup>, assumindo um custo de 40 centavos de dólar por *Gigabyte*, o custo total seria de 10,5 milhões de dólares por ano. No entanto, a maior parte das colisões de pacotes contém interações elásticas ou até mesmo nulas. Para tirar proveito desta natureza, o experimento DZERO emprega um sistema de filtragem em múltiplos níveis para reduzir a taxa de eventos registrada a 50 Hz, implicando em um custo aproximado de 75.000 dólares por ano em fitas magnéticas.

É claro que a economia em fitas magnéticas não é a única razão da existência de um sistema de filtragem. Planejar e projetar um sistema de aquisição de dados com uma taxa de dados de 600 *gigabytes* por segundo seria proibitivo economicamente, especialmente considerando-se o valor relativo dos dados coletados. Os atuais programas de reconstrução de eventos também consomem grande poder de processamento e, se a entrada puder ser sensivelmente reduzida em volume, menor a quantidade de recursos computacionais serão necessários para a análise *offline*.

O enfoque usado para projetar e implementar um sistema de filtragem depende fortemente do domínio do problema. Um experimento moderno em Física de Altas Energias deve estar apto a responder a um novo evento, em geral, a cada dezena ou dezenas de nanossegundos. O tamanho do evento para grandes experimentos (veja a Figura 2.6) também tende a ser grande, chegando a alguns *megabyte* por evento em alguns dos experimentos do LHC (*Large Hadron Collider*).

Experimentos modernos usam um sistema de filtragem dividido em múltiplos níveis para endereçar o problema das taxas de entrada. Os níveis mais baixos implementam algoritmos simples e rápidos, normalmente usando FPGA's e *hardware* personalizado, que rejeitem os processos (*background*) mais simples e abundantes. Os níveis mais altos são progressivamente mais sofisticados e também requerem mais tempo para atingir níveis de rejeição aceitáveis. Ademais, carregando uma menor quantidade ou melhor qualidade de dados nos níveis iniciais de filtragem, estes ex-

---

<sup>2</sup>Este tipo de tecnologia ainda é atualmente utilizado para a armazenagem de grandes quantidades de dados ( $> 10^2$  *Terabytes*).

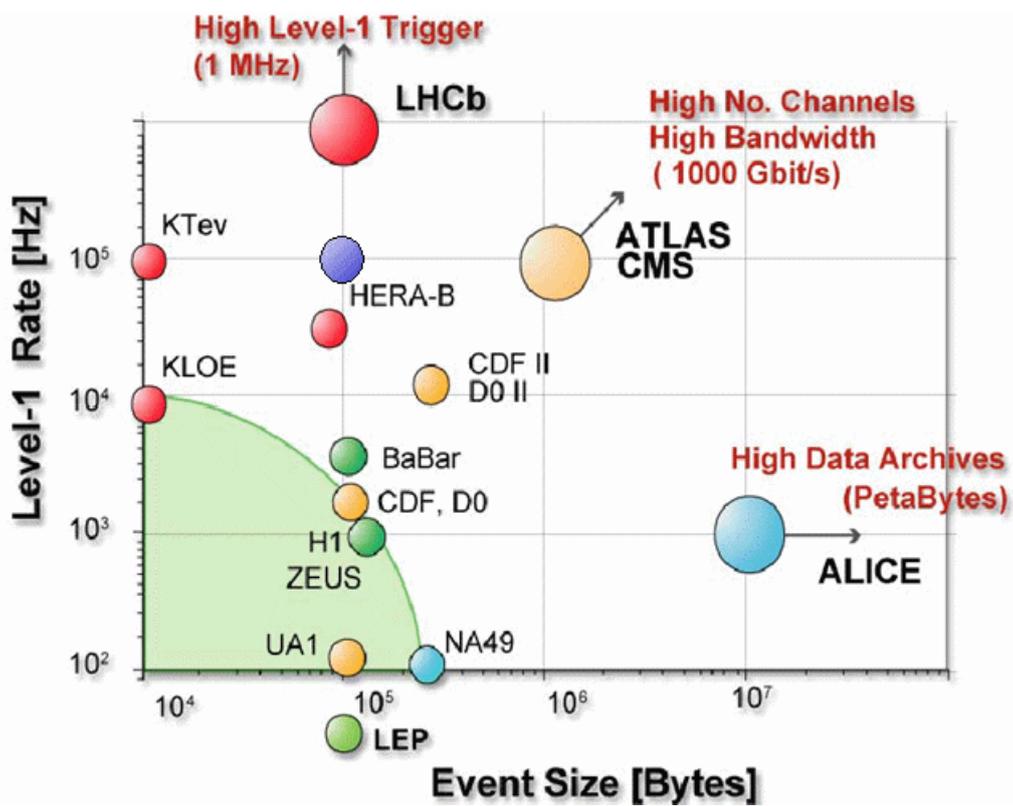


Figura 2.6: O tamanho do evento *versus* a taxa de produção em experimentos atuais. Extraído de [17].



perimentos conseguem reduzir o fluxo de informação no sistema de leitura: tão logo um evento seja rejeitado, seus dados são descartados.

Classicamente, o nível em *hardware* é chamado de Primeiro Nível. O seguinte, composto de partes personalizadas em *hardware* e computadores, é chamado de Segundo Nível. O último estágio, chamado de Terceiro Nível, composto normalmente de redes de computadores tipo PC, realiza a classificação final. Comumente o segundo e terceiro níveis de filtragem são conhecidos como Filtros de Alto Nível ou *High-Level Triggers*, HLT.

## Capítulo 3

# O Experimento ATLAS no CERN

O experimento ATLAS é um dos 4 experimentos que estão sendo construídos ao redor do super-acelerador e colisionador de partículas LHC (do inglês *Large Hadron Collider*), no Laboratório Europeu para a Física de Partículas - CERN em Genebra, Suíça [7]. O CERN é um dos maiores, senão o maior laboratório para a Física de Altas Energias na atualidade, conduzindo vários experimentos em regime de colaboração internacional. A próxima seção introduz um pouco de sua história, que culmina na aprovação para criação do LHC e dos 4 grandes experimentos montados a partir de sua infraestrutura.

### 3.1 O Laboratório CERN

O CERN surgiu em 1952 quando, depois de várias reuniões, 11 Estados europeus aceitaram organizar um Conselho Europeu para a Pesquisa Nuclear (do francês *Conséil Européen pour la Recherche Nucléaire*) Provisório, com sede prevista em um local próximo à Genebra, na Suíça. Dois anos após a ratificação de uma convenção por seus Estados-Membros, no dia 29 de maio de 1954, um centro foi inaugurado. Embora o termo “Provisório” tenha sido banido do nome do laboratório, a sigla CERN continuou a ser empregada, tendo o significado da letra *C* passado a *Centro* ao invés de *Conselho*.

O primeiro acelerador do CERN, um *Synchro-Cyclotron* de prótons a 600 MeV, começou a operar em 1957. Este acelerador foi responsável pela obtenção do primeiro resultado experimental do laboratório: a observação do decaimento de um pión em

um elétron e um neutrino. Em 1959, a primeira “grande” máquina do CERN já estava em operação. Era um *Proton Synchrotron* (PS) de 28 GeV - o acelerador de maior energia na época.

Algumas das tecnologias para colisão e detecção de partículas existentes nos dias de hoje foram inventadas neste laboratório. Dentre as principais, podemos destacar:

- Técnica de Resfriamento Estocástico (do inglês *Stochastic Cooling Technique*), proposta por Simon van der Meer em 1968;
- as Câmaras Proporcionais Multifio (do inglês *Multiwire Proportional Chambers*) e as Câmaras de Arrasto (do inglês *Drift Chambers*). A criação desta tecnologia deu o Prêmio Nobel a Georges Charpack em 1992;

Outras tecnologias associadas também tiveram destaque. Por exemplo, em 1990, Tim Berners-Lee, trabalhando em conjunto com Robert Cailliau, propôs um sistema de informação distribuída, baseado em “hipertextos”, uma forma de descrever ligações a documentos guardados em computadores. Escondendo o endereço de rede através de ítems grafados na tela, a informação pode ser ligada através de vários computadores. O nome World-Wide Web foi escolhido na ocasião.

A injeção de fundos para a pesquisa nuclear alavancou a criação de uma grande infra-estrutura de anéis de aceleração e estocagem de partículas na década de 60, dos quais o mais importante aparato foi o *Super Proton Synchrotron* (SPS), que começou a operar em 1976 com energia inicial de 300 GeV. O SPS têm 7 km de extensão e atravessa a fronteira franco-suíça nos arredores de Genebra, tendo sido o maior instrumento científico da época. A histórica descoberta dos bósons W e Z em janeiro e maio de 1983, respectivamente, foi possível através dos trabalhos no SPS, confirmando a teoria eletrofraca, que unificava as forças fraca e eletromagnéticas. Carlo Rubia e Simon van der Meer receberam o prêmio Nobel por este trabalho em 1984. Este trabalho motivou a construção do Grande Colisionador Elétron-Pósitron (do inglês *Large Electron-Positron Collider*, LEP), com energia inicial prevista para 90 GeV. O LEP estudaria em detalhes as partículas Z em quatro experimentos simultâneos.

Os anos de 1989 a 1994<sup>1</sup> foram marcados pelo sucesso dos experimentos no LEP.

---

<sup>1</sup>Em verdade, as pesquisas com os dados gerados no LEP estendem-se até os dias de hoje.

## Accelerator chain of CERN (operating or approved projects)

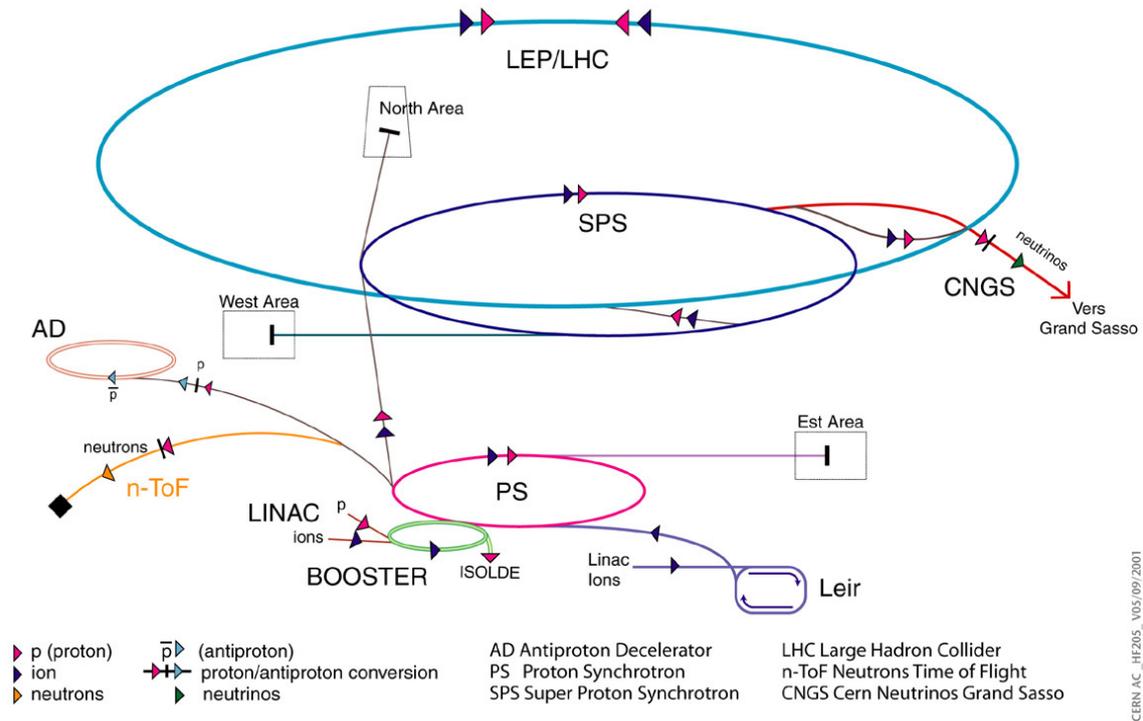


Figura 3.1: Diagrama esquemático dos aceleradores do CERN e suas conexões. Extraído de [18].

Um dos resultados mais importantes se deu com a precisa medição dos parâmetros de ressonância da partícula Z: os 4 experimentos do LEP reconstruíram mais de 10 milhões de decaimentos desta partícula no período. Com base nestes resultados, o Conselho dos Estados-Membro aprovou a construção do Grande Colisionador de Hádrons (do inglês *Large Hadron Collider*, LHC). A Figura 3.1 mostra o atual conjunto de aceleradores do CERN e suas interconexões.

## 3.2 O LHC

No LHC, a energia disponível nas colisões entre os constituintes dos prótons (*quarks* e *glúons*) chegará a faixa dos TeV's, o que representa aproximadamente 10 vezes a potência atingida pelo LEP ou pelo Tevatron no Fermilab [19]. De forma a manter um programa de física efetivo a uma energia mais alta (digamos,  $E$ ), a luminosidade de um colisionador, uma quantidade proporcional ao número de coli-

sões por segundo, deve aumentar proporcionalmente com  $E^2$ . Isto acontece porque o comprimento de onda associado à partícula decresce com  $1/E$  e, portanto, a seção transversal da partícula decrescerá com  $1/E^2$ ! Se o tamanho da partícula decresce, menor a chance de ocorrer uma colisão. Para compensar este efeito, aumenta-se o número de colisões por base de tempo. Enquanto colisionadores mais antigos podem atingir uma luminosidade de  $L = 10^{32} \text{cm}^{-2} \text{s}^{-1}$ , no LHC este parâmetro será de  $L = 10^{34} \text{cm}^{-2} \text{s}^{-1}$ . Este valor será atingido alimentando-se cada um dos dois anéis com 2835 pacotes de  $10^{11}$  partículas cada.

Quando dois pacotes cruzam o centro de um detetor, somente uma pequena fração das partículas colide de forma aproximadamente inelástica, provocando uma aniquilação bastante eficiente da matéria, para produzir os eventos de interesse. Todas as outras são defletidas pelo forte campo magnético do outro pacote. Estas defleções, que são mais intensas para pacotes mais densos, acumulam-se volta após volta e, com isso, podem eventualmente acarretar na perda de partículas. Este efeito do feixe foi estudado em colisionadores mais antigos e a experiência mostrou que não é possível aumentar muito a densidade do pacote de partículas acima de um limite feixe-feixe para preservar uma vida útil suficientemente grande para o feixe. Para atingir a luminosidade desejada, o LHC tem que operar bastante próximo deste limite. Seus injetores, o antigo PS e o SPS, estão sendo re-equipados para prover exatamente a densidade de feixe desejada.

Os feixes do LHC serão armazenados por aproximadamente 10 horas em altas energias. Durante este período, as partículas fazem 400 milhões de revoluções ao redor da máquina, um número verdadeiramente grande. Enquanto isso, a amplitude das oscilações ao redor da órbita central não pode aumentar significativamente, pois isso dissolveria o feixe e degradaria a luminosidade do colisionador. Isto é bastante difícil de ser atingido, já que, além da interação feixe-feixe mencionada anteriormente, pequenas não-linearidades nos componentes geradores dos campos magnéticos para a focalização do feixe podem fazer o movimento se tornar ligeiramente caótico de forma que, depois de várias voltas, partículas podem ser perdidas. No LHC, os efeitos mais desestabilizadores das imperfeições magnéticas são produzidos durante a injeção das partículas no grande anel. Para controlar os estragos deste tipo de efeito, utilizam-se computadores rápidos para rastrear o traçado de centenas

de partículas passo-a-passo, através dos milhares de ímãs espalhados pelo LHC, até completarem cerca de 1.000.000 de voltas, compensando-se o efeito dos campos para estabilizar o sistema.

Quatro grandes experimentos se basearão no LHC:

- **ALICE**: estuda a física da matéria com fortes interações em energias extremamente densas [20];
- **LHCb**: para realizar medidas precisas da violação CP e decaimentos raros de hádrons B (compostos com o quark *beauty*, ou b-quark) [21];
- **ATLAS** [22] e **CMS** [6]: experimentos com propósitos múltiplos, incluindo um programa para a detecção do bóson de Higgs. Estes experimentos também incluem programas para estudo de Física B ou envolvendo íons pesados. Os detetores são bastante complexos e podem ser reconfigurados facilmente para estudo de nova física, na remota possibilidade de não serem encontradas evidências do bóson de Higgs.

Uma vez que o LHC será construído sob a superfície (aproximadamente a 100 metros de profundidade), estes experimentos serão montados em poços comunicantes com a estrutura do acelerador. A Figura 3.2 mostra uma visão esquemática da fronteira entre Suíça e a França, indicando a localização dos poços dos experimentos acima, tal como a localização dos domínios do CERN, em *Meyrin* na Suíça e em *Preessin* na França.

O LHC estará apto a produzir cerca de 40 milhões de interações por segundo (40 MHz) e estará dedicado a cada um dos experimentos de forma alternada. Cada período de aquisição programada ou *run* durará cerca de 10 horas. Os experimentos terão este período para adquirir dados e perseguir o objetivo físico a que se destinam.

### 3.3 O Experimento ATLAS

Um dos objetivos principais do experimento ATLAS é descobrir e estudar o bóson de Higgs. Esta partícula tem importância crítica para o Modelo Padrão, como destacado no Capítulo 2, uma vez que a interação de outras partículas com bósons desta natureza, induz a existência de massa.

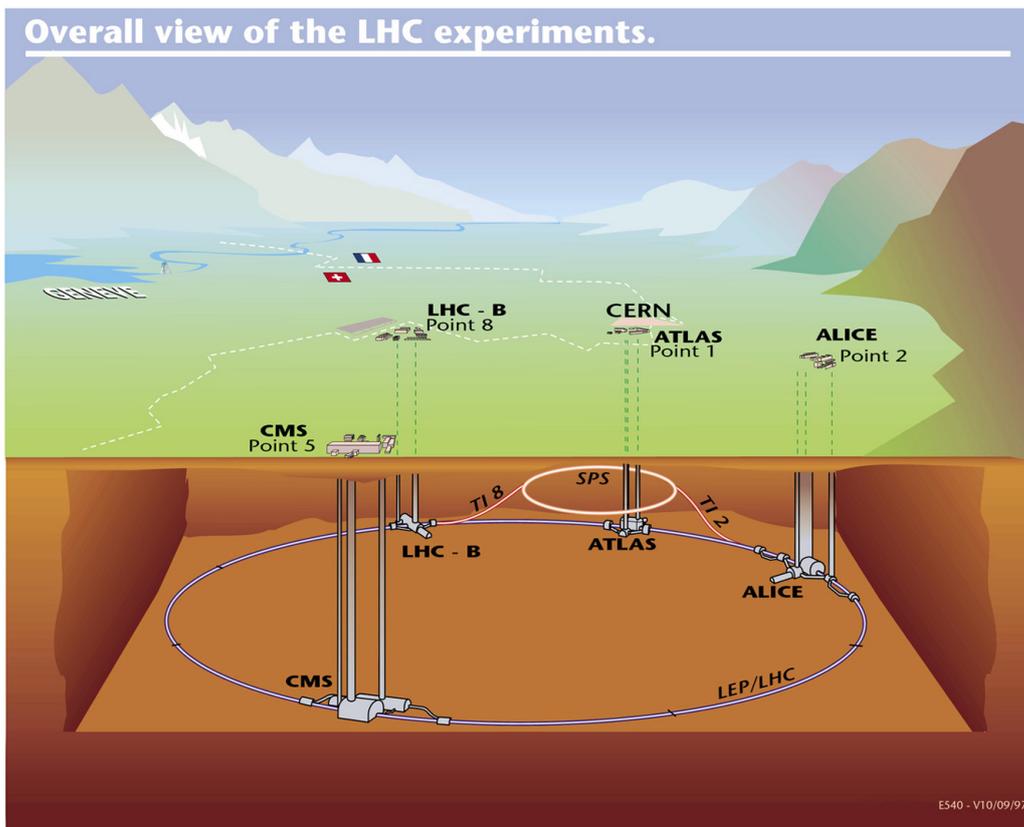


Figura 3.2: Visão esquemática da fronteira franco-suíça mostrando o CERN e os pontos de acesso aos poços dos 4 experimentos do LHC. Extraído de [18].

**A massa** É interessante notar que um conceito tão familiar como a massa não era entendido até a proposição do Modelo Padrão. A maioria de nós está familiarizado com campos magnéticos, elétricos e gravitacionais. Uma pessoa no campo gravitacional da Terra sente a sua força. Ondas eletromagnéticas viajam pelo espaço da mesma forma que ondas se propagam sobre água. Se este balanço fosse descrito pela Mecânica Quântica, a superfície da água que carrega as ondas seria chamada de “campo”.

O Modelo Padrão propõe que exista um outro campo ainda não observado, um campo que é praticamente indistingüível do espaço vazio. Este campo é normalmente chamado de campo de Higgs. Acredita-se que todo o espaço seja preenchido com este campo e, interagindo com ele, partículas adquirem suas massas. Partículas que interajam fortemente com o campo são pesadas, enquanto que aquelas que interajam fracamente tornam-se mais leves.

O campo de Higgs tem ao menos uma nova partícula associada a ele - o bóson de Higgs. O detetor ATLAS será capaz de detetar esta partícula, se existir, tornando-se uma das maiores descobertas físicas de todos os tempos.

### 3.4 O Detetor ATLAS

O detetor ATLAS é um dispositivo em formato cilíndrico, capaz de registrar os subprodutos de colisões próton-próton a 40 MHz. A Figura 3.3 mostra um esquema do detetor. Repare na parte inferior à direita, em escala, o tamanho proporcional das pessoas.

O detetor ATLAS é composto, como mostra a figura, de 4 sistemas de detecção independentes. Na parte mais interna, encontram-se detetores de traço altamente segmentados, que podem medir a trajetória de partículas carregadas. Ao redor deste sistema, ficará instalada a seção eletromagnética do calorímetro. Em seguida, vemos a seção hadrônica dos calorímetros do ATLAS, seguida dos solenóides que curvarão a trajetória de partículas eletricamente carregadas, e, finalmente, englobando toda a estrutura, detetores de múons. A posição do feixe com relação ao detetor também está destacada na figura.



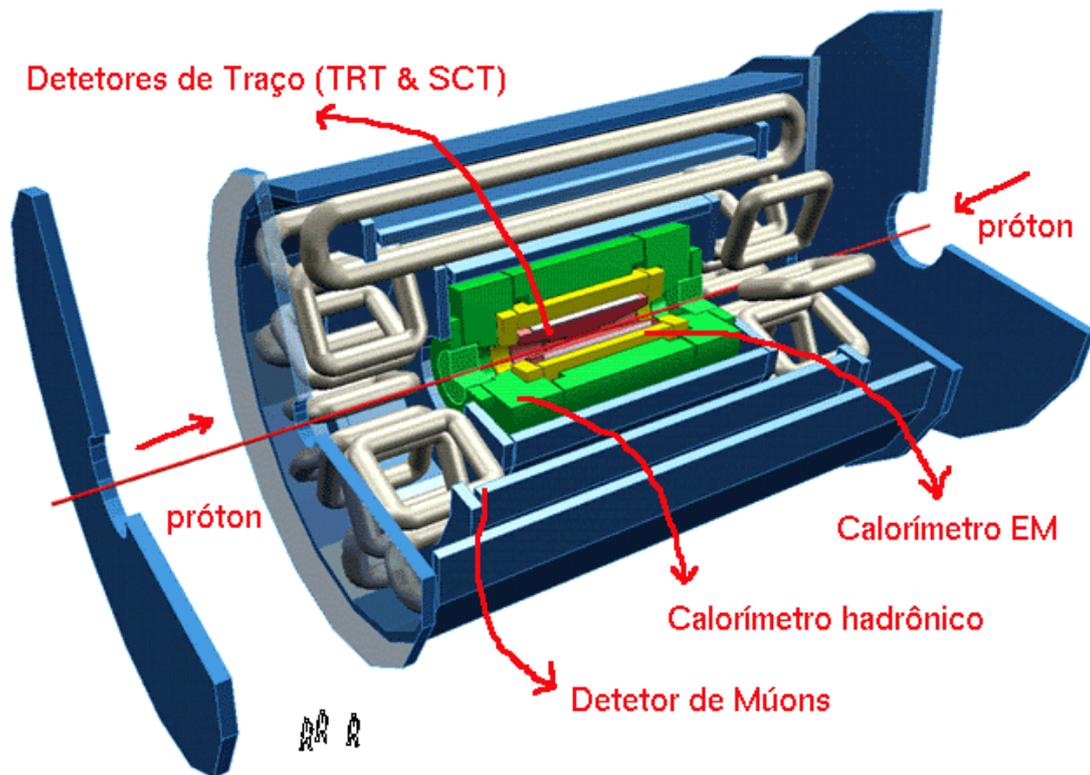


Figura 3.3: O detector ATLAS. Extraído de [18].

### 3.4.1 Os Detetores de Traços

O papel do Detetor Interno (do inglês *Inner Detector*, ID) é reconstruir traços e vértices de um evento com alta eficiência juntamente com o calorímetro (veja a Seção 3.4.2) e os detetores de múons (veja a Seção 3.4.3), para o reconhecimento de elétrons e múons, e suprindo assinaturas extras para vértices provenientes de partículas que decaem rapidamente [23]. Sua aceitação cobre a região de pseudo-rapidez (veja o Apêndice A) de  $[-2, 5; 2, 5]$ , igualando, em alcance, o restante dos sistemas do detector ATLAS.

A Figura 3.4 mostra um corte tridimensional do ID. Este sistema combina detetores de alta resolução na parte interna com elementos para a detecção de traços em modo contínuo, na parte externa, envolvidos por um solenóide com campo central de 2 Tesla.

Os alvos de momento e resolução de vértices requerem que medições de alta precisão sejam feitas com detetores de alta segmentação, dada a grande densidade de traços esperada em experimentos ao redor do LHC. Detetores de traços baseados

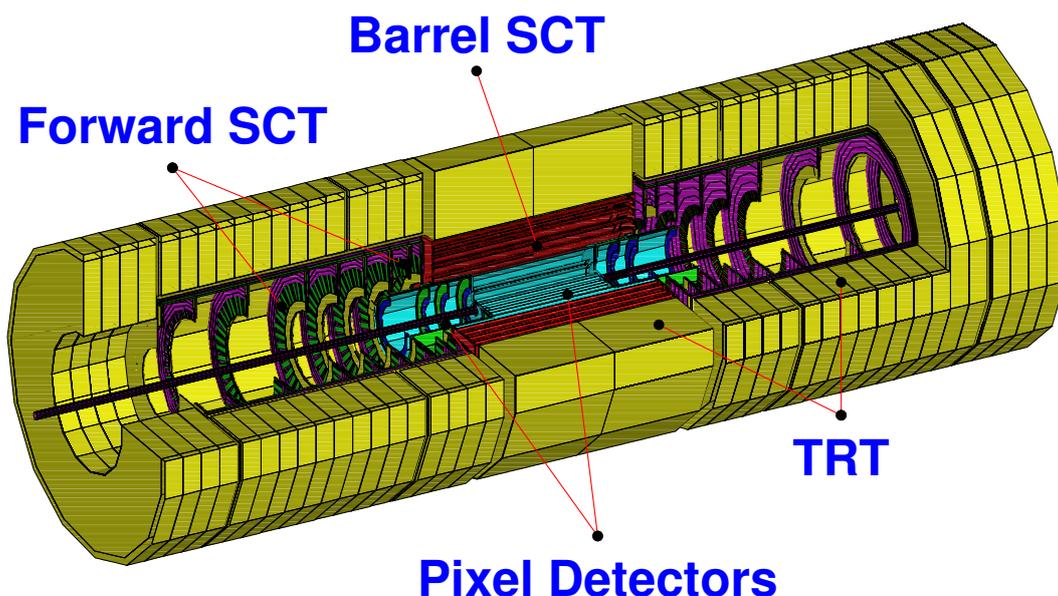


Figura 3.4: Corte tridimensional do Detetor Interno do ATLAS.

em semi-condutores (do inglês, *Semi-Conductor Tracker*, SCT), usando pastilhas e pontos (ou *pixels*) microscópicos de silício oferecem estas características e, portanto, esta tecnologia é utilizada na parte mais interior do ID. Entretanto, o número total de camadas de precisão deve ser limitado por causa do excesso de matéria que estes detetores introduzem e do seu alto custo. Segundo o projeto do ID, a trajetória das partículas cruzará, na pior das hipóteses, 4 camadas com sensores em formato de pastilhas e 3 camadas com sensores em pontos de silício.

Um grande número de pontos da trajetória (tipicamente 36 por traço) será detectado por um Detetor de traços em microtubos (do inglês, *Straw Tube Tracker*, TRT) que possibilita o acompanhamento das trajetórias em modo contínuo com uma quantidade muito menor de material por ponto e baixo custo. A combinação destas duas técnicas propicia um reconhecimento de trajetórias bastante robusto e com alta precisão, tanto em  $\phi$  como em  $z$  (veja a definição completa no Apêndice A). A sensibilização dos microtubos na parte exterior do detetor contribuirá para a detecção do momento da partícula, onde a baixa precisão do sistema, se comparado com o SCT, é compensada pelo grande número de pontos medidos numa maior seção do espaço.

### 3.4.2 Os Calorímetros

Os calorímetros desempenham um papel central na arquitetura do experimento ATLAS [24]. Estes detetores foram projetados para contribuir ativamente na filtração de física interessante e na medição precisa de elétrons, fótons, jatos e energia transversa perdida (do inglês *missing  $E_T$* ) no difícil ambiente do LHC, trabalhando na sua máxima luminosidade.

O sistema pode ser caracterizado funcionalmente ou arquiteturalmente. Funcionalmente é possível classificar os calorímetros do ATLAS em:

- **Eletromagnético ou EM ou e.m.** (*EM Accordion Calorimeters*): Para a detecção de partículas que desenvolvem cascatas e.m. como elétrons ou fótons, situando-se na parte mais interior da estrutura;
- **Hadrônico** (*Hadronic Tile Calorimeter e Hadronic LAr End Cap Calorimeter*): Para a detecção de partículas ou jatos que desenvolvam cascatas baseadas em hádrons, como nêutrons, prótons e píons.
- **Avançado ou *Forward*** (*Foward LAr Calorimeters*): Para a detecção de partículas próximas ao eixo do feixe de colisão. Estes dispositivos tendem a ser menos precisos. Este calorímetro não será abordado neste trabalho, em nenhum de seus níveis, uma vez que é comumente utilizado somente para efeitos de análise *off-line*.

Com relação à arquitetura dos calorímetros, é possível classificar os detetores da seguinte forma:

- **Argônio Líquido com eletrodos de chumbo:** É um calorímetro que utiliza o Argônio Líquido como material absorvedor e eletrodos de chumbo para captar os íons formados durante a interação da partícula com seu volume. Os eletrodos são dispostos em forma de "acordeão", o que confere a esta estrutura um formato peculiar. Esta tecnologia foi escolhida para a maior parte do sistema, cobrindo toda a seção e.m., as tampas hadrônicas e o calorímetro avançado [25].

# ATLAS Calorimetry (Geant)

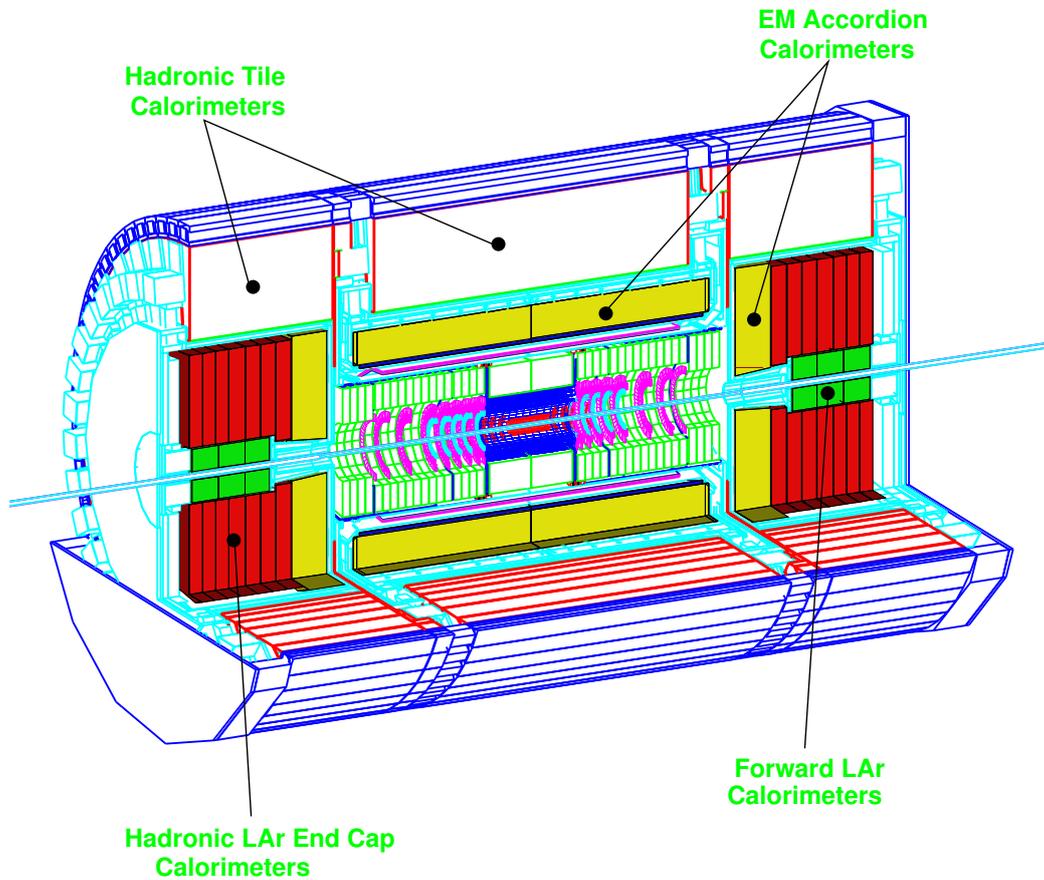


Figura 3.5: Diagrama esquemático dos calorímetros no detetor ATLAS.

- **Telhas Cintilantes ou *Tiles***: É um calorímetro de amostragem que utiliza uma liga de aço como material absorvedor e telhas cintilantes<sup>2</sup> como elemento amostrador. Esta técnica foi escolhida somente para a seção hadrônica do barril [26].

A Figura 3.5 mostra uma esquematização dos calorímetros do ATLAS, indicando as partes mencionadas.

---

<sup>2</sup>Diz-se cintilante o material que emite luz quando uma partícula cruza seu volume. A luz emitida é proporcional à energia da partícula que passa.

## A Seção e.m.

A seção e.m. do calorímetro do ATLAS é dividida em 3 camadas, das quais a segunda é a mais profunda. Cada camada possui uma segmentação<sup>3</sup> específica que otimiza a relação custo-benefício do detetor. As camadas mais internas são mais segmentadas, permitindo a localização precisa das partículas no plano  $\eta \times \phi$ , enquanto as mais externas são concebidas de forma menos segmentada e mais profunda, com o objetivo de absorver toda a energia da partícula incidente, limitando os custos de produção.

A seção e.m. é dividida em três partes: o barril (do inglês *barrel*) e duas tampas (*endcap*). Estas partes fecham quase que hermeticamente o espaço ao redor da colisão até um valor de  $|\eta| = 3,2$  (para maiores referências sobre o sistema de coordenadas do ATLAS, leia o Apêndice A). A Figura 3.6 mostra o posicionamento do calorímetro eletromagnético no detetor ATLAS. Sua segmentação não é mostrada nesta figura. Em vez disso, mostram-se os valores de  $\eta$  que definem os limites geométricos da seção e.m.. Pode-se perceber que a porção do barril de tal calorímetro estende-se de  $\eta = 0$  até  $|\eta| = 1,475$ . Em  $|\eta| = 1,375$ , o barril começa a sobrepor a tampa, que é dividida entre tampa exterior (até  $|\eta| = 2,5$ ) e interior (de  $|\eta| = 2,5$  até  $|\eta| = 3,2$ ).

O calorímetro e.m. do ATLAS também inclui um pré-irradiador (do inglês, *pre-sampler*), que funciona praticamente como um calorímetro muito fino, posicionado antes dos calorímetros de argônio líquido, com a função de recuperar a informação perdida no material *morto* da seção e.m. (ou seja, fios, encapamentos, etc.). O pré-irradiador pode ser observado, na figura, de  $\eta = 0$  até  $|\eta| = 1,5$  no barril, e depois, de  $|\eta| = 1,5$  até  $|\eta| = 1,8$  nas tampas.

Uma observação mais apurada da Figura 3.6 revela um *buraco* entre o barril e a tampa da seção e.m.. Esta falha existe para que seja possível passarem-se os cabos acoplados aos sensores do ID. Para que a perda nessa região seja minimizada, decidiu-se por colocar um cintilador (indicado na figura). Cintiladores são detetores que se excitam pela passagem das partículas energéticas e produzem luz. Cintiladores são normalmente bastante compactos e finos.

A Figura 3.7 exemplifica a diversificação da segmentação ao longo de  $\phi$ . Cada

---

<sup>3</sup>A "segmentação" de um detetor é a resolução deste no plano  $\eta \times \phi$ .

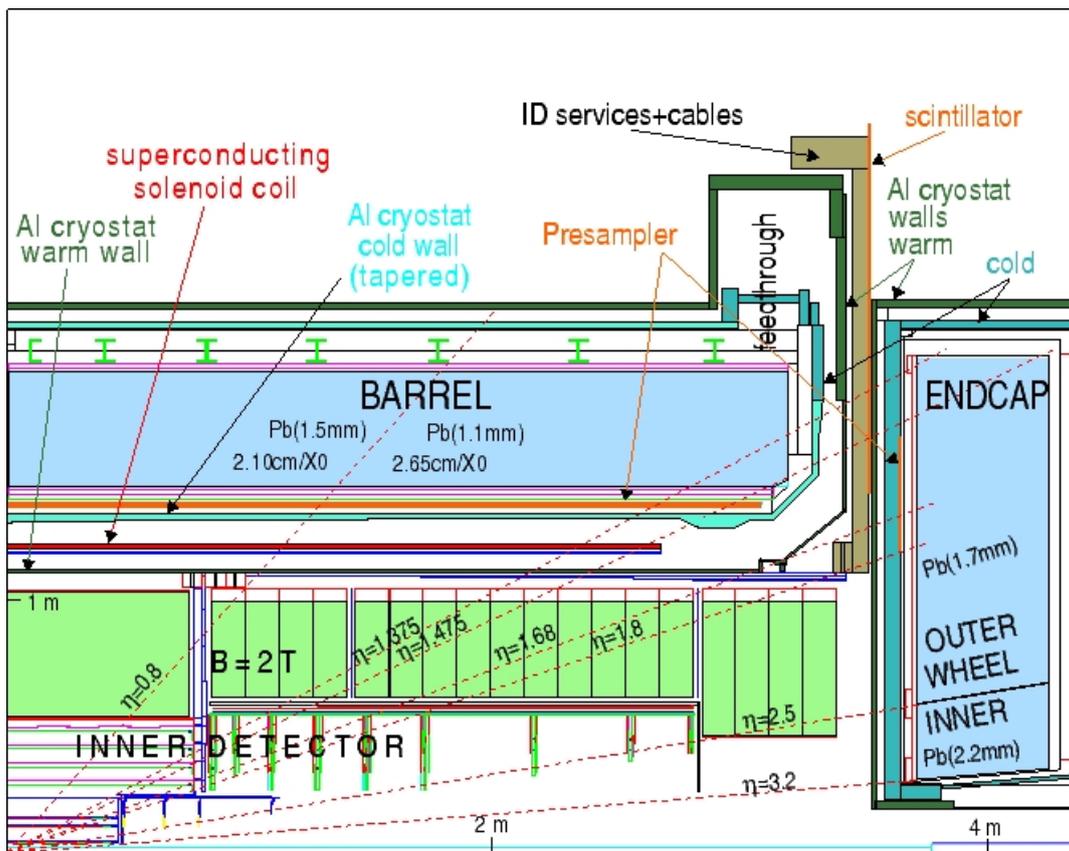


Figura 3.6: O Calorímetro e.m. do ATLAS em um corte transversal.

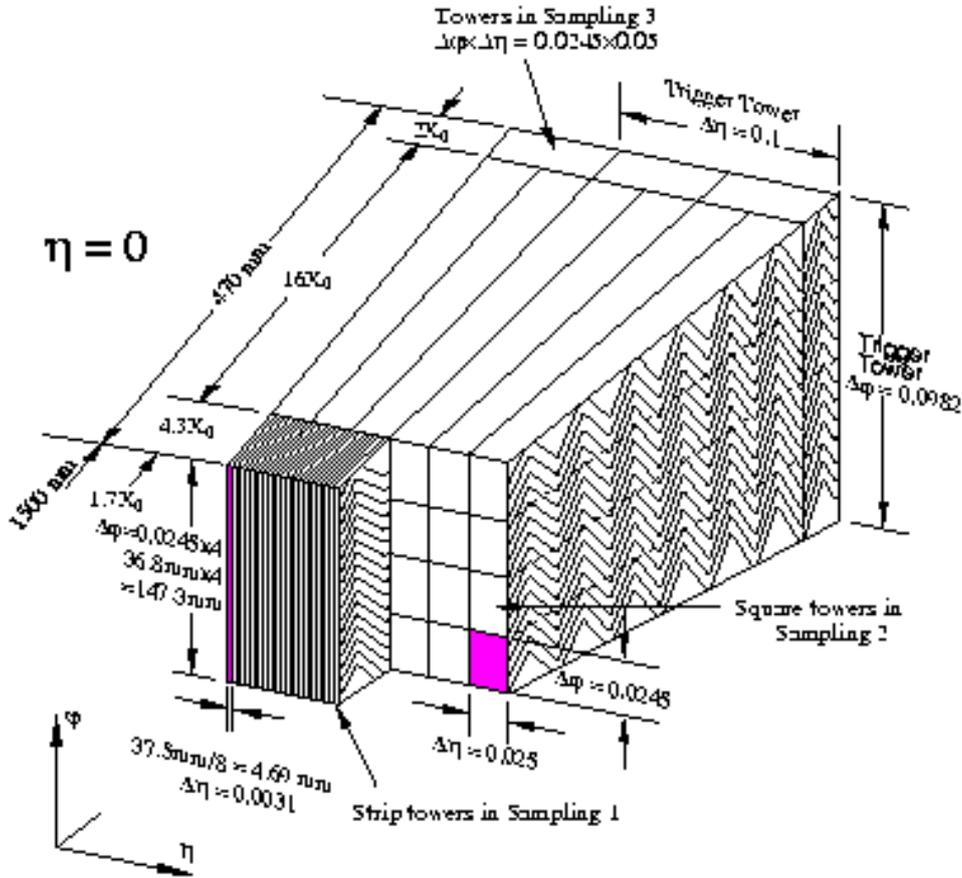


Figura 3.7: Diagrama de um dos gomos do barril da seção e.m. do ATLAS.

camada é formada por células de diferentes tamanhos. Nesta figura, também se verifica que a segunda camada é a que possui células mais profundas. É plausível esperar que mais energia seja amostrada nesta camada.

A Tabela 3.1 resume as informações de segmentação para a seção e.m., incluindo informações sobre o *pre-sampler* e as tampas. Nota-se, a partir da tabela, que a segmentação de algumas camadas varia bastante com  $\eta$ . A coluna da extrema direita indica o número de células numa área de  $0,1 \times 0,1$  no plano  $\eta \times \phi$ . Esta área é uma referência para os níveis de filtragem, como será visto mais adiante.

### A Seção Hadrônica

Os calorímetros hadrônicos do ATLAS são formados pelo Calorímetro de Telhas ou *TileCal* e pela Tampa Hadrônica baseada em Argônio Líquido - a mesma técnica

Tabela 3.1: A segmentação, camada a camada, dos calorímetros e.m. do ATLAS.

Camada	Peça	$\eta_{\text{início}}$	$\eta_{\text{fim}}$	$\Delta\eta \times \Delta\phi$	$N_{\eta} \times N_{\phi}$
<i>Pre-sampler</i>	Barril	0	1,5	$0,025 \times 0,1$	$4 \times 1$
	Tampa	1,5	1,8	$0,025 \times 0,1$	$4 \times 1$
Camada 1	Barril	0	1,4	$0,003125 \times 1$	$32 \times 1$
		1,4	1,475	$0,025 \times 0,025$	$4 \times 4$
	Tampa	1,375	1,5	$0,025 \times 0,1$	$1 \times 4$
		1,5	1,8	$0,003125 \times 0,1$	$32 \times 1$
		1,8	2,0	$0,004167 \times 0,1$	$24 \times 1$
		2,0	2,4	$0,00625 \times 0,1$	$16 \times 1$
		2,4	2,5	$0,025 \times 0,1$	$4 \times 1$
		2,5	3,2	$0,1 \times 0,1$	$1 \times 1$
Camada 2	Barril	0	1,4	$0,025 \times 0,025$	$4 \times 4$
		1,4	1,475	$0,075 \times 0,025$	$1 \times 4$
	Tampa	1,375	2,5	$0,025 \times 0,025$	$4 \times 4$
		2,5	3,2	$0,1 \times 0,1$	$1 \times 1$
Camada 3	Barril	0	1,35	$0,05 \times 0,025$	$2 \times 4$
	Tampa	1,5	2,5	$0,05 \times 0,025$	$2 \times 4$



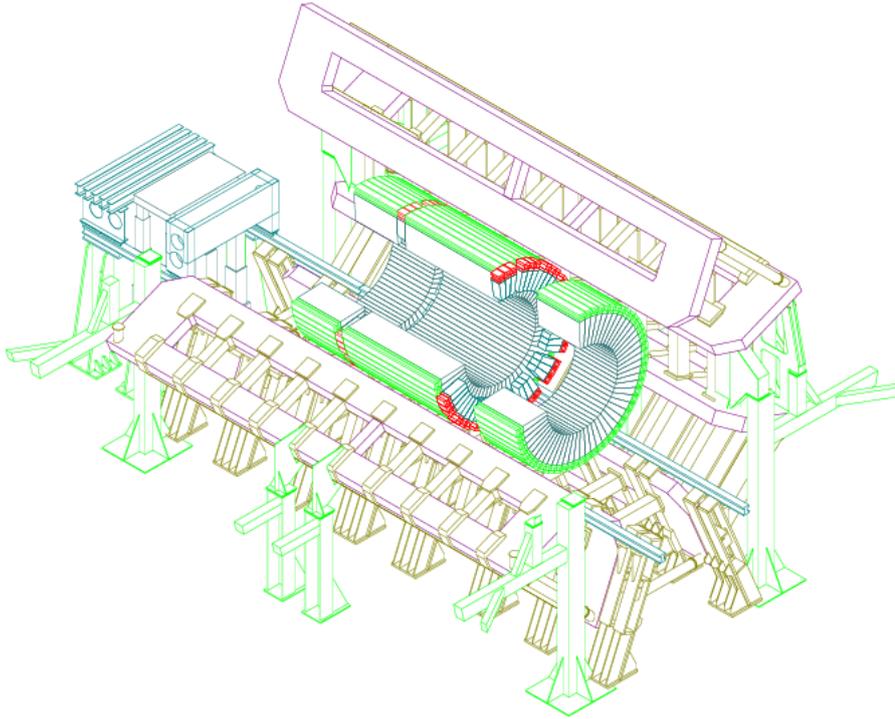


Figura 3.8: O calorímetro de telhas do ATLAS (ao centro), em sua posição final, envolvido pelo detetor de múons. Extraído de [18].

usada para a seção e.m.. O TileCal é um calorímetro de amostragem cujo material absorvedor é uma liga com aço e os elementos amostradores são telhas cintilantes [26]. As telhas são posicionadas em planos perpendiculares aos feixes de partículas colididas, e conectadas a fibras ópticas em duas de suas extremidades. Estas fibras coletam o sinal luminoso, gerado pela telha ao interagir com a partícula, e transportam-no até tubos foto-multiplicadores (PM), onde o sinal é convertido em sinal elétrico. Somadores rápidos [27] se encarregam de adicionar os sinais das telhas entre si formando células de detecção, de forma equivalente à seção e.m.. O TileCal é posicionado após a seção e.m., como é possível verificar na Figura 3.8.

Uma peculiaridade dos calorímetros hadrônicos do ATLAS é que o barril e a tampa são feitos de formas diferentes, ao contrário da seção e.m.. O calorímetro de telhas (TileCal) abrange as porções do barril ( $0 < |\eta| < 1,0$ ) e sua extensão ( $0,8 < |\eta| < 1,7$ ). A tampa desta seção é feita como os calorímetros e.m., no formato de acordeões, usando Argônio líquido.

A Tabela 3.2 resume a informação de segmentação da seção hadrônica dos calo-

Tabela 3.2: A segmentação, camada a camada, dos calorímetros hadrônicos do ATLAS.

Camada	Peça	$\eta_{\text{inicio}}$	$\eta_{\text{fim}}$	$\Delta\eta \times \Delta\phi$	$N_{\eta} \times N_{\phi}$
Camadas 1 e 2	Barril (TileCal)	0	1,0	$0,1 \times 0,1$	$2 \times 2$
	Barril Ext. (TileCal)	0,8	1,7	$0,1 \times 0,1$	$2 \times 2$
	Tampa	1,5	2,5	$0,1 \times 0,1$	$2 \times 2$
	(LAr)	2,5	3,2	$0,2 \times 0,2$	$1 \times 1$
Camada 3	Barril (TileCal)	0	1,0	$0,2 \times 0,1$	$1 \times 2$
	Barril Ext. (TileCal)	0,8	1,7	$0,2 \times 0,1$	$1 \times 2$
	Tampa	1,5	2,5	$0,1 \times 0,1$	$2 \times 2$
	(LAr)	2,5	3,2	$0,2 \times 0,2$	$1 \times 1$

rímetros do ATLAS, de forma equivalente a da Tabela 3.1. Nessa tabela é possível perceber que o tamanho das células, em média, é bem maior que o valor equivalente no calorímetro eletromagnético. A segmentação é também mais uniforme que na seção e.m. dos calorímetros do ATLAS. Isto se deve ao fato dos chuveiros hadrônicos serem mais largos e profundos, provocando maiores flutuações nas medidas de energia e, portanto, não necessitando de uma segmentação tão fina.

Outra diferença é na área de referência. Na seção e.m., considera-se  $0,1 \times 0,1$  - aqui a área de referência é de  $0,2 \times 0,2$ , já que se encontram células maiores que a área de referência no calorímetro e.m..

A Figura 3.9 mostra uma seção transversal da parte do Barril do Calorímetro de Telhas. Nesta figura, verifica-se que o sistema de leitura agrupa as células deste detetor em 3 camadas. A segmentação no sentido de  $\eta$  é mantida constante ainda assim.

### 3.4.3 O Detetor de Múons

Múons com grande momento são assinaturas bastante promissoras e robustas da física de interesse no LHC. Para explorar este potencial, foi projetado um espectrômetro de múons (veja a Figura 3.10) com sistemas de filtragem e medição de

<b>D-1</b>	<b>D-2</b>			<b>D-3</b>			<b>D-4</b>		
X(39),X(40)	X(41),X(42)			X(43),X(44)			X(45),X(46)		
<b>BC-1</b>	<b>BC-2</b>	<b>BC-3</b>	<b>BC-4</b>	<b>BC-5</b>	<b>BC-6</b>	<b>BC-7</b>	<b>BC-8</b>		
X(21),X(22)	X(23),X(24)	X(25),X(26)	X(27),X(28)	X(29),X(30)	X(31),X(32)	X(33),X(34)	X(35),X(36)		
									<b>B-9</b>
									X(37),X(38)
<b>A-1</b>	<b>A-2</b>	<b>A-3</b>	<b>A-4</b>	<b>A-5</b>	<b>A-6</b>	<b>A-7</b>	<b>A-8</b>	<b>A-9</b>	<b>A-10</b>
X(1),X(2)	X(3),X(4)	X(5),X(6)	X(7),X(8)	X(9),X(10)	X(11),X(12)	X(13),X(14)	X(15),X(16)	X(17),X(18)	X(19),X(20)

Figura 3.9: Configuração de leitura das células da seção do barril do Calorímetro de Telhas.

momento com alta resolução independentes do restante do sistema, sensível a este tipo de partículas [28]. A detecção é baseada na deflexão (magnética) de múons provida por um grande toróide com núcleo a ar e detetores de traço bastante precisos. Para  $|\eta| \leq 1,0$ , um grande magneto em formato de barril será construído com oito solenóides circundando a seção hadrônica dos calorímetros. Para  $1,4 \leq |\eta| \leq 2,7$ , a deflexão será garantida por magnetos na forma de tampas ao redor do barril. Esta configuração provê um campo magnético praticamente transversal à trajetória de eventuais múons.

Na região do barril, as trajetórias serão detetadas por câmeras organizadas em três camadas cilíndricas (estações) ao redor do eixo de colisão. Na região da tampa, as câmaras serão montadas verticalmente, também em número de três. As coordenadas dos traços é medida precisamente por Câmaras de Traços Monitoradas (do inglês, *Monitored Drift Chambers*, MDT) na direção principal de deflexão do campo magnético, na maior parte do campo da pseudo-rapidez. Para grandes valores de pseudo-rapidez, Câmaras de Pastilhas Catódicas (do inglês, *Cathode Strip Chambers*, CSC) com alta segmentação serão construídas para garantir a precisão da detecção nesta área bastante ruidosa. Tal como os calorímetros, o Espectrômetro de Múons está diretamente acoplado ao primeiro nível do sistema de filtragem de

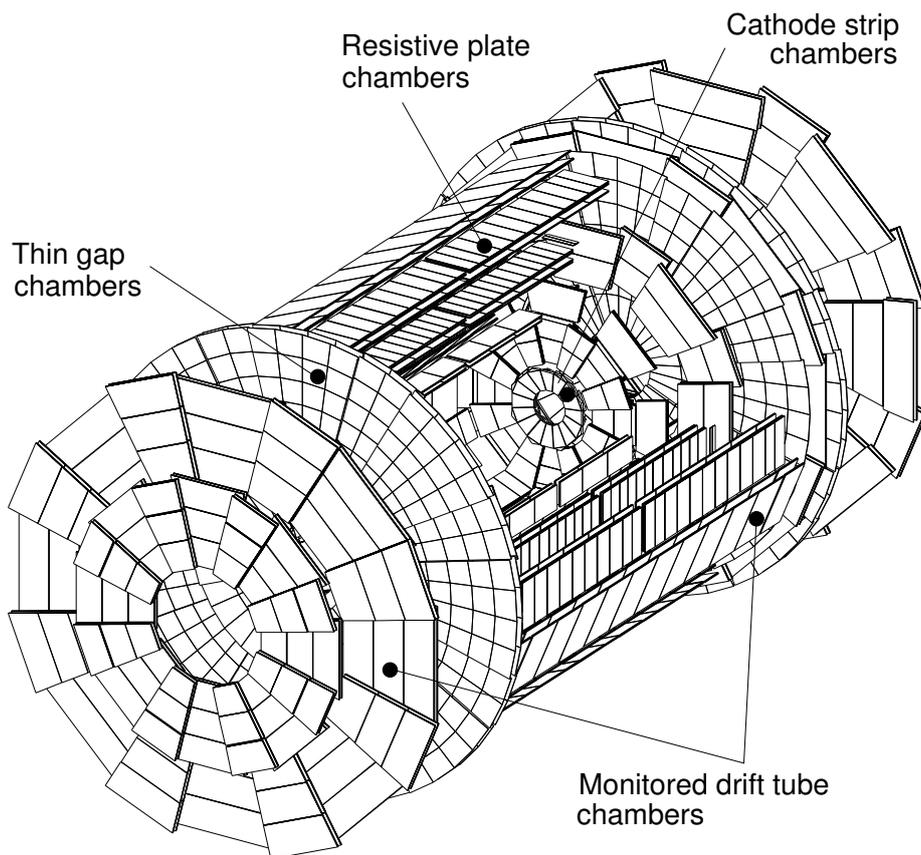


Figura 3.10: Visão tridimensional dos detetores de múons do ATLAS.

eventos do ATLAS (veja o Capítulo 4).

# Capítulo 4

## O Sistema de filtragem e aquisição de dados do ATLAS

O sistema de filtragem do experimento ATLAS tem o encargo de separar *online* a Física considerada como ordinária segundo o experimento, dos raros eventos de interesse. A taxa de eventos produzida pelo LHC será de 40 milhões por segundo, ou 40 MHz, em bateladas que durarão cerca de 10 horas (este período é comumente referido como *run*). Cada evento poderá produzir uma colisão favorável à Física de interesse, embora a grande maioria, mais de 99,9999%, represente canais físicos já bastante estudados em experimentos anteriores, ainda que a luminosidade do sistema seja tão elevada. Ademais, cada evento detetado pelo ATLAS produzirá cerca de 1,5 Megabyte em dados, o que torna o problema da filtragem bastante difícil, já que o fluxo de dados que deve ser analisado por base de tempo está acima da capacidade de qualquer tecnologia de transporte de dados atual.

Para o experimento ATLAS, o sistema de filtragem será construído em 3 níveis sequenciais (veja a Figura 4.1). Cada nível que se sucede refina a decisão do nível anterior através da aquisição de mais dados do detetor. O Primeiro Nível de filtragem (do inglês *First-Level Trigger*, LVL1), deverá reduzir a taxa de eventos do LHC para cerca de 100 kHz apenas, através de técnicas simples de filtragem codificadas em *hardware* dedicado. O Segundo Nível de Filtragem (do inglês *Second Level Trigger*, LVL2), reduzirá a taxa de saída do LVL1 para aproximadamente 1 kHz utilizando aplicações codificadas em *software* para computadores do tipo PC, tão comuns quanto os disponíveis para aplicações domésticas. O Filtro de Eventos (do

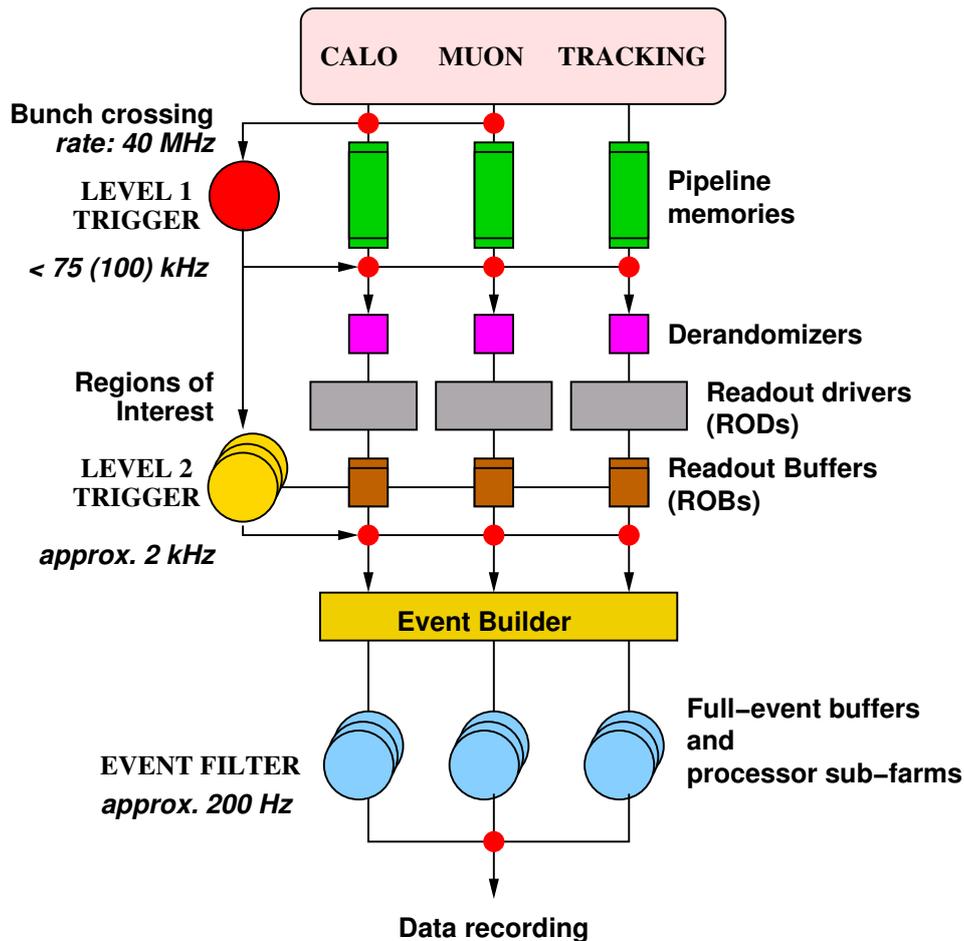


Figura 4.1: Visão funcional simplificada do sistema de filtragem do experimento ATLAS. Extraído de [16].

inglês *Event Filter*, EF), estará baseado na mesma tecnologia escolhida para o LVL2, reduzindo a taxa de entrada de eventos para aproximadamente 100 Hz, através de algoritmos de filtragem mais eficientes, porém mais lentos. Os eventos ultimamente selecionados pelo EF são armazenados em mídia permanente, para posterior análise *offline*.

## 4.1 O primeiro nível de filtragem

O LVL1 realiza uma seleção inicial de candidatos à Física de interesse baseada em informações com segmentação reduzida, fornecidas por um subconjunto dos detectores do ATLAS, os calorímetros e os detectores de múons [29]. Múons com altos valores de momento transversal ( $p_T$ ) são identificados a partir das câmaras de trigger conectadas ao Espectrômetro de Múons (veja a Seção 3.4.3) na região do barril

e da tampa. As seleções baseadas em calorimetria utilizam uma segmentação reduzida das seções e.m. e hadrônica no barril e na tampa. Os objetos procurados usando-se calorímetros no LVL1 são elétrons com altos valores de  $p_T$  e fótons, jatos, taus decaindo em cascatas hadrônicas, grandes quantidades de energia faltante ou energia transversa total. No caso da seleção de elétrons e fótons ou hádrons e taus, isolamento em energia<sup>1</sup> pode ser requisitado.

Para gerar os sinais de filtragem necessários ao LVL1, o sistema de leitura de dados dos calorímetros é equipado com somadores rápidos [27, 25] que aglomeram o sinal de 2 ou mais células destes detetores para gerar macro-células de filtragem denominadas torres de filtragem (do inglês *Trigger Towers*, TT). O sistema de múons também possui circuitos equivalentes que determinam a ocorrência de traços com altos valores de momento transverso ( $p_T$ ) naquele subdetetor.

Os eventos selecionados pelo LVL1 são lidos a partir da eletrônica específica de cada detetor por meio de *drivers* denominados *ReadOut Drivers* (ROD's). Cerca de 1.600 ROD's serão necessários para ler todos os dados do detetor, ou seja, aproximadamente  $10^7$  canais independentes. Vários canais de leitura são multiplexados em cada ROD. Bancos de memória primários, chamados de *derandomizers*, garantem um fluxo contínuo dos dados aos ROD's, apesar da irregularidade da aceitação de eventos proporcionada pelo LVL1.

Os dados completos de eventos selecionados são guardados momentaneamente em bancos de memória denominados *ReadOut Buffers* (ROB's) até que o evento seja rejeitado pelo LVL2 ou, no caso de ser aceito por este nível, até que os dados sejam transferidos para o terceiro nível de filtragem (EF). O processo de mover os dados dos ROB's para o EF é chamado de construção do evento (do inglês, *Event Building*, EB). Porém, até a fase de construção do evento, os dados estarão segmentados em ROB's. Após esta etapa, o evento completo estará guardado em um único banco de memória e acessível a um processador do EF.

**Regiões de Interesse** Além do disparo dado pelo LVL1 à eletrônica de leitura, este nível de filtragem deve enviar ao LVL2 um mapa dos objetos de interesse encontrados no detetor durante a avaliação do evento. Este mapa indica, na precisão do

---

<sup>1</sup>Isolamento em energia significa a ausência de outros picos energéticos adjacentes ao pico energético sendo analisado.



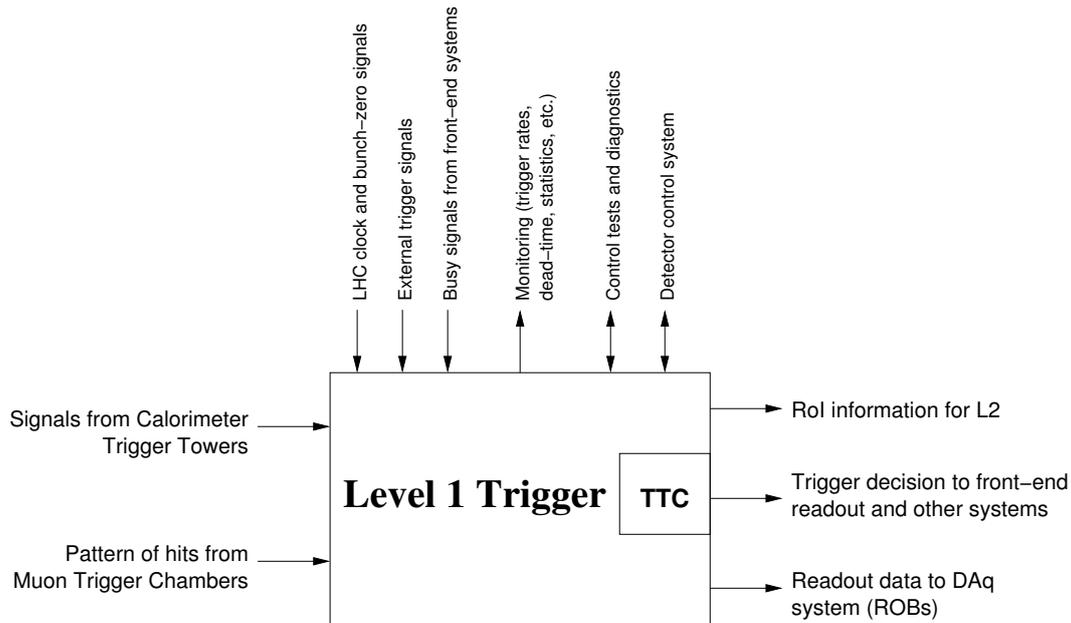


Figura 4.2: Diagrama simplificado dos sinais de entrada e saída do LVL1. Extraído de [16].

LVL1, as regiões de interesse no detetor (do inglês *Region of Interest*, RoI) que contribuíram para a seleção, os tipos de objetos encontrados e o valor mínimo (*threshold*) de  $p_T$  que satisfazem. Informações de RoI's secundárias (que não contribuíram para a decisão do LVL1, mas que excitaram notavelmente o detetor) também podem ser transmitidas ao LVL2, provendo maior flexibilidade de decisão a este nível.

Na Figura 4.2, é possível observar um diagrama do tipo “caixa-preta”, indicando os sinais de entrada e saída do LVL1. A este sistema são injetados os sinais provenientes da eletrônica de leitura e dos sistemas de filtragem diretamente acoplados aos detetores e do LHC (*clock* de 40 MHz). Deste sistema serão colhidos sinais que serão, por sua vez, encaminhados para o LVL2, para a eletrônica de aquisição e para o sistema de monitoração central.

Na Figura 4.3 observa-se um diagrama funcional simplificado do LVL1. Nesta figura é possível identificar processadores específicos para os sinais provenientes dos detetores. Um processador central de filtragem (marcado na figura como *Central Trigger Processor*) é responsável pela decisão final e pela emissão do sinal para o LVL2 e para o sistema de distribuição de tempo, disparo e controle (indicado como *Timing, Trigger and Control distribution*, ou TTC), que gera, finalmente, o sinal para a leitura dos dados do detetor. Dentre os componentes do TTC está

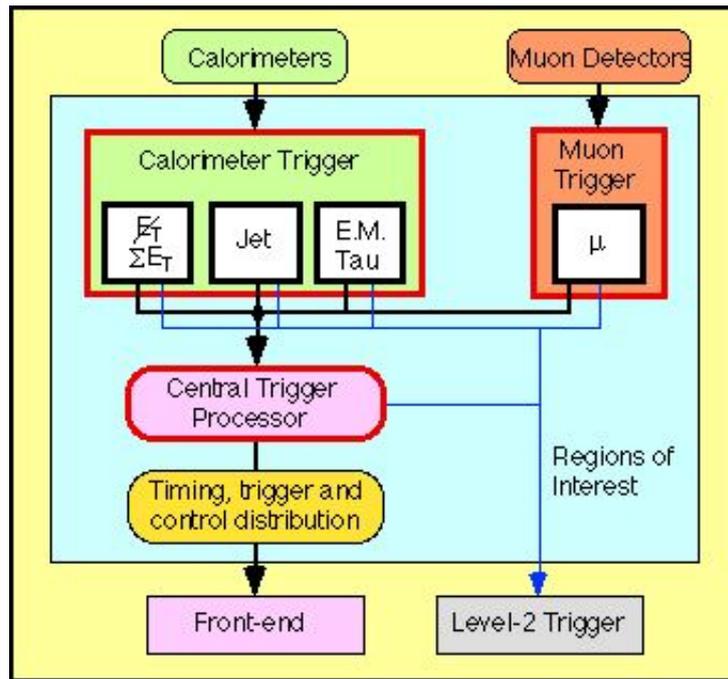


Figura 4.3: Diagrama em blocos indicando as principais funções do LVL1. Extraído de [16].

o Construtor de RoI's (do inglês *RoI Builder*, RoIB) que monta o mapa de RoI's destacadas pelo LVL1 e comunica, através de uma conexão serial rápida tipo S-Link, estes dados ao LVL2.

## 4.2 Os altos níveis de filtragem, aquisição de dados e controle

O sistema de aquisição de dados do ATLAS (do inglês *Data Acquisition System*, DAQ) é composto por processadores que estão encarregados da transmissão e manipulação dos dados produzidos pelo detetor [16]. O DAQ trabalha em cooperação com os diversos níveis de filtragem e controle do experimento, de forma a garantir a correção dos dados transmitidos e a minimização do tempo morto despendido no tráfego de informação pelo sistema. A cadeia de operação do DAQ é iniciada por um disparo do LVL1 para a leitura de dados executada nos ROD's. Cada ROD é ligado, através de uma conexão S-Link<sup>2</sup>, a somente um ROB. Cada ROB pode

<sup>2</sup>Conexão serial ponto-a-ponto com taxa máxima de transferência de 160 megabits por segundo.

receber dados de um ou no máximo 3 ROD's, dependendo da configuração final do experimento.

Cabe aos Altos Níveis de Filtragem (do inglês *High-Level Triggers*, HLT), analisando de forma ótima os dados presentes nos ROB's, decidir, disparando mais uma vez o DAQ, sobre o futuro de cada evento aceito pelo LVL1. O conjunto HLT-DAQ-Control deve funcionar harmoniosamente sobre uma mesma base computacional, utilizando ainda o mesmo conjunto de ferramentas para a codificação, análise e depuração de eventuais problemas que possam ocorrer. Por esta razão, esta união de produtos de *hardware* e *software* é considerada um macro-sistema do experimento. Este sistema pode ser subdividido em 4 partes:

- **Sistema de controle do detetor ou *Detector Control System*, DCS:** responsável pela operação coerente e segura (no que tange a radiação, controle de altas tensões, etc.) do detetor ATLAS e pelo interfaceamento com sistemas externos e serviços incluindo o acelerador LHC. Uma vez que este tópico não está diretamente envolvido neste trabalho, será abordado somente em caráter introdutório;
- ***Online*:** é responsável por todos os aspectos de operação e controle durante a aquisição de dados do experimento e durante operações (*runs*) de teste e calibração;
- **Fluxo de dados ou *Dataflow*:** este sistema é responsável por receber os dados do detetor, servir um subconjunto destes dados ao HLT e transportar os dados de eventos selecionados em última instância para armazenagem em mídia permanente;
- **Altos Níveis de Filtragem ou HLT:** são responsáveis pela seleção de eventos após o LVL1, envolvendo a redução da taxa de eventos e a classificação de todos os eventos aceitos.

#### 4.2.1 O sistema de controle do detetor

O DCS supervisiona todos os componentes de *hardware* do arranjo experimental, incluindo todos os sistemas de detecção do ATLAS. O DCS também se comunica com

sistemas externos, como a infraestrutura de serviços do CERN e, mais notavelmente, com o acelerador LHC.

Aspectos de segurança são tratados pelo DCS somente no nível de menor severidade. Este sistema cuida principalmente de questões ligadas ao seqüenciamento de operações ou à requisição onde prevaleçam condições específicas, antes de permitir que outros procedimentos sejam executados. Ferramentas para a construção de interdependências, tanto em *hardware* quanto em *software*, são providas pelo DCS. Monitoração e prevenção de situações que poderiam causar danos maiores ao detetor e à vida das pessoas são de responsabilidade de sistemas dedicados - o Sistema de Segurança do Detetor (do inglês *Detetor Safety System*, DSS) e do sistema de segurança e alarme do CERN respectivamente. O DCS interage com ambos os sistemas.

Todas as ações iniciadas pelo operador e todos os erros, avisos e alarmes que impliquem em falha de *hardware* do detetor são gerenciadas pelo DCS. Este sistema provê informação *online* da situação presente, com a granularidade necessária para uma operação centralizada do sistema. A interação dos *experts* em detetores com suas máquinas também é feita através do DCS, que continuamente monitora todos os parâmetros operacionais, guiando o operador e sinalizando qualquer comportamento anormal do maquinário. O DCS também será capaz de acionar procedimentos automáticos para trazer o sistema de detecção para um estado seguro, se necessário for.

No que tange à operação do experimento, a interação com o sistema de aquisição de dados é de grande importância. A boa qualidade da Física gravada em mídia permanente depende de uma sincronização absoluta entre o DAQ e o DCS; ambos os sistemas são complementares. O DAQ lida com eventos caracterizados por números e o DCS guarda o estado operacional do detetor enquanto os dados estão sendo adquiridos, correlacionando os números dos eventos com a qualidade da aquisição, que, finalmente, é avaliada *offline*.

Algumas partes do detetor operarão continuamente, pois qualquer interrupção pode ser custosa em tempo, fundos ou desempenho do sistema global. Portanto, a supervisão do sistema de detecção é necessária constantemente. Por outro lado, o DAQ é somente necessário durante a aquisição de dados ou durante *runs* de monito-

ração, calibração ou teste. Assim sendo, o DCS precisa ter completa independência funcional, ainda que, ao mesmo tempo, este quesito não deva, imperativamente, acarretar em limites operacionais com o DAQ.

#### 4.2.2 O sistema *online*

O sistema *online* engloba o conjunto de aplicações para configurar, controlar e monitorar o sistema de filtragem e aquisição de dados do ATLAS (do inglês *Trigger and Data Acquisition*, TDAQ), excluindo o gerenciamento, processamento e transporte de dados físicos. É um conjunto personalizável de ferramentas que provê a “cola” que une entradas e saídas dos outros sistemas. O sistema *online* não contém elementos que sejam específicos aos detetores, pois é também utilizado por todos os sistemas de instrumentação da aquisição de dados de forma transparente. Ele coopera com outros sub-sistemas e interfaces para a leitura do detetor, DCS, LVL1, *Dataflow*, HLT, *Offline* e o interfaceamento com operadores dos *runs*.

Um dos papéis importantes do sistema *online* é prover serviços para guiar o TDAQ na sua inicialização e finalização, de forma que os aplicativos sejam executados de maneira ordenada. Este sistema é responsável pela sincronização dos estados de um *run* para todo o sistema de filtragem, aquisição e supervisão. Os procedimentos de ligamento e desligamento são projetados para se alcançar o menor tempo possível, reduzindo tempos-mortos, pois isto afeta diretamente a quantidade de dados que podem ser adquiridos em um período de funcionamento do LHC. Ferramentas de verificação e diagnóstico ajudam a detetar problemas de maneira precoce e eficiente. Bancos de dados com a configuração do sistema são fornecidos contendo um grande número de parâmetros que descrevem a topologia do sistema, os componentes de *hardware* e *software*, assim como os modos e ordem de execução. Durante a aquisição, componentes especializados capturam e servem dados relacionados à monitoração dos diversos processos do experimento, tais como histogramas e parâmetros de rejeição, assim como mensagens de erro e diagnóstico enviadas pelas diversas aplicações. Interfaces (gráficas) aos operadores permitem configurar e controlar diversas funcionalidades do experimento. Estas interfaces fornecem visões simplificadas de vários sub-sistemas durante um *run*.

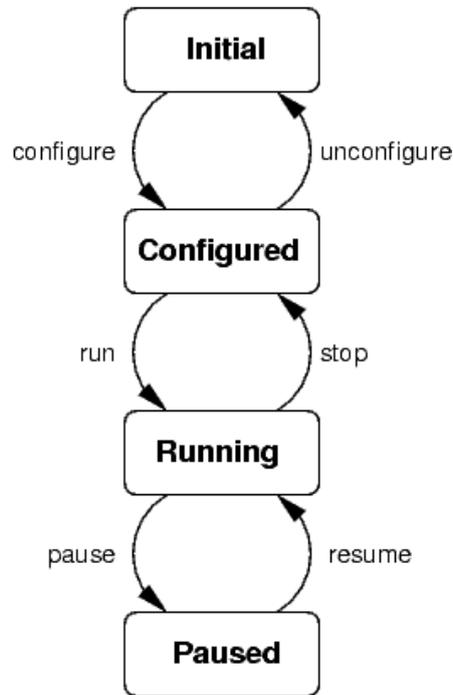


Figura 4.4: A máquina de estados finitos utilizada por todos os componentes do TDAQ e detetores do ATLAS.

### Arquitetura do sistema *online*

O projeto do sistema *online* é baseado numa arquitetura de componentes, que respeitam o modelo cliente-servidor. Os componentes podem ser divididos em 3 grupos, que incluem um conjunto de pacotes cada um. Cada um dos pacotes está associado a um grupo de funções para este sistema, e provê um conjunto definido de serviços. Os serviços possuem interfaces bem definidas e são desacoplados entre si.

**Controle:** contém pacotes para o controle do sistema de aquisição, filtragem e detecção. Os pacotes de controle existem para propiciar ao TDAQ capacidades de inicialização e finalização, distribuição de comandos, sincronização, manuseio de erros e verificação do sistema.

O sistema de controle de dispositivos e aplicações do sistema *online* é baseado em uma máquina de estados finitos global mostrada na figura 4.4. Todos os dispositivos do sistema devem obedecê-la, enviando avisos e erros caso falhem em cumprir as ordens do operador.

Este pacote também provê uma interface gráfica, escrita em Java, para o aciona-

mento e verificação de todas as partes do sistema de filtragem e aquisição de dados, indicando prontamente problemas, ou encaminhando avisos ao operador do *run*.

**Bancos de dados:** contêm pacotes para a configuração do TDAQ e detetores. Os pacotes de configuração existem para propiciar parâmetros de configuração do sistema, sua descrição e acesso e registro operacional de informações durante a aquisição de dados. Há também componentes que provêm acesso de leitura e escrita a bancos de dados que descrevem as condições do feixe provido pelo LHC e do detetor.

Como no caso do pacote de controle, existem ferramentas para a manipulação e verificação dos bancos de dados produzidos pelo usuário, que podem facilmente chegar a dezenas de megabytes em sua versão final, uma vez que englobará a configuração de milhares de componentes. Atualmente, somente uma das interfaces de operação a banco de dados está disponível e é baseada na utilização de arquivos locais, uma vez que o código para o acesso remoto de bancos de dados não está implementado ainda.

**Compartilhamento de informação:** contém pacotes para o compartilhamento de informação no TDAQ. Os componentes destes pacotes podem anunciar erros, publicar estados, estatística e histogramas construídos pelos sub-sistemas do TDAQ e pelos detetores. Há um conjunto de interfaces gráficas para a monitoração do Sistema de Filtragem do ATLAS, que podem ser acessadas através da linha de comando em ambientes apropriadamente configurados.

### 4.2.3 O sistema de fluxo de dados

O Sistema de Fluxo de Dados (do inglês *Dataflow*, DF) do DAQ é encarregado da transmissão correta e coerente dos dados do detetor, garantindo a menor quantidade de tempo morto possível, dentro das expectativas do experimento [30, 31, 32]. A Figura 4.5 define os componentes deste sub-sistema. Os sinais de entrada, como podem ser observados nesta figura, são providos pelo LVL1 e ROD's. Nesta figura, as caixas representam processadores, enquanto as linhas conectando-os representam conexões de rede (neste caso, *gigabit* ou *10-gigabit* ethernet). Próximo aos nós de processamento, há uma estimativa da quantidade destes tipos de processadores que

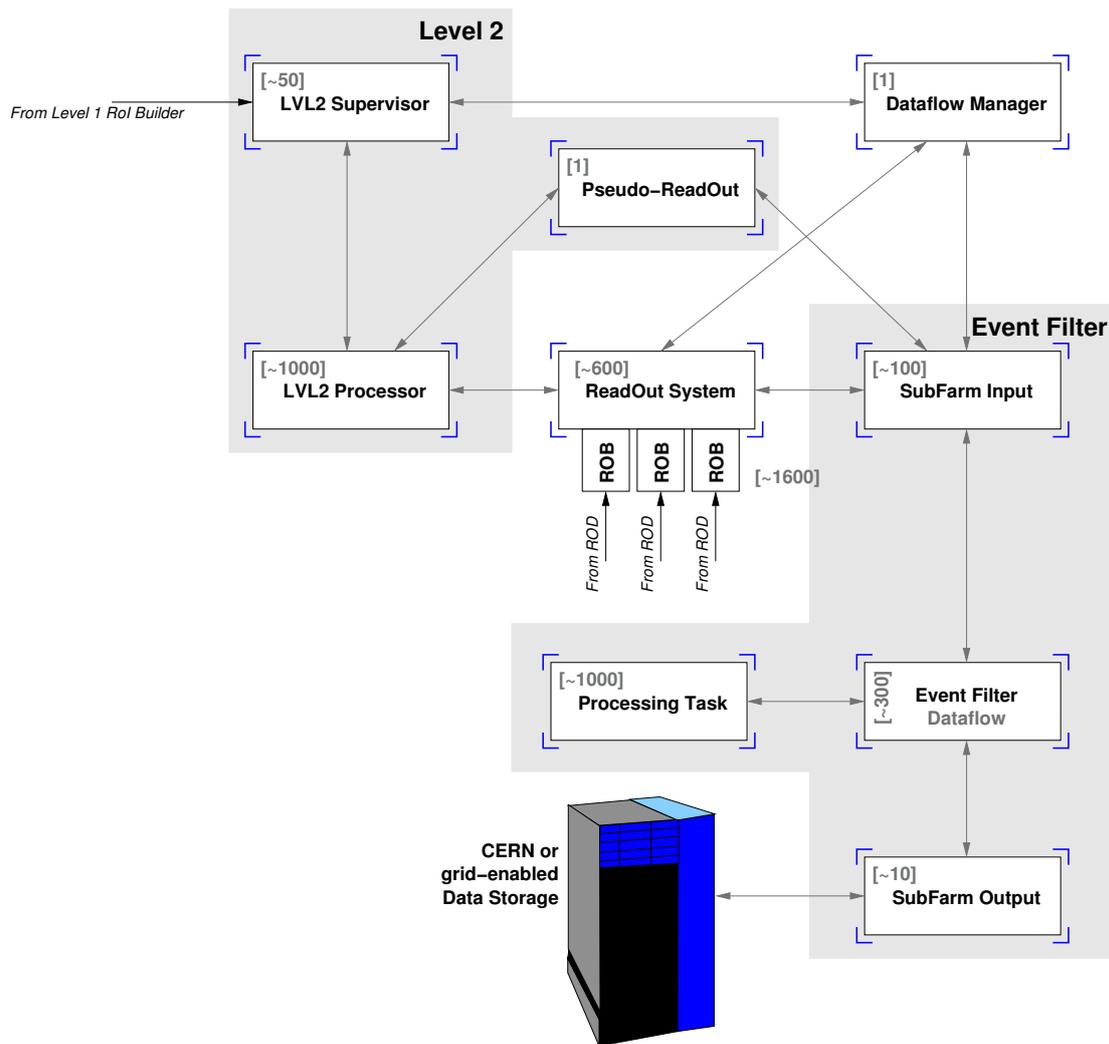


Figura 4.5: Os componentes do DF e suas interconexões.

serão necessárias para manter a taxa de operação de 100 kHz [16].

A seqüência de funcionamento se dá da seguinte forma: o Supervisor do LVL2 (L2SV) recebe uma mensagem do RoIB do LVL1 indicando a ocorrência de um evento interessante e aprovado por este nível. Somente há um RoIB no sistema, que envia para o primeiro Supervisor do LVL2 disponível a mensagem sobre o evento. Cada Supervisor mantém um conjunto de Unidades de Processamento do LVL2 (do inglês *LVL2 Processing Unit*, L2PU) em sua lista de nós supervisionados. Logicamente atrelado a cada L2PU há uma fila de eventos que foram recentemente alocados para processamento. O L2SV localiza a L2PU com menos eventos atribuídos e envia a mensagem do LVL1 para este processador.

A L2PU decodifica a mensagem enviada pelo L2SV e analisa as regiões do detetor de que precisará para processar o evento. O tamanho e a quantidade de dados de



que a L2PU necessitará para processar o evento dependem dos algoritmos que são lá executados e, portanto, este processador tem livre arbítrio na quantidade e qualidade dos dados que irá requisitar. Para coletar os dados, a L2PU computa o número e a localização dos ROB's que são necessários, e encaminha um pedido de dados prontamente ao Sistema de Leitura do Detetor (do inglês, *Read Out System*, ROS). Os ROS's afetados respondem com os dados dos ROB's relevantes e, desta forma a L2PU poderá classificar o evento. Após algum intervalo de tempo necessário à análise da Física do evento, a L2PU responde ao L2SV com um valor binário, indicando rejeição ou aceitação do evento. Se o evento for aceito, a L2PU também envia um diagnóstico do que foi encontrado no evento para dentro de um pseudo-ROS (pROS), que permitirá a anexação desta informação ao evento final.

A segunda parte do processamento do evento nos altos níveis de filtragem é iniciada por uma mensagem do L2SV ao Gerente do Fluxo de Dados (do inglês, *Dataflow Manager*, DFM), indicando o destino de todos os eventos até então aprovados pelo LVL1. No caso de um determinado evento ser desconsiderado, o DFM envia uma mensagem para o ROS indicando que este deverá apagá-lo dos *buffers*, ou em caso contrário enviará uma mensagem indicando ao processador de construção de eventos (*SubFarm Input*, SFI) que deverá coletar os fragmentos do evento. O SFI então envia mensagens a cada um dos ROS's no sistema e requer todas as informações disponíveis para o evento de interesse. O objetivo do SFI é aglutinar todas as partes do evento em um único espaço de memória, liberando as partes anteriores do sistema do encargo do gerenciamento deste mesmo espaço.

Uma vez construído, o evento é guardado em uma pilha lógica dentro do SFI que aguarda um pedido de um Gerente do Filtro de Eventos (do inglês, *Event Filter Dataflow Manager*, EFD), indicando que está disponível para processar mais um evento. Este evento é transferido integralmente para o EFD neste momento. O EFD, de forma análoga ao L2SV, verificará a Unidade de Processamento (*Processing Task*, PT) disponível e a alocará para tratar as informações deste evento. Neste momento, todo o evento (com seus dados) é disponibilizado à PT, que poderá levar mais tempo processando o evento que sua análoga no LVL2, dada a taxa reduzida de eventos remanescentes. Se aprovado, o evento é encaminhado a um processador de saída (*SubFarm Output*, SFO), que coopera com o sistema de gravação de dados na

disponibilização dos eventos colhidos pelo DAQ. O sistema também é equipado com dispositivos que garantem o funcionamento de cada componente na falha de outros ou quando algum nó ou nós começam a responder mais lentamente provocando congestionamentos. Para funcionar plenamente, o *Dataflow* está montado sobre as bases do Sistema *Online*, que provê as ferramentas de controle e configuração necessárias para operar uma cadeia de aquisição completa.

### **Ferramentas do *dataflow***

Para se projetar um conjunto de aplicações que funcionem coerentemente é necessário partir de uma base funcional comum. Desta forma, optou-se pelo projeto e construção de um conjunto de ferramentas que fornecem uma base para o desenvolvimento das aplicações do DF. Estas ferramentas incluem:

- Ferramentas para isolamento do Sistema Operacional: reúnem um conjunto de bibliotecas que executam funções primárias como a criação de tarefas (para ambientes multi-tarefa), relógios e primitivas como filas protegidas. Esta camada de funções também isola, dentro do possível, variações no sistema operacional de alterações no código das aplicações do DF;
- Ferramentas para troca de mensagem: provêem uma interface simples para a troca de mensagens entre aplicações. Os canais de comunicação são configuráveis, permitindo o tráfego (transparente) de dados diretamente sobre pacotes *ethernet*, UDP ou TCP. Este pacote também engloba ferramentas para a codificação e decodificação dos dados provenientes dos detetores.
- Ferramentas para configuração: provêem interfaces para a configuração dos aplicativos utilizando o Sistema *Online*;
- Ferramentas para controle e relatório de erros: permitem o envio de mensagens de erro e depuração para os nós de controle da cadeia de aquisição. Este conjunto de ferramentas também inclui os aplicativos, nos moldes da máquina de estados apresentada na Figura 4.4.

Para aumentar a qualidade dos programas desenvolvidos e a manutenção a longo prazo (o experimento deve operar por 10 anos ou mais), optou-se por utilizar técnicas de orientação a objetos (OO) aplicadas sobre C++. Alguns problemas, e.g.

ineficiência da utilização dos *containers*-padrão da biblioteca padrão (*Standard Template Libraries*, STL) C++ em ambientes multi-tarefa, advieram desta decisão [33]. De fato, a alocação de memória implementada pelos compiladores da linha GCC [34, 35] visa a otimização da alocação de memória por processo enquanto em ambientes multi-tarefas (do inglês *multi-threading*) é necessária a utilização de pilhas separadas para cada tarefa. A solução neste caso foi a adoção de alocadores mais eficientes para este propósito, que não fazem parte do padrão C++.

#### 4.2.4 Fluxo de dados e a arquitetura no LVL2

Especificamente para este trabalho, a arquitetura e o funcionamento do segundo nível de filtragem do experimento são de especial atenção. É possível observar na Figura 4.6 um esquema da troca de mensagens das aplicações no LVL2, como anteriormente indicado na área acinzentada da Figura 4.5. Nesta figura, as linhas tracejadas indicam mensagens de controle do sistema e as linhas cheias, as mensagens que carregam dados do detetor ou resultados da operação de filtragem. Inicialmente, o Supervisor do LVL2 recebe uma mensagem contendo as informações das RoI encontradas pelo LVL1. As RoI's equivalem a objetos encontrados, seja nos detetores de múons ou nos calorímetros, com energia mínima para a ativação dos subsequentes níveis de filtragem. Esta mensagem é chamada de Resultado do LVL1, e abreviadamente de “L1R”. A mensagem que chega ao L2SV representa, portanto, uma avaliação do evento com uma qualidade equivalente à visibilidade que o LVL1 tem do detetor. Para continuar com a avaliação do evento, o L2SV alocará uma L2PU que esteja sob sua guarda para processá-lo.

A L2PU processa eventos por RoI. Para cada RoI encontrada em um resultado do LVL1, a L2PU requisitará os dados ao redor do objeto e avaliará novamente o objeto, levando em conta a máxima granularidade do detetor ao redor daquele ponto. Se o objeto encontrado pelo LVL1 tiver sido achado através de sua interação com o calorímetro, o LVL2 requisitará inicialmente os dados deste sub-detetor. Caso o objeto tenha sido encontrado nos detetores de múons, o processamento começará a partir deste sub-sistema. O processamento de cada evento na L2PU se dá de forma seqüencial. Para cada RoI destacada pelo LVL1 é feita uma confirmação da análise do LVL1 contando com a melhor precisão disponível para os dados avaliados

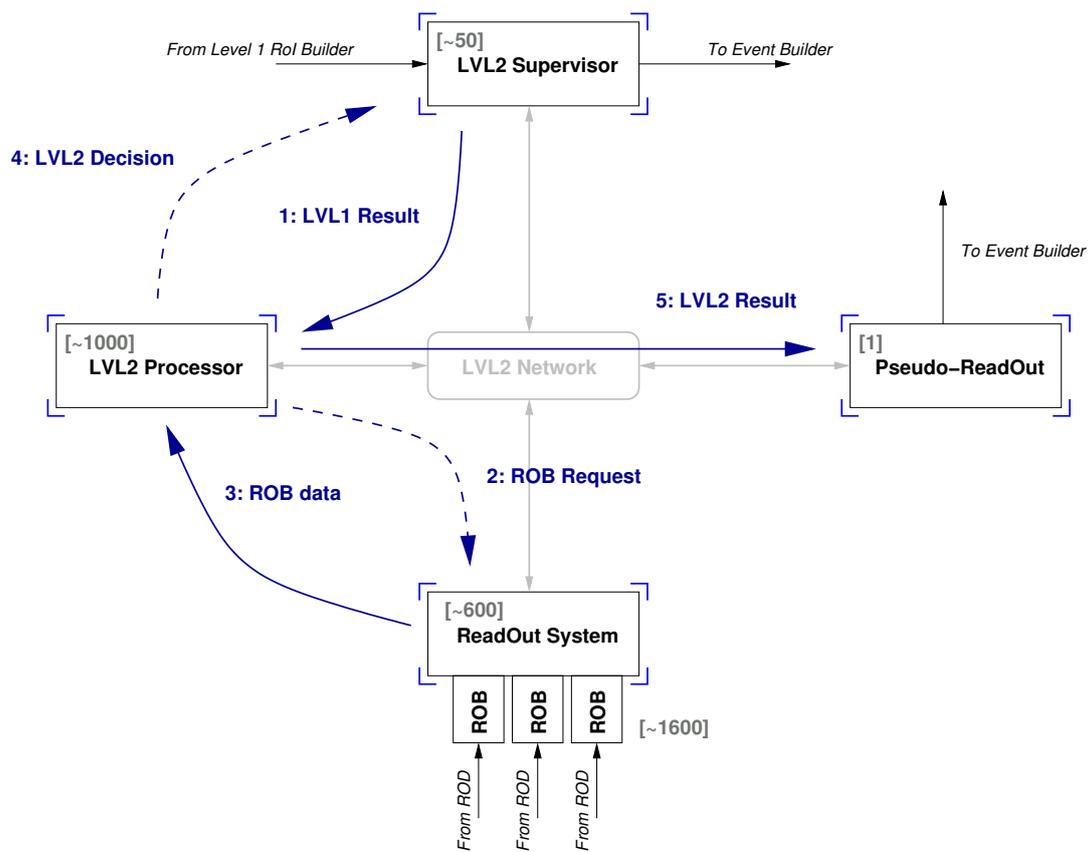


Figura 4.6: A ordem das mensagens trocadas pelas aplicações do LVL2.

e possíveis validações cruzadas, levando-se em conta dados dos detetores internos. Se a RoI for rejeitada, o processamento é abortado e imediatamente uma mensagem de controle é enviada ao L2SV indicando um evento inválido. Caso as diversas etapas de processamento se confirmem ao longo da análise do evento, a resposta ao L2SV é atrasada até que o evento seja confirmado em última instância ou rejeitado no meio do processamento. Nota-se aqui que eventos interessantes terão maior tempo de processamento que eventos ordinários, já que os últimos serão rejeitados assim que possível, enquanto que os primeiros terão que ser aprovados por todos os critérios de processamento no LVL2. O processamento como um todo continua requisitando dados, conforme a necessidade, aos diversos ROS's, até que a L2PU possa concluir, com baixíssima probabilidade de erro, se o evento deve ser aprovado ou rejeitado.

O L2SV opera de forma indiferente à rejeição ou aceitação do evento por um Processador do LVL2, já que repassará todas as mensagens, possivelmente agrupadas para melhorar o desempenho da comunicação de pacotes, ao DFM. Por outro lado, a cadeia de processamento para a L2PU se dá de forma ligeiramente diferente se tiver um evento confirmado. Caso o evento seja rejeitado, somente uma mensagem de controle é enviada ao L2SV. De outra forma a L2PU deverá, antes da mensagem de controle, enviar um extrato completo das operações executadas a um *buffer* indicado na Figura 4.6 por *Pseudo ReadOut*, PROS. Este único processador do sistema opera como um ROS, porém, ao invés de coletar dados de ROB's, armazena os resultados (diagnósticos) de eventos bem-sucedidos do LVL2, servindo estes resultados a um SFI quando o evento é construído. É desta forma que o resultado das operações do LVL2 poderá chegar ao Filtro de Eventos. O PROS é uma aplicação bastante simples que fornece 2 interfaces (lógicas) de rede. A primeira coleta os dados de todas as L2PU's do sistema e a segunda responde a requisições dos SFI's que desejam construir os eventos aprovados.

Após enviar o resultado das operações do LVL2 (L2R) ao PROS, a L2PU poderá enviar a mensagem de aceitação ao L2SV. Desta forma garante-se que a construção do evento operará corretamente, uma vez que o PROS não receberá um pedido para enviar o L2R antes de o ter recebido. O processamento prosseguirá desta forma por períodos que durarão até 10 horas e, portanto, erros de programação e, principalmente, vazamentos de memória são proibitivos neste árduo ambiente.

O orçamento do sistema é otimizado para canalizar a maior quantidade de recursos possível em poder de classificação de eventos. Portanto, a relação entre o número de L2PU's e os demais nós de processamento deve ser tão grande quanto possível. Atualmente estima-se que serão necessários cerca de 1.000 processadores (operando como L2PU's) para suportar a taxa de entrada de eventos de 100 kHz. Dado o paralelismo inerente ao processo, cada L2PU deverá processar um evento em não mais do que 10 milissegundos, i.e., cada um poderá suportar uma taxa de 100 Hz.

### Arquitetura da L2PU

A arquitetura da L2PU pode ser melhor descrita através da Figura 4.7, onde é possível observar três componentes primários:

- Tarefa de Entrada (*input thread*): Centraliza o recebimento de dados dentro da L2PU. Cada tipo de mensagem recebida está associada a um manuseador (*handler*) distinto que especifica o destino a ser dado à mensagem recebida. Por exemplo, mensagens do L2SV indicando resultados do LVL1 são encaminhadas para uma pilha protegida;
- Pilha Protegida: Controla o acesso concorrente das tarefas processadoras (*worker threads*) aos resultados do LVL1 depositados na L2PU;
- Tarefas Processadoras: É onde o processamento (seqüencial) de cada evento é feito. Cada tarefa possui um Coletor de Dados que é responsável pela comunicação com os ROS's disponíveis para reunir os dados necessários a cada passo do processamento de um evento.

Ao receber um L1R do Supervisor, a tarefa de entrada irá colocá-lo numa pilha tipo FIFO (*First In First Out*) protegida ao acesso concorrente de cada uma das tarefas processadoras. As tarefas concorrem na aquisição de um L1R para iniciar o processamento e, caso tenham sucesso, saem do modo de espera, onde praticamente não consomem a CPU do sistema hospedeiro, e iniciam a discriminação do evento. Para isso, requisitam dados aos ROS's disponíveis que enviarão, dentro de suas possibilidades, os fragmentos da RoI equivalentes ao subdetetor de interesse.

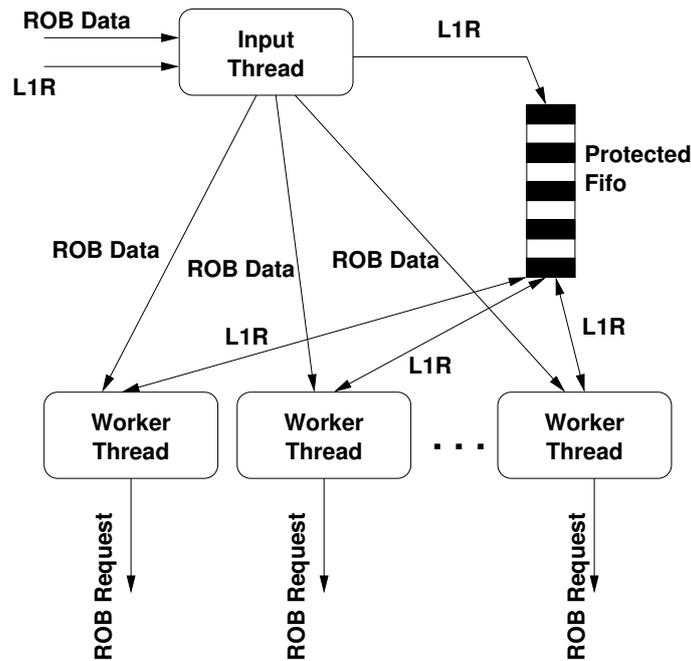


Figura 4.7: A arquitetura da L2PU.

É possível que determinados ROS's não consigam responder ao tempo máximo de espera da L2PU e, neste caso, o evento deve ser processado com dados faltantes. Há outras possibilidades que podem comprometer a integridade dos dados que serão finalmente manipulados na L2PU, já que a cadeia de aquisição, passando pela eletrônica de aquisição acoplada ao detetor, os ROB's e finalmente os ROS's, podem apresentar inúmeros tipos de falhas que resultarão na perda dos dados analisados pela L2PU, total ou parcialmente. Os algoritmos de discriminação devem estar preparados para estas anomalias de funcionamento, e devem ser construídos para consumir o mínimo possível de dados, já que, além da maior probabilidade de falhas, mais dados também significam maior banda passante necessária à rede que interliga estes dois nós de processamento.

Os fragmentos de RoI (e também do detetor) que são carregados pela L2PU são armazenados e transmitidos em um formato particular, e uma biblioteca de codificação e decodificação [36, 37] está disponível para este propósito. Os dados não são somente decodificados, mas também re-arranjados e calibrados antes do uso, o que pode consumir grande parte do tempo de processamento (10 ms) no LVL2.

O L2R juntamente com a decisão do LVL2 são enviados diretamente das tarefas processadoras ao L2SV e ao PROS sem a intervenção de outros componentes dentro

da L2PU.

## Parâmetros de *hardware* do DAQ

A escolha da tecnologia de processamento, tal qual aquela empregada para a rede de interconexão destes nós, é um parâmetro vital para viabilização do Sistema de Aquisição de Dados do ATLAS. Diversos aspectos devem ser considerados:

- **Econômico:** Deve-se maximizar os recursos disponíveis. Desta forma, deseja-se despende o mínimo necessário em recursos para a compra de equipamento. Por “mínimo necessário” entende-se o mínimo necessário para operar, de forma cômoda, o experimento por 10 anos;
- **Operacionais:** O equipamento deve ter sua operabilidade maximizada. Deve ser relativamente fácil repor componentes ou realizar atualizações que não impliquem a reposição de toda a cadeia de aquisição. A manutenção e operação dos programas deve ser acessível a novos usuários.

Por estas razões decidiu-se pela utilização de equipamentos disponíveis no mercado de comodidades atual: PC's operando sobre redes padrão (*gigabit* ou *10-gigabit*) ethernet. O sistema operacional escolhido foi Linux. A razão desta escolha se deve a estabilidade do sistema em operação contínua e a confiabilidade dos módulos para conexão em rede. A utilização de um sistema operacional aberto, gratuitamente distribuído, também facilita o desenvolvimento de *drivers* especializados para atividades específicas.

Para maximizar a operabilidade e diminuir o custo, pretende-se adquirir máquinas com dois ou quatro núcleos de processamento com memória compartilhada (arquitetura SMP, do inglês *Symmetric Multi-Processor*). Esta arquitetura apresenta várias vantagens:

- Para cada 2 ou 4 processadores, somente 1 disco e 2 interfaces de rede serão necessárias<sup>3</sup>. Desta forma, não há somente economia de espaço e dinheiro, mas também no tamanho da rede final que será projetada;

---

<sup>3</sup>Uma das interfaces de rede é alocada para sinais e dados de controle enquanto a outra funciona exclusivamente para o tráfego de dados do detetor. Nesta configuração, somente uma das duas interfaces de rede necessita de uma banda passante maior, a interface por onde trafegarão os dados obtidos no detetor.



- Há também economia de energia durante a operação e, principalmente, durante o momento de ligar as máquinas. A conexão concomitante de 2.500 discos rígidos pode acarretar problemas de super-aquecimento do *cluster* e quedas de tensão em outros pontos do sistema.

A arquitetura da L2PU, apresentada na Seção 4.2.4, deriva desta escolha do equipamento. Uma vez que somente uma das interfaces estará disponível para o tráfego de dados, a utilização de mais tarefas para esta atividade torna-se desnecessária. O número de tarefas processadoras, no entanto, deve ser otimizado para que o sistema atinja o seu desempenho máximo.

#### 4.2.5 Análise funcional do fluxo de dados no LVL2

Para avaliar a correta operação dos diversos processos que compõem o LVL2, é necessário operar os diversos componentes em uma bancada de teste (do inglês *testbed*). Por razões econômicas, o sistema em sua escala de produção (cerca de 2.500 nós de processamentos) somente estará disponível em uma data próxima ao início de operação do detetor. Por outro lado, a transferência na data de compra também possibilita a aquisição de processadores mais potentes pelo mesmo custo. Desta forma, as *testbeds* disponíveis para teste, ainda que representando o estado da arte dos equipamentos escolhidos, são de tamanho reduzido, se comparadas ao sistema final.

De forma a qualificar a avaliação de cada componente, faz-se necessário criar condições de operação que satisfaçam o modelo de operação final. Com relação ao LVL2, muitos modelos foram explorados [38] ao longo de 2 anos de testes. Estes modelos e a verificação do funcionamento coerente desta parte do sistema de aquisição encontram-se sintetizados nesta seção [16].

Como será detalhado mais adiante, na Seção 4.2.6, os módulos que executam a análise da Física de interesse são carregados dinamicamente na L2PU, através de parâmetros de configuração disponíveis ao operador do experimento. Esta estratégia permite que o sistema de fluxo de dados seja testado independentemente dos algoritmos de filtragem, paralelizando o desenvolvimento [39] e reduzindo significativamente as possibilidades de falha devido a componentes que não sejam relevantes para a natureza dos testes. Para os processos descritos nesta seção, foram utilizados

emuladores dos componentes de discriminação, que podem simular a aquisição de dados de RoI's e até o tempo de processamento entre tomadas de dados, possivelmente em múltiplos passos. Este bloco emulador é sintonizado para caracterizar os parâmetros normais de operação do LVL2.

### **Parâmetros críticos do sistema de leitura**

Um dos parâmetros que se deseja investigar é a capacidade da L2PU de coletar dados do sistema de leitura, neste caso representado pelos diversos ROS's. Como colocado anteriormente, ao processar um resultado do L1R, a L2PU começa por confirmar o resultado encontrado no nível precedente, utilizando, para isso, a granularidade máxima do detetor ao redor da RoI. O tamanho da RoI dependerá diretamente dos valores discriminantes que devem ser encontrados. Algumas modalidades de algoritmos podem requerer mais dados e, portanto, aumentar significativamente o tráfego de dados na rede. Algoritmos mais eficientes tenderão a consumir dados em pequenas parcelas, minimizando o tempo de rejeição necessário para cada evento do ruído de fundo. É possível, em todos os casos, adquirir pequenas parcelas de dados inicialmente e aumentar o campo de atuação do algoritmo à medida que o evento não é rejeitado. Por outro lado, também é possível compartimentalizar a aquisição de dados por sub-detetores. Cada estágio do processamento seqüencial do evento adquirirá, neste modelo, dados de diferentes camadas do ATLAS. A solução ótima está provavelmente na utilização das duas técnicas, ou seja, o consumo parcelado de dados para cada sub-detetor com rejeição assim que possível. Uma rejeição rápida garante mais tempo para o processamento de eventos realmente interessantes e minimiza o tráfego de rede.

Cada L2PU, como visto anteriormente, mantém uma fila tipo FIFO com eventos alocados pelo Supervisor do LVL2. A maximização do tempo de processamento disponível depende, portanto, da não-escassez de eventos nesta fila, uma vez que com a fila vazia, uma ou mais das tarefas processadoras serão bloqueadas. Desta forma, o L2SV é sintonizado cuidadosamente para que esta situação nunca ocorra.

Ainda que o número e o tamanho médio das RoI's dependam da Física a ser avaliada, o fator de concentração de ROB's (o número de ROB's que deve ser lido de cada ROS) é um parâmetro que deve ser otimizado. Variações de desempenho

nesta área indicam melhores geometrias para a interconexão do sistema de leitura do detetor e a rede que liga o LVL2. Por outro lado, o número ótimo de tarefas processadoras operando dentro de cada L2PU dependerá do tempo-morto despendido durante a aquisição de dados e do número de processadores disponíveis em cada máquina. Em máquinas com dois núcleos de processamento, por exemplo, o número ótimo de tarefas processadoras deve estar entre 3 e 4, já que o uso de um número menor deste tipo de tarefas implicará na sub-utilização do poder de processamento de cada nó.

Medidas foram realizadas no CERN, variando-se parâmetros críticos para o desempenho da coleta de dados para RoI's. A contribuição gerada pela coleta de dados para uma RoI, ao tempo de processamento de uma L2PU em função do tamanho da RoI e o número de ROS's na qual a RoI está distribuída, é mostrada na Figura 4.8. Estes testes foram executados em uma bancada com apenas um L2SV, doze ROS's e uma L2PU. Todos os nós nesta bancada de teste são PC's com dois núcleos tipo Intel/Xeon's de 2,2 GHz e 1 *gigabyte* de memória RAM, interconectados por uma chave padrão *gigabit* ethernet. Na figura, no eixo vertical é possível observar a taxa de processamento inversa de um L2SV com uma única L2PU sob seu controle. Este parâmetro representa o tempo médio que cada evento levou para ser inteiramente processado pelo LVL2, ou seja:

1. O tempo de envio do resultado do LVL1 para o L2PU;
2. A aquisição do resultado pela tarefa processadora;
3. O mapeamento da RoI nos ROB's de interesse;
4. A aquisição dos dados dos ROS's;
5. A resposta para o L2SV, indicando aceitação ou rejeição.

Neste caso, o tempo de processamento para os dados foi ajustado para zero. A resposta da L2PU ao L2SV é sempre negativa (para evitar o envio do L2R ao PROS).

No eixo horizontal da Figura 4.8, é possível observar o número de ROS's que contribuíram para os dados na RoI. O conjunto de medidas apresentado nesta figura corresponde a valores de tamanhos de RoI realísticos para dados partindo da seção

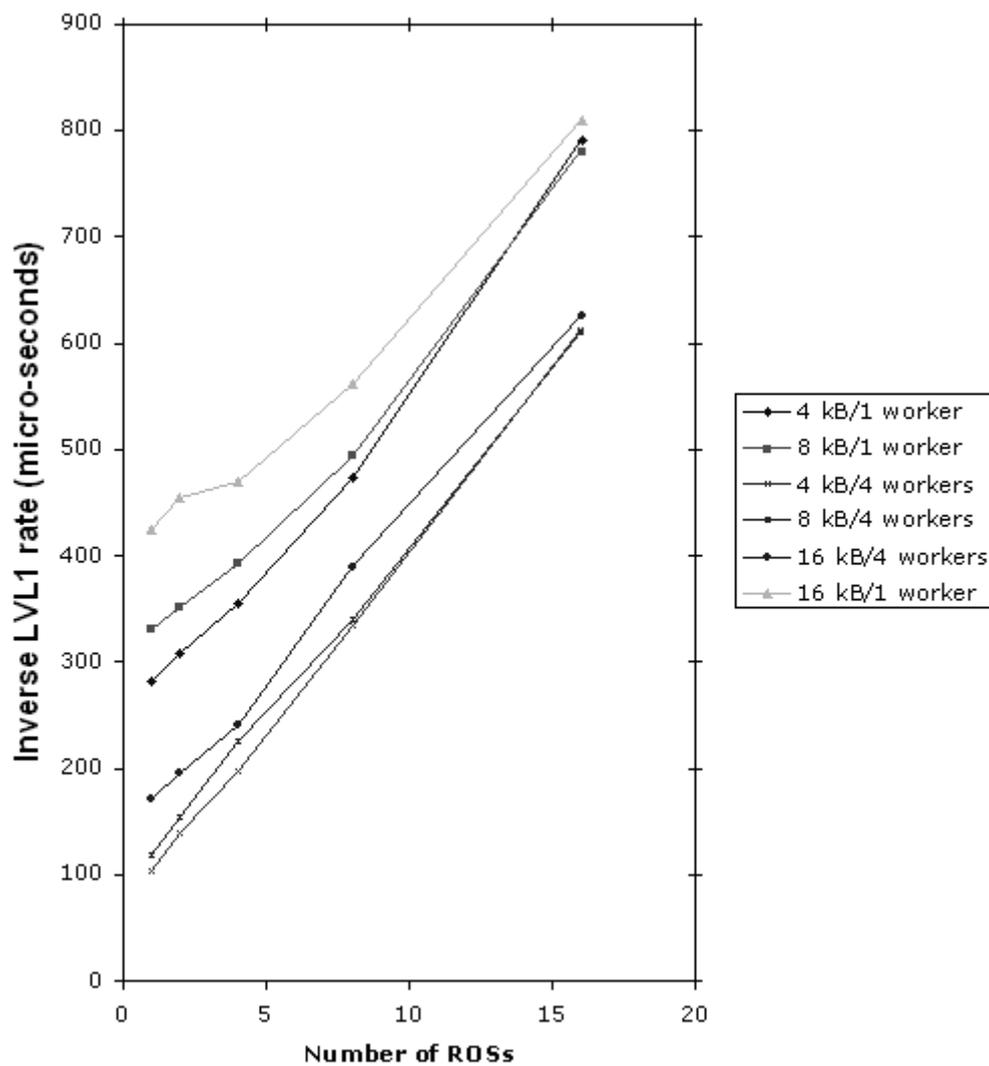


Figura 4.8: Desempenho da coleta de dados de RoI para várias combinações de tamanhos de RoI (em bytes) e tarefas processadoras. Extraída de [38].

e.m. do calorímetro, representando este o pior dos casos. Para este sub-detector, estima-se que a RoI para os dados de todas as três camadas do detector encontrem-se espalhada por 13 a 16 ROB's e tenha um tamanho total, em média, de 16 quilobytes. É possível notar, através deste gráfico, que o tempo de coleta de dados contribui, no pior dos casos, com apenas 10% do tempo de processamento médio estimado para cada evento no LVL2 (10 milissegundos).

Podemos também notar que o aumento no número de tarefas processadoras de 1 para 4 modifica significativamente o perfil do tempo-morto gasto na aquisição de dados, já que não haverá esperas espontâneas no sistema, como colocado anteriormente neste texto. No caso de 4 tarefas processadoras, o tempo de coleta de dados está na ordem de 600 microssegundos. Podemos também observar que com maior distribuição de dados no sistema (mais ROS's), o tempo de coleta de dados de uma RoI aumentará numa relação aproximadamente linear, o que casa bastante bem com o modelo OO apresentado. A minimização do número de ROS's é portanto desejável para um tempo-morto menor.

Na Figura 4.9 observa-se o desempenho de uma única L2PU contra a taxa de operação de um L2SV invertida para um número variável de tarefas processadoras. Neste teste o tamanho da RoI está fixo em 16 quilobytes. As diferentes curvas representam a mesma RoI sendo coletada de 2, 4, 8 ou 16 ROS's. Por exemplo, a curva superior representa os resultados para coletar uma RoI de 16 quilobytes de 16 ROS's (1 quilobyte por ROS). O resultado indica que o número ótimo de tarefas processadoras para esta bancada de testes (sistemas com dois núcleos) é de, aproximadamente, três e é independente do número de ROS's dos quais os dados da RoI estão sendo coletados. Ademais, os resultados mostram que, para as condições deste teste, e para 3 tarefas processadoras, a coleta de dados contribui com menos de 10 % do tempo de processamento médio esperado no LVL2 de 10 ms. Isto concorda com o estudo anterior.

## **Escalabilidade do LVL2**

Os testes apresentados até agora mostraram a taxa com a qual uma L2PU pode coletar dados, se não estiver equipada com algoritmos de discriminação de eventos. O desempenho atingido depende do tamanho da RoI, o fator de concentração de

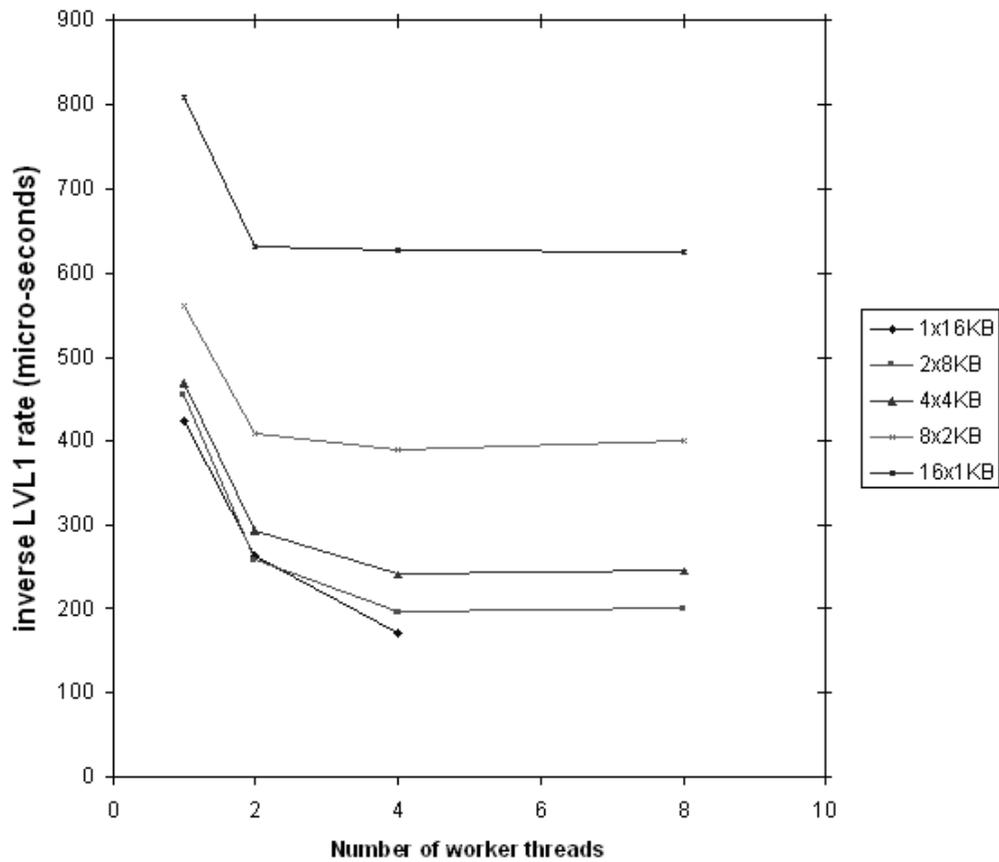


Figura 4.9: A taxa inversa em um L2SV para uma única L2PU funcionando com diferentes números de tarefas processadoras. Extraída de [38].

ROB's por ROS's e do número de tarefas processadoras disponível neste processador. Para o sistema de aquisição como um todo, no entanto, é também necessário demonstrar que, quando muitas L2PU's estão coletando dados ao mesmo tempo do mesmo conjunto de ROS's, o desempenho da coleta de RoI's não degrada de forma inaceitável.

Para este teste, L2PU's não equipadas de algoritmos de filtragem reais foram usadas novamente e, por esta razão, deve ser notado que a taxa de envio de mensagens de cada um destes nós estará operando num patamar bastante acima do que o sistema real necessitará. Para estes testes, cada L2PU estará gerando 10 vezes mais tráfego que o esperado no sistema final. De forma análoga e aditiva, o número de ROS's do sistema também não é ideal, devido à baixa disponibilidade de equipamento para testes. É, portanto, necessário garantir que a taxa de requisições de dados para cada ROS não exceda um valor limite que mascararia quaisquer resultados obtidos.

Estes testes foram executados em uma bancada com três L2SV's, quatro ROS's e até onze L2PU's. Todos os nós nesta bancada de teste são PC's, na mesma configuração de antes, interconectados por uma chave padrão *gigabit* ethernet. Para este teste, cada ROS foi configurado para emular 12 conexões com ROB's, gerando um total de 48 ROB's para este teste. Para cada requisição, a L2PU escolhe um dos 48 ROB's aleatoriamente e um número de ROB's com numeração consecutiva. Por exemplo, no caso em que cada RoI consista de 6 ROB's, a L2PU seleciona o primeiro aleatoriamente e completa o pedido com outros 5 ROB's cuja numeração sucede o primeiro ROB. Se o último ROB fosse selecionado, o algoritmo voltaria ao primeiro ROB disponível (*wrap-around*).

A Figura 4.10 traz os resultados sintetizados desta bancada. Nesta figura é possível observar a taxa de eventos média tratados em cada L2PU em função do número de L2PU's usadas no teste para uma RoI distribuída por 1, 6 ou 24 ROB's (com cerca de 1,5 quiloByte por ROB). Observa-se na figura que a taxa por L2PU diminui quando o número de L2PU's aumenta, por aproximadamente 20% para 1 ROB por RoI e aproximadamente 40% para 24 ROB's por RoI. Entretanto, com 11 L2PU's configurados na bancada, a taxa de requisições por ROS é de 17 kHz para 1 ROB por RoI e de 8 kHz para 24 ROB's por RoI. Estas duas situações representam

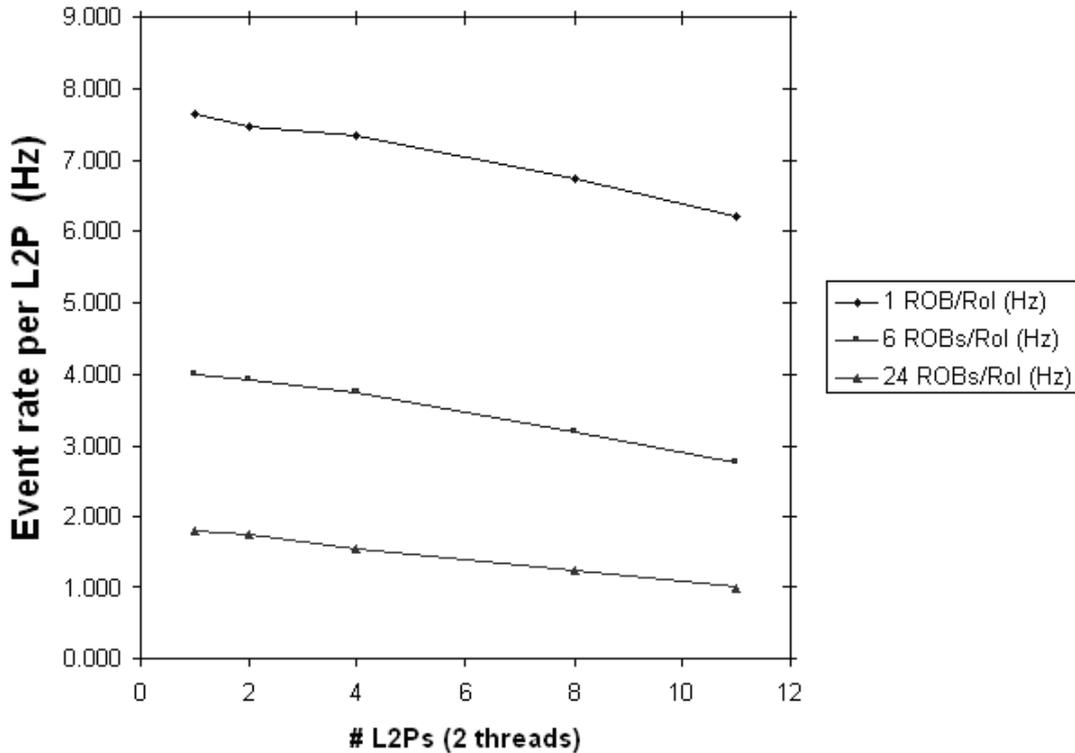


Figura 4.10: A taxa de eventos média para cada L2PU em função do número de L2PU's no sistema para diferentes números de ROB's por RoI. Cada ROB contribui com um montante igual dos dados da RoI. Extraída de [38].

uma sobrecarga jamais atingível no sistema final. A taxa total de requisições de dados de RoI nesta pequena bancada de testes é de 70 kHz e 11 kHz para os dois números de ROB's por RoI indicados.

Na Figura 4.11 é possível observar os resultados de uma bancada de testes equivalente contendo 1, 2, 4 e 8 L2PU's. Porém, desta vez, ajustou-se o tempo de processamento emulado para cada evento em 1 milissegundo. Isto representa apenas a décima parte do esperado no sistema final. Nesta figura, é possível observar um bom escalonamento do sistema. Dadas as extremas condições recriadas nestas bancadas, é razoável assumir que a escalabilidade do sistema final se mostrará aceitável às condições do DAQ.

Observa-se na Figura 4.12 a taxa sustentada por um L2SV em função do número de L2PU's sob seu controle. Nesta bancada, o nó de processamento rodando o L2SV está conectado através de uma chave *gigabit* ethernet, com um número de outras



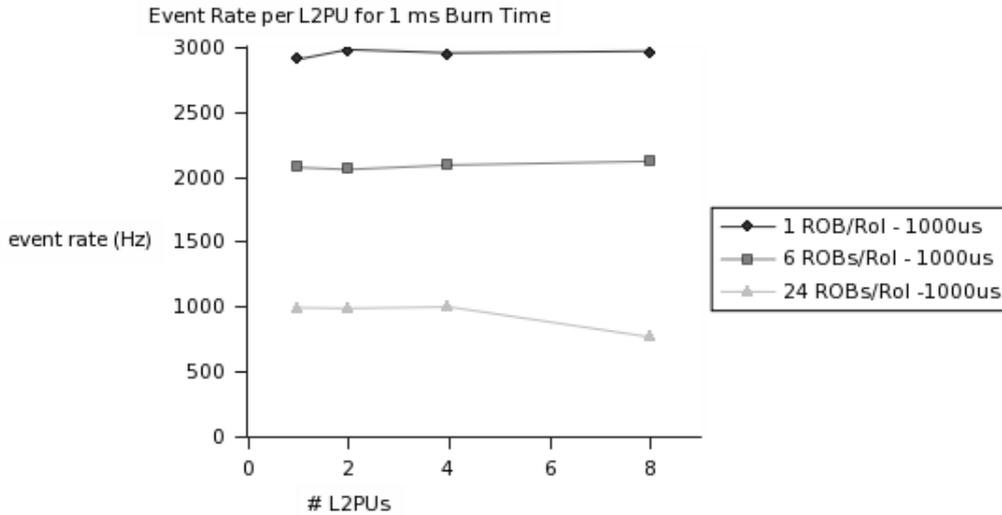


Figura 4.11: A taxa de eventos média para cada L2PU em função do número de L2PU's no sistema. O tempo de processamento emulado para cada evento é de 1 ms. Extraída de [38].

máquinas rodando o código da L2PU sem algoritmos de discriminação carregados e sem coletar dados dos ROS's. Este é portanto um teste extremo, visto que as L2PU's vão responder o mais rápido possível a cada requisição de processamento de eventos feita pelo L2SV. Como pode ser visto pela figura, um único L2SV pode sustentar a taxa de 32 kHz enquanto distribuindo L1R's para apenas uma das L2PU's sob seu controle, e a dependência no número de L2PU's está na ordem de 1%. Desta forma, tomando-se como ponto de partida a implementação do sistema de fluxo de dados e as máquinas, cerca de 10 L2SV's seriam capazes de suprir a taxa de entrada do LVL2. Há, todavia, outros fatores não considerados nesta suposição, tais como robustez e a conexão com o DFM, que certamente penaliza a taxa de processamento. Levando-se todos os pontos em questão, parece ser razoável contar com a presença de 30 a 50 L2SV's no sistema final.

Os testes aqui descritos descrevem e comprovam o funcionamento da infraestrutura que permite o acoplamento dos algoritmos que executam a discriminação dos eventos lidos pelo detetor ATLAS. A eficiência destes mesmos algoritmos é de extrema importância para o desempenho final do sistema de filtragem. Um conjunto de procedimentos ineficaz e lento será mais oneroso economicamente e também tecnicamente, já que a qualidade dos eventos finalmente aprovados se torna duvidosa.

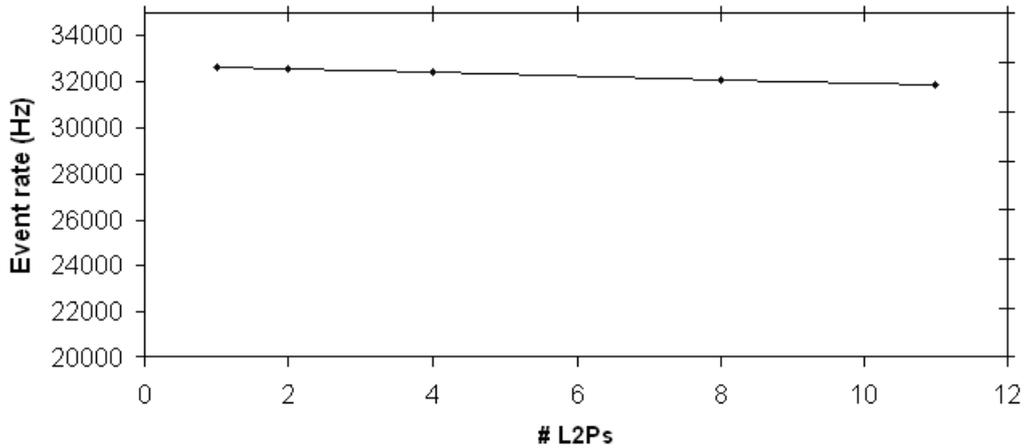


Figura 4.12: A taxa sustentada por L2SV com o aumento de L2PU's no sistema. Extraída de [38].

#### 4.2.6 Altos níveis de filtragem e seus algoritmos

A infraestrutura de seleção de eventos dos Altos Níveis de Filtragem constitui o ambiente de processamento para os algoritmos de discriminação do DAQ. Esta infraestrutura é comum ao LVL2 e ao EF, e é composta de 4 partes distintas. O Gerente do HLT (ou, do inglês, *Steering*) agenda os algoritmos do HLT correspondentes ao tipo de evento que está sendo analisado. Informações sobre quantidades físicas específicas a um evento são trocadas através de componentes do Modelo de Dados do Evento (do inglês *Event Data Model*, EDM). Durante o processamento, os dados gerados são guardados e acessados através de um Gerente de Dados (do inglês, *Data Manager*, DM). Isto permite transparência às escolhas de plataforma de armazenamento e tecnologia utilizadas para executar as operações de filtragem. Os algoritmos do HLT propriamente ditos reconstróem quantidades relativas aos eventos ou confirmam hipóteses através de quantidades previamente calculadas. A Figura 4.13 ilustra o relacionamento entre estes macrossistemas do HLT com alguns exemplos de classes representativas.

Uma vez que o propósito do *software* de discriminação de eventos do HLT é a seleção de eventos, ele deve executar de maneira eficiente e confiável no ambiente do DAQ/Fluxo de Dados. Os componentes críticos para a seleção devem ser portados ao ambiente *offline* para seu desenvolvimento e teste. Provendo um conjunto

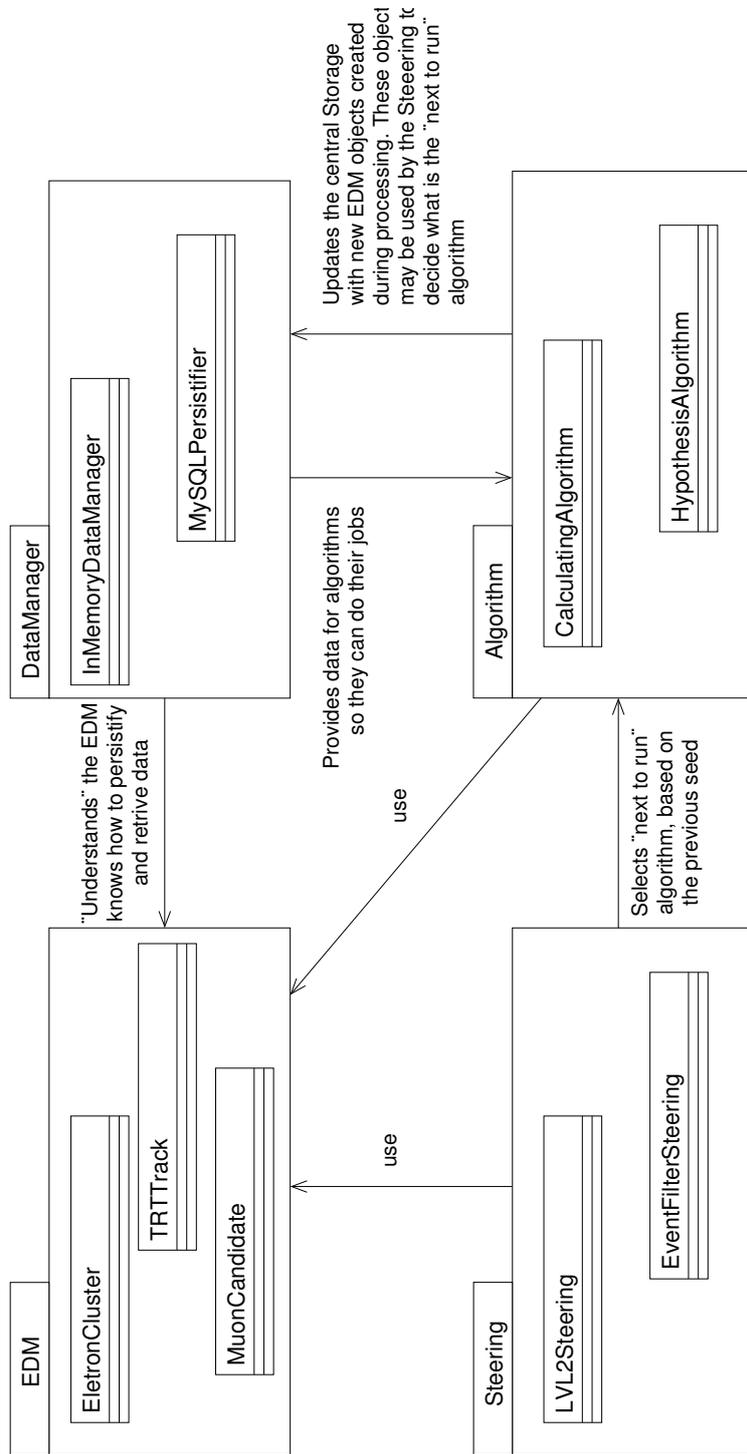


Figura 4.13: Relação entre os macrosistemas do HLT. Extraído de [16].

básico de interfaces para ambos os cenários, o HLT garante a consistência para o desenvolvimento do sistema de filtragem. Ademais, alguns estudos [40] mostram que um grande custo computacional pode ser economizado com a otimização global do sistema de filtragem quanto à ordem e a qualidade dos algoritmos executados. A existência de uma infraestrutura comum também facilita este tipo de estudo.

Uma vez que o Filtro de Eventos provê um ambiente de computação praticamente igual ao ambiente *offline*, o conjunto de rotinas do HLT é naturalmente baseado no ambiente de análise *offline*, Athena, que por sua vez é baseado na infraestrutura de análise física Gaudi [41, 42]. Esta estratégia permite a re-utilização do Gerente de Dados, do EDM, da descrição dos detetores e de muitos algoritmos que já se encontram desenvolvidos por esta comunidade. Apenas o Gerente do HLT (*Steering*) e determinados algoritmos restam como desenvolvimentos específicos a este subsistema. No caso específico do LVL2, um enfoque similar torna-se mais complexo, já que o processo de seleção ocorre em múltiplas tarefas concorrentes às unidades de processamento disponíveis e restrições mais severas tornam-se necessárias. Ainda que os algoritmos destinados ao LVL2 sejam especialmente desenvolvidos para esta plataforma (devido às restrições no desempenho), eles utilizam derivações simplificadas do EDM e da descrição do detetor presentes no EF e *offline*. A utilização transparente de tais componentes é possível e uma implementação comum da infraestrutura do HLT para ambos o LVL2 e o EF pode ser realizada se a mesma interface estiver disponível para o LVL2. Esta funcionalidade é provida pelo Guia ou Gerente do HLT no LVL2 (*LVL2 Steering Controller*, L2SC ou PSC).

## O L2SC

O L2SC é o componente de *software* que faz a interface entre a L2PU, que provê acesso aos dados do detetor através dos ROS's e dos ROB's, e os componentes do HLT. O propósito do L2SC está em 3 aspectos: permitir que a L2PU seja hospedeira do *software* do HLT e sua controladora primária; permitir a re-utilização do mesmo conjunto de componentes (algoritmos, EDM, etc) entre o LVL2, o EF e *offline*; prover um mecanismo de transmissão dos resultados do LVL1 e do LVL2 entre o DAQ e o *software* de seleção de eventos. Por esta razão, da mesma forma que a L2PU, o L2SC deve seguir a máquina de estados descrita na Figura 4.4, ativando

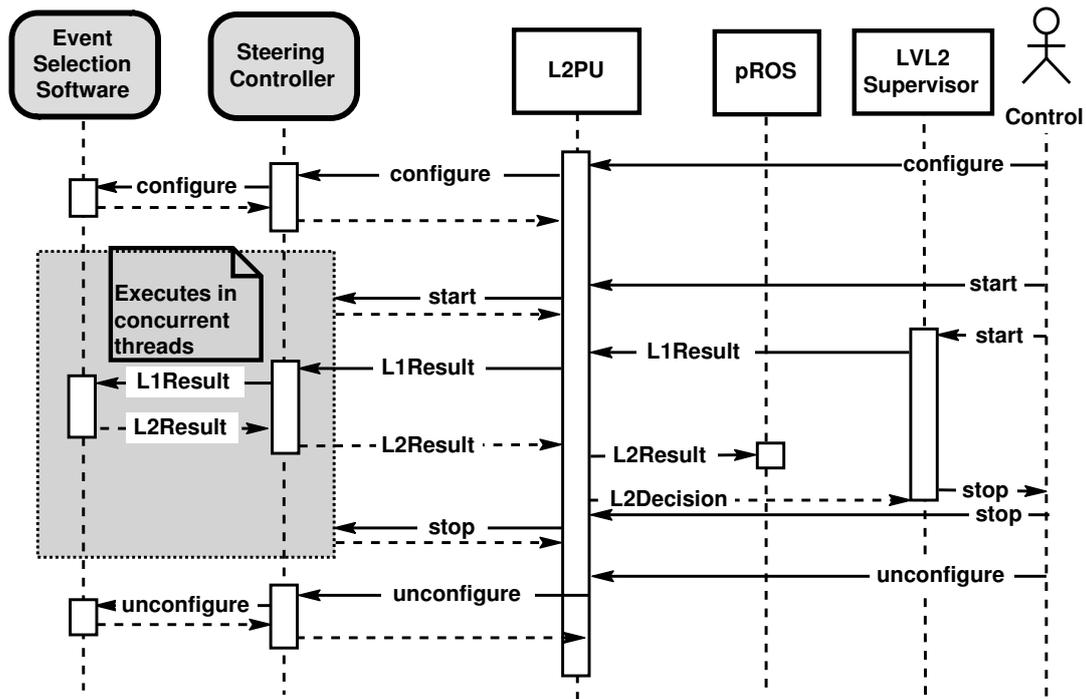


Figura 4.14: A seqüência de interações do L2SC com o L2PU e o *software* de seleção de eventos. Extraída de [39].

seus múltiplos estágios de configuração e processamento conforme requerido pelo operador do DAQ. Um aspecto importante deste enfoque é que o acesso aos dados providos pelo LVL2 é totalmente gerenciado pela base funcional do Sistema de Fluxo de Dados. O L2SC não precisa interagir diretamente com a tarefa de entrada ou quaisquer outros componentes da aplicação hospedeira.

A Figura 4.14 ilustra a seqüência de interações do L2SC com a L2PU e com alguns dos componentes pertinentes do HLT. Durante a fase de configuração, parâmetros de execução e dados atrelados às condições de detecção são obtidos de bancos de dados externos através desta interface. Estes dados são posteriormente utilizados para configurar o sistema de seleção de eventos e componentes associados. Três estados são mostrados: *Configuração*, *Inicialização* e *Término*. A área em cinza mostra interações que ocorrem em tarefas processadoras (concorrentes).

Após a *Inicialização* o L2SC recebe uma chamada `Execute Event` com um resultado do LVL1 como argumento. O resultado do processamento do evento é automaticamente retornado como resultado do LVL2 para a L2PU. O comando *Término* termina a execução dos algoritmos e produz um sumário do período de execução.

## Modelo de desenvolvimento do L2SC

Uma vez que as mesmas interfaces estão disponíveis no LVL2 e no EF, o código desenvolvido no ambiente *offline* pode ser diretamente carregado em seu formato binário nestes processadores. Para o LVL2, no entanto, o desenvolvedor deve seguir um conjunto de regras simples [43] durante o ciclo de desenvolvimento para produzir código seguro ao acesso concorrente (do inglês *thread-safe*) ou serviços compatíveis com a criação automática de múltiplas cópias na L2PU. Não deve ser necessário utilizar trancas (*locks*) ou *mutexes* para adaptar o código a este ambiente multi-tarefas. Para respeitar as restrições no tempo de execução para o LVL2, o número e o tipo de serviços disponíveis estão restritos ao mínimo necessário, sendo este um subconjunto do serviços disponíveis no EF e *offline*. Desta forma é sempre possível mover um componente do LVL2 para executar no EF ou até *offline*. A movimentação de algoritmos no sentido oposto, no entanto, deve observar as restrições de operação no LVL2.

## Desempenho do HLT

Após integrar, o L2SC com o Sistema de Fluxo de Dados, testes de desempenho e robustez foram conduzidos no CERN em uma máquina com dois núcleos tipo AMD/Athon's de 1,533 GHz. O L2SC rodou por mais de 50 horas com três tarefas processadoras. O protótipo executou sem problemas aparentes (*dead-locks*, vazamentos de memória, etc.) em nós simples ou com dois núcleos, mostrando a segurança ao acesso concorrente. Medidas do desempenho desta infraestrutura exibiram um *overhead* de 13  $\mu$ s por evento. Este valor foi estimado comparando-se o número de eventos por segundo que uma L2PU pode manusear quando operando com e sem o L2SC e executando um algoritmo simples (primeira e segunda linhas da Tabela 4.1). Esta tabela também mostra a taxa de processamento obtida quando, adicionalmente, o resultado do LVL1 é transferido para o Gerente de Dados. O valor de *overhead* mencionado inclui toda a infraestrutura Athena para agendar e executar os algoritmos e os serviços básicos. Estes valores são baseados em execuções de, no mínimo, 100.000 eventos. Escalonamento quase perfeito da latência medida foi observado quando o tempo de execução dos algoritmos era variado com um ciclo de processamento fictício de 0 a 8 milissegundos em configurações distintas, indi-

Tabela 4.1: Taxas de processamento em um sistema com dois núcleos tipo AMD-Athlon 1,533 GHz.

Configuração do Protótipo	Taxa Medida	<i>Overhead</i> por Evento
L2PU	21,7 kHz	46 $\mu$ s
L2PU + L2SC	17,0 kHz	59 $\mu$ s
L2PU + L2SC + Gerente de Dados	15,3 kHz	65 $\mu$ s

cando que o *overhead* por evento introduzido pelo L2SC é independente do tempo de execução observado.

Na seqüência, testes mais complexos foram feitos com uma fatia completa do sistema de seleção para elétrons e fótons ( $e^-/\gamma$ ). Estes testes continham o Gerente do HLT, um algoritmo decodificador do resultado do LVL1, agendamento de algoritmos para o LVL2 e o envio dos resultados do processamento do LVL2 para o EF. O Gerente de Dados, o EDM e a descrição dos detetores para os calorímetros do ATLAS foram também carregados no conjunto de teste, que estava funcionando sobre uma bancada que consistia de três nós de processamento: dois com dois núcleos (o AMD/Athlon anteriormente mencionado e um Intel/Xeon de 2,2 GHz) hospedando o ROS e a L2PU respectivamente, interconectados por *Gigabit* ethernet e um nó com apenas um processador (Intel/Pentium 4 de 2,2 GHz) para hospedar o L2SV. O algoritmo de extração de características para dados do calorímetro foi especialmente desenvolvido para o LVL2, utilizando toda a infraestrutura disponível mencionada anteriormente. Para este teste, 95 % dos eventos foram processados em menos de 5 milissegundos para uma amostra simulada de jatos duplos (*di-jets*) em baixa luminosidade, para um tamanho de RoI de  $\Delta\eta \times \Delta\phi = 0,3 \times 0,3$ . A Figura 4.15 exhibe resultados similares para uma fatia completa de seleção para múons. Nesta figura é possível, ver que as contribuições, em ordem decrescente, para o tempo de processamento total de cada evento são pré-processamento dos dados e sua conversão para o EDM (*Data conversion*), os *overheads* da infraestrutura provida pelo HLT (*Steering*), processamento do algoritmo de extração de características (denotado por **muFast** na figura) e o tempo de acesso dos dados pela rede (*Network latency*). As curvas mostram que, neste exemplo, o processamento termina, para 95% dos even-

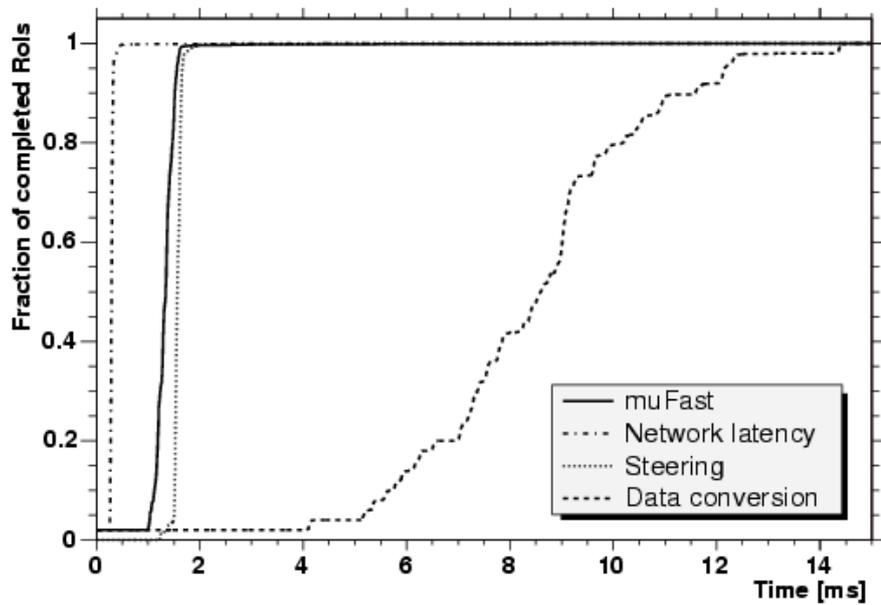


Figura 4.15: As principais contribuições para o tempo de processamento mostradas como integrais para a fatia completa do sistema de seleção para RoI's de múons. Extraída de [16].

tos, em menos de 2 milissegundos. A amostra de dados analisada consiste de múons com massa de 200 GeV em um ambiente de alta luminosidade do feixe. Esta amostra, que é simulada através de processos de Monte Carlo, inclui um ruído de fundo equivalente ao que será observado na caverna do ATLAS em sua disposição final.



## Capítulo 5

# Filtragem baseada em calorimetria no experimento ATLAS

Sistemas de calorimetria são habitualmente construídos com forma e leitura segmentada de maneira que possam ser usados diretamente na detecção de partículas. Isto é possível uma vez que muitos processos interessantes são distinguíveis pela forma com que depositam energia à medida que interagem com estes detetores, tanto de forma radial quanto longitudinal [44]. Cada segmento ou célula de um calorímetro contém a informação de energia depositada por uma partícula ou conjunto de partículas que interagiu com o detetor naquela área específica. A segmentação destes instrumentos dependerá da Física do experimento e também, normalmente, do eixo de dispersão das partículas de interesse.

Algoritmos de detecção são normalmente concebidos por especialistas em Calorimetria que compreendam os diferentes perfis de interação para a Física de interesse. O primeiro passo é definir um processo de redução de dimensionalidade, de acordo com os processos que se deseja detetar, ainda que levando-se em consideração à velocidade requerida na detecção. Quanto menor o tempo disponível à detecção, maior a compressão a ser realizada de forma a simplificar este processo. Este pré-processamento para a compressão dos dados ou Extração de Características (do inglês, *Feature Extraction* ou FEx) é, em geral, feito com perda de informação do perfil de deposição, mas de forma a manter as principais características do objeto que se deseja detetar. Em seguida, tendo por base simulações ou dados reais da Física de interesse, define-se um conjunto de cortes por métodos com conhecimento

da informação *a priori* que maximizem a probabilidade de detecção das partículas interessantes.

O Sistema de Filtragem *online* do experimento ATLAS deverá prover uma seleção de eventos muito eficiente e desprovida de tendências, mantendo o potencial de descoberta do detetor. Deve ser extremamente flexível, de forma a operar no ambiente desafiador do LHC, com até 23 colisões inelásticas por interação, 14 TeV de energia no centro de massa e apenas 25 ns (40 MHz) entre colisões sucessivas. Além disso, deve prover critérios de seleção robustos e, onde possível, redundantes. É altamente desejável a rejeição de canais ruidosos, não interessantes ou falsos o mais rápido possível, de forma a otimizar o uso dos recursos computacionais disponíveis.

Para trazer uma eficiência compatível com este ambiente, o Sistema de Filtragem fará uso de um enfoque baseado na **inclusão** para a seleção *online*, garantindo assim um conjunto de *triggers* ótimo para a nova Física. O critério da inclusão tem por objetivo selecionar um conjunto de fenômenos físicos baseados em características comuns entre eles, que ainda sejam claramente distinguíveis dentro do ruidoso ambiente produzido pelo acelerador. Com este objetivo, o Sistema de Filtragem estará sendo disparado para eventos que contenham assinaturas baseadas em objetos (partículas) simples ou duplos com alto valor de momento transverso. Neste contexto, objetos com alto valor de momento transverso são, entre outros, *léptons* com carga e com momento transverso acima de  $\sim 10$  GeV [16].

Neste capítulo, fazemos uma análise do processo de filtragem de elétrons a ser conduzido no Sistema de Filtragem do experimento ATLAS. Como em muitos outros experimentos em Física de Altas Energias [45, 46], este sistema é seccionado em níveis que realizam o processo de filtragem em passos com complexidade e tempo de execução crescentes. O primeiro dos níveis é normalmente implementado em *hardware*, para atender as demandas de tempo do feixe, enquanto que os demais em *software*, maximizando a flexibilidade. Em particular, atenção será dada à definição do funcionamento do Primeiro e Nível de Filtragem (LVL1) com relação a detecção de elétrons e, em seguida, o sistema desenvolvido atualmente para o LVL2. Finalmente, uma caracterização dos dados disponíveis segundo as premissas desenvolvidas neste capítulo é estabelecida como base de comparação para o restante do trabalho.

## 5.1 Objetos de interesse e RoI's

Foi visto no Capítulo 4 que o Primeiro Nível de Filtragem do ATLAS define, para objetos encontrados seja nos Calorímetros ou nos Detetores de Múons, regiões de interesse onde objetos relevantes tenham sido observados. Para cada objeto interessante, a região aproximada de interação do objeto com o detetor é anotada e repassada aos demais níveis de filtragem, no caso da aprovação do evento pelo LVL1. Este processo define as Regiões de Interesse ou RoI's, que guiarão o processo de seleção no HLT.

De fato, dadas as condições do feixe provido pelo LHC, múltiplos objetos poderão ser detetados a cada evento. Porém, quando um ou mais objetos detetados ativam a aceitação do evento, o LVL1 marca o dado de forma especial, distinguindo este objeto como primário, em contraste com aqueles que também foram detetados, mas não fizeram objeto da aceitação deste nível de filtragem. O processo de detecção no LVL2 começa por uma confirmação dos objetos primários detetados pelo LVL1, utilizando-se da granularidade completa do detetor ao redor da RoI. Em seguida, o processo de filtragem poderá, eventualmente, ser refinado ainda neste nível de filtragem ou mais a frente, no EF, complementando-se a análise com base em RoIs secundárias. Este sistema de filtragem *online* baseado em RoIs é um elemento de diferenciação do experimento ATLAS [7], sendo esta a primeira vez que é implementado em um ambiente em Física de Altas Energias.

A Tabela 5.1 apresenta as taxas (simuladas) para as diferentes assinaturas de interesse, ou conjuntos de objetos primários, que serão entregues pelo LVL1 [16] ao HLT, encabeçados pelo LVL2, para a luminosidade inicial de  $2 \times 10^{33} \text{cm}^{-2} \text{s}^{-1}$ . Este valor de luminosidade está cerca de uma ordem de magnitude abaixo do valor final de operação do LHC, de  $\approx 10^{34} \text{cm}^{-2} \text{s}^{-1}$ . Com o aumento da luminosidade do feixe, a relação entre as taxas ou até o tipo das assinaturas nesta tabela sofrerá alterações tendo em vista a mudança na probabilidade de ocorrência de fenômenos mais raros.

Nota-se que, aproximadamente, 16 dos 25 kHz (i.e.  $\sim 65\%$  das assinaturas) entregues pelo LVL1 conterão objetos tipo e.m.. Estes objetos são assim classificados por representarem elétrons **ou** fótons, que interagem com os detetores a partir de interações eletromagnéticas em vez das componentes hadrônicas presentes em outros casos. Em algumas das assinaturas, exige-se que o objeto seja *isolado*. Isto quer

Tabela 5.1: Taxas de saídas para as assinaturas reconhecidas pelo LVL1, para a luminosidade de pico inicial de  $2 \times 10^{33} \text{cm}^{-2} \text{s}^{-1}$ . Dados baseados em uma simulação do LVL1.

Assinatura do LVL1	Codinome	Taxa (kHz)
e.m., 25 GeV, com isolamento	EM25i	12
$2 \times$ e.m., 15 GeV, com isolamento	2EM15i	4
Múon, 20 GeV	MU20	0.8
$2 \times$ múon, 6 GeV	2MU6	0.2
Jato, 200 GeV	J200	0.2
$3 \times$ jato, 90 GeV	3J90	0.2
$4 \times$ jato, 65 GeV	4J65	0.2
Jato, 60 GeV + <i>Missing Energy</i> , 60 GeV	J60 + XE60	0.4
$\tau$ , 25 GeV, com isolamento + <i>Missing Energy</i> , 30 GeV	TAU25i + XE30	2
Múon, 10 GeV + e.m., 15 GeV, com isolamento	MU10 + EM15i	0.1
Outros disparos (pré-escalonados, aleatórios, calibração, monitoração)	–	5
Total	–	$\sim 25$

dizer que a quantidade de energia no centro do objeto deverá manter uma razão mínima com relação a energia depositada em sua periferia, como ficará claro mais a frente.

## 5.2 Análise do sistema de filtragem com relação à Física

A Figura 5.1 apresenta um diagrama esquemático que despreza as interfaces em *hardware* propostas pela construção do Sistema de Filtragem e tenta enfocar melhor o problema do ponto de vista da análise dos canais físicos. De cima para baixo, indica-se o tráfego de mensagens até a completa aceitação de um evento por parte do Sistema de Filtragem.

O processamento começa à medida que o LVL1 recebe o sinal do LHC acusando o iminente cruzamento de pacotes (do inglês *Bunch Crossing*) no ponto de impacto (mensagem ❶). Uma janela de aquisição de alguns nanossegundos fará com que o Sistema de Leitura do LVL1 registre o evento (mensagem ❷). Os sistemas de pré-processamento são ativados em seguida.

No caso dos Calorímetros, somadores rápidos [27] agrupam as informações das células dos calorímetros e.m. e hadrônicos, formando, para cada seção, uma matriz de dados com granularidade reduzida [29]. Estas macro-células ou “torres” de filtragem (TT’s) agrupam os dados formando elementos com tamanho igual a  $0,1 \times 0,1$  no plano  $\eta \times \phi$ . Estes dois planos com granularidade minimizada são varridos pelos diversos processadores do LVL1 e os objetos de interesse, ou seja RoI’s, localizados no detetor. Um processo análogo é conduzido tendo por base os detetores de múons, com o objetivo de encontrar objetos deste tipo (mensagem ❸).

No final do processamento pelo LVL1, rapidamente depois da colisão, e no caso de tratar de um evento interessante, uma mensagem contendo a localização dos objetos encontrados por este nível de filtragem e variáveis indicando as assinaturas pelo qual o evento foi aprovado, é repassada ao LVL2 (mensagem ❹). Esta mensagem é chamada de Resultado do LVL1 (do inglês *LVL1 Result*) ou L1R, como visto na Seção 4.2.4. Nota-se que o L1R pode conter múltiplas assinaturas em um único evento, no caso deste conter elementos que confirmem mais de uma das linhas da

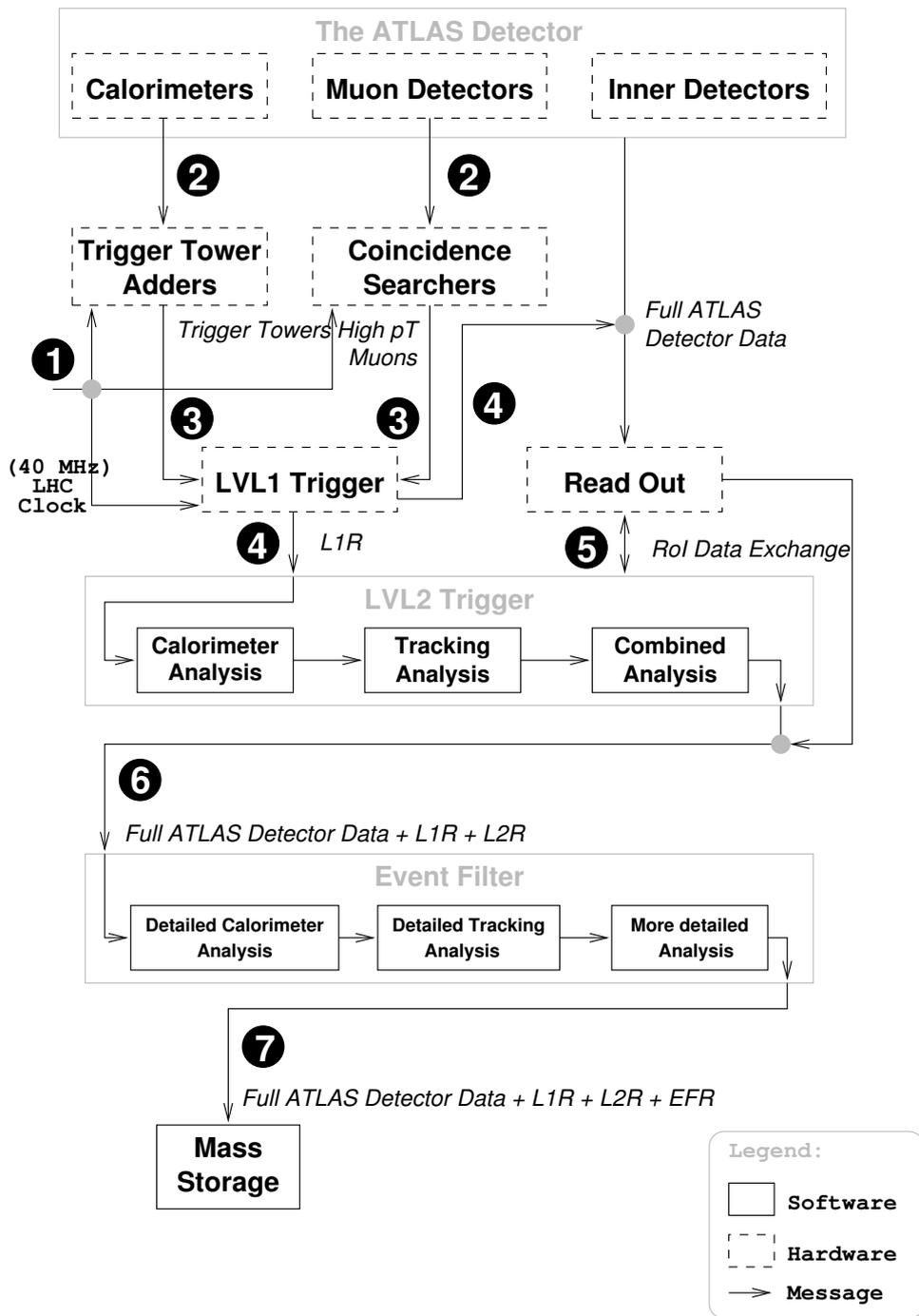


Figura 5.1: Diagrama esquemático do Sistema de Filtragem do Experimento ATLAS segundo suas funções de filtragem.

Tabela 5.1. Ademais, múltiplas RoI's secundárias podem estar presentes na lista de objetos na mensagem. Em paralelo ao envio da mensagem em direção ao LVL2, o LVL1 indica ao sistema de aquisição de dados (marcado na figura como *Read Out*) que carregue todos os dados do detetor em seus *buffers*.

O processamento no LVL2 é seqüencial. Ao receber um L1R, o Coordenador de Execução do LVL2 (do inglês, *Steering*) chamará um algoritmo de filtragem para a confirmação do disparo do LVL1. O algoritmo em questão poderá transferir, do sistema de leitura do detetor, a quantidade de dados que achar necessária, ao redor da RoI, de forma a processá-la (mensagem ⑤). Se não for possível rejeitar a assinatura do LVL1, o *Steering* poderá chamar mais e mais algoritmos até que o evento seja rejeitado ou definitivamente aprovado ao Terceiro Nível de Filtragem ou Filtro de Eventos (EF). Neste último caso, de forma equivalente à relação entre o LVL1 e o LVL2, o LVL2 enviará um sumário indicando a razão da aceitação do evento ao EF (mensagem ⑥). Este sumário contém valores refinados de energia, momento e localização dos objetos explorados no contexto LVL2. Analogamente ao resultado do LVL1, o resultado do LVL2 é chamado L2R (do inglês *LVL2 Result*). Espera-se que o LVL2 atinja uma redução da taxa de eventos de aproximadamente 25 vezes, de forma que o EF possa operar a uma taxa de apenas alguns (1 ou 2) kHz. O tempo médio de processamento para o LVL2 deve estar na ordem dos 10 ms e para o EF, na ordem de alguns segundos, levando-se em consideração o número de processadores disponíveis nestes níveis de filtragem. No caso do evento ser aprovado pelo EF, é armazenado em mídia permanente (mensagem ⑦).

Os tempos de processamento-alvos representam a média de processamento por evento. De fato, espera-se que eventos de interesse utilizem mais tempo de processamento, enquanto que eventos pouco ou nada interessantes sejam rapidamente descartados. Neste contexto, entende-se que, quanto menor o tempo necessário e maior a qualidade da detecção dos diversos componentes do sistema de filtragem, mais tempo computacional será despendido com a Física de interesse e menos recursos com eventos que representem Física ordinária.

Diferentemente do LVL2, a análise no EF começa com 100% dos dados do evento e os resultados dos nível precedentes (L1R e L2R) já disponíveis na memória. Neste nível de filtragem, correlações mais complexas podem ser feitas entre as diferentes

RoI's, e uma análise mais depurada do evento poderá ser conduzida. Por exemplo, é possível fazer uso das RoI's secundárias para atingir maior redução da taxa de eventos a serem registrados em mídia permanente. Se aprovado, o evento é guardado, devidamente “etiquetado” e repassado às fazendas (*farms*) de processamento *offline* para posterior reconstrução.

### 5.3 Detecção de elétrons baseada em calorimetria

O *trigger* de elétrons ou fótons para o ATLAS conterà forte contaminação de jatos com componentes hadrônicas, que formam o ruído de fundo na detecção de elétrons [47]. Estima-se que o ruído de contaminação original esteja na ordem de  $10^6$ . O LVL1 do ATLAS poderá excluir grande parte dos jatos QCD, se as conversões hadrônicas forem importantes neste objeto, e apenas observando a quantidade de energia hadrônica depositada na seção de mesmo nome dos calorímetros, mas terá pouca eficiência em conversões com menos componentes hadrônicas. Estima-se em [48], que a cada 25.000 objetos do tipo e.m., aprovados pelo LVL1, apenas 1 será um elétron ou fóton verdadeiro. Somente no LVL2, o Sistema de Filtragem, disporá do tempo e da completa granularidade do detetor para executar uma seleção mais criteriosa, refinando a decisão obtida no LVL1 [16] (mas originalmente discutida em [49]). O algoritmo de calorimetria também refina a posição do objeto originalmente determinada pelo LVL1, de forma que algoritmos de detecção de traços, reconhecidamente menos eficientes, possam reduzir sua área de procura. Por estas razões, a correta identificação de elétrons é de suma importância para o experimento.

A Figura 5.2 mostra a interação de um elétron com um detetor de traços (*Cloud Chambers*) e absorvedores de chumbo. Nesta figura, é possível observar a formação de um chuveiro de partículas que diverge, aproximadamente de forma isotrópica, do eixo de penetração do elétron. No caso de elétrons [44], a dispersão é menor que no caso de jatos de partículas, já que há uma grande probabilidade de espalhamento múltiplo (do inglês *Multiple Scattering*) causado pelas componentes hadrônicas do conjunto de partículas. A distância de penetração é diretamente proporcional à energia do objeto em análise [12, 11], ainda que, para os mesmos valores energéticos, jatos tendam a penetrar mais profundamente nos dispositivos.



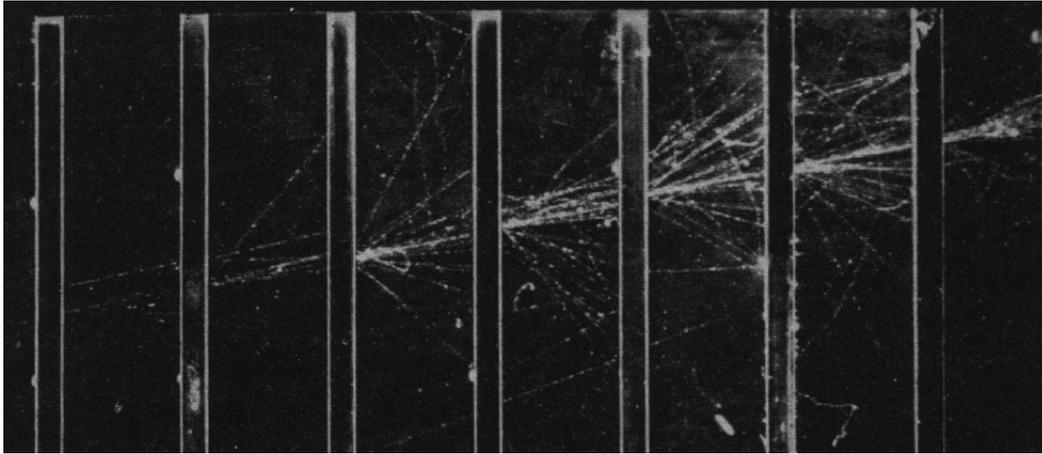


Figura 5.2: Interação de um elétron com um detetor de traços com absorvedores em chumbo.

A detecção de elétrons beneficia-se deste conhecimento básico para distingui-los de sua contaminação natural de jatos (e píons). Numa primeira fase, chamada de Extração de Características, avalia-se, para uma região do detetor, normalmente centralizada em um ponto pré-determinado, um conjunto de valores que representem, o melhor possível, os perfis de deposição lateral (ou radial) do objeto sendo estudado e sua penetração no aparato. Este processo comprime<sup>1</sup> a informação de entrada, que normalmente possui dimensionalidade bastante elevada ( $O(100)$ - $O(1000)$ ), para um conjunto pequeno ( $O(10)$ ) de variáveis que possam ser analisadas mais facilmente.

A questão da dimensionalidade da entrada está intimamente correlacionada com a capacidade discriminatória do calorímetro. Quanto mais granular, maior precisão pode ser obtida na definição de pontos de impacto e detecção de partículas. O revés é o custo - detetores multi-segmentados são mais difíceis e custosos de serem construídos.

No caso do experimento ATLAS (veja a Seção 3.4.2 para mais detalhes), os calorímetros são segmentados em diversas camadas, e, camada a camada, granularizados em células de deposição energética. A granularidade varia camada a camada, pois cada uma delas possui um objetivo distinto de emprego: as camadas iniciais têm

---

<sup>1</sup>Para os fins deste trabalho define-se “compressão” ou compactação a técnica pelo qual reduz-se a dimensionalidade de um sinal em um espaço bem definido, preservando ou tentando preservar a informação relevante que caracteriza este sinal de interesse. A transformação do espaço original para o espaço “comprimido” é irreversível, salvo onde especificado.

por objetivo a discriminação de elétrons e fótons e estimativa de posicionamento, as mais traseiras a detecção de hádrons. As camadas da seção e.m. têm profundidade variante, sendo que a segunda (EM2, a contar do zero, ou seja o pré-irradiador), tem a maior profundidade de todas, ocupando mais de 80% do volume total desta seção.

Para elétrons com valores energéticos até algumas dezenas de gigaelétron-volts, espera-se que a maior parte da energia do objeto esteja depositada nesta camada. A granularidade é bastante regular, tanto ao longo do eixo  $\eta$  quanto ao longo de  $\phi$ , como é possível observar na Tabela 3.1. Esta tabela também mostra o fator de compressão (ou tamanho da torre de filtragem)  $N_\eta \times N_\phi$  aplicado ao processo de agrupamento necessário ao pré-processamento antes da análise feita pelo LVL1.

A seção hadrônica do calorímetro é bastante menos granular. O tamanho padrão da célula ( $0,1 \times 0,1$  no espaço  $\eta \times \phi$ ) foi otimizado para detetar hádrons. Esta classe de partículas penetra mais profundamente na matéria e desenvolve complexas cascatas ao longo da penetração, normalmente pouco interessantes. Entretanto, devem estar contidas no espaço do detetor. O agrupamento padrão (torre de filtragem) para o LVL1 é de  $2 \times 2$  células.

A Figura 5.3 mostra histogramas da relação da deposição energética na seção hadrônica e a energia total do objeto considerando-se uma área  $0,4 \times 0,4$  no plano  $\eta \times \phi$ . Os eventos considerados são simulações de elétrons e jatos cobrindo um campo energético de algumas dezenas de GeV até 90 GeV interagindo com os calorímetros do ATLAS. Os eventos selecionados passaram uma simulação realística do LVL1 e portanto representam, todos, eventos que seriam classificados como elétrons por este nível de filtragem. Observa-se que para jatos há uma probabilidade maior de depósito na seção hadrônica que para elétrons.

Em seguida, aplica-se a discriminação da Física de interesse. Tipicamente [47, 50, 51], a detecção é realizada por meio da definição de patamares de separação (também ditos *cortes*) com informação *a priori*. Este trabalho é normalmente realizado por um físico experiente, de acordo com os canais físicos de maior interesse para o experimento.

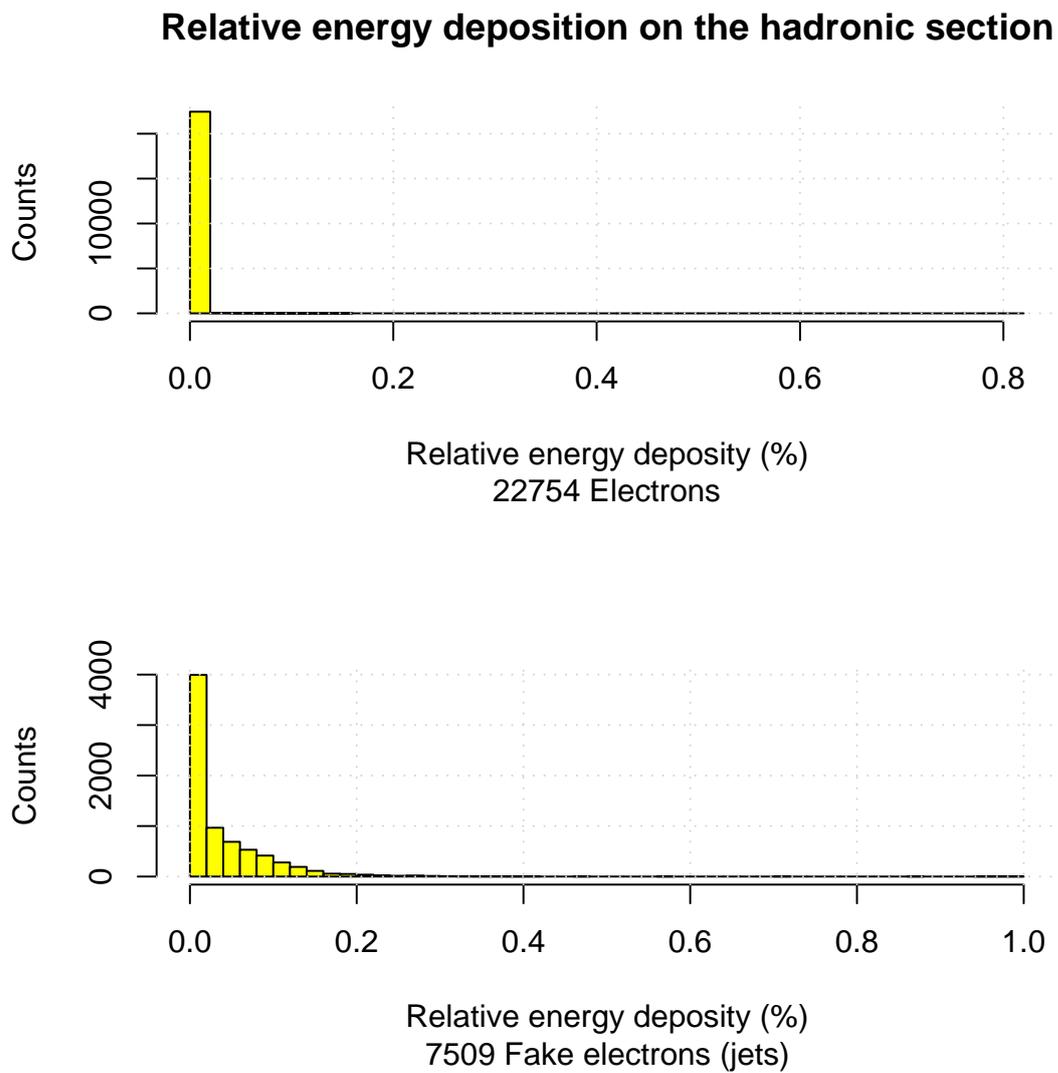


Figura 5.3: Exemplo da relação de deposição energética na seção hadrônica e energia total para elétrons (topo) e jatos (baixo).

## 5.4 Detecção de elétrons no LVL2

O LVL1 (Seção 4.1) executa seu procedimento de filtragem procurando elementos nos calorímetros e detetores de múons que obedecem a certos critérios de classificação. A Tabela 5.1, mostrada anteriormente, indica os principais objetos de interesse. No caso de objetos e.m. (elétrons e fótons), o algoritmo pode ser resumido da seguinte forma:

1. Os sinais analógicos provenientes das células do detetor são agrupados em macro-células chamadas Torres de Filtragem (TT). A taxa de agrupamento não é uniforme, mas, em geral, forma TT's com  $0,1 \times 0,1$  no plano  $\eta \times \phi$  para a seção eletromagnética do calorímetro e  $0,2 \times 0,2$  na seção hadrônica;
2. Os sinais somados das TT's são disponibilizados para a lógica de leitura do LVL1;
3. Processadores especializados (4 no total) cobrem diferentes áreas do detetor e utilizam um algoritmo de agrupamento, baseado em uma janela deslizante [29] cobrindo uma região de  $4 \times 4$  TT's (16 no total), aqui resumido:
  - (a) Se o núcleo de  $2 \times 2$  TT's contiver energia maior que a periferia, para um determinado patamar programável, um candidato a RoI é definido;
  - (b) Em seguida, somam-se, duas a duas, as energias transversas das TT's do núcleo de  $2 \times 2$  TT's e utiliza-se a maior das somas em uma comparação com um patamar de energia EM;
  - (c) Caso passe este critério, a energia na periferia da RoI-candidata é verificada para assegurar o isolamento do objeto;
  - (d) O último critério é o isolamento hadrônico, onde o LVL1 verifica se não há grande depósito de energia (transversa) na seção hadrônica.

No caso de atender a todos os pré-requisitos, o objeto candidato é promovido a uma RoI tipo e.m.. Este algoritmo está representado na Figura 5.4.

Os diversos sub-módulos do LVL1 irão detetar e pré-classificar os elementos de interesse para jatos, múons e energia faltante. Em seguida:

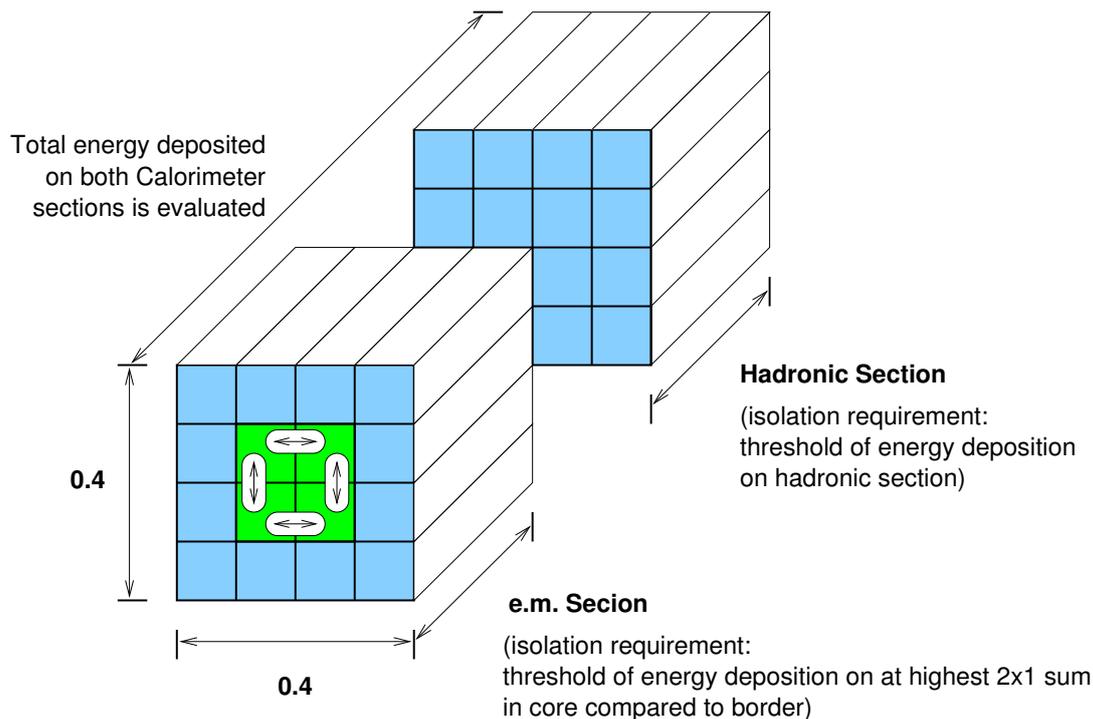


Figura 5.4: Representação gráfica do algoritmo de detecção de objetos e.m. no LVL1.

1. Os objetos detetados e seus pontos **centrais** de impacto estimados no plano  $\eta \times \phi$ , ou seja, as RoIs, são encaminhados ao Processador Central do Sistema de Filtragem do LVL1 (do inglês *Central Trigger Processor*, CTP) e uma decisão é feita levando-se em consideração as assinaturas de interesse e resultados dos processamentos de outros módulos (jatos, energia faltante e múons). A decisão de aprovar um evento é tomada baseando-se em RoI's de acordo com critérios de preferência estipulados *a priori*. As RoI's que são utilizadas para a decisão de aprovação são rotuladas “primárias” ao passo que as demais, “secundárias”;
2. Um sumário do resultado e das RoI's encontradas, já rotuladas (primárias e secundárias) é repassado, através do *RoI Builder* ao LVL2.

A primeira vez onde o evento será analisado tendo por base a granularidade máxima do detetor é no Segundo Nível de Filtragem do experimento ATLAS. O objetivo deste nível de filtragem, de uma forma geral, é a confirmação das informações do LVL1 e eventual depuração e extensão. Aproveitando-se de toda a granularidade disponibilizada pelo sistema de leitura do detetor, o LVL2 começa analisando o evento através das RoI's classificadas como primárias pelo LVL1.

Ao receber as informações das RoI's o *Steering* ativará o passo de decodificação do resultado do LVL1 e, em seguida, selecionará a próxima fase de análise, que dependerá diretamente do tipo das RoI's primárias do evento. No caso de RoI's do tipo e.m., um algoritmo de detecção de elétrons (e fótons) será ativado.

## 5.5 Detecção de elétrons no LVL2 do ATLAS

Como discutido anteriormente, na Seção 5.3, o processo de detecção veloz baseado em calorimetria pode ser sub-dividido nas seguintes partes:

1. Compactação do sinal de entrada;
2. Detecção da Física de interesse baseada no espaço comprimido.

Para o ATLAS, a fase de compactação irá mapear o espaço de entrada definido pelas células da RoI sendo tratada em 4 variáveis com alto poder discriminatório, baseado na percepção da Física de interesse e da contaminação esperada. No caso de elétrons, espera-se que haja forte contaminação de jatos com muitas componentes e.m.. Desta forma, a compactação da informação de entrada tentará realçar informações como a dispersão do objeto na RoI ou a profundidade de interação do objeto com as diversas camadas do detetor.

O nome do algoritmo usado para a FEx de RoI's primárias tipo e.m. é *T2Calo*, uma abreviação composta das palavras *Trigger*, LVL2 e Calorimetria. As variáveis extraídas do *cluster* definido pela RoI do LVL1, por este algoritmo, estão descritas a seguir. Cada uma destas variáveis define uma etapa ou mais etapas do processamento do T2Calo que está esquematizada nas figuras correspondentes:

1. **Energia do núcleo e.m.**, Figura 5.5: O primeiro passo do algoritmo é um refinamento do centro da RoI. Isto é feito encontrando-se o pico de deposição energética na segunda camada da seção eletromagnética do calorímetro, que é também a mais profunda desta seção (veja uma discussão mais detalhada da geometria do detetor na Seção 3.4.2). O valor do centro da célula é tomado como uma nova estimativa de posicionamento da RoI e será utilizado para os cálculos seguintes.

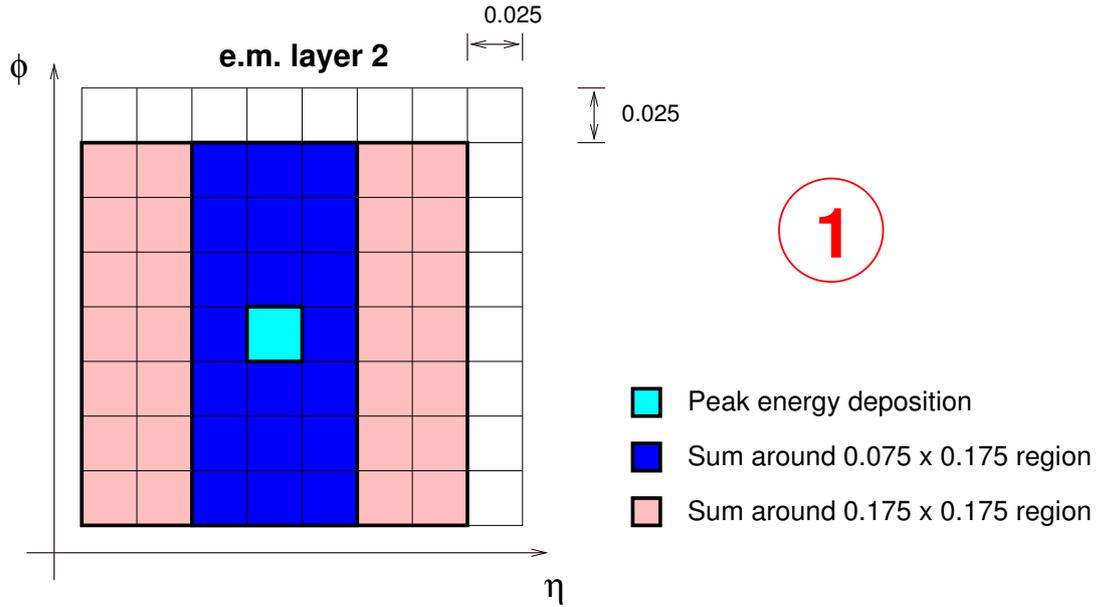


Figura 5.5: T2Calo, Etapa 1: Cálculo do centro refinado e deposições de energia na segunda camada e.m., para regiões de tamanho  $\Delta_\eta = 0,075 \times \Delta_\phi = 0,175$  e  $\Delta_\eta = 0,175 \times \Delta_\phi = 0,175$ .

Em seguida, ainda utilizando-se dos dados da segunda camada e.m., calcula-se a soma dos valores de deposição energética numa região de tamanho  $\Delta_\eta = 0,075 \times \Delta_\phi = 0,175$  e numa região de tamanho  $\Delta_\eta = 0,175 \times \Delta_\phi = 0,175$ , ao redor do centro de impacto encontrado anteriormente. De posse destes valores, a primeira característica será extraída:

$$R_{e.m.2}^{3 \times 7 / 7 \times 7} \text{ ou } R_{core} = \frac{\text{Energia}_{\Delta_\eta=0,075 \times \Delta_\phi=0,175}^{e.m.2}}{\text{Energia}_{\Delta_\eta=0,175 \times \Delta_\phi=0,175}^{e.m.2}} \quad (5.1)$$

Esta quantidade tenta estimar o espalhamento da cascata formada pelo decaimento do objeto na RoI. No caso do objeto ser um jato, espera-se que a cascata tenha um espalhamento maior que no caso de elétrons, resultando em um valor menor que 1 para esta quantidade;

2. **Máximos na primeira camada e.m.**, Figura 5.6: Os dois maiores picos de energia na primeira camada da seção e.m. (denominados  $E_1$  e  $E_2$  respectivamente) são detetados. De posse destes valores, a segunda característica é

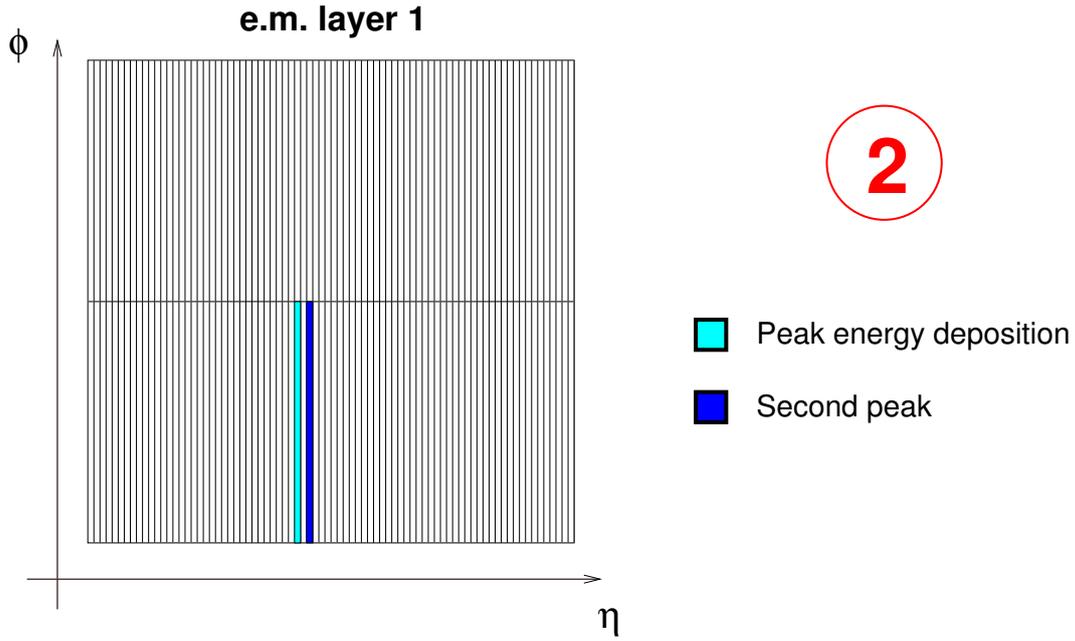


Figura 5.6: T2Calo, Etapa 2: Procura dos dois maiores picos na primeira camada e.m..

extraída:

$$R_{e.m.1}^{2\text{-máximos}} \text{ ou } E_{\text{ratio}} = \frac{E_1 - E_2}{E_1 + E_2} \quad (5.2)$$

Uma vez que jatos de partículas interagem de forma mais espalhada do que elétrons, espera-se que, na primeira camada e.m., sejam observados vários picos. Desta forma, se o objeto for um elétron (objeto único), espera-se que  $E_2 = 0$ , já que a cascata de partículas que o elétron forma na sua interação com o calorímetro é, tipicamente, bastante estreita. Assim sendo, para elétrons,  $R_{e.m.1}^{2\text{-máximos}}$  tenderá a 1. Para jatos, normalmente,  $R_{e.m.1}^{2\text{-máximos}}$  será menor que 1, já que  $E_2 \neq 0$ .

3. **Energia e.m.**, Figura 5.7: Os valores parciais de deposição energética para cada camada da seção e.m., numa região de  $\Delta_\eta = 0,075 \times \Delta_\phi = 0,175$  ao redor do pico de deposição energética na segunda camada, também são calculados. Destas somas parciais, a terceira característica da RoI é calculada somando-se estes 4 valores. O valor original calculado é dividido pelo  $\cosh |\eta|$  para



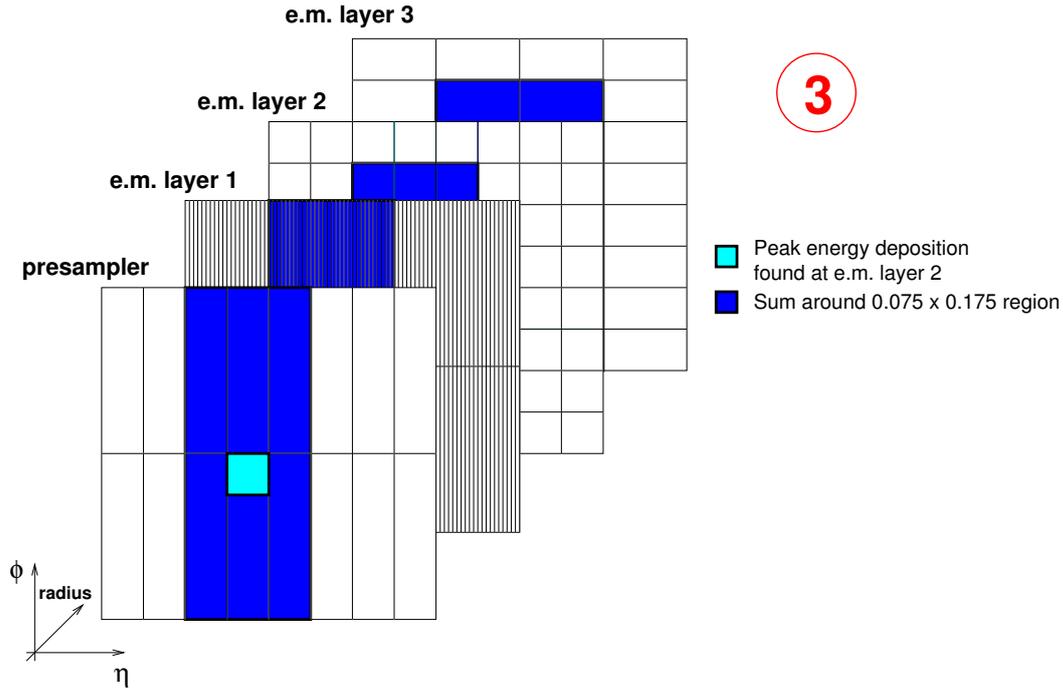


Figura 5.7: T2Calo, Etapa 3: Cálculo dos somatórios parciais de energia por camada, em uma área equivalente a  $\Delta_\eta = 0,075 \times \Delta_\phi = 0,175$ .

que obtenha-se a energia transversa<sup>2</sup> (projetada no plano  $x \times y$  do detector). Espera-se que a energia de elétrons esteja totalmente contida nesta janela; por sua vez, jatos devem apresentar uma fração menor de sua energia depositada nesta área;

4. **Energia Hadrônica** $_{\Delta_\eta=0,2 \times \Delta_\phi=0,2}$ : Calcula-se a energia total numa região  $\Delta_\eta = 0,2 \times \Delta_\phi = 0,2$ , centrado no pico de deposição energética na segunda camada e.m.. Este procedimento é executado para cada camada na seção hadrônica, e depois um somatório do total transversal é executado. Para elétrons, a quantidade de energia depositada na camada hadrônica deve ser próxima de zero, enquanto que, para jatos, espera-se que seja diferente de zero.

O algoritmo de detecção propriamente dito segue a execução do T2Calo e é chamado *EGammaHypo*. Este algoritmo tem a função exclusiva de combinar as informações disponibilizadas a partir do processo de extração de características e tomar uma decisão simples baseada nas propriedades físicas das partículas de interesse. A

<sup>2</sup>O termo “transverso” neste contexto refere-se ao feixe de partículas.

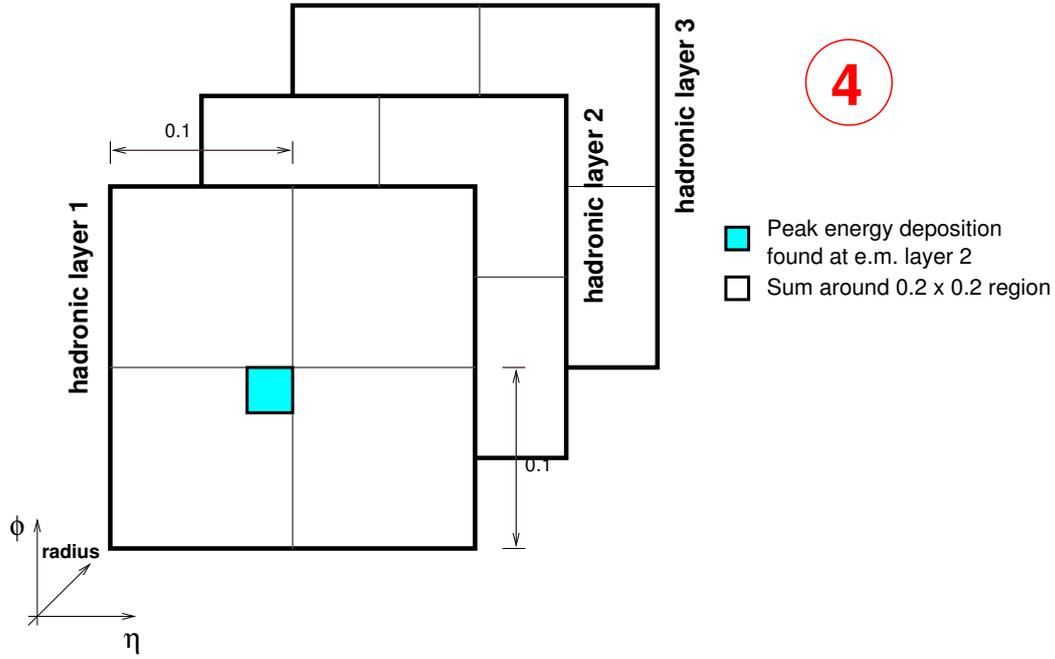


Figura 5.8: T2Calo, Etapa 4: Valores parciais e somatório das energias, em uma área equivalente a  $0,2 \times 0,2$  no plano  $\eta \times \phi$ , para as camadas da seção hadrônica.

seqüência discriminatória empregada pelo EGammaHypo está representada a seguir, em pseudo-código:

- ❶ **se**  $R_{e.m.2}^{3 \times 7 / 7 \times 7} \lesssim \text{corte}_{R_{e.m.2}^{3 \times 7 / 7 \times 7}}$ , **então**  
**retornar não é elétron;**
- ❷ **se**  $R_{e.m.1}^{2\text{-máximos}} \lesssim \text{corte}_{R_{e.m.1}^{2\text{-máximos}}}$ , **então**  
**retornar não é elétron;**
- ❸ **se**  $E_{T_{3 \times 7}}^{e.m.} \lesssim \text{corte}_{E_{T_{3 \times 7}}^{e.m.}}$ , **então**  
**retornar não é elétron;**
- ❹ **se**  $E_{T_{0,2 \times 0,2}}^{\text{HAD-1}} \gtrsim \text{corte}_{E_{T_{0,2 \times 0,2}}^{\text{HAD-1}}}$ , **então**  
**retornar não é elétron;**

Se o objeto definido pelo LVL1 conseguir passar todos os critérios definidos neste procedimento, é promovido a “candidato à elétron” dentro do LVL2. As fases seguintes de discriminação tentarão encontrar um traço, com momento equivalente a energia do *cluster*, e uma nova rodada de hipóteses confirmará (ou não) o possível elétron. A procura de traços nos detetores internos é um procedimento tipicamente

lento (na ordem de dezenas, dependendo centenas de milissegundos [16]), e deve ser evitada tanto quanto possível. Por outro lado, deseja-se que o sistema de detecção preliminar, baseado em calorimetria, seja o mais rápido possível, e mantenha bons níveis de discriminação, como já discutido.

Uma das estratégias em consideração atualmente [52], pondera sobre a utilização dos diversos estágios de hipótese de forma seqüencial, durante a extração de característica. Neste caso, após interagir com a segunda camada e.m., o T2Calo seria brevemente suspenso e a primeira das hipóteses verificada, criando recursos para uma rejeição mais antecipada. Em seguida, os dados da RoI para a primeira camada e.m. seriam carregados no processador, a segunda variável seria calculada e mais uma rodada de decisão seria tomada. As próximas etapas seriam a carga dos dados da terceira camada e.m. e do pré-irradiador, seguindo-se da análise de dados na seção hadrônica.

### 5.5.1 Ajuste fino para cada canal

Para cada uma das assinaturas de interesse aprovadas pelo LVL1, o segundo nível contará com algoritmos e sistemas de hipótese adequados àquela Física. É interessante notar que, mesmo para os canais similares, e.g. (veja a Tabela 5.1) EM25i e EM15i, os cortes impostos pelo algoritmo de hipótese serão otimizados para estas categorias de energia.

## 5.6 Caracterização do T2Calo e do EGammaHypo

Para a caracterização de operação do T2Calo e do algoritmo de hipótese associado, o EGammaHypo, os seguintes grupamentos de dados foram considerados [53, 54]:

- Cerca de 23.000 elétrons simulados via Monte Carlo para o *Data Challenge* 1, no CERN, proveniente de interações tipo  $H(130\text{GeV}) \rightarrow ZZ \rightarrow 4e^-$  e  $H(130\text{GeV}) \rightarrow ZZ \rightarrow 2e^- + 2\mu$ . Os elétrons de cada evento simulado interagem com diferentes regiões do detetor;
- Cerca de 250.000 jatos-duplos, simulados via Monte Carlo (*Pythia*) para o

*Data Challenge 1*, no CERN, com energia total, fixa em 25 GeV, também interagindo com várias partes do detetor.

Nos dois casos, considerou-se simulações com ruído, proveniente da eletrônica, e uma planta de detetor compatível com o que está sendo instalado atualmente na caverna do experimento, provendo um ambiente bastante acurado dentro do que é possível prever-se em termos de operação, mas ainda exequível para um ambiente de simulação.

Esta massa original de dados foi submetida a uma simulação do LVL1, que implementa o algoritmo discutido na Seção 5.4, aceitando todos os eventos com objetos tipo EM com as seguintes características<sup>3</sup>:

- ao menos 10 GeV numa região de  $2 \times 2$  torres de filtragem na seção e.m.;
- no máximo 4 GeV na região periférica ao núcleo de  $2 \times 2$  torres de filtragem;
- no máximo 2 GeV de depósito na seção hadrônica;
- qualquer valor para o centro em  $\phi$ ;
- $\eta$  variando de  $-2,5$  a  $+2,5$ ;

Um total de cerca de 22.000 elétrons e 7.000 fazem parte da massa de dados remanescente após a aplicação dos cortes acima. A Figura 5.9 contém os histogramas de deposição energética total transversa de elétrons e jatos, em uma janela de  $0,4 \times 0,4$  no plano  $\eta \times \phi$ , considerando-se as duas seções do calorímetro (e.m. e hadrônica). É possível distinguir um corte acentuado nas redondezas de aproximadamente 10 GeV, que equivale à ação da simulação do LVL1, como esperado. Para o histograma de jatos a média aproxima-se ao valor esperado de 25 GeV, exibindo um pico pronunciado ao redor deste valor. Como observa-se, grande parte dos jatos encontram-se com valor de energia transversa entre 10 e 40 GeV, enquanto que para elétrons, a distribuição decai suavemente até  $\approx 91$  GeV (aproximadamente a massa de repouso do bóson Z).

As Figuras 5.10 e 5.11 mostram histogramas da energia transversa total por seção de calorimetria para elétrons e jatos. Nota-se que, para elétrons, quase 100%

---

<sup>3</sup>Os cortes utilizados neste trabalho são “relaxados”, no sentido que deixariam passar mais candidatos a elétrons que os cortes tipo e25i ou, ainda, e30i.

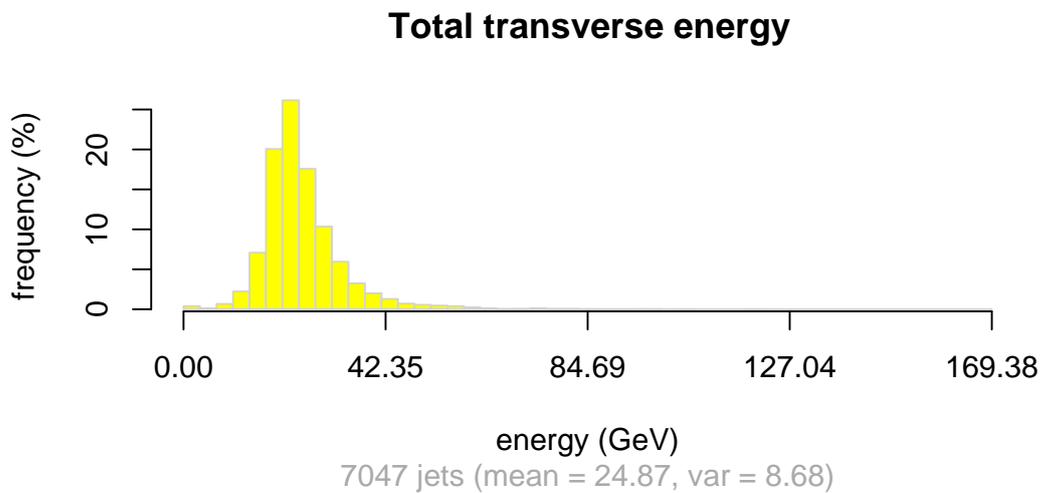
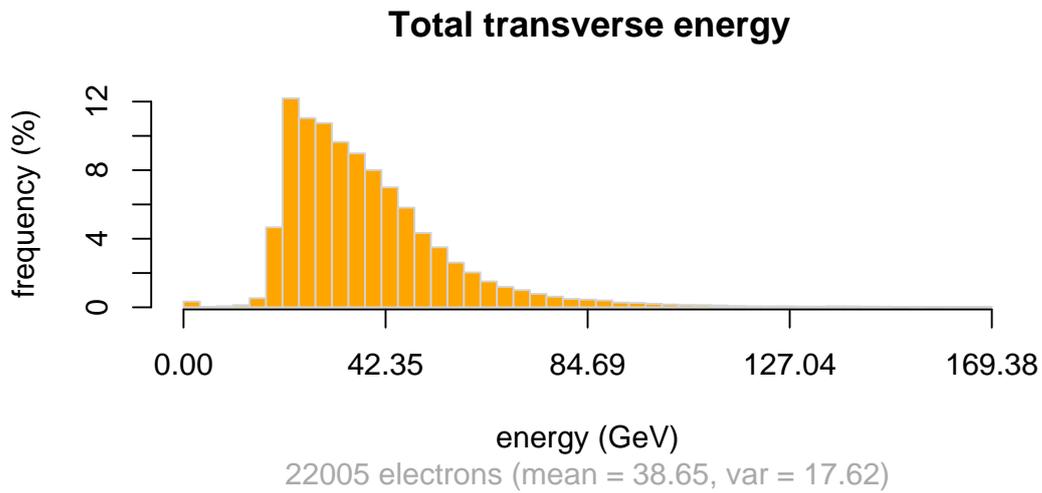


Figura 5.9: A deposição total de energia em uma região  $0,4 \times 0,4$  no plano  $\eta \times \phi$  para elétrons (em cima) e jatos (em baixo) para a massa de dados disponível para o estudo.

da energia total do objeto é retida na seção e.m., enquanto que, para jatos, observa-se algum vazamento de energia na seção hadrônica. Embora o corte realizado pelo LVL1 tenha sido em 2 GeV para o vazamento de energia nesta seção, observa-se uma quantidade não desprezível de eventos com energia hadrônica além deste valor. Por outro lado, a energia total nesta seção decai rapidamente e assume-se que a existência desta cauda esteja relacionada à (baixa) qualidade da calibração dos dados utilizada no LVL1.

As Figuras 5.12 e 5.13 mostram, respectivamente, a fração  $R_{e.m.2}^{3 \times 7/7 \times 7}$  e  $R_{e.m.1}^{2\text{-máximos}}$ , tal como utilizada pelo EGammaHypo para definir a eficiência de detecção de elétrons e jatos. Na Figura 5.12, observa-se que a distribuição para elétrons tem média bastante próxima a 1 e baixíssima variância. Para jatos, a média é mais baixa e a distribuição apresenta longa cauda em direção à origem. No caso da variável  $R_{e.m.1}^{2\text{-máximos}}$ , que tem por objetivo detetar picos de deposição energética na primeira camada e.m., para elétrons observa-se um pico dominante em 1, ao passo que, para jatos, há uma distribuição mais uniforme, indicando uma separabilidade linear.

As Figuras 5.14 e 5.15 mostram histogramas para elétrons e jatos dos centros refinados pelo T2Calo das RoI's em questão, tanto em relação à variável  $\eta$  quanto à variável  $\phi$ . É possível distinguir uma uniformidade na distribuição em  $\phi$  e uma tendência a concentração de eventos nas proximidades de  $\eta = 0$ , como é aguardado no experimento.

Exceto por eventuais problemas na lógica de simulação (Monte Carlo) e filtragem (pelo LVL1), estima-se que a massa de dados seja representativa do problema da separação entre jatos e elétrons. Deve ser levado em consideração que, para jatos, o pico ao redor de 25 GeV poderá introduzir imprecisões na definição da capacidade de detecção de elétrons bastante energéticos, presentes na massa de estudo. Sempre que possível, tentar-se-á levar estas restrições em consideração.

### 5.6.1 Detecção de elétrons com o EGammaHypo e sua otimização

Para que a eficiência de detecção do algoritmo proposto pelo EGammaHypo seja estimada, deve-se passar a massa de dados de estudo por um processo de otimização que ajude o especialista a selecionar os patamares de corte que definem o algoritmo,

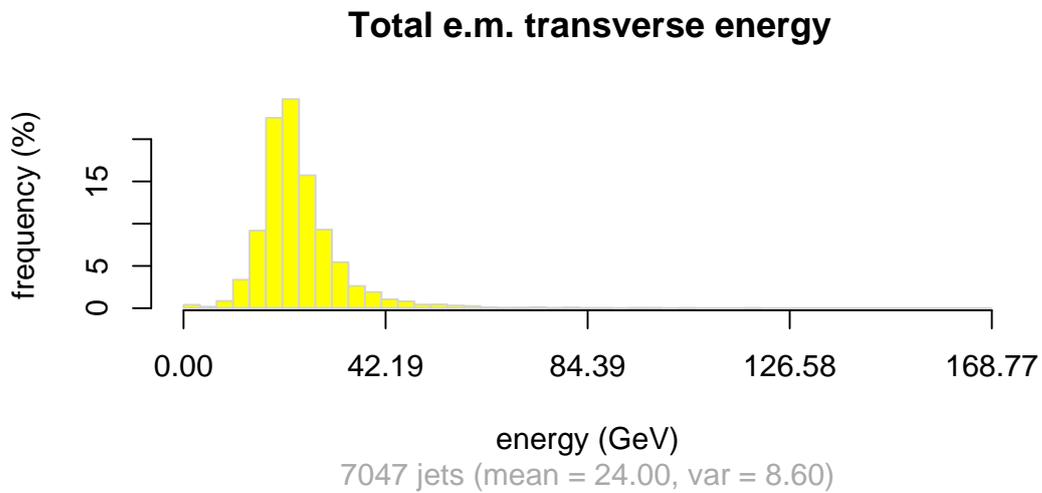
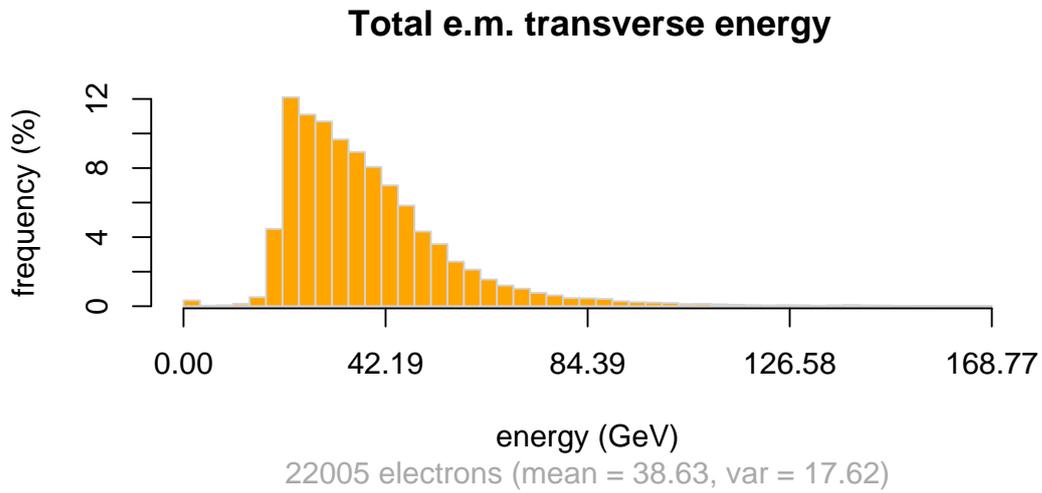


Figura 5.10: A deposição total de energia na seção e.m. em uma região  $0,4 \times 0,4$  no plano  $\eta \times \phi$  para elétrons (em cima) e jatos (em baixo) para a massa de dados disponível para o estudo. As contagens estão normalizadas.

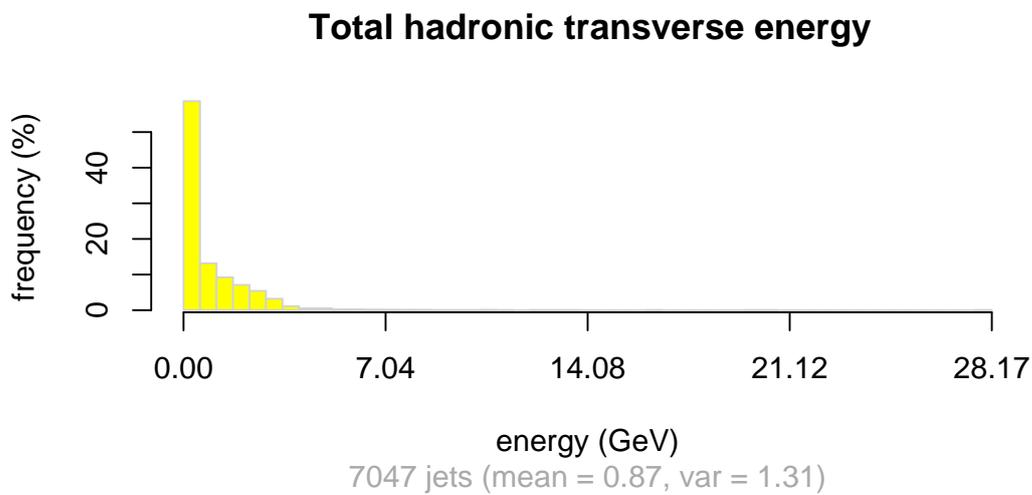
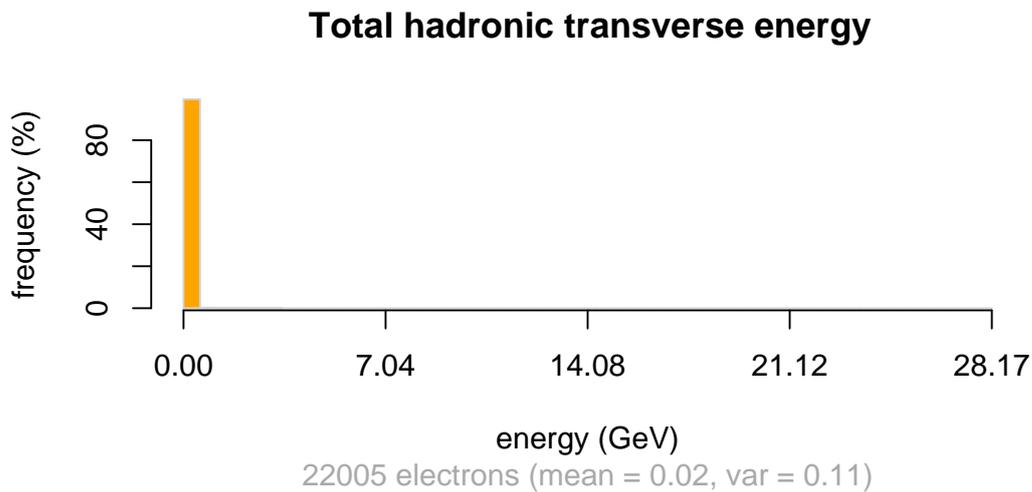


Figura 5.11: A deposição total de energia na seção hadrônica em uma região  $0,4 \times 0,4$  no plano  $\eta \times \phi$  para elétrons (em cima) e jatos (em baixo) para a massa de dados disponível para o estudo. As contagens estão normalizadas.



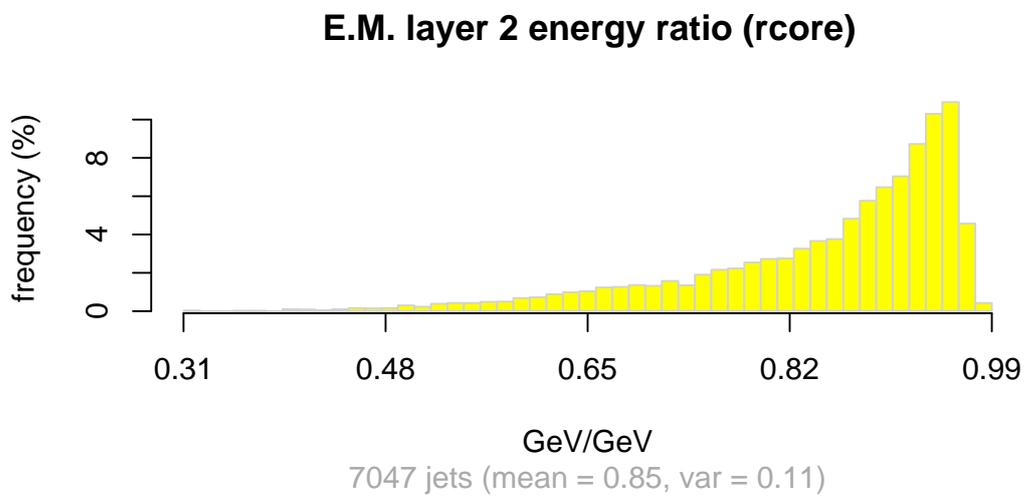
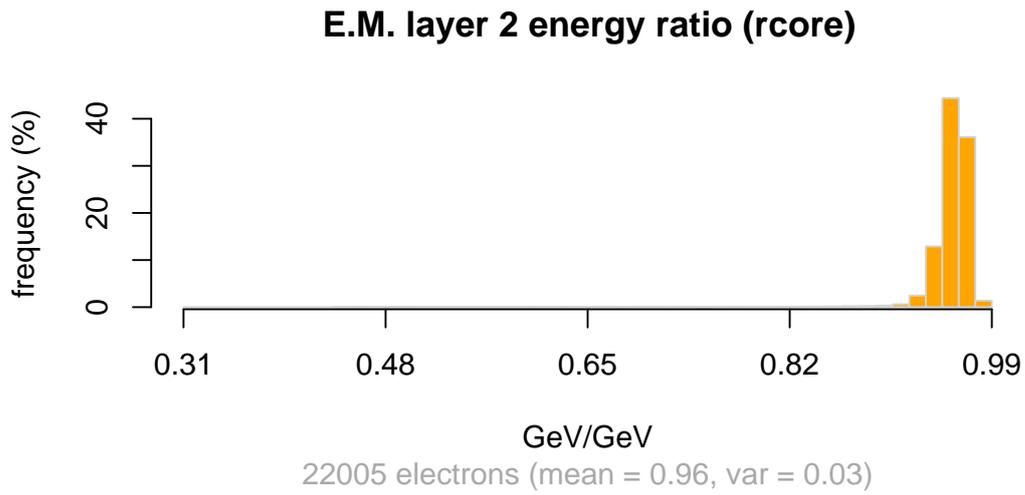


Figura 5.12: Distribuição da variável  $R_{e.m.2}^{3 \times 7 / 7 \times 7}$  para elétrons (em cima) e jatos (embaixo). As contagens estão normalizadas.

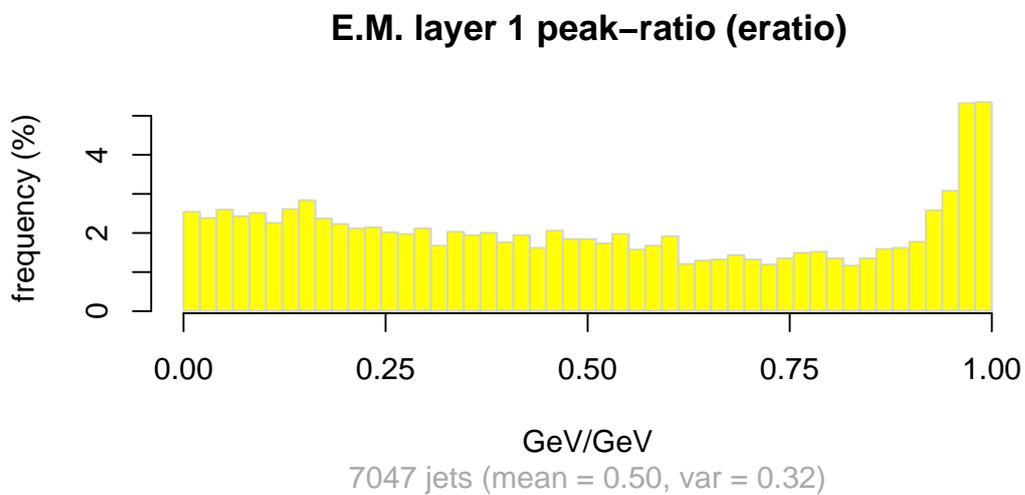
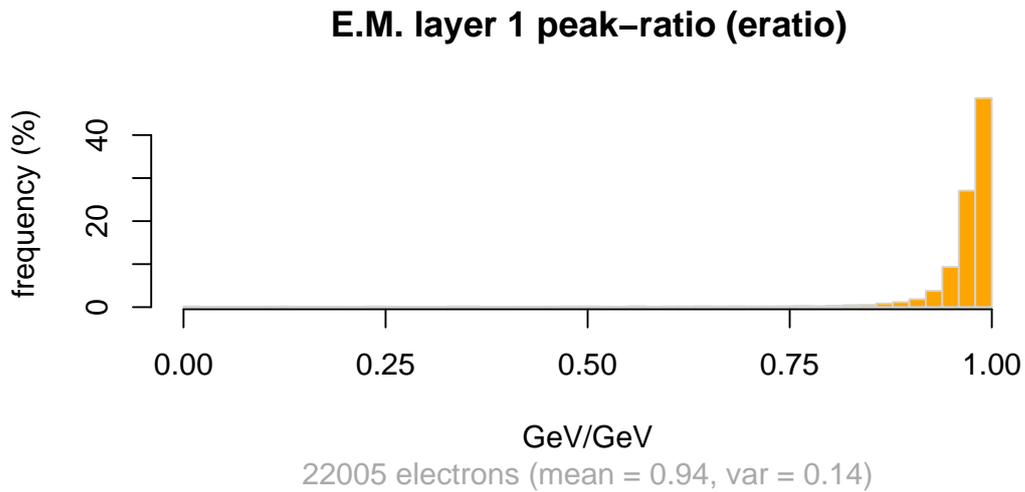
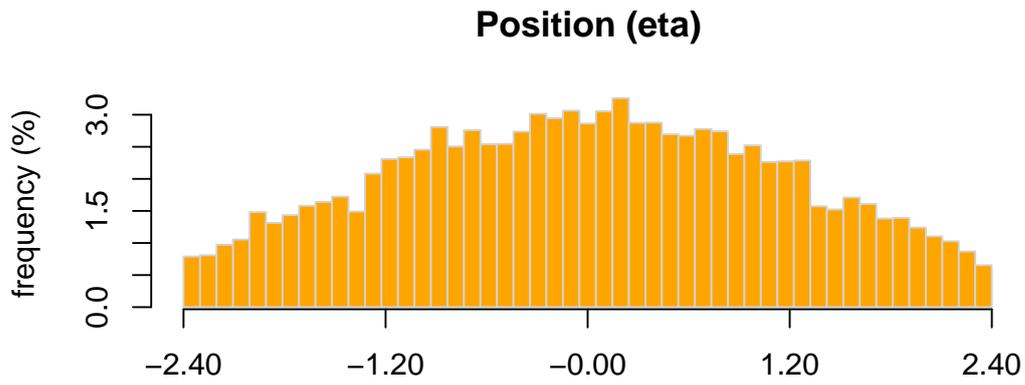
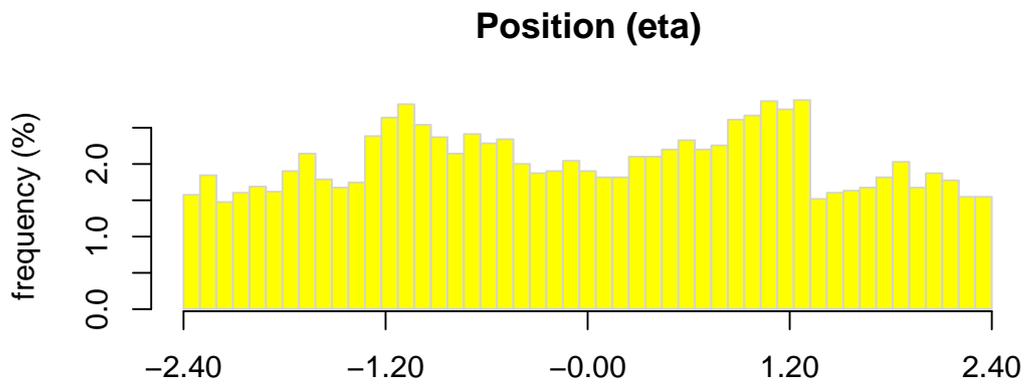


Figura 5.13: Distribuição da variável  $R_{e.m.1}^{2\text{-máximos}}$  para elétrons (em cima) e jatos (em-baixo). As contagens estão normalizadas.

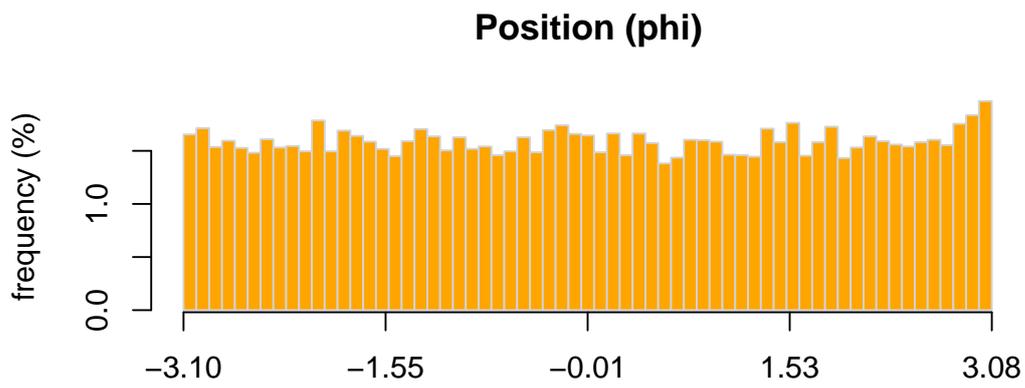


22005 electrons (mean = -0.00, var = 1.14)

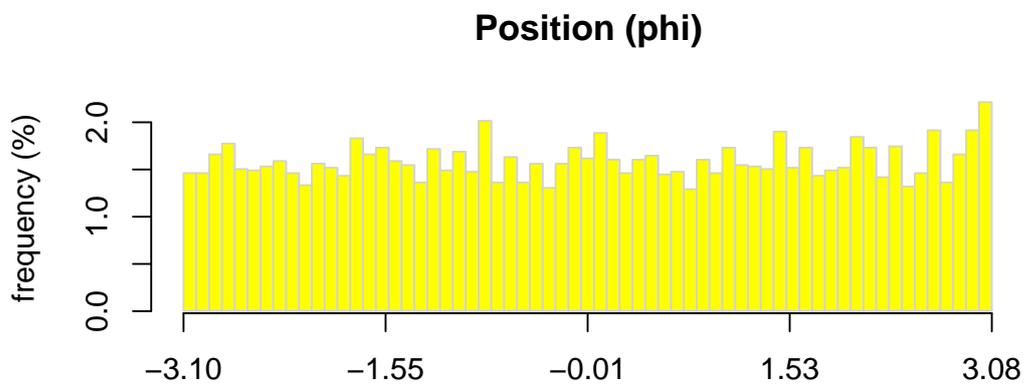


7047 jets (mean = 0.01, var = 1.32)

Figura 5.14: Distribuição em  $\eta$  dos centros refinados das RoI's para elétrons (em cima) e jatos (embaixo). As contagens estão normalizadas.



22005 electrons (mean = 0.01, var = 1.80)



7047 jets (mean = 0.03, var = 1.80)

Figura 5.15: Distribuição em  $\phi$  dos centros refinados das RoI's para elétrons (em cima) e jatos (embaixo). As contagens estão normalizadas.

como exposto na Seção 5.5. Para tal, dividiu-se o conjunto de dados em 2 metades com aproximadamente o mesmo número de RoI's. A primeira metade será utilizada para o “treinamento” ou otimização dos cortes, enquanto que a segunda será utilizada para o teste dos cortes de forma que seja possível testar se, para uma massa de dados jamais apresentada ao sistema, a eficiência de detecção está de acordo com aquela encontrada durante a determinação dos cortes. Desta forma, saberemos se o detetor está demasiado especializado nos dados presentes no conjunto de treinamento ou pode ser utilizado *genericamente* para a detecção de elétrons e jatos.

O algoritmo de otimização não será exaustivo, varrendo todo o espaço de dados, mas partirá de patamares pré-fixados em valores intuitivos e definirá uma sub-área de variação por onde testará exaustivamente as combinações dos 4 cortes necessários, como é normalmente feito atualmente. Eis aqui os patamares e passos de busca (ou *possibilidades*) que utilizaremos:

1. Corte em  $R_{e,m,2}^{3 \times 7 / 7 \times 7}$ : de 0,6 a 1,0, em passos de 0,01 (41 possibilidades);
2. Corte em  $R_{e,m,1}^{2\text{-máximos}}$ : de 0,6 a 1,0, em passos de 0,01 (41 possibilidades);
3. Corte em  $E_{T_{3 \times 7}}^{e.m.}$ : de 5000 a 30000 MeV em passos de 500 MeV (51 possibilidades);
4. Corte em  $E_{T_{0,2 \times 0,2}}^{\text{HAD-1}}$ : de 0 a 4000 MeV em passos de 500 MeV (9 possibilidades).

Para cada corte, a totalidade da massa de dados de treinamento do método é avaliada e as RoI's que *sobrevivem* ao corte são levadas à próxima etapa. As eficiências relativas e a quantidade de dados utilizadas para o teste da fase seguinte são acumulados para posterior análise. Levando-se em consideração o número de combinações dadas as possibilidades do problema de otimização e considerando-se que a superfície de erro não possua um único mínimo, há de se tentar  $41 \times 41 \times 51 \times 9 = 771579$  diferentes combinações, neste caso, antes de qualquer conclusão.

Está claro que uma varredura completa do espaço de parâmetros, i.e. buscando-se todo o espaço de possibilidades ao invés dos sub-conjunto utilizados, estaria fora de questão para uma utilização em condições que possam variar em questão de horas, como é o caso do experimento ATLAS. Desta forma, seria necessária uma antecipação para os cortes ou uma redução do espaço de busca, permitindo uma

otimização mais rápida. A utilização de programas especialmente codificados para a tarefa também poderia diminuir o tempo de teste, tornando o método mais atraente para uma utilização prática.

A Figura 5.16 mostra os resultados da busca exaustiva definida anteriormente. Esta figura de mérito, normalmente chamada de *Característica de Operação do Receptor* (do inglês, *Receiver Operating Characteristics*, ROC [55]), contém os resultados de cada ponto de operação do algoritmo EGammaHypo. O eixo vertical denota a eficiência na detecção do sinal de interesse (elétrons), enquanto o eixo horizontal, a taxa de falso-alarme (erro em jatos) naquele ponto de operação. A taxa de falso-alarme é escalonada em 25 kHz, que é a taxa máxima de ruído de fundo esperada. O valor do eixo horizontal portanto, denota a quantidade de jatos que serão aprovados pelo LVL2, como elétrons e, portanto, representa uma medida direta da taxa de eventos descarregados no Filtro de Eventos.

Observa-se que, para o conjunto de intervalos testado, a massa de resultados assume uma forma angular. Ao redor do ponto de flexão, na parte exterior da massa de resultados, encontraremos as melhores relações de eficiência *versus* falso-alarme para este discriminador. O ponto destacado nesta figura representa o conjunto de cortes que maximiza a multiplicação da soma das eficiências de detecção pelo produto entre elas, ou seja:

$$SP = (\text{efic.}_{\text{classe}_1} + \text{efic.}_{\text{classe}_2}) \times (\text{efic.}_{\text{classe}_1} \times \text{efic.}_{\text{classe}_2}) \quad (5.3)$$

O produto SP de um detetor para 2 classes de eventos apresenta um máximo ( $= 2, 0$ ) quando a eficiência na detecção de ambas as classes é máximo (i.e.  $= 1, 0$ ), e um mínimo em 0 quando a eficiência de detecção de **qualquer** uma das duas classes de eventos é 0. É interessante notar que, para um determinado classificador, o valor do máximo da Equação 5.3 indicará o ponto de inflexão da curva R.O.C.. Este ponto determina a melhor relação de detecção entre as duas classes de eventos para o classificador.

À medida que a curva R.O.C. se aproxima, em detrimento de uma melhora na eficiência de detecção, do eixo vertical para  $\text{Prob.}_{\text{falso-alarme}} = 0$  e do eixo horizontal para  $\text{Prob.}_{\text{detecção}} = 1, 0$ , o máximo da Equação 5.3 aumentará. Por esta razão utilizar-se-á o valor máximo produto SP de um detetor como uma figura de mérito

de sua eficiência de classificação.

No caso específico em análise, o ponto onde produto  $SP$  atinge seu máximo determina uma eficiência de detecção de elétrons de 91,85% contra 10,19% de falso alarme na detecção de jatos, ou 2,55 kHz, levando-se em consideração a contaminação do canal de elétrons na saída do LVL1. O valor do produto  $SP$  no ponto em questão é de 1,50.

Os seguintes valores de corte determinam o detector marcado na Figura 5.16:

1.  $R_{e.m.2}^{3 \times 7 / 7 \times 7}$ : 0,93;
2.  $R_{e.m.1}^{2\text{-máximos}}$ : 0,82;
3.  $E_{T_{3 \times 7}}^{e.m.}$ : 14.500 MeV;
4.  $E_{T_{0,2 \times 0,2}}^{HAD-1}$ : 500 MeV.

Aplicando-se estes cortes ao conjunto de teste, de forma análoga, obtém-se 91,62 % de eficiência na detecção de elétrons contra 10,45 % de falso-alarme ( $SP = 1,49$ ). Este resultado indica também que os dois sub-conjuntos de dados (treino e teste) são estatisticamente semelhantes. Durante a busca exaustiva deste resultado, acumularam-se os resultados de eficiência e falso-alarme parciais entre operações de corte. Estes valores podem ser vistos na Tabela 5.2. Como é possível avaliar, a partir desta tabela, as variáveis  $R_{e.m.2}^{3 \times 7 / 7 \times 7}$  e  $R_{e.m.1}^{2\text{-máximos}}$  são as mais discriminatórias, deixando passar apenas cerca de 30 % e 45 % dos jatos avaliados, respectivamente. O corte em  $E_{T_{3 \times 7}}^{e.m.}$  não reduz a taxa de eventos, sendo praticamente irrelevante a este processo de discriminação. Este comportamento é esperado, para a massa de dados sendo avaliada, uma vez que os jatos disponíveis tem energia fixa em 25 GeV e que elétrons estão espalhados no espectro de energia. O vazamento de energia na seção hadrônica (último corte), ainda conseguirá diminuir a taxa de falso-alarme na saída do detector.

As Figuras 5.17, 5.18 e 5.19 mostram os valores parciais de eficiência e falso-alarme considerando-se uma divisão dos dados por  $\eta$ ,  $\phi$  e por energia total transversa na seção e.m. dos calorímetros. No caso da varredura em  $\eta$ , é possível notar que a eficiência de detecção do método apresenta uma queda abrupta próximo ao vão entre a seção do barril e da tampa ( $\eta \approx 1,5$ ), o que é esperado devido a ausência de

### EGammaHypo parameter optimization (train set)

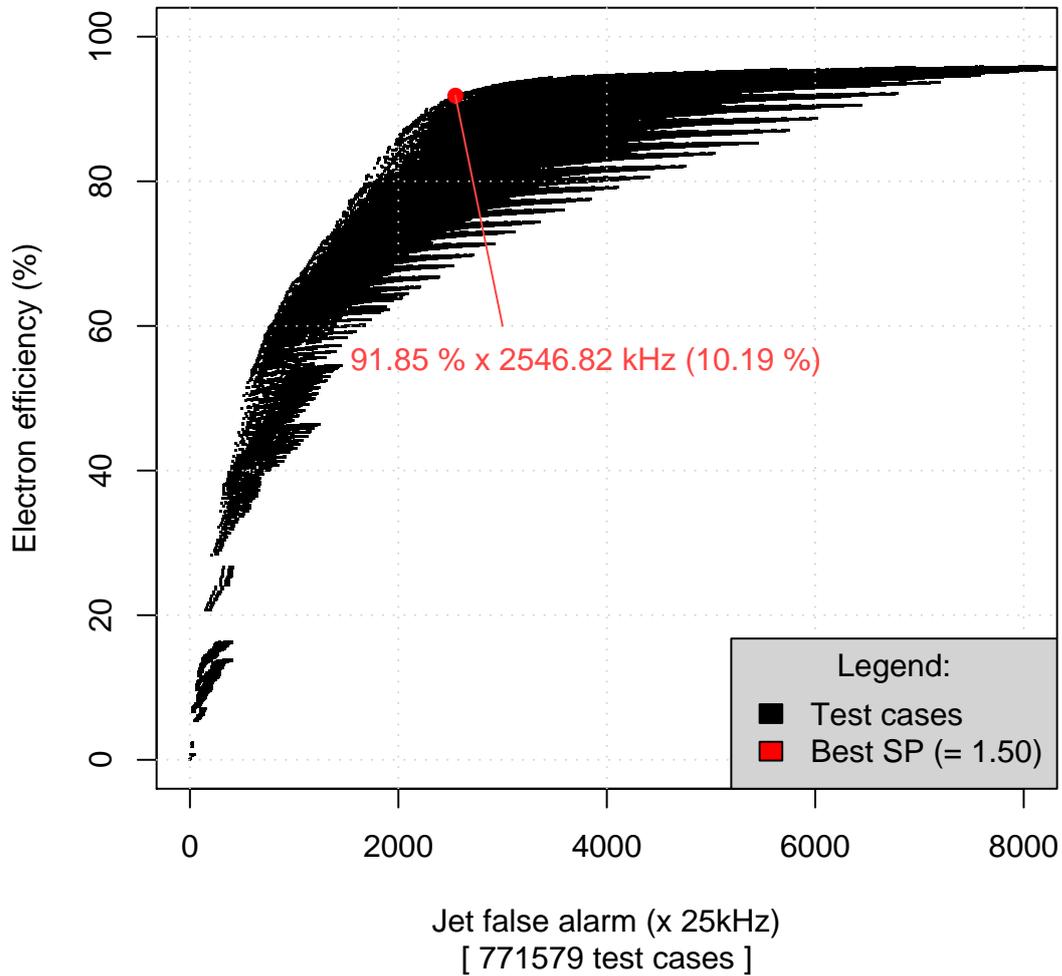


Figura 5.16: A curva ROC para 771.579 combinações de valores de corte para o algoritmo EGammaHypo.

Tabela 5.2: Valores parciais de detecção e falso-alarme para o detetor baseado no algoritmo EGammaHypo.

Ordem	Variável	Eficiência (%)	Falso-alarme (%)
1	$R_{e.m.2}^{3 \times 7 / 7 \times 7}$	96,4	30,5
2	$R_{e.m.1}^{2\text{-máximos}}$	95,7	43,6
3	$E_{T_{3 \times 7}}^{e.m.}$	99,9	<b>98,9</b>
4	$E_{T_{0,2 \times 0,2}}^{HAD-1}$	91.8	77.4



elementos de detecção nesta área. Para a varredura em  $\phi$ , observa-se uma distribuição bastante homogênea para a eficiência de detecção de elétrons e jatos, apesar da baixa estatística para jatos em alguns intervalos em  $\phi$  (muitos canais possuem apenas 30 a 40 jatos, como é possível observar na Figura 5.9). A varredura por  $E_{T_{3\times 7}}^{e.m.}$  indica que o método é bastante robusto na detecção de elétrons, aumentando sua eficiência suavemente com o valor de energia da RoI. Este comportamento é esperado pois sabe-se que a resolução em energia de um calorímetro aumenta com a energia do objeto [44] e, portanto, a capacidade discriminante do sistema. Por outro lado, jatos mais energéticos tendem a penetrar ainda mais no calorímetro, fazendo com que a detecção de elétrons se torne mais fácil. O falso-alarme em jatos também aumenta com a energia do objeto analisado, embora seja difícil estimar com precisão a qualidade do valor de falso-alarme determinado, uma vez que a estatística para jatos acima de 50 GeV é praticamente inexistente.

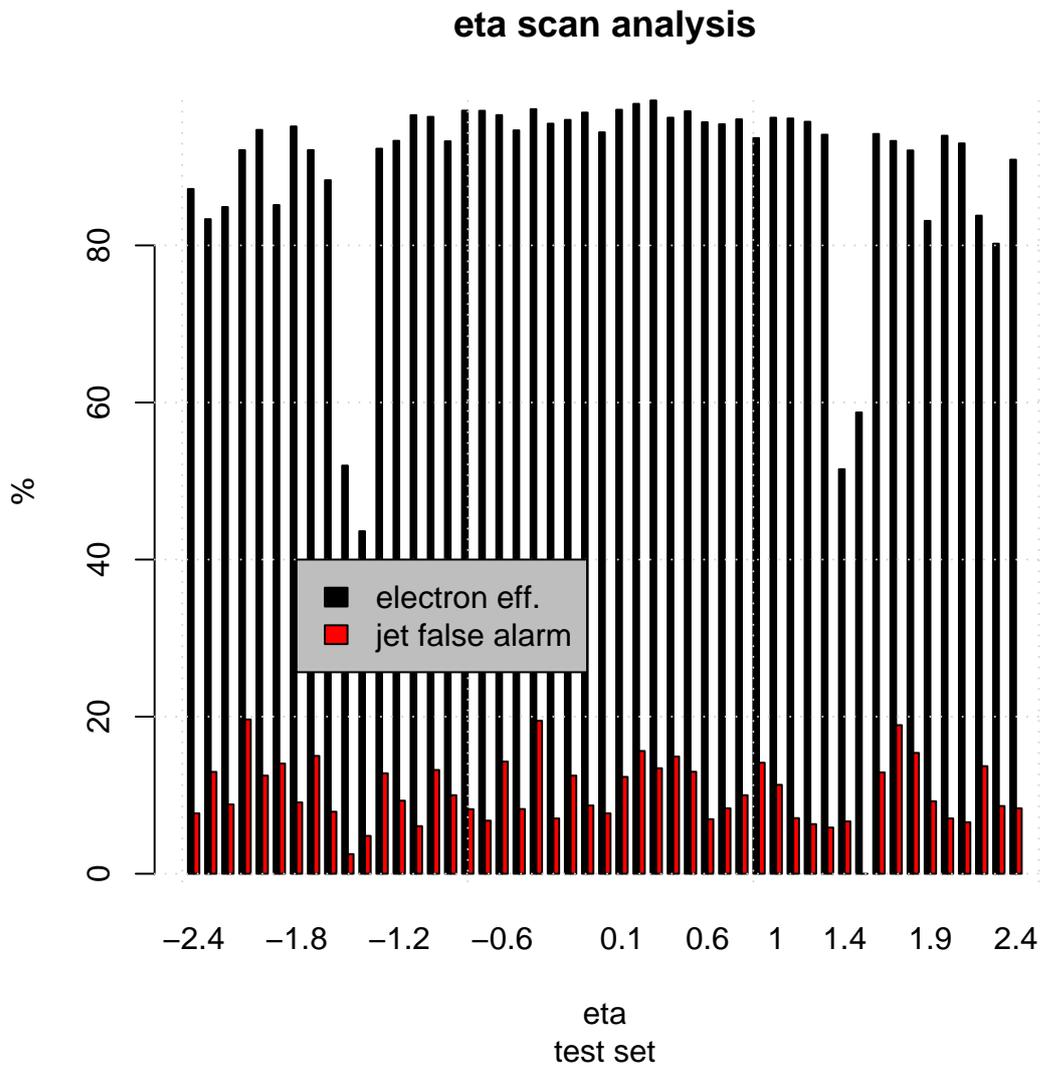


Figura 5.17: Eficiência de detecção contra falso-alarmede por  $\eta$  para o conjunto de testes utilizando o algoritmo EGammaHypo.

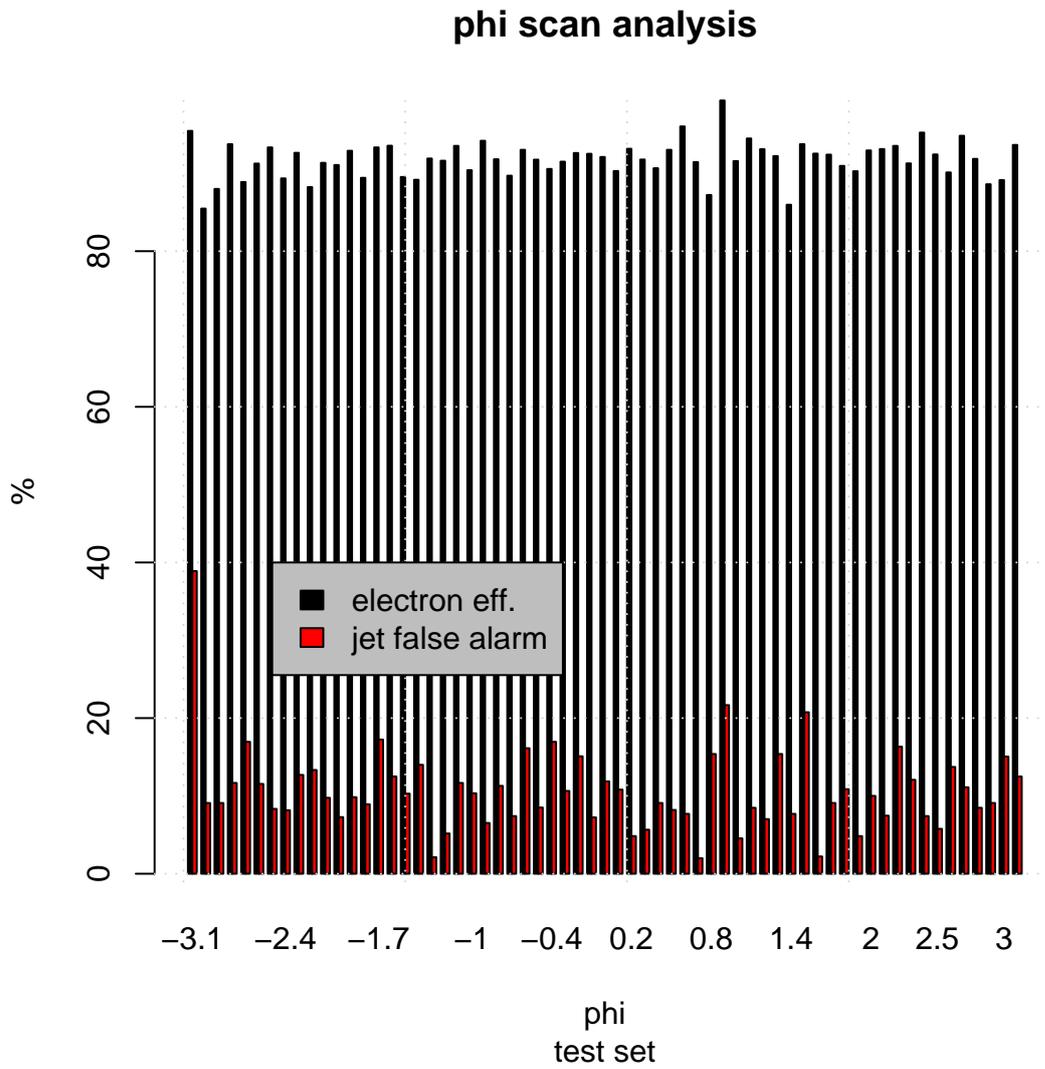


Figura 5.18: Eficiência de detecção contra falso-alarme por  $\phi$  para o conjunto de testes utilizando o algoritmo EGammaHypo.

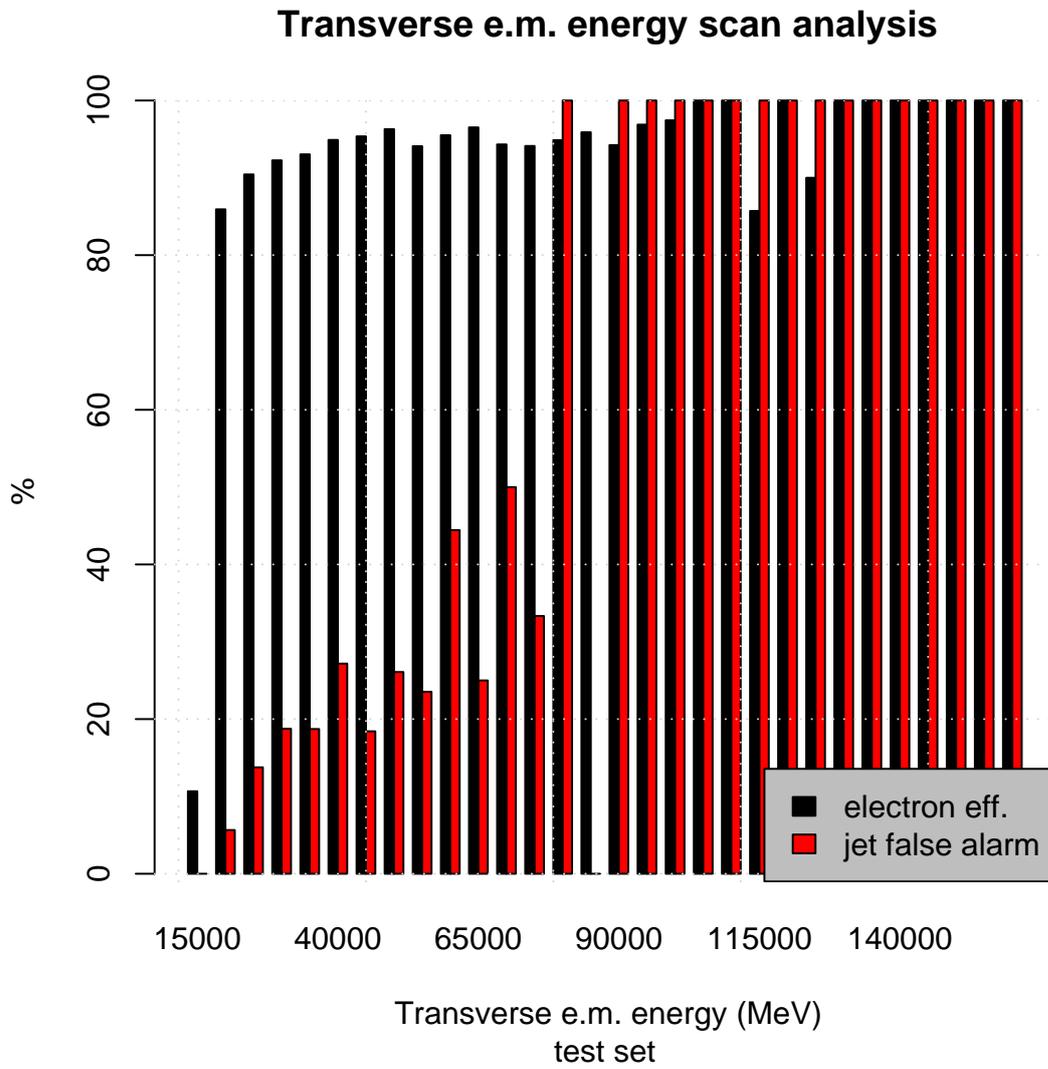


Figura 5.19: Eficiência de detecção contra falso-alarme por  $E_{T_{3 \times 7}}^{e.m.}$  para o conjunto de testes utilizando o algoritmo EGammaHypo.

# Capítulo 6

## Análise neural baseada em calorimetria para a detecção elétron/jato

Neste capítulo desenvolvem-se sistemas de classificação de simples treinamento e emprego, baseados em redes neurais artificiais. Estes classificadores definem um patamar de detecção melhor que o sistema discutido no capítulo anterior, superando-o em robustez e flexibilidade e ainda atingindo uma velocidade de execução compatível com a operação dentro do Segundo Nível do Sistema de Filtragem do experimento ATLAS.

### 6.1 Revisão da literatura

Dentre os métodos modernos de análise multi-variável mais empregados em Física de Altas Energia, Redes Neurais Artificiais possuem um papel importante ao lado de outras técnicas também eficazes como a Análise de Componentes Principais e classificadores bayesianos. Nesta seção, conduz-se uma revisão da literatura deste tipo de análise, empregada atualmente neste campo da física.

Os primeiros sistemas de detecção baseados em análise neural, aplicados à Física de altas energias, surgiram em 1987, com o emprego de redes neurais artificiais na reconstrução de partículas carregadas e píons neutros, como está bem resumido em [56]. Hoje em dia [57], cerca de 20 anos passados, tais sistemas são empregados

em praticamente todas as áreas neste campo, desde a concepção de detetores [58], filtragem e aquisição de dados [59, 60, 61], passando pela reconstrução *offline* [62] e análise [63, 64]. Nestes ambientes, o reconhecimento de padrões pode ocorrer tanto num plano mais elementar, onde deseja-se identificar os parâmetros de cada partícula que surge de uma reação sub-atômica, quanto num plano mais avançado, onde combinam-se as informações detetadas para se tentar definir as características globais de um evento.

### 6.1.1 Métodos neurais para análise *offline*

Em análise, o objetivo de classificadores em Física de Altas Energias é o de destacar o sinal de interesse, normalmente inserido em uma massa de eventos ordinários de proporções colossais. Em [65] encontra-se a descrição de um sistema nestas características: A produção de pares de bósons  $W$  pode ser observada através das colisões elétron-pósitron providas pelo LEP e detetadas no experimento DELPHI, no CERN. Os pares  $W^+W^-$  decaem rapidamente em quarks segundo a equação  $W^+W^- \rightarrow q\bar{q}q\bar{q}$ . Os quarks, por sua vez, são detetados por decaimento em jatos com fortes componentes hadrônicas, pelos calorímetros do experimento.

Este tipo de processo, característico das observações do DELPHI, estão imersos em uma massa de dados contendo outros tipos de evento que se assemelham à geração de pares  $W$ . Dentre eles, os mais importantes são  $e^+e^- \rightarrow Z^0/\gamma^* \rightarrow q\bar{q}$  e  $e^+e^- \rightarrow ZZ \rightarrow q\bar{q}q\bar{q}$ , sendo o segundo tipo, bastante difícil de detetar, já que decai no mesmo tipo de fenômeno que o sistema que se deseja discriminar. Para executar esta separação, definiram-se quatro variáveis de interesse: a energia no centro de massa da reação detetada, o número de jatos encontrados, a quantidade de traços por jato e uma variável combinada que depende dos ângulos e energias entre os jatos detetados. Estas quatro variáveis são submetidas a um processo de cortes sucessivos para que se elimine o maior número possível eventos-ruído, ainda que mantendo-se um máximo de eventos de interesse.

Para melhorar a taxa de separação, utiliza-se uma rede neural com 13 entradas equivalentes às primitivas utilizadas para encontrar as 4 variáveis utilizadas para os cortes sucessivos. As 13 entradas são encaminhadas à uma rede neural, sem realimentação, com 7 neurônios escondidos e apenas um neurônio na camada de saída.

O sistema é treinado com eventos simulados (*Monte Carlo*) e, finalmente, empregado para a detecção de pares  $W$ . O resultado final mostra uma melhora significativa na detecção deste fenômeno, apontando para uma eficiência de 88,74% contra 85,58% do método anterior, para um falso-alarme cerca de 15% menor.

Em [66], sugere-se que utilização de redes neurais possa aumentar o potencial de descobrimento de bósons de Higgs no experimento D0, uma vez que possa aumentar a relação sinal/ruído observada. Neste caso, deseja-se observar bósons de Higgs através da reação  $H(130\text{GeV}) \rightarrow b\bar{b}$ , que ocorre com a probabilidade de 85%, sendo a assinatura dominante no contexto do Modelo Padrão atualmente aceito. Desta forma, é essencial um método eficiente de detecção de quarks tipo *bottom*. Neste caso, deseja diminuir a quantidade de eventos ruidosos tipo  $p\bar{p} \rightarrow Wb\bar{b}$ , que contaminam o sinal de interesse  $p\bar{p} \rightarrow WH \rightarrow l\nu b\bar{b}$ .

A idéia central desse trabalho é definir um sistema neural especializado nas assinaturas de interesse destacadas, que consiga reduzir a taxa de eventos ordinários maximamente, diminuindo, por sua vez, a luminosidade necessária para a detecção do Higgs. Isto se faz necessário pois o experimento D0 encontra-se em uma posição limite para o estudo do bóson de Higgs.

Eventos simulados de bósons de Higgs com massa em 100, 120 e 130 GeV e eventos típicos do processo natural de contaminação foram utilizados para treinar um processador neural com 3 entradas, 6 neurônios escondidos e apenas um neurônio de saída. As três variáveis de entrada escolhidas foram a energia transversa de jatos rotulados como provenientes de decaimentos de quarks tipo *bottom*, o número de traços associados a este jato e sua largura detetada nos calorímetros. Fixando-se um corte na saída da rede, obtém-se cerca de 65% de eficiência na detecção de bósons de Higgs com 100 GeV, 74% para Higgs a 120 GeV e 58% para aqueles com 130 GeV, para uma taxa de falso-alarme de 24%. Utilizando-se técnicas tradicionais, a eficiência máxima de detecção é de apenas 45%, para a mesma taxa de falso-alarme.

Ainda em outro trabalho [67], empregam-se redes neurais artificiais para a detecção de *top* quarks no mesmo experimento. De forma semelhante ao trabalho anteriormente exposto, identificam-se as assinaturas de interesse que contenham quarks deste tipo e sua contaminação natural. A partir destes eventos, define-se um conjunto de variáveis que contenham a informação de detecção necessária e aplica-se um

treinamento à base da retro-propagação de erros para atingir uma melhor separação sinal de interesse/física ordinária. No caso específico deste trabalho, uma rede neural “especialista” é dedicada a cada tipo de assinatura de interesse, de forma a maximizar o potencial discriminante com relação ao *ruído* específico daquele canal. As eficiências finais de detecção do *top* quark são baseadas nas eficiências acumuladas de cada rede especialista.

### 6.1.2 Redes neurais em sistemas de filtragem

O número de falsos-positivos em um experimento em física de altas energias pode ser bastante grande, chegando muitas vezes a ser ordens de magnitude mais elevadas que o sinal que se deseja detectar. Somado a este fator, incluem-se as restrições de tempo de processamento de um sistema de filtragem que deva operar em tempo real, processando milhões de eventos por segundo.

O primeiro sistema de filtragem baseado em redes neurais foi construído e testado em 1992, no Fermilab, Estados-Unidos, para o experimento *D0* no Tevatron [68]. Este sistema pode determinar os parâmetros da trajetória de múons levando em consideração valores de tensão digitalmente amostrados de uma câmara dividida em três planos, cada um contendo várias células de detecção. Cada célula é equipada com placas coletoras de tensão que permitem a amostragem da distância relativa da partícula com relação aos dipólos através do atraso do sinal de avalanche disparado pela interação do múon com a célula de detecção.

Levando-se em consideração os valores de atraso em cada célula e que a partícula atravessará todas as camadas do detetor, é possível calcular o ponto de impacto aproximado e o ângulo da trajetória da partícula. Para tal, uma rede neural tipo MLP com 64 neurônios escondidos e 64 saídas, baseada em *hardware* especializado (chip Intel ETANN), é treinada levando-se em consideração os sinais de interação de cerca de 10.000 múons com este detetor. Estes eventos foram simulados através de processos de Monte-Carlo. O treinamento em si ocorre em um PC que emula o sistema em *hardware*. Os pesos podem então ser descarregados no processador e utilizados *online*.

Os resultados encontrados por este sistema são surpreendentemente positivos em comparação com métodos de reconstrução *offline*. Considerando-se um conjunto de



testes de alguns milhares de traços, o sistema baseado no processador neural obtém um erro menor que 5 mm no ponto de impacto na primeira camada de 96% (contra 98% de um sistema de *fitting offline*). Para o coeficiente angular, este sistema possui um erro menor que 5,7° para 93% (contra 97% *offline*) para os padrões de teste. O tempo total de processamento do sistema é de apenas 8  $\mu$ s.

Já no experimento H1, no acelerador HERA, do laboratório DESY em Hamburgo, Alemanha, estuda a distribuição do momento dos constituintes de prótons e mede a força de acoplamento do glúon aos diferentes quarks. Devido às interações tais como as que ocorrem entre o feixe e resíduos no tubo do acelerador, a taxa de eventos que constituem ruído de fundo chega a uma ordem de  $10^5$  comparada à ocorrência de física interessante. Para filtrar a taxa de eventos inicial de cerca de 10 milhões por segundo para apenas 100, este experimento utiliza um sistema de filtragem dividido em 4 níveis. O primeiro nível de filtragem reduz a taxa de eventos em cerca de 100 vezes. Desta forma, o segundo nível de filtragem tem apenas cerca de 10  $\mu$ s para tomar uma decisão. Este nível deve reduzir a taxa de eventos aproximadamente 10 vezes. Uma rede neural também foi utilizada para resolver este problema [60].

No primeiro nível de filtragem, são calculadas 16 variáveis, tais como a energia total, transversal e nas partes central do calorímetro, dentre outras. Este nível de filtragem ignora, no entanto, quaisquer relações entre estas variáveis. Cortes mais sofisticados podem ser executados no segundo nível adicionando-se a estas variáveis outras quantidades que somente estão disponíveis após o tempo de decisão do primeiro nível e alimentando-se uma rede neural tipo MLP.

As variáveis selecionadas (somadas de energias, número de partículas com carga detetadas, etc.) são utilizadas para determinar se o evento em questão advém de uma interação elétron-próton (Física de interessante), de uma interação do feixe com uma molécula no tubo do acelerador ou qualquer outro fenômeno a ser desprezado. O sistema é implementado em *hardware*, utilizando o processador CNAPS 1064, da empresa Adaptive Solutions, executando em um tempo total de 20  $\mu$ s. Cada canal de decaimento é estudado separadamente e uma rede especialista utilizada para detetar sua ocorrência.

A eficiência do sistema é variável conforme o canal em análise e apresentada

neste trabalho levando-se em consideração a taxa de redução de eventos representando falso-alarme com relação ao primeiro nível de filtragem. Esta taxa chega a valores entre 40 e 160 para eventos tipo  $\gamma p \rightarrow X + J/\Psi$  e de 10 a 20 para eventos tipo  $\gamma p \rightarrow$  jatos, o que está acima das especificações originais para o sistema de filtragem. Nestes casos, o sistema neural permitiu que se elimine completamente pré-escalamento no primeiro nível, aumentando substancialmente o número de eventos interessantes registrados em mídia permanente.

### 6.1.3 Redes neurais e calorimetria em sistemas de filtragem

Em [61], descreve-se um sistema de detecção neural de elétrons em tempo real que utiliza-se das informações dos calorímetros e.m. do experimento CMS [13]. O primeiro nível de filtragem desta experiência não exaure a informação provida pelo calorímetro e.m. e portanto há oportunidade para que um segundo nível de filtragem, apoiando-se na informação não aproveitada pelo LVL1, consiga atingir alguma redução na taxa de eventos que será repassado ao terceiro nível de filtragem.

A informação utilizada para a detecção de elétrons são os valores de energia de um conjunto de células provenientes dos cristais do calorímetro e.m.. Este conjunto define um “quadrado” de aproximadamente 7 por 7 células deste detetor, cobrindo uma área de aproximadamente  $0,1 \times 0,1$  no plano  $\eta \times \phi$ . O algoritmo escolhido inicia o processamento localizando o pico de deposição de energia na região. Em seguida, de forma a tornar o processo de detecção independente da energia, normaliza-se o valor de deposição energética em cada célula pelo valor de energia na célula mais energética. Para amplificar os valores de deposição energética nas bordas da região em análise, tira-se a raiz quadrada do valor normalizado das células.

Cada região analisada, composta de 49 valores de energia normalizada, é disponibilizada como entrada a uma rede neural tipo MLP, com 8 neurônios escondidos, treinada para detetar jatos e elétrons. A base de dados inicial consistia de 50.000 jatos e 15.000 elétrons simulados com energia entre 5 e 225 GeV. Estes eventos são submetidos a uma filtragem por um sistema de simulação do LVL1 do CMS, definindo-se a base de dados disponível para treinamento: 1.500 jatos e por volta de 13.000 elétrons. Para uma taxa constante de detecção de elétrons de 98%, obteve-se uma taxa de detecção de jatos de 71% para eventos com ruído de fundo equivalente a

um ambiente simulado de alta luminosidade ( $10^{34}\text{cm}^{-2}\text{s}^{-1}$ ) e 78%, para um ambiente em baixa luminosidade ( $10^{33}\text{cm}^{-2}\text{s}^{-1}$ ).

Em [69], sugere-se a utilização de redes neurais artificiais para detecção de elétrons na parte central do detetor D0, Fermilab. Especificamente, para elétrons com energia superior a 10 GeV, que encontram-se normalmente embebidos em uma massa de jatos com fortes componentes eletromagnéticas. O conjunto de treinamento disponível consiste de cerca de 2.600 elétrons simulados, com energia entre 10 e 100 GeV e cerca de 1600 jatos, com ao menos um dos dois jatos hadronizando em um estado final  $\pi^0$ , o que confundirá, provavelmente, o sistema de detecção. Um conjunto similar de eventos é usado para o teste do sistema.

Sete variáveis que determinam a penetração e largura do objeto observado nos calorímetros são fornecidas a um processador neural. Estas variáveis são, em sua maioria, valores energéticos, normalizadas pela energia total do agrupamento observado. O sistema de detecção neural é composto de 7 entradas, 16 neurônios escondidos e apenas um neurônio de saída, treinado usando a técnica clássica de retro-propagação de erros. Para esta bancada, observa-se que para níveis aceitáveis de detecção do sinal, o sistema neural provê uma supressão de ruído de fundo 5 a 10 vezes melhor.

A mesma técnica, segundo o artigo, pode ser empregada na detecção de léptons tau ( $\tau$ ). Um novo conjunto de 9 variáveis de detecção é determinado, exaltando as características da partícula de interesse. As 9 variáveis são fornecidas a uma rede neural, treinada para reconhecer este tipo de partícula embebido em um ruído de fundo ocasionado mais uma vez por jatos com fortes componentes eletromagnéticas. O sistema neural possui 20 nós escondidos e apenas um neurônio na camada de saída, tal como o sistema de detecção para elétrons descrito. Após ser treinada, a rede mostra, para uma eficiência na detecção de *taus* fixa, um nível de rejeição até 2 vezes melhor que o método atualmente empregado, baseado em uma análise de componentes principais. Por utilizar uma rede neural, a implementação do sistema de detecção é trivial e poderia ser utilizado em substituição ao método atual empregado no experimento.

No caso em questão, a análise neural é sugerida como forma de substituição de uma análise por Componentes Principais (do inglês, *Principal Component Analysis*,

PCA), que aplicada ao mesmo problema resulta em um resultado inferior, indicando a possível existência de correlações de ordem superior a 2.

Ao invés de estabelecer um conjunto de variáveis que acentuem o poder discriminatório da rede, [70] sugere que as informações de células homogeneamente distribuídas em um determinado agrupamento calorimétrico possam ser pré-processadas através de uma análise de componentes principais para a identificação de um sub-espaço ótimo de filtragem. Este artigo compara diferentes maneiras de executar a extração de componentes, levando-se em consideração todos os dados disponíveis em uma única extração ou executando extrações parciais que preservem o perfil de deposição energético do sistema. Os resultados destes sub-espaços definidos pelas componentes principais do agrupamento são abastecidos à um discriminador neural que realiza o papel de detecção

Levando-se em consideração as possibilidades de segmentação da extração, se obtêm 72,4% de eficiência na detecção de elétrons contra 39% de falso-alarme para uma análise não segmentada e 98,9% para apenas 1% de falso-alarme preservando-se a informação de penetração longitudinal do objeto, ou seja, extraíndo-se as componentes principais camada a camada.

#### 6.1.4 Análise neural resumida

A análise neural, como indicado anteriormente, vem sendo sugerida como interessante candidata à análise bayesiana clássica ou à análise da matriz de correlação (PCA) nos trabalhos revisados. Em muitos dos casos, os métodos empregados podem ser simplesmente substituídos pela análise neural. Em outros, observa-se uma tendência ao questionamento da qualidade das variáveis produzidas pela extração de características, onde o autor tenta explorar o espaço original, uma vez que dispõe de ferramentas mais sofisticadas que simples cortes uni ou bi-dimensionais para executar a seleção. As variáveis de interesse advêm diretamente do detetor, calibradas ou a serem calibradas [71], ou são definidas através de equações ou transformações simples e normalizadas de acordo com parâmetros do objeto que se deseja discriminar.

Redes neurais artificiais resolvem um problema latente em Física de Altas Energias, inerente aos avanços na área de detecção e à explosão do tamanho dos eventos (veja a Figura 2.6) e por consequência do número de variáveis que estão disponíveis

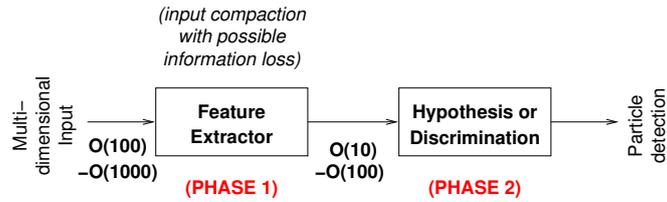


Figura 6.1: Um sistema genérico de detecção em um problema de Física de Altas Energias.

para o discriminador. Por outro lado, fornecem uma base matemática sólida que possa ser utilizada para analisar e transformar os dados disponíveis em projeções onde o problema de detecção torna-se mais óbvio.

Duas fases de processamento podem ser claramente definidas em muitos dos trabalhos nesta área, como indica a Figura 6.1:

1. A **extração de características**, onde deseja-se compactar a informação disponível para análise em um mínimo de variáveis que consigam extrair toda (ou na maior parte) a informação discriminante disponível na RoI. Esta fase é normalmente essencial uma vez que dispõe-se de milhares de variáveis para cada evento, contando com os elementos de detecção providos pelo experimento até variáveis complexas que podem ser acessadas após a reconstrução como o número de jatos encontrados ou o ângulo entre léptons;
2. A **classificação ou hipótese**, seguindo-se à compressão do espaço original de entrada, representa a detecção da física de interesse. Observa-se uma tendência ao desenvolvimento dos sistemas utilizando-se dados de simulação acurados, mas onde possível, dados provenientes diretamente do detetor. É interessante perceber que sistemas desenvolvidos através de simulações ainda comportam-se de forma satisfatória ou adaptativa em seu emprego final, com dados reais.

Os métodos propostos possuem maior robustez que sua contrapartida e podem ser executados de forma bastante rápida, em muitas instâncias sendo propostos para execução em *hardware* especializado. Sistemas especialistas podem ser empregados em problemas mais complexos, onde deseja-se dedicar cada porção do detetor a uma atividade bem definida, combinando-se ao final as respostas de cada sub-sistema para obter um sistema de classificação com uma única saída.

No que se segue deste capítulo, abordaremos sistemas de classificação que substituem o classificador atualmente empregado na discriminação elétron/jato para o Segundo Nível do Sistema de Filtragem do experimento ATLAS, introduzindo técnicas automáticas de detecção. Após definirmos os limites do processo de compressão estabelecidos pelo extrator de características T2Calo, tentar-se-á abordar a discriminação elétron/jato através de outras variáveis que possuam um poder discriminante maior que o sistema proposto atualmente no CERN.

### 6.1.5 Discriminador de Fisher na separação elétron/jato

O discriminador de Fisher [72] ou a Análise de Discriminação Linear descreve um algoritmo para que se maximize a capacidade discriminante de um corte num plano com  $N$  dimensões, que separa duas classes de dados. O resultado do discriminante é ótimo no caso de dados apresentarem uma distribuição gaussiana e as matrizes de covariância, para ambas as classes em separado, serem idênticas<sup>1</sup>. Em resumo, o discriminante de Fisher maximiza, através de uma transformação linear, a distância entre as classes que se deseja detetar, ao mesmo tempo diminuindo a distância de elementos intra-classe.

Levando-se em consideração as 4 variáveis definidas pelo T2Calo, é possível encontrar o plano quadri-dimensional que defina um discriminador linear ótimo, usando-se o algoritmo de retro-propagação resiliente a erros, como descrito no Apêndice B. A separação em classes de treinamento e teste, como apresentada na Seção 5.6.1, será reaproveitada aqui, para que seja possível a comparação dos resultados nos dois casos. O fluxo de informação deste novo sistema está apresentado na Figura 6.2. Nesta nova implementação do algoritmo de filtragem, substitui-se o discriminador por cortes proposto originalmente, por um sistema simples de detecção linear.

O sistema de treinamento e teste foi implementado em um ambiente de trabalho C++, seguindo o paradigma da orientação a objetos (OO) [73, 74]. Este ambiente,

---

<sup>1</sup>A Análise de Discriminação Quadrática, no entanto, demonstra que esta característica pode ser relaxada considerando-se que seja sempre possível projetar, através de uma transformação linear, o espaço de entrada em um outro espaço onde as matrizes de covariância sejam iguais e portanto recaindo no caso simples.

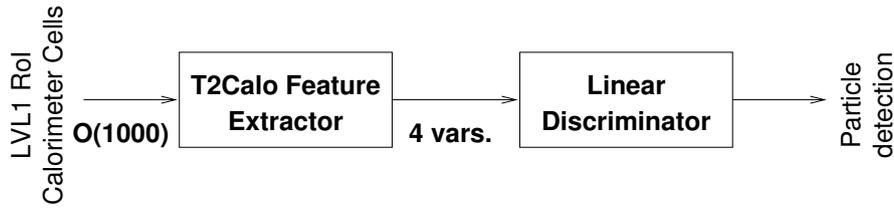


Figura 6.2: Diagrama de blocos de um sistema de detecção elétron/jato baseado no Discriminador de Fisher associado ao extrator de características T2Calo.

que será re-utilizado em várias partes deste trabalho, é discutido em detalhes no Apêndice C. A Figura 6.3 contém um diagrama de fluxo com os diversos passos do sistema de treinamento implementado. Inicialmente, os bancos de dados usados para o treinamento e teste da rede são carregados. Os valores de média e variância experimentais, são extraídos do conjunto de treinamento e guardados juntos aos pesos  $\hat{w}(n)$ , para que sejam aplicados durante o treinamento<sup>2</sup>. Os pesos sinápticos são aleatoriamente inicializados (entre  $-1$  e  $+1$ ). Valores-alvo, para cada uma das classes, são pré-fixados:  $-1$  para elétrons e  $+1$  para jatos. O processo de treinamento é então disparado. Para cada época ou batelada<sup>3</sup> de treinamento, um conjunto de elétrons e jatos é escolhido aleatoriamente à partir dos bancos-de-dados disponíveis. Os valores de erro são calculados e seu valor médio é aplicado para a correção dos pesos sinápticos, considerando a taxa de aprendizagem  $\alpha$ . A rede resultante no final do processo de treinamento é salva e utilizada como o discriminador treinado.

Um sistema configurável é utilizado para detetar a estagnação do treinamento e parar o treinamento do detetor. Para isto, tal sistema baseia-se na observação do valor do produto SP para o conjunto de teste. Quando este valor atinge um estado no qual as alterações sinápticas não o modificam significativamente (levando-se em consideração uma margem de erro configurável e obtida empiricamente), e por um número de iterações, o treinamento é automaticamente encerrado. O estado final do discriminador é recarregado. Um estudo dos valores de Erro Médio Quadrático

<sup>2</sup>De fato, seria mais eficiente que os fatores de normalização fossem aplicados previamente ao treinamento. No entanto, uma vez que deseja-se re-aproveitar a base de código para rodar o discriminador em uma etapa seguinte, é mais simples, do ponto de vista da implementação, que a normalização seja aplicada como parte do passo de execução do discriminador.

<sup>3</sup>Os termos *época* e *batelada* serão usados indistintamente no contexto deste trabalho.

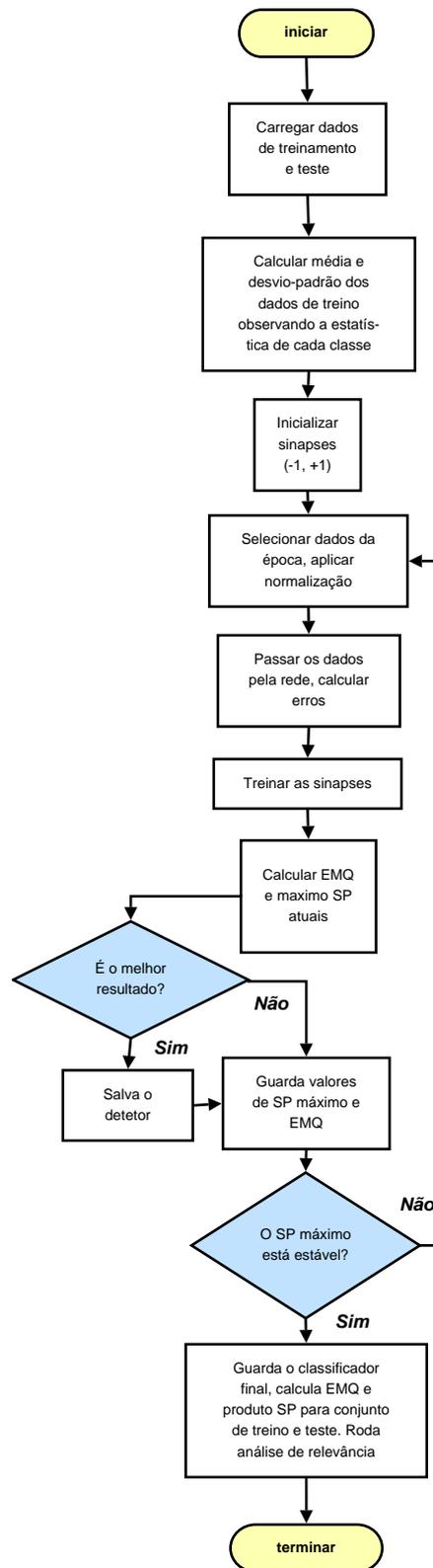


Figura 6.3: O fluxo implementado para o treinamento do discriminador linear.



(EMQ, do inglês *Mean-square error*, MSE) é feito para os conjuntos de treinamento e teste. O produto SP máximo também é avaliado para os dois conjuntos de dados. Esta técnica tem o objetivo de evitar um treinamento excessivo da rede (do inglês *overtraining*) fazendo com que ela se especialize em demasiado nestes dados sem atingir uma capacidade de discriminação generalizada. O valor da variação do produto SP, a partir do qual considera-se que o sistema esteja estagnado, dependerá do problema. Valores típicos estão na ordem de  $10^{-3}$ , para o problema abordado utilizou-se um valor igual a 0,0001.

Neste trabalho, optou-se pela não utilização de um conjunto de testes distinto do conjunto de validação. Este conjunto poderia ser utilizado para verificar a operação do sistema após o treinamento. No entanto, a física disponível para o estudo é limitada devido a natureza de geração de eventos. Para se realizar uma análise baseada no LVL2, elétrons e jatos que sejam selecionados por um corte do LVL1 devem ser simulados, uma vez que o LHC ainda não encontra-se operacional. Desta forma, os termos *conjunto de teste* e *conjunto de validação* serão utilizados de forma intercambiável no contexto deste trabalho.

A taxa de redução do LVL1 é bastante expressiva, atingindo uma supressão de aproximadamente 1000 vezes na taxa de contaminação de elétrons. Desta forma, para produzir 1000 jatos que seriam selecionados por um corte do LVL1 é necessário que se produza cerca de 1 milhão de eventos deste tipo. Cada simulação de um evento pode durar muitos minutos, às vezes horas de processamento do poder de computação disponível. Por outro lado, dada a natureza da geração dos eventos (simulação por processos de Monte Carlo), a estatística contida no conjunto é tida como suficiente para expressar bem as características do tipos de evento analisados.

### **6.1.6 Resultados para um detetor linear aplicado à separação elétron-jato**

Utilizando o sistema descrito anteriormente e a separação dos dados (em classes de treinamento e teste) proposta no início do estudo, realizou-se um conjunto de 5 testes, com a finalidade de verificar que o mínimo tenha sido atingido durante o processo de treinamento. Em cada um dos cinco testes os pesos sinápticos são inicializados aleatoriamente com um valor dentro do intervalo  $[-1, +1]$ . O tamanho

da época é fixado em 22000 (duas vezes o número total de elétrons disponíveis para o treinamento), de forma que cada passo de treinamento represente a informação contida em toda a massa de dados disponível.

Dependendo da inicialização, a convergência do sistema (estagnação detetada no conjunto de teste a partir do produto SP) poderá demorar de 50 a 70 passos para ser estabelecida pelo programa. O valor do EMQ na saída do detetor treinado é ligeiramente diferente para os 5 casos, variando entre 0,3818 e 0,3825. O valor do produto SP no final do treinamento está entre 1,477 e 1,479, também dependendo do ponto de parada.

A Figura 6.4 mostra a evolução do EMQ para a melhor das 5 redes, tanto para o conjunto de treinamento quanto para o conjunto de teste. A Figura 6.5 mostra um gráfico equivalente, mas para o produto SP, ao invés do EMQ. Como é possível verificar pelo valor desta variável, o sistema converge a um valor estável depois de pouco menos que 20 passos de treinamento. O produto SP oscila bruscamente no início do treinamento, por cerca de 10 passos, e depois se estabiliza, em aproximadamente 1,48, para o restante do treinamento. O mesmo ocorre para o EMQ para um valor de aproximadamente 0,38. O perfil de evolução do produto SP e a minimização do EMQ é seguido pelo conjunto de teste, de forma bastante semelhante, indicando que a estatística disponível no conjunto treino é representativa dos dados no conjunto de teste.

A Figura 6.6 mostra os histogramas da saída da rede para elétrons e jatos, considerando-se o conjunto de teste. O ponto ótimo de corte, determinado para que se maximize o produto SP da rede para o conjunto de treinamento para este sistema é  $-0,436$ . No caso em questão, observa-se que, possivelmente, os conjuntos de dados não sejam linearmente separáveis, dado que a saída do detetor, tanto para elétrons quanto para jatos, se superpõe em parte do espectro de saída do sistema. Nestas condições, o máximo produto SP para o conjunto de teste é 1,48, definido no ponto da curva R.O.C. (veja Figura 6.7) onde a eficiência para a detecção de elétrons é 90,77%, enquanto que a eficiência para a detecção de jatos é de 90,21%. A Figura 6.8 mostra uma comparação entre o resultado da otimização proposto atualmente no experimento ATLAS contra os resultados obtidos para este classificador. Através desta figura é possível observar que o discriminador linear praticamente tangencia

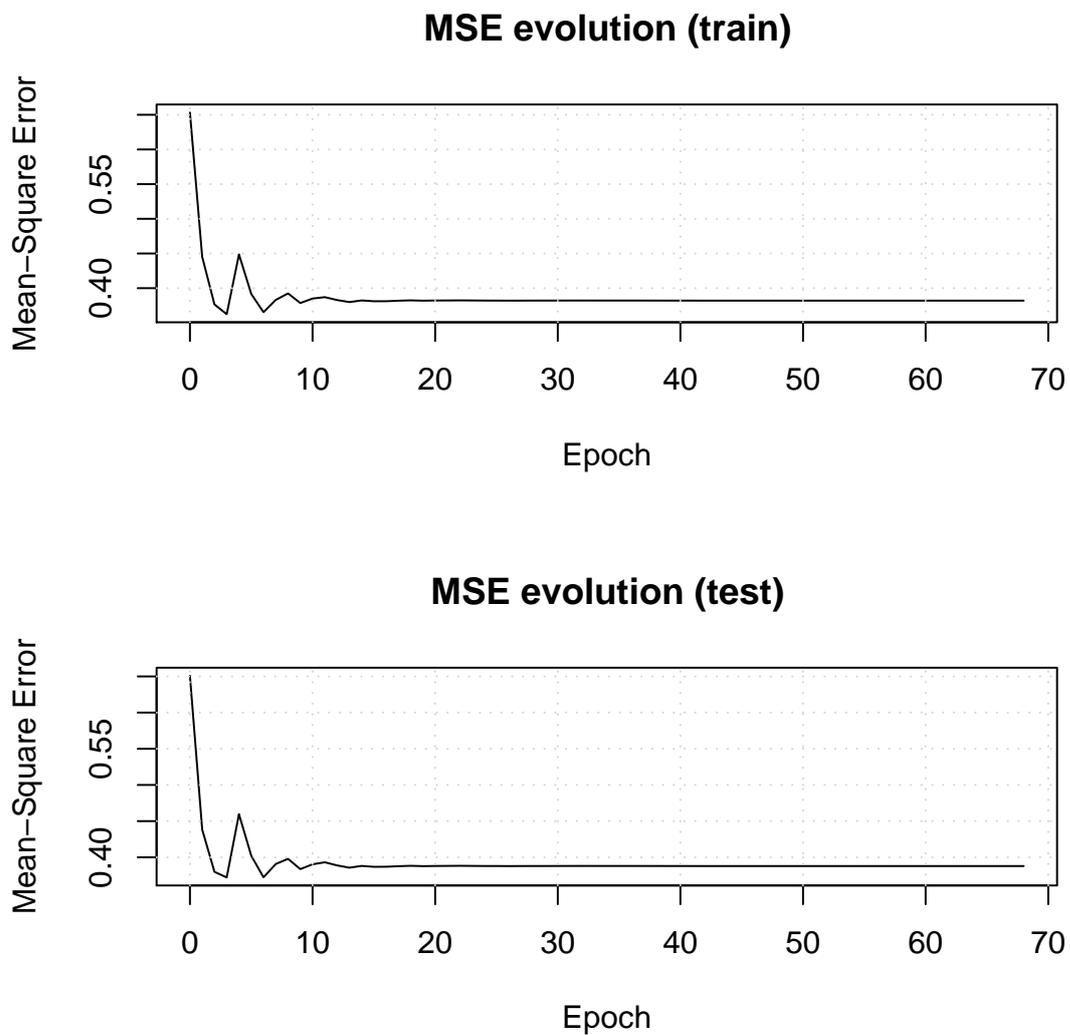


Figura 6.4: Evolução dos valores do EMQ ao longo do treinamento do detetor linear, para o conjunto de treinamento e de teste.

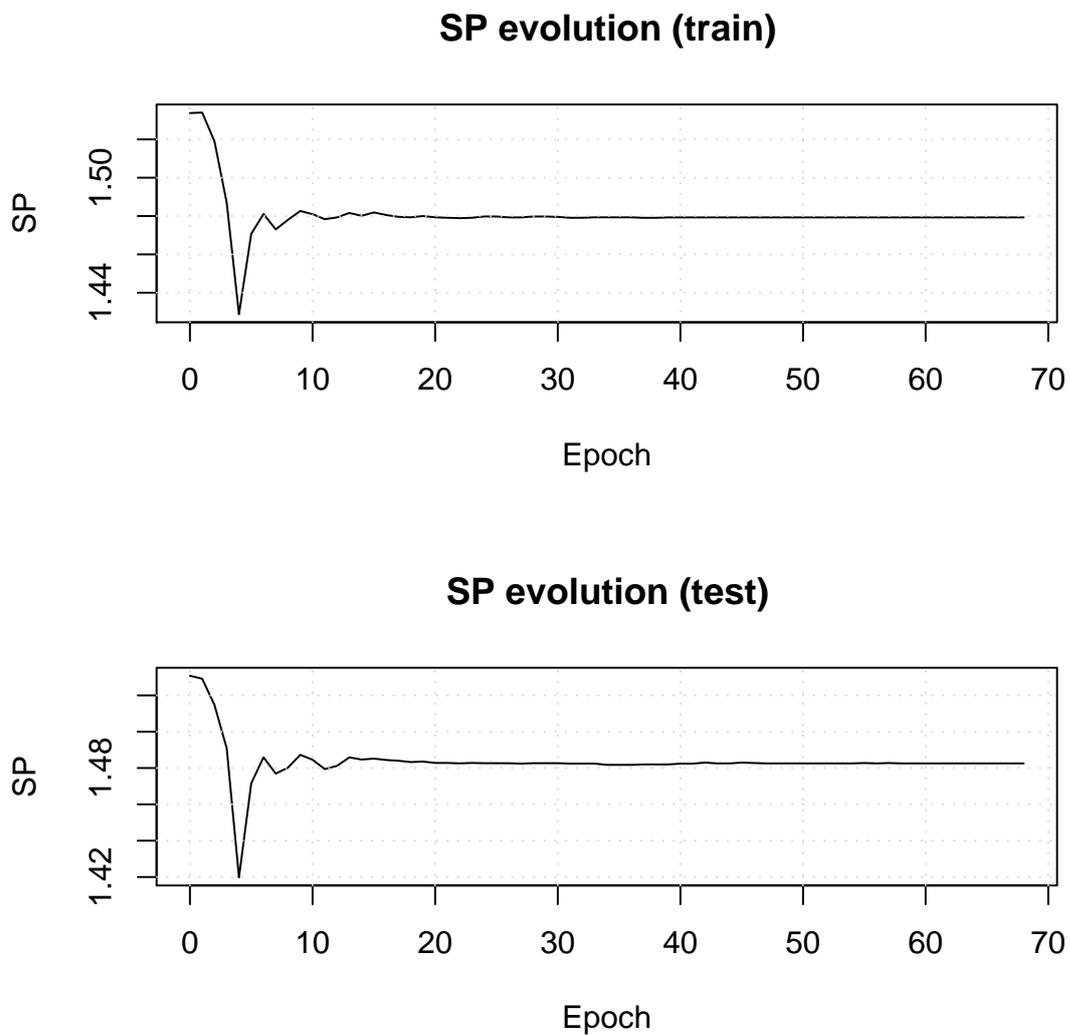


Figura 6.5: Evolução dos valores do produto SP ao longo do treinamento do detetor linear, para o conjunto de treinamento e de teste.

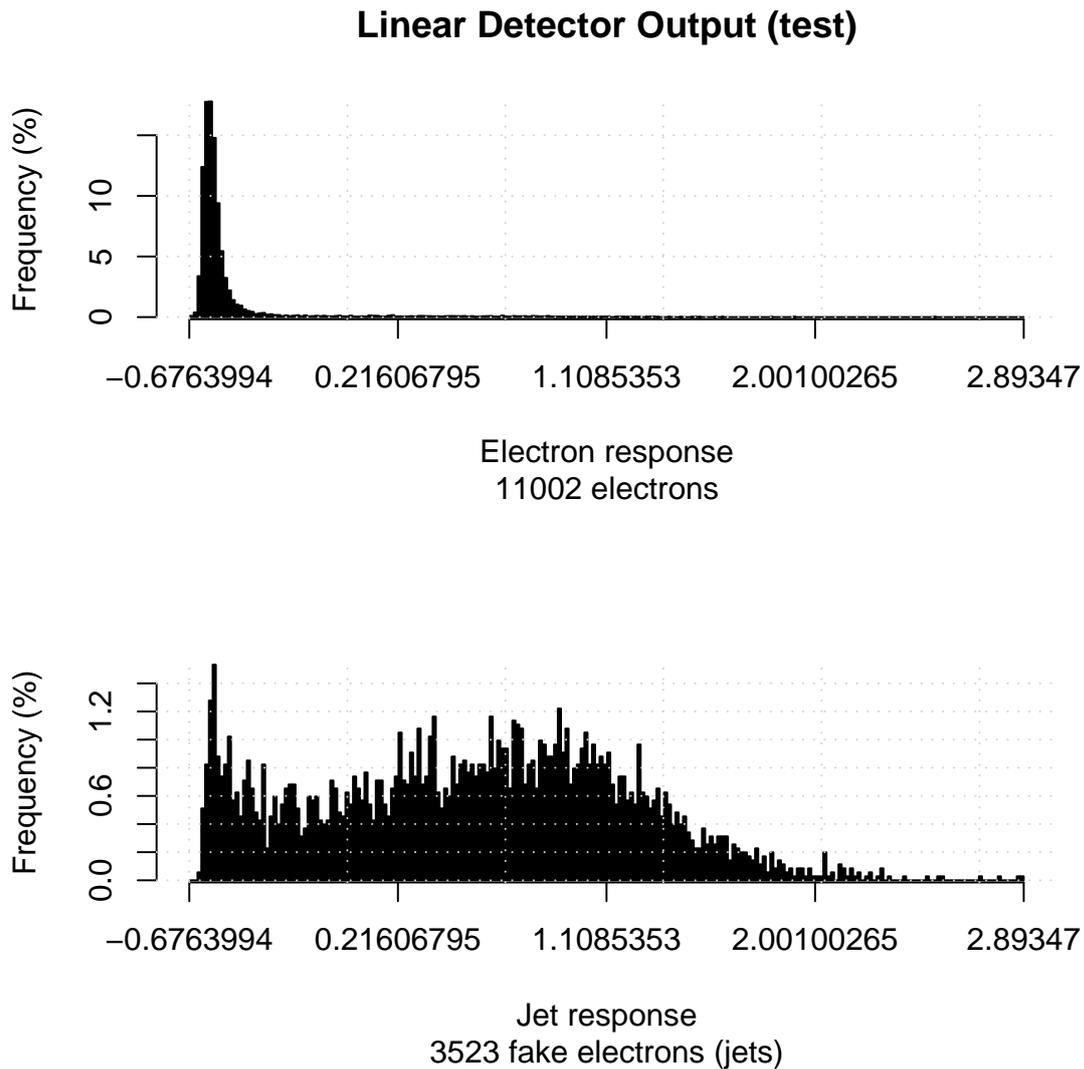


Figura 6.6: Saída do discriminador linear (após treinamento) para elétrons e jatos.

a parte exterior dos pontos da otimização proposta atualmente no experimento. O resultado obtido com este sistema de discriminação, ao invés do longo processo de otimização proposto originalmente, é atingido depois de apenas alguns minutos de treinamento em um computador dotado de um processador AMD 64 com *clock* de 2 GHz e 1 Gb de memória RAM. Ademais, tal sistema apresenta uma capacidade discriminatória equivalente após apenas 10 passos de treinamentos, isto é,  $\approx 25$  segundos na máquina citada. Esta nova ferramenta oferece algo que o EGammaHypo jamais poderá: a atualização dos pesos sinápticos poderá ser feita rapidamente para se arcar com a mudança de condições no experimento.

As Figuras 6.9 e 6.10 mostram, respectivamente, o valor do produto SP segundo

### R.O.C. for e/jet discrimination

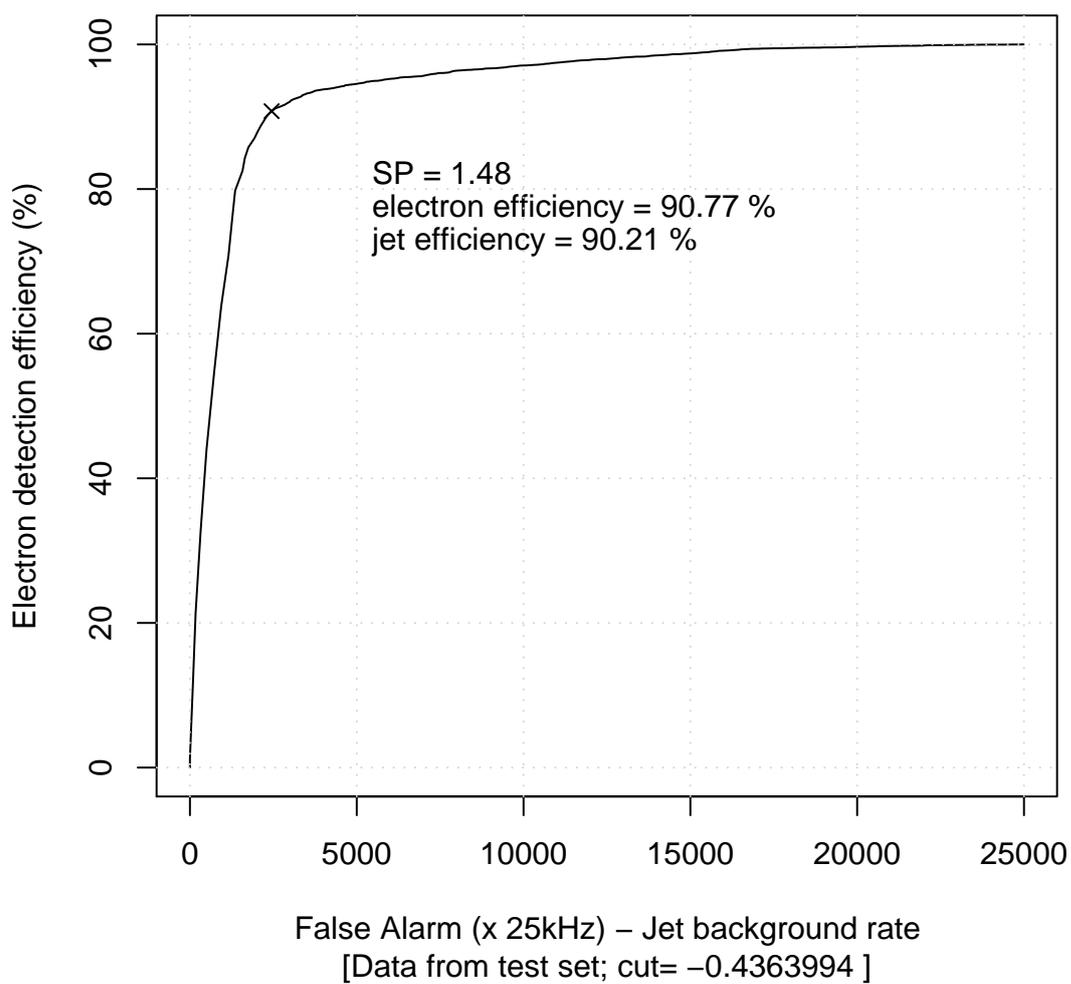


Figura 6.7: Curva R.O.C. para um detetor linear para elétrons e jatos baseado nas 4 variáveis do T2Calo, usadas para discriminação pelo EGammaHypo.

### EGammaHypo parameter optimization (train set)

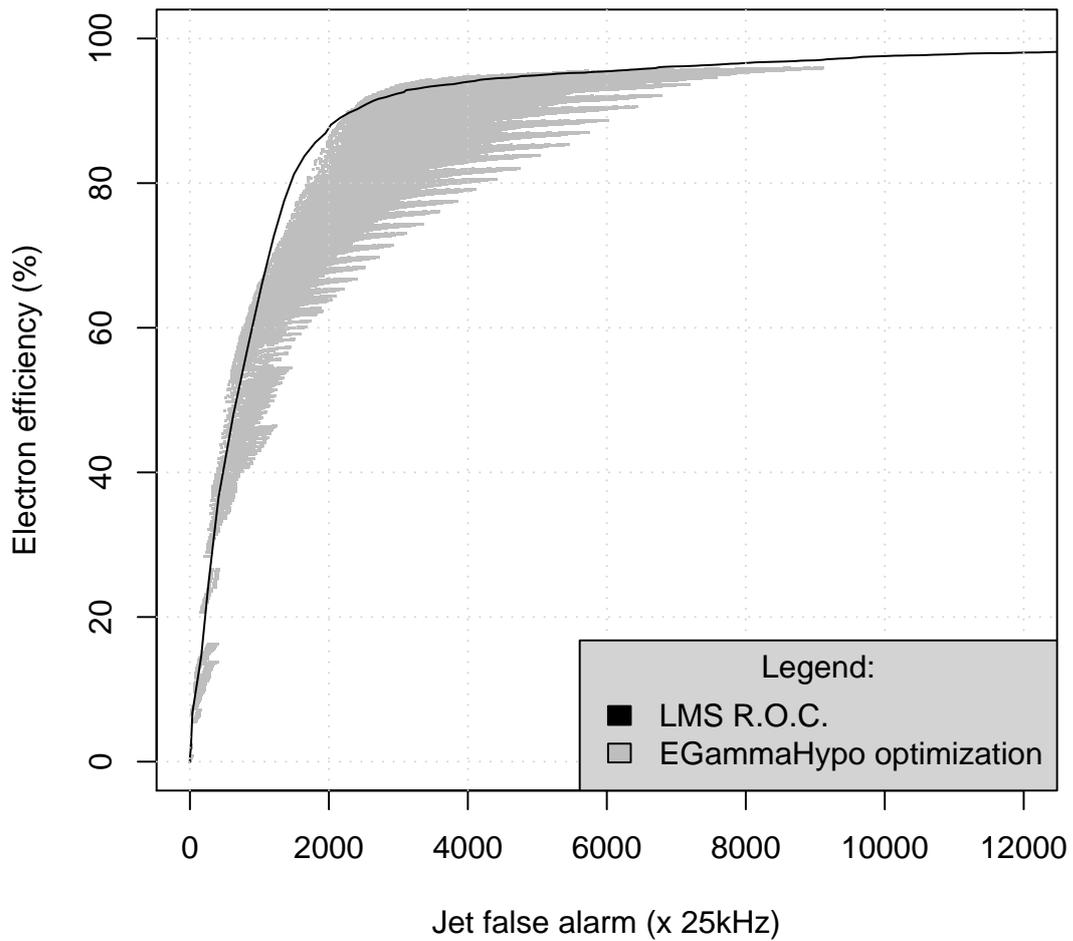


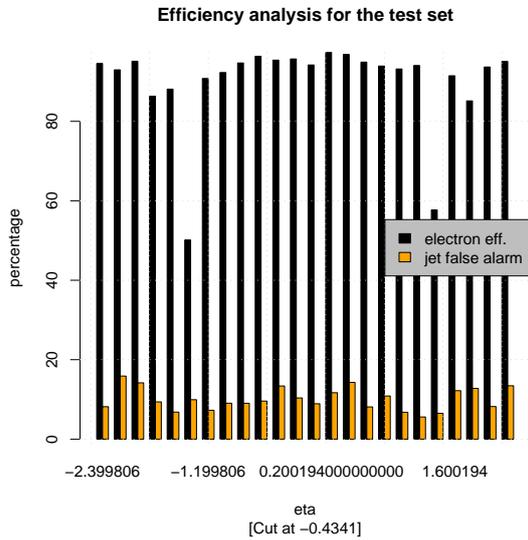
Figura 6.8: R.O.Cs comparativas entre a otimização atual para o EGammaHypo e um detetor linear.

a distribuição dos dados em  $\eta \times \phi$  e  $E_{T_{3 \times 7}}^{e.m.}$ . O número anexo ao topo de cada barra (gráficos do produto SP) indica a quantidade, separadamente, de elétrons e jatos em cada canal e é útil para que se considere a importância de um valor parcial no desempenho do discriminador. Nota-se que na distribuição por  $\eta$ , há uma clara perda de eficiência na região  $|\eta| \approx 1,5$ . Esta notável queda no desempenho do classificador é devido a um espaço sem elementos de detecção nesta área (também conhecido como *gap* ou *crack* dos calorímetros), por onde passam os cabos de leitura e manutenção dos detetores internos. Como mostra a Figura 6.9(a), a perda devido à esta região do sistema sem elementos de detecção, está diretamente relacionada à eficiência de detecção de elétrons. Isto acontece pois elétrons desenvolvem uma cascata de deposição energética tipicamente mais fina e de pouca penetração, estando, adicionalmente, sujeitos à interação com os cabos do detetor. Jatos, por sua vez, tenderão a penetrar mais no detetor e se radializar, o que facilita a recuperação da informação mesmo em regiões mais difíceis, como o *crack*.

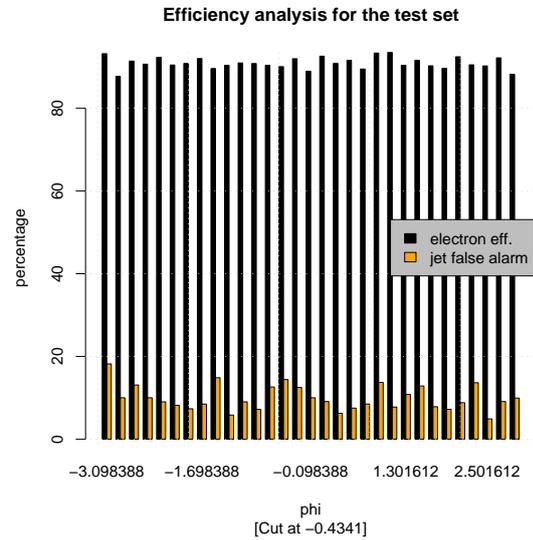
A eficiência é recuperada logo após a região do *crack*. O gráfico de barras para a distribuição  $\phi$  mostra-se praticamente uniforme, indicando que o sistema funciona sem tendências para todos os valores desta variável. Este resultado é esperado, já que o detetor é simétrico neste eixo. A análise por  $E_{T_{3 \times 7}}^{e.m.}$  demonstra uma predominante qualidade de detecção para valores mais baixos de energia do que para valores mais altos. Deve-se levar em conta que, embora a eficiência de detecção de elétrons ainda seja máxima nesta última região, como mostra a Figura 6.10(a), o falso alarme na detecção de jatos também apresenta-se máximo. Uma vez que o produto SP é uma figura de mérito da eficiência agregada de ambas as classes de eventos do discriminador, ela quantificar-se-á em zero nesta região. Este resultado está, possivelmente, associado à baixa estatística de dados disponível na região, como indicado no texto sob as barras. Desta forma, a análise de valores de eficiência de detecção para objetos com mais de 50 GeV deverá ser tomada com esta ressalva.

A Figura 6.11 mostra uma comparação entre as duas técnicas de detecção, por energia transversa na seção e.m.. Por inspeção desta figura, distingue-se que os dois sistemas possuem respostas bastante próximas. Na primeira (mais à esquerda) faixa de energia considerada, o EGammaHypo possui uma resposta melhor, enquanto que com o aumento da energia transversa o detetor linear apresenta-se mais eficiente, e

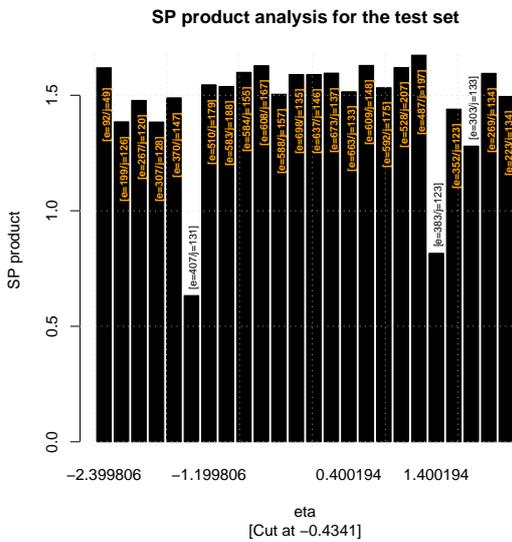




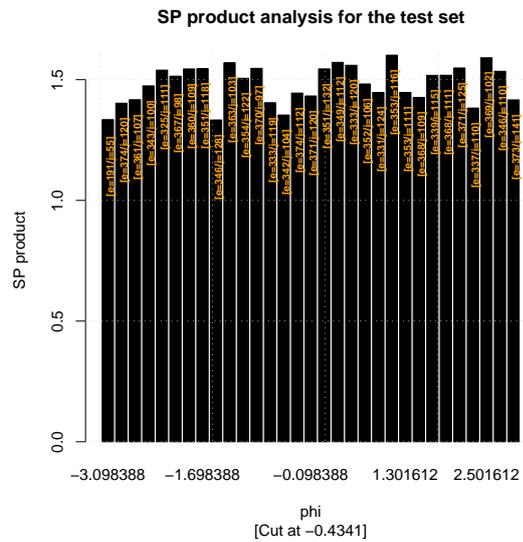
(a) Eficiência por  $\eta$



(b) Eficiência por  $\phi$

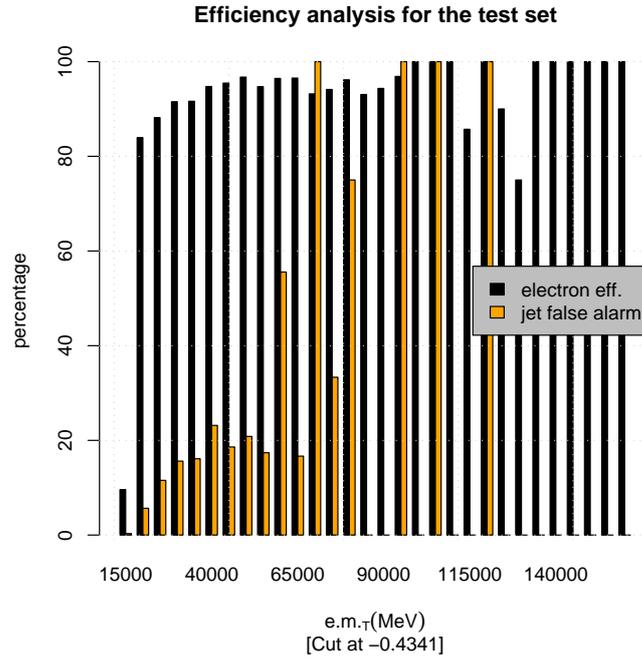


(c) Produto SP por  $\eta$

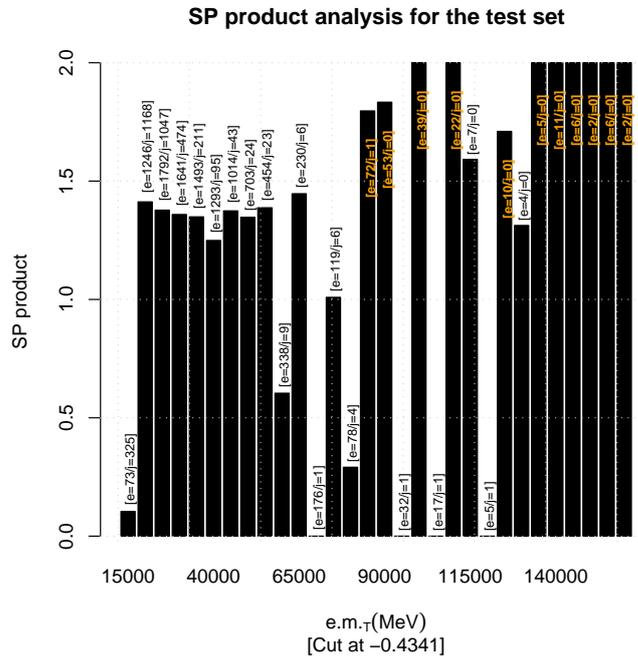


(d) Produto SP por  $\phi$

Figura 6.9: Análise da eficiência de classificação e do produto SP para o classificador linear, para os dados do conjunto de teste ao longo de  $\eta$  em (a) e (c) e por  $\phi$ , em (b) e (d).



(a) Eficiência



(b) Produto SP

Figura 6.10: Análise da eficiência de detecção de elétrons e falso-alarme em jatos (a) e do produto SP (b) por  $E_{T_{3 \times 7}}^{e.m.}$  para o classificador linear, tendo por base os dados do conjunto de teste.

segue este padrão para grande parte das faixas onde se diferenciam (15 pontos no total). Na região onde há maior concentração de eventos de teste, como é possível determinar a partir da Figura 6.10, o desempenho dos detetores é muito próximo. Com o aumento da energia, onde a eficiência do EGammaHypo começa a cair, o sistema baseado no detetor linear consegue obter bom desempenho superando o primeiro.

Levando-se em consideração a saída do detetor, na Figura 6.6, compreende-se que este método de detecção não apresente a robustez necessária ao experimento. Conclui-se, a partir desta saída, que os objetos das duas classes de dados não sejam linearmente separáveis.

## 6.2 Métodos neurais de discriminação

Redes neurais artificiais (RNA) vêm sendo atualmente utilizadas com sucesso em problemas de otimização, controle e reconhecimento de padrões [75]. RNA's são modelos matemáticos inspirados no conhecimento limitado que temos do cérebro animal. Neste contexto, entende-se que uma rede neural pode aprender através do contato com os elementos de interesse de um determinado espaço de entrada e que este conhecimento é armazenado através das conexões sinápticas que conectam os elementos processadores (neurônios) da rede. Dentre as características interessantes destacam-se para o problema atual:

**Robustez** Sistemas neurais, por tentarem recortar o espaço de variáveis de entrada de forma mais sofisticada que sistemas lineares, normalmente atingem um nível de separação entre as classes de objetos superior a outros detetores. Esta qualidade é fundamental em ambientes onde a informação de entrada no sistema de detecção nem sempre está completa. Especificamente no experimento ATLAS, dada a grande complexidade do sistema, é possível que determinadas regiões do detetor sejam omitidas pelo sistema de leitura devido a panes de naturezas diversas. Neste contexto, contar com um sistema que ainda conseguirá obter um bom nível de discriminação apesar destas condições torna-se muito vantajoso;

**Generalização** RNA's podem extrair a informação relevante escondida sob uma

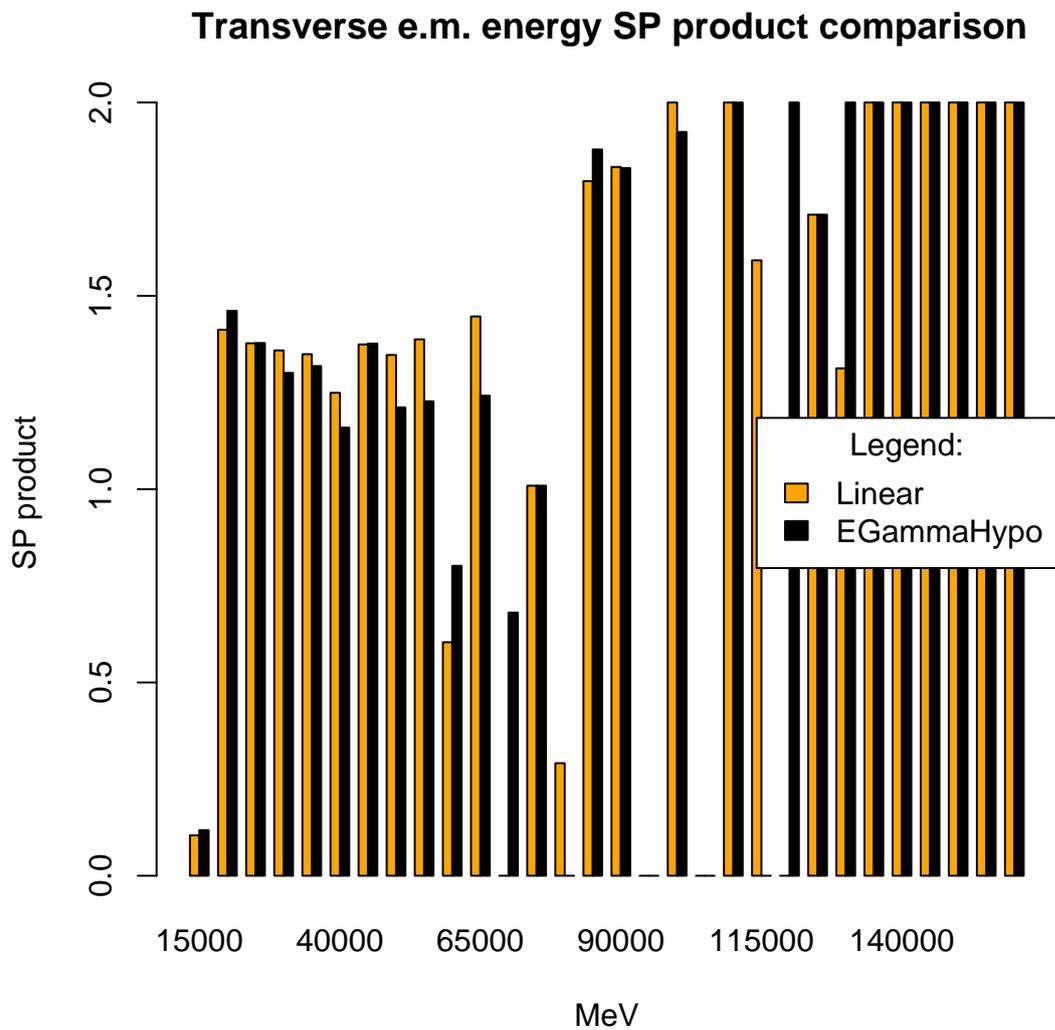


Figura 6.11: Comparação do produto SP para o classificador linear e o EGamaHypo por energia transversa na seção e.m..

grande quantidade de ruído e portanto poderão detetar canais físicos não simuláveis no caso de exibirem propriedades equivalente àquelas para o qual foram treinadas. Num ambiente de descobertas, como é o experimento ATLAS, esta característica torna-se fundamental uma vez que a Física de interesse pode se apresentar num formato pouco evidente a sistemas clássicos de deteção;

**Simples implementação** RNA's podem ser facilmente descritas em linguagens de programação convencionais, atingindo desempenhos satisfatórios para um grande conjunto de aplicações em tempo real. Esta característica torna RNA's atrativas para implementação em ambientes como o Sistema de Filtragem do experimento ATLAS, onde os tempos de processamento e a utilização de recursos são bastante reduzidos.

RNA's podem ser implementadas de várias maneiras, incluindo sistemas com percéptrons com apenas uma ou múltiplas camadas (do inglês, *multi-layer perceptron*, ou MLP), incluindo sistemas realimentados ou híbridos [75]. O processo de aprendizagem ou aproximação da resposta à saída desejada no sistema é conduzido normalmente por uma fase de treinamento, que poderá ser realizado, igualmente, de maneiras distintas. Duas grandes sub-categorias são reconhecíveis: métodos supervisionados e não-supervisionados.

Na aprendizagem supervisionada, o desenvolvedor utiliza-se de uma informação de conhecimento *a priori* da função que deseja executar para reprimir ou reforçar a migração do sistema na direção adequada. Uma das grandes dificuldades neste tipo de aprendizagem é a definição de um algoritmo de treinamento que convirja de forma adequada ao problema que se deseja atender.

Muito algoritmos foram desenvolvidos para as diversas estruturas de rede, dentre os quais, a retropropagação de erros tornou-se um dos mais populares em aplicações ligadas a Física de Altas Energias. No treinamento por retropropagação de erros, os sinais de erro são propagados a partir da saída da rede, onde são medidos, para os neurônios escondidos através de um mecanismo análogo ao algoritmo LMS, para o treinamento do sistema linear estudado anteriormente. A retropropagação de erros aplicada a redes neurais artificiais é revista com mais detalhes no Apêndice B. Este método possui, no entanto, muitos problemas para a determinação dos parâmetros ótimos de treinamento. Como variação dos métodos explorados normalmente em

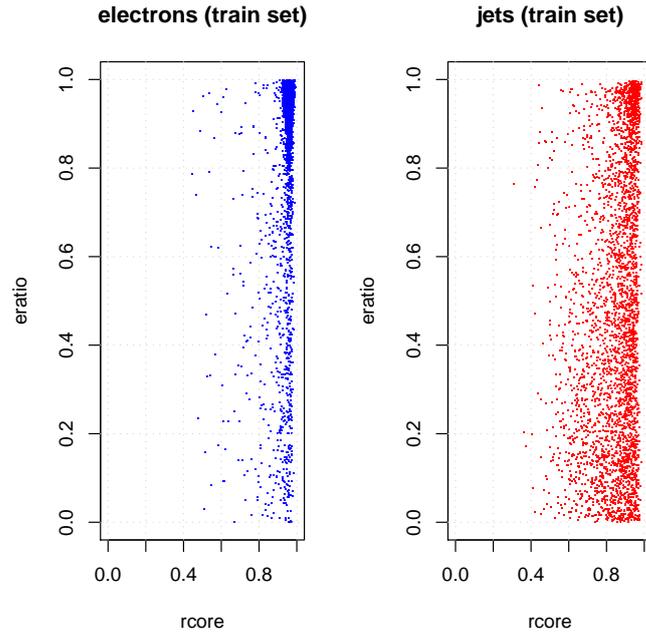


Figura 6.12: Gráfico de dispersão mostrando a distribuição das duas variáveis mais discriminantes para o T2Calo, para as classes de elétrons e jatos.

Física de Altas Energias, e no intuito de buscar-se um método mais robusto e livre de parâmetros de treinamento, utilizaremos uma variante do treinamento por retropropagação de erros conhecida como *retropropagação de erros resiliente* (do inglês *resilient backpropagation*, RPROP) [76]. Esta técnica também encontra-se descrita no apêndice.

Em específico, o problema da separação elétron/jato como se apresenta ao detector EGammaHypo e ao sistema de detecção linear (treinado pelo LMS), indica que correlações não lineares estejam presentes no sistema. De fato, a inseparabilidade linear através das variáveis do T2Calo pode ser facilmente inspecionada através das distribuições das duas variáveis mais discriminantes deste extrator de características, na Figura 6.12.

Não obstante, a qualidade da discriminação é comparável ao sistema de detecção proposto pelo algoritmo EGammaHypo, não necessitando da intervenção de um especialista para seu ajuste. Se por um lado são compatíveis, estes dois sistemas de detecção mostram-se pouco robustos a mudanças de condição como as que serão observadas no experimento ATLAS, onde a luminosidade do feixe está sujeita a brus-

cas variações em questão de minutos. Na seção que se segue, analisa-se o potencial discriminatório que redes neurais artificiais adicionariam ao sistema de filtragem do ATLAS.

Para o treinamento e implementação dos sistemas de discriminação neural, fez-se uso das bibliotecas e programas do *NeuralRinger*. O *NeuralRinger* é um pacote totalmente desenvolvido usando o paradigma da orientação a objetos, implementado em C++ [73, 74] e especialmente concebido para executar a operação de filtragem de partículas dentro do LVL2. Outras soluções para treinamento e execução de redes neurais, tais como o pacote SNNS [77] foram investigadas antes do desenvolvimento deste sistema. Entretanto, os pacotes em questão são normalmente utilizados para uma grande variedade de sistemas neurais. Esta flexibilização muito freqüentemente ocorre em detrimento do desempenho em casos específicos. Ademais, ao usar um sistema tão genérico, estaria-se inserindo uma grande quantidade de código dentro do sistema de filtragem, que seria, em sua maior parte, pouco ou não utilizada.

As escolhas do paradigma de programação e da linguagem de implementação C++ ocorreram em função do ambiente-alvo para este sistema. Como colocado na Seção 4.2.6, o LVL2 de filtragem rodará utilizando como base as bibliotecas do Projeto Athena. O Athena é o ambiente padrão para a análise da Física dos eventos no ATLAS, tanto no Sistema de Filtragem, como na Reconstrução pós-filtragem nas diversas *farms* de análise *offline*. Desta forma, a integrabilidade ao Athena constitui-se um passo essencial para a utilização da técnica no LVL2.

O projeto foi desenvolvido tendo por base as seguintes prerrogativas, em ordem:

1. Otimização da execução de redes neurais: É importante que o sistema tenha um bom desempenho no processamento neural, de tal forma que seja viável sua implementação no sistema de filtragem. Preferencialmente, deseja-se que um passo de execução da rede não dure mais que alguns milissegundos, já que o tempo total de processamento no LVL2 deve ser em média 10 ms, contando com o acesso aos dados;
2. Interoperabilidade: O treinamento do sistema deve ser executado *offline*, dada a disponibilidade dos dados. Portanto, é necessário que o sistema opere tanto dentro do ambiente Athena, quanto em modo desacoplado;

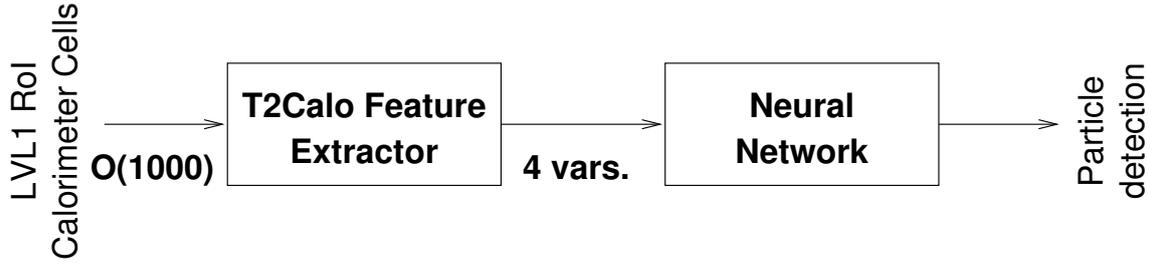


Figura 6.13: Um discriminador neural para as características definidas pelo T2Calo.

3. Multi-tarefas: Respeitando o projeto do LVL2, o *NeuralRinger* não deve ter construções que impossibilitem ou dificultem a sua utilização concorrente em múltiplas tarefas;
4. Configurabilidade: O *NeuralRinger* deve ser, dentro de determinadas medidas, configurável e flexível com relação à configuração da rede que será usada na detecção dos alvos. O método de treinamento a ser utilizado será a retropropagação de erros resiliente, tendo por base redes totalmente conectadas e sem re-alimentação.

O Apêndice C traz detalhes do projeto e implementação desta ferramenta.

### 6.2.1 Discriminação neural aplicada às saídas do T2Calo

É possível substituir o discriminador linear definido na Seção 6.1.5 por um discriminador baseado em uma rede neural tipo MLP como a da Figura B.3. A implementação deste novo sistema utilizando-se do pacote *NeuralRinger* segue o fluxo de treinamento foi detalhado na Figura 6.3. A Figura 6.13 contém um diagrama de blocos que ilustra tal sistema. Os conjuntos de treinamento e teste serão mantidos, o que simplificará a comparação das duas técnicas.

Uma vez que estamos utilizando o algoritmo de retropropagação de erros resiliente, o único parâmetro a ser otimizado é o número de neurônios escondidos. Para procedermos com a escolha, a seguinte heurística será utilizada: realizam-se 5 testes com um número fixo de neurônios na camada escondida. Em seguida, avalia-se o número de passos de treinamento para um nível de estabilização fixo do produto SP, o valor final do EMQ e do produto SP da rede. A média dos cinco valores é extraída,



Tabela 6.1: Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado na saída do T2Calo.

N	Passos	EMQ/teste	EMQ/treino	SP/teste	SP/treino
2	232 ± 54	0,261 ± 0,002	0,252 ± 0,003	1,543 ± 0,004	1,542 ± 0,003
3	232 ± 66	0,262 ± 0,002	0,253 ± 0,002	1,543 ± 0,002	1,543 ± 0,001
4	154 ± 42	0,261 ± 0,003	0,253 ± 0,003	1,541 ± 0,002	1,541 ± 0,002
5	206 ± 130	0,263 ± 0,003	0,254 ± 0,003	1,545 ± 0,001	1,545 ± 0,003
6	332 ± 138	0,258 ± 0,004	0,249 ± 0,004	1,545 ± 0,001	1,546 ± 0,003

assim como o desvio padrão destes valores em relação à média. O número de neurônios escondidos a ser considerado está no conjunto [2, 3, 4, 5, 6]. A Tabela 6.1 contém os resultados deste estudo. O valor do desvio padrão está anotado após o sinal “±” em cada entrada da tabela. Como é possível observar, com 4 neurônios escondidos o sistema tende a ser treinado em um número menor de passos. Os valores de EMQ e o produto SP encontram-se em faixas de valores muito próximas, sendo difícil sua comparação para fins de escolha. Desta forma, escolhe-se o sistema com 4 neurônios na camada escondida.

Em seguida, treina-se um total de 10 redes com os parâmetros selecionados. Para nos assegurarmos da estabilidade da rede, estabelece-se como critério de parada, uma variação menor que 0,0001 para o produto SP, por pelo menos 20 passos de treinamento. As Figuras 6.14 e 6.15 mostram, respectivamente, o desenvolvimento ao longo do treinamento, dos valores EMQ e do produto SP para um dos 10 resultados obtidos.

O sistema estabiliza após cerca de 40 passos, considerando a evolução do EMQ. No que tange à capacidade discriminante, aferida através do produto SP, após cerca de 60 passos, a qualidade discriminativa parece estar estável. A Figura 6.16 mostra as saídas da rede, no final do treinamento, para elétrons (em cima) e jatos (em baixo). A saída para elétrons é fixada em  $-1$  e para jatos, em  $+1$ . Nota-se que, diferente do sistema linear, há uma notável separação entre as classes, proporcionada pelo emprego de um discriminador neural. O ponto ótimo de separação pode ser mais facilmente determinado, inclusive por inspeção visual, o que torna este sistema

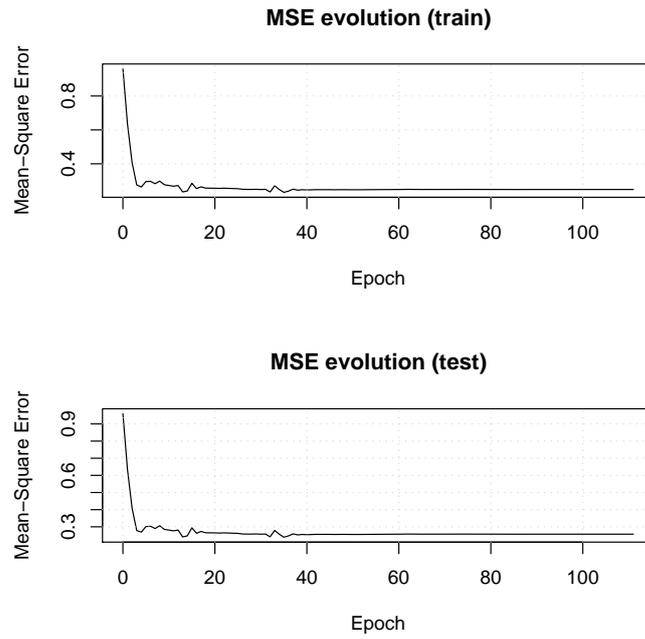


Figura 6.14: Evolução do EMQ para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do T2Calo.

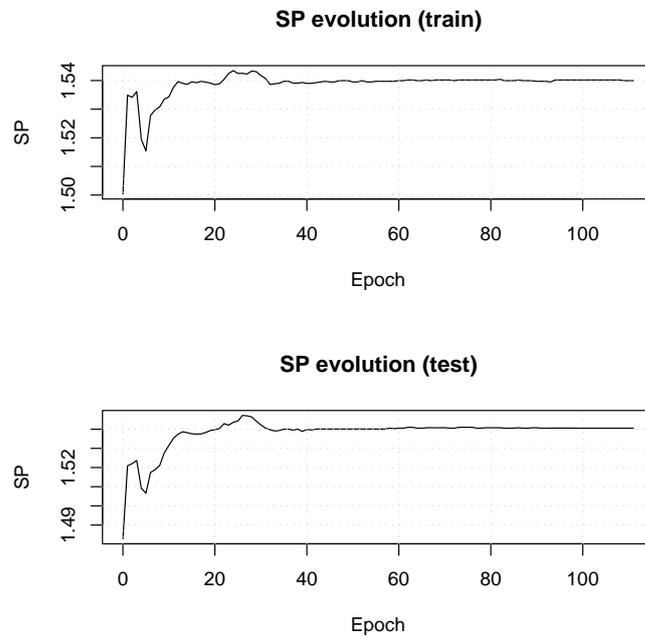


Figura 6.15: Evolução do produto SP para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do T2Calo.

mais robusto que os propostos anteriormente.

Diferentemente do sistema linear, uma rede neural poderá traçar uma superfície curva no espaço de entrada (número de dimensões = 4), melhorando a separação entre as classes que se deseja detectar. A Figura 6.16 mostra o resultado desta capacidade discriminativa. Na Figura 6.17 vemos uma comparação entre as técnicas de detecção vistas até aqui. Nota-se que o detetor baseado em uma rede neural com 4 neurônios escondidos exibe superior poder de detecção comparando-se às outras duas técnicas. A curva para este detetor está sempre acima da curva dos outros sistemas, indicando superior eficiência de detecção e rejeição de falso alarme.

O valor do produto SP máximo na curva para o detetor neural está nas coordenadas onde a detecção de elétrons é igual a 92,38% e o falso-alarme em jatos é apenas 9,05% ( $\approx 2,20$  kHz de falso-alarme). Estes valores correspondem a um produto SP de 1,54, tal qual ao caso do discriminador LMS. Uma verificação das eficiências parciais por região em  $\eta$  e em  $\phi$  revelam uma estrutura bastante semelhante aos sistemas de detecção precedentes, como mostra a Figura 6.18. A única ressalva é a na qualidade de detecção no *crack* dos calorímetros, onde o detetor neural mostra-se ligeiramente superior, indicando que este sistema consegue recuperar informações vitais à detecção que os outros dois métodos não conseguem.

A Figura 6.19 mostra o valor relativo do produto SP por faixa de energia transversa na seção e.m.. Uma comparação entre os 3 métodos do produto SP por faixas de energia descritos até agora se segue na Figura 6.20. Os dados nesta figura correspondem aos do conjunto de teste, para um corte efetuado em  $-0,243$  para o discriminador neural, e para os melhores resultados obtidos com os outros dois sistemas de detecção de forma equivalente.

É possível notar que o sistema neural ganha em praticamente todas as faixas energéticas, sobretudo onde existe maior concentração de eventos ( $e_T < 50\text{GeV}$ ). No final do espectro observa-se que o detetor neural possui eficiência de detecção igual a 100% para ambas as classes de partículas. Como colocado anteriormente, este resultado é questionável, em virtude da carência de dados nesta região. A Figura 6.21 mostra o estado final da rede neural, após o treinamento, a título ilustrativo.

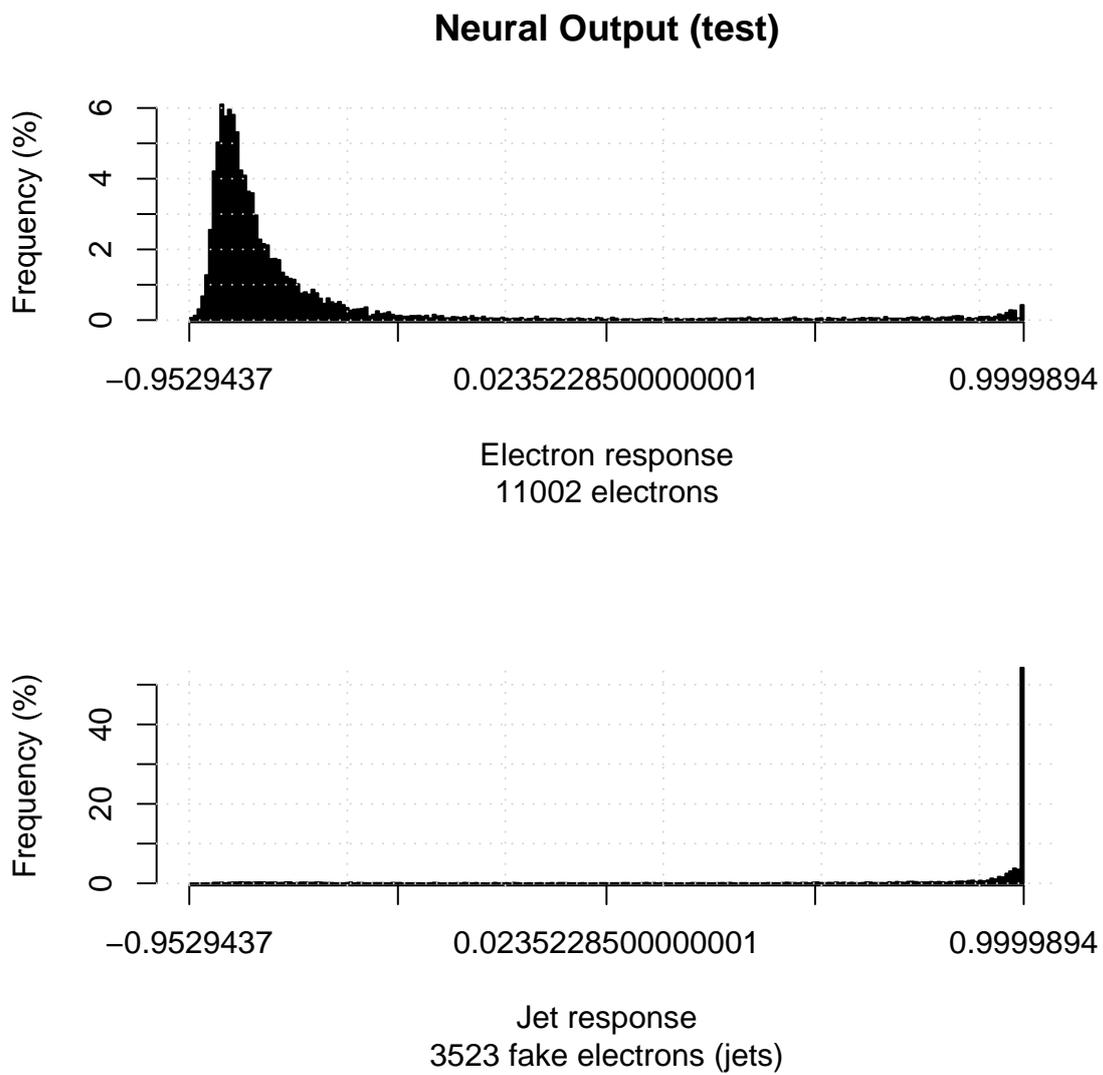


Figura 6.16: Saída para o conjunto de teste, do discriminador neural baseado nas características extraídas pelo T2Calo.

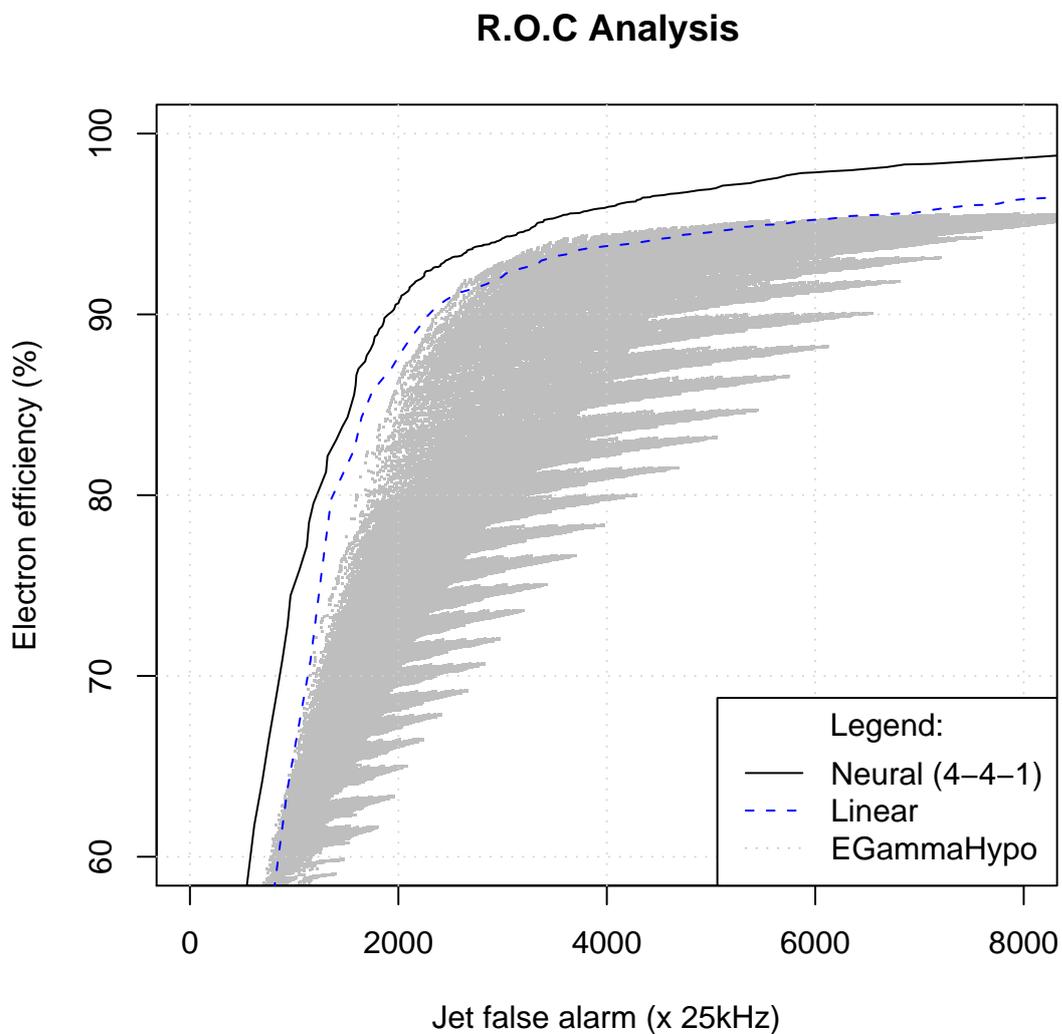
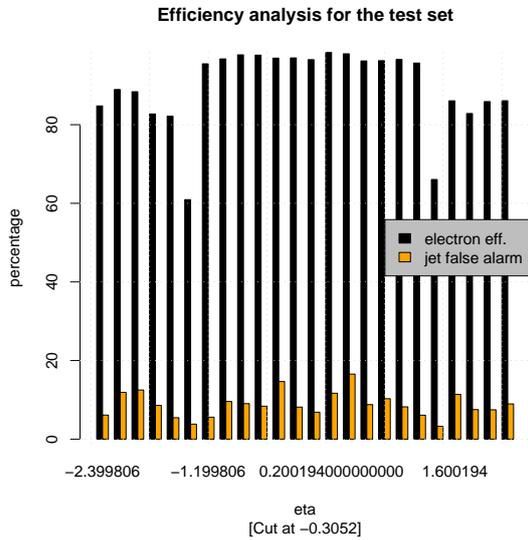
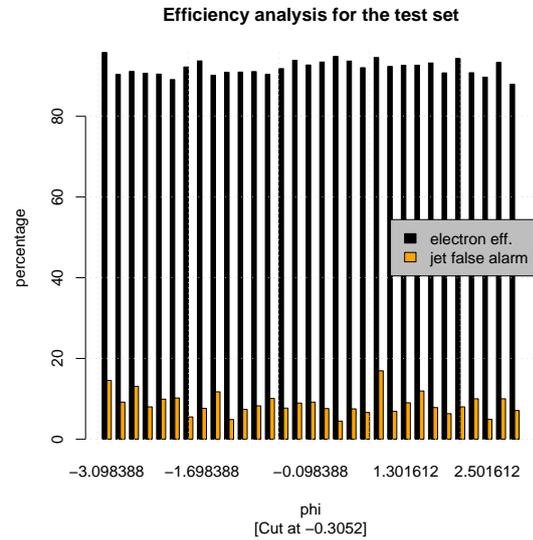


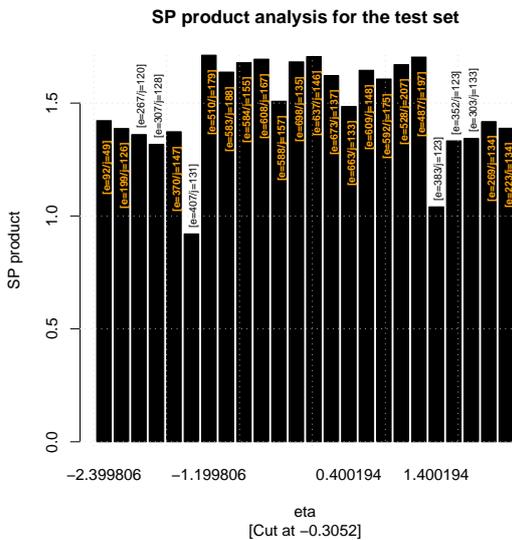
Figura 6.17: R.O.C. para o conjunto de teste, do discriminador neural baseado nas características extraídas pelo T2Calo, comparado com o detetor linear e a técnica de otimização baseado no algoritmo EGammaHypo.



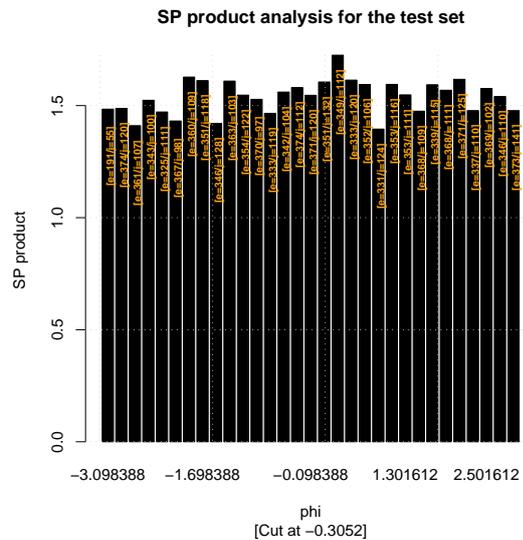
(a) Eficiência por  $\eta$



(b) Eficiência por  $\phi$

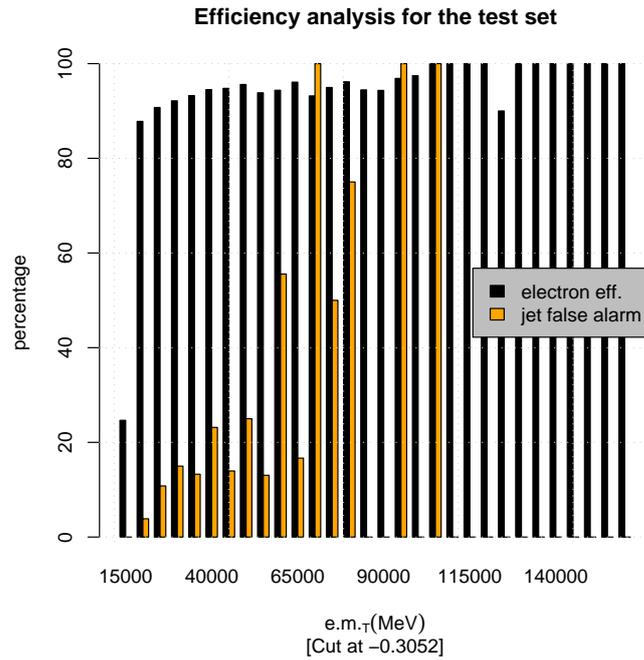


(c) Produto SP por  $\eta$

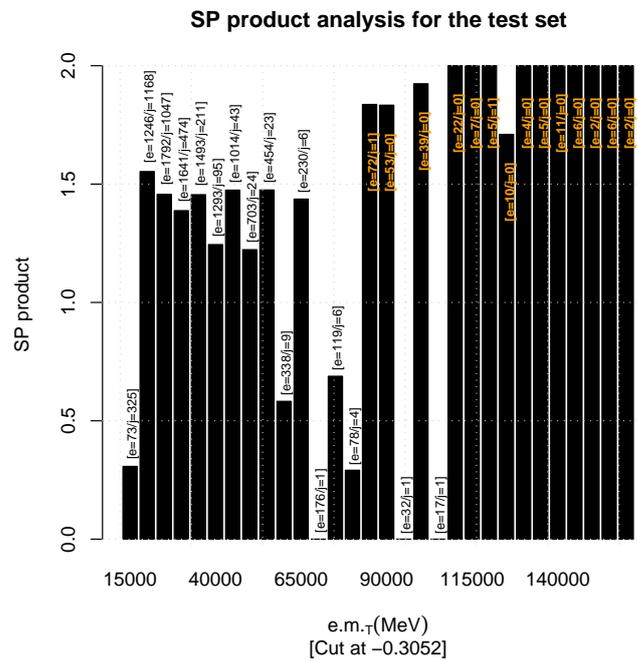


(d) Produto SP por  $\phi$

Figura 6.18: Análise da eficiência de classificação e do produto SP para o discriminador neural baseado nas saídas do T2Calo, para os dados do conjunto de teste ao longo de  $\eta$  em (a) e (c) e por  $\phi$ , em (b) e (d).



(a) Eficiência



(b) Produto SP

Figura 6.19: Análise da eficiência de detecção de elétrons e falso-alarme em jatos (a) e do produto SP (b) por  $E_{T_{3 \times 7}}^{e.m.}$  para o discriminador neural baseado nas saídas do T2Calo, tendo por base os dados do conjunto de teste.

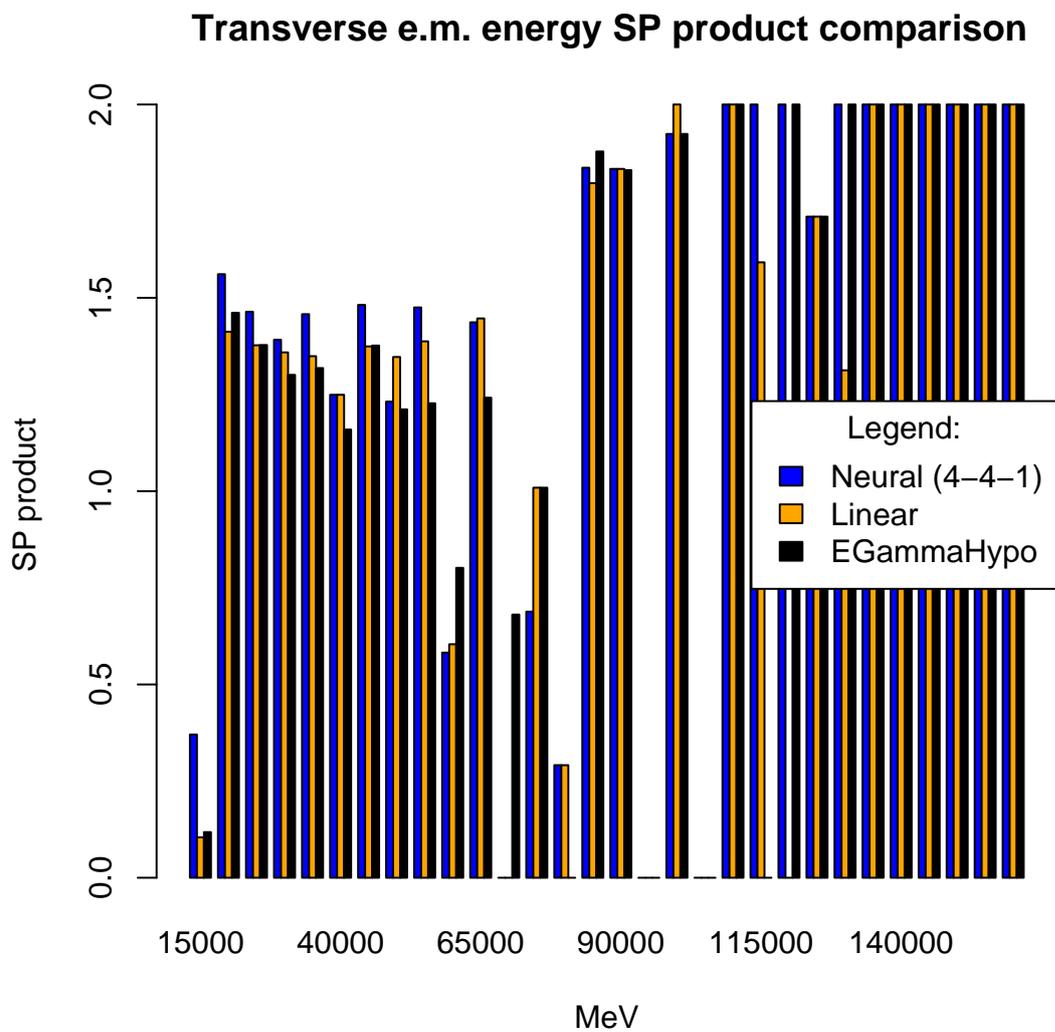


Figura 6.20: Comparação do produto SP para os 3 detetores abordados até aqui, por faixa de energia e.m. transversa.



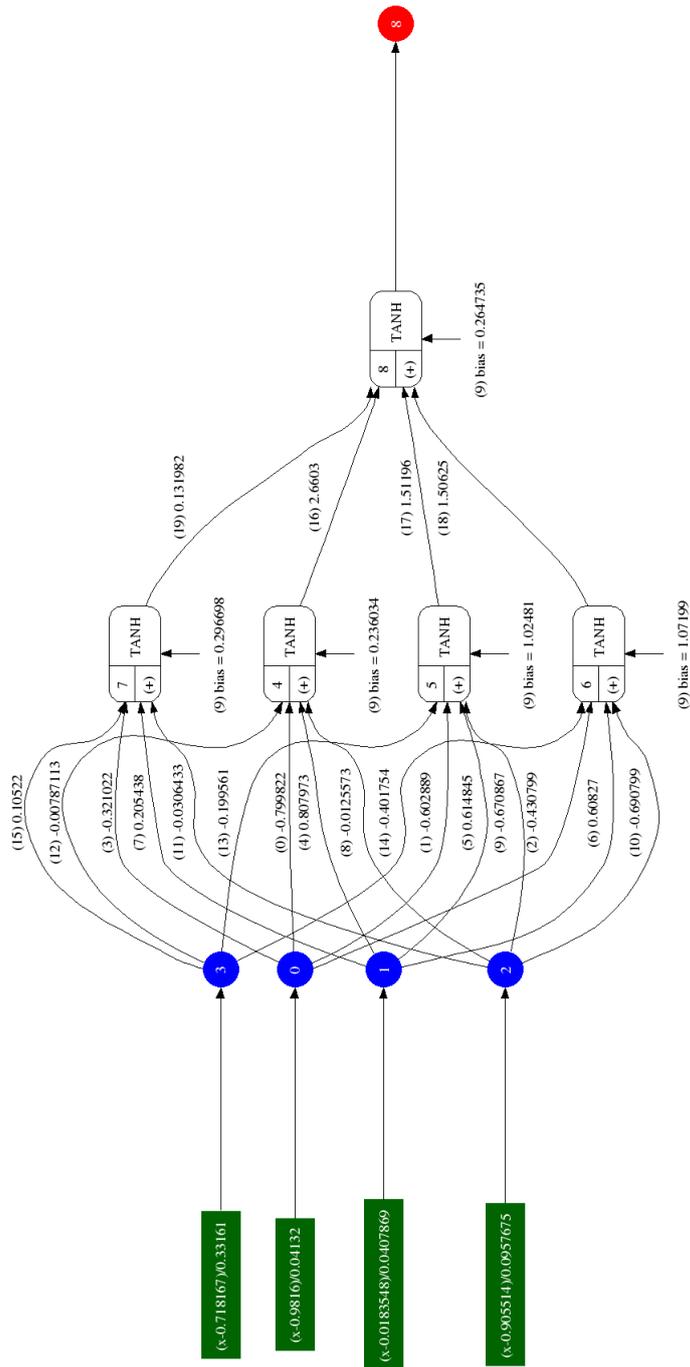


Figura 6.21: Diagrama de fluxo do sistema de deteção elétron/jato, neural, usando as características do T2Calo como entrada.

## 6.2.2 Adicionando mais variáveis

Na utilização de discriminação baseada em redes neurais para a Física de Altas Energias, observa-se uma tendência à re-análise das características utilizadas pelo sistema de detecção. Isto se deve a capacidade que redes neurais possuem em extrair correlações de ordem elevada entre as entradas de um detetor, atingindo desempenhos de discriminação superiores a sistemas baseados em cortes uni ou bi-dimensionais. Nestes casos, o especialista deve tomar cuidado em não povoar seu espaço de discriminação com muitas variáveis, o que tornaria o processo de decisão demasiado complicado. Uma vez que propõe-se o uso de redes neurais como sistema de discriminação, é possível, até certo ponto, ignorar esta limitação.

Como sub-produto do cálculo das 4 variáveis utilizadas pelo EGammaHypo, o T2Calo também avalia, dentre outras, a energia, camada a camada do calorímetro, depositada pela partícula na RoI sendo analisada. Estas variáveis, normalmente desprezadas para fins de filtragem, podem conter informação não originalmente aproveitada pelo EGammaHypo, mas que revele novos detalhes sobre a natureza da partícula. No total 10 outras variáveis estão disponíveis:

- Energias parciais nas 4 camadas da seção e.m. dos calorímetros;
- Energias parciais nas 4 camadas da seção hadrônica dos calorímetros;
- A largura do objeto, medida em função da distribuição energética na segunda camada e.m. dos calorímetros;
- A relação de deposição energética entre uma região de  $3 \times 7$  células na direção  $\eta \times \phi$  e uma região de  $7 \times 7$  células na segunda camada e.m..

Estas variáveis são usadas normalmente *offline* para a depuração do comportamento do T2Calo durante sua execução no sistema de filtragem. Enquanto combinar estas variáveis no algoritmo EGammaHypo seria difícil, o sistema de discriminação neural proposto na seção anterior poderia ser re-utilizado para montar-se um novo sistema de detecção baseado em um universo de variáveis maior e com mais informações. A Figura 6.22 ilustra este novo sistema, em blocos.

Para o treinamento deste novo detetor, faz-se necessária uma nova avaliação do número de neurônios ótimos a ser empregado na camada escondida do discriminador.

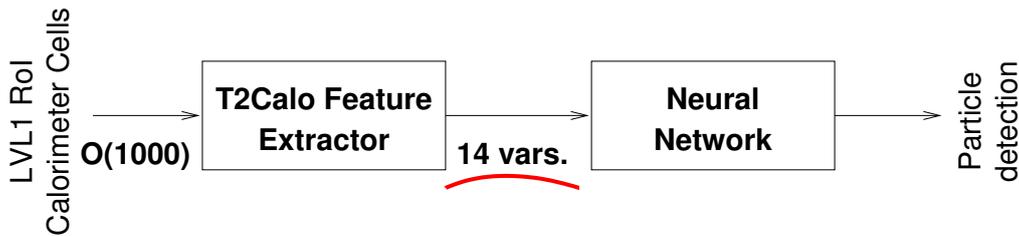


Figura 6.22: Diagrama em blocos de um sistema de detecção utilizando toda a informação produzida pelo T2Calo.

A heurística de otimização é usada mais uma vez: treinam-se 5 redes com o mesmo número de neurônios na camada escondida, com seus pesos sinápticos inicializados aleatoriamente a cada iteração. O número de neurônios escondidos foi variado de 2 a 12 (número de entradas), observando-se à média do produto SP, EMQ e o número de passos de treinamento obtido para cada grupo de testes. O desvio-padrão dos testes com relação à média do produto SP foi utilizado como medida de erro. A Tabela 6.2 contém os resultados desta heurística de otimização.

A partir de 6 neurônios na camada escondida, o sistema apresenta uma superior qualidade de detecção. Escolhe-se  $N = 10$  pois apresenta melhor relação entre o número de passos de treinamento e o produto SP final. Treinaram-se 10 redes neurais com a seguinte parametrização: 10 neurônios escondidos e 22000 eventos por época de aprendizado. A Figura 6.23 mostra a saída do melhor dos 10 testes com 10 neurônios na camada escondida. Este sistema apresenta uma saída ainda mais próxima do alvo desejado que aquele baseado em apenas 4 das variáveis do T2Calo. A evolução (e estagnação) do EMQ e do produto SP para a mesma rede pode ser vista nas Figuras 6.24 e 6.25 O sistema encontra-se completamente treinado após cerca de 1000 passos, quando é parado. O valor do produto SP máximo atingido para o conjunto de teste (1,63) é superior àquele para o sistema usando somente 4 das variáveis do T2Calo (1,54). Para este valor de produto SP, o discriminador possui uma eficiência de detecção de elétrons de 94,79% contra uma taxa de falso-alarme de apenas 7,93%.

A Figura 6.26 mostra a ROC dos sistemas vistos até agora. Como é possível distinguir, o sistema neural utilizando todas as variáveis produzidas pelo T2Calo apresenta uma qualidade superior de detecção para qualquer valor razoável de efi-

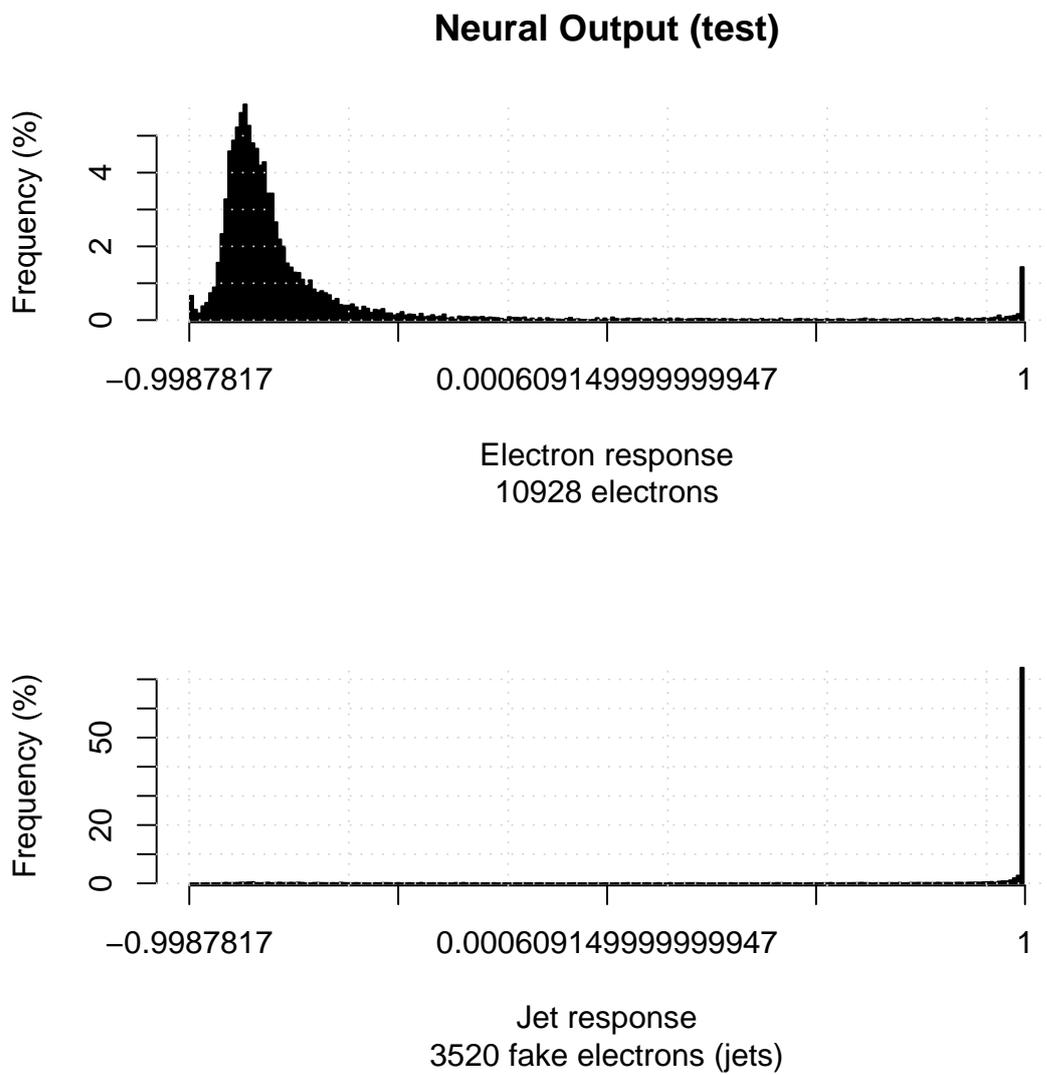


Figura 6.23: Saída para o conjunto de teste, do discriminador neural baseado nas 14 características extraídas pelo T2Calo.

Tabela 6.2: Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado em todas as variáveis produzidas pelo T2Calo.

N	Passos	EMQ/teste	EMQ/treino	SP/teste	SP/treino
2	410 ± 217	0,225 ± 0,003	0,228 ± 0,004	1,598 ± 0,004	1,585 ± 0,008
3	361 ± 153	0,221 ± 0,003	0,225 ± 0,003	1,606 ± 0,003	1,592 ± 0,003
4	541 ± 162	0,216 ± 0,003	0,218 ± 0,002	1,609 ± 0,005	1,604 ± 0,006
5	594 ± 111	0,215 ± 0,002	0,215 ± 0,002	1,612 ± 0,003	1,609 ± 0,004
6	557 ± 140	0,214 ± 0,004	0,217 ± 0,004	1,613 ± 0,004	1,607 ± 0,007
7	566 ± 270	0,210 ± 0,004	0,211 ± 0,004	1,618 ± 0,005	1,616 ± 0,003
8	670 ± 133	0,211 ± 0,004	0,212 ± 0,005	1,621 ± 0,007	1,617 ± 0,013
9	658 ± 180	0,211 ± 0,003	0,210 ± 0,004	1,616 ± 0,008	1,619 ± 0,008
10	567 ± 255	0,209 ± 0,004	0,210 ± 0,004	1,626 ± 0,006	1,622 ± 0,009
11	654 ± 245	0,205 ± 0,003	0,206 ± 0,002	1,628 ± 0,009	1,624 ± 0,003
12	683 ± 150	0,207 ± 0,003	0,209 ± 0,004	1,621 ± 0,010	1,618 ± 0,008

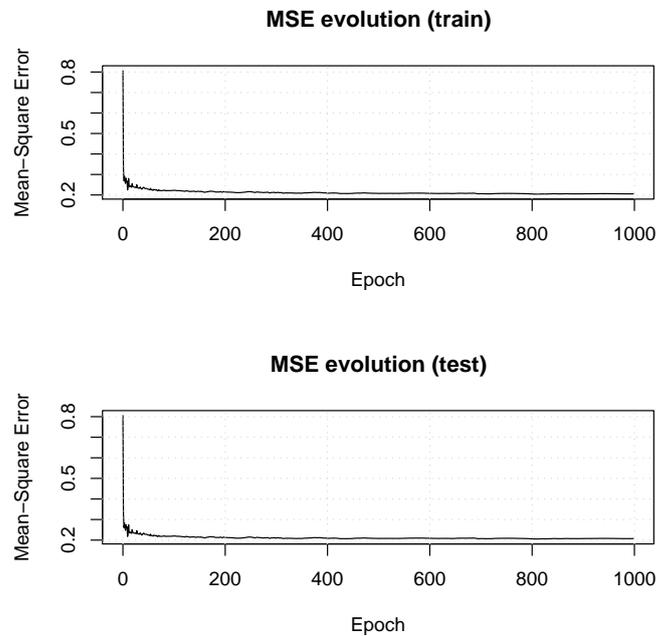


Figura 6.24: Evolução do EMQ para os conjuntos de treino e teste de um discriminador elétron/jato neural para as 14 variáveis extraídas pelo T2Calo.

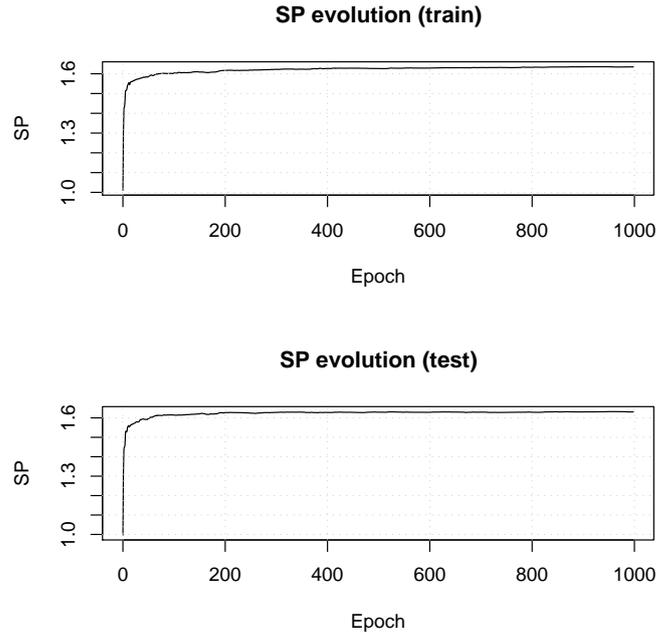


Figura 6.25: Evolução do produto SP para os conjuntos de treino e teste de um discriminador elétron/jato neural para as 14 variáveis extraídas pelo T2Calo.

ciência em elétrons (acima de 80%). Este resultado sugere que exista informação importante para o processo de detecção que é ignorada por processos de discriminação como o sugerido pelo EGammaHypo.

A Figura 6.27 mostra a eficiência de classificação deste novo sistema de detecção por  $\eta$  e  $\phi$ . Como é possível notar, este sistema recupera, quase que completamente, a perda de informação na região de *crack* do ATLAS, que podia ser claramente visível em detetores anteriores (veja novamente as Figuras 5.17, 6.9 e 6.18). Isto indica que a introdução de mais variáveis na detecção elétron/jato é benéfica no que tange à recuperação de informação, mesmo em regiões onde existe pouca informação disponível ou onde há descontinuidades no sistema de detecção.

A Figura 6.28 mostra uma comparação do produto SP por faixa energética para os dois sistemas neurais e o detetor linear. Como é possível verificar, este novo sistema apresenta uma qualidade de detecção muito superior no início e no final da faixa energética, onde os sistemas de detecção anteriores obtinham resultados bastante ruins, ou quase nulos. Nas demais faixas energéticas o desempenho do novo sistema continua superior aos dos demais detetores, ainda que em menor escala.

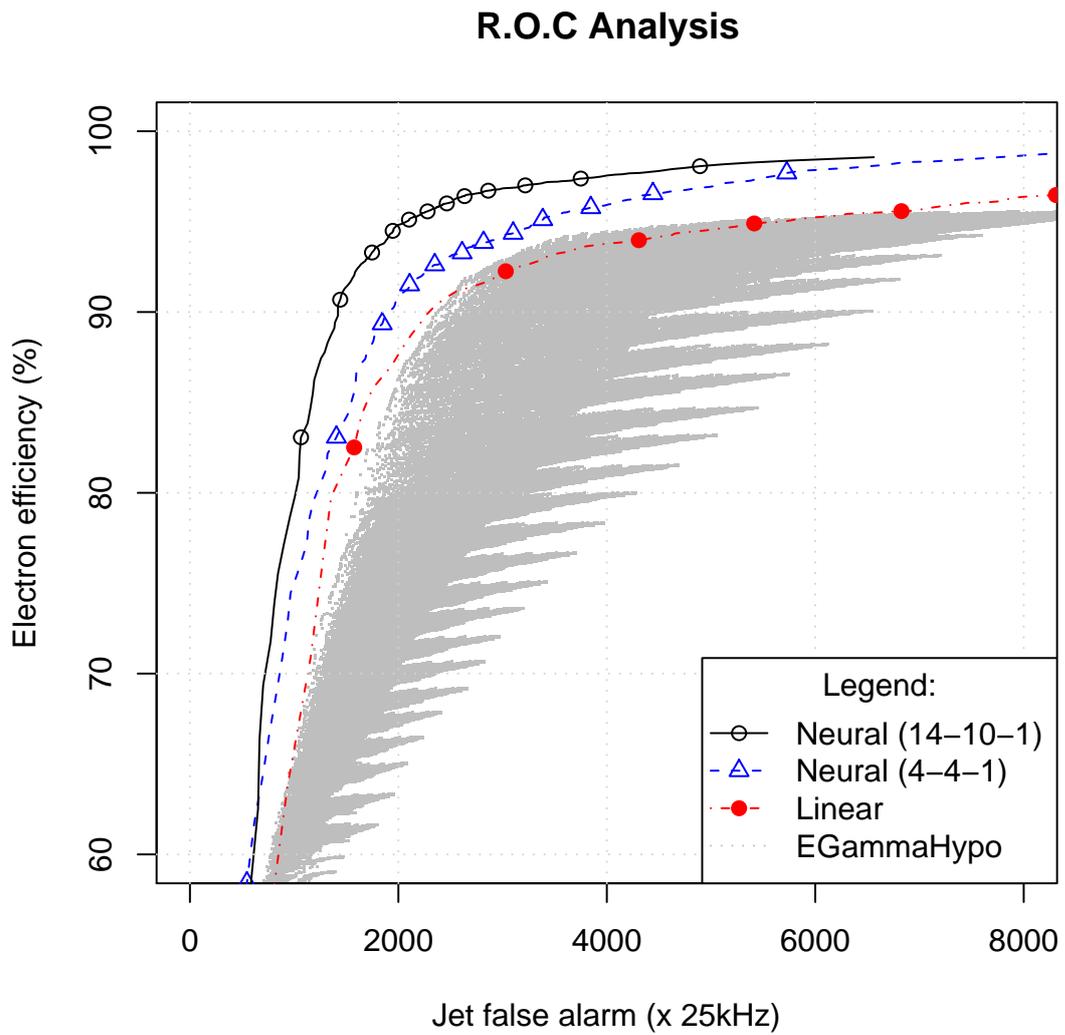
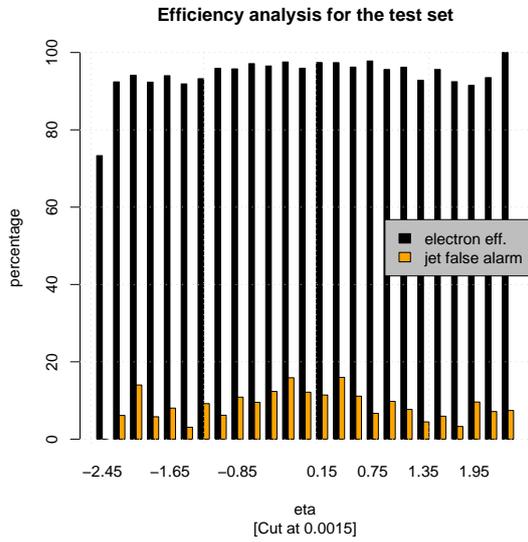
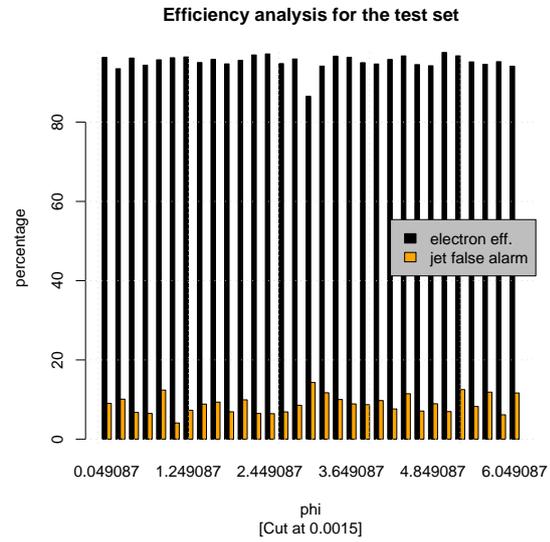


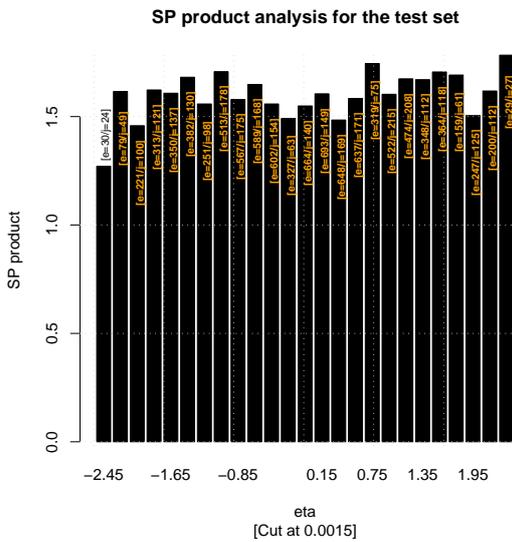
Figura 6.26: Comparação das ROC's para os 4 detetores abordados até aqui.



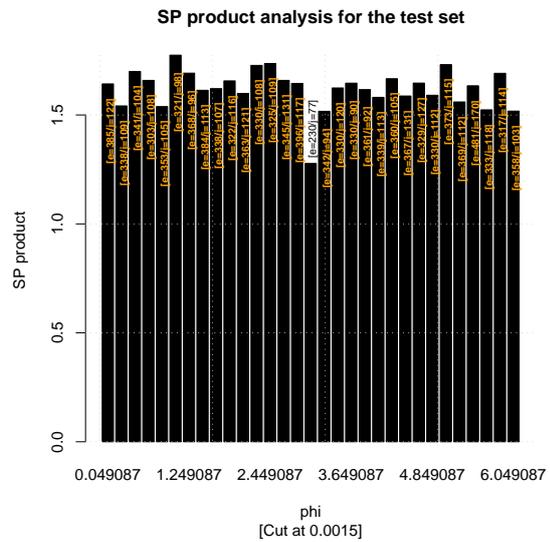
(a) Eficiência por  $\eta$



(b) Eficiência por  $\phi$



(c) Produto SP por  $\eta$



(d) Produto SP por  $\phi$

Figura 6.27: Análise da eficiência de classificação e do produto SP para um detetor neural baseado nas 14 saídas do T2Calo, para os dados do conjunto de teste ao longo de  $\eta$  em (a) e (c) e por  $\phi$ , em (b) e (d).



Outra tendência que pode ser observada é um aumento substancial na qualidade de generalização do sistema neural, agora dotado de mais informações, conseguindo distinguir elétrons de jatos onde outros sistemas falham, como é o caso da região energética entre 90 e 120 GeV. Este resultado indica que, apesar da carência de dados na região, a rede conseguiu aprender corretamente como detetar as partículas de interesse. Esta característica é de extrema importância para o ambiente do ATLAS, onde a Física de interesse poderá apresentar características desconhecidas e não-simuláveis utilizando as ferramentas disponíveis atualmente.

Embora o sistema de detecção utilizando as 14 variáveis do T2Calo apresente uma alternativa com qualidade de detecção superior e não equiparável por nenhum dos outros sistemas abordados até aqui, exige que uma grande quantidade de dados seja buscada no sistema de leitura do detetor. Os dados de todas as camadas dos calorímetros, para a RoI em análise, devem ser carregados no processador da L2PU para obter-se as 14 variáveis. Ademais, acredita-se que haja correlações diretas ou de ordem superior entre os valores das variáveis. Por exemplo, existe uma correlação entre a energia total na seção e.m. dos calorímetros com aquela depositada na segunda camada da seção e.m.. Se fosse possível determinar, dado o sistema de detecção empregado, quais variáveis são as mais relevantes ao processo de filtragem, seria possível desenvolver um detetor mais eficiente.

### **6.2.3 Relevância das 14 características do T2Calo**

A técnica da análise de relevância [78] tem por objetivo medir a importância de cada uma das variáveis de entrada para um classificador. Nesta técnica, avalia-se a contribuição da variável à composição da saída substituindo-se seu valor, a cada evento, pelo valor médio da variável com respeito a todos os eventos disponíveis na entrada do sistema de discriminação. Quanto maior a variação (maior é o valor absoluto da medida de relevância), mais relevante é considerada a variável. Assim sendo, observando-se a variação da saída, é possível estimar a contribuição daquela variável ao processo discriminatório.

A estimativa de relevância da  $i$ -ésima componente de entrada pode, então ser

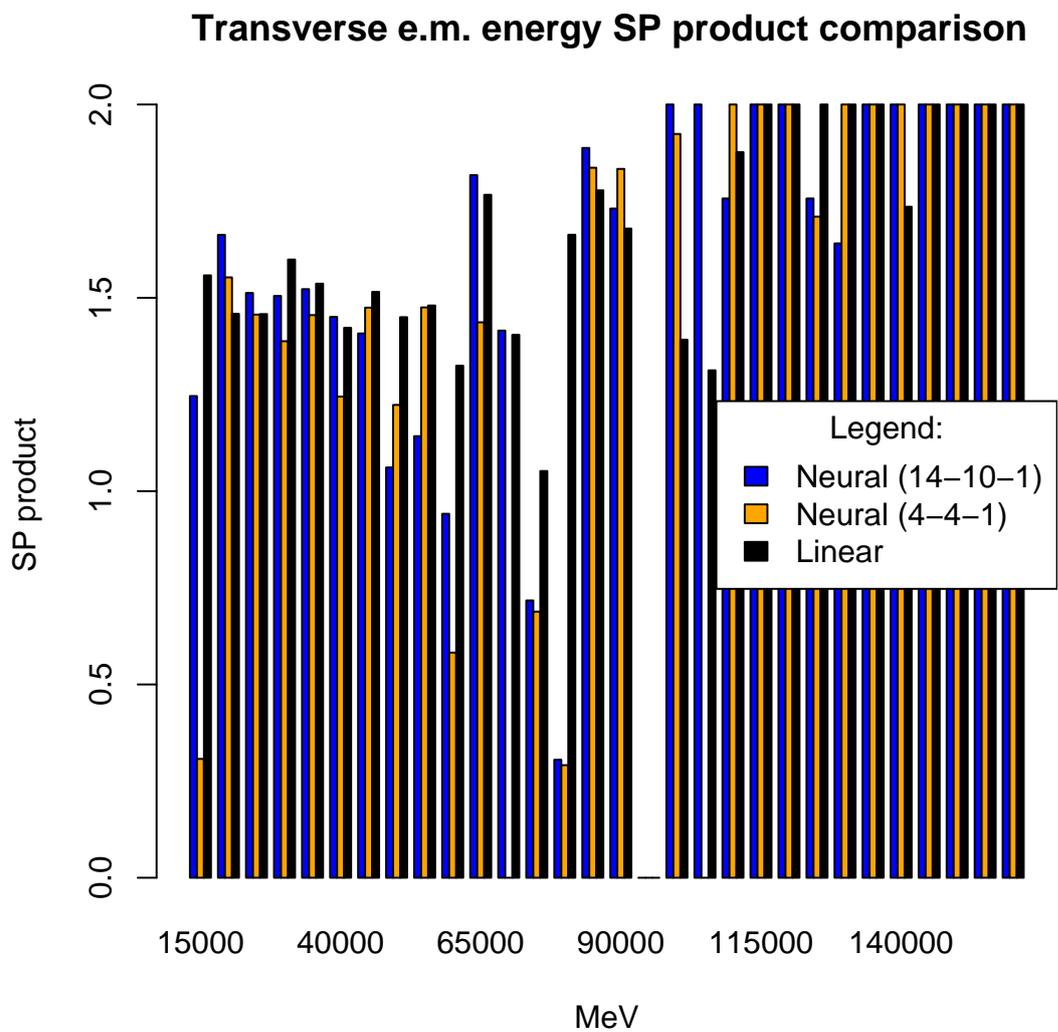


Figura 6.28: Comparação do produto SP máximo por faixa energética para os 2 detectores neurais abordados até aqui e o detector linear.

calculada pela fórmula:

$$R_i = \frac{1}{N} \sum_{j=1}^N [\text{saída}(\vec{x}_j) - \text{saída}(\vec{x}_j |_{x_{j,i}=\bar{x}_i})]^2 \quad (6.1)$$

Numa segunda interpretação, é possível definir a relevância da variável  $x_i$ ,  $R_i$ , como o EMQ da saída de um discriminador comparada à mesma saída quando faz-se a variável assumir o valor de sua média. Nessa equação,  $N$  representa o número total de eventos (padrões) disponíveis para o classificador.

A Figura 6.29 mostra os valores de relevância para cada uma das variáveis do classificador em análise, tendo por base os conjuntos de treino e teste. Esta figura mostra, mais uma vez, que o conjunto de treino permite generalizar o comportamento do discriminador, apresentando valores de relevância, para cada uma de suas variáveis, bastante próximos daqueles valores calculados para o conjunto de teste. Esta figura também indica que a variável que contém a energia total na seção hadrônica (T2Cahade) é de extrema importância na tarefa de discriminação, como é esperado na discriminação elétron/jato. Quando a substituímos pelo seu valor médio, o erro na saída aumenta significativamente, enquanto que, para as demais variáveis, o impacto é nitidamente menor.

Aproveitando a figura de mérito utilizada para o critério de parada do treinamento, o produto SP, seria possível desenvolver um critério de análise de relevância que fosse compatível com o potencial discriminante de um detetor. Ao invés de se calcular a variação do EMQ na saída do sistema, ao fazermos uma variável assumir seu valor médio, calcula-se a variação do produto SP na saída do detetor para as mesmas condições, assim descrita:

$$R_{d_i} = \max(\text{SP}_{\text{original}}) - \max(\text{SP}(\vec{x}_j |_{x_{j,i}=\bar{x}_i})) \quad (6.2)$$

Define-se então a relevância  $R_{d_i}$  da variável  $i$  como a diferença produzida no produto SP máximo no evento da substituição desta variável pelo seu valor médio. De fato, espera-se que a maior parte dos valores de relevância sejam positivos, indicando uma degradação do desempenho do classificador seguindo a neutralização de uma variável. Dada a natureza do cálculo, no entanto, é possível que se encontrem valores de relevância de discriminação negativos, indicando que uma variável esteja, na verdade, atrapalhando o processo de classificação ao invés de melhorá-lo.

### Relevance analysis (MSE)

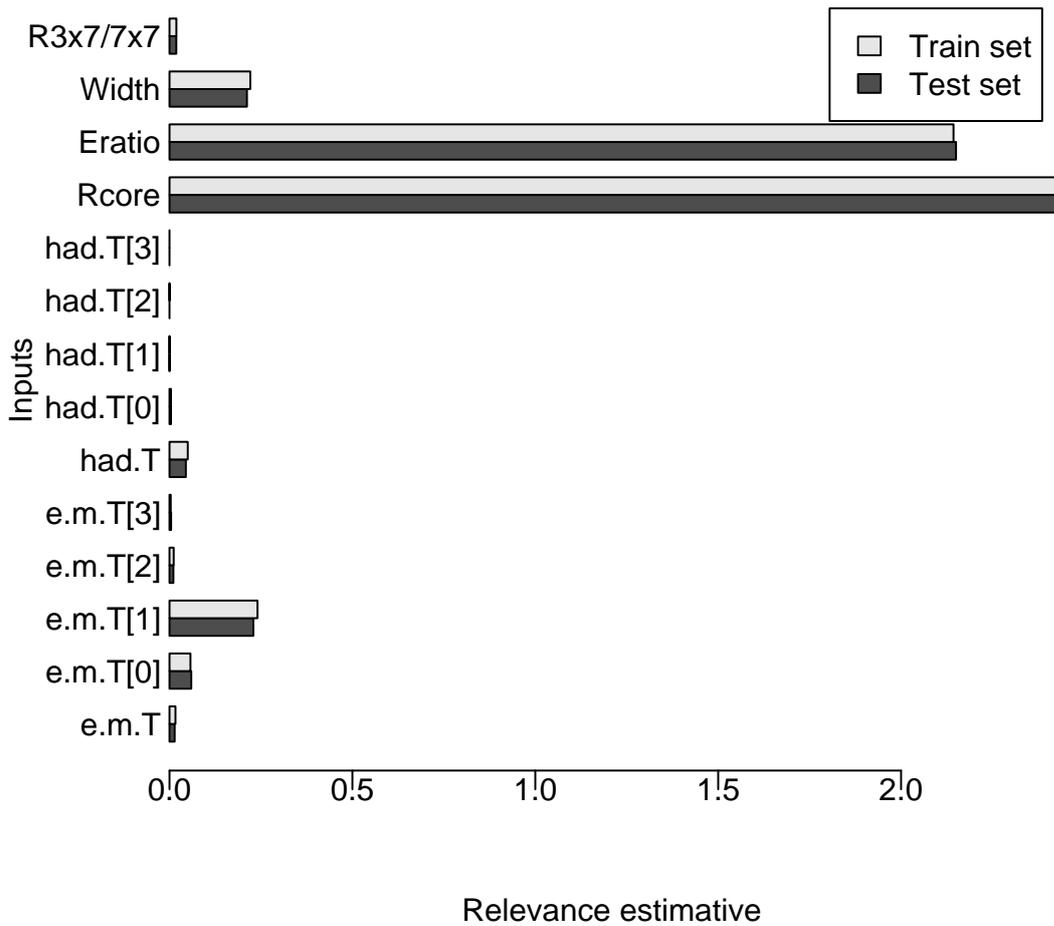


Figura 6.29: Os valores de relevância para os conjunto de treino e teste, para as 14 variáveis do T2Calo e considerando-se o classificador neural em estudo.

A Figura 6.30 mostra os valores de  $R_d$  para o classificador LMS em estudo. Como é possível ver nesta figura, o quadro é apenas marginalmente diferente daquele mostrado pela Figura 6.29. A variável mais importante para o processo discriminatório ainda é  $R_{e.m.2}^{3 \times 7 / 7 \times 7}$ , seguindo-se de  $R_{e.m.1}^{2\text{-máximos}}$ . É interessante notar que o processo de análise de relevância, baseado no conhecimento do classificador neural de melhor desempenho, indique a mesma ordem de separação que especialistas determinam para a filtragem baseada no EGammaHypo. O que escapa a este sistema de detecção é que existe informação relevante para a discriminação escondida em outras variáveis que não são contempladas pelo método, como a largura do agrupamento no calorímetro.

A análise de relevância pode ser utilizada para compactar a informação de entrada de um detetor de forma a obter melhores desempenhos tentando-se minimizar a perda de informação. Ao mesmo tempo, a análise de relevância pode ser usada como medida de robustez de um sistema de detecção: quanto mais distribuída a relevância de cada variável, menor a importância individual de cada variável ao processo e, portanto, menor será a perda na capacidade discriminante no caso da variável ser suprimida. A técnica de poda começa por eliminar as variáveis que apresentam menor valor de relevância, seguindo-se para aquelas com maior valor até que o desempenho de detecção apresente uma degradação inaceitável ao projeto.

Para o caso em estudo, embora as relevâncias baseadas no desvio do EMQ e do produto SP concordem, como se espera, em grande parte, na ordem de relevância, discordam sobre a importância das variáveis entre si. Por exemplo, a variável  $E_{T_{0,2 \times 0,2}}^{\text{HAD-1}}$  (had.T nos gráficos), parece apresentar uma importância relativa menor considerando-se a relevância baseada no desvio do EMQ do detetor. No caso da análise da relevância discriminativa desta variável, nota-se que ela está melhor destacada entre as 5 variáveis mais relevantes ao processo discriminatório. De fato, uma simples poda por patamar de corte nos dois casos, de forma a preservar as 5 variáveis mais relevantes, escolherá conjuntos ligeiramente diferentes se nos basearmos em um ou em outro método. Para averiguar qual dos métodos gerará um detetor mais eficiente, os dois casos são estudados:

- Poda baseada na relevância pelo EMQ, preservando 5 variáveis:  $R_{\text{core}}$ ,  $R_{e.m.1}^{2\text{-máximos}}$ , largura do chuveiro ( $Width$ ),  $E_{T_{3 \times 7}}^{e.m.}[0]$ ,  $E_{T_{3 \times 7}}^{e.m.}[1]$ ;
- Poda baseada na relevância discriminativa, preservando 5 variáveis:  $R_{\text{core}}$ ,

### Relevance analysis (SP product)

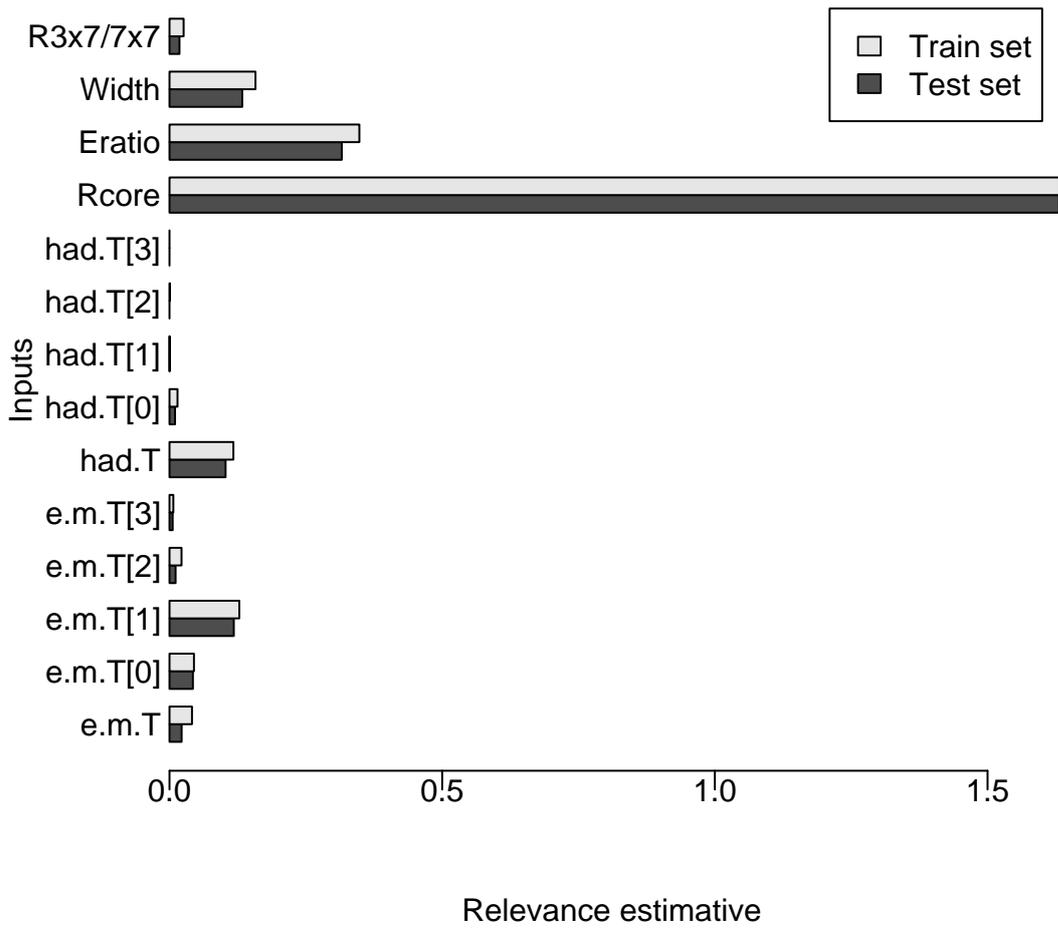


Figura 6.30: Os valores de relevância de discriminação para os conjunto de treino e teste, para as 14 variáveis do T2Calo e considerando-se o classificador neural em estudo.

$R_{e.m.1}^{2\text{-máximos}}$ , largura do chuveiro (*Width*),  $E_{T_{3\times 7}}^{e.m.}[1]$ ,  $E_{T_{0,2\times 0,2}}^{HAD-1}$ ;

Por inspeção visual, uma poda de mais de 5 variáveis começaria a cortar variáveis relevantes tanto ao mapeamento da entrada do sistema neural nos alvos quanto ao poder discriminante da rede. Treinam-se 10 redes neurais para cada um dos casos na listagem acima, observando-se os melhores resultados em ambos. A Figura 6.31 mostra curvas ROC para os dados de testes dos melhores detetores nos dois casos. Como é possível notar, há uma perda de eficiência de detecção do processo quando se podam 9 variáveis das 14 utilizadas pelo discriminador neural com melhor resultado. No entanto, a poda baseada na relevância discriminativa baseada na variação do produto SP produz melhores resultados, impedindo substancial degradação do desempenho de classificação, como é o caso da poda pela relevância clássica.

O classificador neural com 5 neurônios na camada de entrada, propiciado pela poda baseada na relevância de discriminação, atinge um valor máximo para o produto SP quando a eficiência de detecção de elétrons é igual a 93,94% contra 9,09% de falso-alarme na detecção de jatos. Este valor está somente cerca de 1 ponto percentual, na classificação elétrons, abaixo do valor obtido para um classificador baseado nas 14 variáveis, para deixar passar cerca de 1,4 ponto percentual a mais de ruído, embora conte com apenas um terço das variáveis originais. A Figura 6.32, mostra outra faceta da degradação da eficiência de detecção no espaço reduzido de variáveis gerado através da poda por relevância de discriminação. Como é possível notar através desta figura, as perdas relativas ao *crack* entre a tampa e o barril dos calorímetros pode ser novamente identificada.

Das 5 variáveis de entrada selecionadas pela poda baseada na relevância discriminativa, 3 correspondem às variáveis utilizadas pelo EGammaHypo, escolhidas por especialistas. A análise de relevância indica que a informação de largura do agrupamento seja mais importante à detecção que, por exemplo, a energia total na seção e.m. dos calorímetros e, portanto, esta variável é incluída no processo de classificação no lugar da última. A quinta variável de interesse é a energia na primeira camada da seção e.m., que ajuda a compensar a informação perdida ao removermos a energia total nesta seção. Este detetor é quase tão compacto quanto o sistema T2Calo+Rede-Neural (4-4-1) e exibe uma qualidade de discriminação compatível ou até superior (se consideramos valores de eficiência na detecção de elétrons em

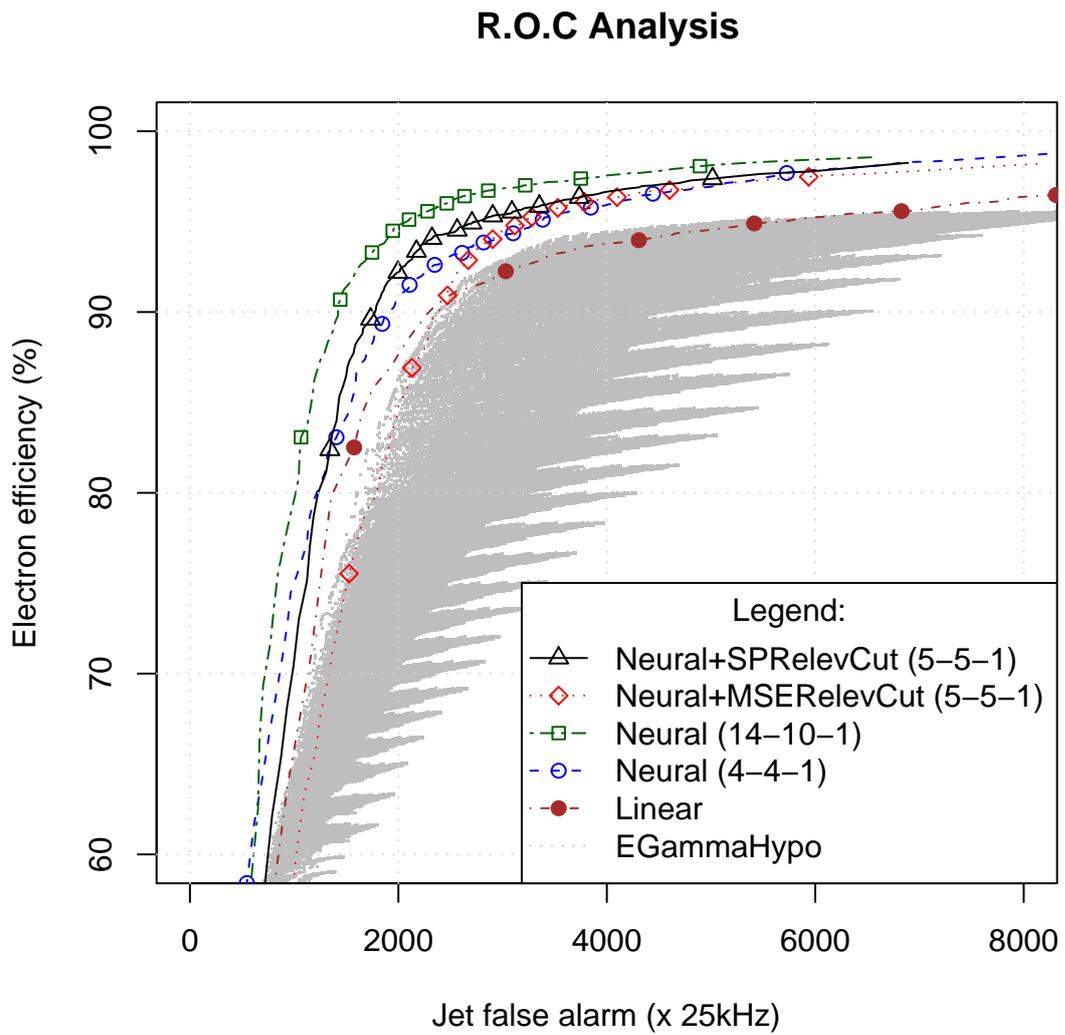


Figura 6.31: Comparação das ROC's para os 6 detetores abordados até aqui.



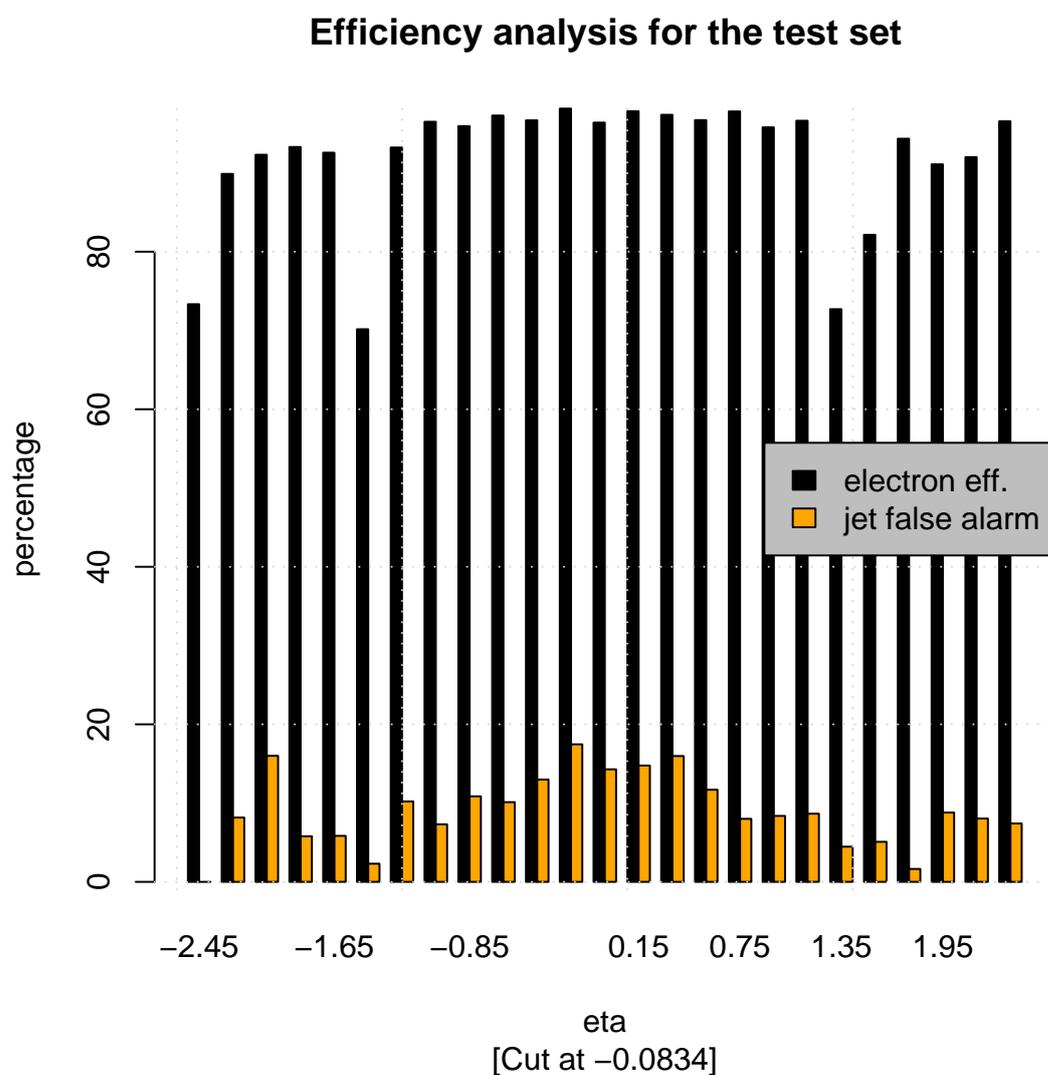


Figura 6.32: Eficiência de detecção e falso-alarme para o detetor baseado nas 5 mais relevantes (por discriminação) do espaço das 14 variáveis produzidas pelo T2Calo.

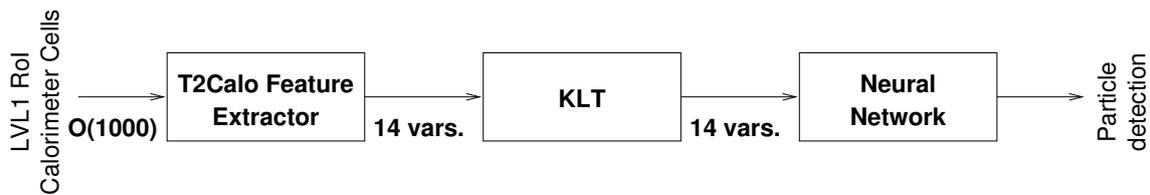


Figura 6.33: Um discriminador neural para as características definidas pela KLT das saídas do T2Calo.

patamares realísticos, i.e., acima de 80%) a este sistema. Ainda, poupa o cálculo das energias parciais na seção hadrônica e de algumas das camadas na seção eletromagnética. Este sistema também evita o carregamento dos dados de energia nas camadas 0 e 3 da seção e.m., sendo portanto mais econômico.

## 6.2.4 Análise de Componentes Principais

As variáveis utilizadas neste último e melhor sistema de discriminação possuem, por sua própria natureza, uma correlação bastante elevada. Por exemplo, há correlação entre os valores de energia totais e em cada camada de uma determinada seção do calorímetro. Tentando aproveitar este conhecimento, é possível encontrar, uma transformação linear (denotada por Transformação de Kharhunen-Loève ou KLT) que projeta um conjunto de variáveis em um segundo espaço onde cada componente seja descorrelacionada de cada uma das outra [75]. Espera-se que operando sob o espaço transformado, onde as variáveis são descorrelacionadas, processadores neurais consigam apresentar melhor qualidade de separação elétron/jato e que a compactabilidade deste novo sistema esteja mais evidente. A Figura 6.33 mostra o diagrama em blocos para este novo sistema de extração de características e detecção.

Para encontrarmos a KLT no problema da separação elétron/jato faz-se uso do método da covariância. A vantagem neste método é que é possível, a partir dos autovalores da matriz de covariância, fazer diretamente uma avaliação da energia de cada componente, o que facilitará o processo de compactação em seguida. Para a extração baseada no método da covariância remove-se o valor médio de cada variável de entrada baseando-se na média empírica obtida a partir das variáveis no conjunto de treinamento, de forma a obter-se um conjunto de entrada com média

nula. Considerando:

$$Z = [z_1, z_2, z_3, \dots, z_{14}] \quad (6.3)$$

como o vetor de entrada, onde  $z_n$  são os vetores-coluna contendo os dados para cada uma das observações em  $Z$ , e o vetor-linha  $E[Z]$  como representativo da média (ou esperança) de cada uma das colunas:

$$\mathbb{E}[Z] = [\mathbb{E}[z_1], \mathbb{E}[z_2], \mathbb{E}[z_3], \dots, \mathbb{E}[z_{14}]] \quad (6.4)$$

então podemos definir a matriz  $X$ , que representa  $Z$  com o valor médio de cada coluna extraído:

$$X = Z - [1, 1, 1, \dots, 1] \cdot \mathbb{E}[Z] \quad (6.5)$$

Para encontramos a KLT de  $X$ , deseja-se encontrar a transformação linear  $P$  de tal forma que:

$$Y = P^T \cdot X \quad (6.6)$$

com a restrição que  $\text{cov}(Y)$ , a covariância da matriz de dados transformados  $Y$ , seja uma diagonal, isto é, não haja correlações entre as variáveis do espaço transformado, e que  $P^{-1} = P^T$ . Por substituição:

$$\text{cov}[Y] = \mathbb{E}[Y \otimes Y] = \mathbb{E}[YY^*] \quad (6.7)$$

$$= \mathbb{E}[YY^T] = \mathbb{E}[(P^T X)(P^T X)^T] \quad (6.8)$$

$$= \mathbb{E}[(P^T X)(X^T P)] = P^T \mathbb{E}[XX^T] P \quad (6.9)$$

$$= P^T \text{cov}[X] P \quad (6.10)$$

Desta forma temos:

$$P \text{cov}[Y] = P P^T \text{cov}(X) P = \text{cov}[X] P \quad (6.11)$$

de onde finalmente, por inspeção, nota-se que, se  $\text{cov}[Y]$  é uma matriz diagonal, como se requer, então a Equação 6.11 assemelha-se a definição de auto-valores e auto-vetores:

$$V \cdot D = C \cdot V \quad (6.12)$$

onde  $V$  e  $D$  representam, respectivamente, a matriz de auto-vetores e a matriz de auto-valores de  $C$ . Uma vez que a matriz de auto-vetores representa uma transformação ortonormal e que a  $P^{-1} = P^T$ , então tem-se que a matriz de auto-vetores de  $\text{cov}[X]$  seja a procurada KLT.

Para projetar uma observação (linha) do conjunto de dados  $Z$ , remove-se, desta observação, o valor médio das variáveis de  $Z$ , em seguida multiplicando-se o resultado pelos auto-vetores de  $\text{cov}[X]$ . Os auto-valores de  $\text{cov}[Y]$  são utilizados para organizar os auto-vetores de  $\text{cov}[X]$  do mais (maior auto-valor) para o menos energético, propondo uma forma intuitiva de compactação do espaço de variáveis transformado. Quanto maior a energia (maior o auto-valor), mais relevante a variável seria ao processo de detecção. Esta análise por energia, conduzida utilizando-se os auto-valores da matriz de covariância de  $X$ , é denominada Análise de Componentes Principais.

A Figura 6.34 mostra a correlação normalizada entre as 14 variáveis do T2Calo, com seu valor médio removido. Como é possível ver, muitas das variáveis estão correlacionadas e portanto a Análise de Componentes Principais pode se mostrar benéfica na procura de um discriminador mais eficiente e compacto.

Para a Análise de Componentes Principais, calculou-se o valor médio das 14 variáveis baseando-se exclusivamente nos dados do conjunto de treinamento, utilizados também nos outros sistemas de detecção abordados até aqui. Em seguida, através do método da covariância definido anteriormente, extraiu-se a matriz de auto-vetores (KLT) e auto-valores (energia) baseando-se igualmente na matriz de covariância dos dados de treinamento. A Figura 6.35 mostra a curva de carga em energia baseando-se no vetor de auto-valores normalizado por sua soma. Como é possível ver, cerca de 90% da energia total disponível é atingida com a utilização de apenas 7 das 14 variáveis no espaço transformado. As últimas 3 componentes parecem não apresentar nenhum adicional energético ao espaço de entrada. Para uma detecção baseada em apenas 5 das variáveis, como foi o caso da compressão atingida empregando-se

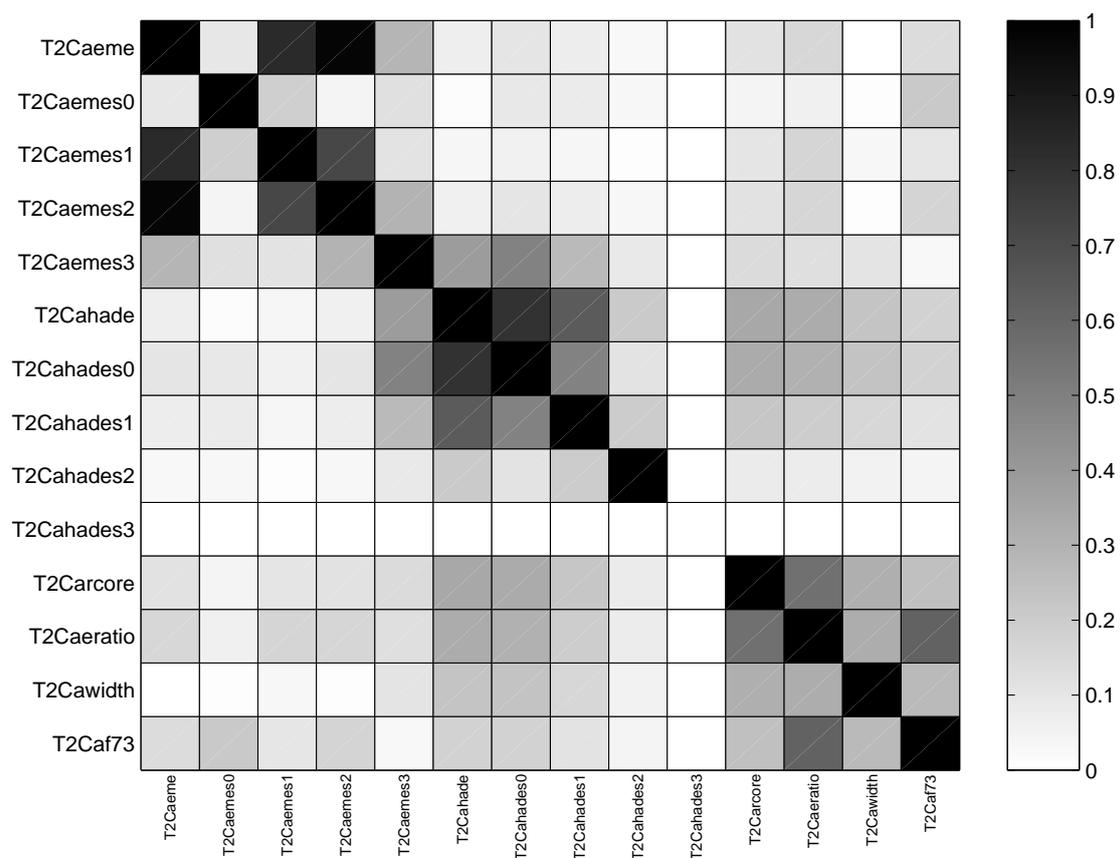


Figura 6.34: Correlação (normalizada) entre as 14 variáveis do T2Calo.

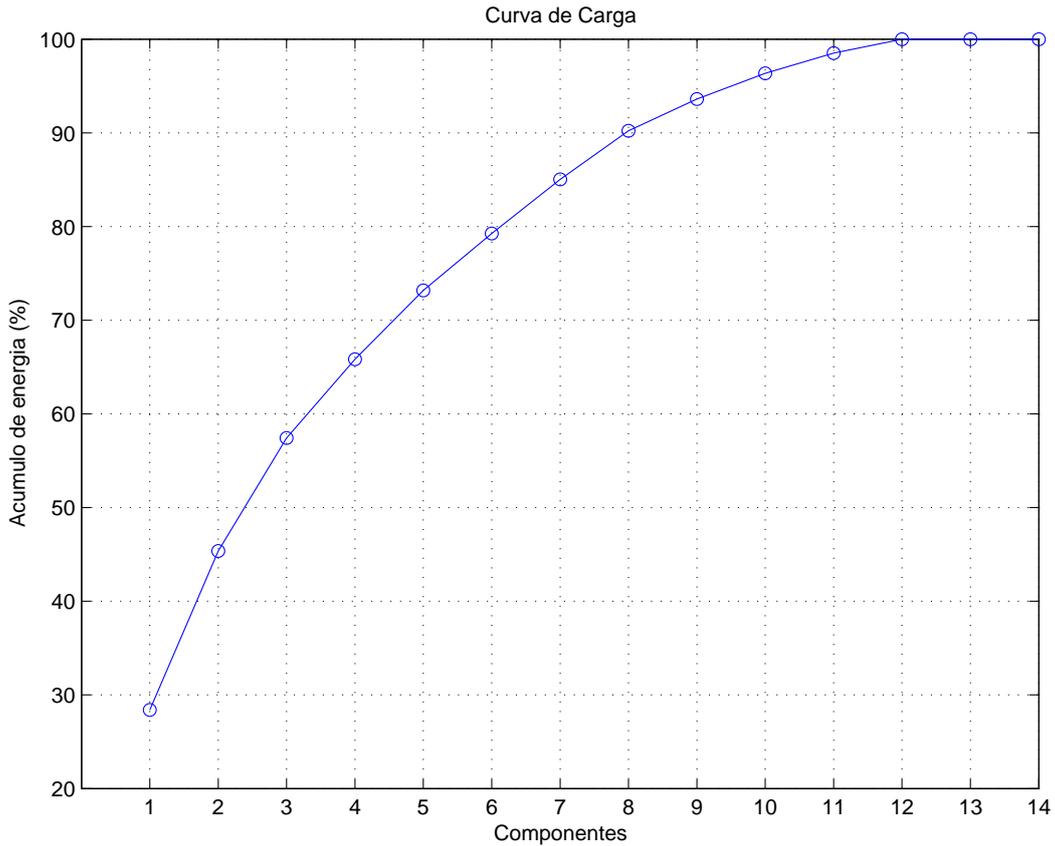


Figura 6.35: A curva de carga de energia para as 14 variáveis do novo espaço criado pela KLT.

o método da compactação por relevância de discriminação na Seção 6.2.3, apenas cerca de 73% da energia total no espaço transformado seria considerada.

### Discriminação baseada na Análise de Componentes Principais

Tendo por base as 14 novas variáveis do espaço provido pela PCA, é possível definirmos um detetor neural que discrimine elétrons de jatos. Desta forma, uma avaliação do número ideal de neurônios escondidos faz-se necessária. O mesmo procedimento de outras seções foi tomado: treina-se 5 redes com um número de neurônios escondidos fixo, avaliando-se a variação da eficiência de discriminação do detetor e o tempo de treinamento até que seja automaticamente interrompido. A Tabela 6.3 mostra os resultados desta otimização.

A partir desta tabela, observa-se que o número de neurônios ótimo está entre 10 e 12. Também nota-se que, comparativamente aos resultados da Tabela 6.2, as médias

Tabela 6.3: Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado na aplicação da KLT (PCA) nas 14 variáveis originais do T2Calo.

N	Passos	EMQ/teste	EMQ/treino	SP/teste	SP/treino
2	239 ± 93	0,223 ± 0,004	0,228 ± 0,004	1,601 ± 0,004	1,587 ± 0,004
3	239 ± 93	0,220 ± 0,004	0,224 ± 0,004	1,605 ± 0,010	1,595 ± 0,012
4	270 ± 81	0,213 ± 0,003	0,216 ± 0,005	1,614 ± 0,007	1,605 ± 0,010
5	281 ± 92	0,213 ± 0,004	0,216 ± 0,004	1,612 ± 0,006	1,604 ± 0,010
6	280 ± 85	0,213 ± 0,004	0,216 ± 0,007	1,617 ± 0,013	1,606 ± 0,011
7	450 ± 181	0,209 ± 0,004	0,212 ± 0,004	1,625 ± 0,013	1,612 ± 0,010
8	327 ± 171	0,211 ± 0,004	0,213 ± 0,003	1,617 ± 0,008	1,612 ± 0,007
9	350 ± 173	0,209 ± 0,002	0,211 ± 0,003	1,622 ± 0,006	1,613 ± 0,010
10	315 ± 83	0,206 ± 0,004	0,207 ± 0,004	1,630 ± 0,008	1,624 ± 0,009
11	558 ± 250	0,206 ± 0,007	0,206 ± 0,007	1,631 ± 0,015	1,627 ± 0,019
12	381 ± 102	0,204 ± 0,002	0,205 ± 0,002	1,641 ± 0,013	1,630 ± 0,010

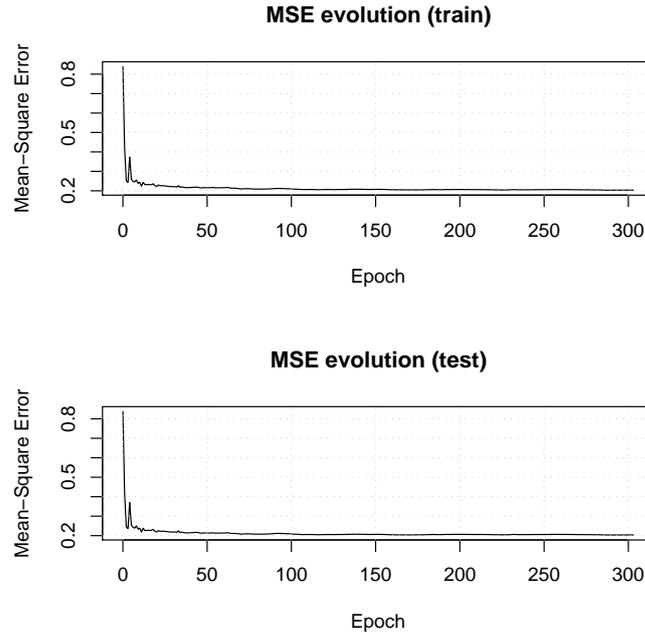


Figura 6.36: Evolução do EMQ ao longo do treinamento para um classificador baseado na KLT das 14 variáveis extraídas pelo T2Calo.

de detecção estão ligeiramente mais altas, atingindo 1,641 no máximo, comparado à 1,628 no caso das variáveis originais do T2Calo. Escolhe-se  $N = 12$ , treinando-se 10 redes neurais para testarmos se estas tendências são mantidas. Este exercício é realizado com sucesso.

As Figuras 6.36 e 6.37 mostram a evolução do EMQ e do produto SP ao longo do treinamento, que neste caso dura cerca de 300 passos. Um bom grau de concordância pode ser observado entre os dois conjuntos de dados. A ROC deste classificador pode ser vista, em comparação com aquelas do melhor detetor baseado nas 14 variáveis originais do T2Calo, dos detetores neural e linear baseado em apenas 4 variáveis do T2Calo e do sistema de detecção proposto pelo EGammaHypo na Figura 6.38.

O ponto onde o produto SP é máximo ( $= 1,64$ ), no caso do detetor em estudo, apresenta uma eficiência para a detecção de elétrons de 94,75% contra 7,59% de falso-alarme. Este novo sistema de detecção possui uma característica de detecção bastante próxima do sistema baseado diretamente na saída do T2Calo, usando 14 variáveis. A Figura 6.39 mostra uma varredura da eficiência de detecção e do falso-alarme por  $\eta$  e  $\phi$ . Neste novo sistema, volta-se a perceber uma ligeira queda de eficiência



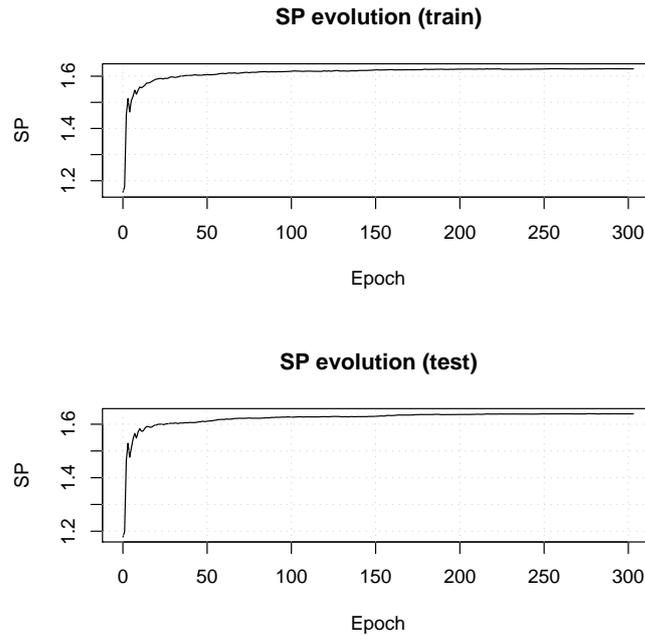


Figura 6.37: Evolução do Produto SP ao longo do treinamento para um classificador baseado na KLT das 14 variáveis extraídas pelo T2Calo.

na deteção de elétrons próximo à região do *crack*, que não era perceptível sem a aplicação da transformação. Uma vez que a Análise de Componentes Principais nos permite descorrelacionar totalmente a informação no espaço projetado, e levando-se em conta que a eficiência de deteção não aumenta significativamente a partir da utilização da KLT, conclui-se que existem correlações de ordem mais elevada entre as 14 variáveis do T2Calo. Por esta razão, a transformação linear provida por este método não melhora significativamente a qualidade de deteção.

### Compactação baseada em PCA

A Análise de Componentes Principais oferece um conjunto de ferramentas onde a compactação do espaço transformado pode ser feita de forma natural. Ainda que o detetor baseado na KLT não tenha mostrado ganho em eficiência de deteção, é possível utilizar a informação de energia dos autovalores da matriz de correlação para avaliar possíveis ganhos na compactabilidade do espaço das novas 14 variáveis. A Figura 6.35, mostrada anteriormente, contém a curva de carga em energia das componentes geradas por PCA. Um fato interessante é que, com o aumento do índice

### R.O.C Analysis

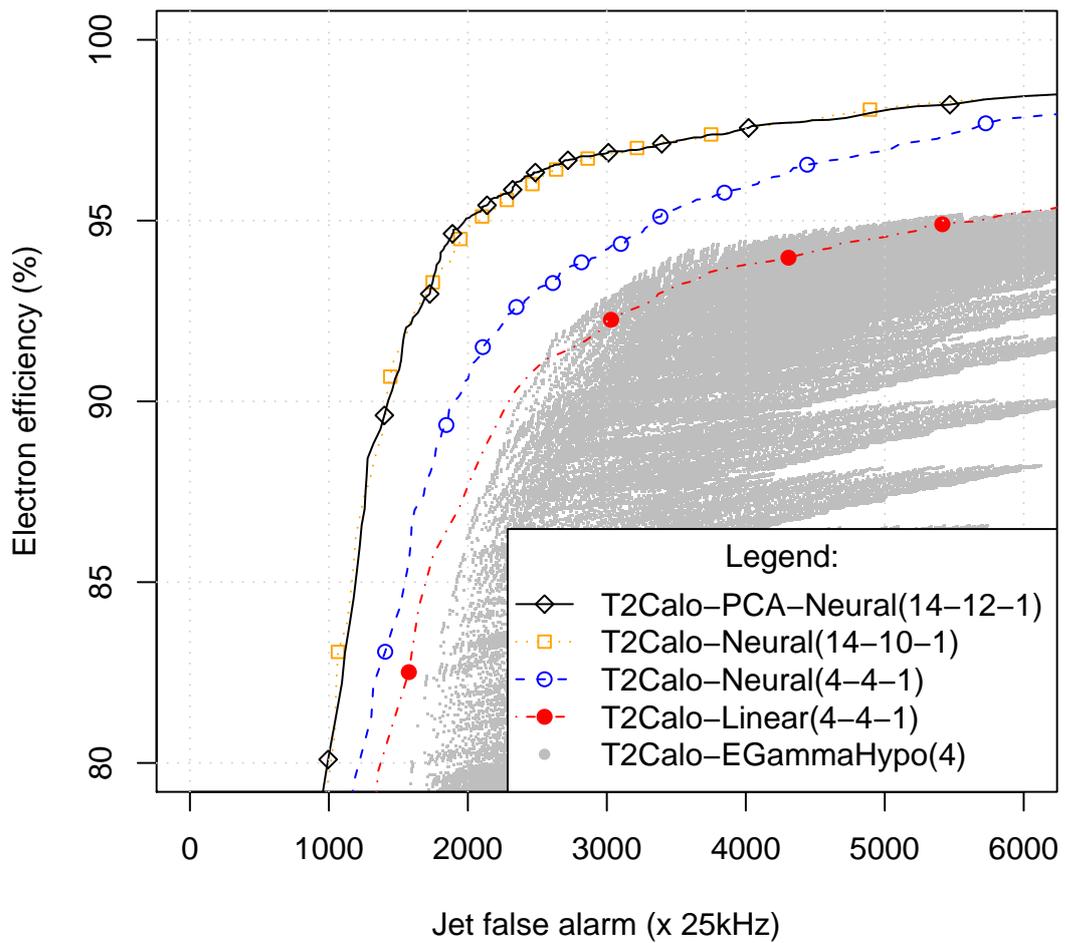
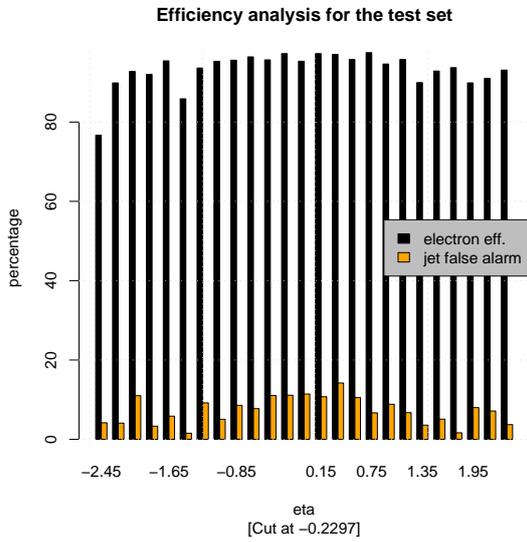
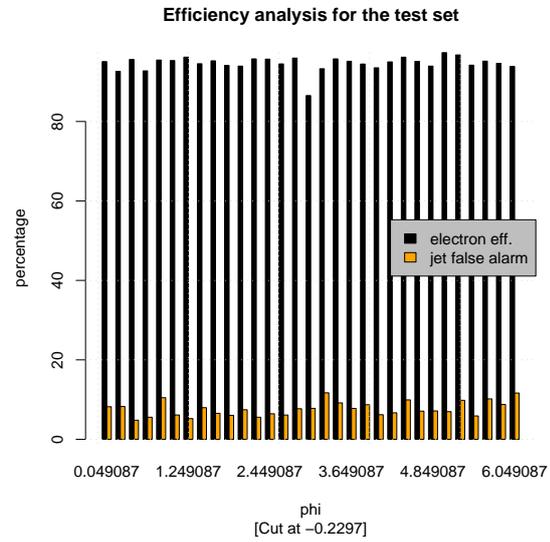


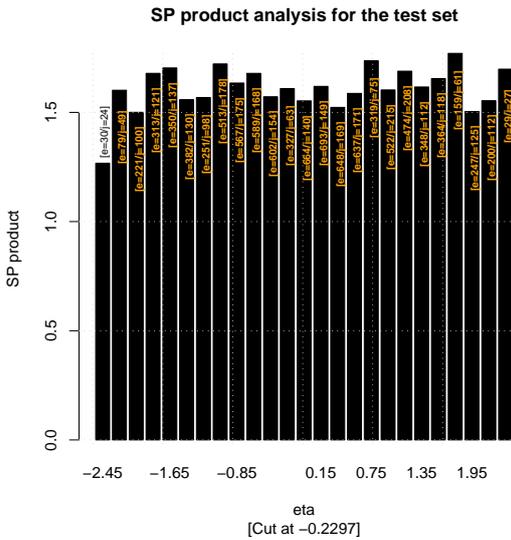
Figura 6.38: Curva ROC comparativa entre todos os métodos de detecção avaliados até este momento e incluindo a detecção baseada em PCA.



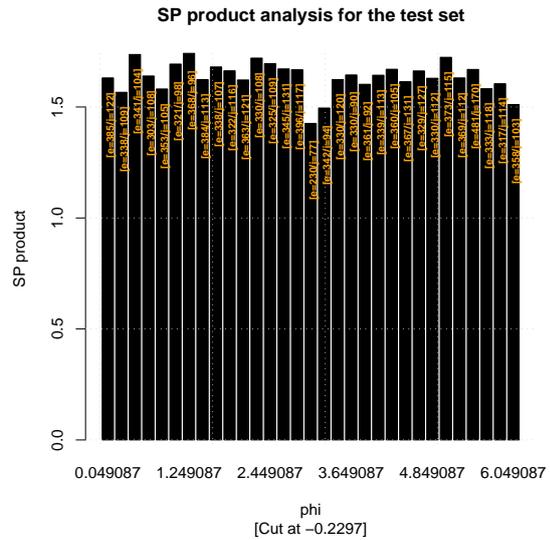
(a) Eficiência por  $\eta$



(b) Eficiência por  $\phi$



(c) Produto SP por  $\eta$



(d) Produto SP por  $\phi$

Figura 6.39: Análise da eficiência de classificação e do produto SP para o detetor elétron/jato baseado na transformação linear (via KLT) do espaço das 14 variáveis produzidas pelo T2Calo, para os dados do conjunto de teste ao longo de  $\eta$  em (a) e (c) e por  $\phi$ , em (b) e (d).

## Relevance analysis (SP product)

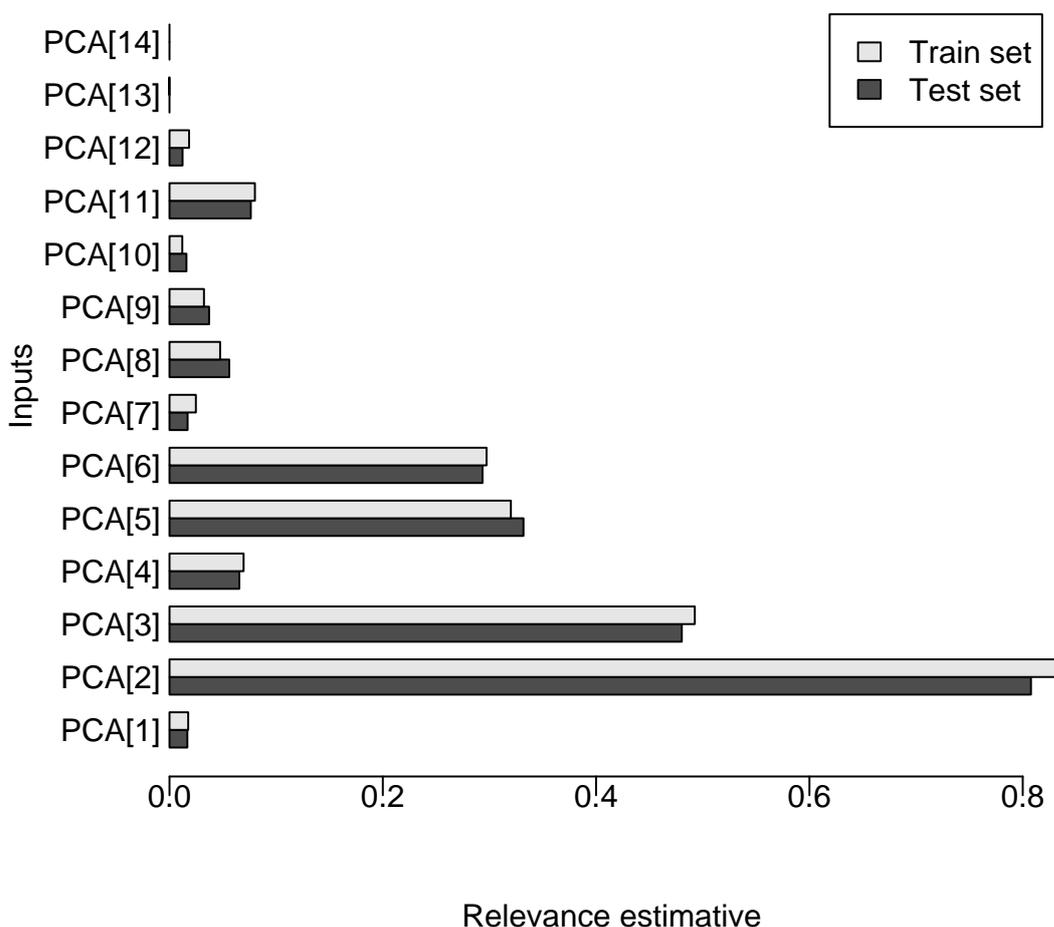


Figura 6.40: Análise de relevância de discriminação para detetor baseado em uma KLT das 14 variáveis do T2Calo.

da componente, a quantidade de acréscimo em energia decai, finalmente chegando a zero, preferencialmente, em algum ponto antes do último índice.

A análise de relevância conduzida em seções anteriores também nos oferece um método de compactação do espaço de entrada ao sistema de detecção. Se nos referirmos à análise de relevância de discriminação deste detetor, na Figura 6.40, observaremos que, embora haja uma tendência a haver uma diminuição da relevância com o aumento do índice da componente, a primeira componente, que é a mais energética, é determinada como sendo uma das menos relevantes.

Para testarmos qual dos dois métodos é o mais interessante para o problema em

análise, propõe-se uma verificação em 3 podas que resultem nas mesmas dimensões de conjuntos de dados, baseados na energia da PCA ou na análise por relevância de discriminação<sup>4</sup>:

- Podar 3 variáveis, resultando em praticamente nenhuma perda em energia segundo a PCA;
- Podar 7 variáveis, resultando em aproximadamente 10% de perda em energia segundo a PCA;
- Podar 9 variáveis, resultando em aproximadamente 27% de perda em energia segundo a PCA. Esta margem de poda também foi testada para as 14 variáveis do T2Calo na seção anterior e será possível uma comparação direta dos dois sistemas.

Treina-se 5 redes neurais (com inicializações diferentes) para cada uma das 6 possibilidades apresentadas acima, otimizando-se o número de neurônios escondidos para cada uma das possibilidades. A Tabela 6.4 resume os resultados da otimização do número de neurônios escondidos para cada uma das 6 possibilidades. A coluna marcada “Sistema” indica o tipo de poda executada e no número de variáveis que sobreviverão ao corte. A coluna “N” indica o número de neurônios escondidos que maximiza o poder de discriminação do conjunto de entrada, baseado em 5 testes com inicializações diferentes. Os valores entre parênteses indicam números de neurônios escondidos utilizados na varredura que culminou nos valores otimizados desta tabela. Como é possível notar, as duas análises apresentam tendências, no que tange à eficiência final de detecção, bastante próximas, exceto para a última (e mais dura) poda. Neste caso, detetores baseados na poda por relevância de discriminação apresentam resultados nitidamente melhores. Os detetores baseados na poda por relevância também treinam mais rapidamente, em geral. Esta tendência fica evidenciada com o aumento do número de variáveis podadas. A Figura 6.41 traz uma comparação das diversas ROC’s dos seis sistemas podados e do sistema de classificação baseado nas 14 componentes.

---

<sup>4</sup>A relevância clássica demonstra as mesmas tendências da relevância de discriminação neste caso, inclusive na ordem das variáveis.

### R.O.C Analysis

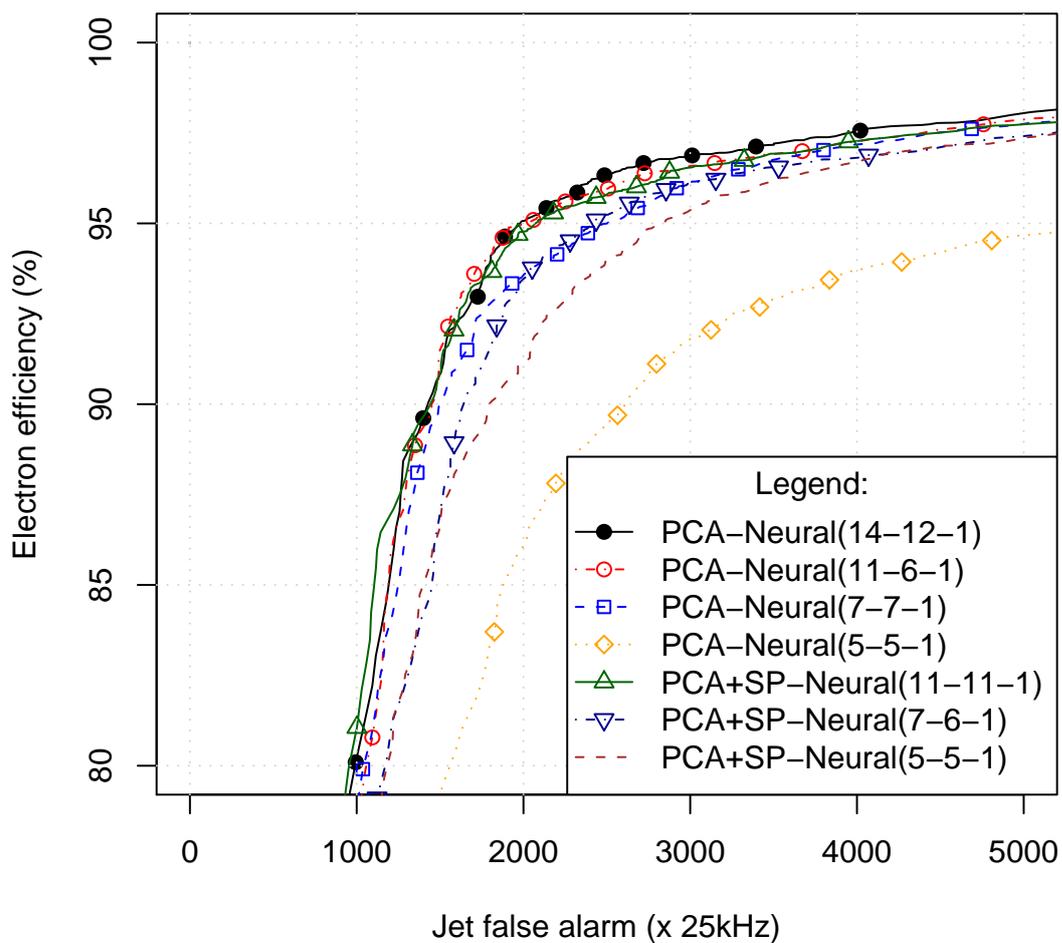


Figura 6.41: Comparação das ROC's de diversos detetores baseados da poda (segundo diferentes critérios) de um sistema com 14 variáveis na entrada, resultantes da aplicação da KLT nas 14 variáveis disponíveis na saída do T2Calo.

Tabela 6.4: Resultados da otimização do número de neurônios escondidos para um discriminador neural elétron/jato baseado na aplicação da KLT (PCA) nas 14 variáveis originais do T2Calo.

Sistema	N	Passos	EMQ/teste	SP/teste
11 var/PCA	6 (2-11)	$425 \pm 136$	$0,220 \pm 0,005$	$1,619 \pm 0,014$
11 var/Rel.SP	11 (2-11)	$356 \pm 204$	$0,220 \pm 0,010$	$1,618 \pm 0,022$
7 var/PCA	7 (2-7)	$511 \pm 220$	$0,233 \pm 0,008$	$1,588 \pm 0,008$
7 var/Rel.SP	6 (2-7)	$258 \pm 125$	$0,230 \pm 0,001$	$1,595 \pm 0,004$
5 var/PCA	5 (2-5)	$181 \pm 67$	$0,347 \pm 0,011$	$1,443 \pm 0,016$
5 var/Rel.SP	5 (2-5)	$346 \pm 169$	$0,249 \pm 0,004$	$1,552 \pm 0,007$

### 6.3 Mapeamento topológico

Enquanto a utilização das 14 variáveis do T2Calo aplicadas a um processador neural parece elevar o poder de discriminação do LVL2, indicando que existem informações perdidas no processo de detecção T2Calo+EGammaHypo, pergunta-se quanta informação é removida durante a compactação promovida pelo algoritmo. Por outro lado, tendo-se estabelecido que sistemas de discriminação com mais variáveis são melhores classificadores para o problema abordado, seria possível explorar o universo de entrada original ao T2Calo, ou seja, as células que compõem uma RoI, e se tentar recuperar mais informações sobre a natureza da partícula.

A idéia desenvolvida nesta seção segue a seguinte linha: expande-se o número de componentes na entrada do detetor neural, por um método que será descrito a seguir, diminuindo-se o grau de compactação dos dados disponíveis. Observaremos se a capacidade de detecção aumentará nestas condições. Caso positivo, utilizaremos um estudo baseado na relevância de discriminação para compactar a informação de entrada de tal sistema de detecção, a fim de obter um discriminador estruturalmente menor, ainda guardando melhor capacidade discriminatória.

O algoritmo T2Calo, durante a análise conduzida no LVL2, compacta a informação da RoI representando um candidato a elétron em quatro quantidades altamente discriminantes. Embora as técnicas de detecção linear ou neurais propostas sejam

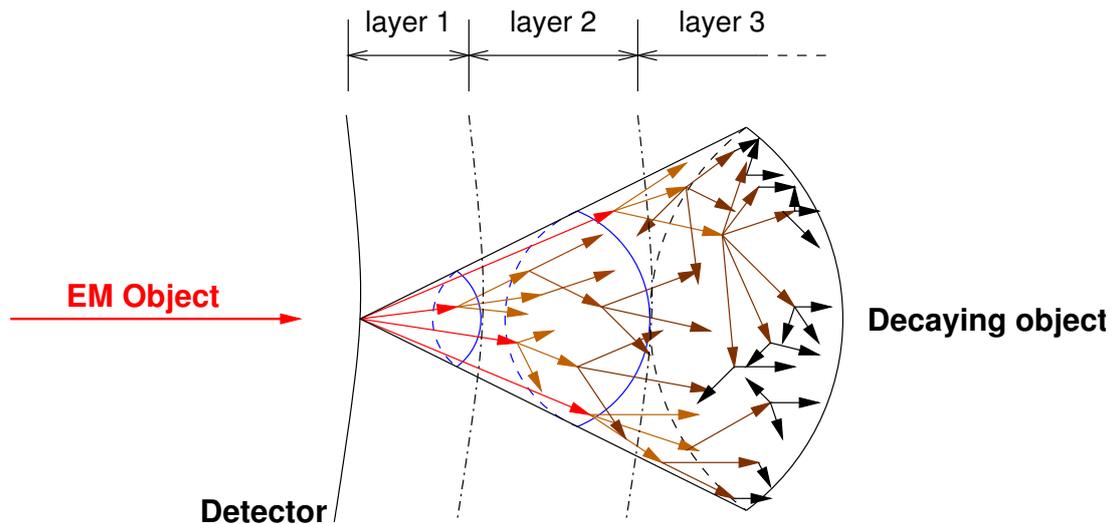


Figura 6.42: Modelo de um objeto e.m. interagindo com um calorímetro.

mais robustas quanto à operação e à manutenção, a qualidade física da separação é menos afetada do que desejaríamos. Isto se deve, principalmente, à agressiva compressão nos dados proporcionada pelo algoritmo de extração de características. Ainda que não seja viável a detecção de objetos usando toda a informação da RoI (cerca de 1300 células), é possível encontrar um método que se situe a meio caminho entre estes dois extremos.

Nesta seção, propõe-se um método de compactação da informação da RoI que utiliza a topologia do evento para gerar um conjunto de características que pode ser usado para atingir maiores eficiências de detecção de elétrons e menores valores de falso-alarme.

### 6.3.1 Anelamento dos dados de uma RoI

Observando-se a interação de elétrons e jatos (enquanto falsos-elétrons) com os calorímetros, é possível assumir que estes objetos interagirão de forma aproximadamente isotrópica em relação ao eixo  $\eta$  do detetor. Ou seja, dado um ponto de impacto inicial, a partícula tende a decair em objetos menos energéticos ao redor do eixo de penetração. Este fenômeno ocorrerá de forma expansiva, formando finalmente um cone de deposição energética, como o exemplificado na Figura 6.42.

A Figura 6.43 mostra um gráfico com a deposição energética na segunda camada e.m. de uma RoI proveniente de um elétron simulado. As partes acinzentadas



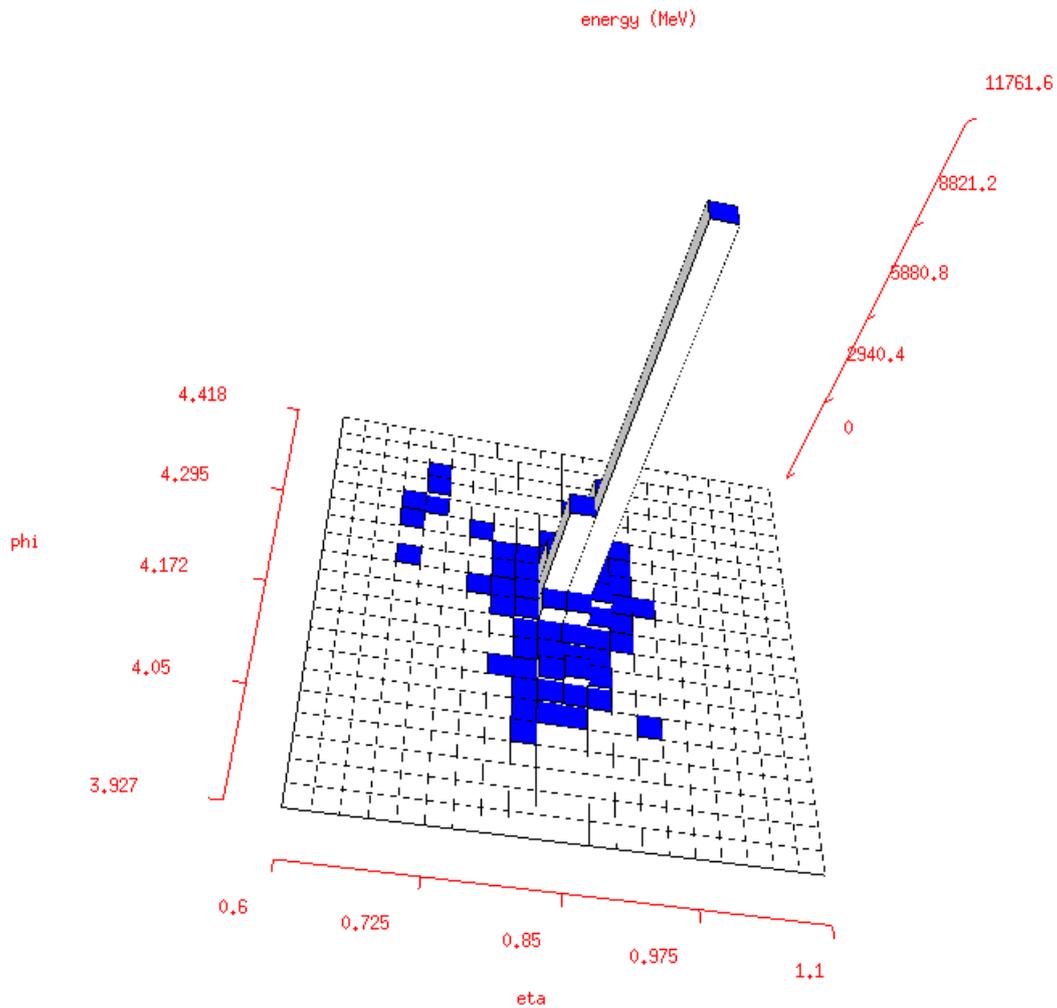


Figura 6.43: Imagem tridimensional mostrando a deposição energética de um elétron simulado na segunda camada e.m. do calorímetro do ATLAS.

indicam deposição energética, enquanto que as partes em branco indicam que houve pouca ou nenhuma deposição energética na região.

Como colocado na Seção 5.3, a topologia da interação do objeto em análise com os calorímetros define, com boa margem de sucesso, o tipo do mesmo. Elétrons tendem a um menor espalhamento que jatos, percorrendo menores distâncias ao longo de  $\eta$ . Por outro lado, jatos tendem a radializar a deposição energética, formando cones mais largos e possivelmente mais profundos.

Levando-se em consideração a formatação dos calorímetros do ATLAS e sua granularidade, sugere-se o seguinte algoritmo de compactação:

- ❶ Define-se, para cada camada, o centro de posição energética. Isto é feito de forma

simplificada, procurando para todas as células dentro de uma mesma camada, aquela célula com maior deposição energética;

- ② Uma vez que o centro de interação esteja definido, formar um conjunto de anéis com formato retangular que circundem este pico. A largura destes anéis é ajustada de forma que a largura do anel de energia seja a mesma que de uma célula dentro da camada estudada;
- ③ Para cada anel, somar os valores energéticos das células que recaem sobre o seu interior.

Por estar adaptado ao formato da cascata energética, este método de compressão preserva a informação da estrutura de um evento, melhor que a compactação provida pelo T2Calo. A estrutura segmentada dos calorímetros é preservada, de forma que tanto os valores de energia parciais quanto totais poderão ser recuperados no final do processamento. Ainda sim, obtém-se um nível de compressão de aproximadamente 10 vezes, sem demandar altos tempos de processamento, como será visto no Capítulo 7.

A Figura 6.44 esquematiza o algoritmo. Nesta figura, encontram-se exemplos de anéis de energia formados em 4 diferentes camadas dos calorímetros do detetor ATLAS. O processo de anelamento para as outras camadas é feito de forma idêntica. Próximo ao centro da RoI, nota-se uma célula que destaca, hipoteticamente, o centro de deposição energética conforme a etapa ① do algoritmo viria a encontrar. Em seguida, as células de cada segmento são somadas para a obtenção das “características” (somas em anel) do objeto a ser discriminado.

A granularidade dos anéis respeita a granularidade *típica* das células na camada. Sabe-se, no entanto, que a granularidade das células varia com  $\eta$  no detetor. Nos casos onde a granularidade é diferente da granularidade padrão na camada, atribuir-se-á a energia da célula ao anel onde recai o centro da mesma. Este algoritmo também é resiliente a dados faltantes<sup>5</sup>, uma vez que células que não puderem ser lidas apenas não serão contabilizadas. Isto evita procedimentos de verificação especiais que poderiam comprometer o desempenho do algoritmo de compactação proposto.

---

<sup>5</sup>Ocorrendo, potencialmente, com a falha de um canal de leitura ou do sistema de aquisição de dados.

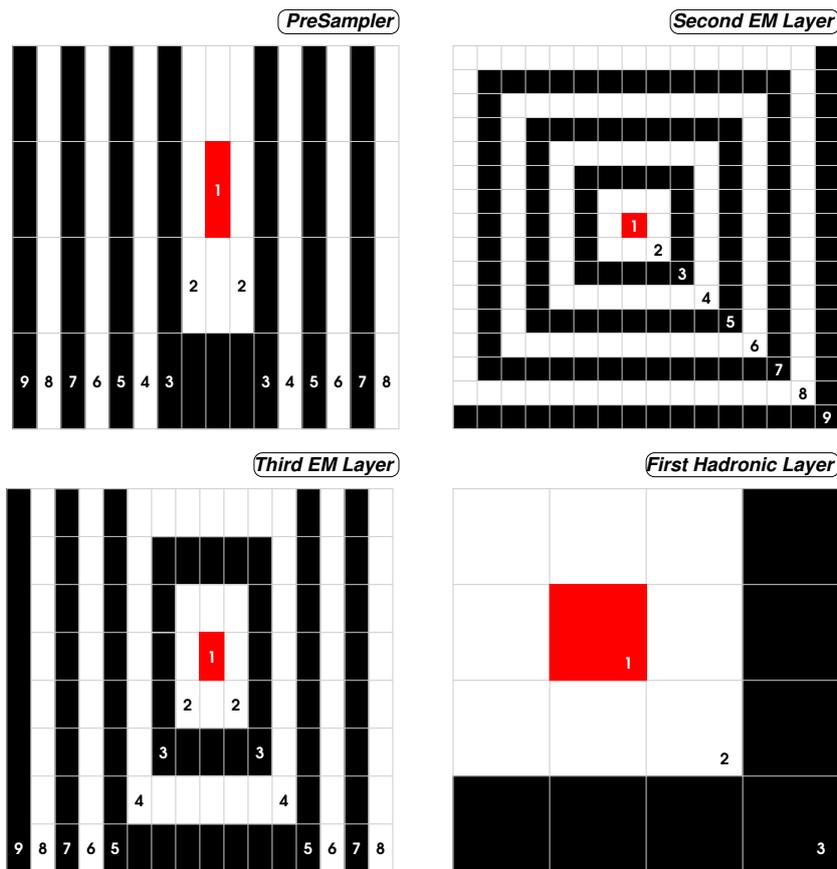


Figura 6.44: Esquema do processo de anelamento para as diferentes camadas dos calorímetros do ATLAS.

Através desta figura, observa-se igualmente que algumas camadas, devido ao formato não simétrico das células (maiores numa direção que na outra), terão anéis abertos ou incompletos. Por isso, marcaram-se os segmentos que pertencem ao mesmo anel com um número indicando a que anel determinado segmento pertence.

Um problema recorrente nos algoritmos de filtragem do ATLAS é a região de *wrap-around* da variável  $\phi$ , já que o detetor tem formato cilíndrico. Um cuidado especial deve ser tomado na confecção dos anéis, levando-se em consideração que, por exemplo, a região com  $\phi = 6,18$  radianos está bastante próxima a região com  $\phi = 0,1$  radianos, apesar da diferença numérica. Outro problema encontrado é que diferentes partes do sistema de localização dos dados dentro do Athena identificam a região do detetor onde  $\phi > \pi$  de maneiras distintas. Para parte do código (relativa à seção e.m.), o detetor se encontra na área entre  $[-\pi, \pi]$  enquanto para a outra (parte hadrônica), o detetor se encontra na área entre  $[0, 2\pi]$ . Para resolver este problema, implementou-se um algoritmo que normaliza as diferenças em ambos os casos.

Ademais, por se tratar de elétrons ou jatos na qualidade de falsos-elétrons, é bastante provável que a razão sinal-ruído seja pequena nas camadas traseiras do detetor, principalmente na região hadrônica. A Figura 6.45 mostra a deposição energética de um elétron na primeira camada do detetor hadrônico. Nesta figura este fenômeno pode ser bem observado.

Para evitar que o ruído inerente à eletrônica do detetor possa interferir com a determinação do centro de deposição energética, a procura do pico em uma camada está restrita a uma região de  $\eta = 0,1 \times \phi = 0,1$  ao redor do pico de deposição na segunda camada e.m.. Como já colocado em um capítulo anterior deste trabalho, a segunda camada e.m. é a mais profunda do sistema de calorimetria, concentrando a maior parte da energia dos elétrons que interagem com este detetor. Desta forma, se há um pico de deposição energética, este é considerado para a formação dos anéis. De outra maneira, um ponto arbitrário (aquele com maior energia) dentro da sub-janela definida pelo pico de deposição energética na segunda camada e.m. é escolhido.

Tomando-se em consideração a granularidade padrão de todas as camadas do calorímetro (7 no total, incluindo o pré-irradiador), a configuração sugerida para o número de anéis pode ser encontrada na Tabela 6.5. O número de anéis em cada

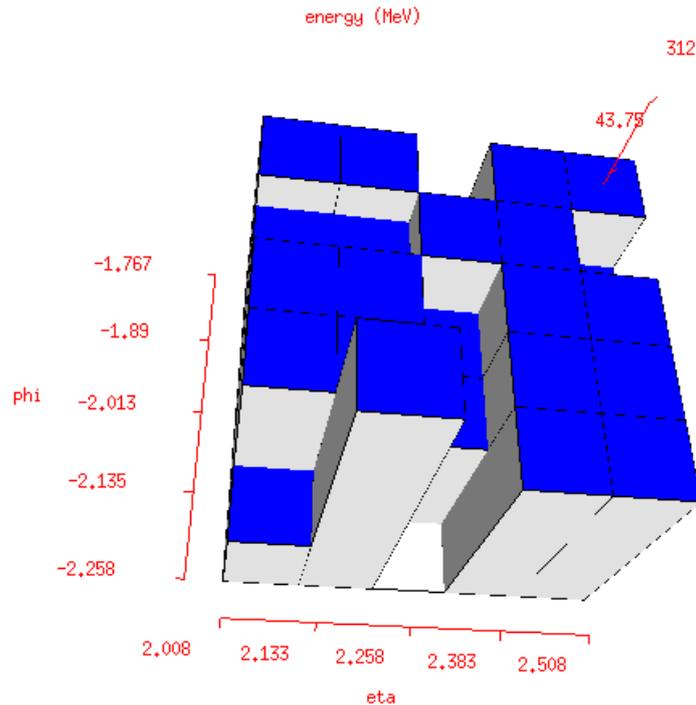


Figura 6.45: Imagem tri-dimensional mostrando a deposição energética de um elétron simulado na primeira camada hadrônica do calorímetro do ATLAS.

camada leva em consideração o número máximo de anéis que podem ser formados, considerando-se uma sub-região de  $0,4 \times 0,4$  no plano  $\eta \times \phi$ , uma vez que este é o tamanho padrão da região de interesse utilizado pelo T2Calo para a detecção de elétrons. É interessante notar que nem sempre a região que estaria demarcada por anéis concêntricos de uma largura determinada, excederá uma região com tamanho  $0,4 \times 0,4$ , já que o posicionamento dos anéis dentro da RoI dependerá do centro de deposição energética em cada camada. No entanto, uma região de  $0,3 \times 0,3$  está garantida ao redor do pico de energia, o que é suficiente para a análise de elétrons [48].

Um total de 100 anéis forma o conjunto de entrada para o novo sistema de discriminação elétron/jato. Nota-se que o primeiro anel de cada camada será considerado o pico de deposição energética na camada. Para classificar os padrões neste espaço de alta dimensionalidade, utilizaremos detetores baseados em redes neurais tipo MLP como descrito nas seções anteriores.

Tabela 6.5: Configuração para os anéis formados a partir de uma RoI de tamanho  $\eta = 0,4 \times \phi = 0,4$ .

Camada	Número de Anéis	Larg. do anel ( $\eta$ )	Larg. do anel ( $\phi$ , rad)
<b>Pré-irradiador</b>	8	0,025	$\frac{\pi}{32} \approx 0,098174$
<b>1<sup>a</sup> e.m.</b>	64	0,0031245	$\frac{\pi}{32} \approx 0,098174$
<b>2<sup>a</sup> e.m.</b>	8	0,025	$\frac{\pi}{128} \approx 0,024544$
<b>3<sup>a</sup> e.m.</b>	8	0,05	$\frac{\pi}{128} \approx 0,024544$
<b>1<sup>a</sup> hadrônica</b>	4	0,1	$\frac{\pi}{32} \approx 0,098174$
<b>2<sup>a</sup> hadrônica</b>	4	0,1	$\frac{\pi}{32} \approx 0,098174$
<b>3<sup>a</sup> hadrônica</b>	4	0,1	$\frac{\pi}{32} \approx 0,098174$
<b>Total</b>	100		

## Normalização

Uma vez que estaremos lidando com os diferenciais energéticos dentro da mesma camada, por muitas vezes sutis, faz-se necessário que a normalização aplicada aos conjuntos de anéis seja compatível com a detecção destas diferenças. Ademais, outros estudos [70] já demonstraram que a informação periférica dentro de cada camada de uma RoI é tão ou mais importante para a discriminação que a informação de energia no centro da RoI. Por estas razões se propõe um sistema de normalização dos dados da RoI baseado em um conjunto de razões cujo denominador diminui, de forma relacionada com a energia na camada, do centro para as bordas da região de interesse. Desta forma, o pico de deposição energética na camada (ou “primeiro anel”) teria um fator de normalização maior que o segundo anel, que seria por sua vez maior que o fator de normalização aplicado ao terceiro anel, e assim sucessivamente. O algoritmo está anotado na Tabela 6.6.

Neste algoritmo, o primeiro anel (pico de deposição energética) será normalizado pelo valor da energia total na camada. O segundo anel pelo valor da camada subtraindo-se o valor de deposição energética no primeiro anel. O terceiro anel, pelo valor da energia na camada menos o valor do primeiro e segundo anéis, e desta forma recursivamente, até que todos os anéis estejam normalizados. Referir-nos-emos à este algoritmo a partir de agora como algoritmo de “Normalização Sequencial”.

Tabela 6.6: Algoritmo para a “Normalização Seqüencial”.

<b>Anel</b>	Fator de Normalização	Soma em anel
<b>1</b>	$E$ (total da camada)	$E_1$
<b>2</b>	$E - E_1$	$E_2$
<b>3</b>	$E - E_1 - E_2$	$E_3$
<b>4</b>	$E - E_1 - E_2 - E_3$	$E_4$
...	...	
<b>N-4</b>	$E - E_1 - E_2 - \dots - E_{N-5}$	$E_{N-4}$
<b>N-3</b>	$E - E_1 - E_2 - \dots - E_{N-5} - E_{N-4}$	$E_{N-3}$
<b>N-2</b>	$E - E_1 - E_2 - \dots - E_{N-4} - E_{N-3}$	$E_{N-2}$
<b>N-1</b>	$E - E_1 - E_2 - \dots - E_{N-3} - E_{N-2}$	$E_{N-1}$
<b>N</b>	$E - E_1 - E_2 - \dots - E_{N-2} - E_{N-1}$	$E_N$

Nota-se que este tipo de normalização poderá amplificar em demasiado os valores periféricos. Isto é desejável para os anéis que se sucedem ao pico de energia na camada, porém nem tanto para os anéis da periferia da RoI, mais distantes, onde a razão sinal-ruído é praticamente nula. Desta forma, propõe-se que, a partir de um determinado limiar de energia, o fator de normalização torne-se constante. Desta forma evita-se que o sinal ruidoso das bordas influencie o sistema de discriminação de maneira significativa.

O sub-pacote `rbuild` do *NeuralRinger* contém a implementação de referência para este processo de anelamento descrito (ver Apêndice C). Ademais, este pacote pode ser reconfigurado de forma a definir a largura dos anéis de cada camada ou o número máximo de anéis por camada. Assim, o sistema de anelamento torna-se independente do tamanho real da RoI e, portanto, independente de quão completos são os dados recebidos pelo processador do LVL2. Anéis sem dados recebem o valor *default*, que é igual a zero. O formato do arquivo de configuração é também XML, o que simplifica a leitura e escrita dentro do contexto do trabalho.

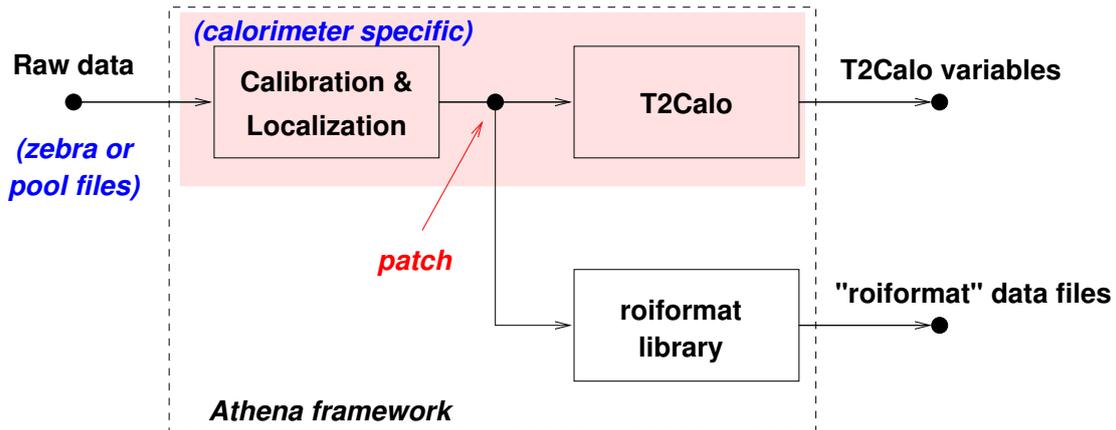


Figura 6.46: Estratégia para obter acesso aos dados tais como processados pelo algoritmo T2Calo.

## Dados

Os dados em seu formato utilizado nas seções anteriores não podem ser reaproveitados para esta porção do projeto, uma vez que necessita-se do acesso aos dados “brutos” da RoI. Em outras palavras, deseja-se acesso aos mesmos dados utilizados pelo T2Calo para a operação de extração de características que este algoritmo realiza. De posse destes dados, será possível definir a operação de compactação topológica proposta. A Figura 6.46 mostra o sistema que foi utilizado para a obtenção dos dados. Para cada RoI que é analisada pelo T2Calo, um conjunto de células é carregado dos arquivos de dados iniciais. Em seguida, os valores das células são calibrados e mapeados na geometria do detetor de forma que possam ser mais facilmente acessados pelo algoritmo de extração de características. A partir deste ponto, o T2Calo tem acesso aos dados e, é exatamente neste ponto que inserimos um *patch*, que permite que os valores das células “vistas” pelo algoritmo sejam guardados em um arquivo separado.

O formato deste arquivo é legível pelo sub-pacote `roiformat` do *NeuralRinger* e pode ser transformado em uma base de dados de anéis utilizando-se o programa `ringer`, descrito no Apêndice C. Desta forma, criou-se uma base de dados a partir do mesmo conjunto de eventos analisados na caracterização do desempenho físico do T2Calo. Esta nova base de dados tem cerca de 30000 eventos com a mesma proveniência e características dos eventos analisados até o momento.



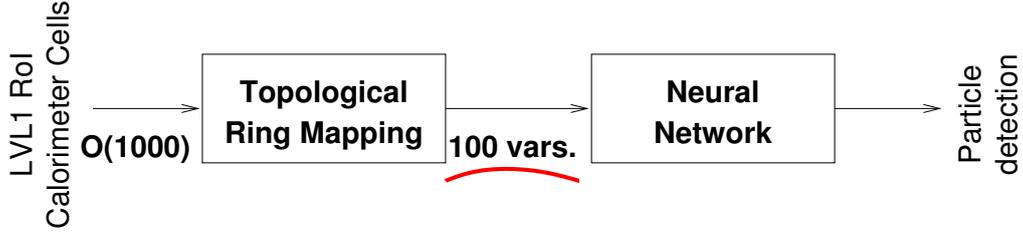


Figura 6.47: Diagrama em blocos de um sistema de detecção utilizando o anelador como extrator de características e um detetor neural.

### 6.3.2 Discriminação baseada em anelamento

A Figura 6.47 mostra o diagrama de blocos do novo sistema de detecção elétron/jato proposto. Neste caso, substitui-se o bloco de extração de características pelo extrator em anéis. O processo de anelamento reduz o espaço de dados de entrada em uma RoI (cerca de 1300 células) para 100 valores correspondentes aos anéis formados com os dados das 7 camadas dos calorímetros. Adota-se um sistema de discriminação neural, conectado à saída do sistema de anelamento, que executará o processo de detecção.

Tal qual nos sistemas anteriores, executou-se uma análise do número de neurônios ótimo para a camada escondida. Um cuidado deve ser tomado: com o aumento de neurônios no nó de entrada do discriminador neural, o número de parâmetros livres, que se deseja sintonizar, aumenta bruscamente. Finalmente, deve ser levado em consideração que cerca de 3500 jatos são usados para treinamento do sistema. Considerando-se  $N$ , o número de neurônios na camada escondida do discriminador neural, o número de parâmetros livres ( $K$ ) a serem treinados pode ser computado da seguinte forma:

$$K = (100 \times N) + N(\text{polarização}) + N(\text{con. saída}) + 1(\text{pol. saída}) \quad (6.13)$$

$$= (102 \times N) + 1 \quad (6.14)$$

Desta forma, para, por exemplo, 12 neurônios na camada escondida, o número de parâmetros livres é 1225; para 11, 1123 e para 10, 1003. Estes números estão próximos de um limite sensível para o treinamento do sistema e, desta forma, limitar-nos-emos a este patamar (12 neurônios escondidos). Portanto, varia-se o número de

Tabela 6.7: Análise dos parâmetros de treinamento para um detetor neural cuja entrada são os 100 anéis resultantes do processamento de uma RoI.

N	Passos	EMQ/teste	EMQ/treino	SP/teste	SP/treino
2	$242 \pm 59$	$0,133 \pm 0,002$	$0,098 \pm 0,002$	$1,763 \pm 0,005$	$1,821 \pm 0,008$
3	$473 \pm 128$	$0,129 \pm 0,003$	$0,089 \pm 0,004$	$1,764 \pm 0,006$	$1,840 \pm 0,011$
4	$654 \pm 283$	$0,123 \pm 0,004$	$0,079 \pm 0,002$	$1,781 \pm 0,011$	$1,861 \pm 0,007$
5	$531 \pm 124$	$0,124 \pm 0,006$	$0,077 \pm 0,005$	$1,774 \pm 0,012$	$1,864 \pm 0,011$
6	$808 \pm 220$	$0,124 \pm 0,003$	$0,070 \pm 0,004$	$1,776 \pm 0,007$	$1,877 \pm 0,012$
7	$898 \pm 217$	$0,124 \pm 0,003$	$0,065 \pm 0,004$	$1,777 \pm 0,008$	$1,890 \pm 0,006$
8	$910 \pm 340$	$0,122 \pm 0,005$	$0,063 \pm 0,004$	$1,777 \pm 0,012$	$1,894 \pm 0,008$
9	$930 \pm 207$	$0,120 \pm 0,004$	$0,060 \pm 0,003$	$1,780 \pm 0,010$	$1,900 \pm 0,006$
10	$1426 \pm 453$	$0,120 \pm 0,002$	$0,054 \pm 0,002$	$1,782 \pm 0,009$	$1,910 \pm 0,002$
11	$990 \pm 343$	$0,120 \pm 0,002$	$0,055 \pm 0,002$	$1,782 \pm 0,006$	$1,908 \pm 0,004$
12	$930 \pm 240$	$0,116 \pm 0,004$	$0,050 \pm 0,003$	$1,785 \pm 0,007$	$1,918 \pm 0,006$

neurônios escondidos entre 2 e 12. A Tabela 6.7 resume os resultados, indicando o número de neurônios escondidos utilizados e os resultados para cada agrupamento de teste.

Através desta tabela, nota-se que o número mínimo de neurônios na camada escondida deve ser maior ou igual a 4, já que a partir deste número o desempenho da rede melhora tanto com relação ao mapeamento da entrada nos alvos de treinamento (EMQ menor) quanto com relação ao poder discriminante. Quando  $N = 12$  o sistema aparenta atingir um máximo, registrando a menor média para o EMQ do conjunto de testes, com baixo desvio assim como bons patamares de detecção. O número de épocas de treinamento é maior a partir de  $N = 6$ , chegando a ser bem maior que 1000 quando  $N = 10$ . Desta forma, escolhe-se  $N = 12$ .

### Resultados para o *NeuralRinger*

Baseando-se no número de neurônios destacado anteriormente, treinaram-se dez redes que tiveram seus pesos sinápticos iniciais fixados em valores aleatórios (no in-

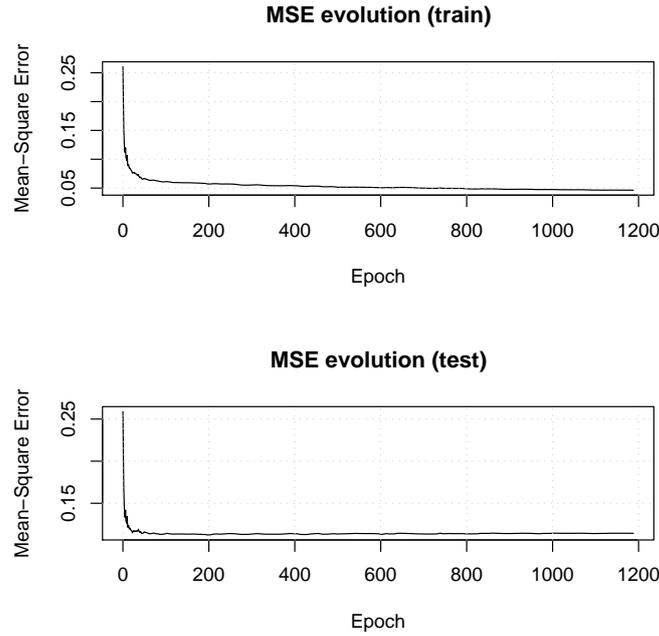


Figura 6.48: Evolução do EMQ para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do anelador.

tervalo  $[-1, +1]$ ) e diferentes para cada uma destas iterações. A Figura 6.48 mostra a evolução do EMQ da rede com melhor desempenho discriminatório. A Figura 6.49 mostra um gráfico equivalente para a evolução do produto SP ao longo do treinamento da rede. Como é possível distinguir, após cerca de 100 passos de treinamento a capacidade discriminante da rede para o conjunto de teste estagna. A capacidade de mapeamento nos alvos (medido pelo EMQ) da rede, com relação ao conjunto de treinamento, apesar da estagnação do conjunto de teste, ainda continua a melhorar após os 100 passos, chegando, no momento onde o treinamento é automaticamente interrompido, a um valor de pouco menor que 0,05, indicando um produto SP de 1,91. Ainda assim, não se observa, para o sistema em análise (também levando em consideração os outros 9 testes), o fenômeno conhecido como *overtraining*, onde a capacidade de generalização da rede é penalizada por uma excessiva especialização no conjunto de treino, normalmente fazendo com que ocorra um aumento no EMQ do discriminador para o conjunto de teste.

A Figura 6.50 contém a saída da rede após o treinamento do detetor neural. Diferente dos casos anteriores, o sistema parece identificar com bastante precisão

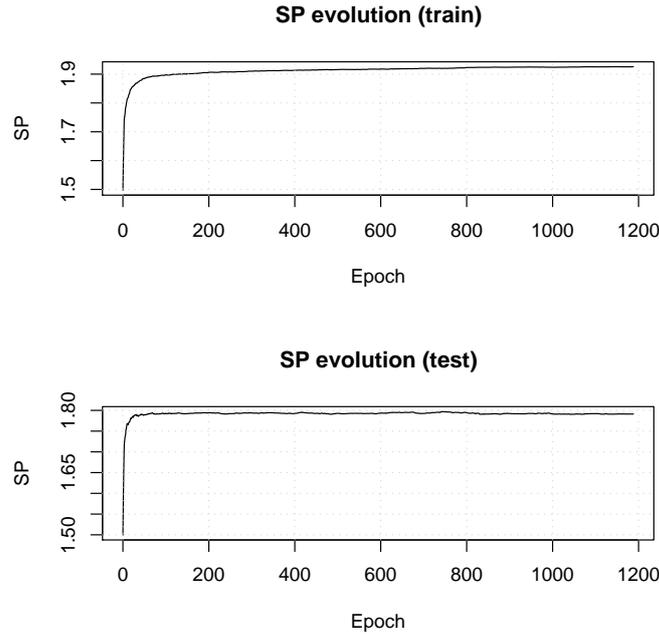


Figura 6.49: Evolução do produto SP para os conjuntos de treino e teste de um discriminador elétron/jato neural para as variáveis do anelador.

elétrons e jatos, mostrando nítidas “torres” de concentração de eventos nos pólos do gráfico. Um corte em  $-0,07$  estabelece o máximo do produto SP (conjunto de teste) para esta rede, com uma eficiência de 96,21% na detecção de elétrons contra apenas 3,44% de falso-alarme em jatos. O produto SP para estes valores é  $\approx 1,79$ , representando o melhor valor para este índice obtido até o momento. Este valor está 0,16 ponto acima do valor obtido para um detetor neural baseado em todas as características extraídas pelo T2Calo, 0,25 ponto acima do valor obtido para as 4 saídas (de discriminação) do T2Calo usando um detetor neural e 0,31 ponto (do produto SP) para um detetor baseado no algoritmo LMS! Em termos de rejeição de ruído, este novo sistema apresenta-se cerca de 6 vezes melhor que o proposto pela combinação do T2Calo aliado à discriminação utilizando o algoritmo EGammaHypo.

A Figura 6.51 contém uma comparação das R.O.C. entre todos métodos estudados até o momento. Como é possível perceber, o método proposto de anelamento, associado a um detetor neural produz um resultado significativamente melhor para a massa de dados avaliada. Para um mesmo valor de eficiência na detecção de elétrons, por exemplo 90%, o sistema baseado no anelador aprovará apenas cerca de

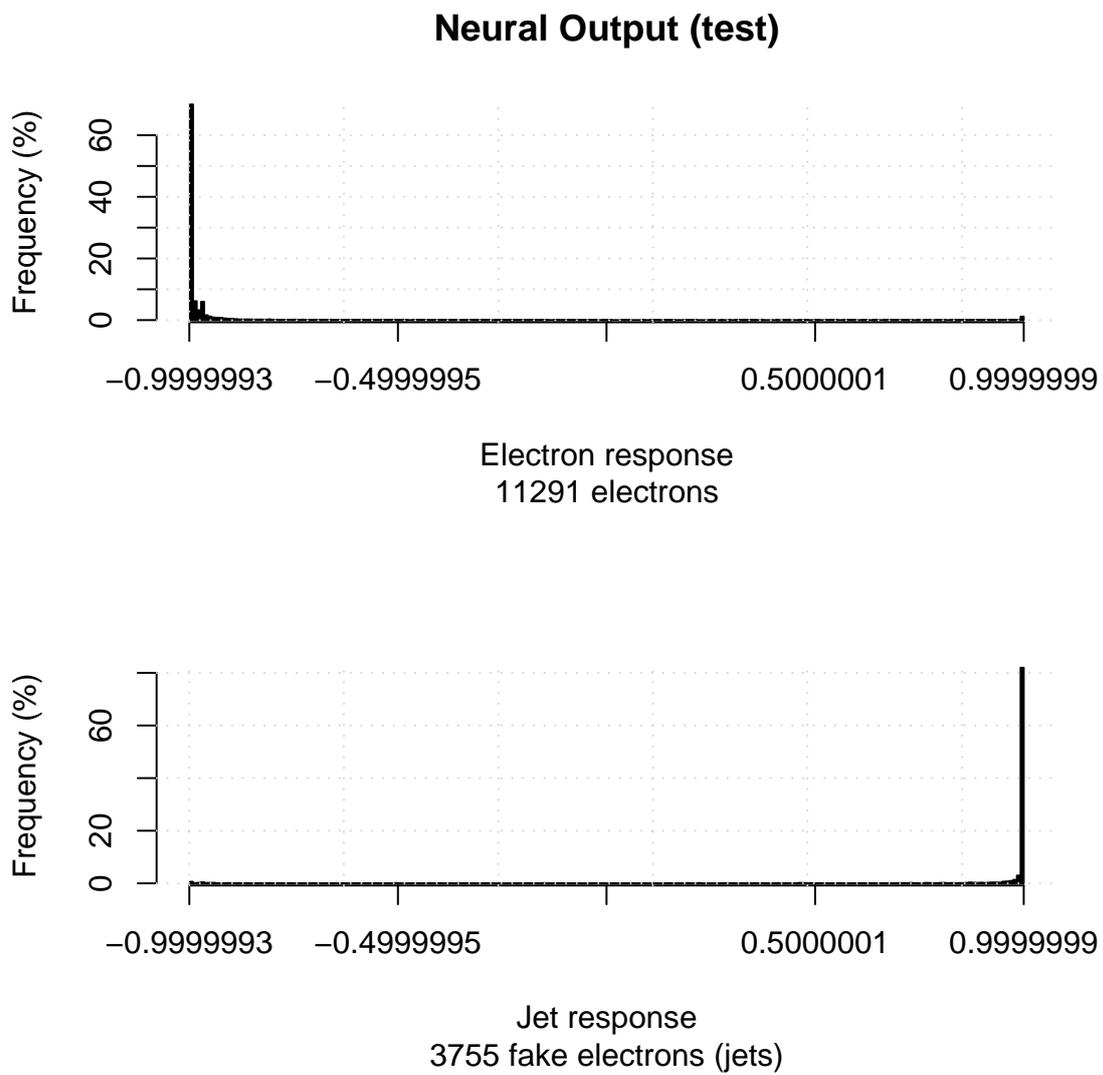


Figura 6.50: Saída para o conjunto de teste, do discriminador neural baseado nas características extraídas pelo anelador.

400 Hz em jatos ao passo que o sistema baseado na solução atual (EGammaHypo) para o LVL2, cerca de 2,4 kHz. Por outro ângulo este resultado traz que, para uma mesma capacidade de aprovar elétrons, um sistema baseado no anelamento e redes neurais poupará 6 vezes mais o terceiro nível de filtragem do experimento ATLAS. Se, por outro lado, deseja-se aumentar a eficiência na detecção de elétrons, o anelador consegue, até cerca de 96% de eficiência, manter uma taxa de falso-alarme igual ou inferior a 1 kHz. Para o sistema atual baseado no EGammaHypo, para uma eficiência na detecção de elétrons de 97%, 100% dos jatos deveriam ser aprovados (como elétrons), o que seria inviável.

A Figura 6.52 mostra os valores do produto SP por intervalo em  $\eta$  e em  $\phi$ . Observa-se neste caso que a recuperação do sinal perdido na região do *crack* é excelente, sendo difícil notar onde realmente ocorre, o que pode ser facilmente percebido nos outros métodos. Na parte central do detetor, onde a maior parte dos eventos de interesse se concentram, o valor do produto SP é bastante alto, chegando, em alguns pontos, a aproximar-se de 2,0. A eficiência de detecção para elétrons é de praticamente 100% nesta região. O gráfico equivalente em  $\phi$  mostra a homogeneidade esperada para todo o espectro.

A Figura 6.53 contém uma análise da eficiência de detecção (a) e do produto SP (b) por intervalo em  $E_{T_{3\times 7}}^{e.m.}$ . A eficiência de detecção neste caso apresenta-se aproximadamente constante, até a faixa de energia em torno de 100 GeV, caindo abruptamente em seguida. Os números no topo de cada barra indicam a quantidade de elétrons e jatos analisadas no intervalo e servem como uma medida da significância do resultado. Para valores de energia acima de  $\approx 100$  GeV, o número de eventos disponíveis para treino e análise cai significativamente, devido à natureza dos eventos estudados. Por esta razão, os resultados para altos valores energéticos devem ser considerados com esta ressalva. A Figura 6.54 traz uma comparação da capacidade discriminativa deste novo sistema, com os métodos de discriminação neurais baseados na saída do T2Calo (parcial, com 4 variáveis e completa com 14 variáveis). Este novo sistema possui melhor discriminação nas faixas de energia com mais eventos, superando em muitos pontos os outros dois detetores.

As Figuras 6.55 e 6.56 trazem as análises de relevância para o conjunto de teste, das 100 variáveis do anelador, tendo em vista a variação do EMQ e do produto

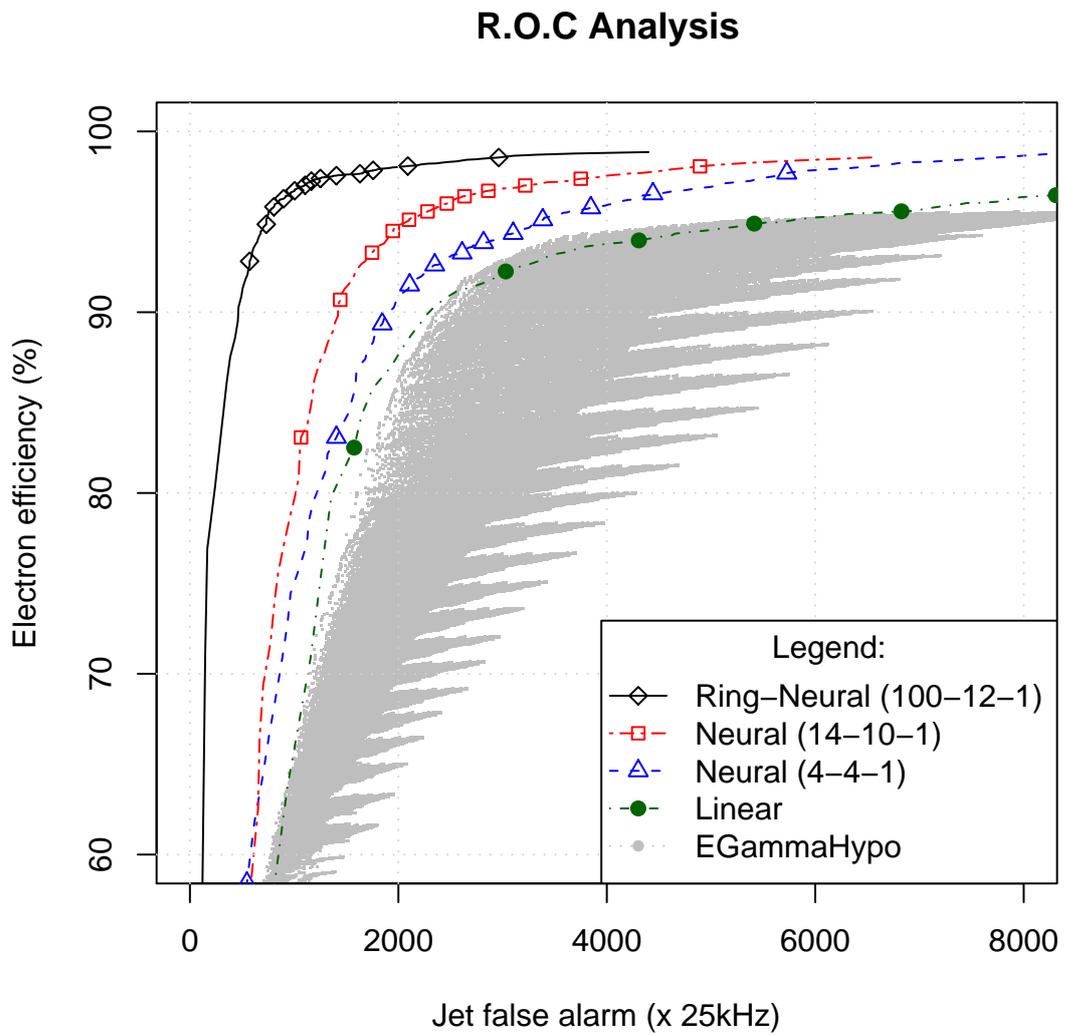
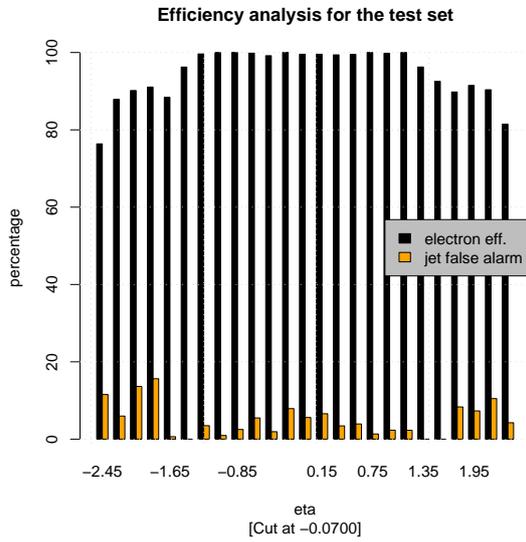
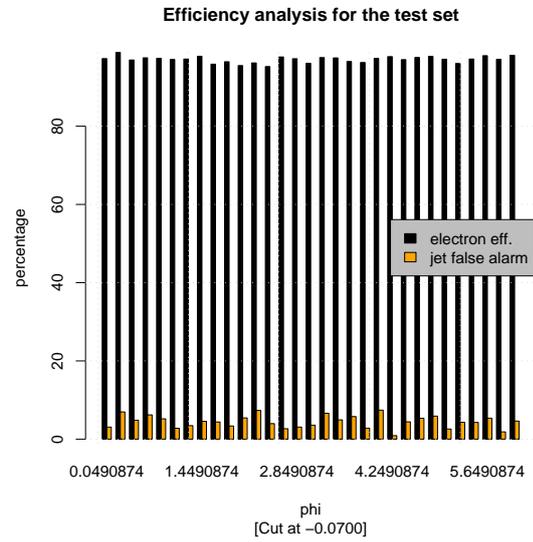


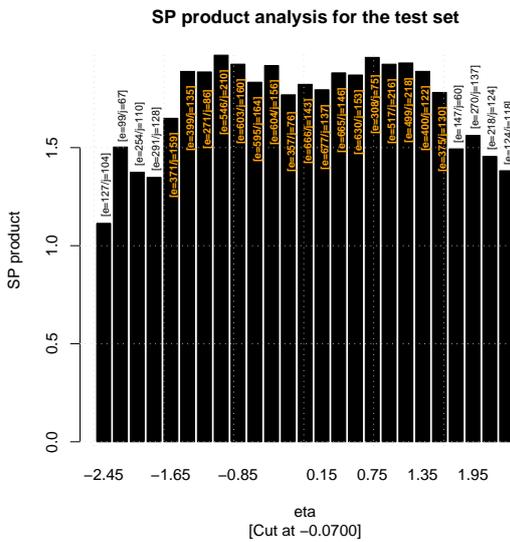
Figura 6.51: Comparação da R.O.C. dos diversos sistemas de filtragem abordados até agora com o sistema neural baseado em anéis de deposição energética.



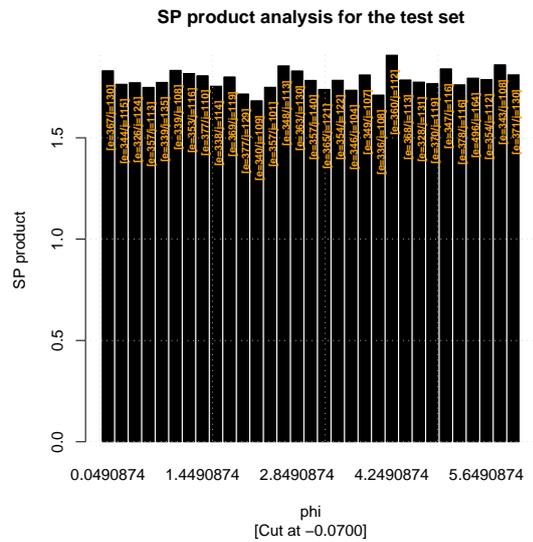
(a) Eficiência por  $\eta$



(b) Eficiência por  $\phi$



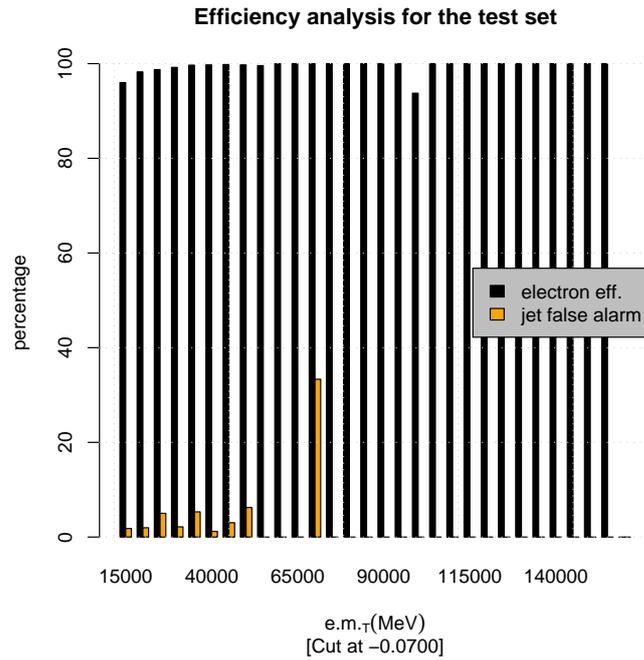
(c) Produto SP por  $\eta$



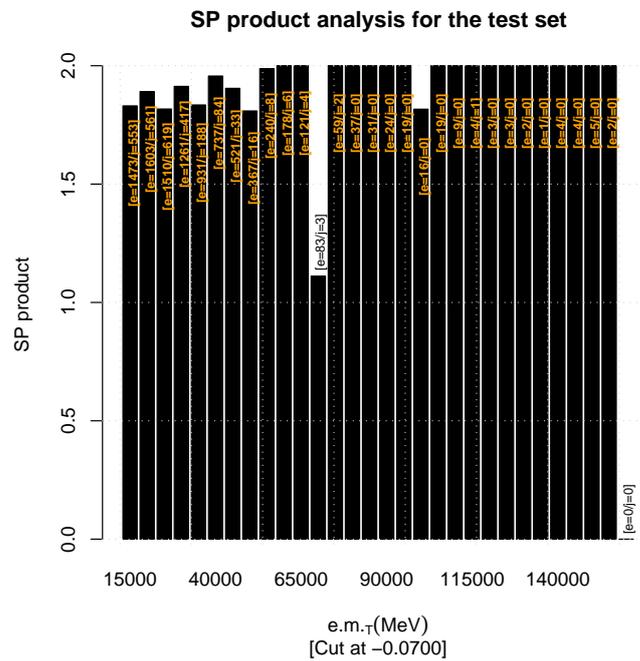
(d) Produto SP por  $\phi$

Figura 6.52: Análise da eficiência de classificação e do produto SP para o discriminador neural baseado nas saídas do anelador, para os dados do conjunto de teste ao longo de  $\eta$  em (a) e (c) e por  $\phi$ , em (b) e (d).





(a) Eficiência



(b) Produto SP

Figura 6.53: Análise da eficiência de detecção de elétrons e falso-alarme em jatos (a) e do produto SP (b) por  $E_{T_{3 \times 7}}^{e.m.}$  para o discriminador neural baseado nas saídas do anelador, tendo por base os dados do conjunto de teste.

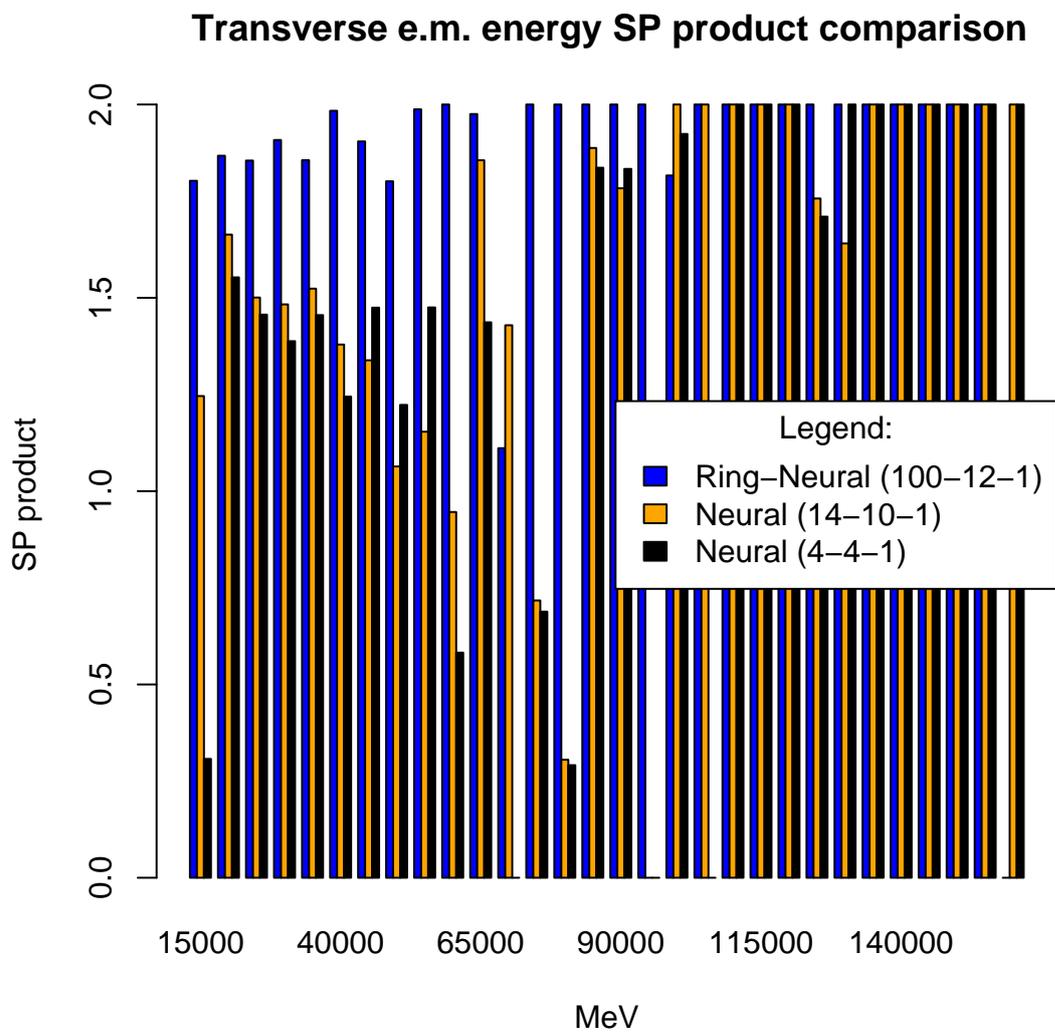


Figura 6.54: Gráfico comparativo do produto SP ao longo de  $E_T^{e.m.}$  para os discriminador neurais baseados nas saídas do anelador, parcial e completa do T2Calo.

SP respectivamente. Os dois gráficos parecem indicar, igualmente, uma grande importância nos primeiros anéis de cada camada (veja Tabela 6.8). A importância de cada anel diminui, em geral, com o afastamento do centro, onde pressupõe-se que a razão sinal ruído seja inferior. A importância relativa entre os anéis mais importantes e aqueles de menor importância parece ser menor da perspectiva da relevância baseada no produto SP que da relevância clássica, como também observado na análise das 14 variáveis produzidas pelo T2Calo. Outra informação que pode ser aferida é a importância da informação na primeira camada e.m.. Para a relevância clássica, as variáveis desta camada constituem uma informação essencial na determinação da saída, indicando ser duas vezes mais relevantes que os dados de outras camadas. Na avaliação baseada na capacidade discriminante do sistema, a parte hadrônica é a mais importante, sobretudo a primeira camada indica ser a mais relevante. É interessante notar mais uma vez que a informação de relevância está ligada diretamente ao caráter físico do evento: para a discriminação, os valores mais relevantes na detecção elétron/jato são, sem dúvida, a informação de depósito de energia ao longo das camadas e, em especial, a existência de energia na primeira camada hadrônica, indicando que o evento contenha componentes desta natureza (aumentando a probabilidade de não ser um elétron<sup>6</sup>). O detetor baseado nos anéis de deposição energética “aprende” este conceito através da análise do conjunto de dados, sem a necessidade da intervenção de um especialista.

### 6.3.3 Compactação baseada na relevância

A informação de relevância estudada até aqui pode ser aproveitada para que se aplique sobre as entradas do discriminador neural uma poda baseada na capacidade discriminante de uma variável. Esta poda pode ser feita automaticamente [79] em muitos casos, ou manualmente, se forem necessárias otimizações no processo de poda.

Um fator interessante e novo até o momento é o aparecimento de valores negativos de relevância, cuja possibilidade havia sido prevista anteriormente. Esta informação indica que existem variáveis que dificultem o treinamento da rede e, por

---

<sup>6</sup>Elétrons com energia superior a 90 GeV podem vir a interagir com a seção hadrônica dos calorímetros.

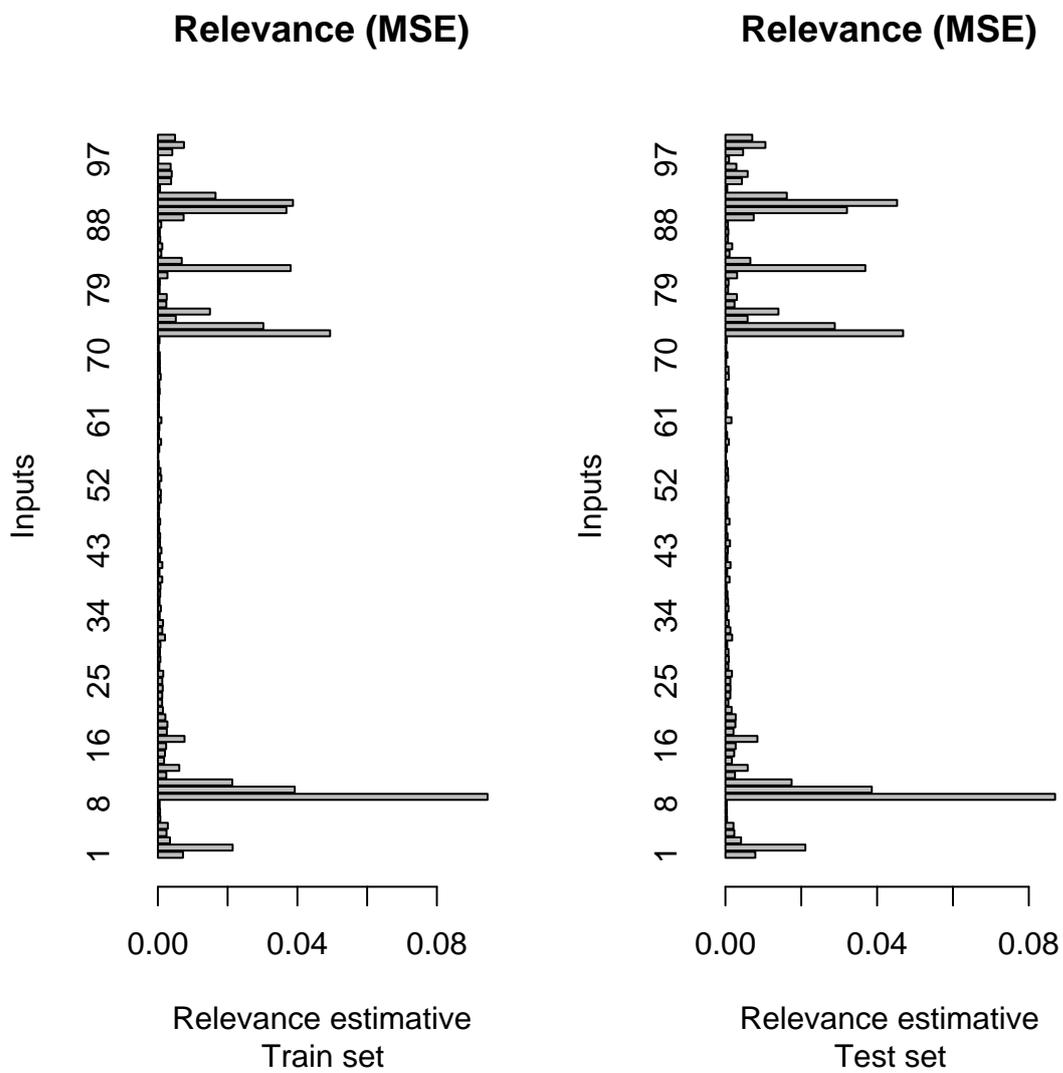


Figura 6.55: Os valores de relevância para os conjunto de teste, para as 100 variáveis do anelador e considerando-se o classificador neural em estudo, como indica a Tabela 6.8.

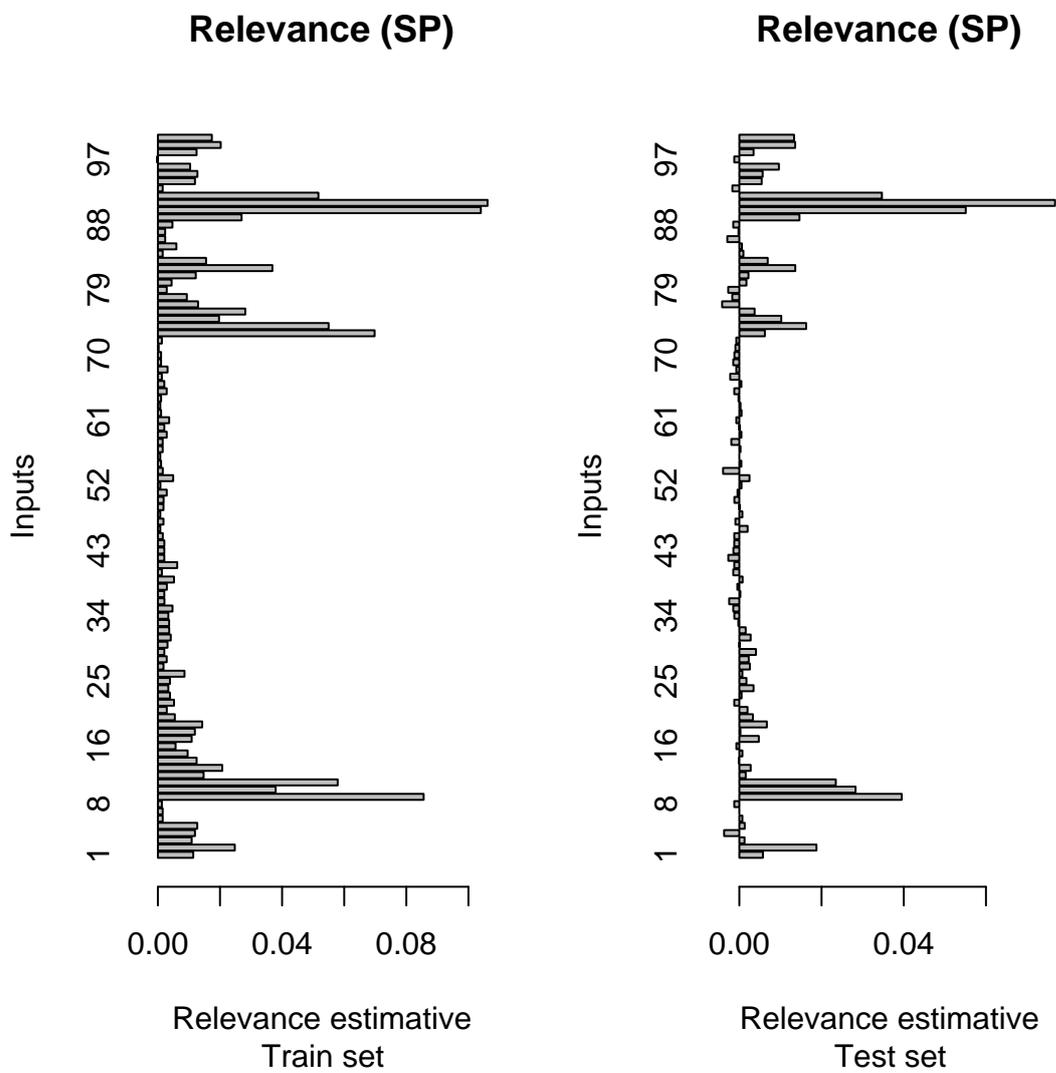


Figura 6.56: Os valores de relevância de discriminação para os conjunto de teste, para as 100 variáveis do anelador e considerando-se o classificador neural em estudo, como indica a Tabela 6.8.

Tabela 6.8: Posicionamento relativo dos anéis produzidos pelo anelador no vetor de entrada ao sistema de detecção neural.

<b>Camada</b>	<b>Número de anéis</b>	<b>Primeiro</b>	<b>Último</b>
Pré-irradiador	8	1	8
1 <sup>a</sup> e.m.	64	9	72
2 <sup>a</sup> e.m.	8	73	80
3 <sup>a</sup> e.m.	8	81	88
1 <sup>a</sup> hadrônica	4	89	92
2 <sup>a</sup> hadrônica	4	93	96
3 <sup>a</sup> hadrônica	4	97	100

consequência, o processo de discriminação proporcionado pela combinação anelador + redes neurais. Para validarmos esta tendência, eliminam-se do conjunto de treinamento e testes as variáveis com relevância de discriminação negativas, aplicando-se um patamar de poda em 0 (zero). Neste modo, 62 variáveis sobrevivem ao corte, como indicado na Tabela 6.9.

Nota-se que a camada que teve mais cortes foi a 1<sup>a</sup>camada e.m.. Esta camada tem uma geometria de células muito particular: 4 vezes mais compridas ao longo da variável  $\phi$  e 8 vezes mais finas ao longo de  $\eta$ , comparativamente à 2<sup>a</sup>camada e.m.. Ela é utilizada para detetar a multiplicidade da cascata, indicando a presença de muitas partículas ao invés de apenas uma. A informação de detecção entre elétrons e jatos encontra-se, naturalmente, próxima ao centro da RoI para esta camada, uma vez que elétrons (partículas simples) exibirão um pico de deposição energética pronunciado nos arredores do centro, ao passo que jatos, na qualidade de falsos-elétrons, vários picos situados também nesta região. Caso esta afirmação não fosse verdadeira, o LVL1 provavelmente não teria selecionado o evento, pois a RoI não passaria o teste de isolamento em energia na seção e.m.. Desta forma, como é esperado, o sistema de detecção indica que os anéis na parte exterior (em sua maioria) sejam pouco relevantes ao processo discriminatório, dificultando o processo de aprendizado do detetor neural.

No caso das outras camadas, as informações da periferia mais distante são con-

Tabela 6.9: Anéis sobreviventes de uma poda baseada na relevância de discriminação.

Camada	Número de anéis	Primeiro	Último	Anéis removidos
Pré-irradiador	6	1	6	2
1 <sup>a</sup> e.m.	36	7	42	28
2 <sup>a</sup> e.m.	5	43	47	3
3 <sup>a</sup> e.m.	5	48	52	3
1 <sup>a</sup> hadrônica	4	53	56	0
2 <sup>a</sup> hadrônica	3	57	59	1
3 <sup>a</sup> hadrônica	3	60	62	1

sideradas menos relevantes, como é de se esperar. Isto não quer dizer, em absoluto, que a informação periférica não seja importante. Quer dizer, simplesmente que, muito longe do centro, chega-se a um ponto onde não existe mais interação da partícula com o detetor, e portanto a informação contida nesta região não ajuda a discriminação.

De posse dos 62 anéis que sobreviveram ao corte de discriminação, treinam-se 10 redes neurais, ainda com 12 neurônios escondidos, partindo-se de inicializações distintas. Observa-se a variabilidade de parâmetros como o número de épocas necessárias ao treinamento (espera-se que convirja mais rápido) e os valores finais de EMQ e produto SP que se espera, sejam maiores que no caso anterior utilizando-se os 100 anéis. Em média, este novo sistema toma cerca de 552 passos de treinamento para atingir a estagnação do produto SP. O desvio-padrão para o número de passos de treinamento é de apenas 130 passos, indicando boa concordância entre os detetores destes últimos testes, no que tange este aspecto; sobretudo em comparação ao sistema anterior, onde observaram-se  $930 \pm 240$  passos de treinamento. O EMQ do conjunto de teste cai ligeiramente, na média, passando de  $0,116 \pm 0,004$  para  $0,110 \pm 0,002$ . O produto SP de teste aumenta 0,032 ponto, passando de  $1,785 \pm 0,007$  para  $1,807 \pm 0,004$ . O produto SP máximo determina uma eficiência na detecção de elétrons de 96,64% contra apenas 3,17% de falso-alarme. A Figura 6.57 mostra uma comparação das ROC's de alguns dos discriminadores analisados até o

Tabela 6.10: Anéis sobreviventes de uma segunda poda baseada na relevância de discriminação.

Camada	Número de anéis	Primeiro	Último	Anéis removidos
Pré-irradiador	6	1	6	0
1 <sup>a</sup> e.m.	25	7	31	11
2 <sup>a</sup> e.m.	3	32	34	2
3 <sup>a</sup> e.m.	4	35	38	1
1 <sup>a</sup> hadrônica	4	39	42	0
2 <sup>a</sup> hadrônica	3	43	45	0
3 <sup>a</sup> hadrônica	3	46	48	0

momento.

A Figura 6.58 traz, mais uma vez, a análise de relevância de discriminação do sistema baseado em 62 dos 100 anéis originais. Como é possível determinar, ainda existem cerca de 14 componentes que apresentam relevância negativa. Um segundo corte baseado nesta observação será realizado. A Tabela 6.10 contém a informação dos anéis sobreviventes. Após reajustar os pesos sinápticos, o discriminador neural, mais uma vez tende a remover anéis excedentes da periferia da 1<sup>a</sup>camada e.m.. O número total de anéis neste momento é menor que metade dos anéis originais.

Um novo conjunto de 10 testes é executado para se determinar se a poda é revertida em benefício do processo de treinamento. Para estes novos testes, o número médio de passos de treinamento para a rede neural cai, de 552 passos para 426, com desvio padrão aproximadamente igual (127 para estes 10 testes). A capacidade discriminante da rede para o conjunto de testes sobe, marginalmente, de  $1,807 \pm 0,004$  para  $1,817 \pm 0,003$ , indicando, mais uma vez, que a poda foi benéfica ao processo de discriminação a que se destina o classificador neural. O melhor dos detetores consegue identificar corretamente 96,89% dos elétrons contra 2,93% de falso-alarme (o produto SP é 1,82). A Figura 6.59 mostra uma comparação deste classificador com outros, também baseados em redes neurais obtidos até o momento. Como é possível determinar por inspeção desta figura, o comportamento deste detetor é quase idêntico ao detetor baseado em 62 dos 100 anéis, superando-o em alguns pontos.



## R.O.C Analysis

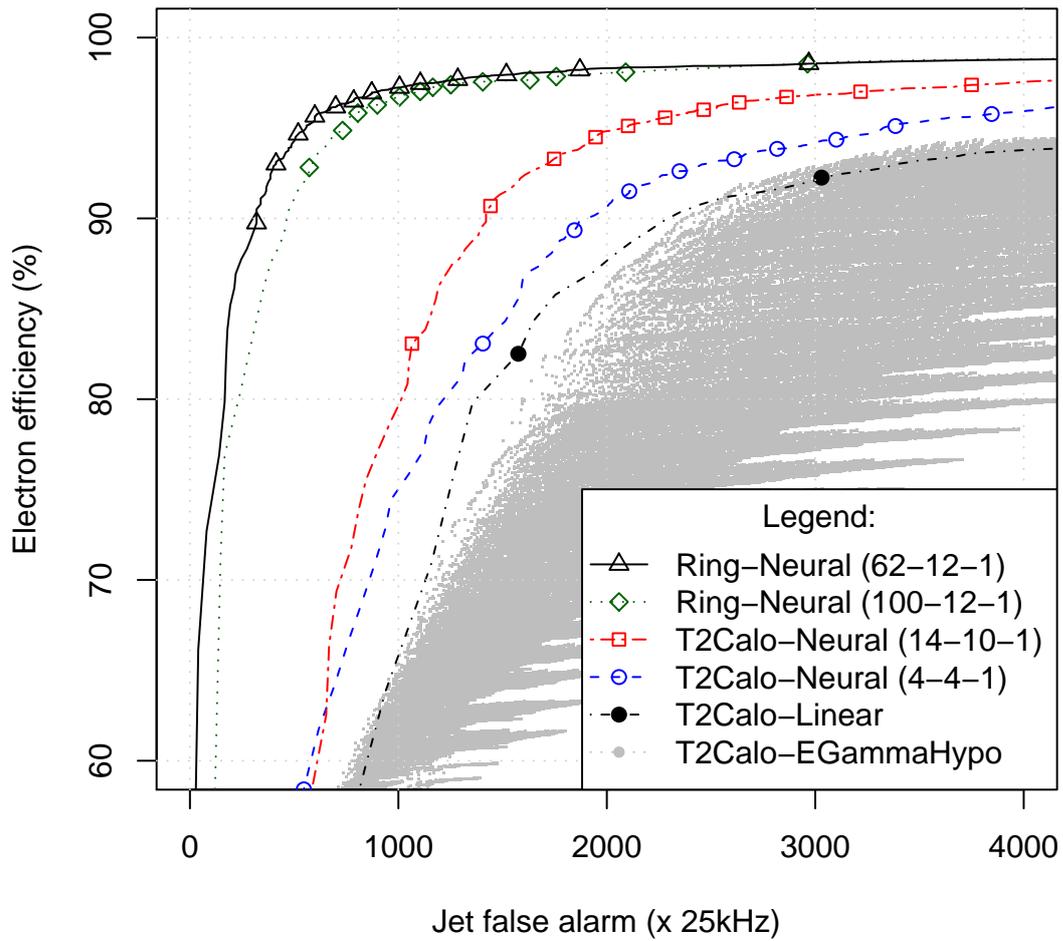


Figura 6.57: Comparação entre as ROCs de alguns dos sistemas de discriminação abordados até aqui. Em destaque, a ROC do melhor discriminador (linha cheia), determinado a partir da análise de relevância.

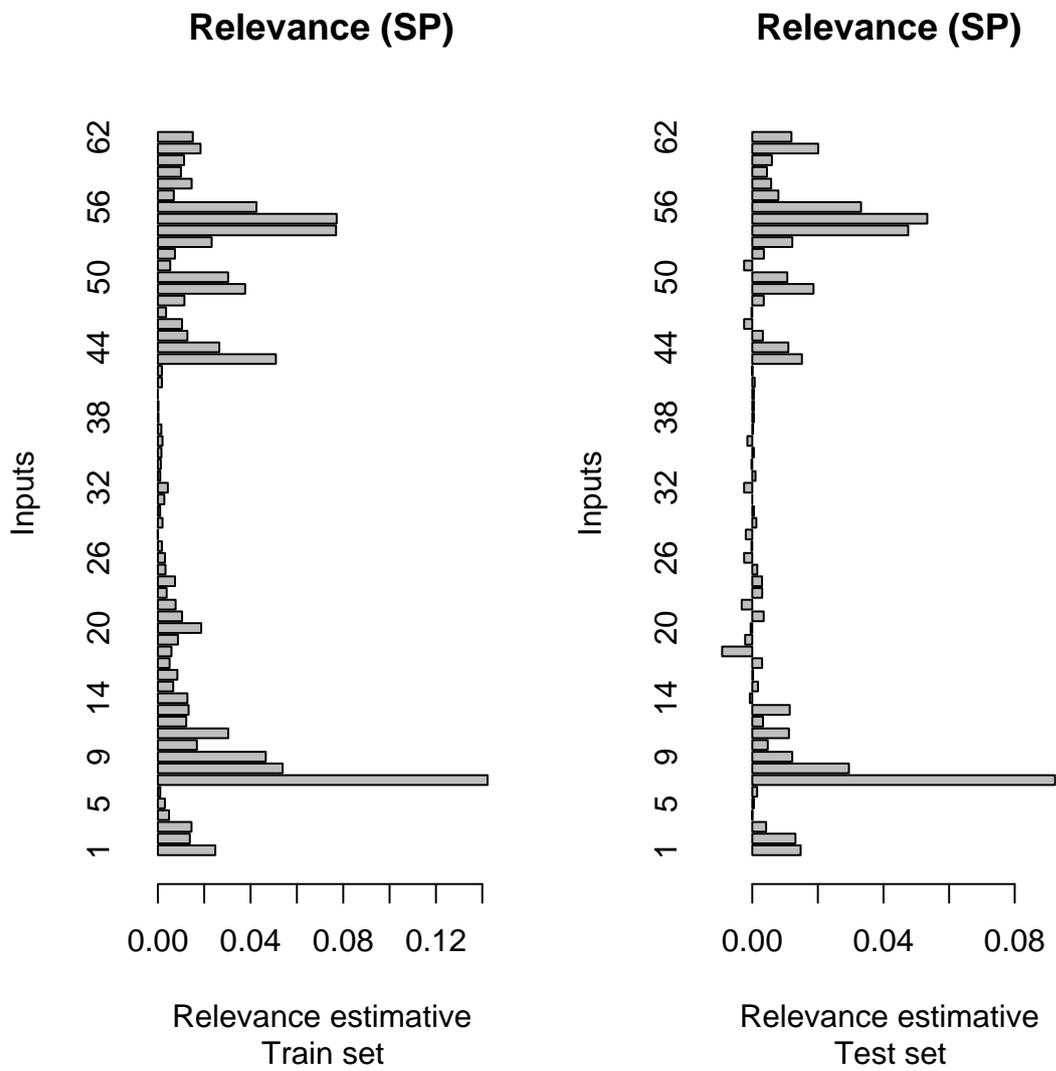


Figura 6.58: Os valores de relevância de discriminação para os conjunto de treino e teste, para as 62 variáveis que sobreviveram a uma poda baseada na relevância de discriminação.

A Figura 6.60 mostra os resultados de relevância de discriminação para este último discriminador.

### **Considerações com relação à normalização seqüencial**

Até o presente momento, a poda está sendo realizada sem nenhuma consideração ao processo de normalização, ainda que exista uma dependência. De fato, a técnica de normalização seqüencial dos anéis proposta neste trabalho assume a disponibilidade de um valor de energia total associado a cada camada para o cálculo do fator de normalização de cada anel. Ademais, para calcular o fator de normalização para o anel  $N + 1$  em uma mesma camada, necessita-se do valor de energia do anel  $N$ . Portanto, para manter o anel  $N + 1$  de uma camada, é necessário que se mantenha o anel  $N$ , ao menos para o cálculo do fator de normalização. De fato, como veremos mais a frente, esta última sugestão seria uma prática que compensaria em pouco, visto que a maior parte do tempo de processamento não está na discriminação neural, e sim na aquisição e preparação dos dados (calibração). Por outro lado, seria possível considerar que uma poda sempre seria positiva no que tange à complexidade do discriminador neural que segue o sistema de anelamento. Ao retirar uma variável da entrada, ao menos 12 conexões sinápticas do detetor final seriam removidas simultaneamente, levando-se em consideração a configuração exemplificada anteriormente.

Nesta seção proporemos uma técnica híbrida de compactação do anelador, que observa as informações de relevância levando em conta a continuidade dos anéis disponíveis para a aplicação da normalização. Espera-se que seja possível obter, simultaneamente, detetores mais compactos e velozes no que tange à aquisição e ao pré-processamento dos dados. Uma das técnicas que empregaremos neste caso será a combinação de camadas menos relevantes, reduzindo o número de entradas do processador neural, sem que se perca, necessariamente, a informação de energia presente nos anéis removidos. Uma outra forma de manter a informação de energia, mas compactando ainda mais a entrada, seria a utilização de anéis cuja largura dependesse da relevância dos dados naquela região. Este estudo pode vir a ser realizado no futuro. Tendo em vista a discussão no início desta seção, propõe-se a seguinte variante do processo de poda, baseando-se na relevância de discriminação

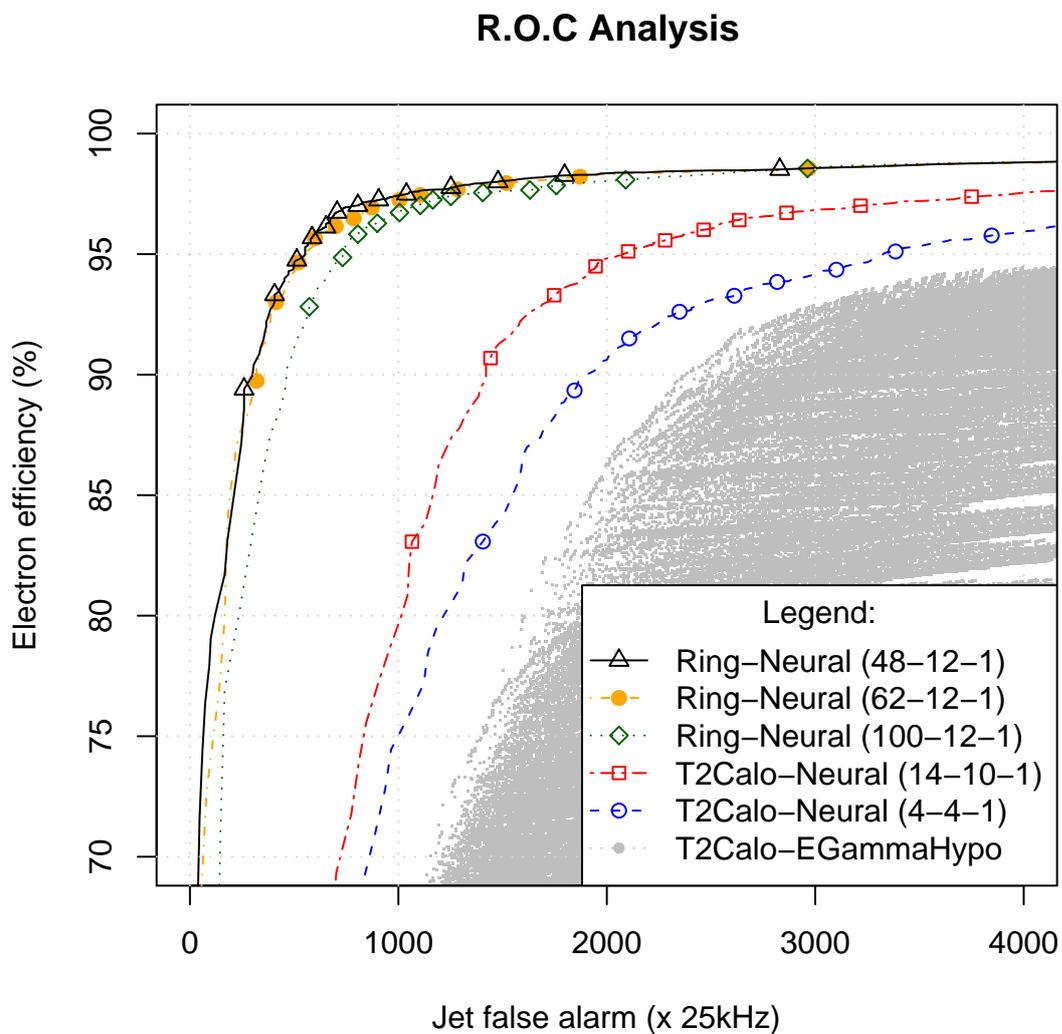


Figura 6.59: Comparação entre as ROCs de alguns dos sistemas de discriminação abordados até aqui. Em destaque, a ROC do melhor discriminador (linha cheia), determinado a partir de uma segunda análise de relevância.

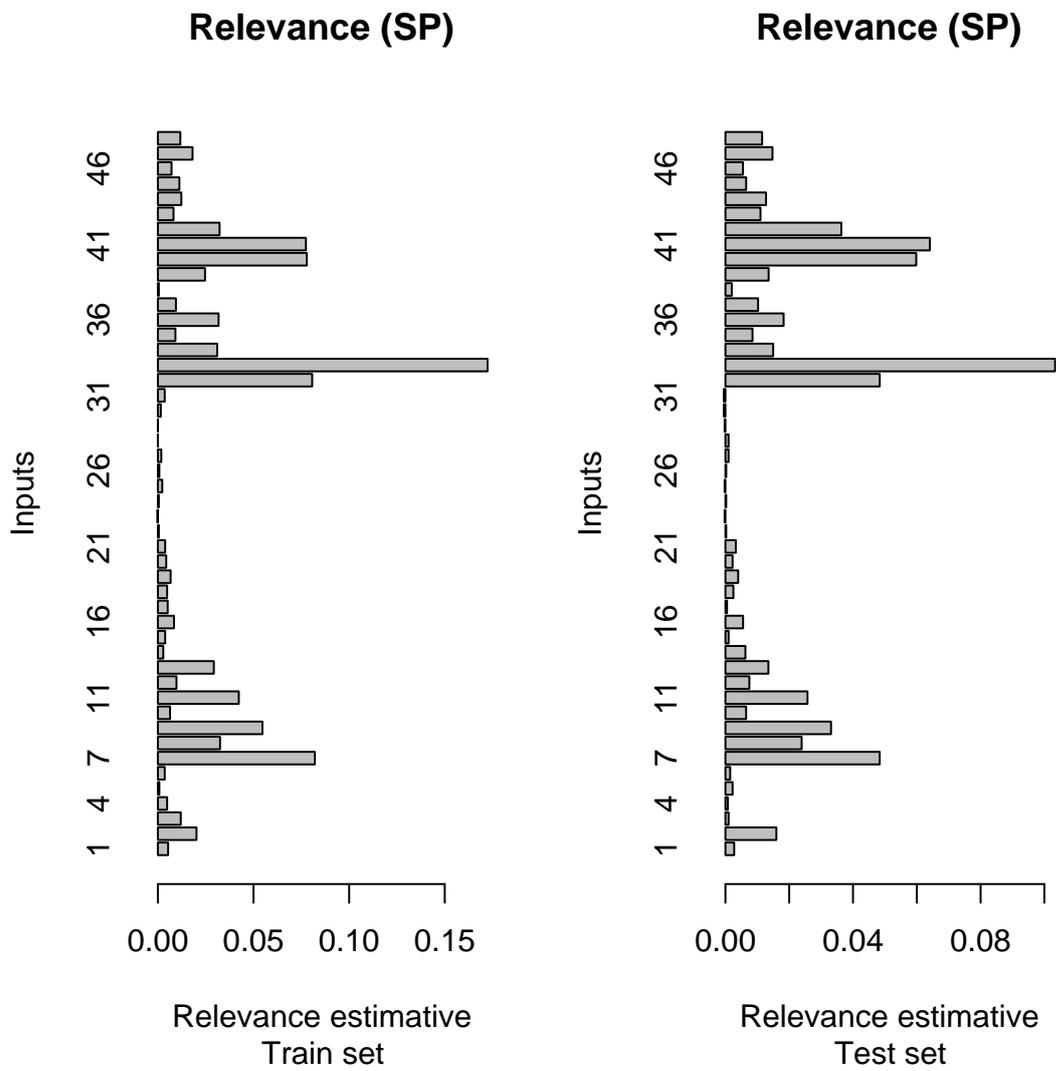


Figura 6.60: Os valores de relevância de discriminação para os conjunto de treino e teste, para as 62 variáveis que sobreviveram a uma poda baseada na relevância de discriminação.

da Figura 6.56.

1. **Pré-irradiador:** utilizar apenas 4 ao invés de 8 anéis podados, com a mesma granularidade anterior;
2. **1<sup>a</sup> camada e.m.:** utilizar apenas 12 dos 64 anéis podados, com a mesma granularidade anterior;
3. **2<sup>a</sup> camada e.m.:** utilizar apenas 6 dos 8 anéis iniciais, com a mesma granularidade anterior;
4. **3<sup>a</sup> camada e.m.:** utilizar apenas 4 dos 8 anéis iniciais, com a mesma granularidade anterior;
5. **1<sup>a</sup> camada hadrônica:** manter a configuração atual;
6. **2<sup>a</sup> e 3<sup>a</sup> camadas hadrônica:** aglutinar em uma mesma “camada” virtual, mantendo a granularidade anterior.

Um novo total de 34 anéis são extraídos da massa de dados originais, com a preocupação da eliminação das variáveis menos relevantes (ou com relevância negativa), sem que se criem “buracos” entre os anéis. Em seguida, dez testes, mantendo a mesma parametrização do sistema anterior, foram conduzidos. A R.O.C. do melhor destes testes, junto àquelas dos testes anteriores, pode ser vista na Figura 6.61. A figura mostra que o sistema podado se comporta melhor que o sistema original com 100 anéis ou tal como um dos sistemas otimizados com 62 ou 48 anéis. Este novo sistema possui apenas a terça parte do número de anéis do sistema original, mas praticamente mantém o desempenho do sistema discriminador com 100 anéis, seguindo a linha dos sistemas esculpidos pela relevância de discriminação. O discriminador neural, neste caso, tem  $66 \times 12 = 792$  menos conexões sinápticas, somando um total de 433 sinapses. O máximo do produto SP (1, 81) é atingido para uma eficiência na detecção de elétrons de 96,64% contra um falso-alarme em jatos de apenas 3,09% (aproximadamente 770 Hz de ruído de fundo).

A primeira camada e.m., como mostra a Figura 3.7 no Capítulo 3, é uma camada mais fina, originalmente concebida para a detecção de jatos de partículas [25]. Estes objetos podem ser também caracterizados analisando-se os padrões de deposição

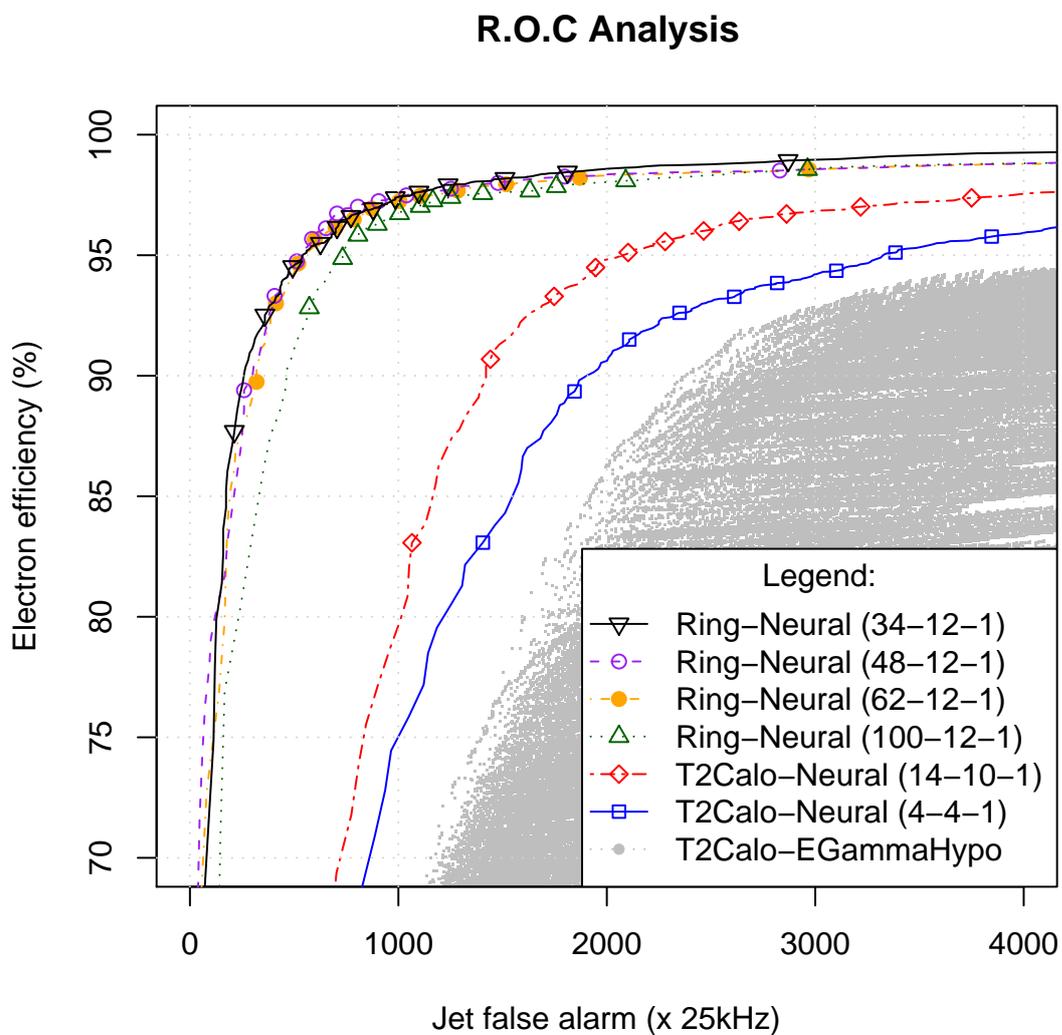


Figura 6.61: R.O.C. comparativa entre os sistemas de detecção de elétrons e jatos analisados até agora e o novo sistema de detecção gerado a partir da poda dos canais de entrada baseada na relevância.

Tabela 6.11: Posicionamento relativo dos anéis produzidos pelo sistema de anelamento podado para ter apenas 34 (dos 100) anéis.

Camada	Número de anéis	Primeiro	Último
Pré-irradiador	4	1	4
1 <sup>a</sup> e.m.	12	5	16
2 <sup>a</sup> e.m.	6	17	22
3 <sup>a</sup> e.m.	4	23	26
1 <sup>a</sup> hadrônica	4	27	30
2 <sup>a</sup> e 3 <sup>a</sup> hadrônicas	4	31	34

nesta camada: múltiplas células sensibilizadas dentro de uma mesma RoI aumentam a probabilidade do objeto em análise não ser um elétron. De fato, a variável  $R_{e.m.1}^{2\text{-máximos}}$  do T2Calo assume exatamente este papel, como explicado no início do Capítulo 5. Por esta razão, é bastante intuitivo que os dados na primeira camada e.m. apresentem um padrão de relevância distinto aos das demais camadas. De fato, o mesmo fenômeno pode também ser observado na Figura 6.56; embora, neste contexto, a informação esteja menos evidente, dado o contraste entre valores de relevância baixos para os outros anéis. Para as outras camadas, a relevância parece ser maior para o segundo anel que para o primeiro, caindo à medida que se afasta do centro.

Uma vez que a eficiência deste novo sistema de detecção não foi afetada pela poda e, que observa-se a existência de relevâncias de discriminação negativas, propõe-se uma quarta estratégia de poda com a seguinte configuração:

1. **Pré-irradiador:** não se utilizarão os dados desta camada;
2. **1<sup>a</sup> camada e.m.:** utilizar apenas 6 dos 64 anéis iniciais, com a mesma granularidade anterior;
3. **2<sup>a</sup> camada e.m.:** utilizar apenas 4 dos 8 anéis iniciais, com a mesma granularidade anterior;
4. **3<sup>a</sup> camada e.m.:** não se utilizarão os dados desta camada;



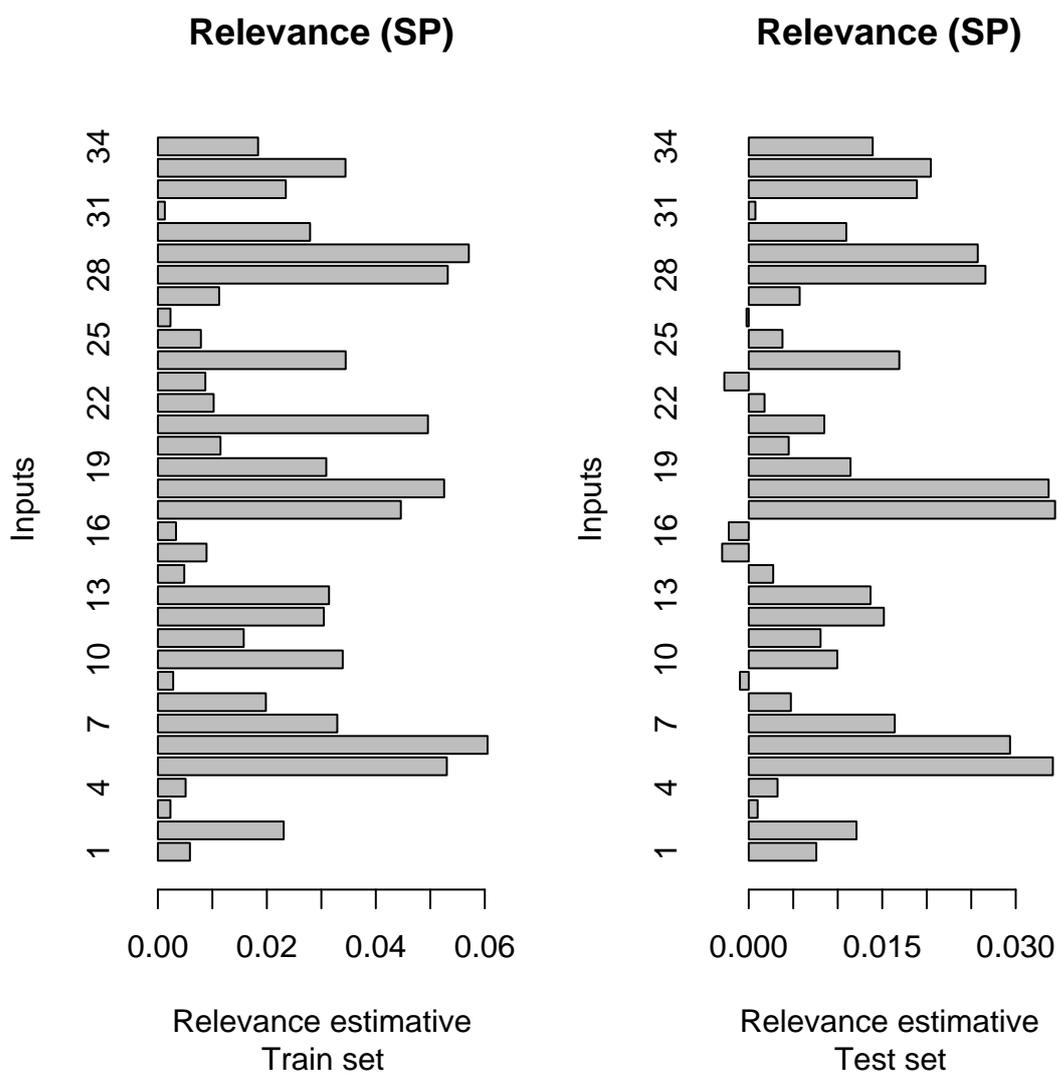


Figura 6.62: Relevâncias de discriminação para o detetor neural baseado em 34 (dos 100) anéis resultantes da poda por relevância.

5. **1<sup>a</sup> camada hadrônica, 2<sup>a</sup> e 3<sup>a</sup> camadas hadrônica:** aglutinar em uma mesma “camada” virtual, mantendo a granularidade anterior.

Um total de 14 anéis restam nesta nova configuração. Este número de variáveis é idêntico à situação de melhores resultados com as características extraídas pelo T2Calo. Mantém-se os parâmetros de treinamento e o número de neurônios na camada escondida do detetor neural inicial. A Figura 6.63 mostra a curva R.O.C. de um dos 10 testes realizados, junto à curva R.O.C. de outros detetores. Mais uma vez o sistema não apresenta perda na eficiência de classificação, comparado ao sistema original com 100 anéis. O ponto de máximo do produto SP (impressionantes 1,83!) ocorre quando a eficiência na detecção de elétrons é de 97,24% contra um falso-alarme de apenas 2,93% (730 Hz).

A Figura 6.64 contém os valores de relevância de discriminação para a segunda poda baseada no critério da relevância de discriminação. A Tabela 6.12 mostra a relação entre os números neste gráfico e a informação dos anéis. Neste novo sistema, a parte hadrônica assume uma importância secundária. A informação periférica na segunda camada e.m. parece ser a mais relevante para este último classificador, que também teve o melhor desempenho até o momento. Como colocado anteriormente, espera-se que a informação de detecção esteja realmente na periferia da RoI e, neste exemplo, é possível distinguir este efeito de forma bastante evidente. Não mais se vêem níveis de relevância de discriminação negativos, o que indica que estamos no limite do que é possível com esta configuração de anéis. Ademais, nota-se que a maior parte da informação de interesse está na periferia do objeto considerado. Por exemplo, o anel 5 é o mais relevante na 1<sup>a</sup>camada e.m., o anel 10, o mais relevante na 2<sup>a</sup>camada e.m. e, finalmente, os anéis 13 e 14 para a seção hadrônica combinada.

Desta forma, uma otimização que remova as componentes menos relevantes, desvinculando-se do problema de recuperação dos dados do sistema de leitura, e assumindo-se que os dados para os 14 anéis serão recuperados de qualquer maneira, não parece interessante. Ademais, como a análise de relevância indica, haverá possível perda de eficiência. Em troca, eliminar-se-ão algumas poucas conexões sinápticas, o que, finalmente, como iremos ver no Capítulo 7, não impactará fundamentalmente no tempo de execução do algoritmo. Por outro lado espera-se que, preservando anéis menos relevantes, construiremos um detetor fundamentalmente

### R.O.C Analysis

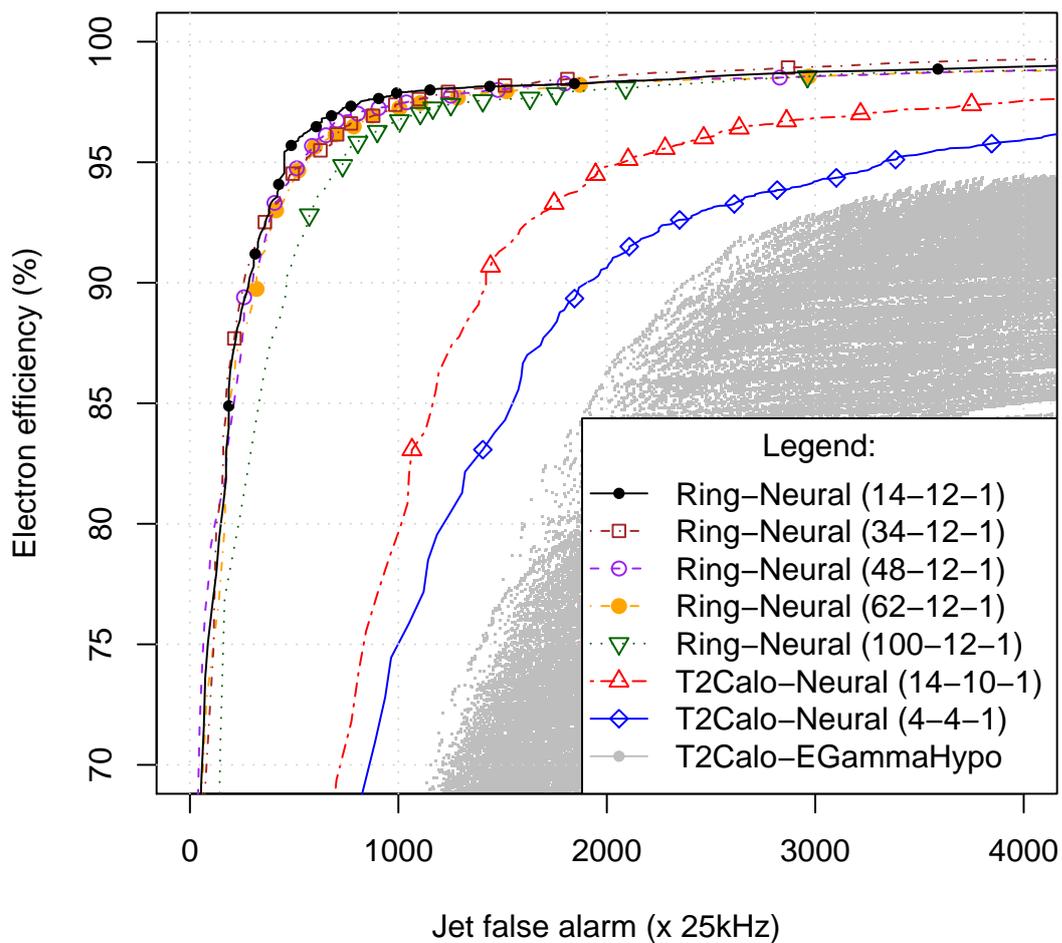


Figura 6.63: R.O.C. comparativa entre alguns dos sistemas de detecção de elétrons e jatos analisados até agora e o novo sistema de detecção baseado na poda da relevância, com apenas 14 (dos 100) anéis.

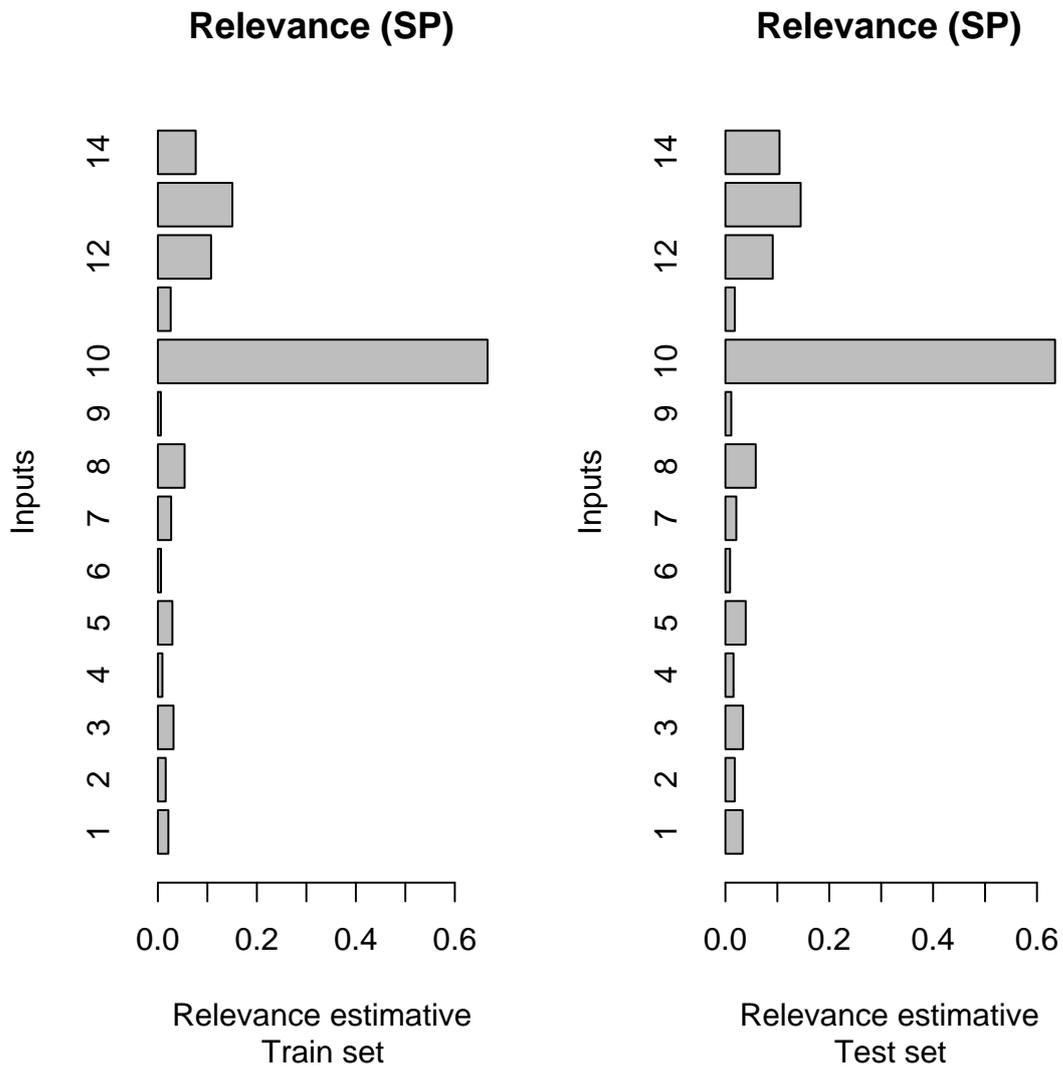


Figura 6.64: Relevâncias de discriminação para o detetor neural baseado em 14 (dos 100) anéis resultantes da poda por relevância.

Tabela 6.12: Posicionamento relativo dos anéis produzidos pelo sistema de anelamento podado usando a análise de relevância discriminante, para ter apenas 14 (dos 100) anéis.

Camada	Número de anéis	Primeiro	Último
1 <sup>a</sup> e.m.	6	1	6
2 <sup>a</sup> e.m.	4	7	10
1 <sup>a</sup> 2 <sup>a</sup> e 3 <sup>a</sup> hadrônicas	4	11	14

mais robusto. Para averiguarmos se a informação de anéis relevantes pode ser extraída de outros anéis menos relevantes, artificialmente podamos a variável mais importante (de número 10 na Figura 6.64), representando o anel mais periférico na segunda camada e.m..

Treina-se novamente um conjunto de 10 redes neurais com base nas 13 entrada menos relevantes. O resultado é como esperado: este novo sistema consegue recuperar o sinal perdido da variável mais relevante apresentando, no melhor dos casos, uma eficiência na detecção de elétrons de 97,59% para um falso-alarme na detecção de jatos de 3,22% (SP=1,84). A Figura 6.65 mostra o gráfico da relevância de discriminação deste novo detetor. A ordem de importância das relevâncias se mantém aproximadamente igual, resguardado o fato de que se perdeu a informação do 4ºanel da 2ªcamada e.m..

Desta forma conclui-se que o sistema baseado em 14 dos 100 anéis seja mais que o necessário para implementar um discriminador elétron/jato mais eficiente que a técnica atualmente proposta no CERN (T2Calo+EGammaHypo). Este sistema possui uma capacidade de discriminação com quase 98% de acerto em elétrons para uma taxa de falso-alarme de pouco mais que 3%. Estes valores representam uma redução na taxa de ruído de fundo 3 vezes maior e, desta forma, economizando tempo de processamento para atividades inexoravelmente mais complexas, como a detecção de traços ou a *quasi*-reconstrução executada no Filtro de Eventos do experimento ATLAS.

Este sistema contém elementos que garantem sua robustez à falhas de leitura e processamento dos dados do detetor e, como veremos mais adiante, pode ser implementado de forma extremamente veloz. A técnica de análise por relevância de discriminação permitiu a busca de tal sistema de detecção, fornecendo um conjunto de medidas que otimizam a poda de variáveis pouco interessantes ou que atrapalhem o processo de seleção. O detetor final utiliza-se da informação de um sub-espaco da RoI, para apenas duas das quatro camadas da seção e.m. dos calorímetros do detetor ATLAS, mantendo a utilização de todas as camadas na seção hadrônica, com um formato compacto.

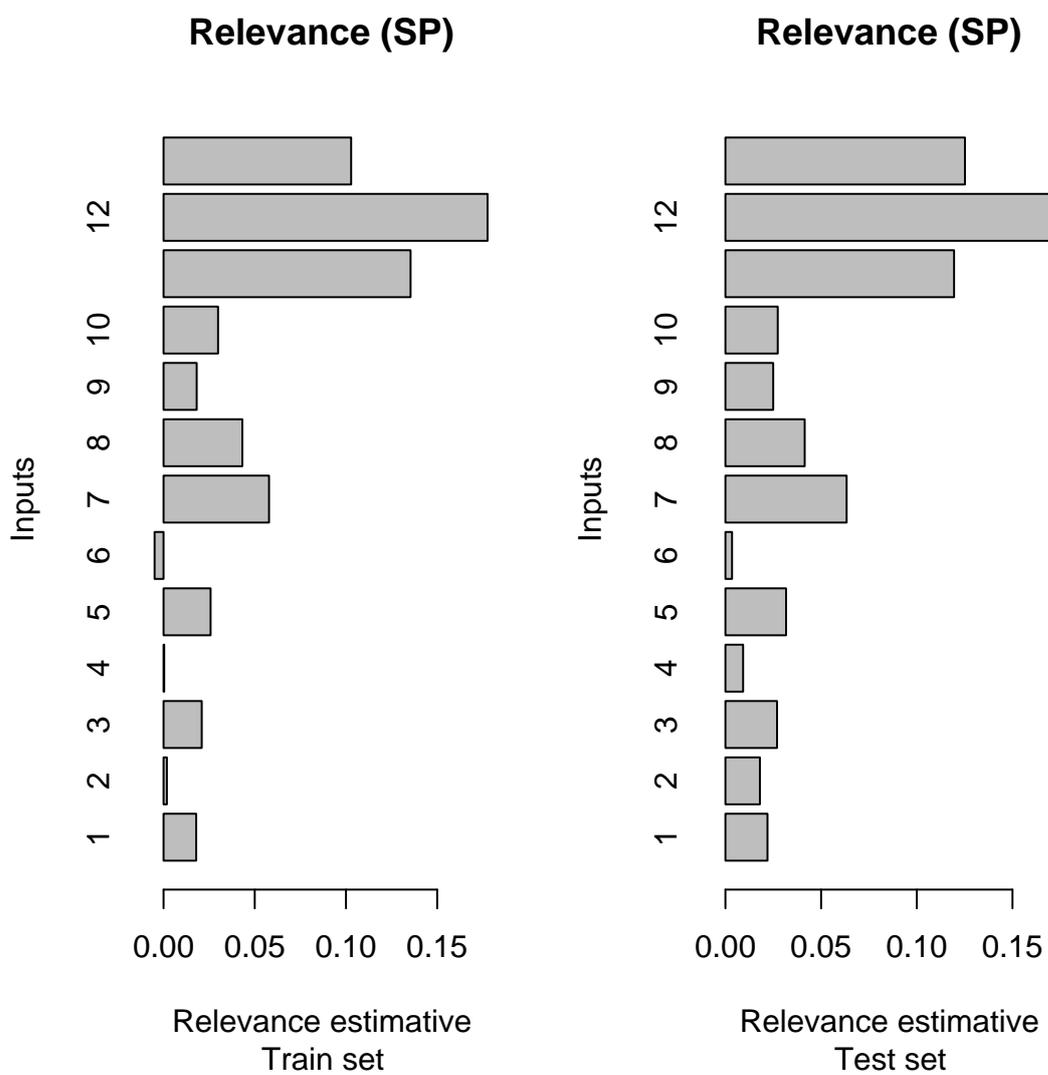


Figura 6.65: Relevâncias de discriminação para o detetor neural baseado em 13 (dos 14) anéis menos relevantes do último detetor.

## Capítulo 7

# Desempenho computacional do sistema de discriminação

O algoritmo apresentado nesta seção deve respeitar as normas de operação do Segundo Nível de Filtragem do experimento. Neste nível de filtragem, espera-se que o tempo médio de processamento para cada evento, incluindo acesso aos dados no sistema de leitura, seja de 10 milissegundos. No entanto, espera-se que o sistema opere rejeitando eventos ordinários o mais rápido possível, de tal forma que, estatisticamente, a maior parte do tempo despendida nos processadores do LVL2 seja dedicada a eventos interessantes.

Levando-se em consideração que o tempo de acesso aos dados no sistema de leitura está na ordem de 1 ms, como descrito no Capítulo 4, deseja-se minimizar o tempo de processamento despendido no algoritmo. Desta forma, medidas de desempenho são normalmente executadas nos candidatos a algoritmos ao sistema de filtragem, de modo a determinar seu tempo de execução e possíveis pontos a serem otimizados. O correto balanço entre a rejeição prematura de um evento e o número de requisições de dados deve ser feita, de forma a otimizar tanto seu acesso quanto seu processamento. Isto ocorre pois existe uma dependência direta entre os dois parâmetros a serem considerados: o número de requisições de dados interrompe o processamento para se aguardar a informação requisitada. Quanto mais informação, mais processamento deverá ser realizado para o tratamento dos dados. Deseja-se, desta forma, minimizar o acesso a dados e maximizar a capacidade discriminativa do sistema.

Tabela 7.1: Tempo de processamento médio (em milissegundos) das diversas fases do algoritmo T2Calo em uma máquina com processadores Intel Xeon de 2,4 GHz e 1 Gb de memória RAM.

Fase	Prep. Dados	Algoritmo	Tot. parcial
$R_{e.m.2}^{3 \times 7 / 7 \times 7}$	0,45	0,49	0,94
$R_{e.m.1}^{2\text{-máximos}}$	0,48	0,57	1,05
$E_{T_{3 \times 7}}^{e.m.}$	0,56	0,17	0,73
$E_{T_{0,2 \times 0,2}}^{HAD-1}$	0,59	0,33	0,92
<b>Total</b>	<b>2,08</b>	<b>1,56</b>	<b>3,64</b>

Nossa referência, o algoritmo do T2Calo, foi completamente descrita na Seção 5.4. Neste algoritmo, inicia-se o processamento detetando-se o pico na segunda camada e.m., seguindo-se do cálculo da variável  $R_{e.m.2}^{3 \times 7 / 7 \times 7}$ . A variável  $R_{e.m.1}^{2\text{-máximos}}$  é calculada em seguida, utilizando-se dos dados da primeira camada e.m.. De posse dos valores parciais de energia nestas duas camadas, os dados do pré-irradiador e da terceira camada e.m. são utilizados para definir o valor da variável  $E_{T_{3 \times 7}}^{e.m.}$ . Na última parte do processamento define-se o valor da variável  $E_{T_{0,2 \times 0,2}}^{HAD-1}$ , acessando-se os dados da parte hadrônica. A Tabela 7.1 contém as mais recentes medidas realizadas na avaliação do desempenho do T2Calo [80]. Os valores nesta tabela representam o tempo médio de processamento, em milissegundos, para cada uma das fases. Estes tempos também desconsideram a execução do algoritmo de hipótese EGammaHypo. Assume-se no entanto, que seja desprezível, dado sua simplicidade.

Tomando-se por base que um sistema baseado no anelamento e detecção neural terá que acessar os mesmos dados das mesmas fontes que aquelas acessadas pelo T2Calo, em uma primeira aproximação e sem se considerar quaisquer processos de otimização, a primeira parte do tempo (Calibração e Localização na tabela) deve ser tomada como constante. A segunda parte, no entanto, é completamente dependente do algoritmo e devemos compará-la ao tempo de execução do *NeuralRinger*.

A complexidade de execução do método baseado no anelador e detecção neural é variável de acordo com dois parâmetros principais:

1. Com o número de anéis que se deseja extrair: neste caso, quanto mais anéis



Tabela 7.2: Configurações das máquinas utilizadas para o teste de desempenho do *NeuralRinger*.

Processador	<i>Clock</i>	<i>Cache</i> de L2	Memória RAM
AMD Opteron (250)	2,4 GHz	1 Mb	4 Gb
AMD Opteron (275)	2,2 GHz	1 Mb	4 Gb
Intel Xeon	2,4 GHz	512 kb	1 Gb

deseja-se extrair dos dados, maior o tempo de execução que será necessário para cumprir o processo de anelamento. Ademais, com o aumento do número de anéis aumenta-se igualmente o tamanho da entrada do discriminador neural e, portanto, o número de operações aritméticas para esta última fase do processamento;

2. O número de neurônios escondidos na rede neural também influenciará o tempo de execução do discriminador.

Para determinar a viabilidade em termos de desempenho, para o algoritmo proposto por este trabalho, executou-se a aplicação `ringer-run` em um conjunto de máquinas de referência ao experimento ATLAS. A Tabela 7.2 resume as características destas máquinas. A primeira das máquinas listadas está sendo utilizada em uma avaliação de arquiteturas para o sistema de filtragem e aquisição de dados do ATLAS, conhecido como *pré-série* [81]. Esta é a máquina de referência máxima para estudos de desempenho no ATLAS. A segunda máquina de teste é um modelo com dois núcleos (*dual-core*). A arquitetura desta máquina está sendo considerada como opção às máquinas atualmente empregadas nas pré-séries, sendo este modelo de processador o mais recente disponibilizado pelo fabricante. A terceira e última máquina faz parte de um conjunto de máquinas dedicado a estudos do sistema de filtragem. Embora possua uma arquitetura mais antiga, ainda é usada como referência em muitos trabalhos, sendo este o caso das medidas do T2Calo apresentadas acima.

Estas máquinas contêm uma instalação completamente funcional do sistema operacional atualmente empregado no CERN, o SLC3 [82]. Embora alguns dos proces-

sadores possam operar em 64-bits (assim como o *NeuralRinger*), estas máquinas são operadas em modo de compatibilidade para 32-bits, já que grande parte do *software* do ATLAS ainda não opera corretamente em 64-bits.

Um conjunto idêntico, composto de três diferentes testes foram executados em cada uma das máquinas, utilizando todos os dados disponíveis (cerca de 22.600 elétrons e 7500 jatos) para este estudo. Para estes testes, utilizou-se o programa *ringer-run*, que executa as fases de anelamento e discriminação em seqüência, como aconteceria num sistema *online*. Os testes realizados em uma mesma máquina diferem entre si somente pela configuração de anelamento e da rede neural que discriminará os dados após a extração dos anéis. As combinações utilizadas equivalem àquelas dos testes mostrados no Capítulo 6, para o sistema com 100 anéis, um sistema podado com 34 anéis e o último, utilizando apenas 14 anéis mais relevantes. Durante os testes, observou-se a quantidade de memória RAM disponível na máquina de tal forma que o carregamento dos dados não ativasse o mecanismo de troca de disco (do inglês, *swap*), já que isso faria com que a aplicação tivesse um desempenho abaixo do esperado para o experimento.

As Figuras 7.1, 7.2 e 7.3 contém os resultados dos 3 testes com 100, 34 e 14 anéis, respectivamente, nas três arquiteturas. Das três arquiteturas escolhidas, a máquina da pré-série parece ser a mais rápida, apresentando um tempo de processamento de apenas 445 microssegundos para o mais complexo dos testes, utilizando 100 anéis. Em seguida, o sistema baseado no processador Opteron-275, cerca de 8% mais lento para o mesmo teste. A máquina de testes do sistema de filtragem é a mais lenta de todas as plataformas, apresentando um tempo médio de processamento de 542 microssegundos para cada RoI. O desvio-padrão está na faixa dos 10%, indicando uniformidade nos resultados.

Nos testes seguintes, a relação entre o desempenho das máquinas se mantém. Para o teste com 34 anéis o tempo de processamento médio na máquina mais rápida é de 215 microssegundos, tendo caído aproximadamente a metade do valor para os 100 anéis. Para o teste com apenas 14 dos 100 anéis originais, o tempo médio de processamento é de apenas 125 microssegundos para a máquina mais veloz, representando quase um-quarto do tempo de processamento para o sistema de anelamento completo.

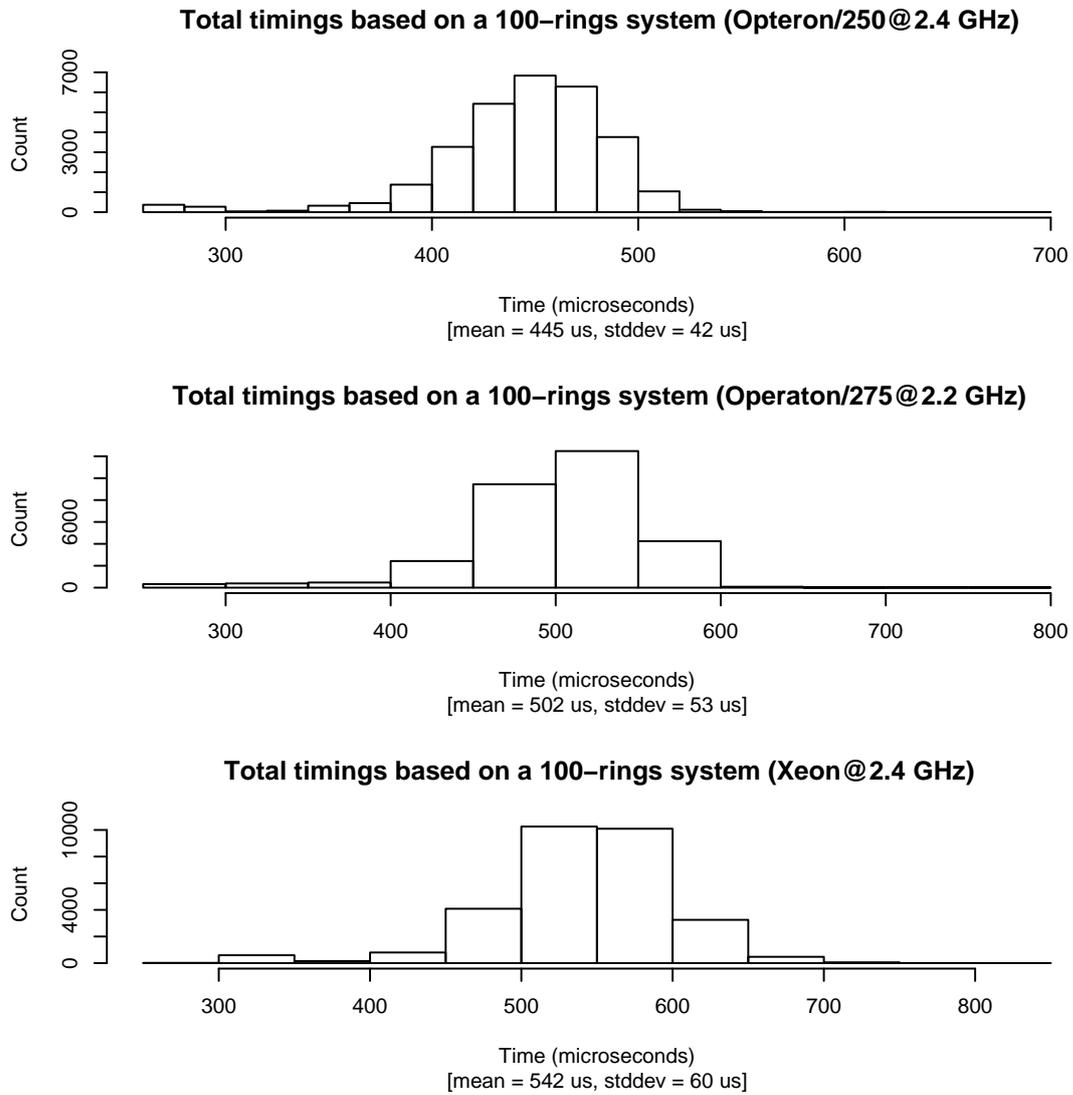


Figura 7.1: Tempos totais de execução do *NeuralRinger* em três plataformas distintas, utilizando 100 anéis.

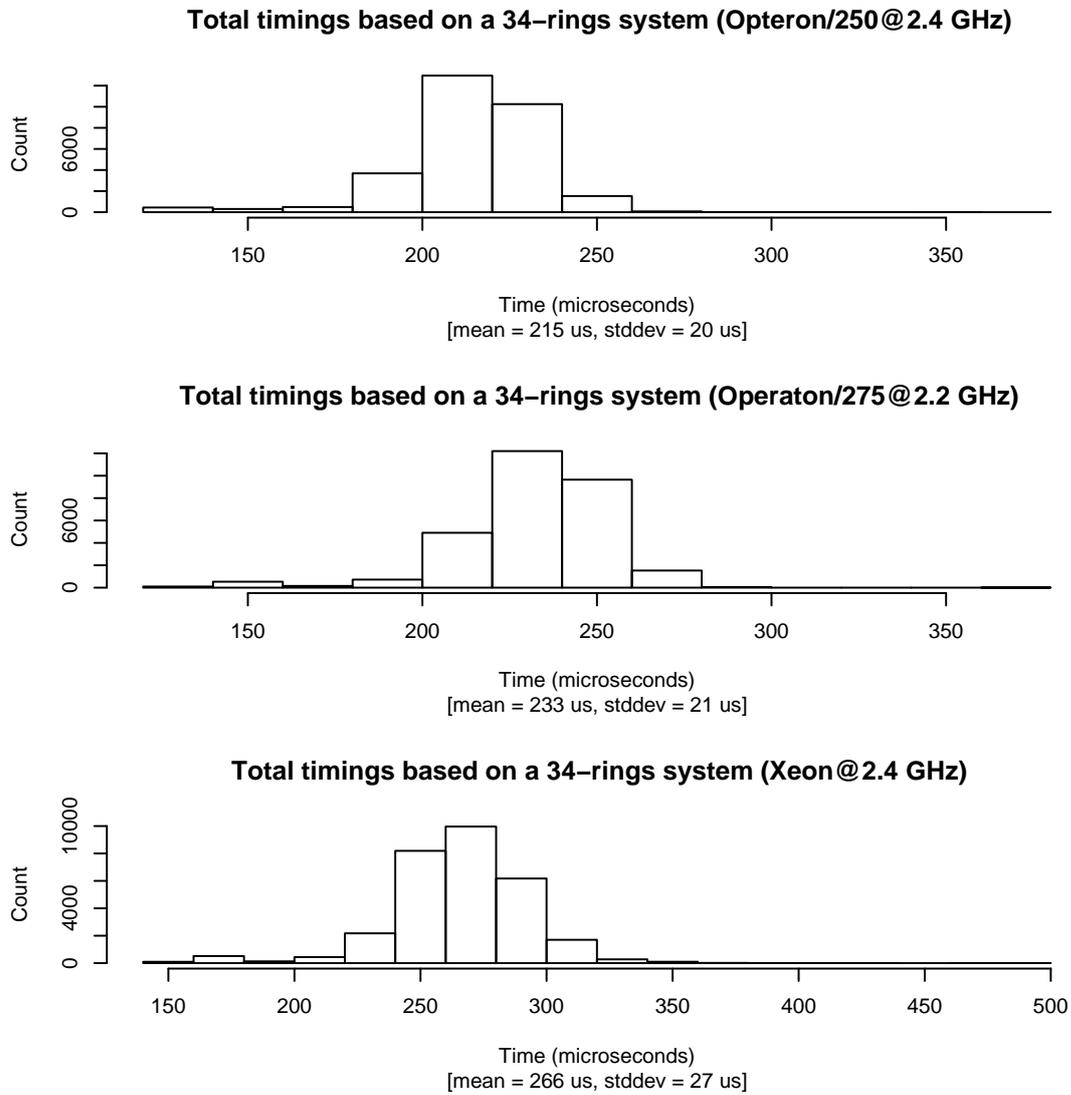


Figura 7.2: Tempos totais de executava do *NeuralRinger* em três plataformas distintas, utilizando 34 anéis.

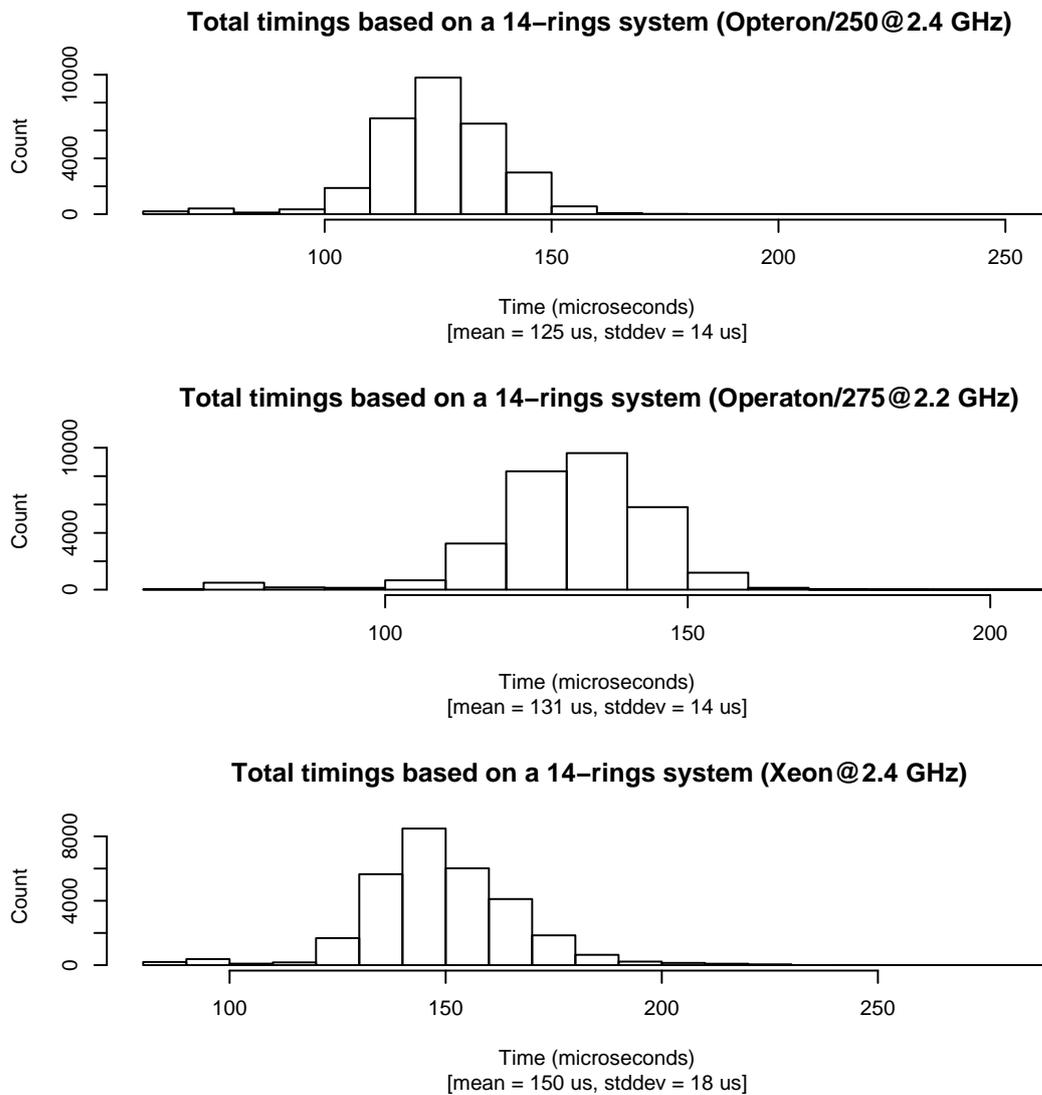


Figura 7.3: Tempos totais de execução do *NeuralRinger* em três plataformas distintas, utilizando apenas 14 anéis.

Ao compararmos os resultados dos testes com 100 anéis ao tempo de referência do T2Calo (1,56 ms) para a mesma plataforma, observa-se que o sistema proposto executa em apenas um-terço do tempo requerido para o T2Calo nesta arquitetura. Esta análise não leva em conta o acesso aos dados e sua calibração e localização dentro das bibliotecas de infraestrutura do LVL2. No entanto, por inspeção à Tabela 7.1, nota-se que cerca de 57% do tempo total do processamento de uma RoI, isto é, aproximadamente 2,1 ms é devido a esta parte dos dados. Desta forma, conclui-se que o sistema proposto pelo anelador está dentro das especificações de desempenho para o sistema de filtragem do ATLAS.

A Figura 7.4 mostra os histogramas individuais para cada uma das fases de processamento, para o teste usando 100 anéis rodando na máquina mais rápida (Opteron/250, 2,4 GHz). Desta figura conclui-se que, no caso de tempos de processamento menores se tornarem imperativos, uma implementação dedicada do processo de anelamento (originalmente implementado na biblioteca `rbuild`) poderá ser realizada de forma que o tempo total de deste algoritmo seja otimizado. Junto à discriminação neural as duas fases representam mais de 90% do tempo de processamento total. Otimizações na fase de procura do pico de deposição energética na segunda camada e.m. ou na fase de normalização dos anéis teriam quase ou nenhum impacto considerando-se a arquitetura atual do *NeuralRinger*.

## 7.1 Migração para o Sistema de Filtragem do ATLAS

Algoritmos candidatos ao Sistema de Filtragem do ATLAS não devem somente ter bom desempenho, mas também atender todos os critérios e restrições de operação neste árduo ambiente. Em específico, o ambiente de trabalho deste sub-sistema do experimento contém as primitivas de acesso aos dados e calibração assim como a lógica que coordena o processo de filtragem. Mesmo embebido dentro desta, muitas vezes pesada, arquitetura, o algoritmo deve manter um desempenho compatível seu propósito: filtragem ultra-veloz. Nesta seção abordamos a arquitetura e implementação do sistema de anelamento e processamento neural dentro do Sistema de Filtragem do experimento ATLAS e, sobretudo, ao ambiente de desenvolvimento Athena [41, 42], que permite o transporte destas ferramentas entre os dois mundos

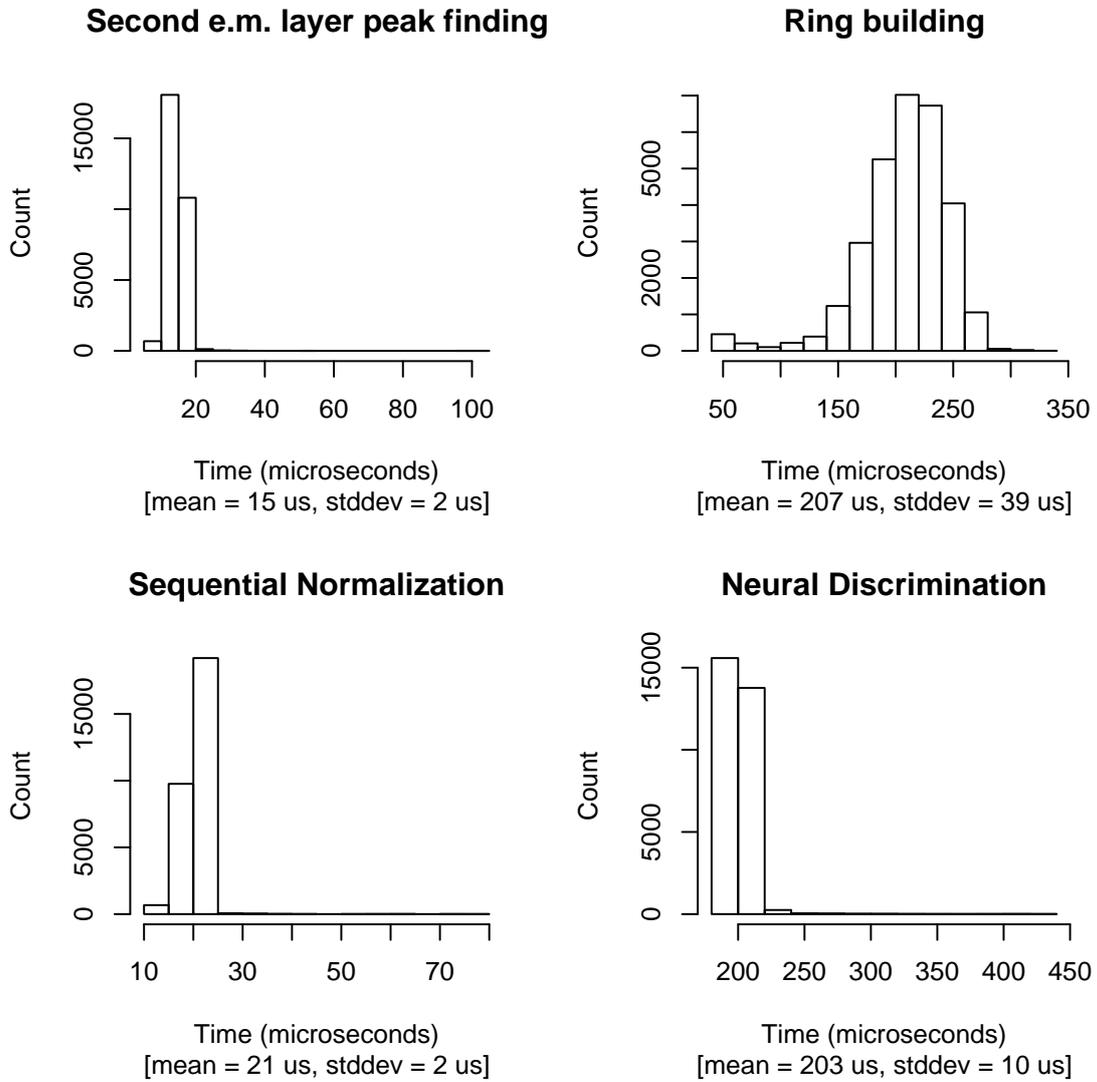


Figura 7.4: Tempos individuais para cada fase de execução do *NeuralRinger*, levando-se em consideração a extração e discriminação baseada em 100 anéis. O teste foi executado na plataforma Opteron/250 (pré-serie).

do experimento: filtragem e análise *offline*.

O *NeuralRinger* evoluiu ao longo de 2 anos de desenvolvimento de forma que se tornasse facilmente integrável ao ambiente de funcionamento *online*, que é baseado no ambiente de desenvolvimento Athena. Uma vez integrado, o *NeuralRinger* poderá extrair anéis da RoI seguindo uma determinada configuração de anelamento e aplicar um processo de decisão baseado em uma rede neural pré-configurada. A implementação de um detetor neural baseado na técnica de anelamento, neste contexto, deve seguir os seguintes passos:

1. Encontrar os dados que serão utilizados para treino e teste do discriminador neural. Rodar o programa `athena`, de forma a extrair dados no formato nativo do *NeuralRinger*;
2. Projetar e treinar o sistema de anelamento e detetor neural no ambiente do *NeuralRinger*, totalmente desconectado do Sistema de Filtragem do ATLAS;
3. Aplicar as configurações de anelamento e detecção neural no sistema *online*.

Desta forma, é possível desacoplar o desenvolvimento de detetores ao uso do ambiente Athena, o que é uma vantagem considerável: diminui-se o tempo de desenvolvimento e permite-se que o usuário foque sua atenção ao problema da classificação sem se importar com os detalhes, muitas vezes complexos, deste ambiente. Os resultados da aplicação de um detetor específico, dentro e fora do ambiente Athena devem ser idênticos.

### 7.1.1 *NeuralRinger* e o Athena

Para que seja facilmente integrável dentro do ambiente Athena, o *NeuralRinger* segue a filosofia de funcionamento deste ambiente. Cada uma das fases de processamento, i.e., a criação dos anéis de energia e a classificação neural estão codificadas em bibliotecas individualizadas e podem ser utilizadas tanto conjunta quanto separadamente. No ambiente *online*, a extração de características e detecção são normalmente separadas de forma que possam ser desenvolvidas independentemente.

O processamento do evento dentro do sistema filtragem ocorre de forma seqüencial, como explicado na Seção 4.2.6. Para cada objeto ou RoI destacado pelo nível



de filtragem antecedente, o componente do HLT conhecido como *Steering* (que poderia ser traduzido, neste contexto, como Coordenador ou Agendador) irá definir o algoritmo que será chamado para tratá-lo. O algoritmo poderá realizar um conjunto de operações no objeto em questão, aumentando a quantidade de informações disponíveis para análise deste ítem ou interrompendo o processamento.

Desta forma, é possível definir duas classes de algoritmos:

- **Extração:** Algoritmos que apenas adicionam informação a um objeto no evento, através da extração de características nos dados do detetor;
- **Hipótese:** Algoritmos que interrompem o processamento de um objeto por considerarem-no como física ordinária que não deve ser registrada em mídia permanente.

Sendo, o processo de hipótese, separado do processo de extração, é possível desenvolver sistemas de decisão mais e mais complexos conforme se avança dentro da análise do evento. A Figura 7.5 exemplifica este cenário para uma configuração fictícia para o *Steering*. Este é um exemplo que pode ser reproduzido no LVL2: um objeto tipo e.m. é encontrado no resultado do LVL1 passado ao *Steering*. Este objeto causa o agendamento de um algoritmo que possa tratá-lo. A primeira fase do processamento é a extração de características do objeto. Neste exemplo, a primeira extração de característica considerada é a extração dos anéis. No passo seguinte a extração é a hipótese (neural). Se o objeto atender às características de um elétron, o *Steering* irá agendar o passo seguinte. Caso, não, a hipótese de que o objeto seja um objeto tipo e.m. é rejeitada e o processamento é parado, causando a rejeição do evento pelo sistema de filtragem. Caso a hipótese seja afirmativa, neste exemplo, tentar-se-á encontrar um traço relativo ao objeto nos detetores de traço. Caso isto falhe, o objeto será rejeitado como indica a figura. Se, finalmente, nenhuma das fases de processamento conseguir rejeitar o objeto, ele será rotulado como um elétron, os valores extraídos das características do objeto serão adicionados a este novo elemento e o evento será aprovado.

Para atender à alta complexidade do experimento, muitas seqüências como a descrita na Figura 7.5 podem co-existir dentro do *Steering*. Cabe a este sistema de

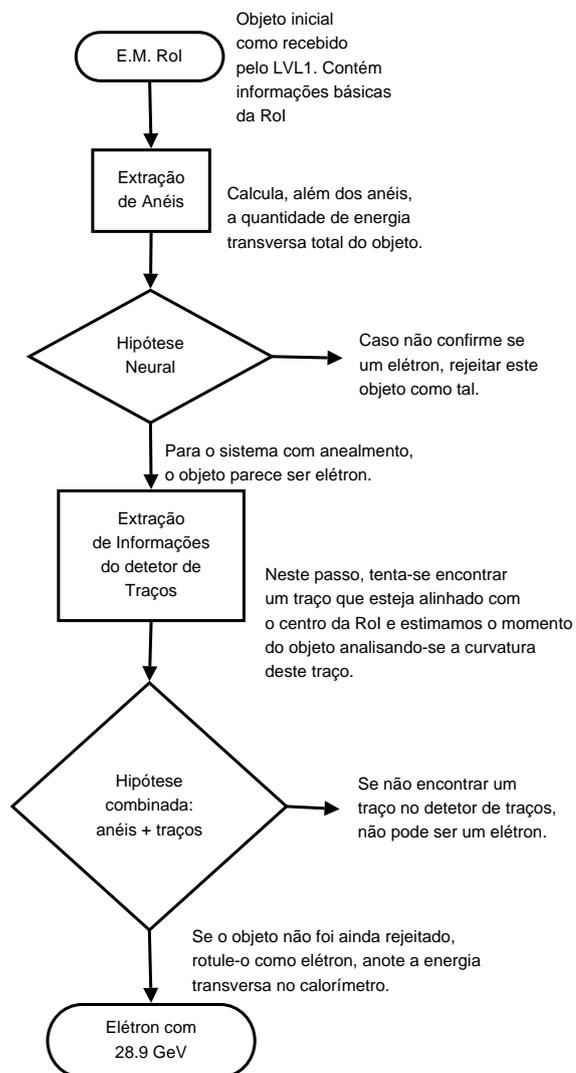


Figura 7.5: Exemplo de um cenário de seleção simples para o *Steering* operando no LVL2, baseado em uma Rol tipo E.M..

agendamento escolher a seqüência de passos e combinações de objetos apropriadas para que se possa analisar o evento adequadamente.

Para a integração do *NeuralRinger* e observando-se a estrutura de processamento *online*, divide-se o anelamento e a decisão neural em dois passos distintos, que podem ser executados independentemente. Para evitar a dispersão de código, foram criados 3 pacotes:

- **TrigRingerTools**: Este pacote contém as bibliotecas originais do *NeuralRinger* e é usado como base funcional dos outros pacotes;
- **TrigCaloRinger**: Este pacote implementa um algoritmo de extração de características. As características neste caso são os valores energéticos dos anéis;
- **TrigMultiVarHypo**: Este pacote implementa um algoritmo de hipótese baseada em redes neurais. Nele encontra-se a implementação do algoritmo de hipótese baseada em redes neurais.

Com esta configuração é possível manter o conjunto de bibliotecas do *NeuralRinger* coeso em um único pacote (**TrigRingerTools**) e apenas implementar um conjunto de algoritmos que usam a funcionalidade de base deste pacote. Os algoritmos de extração de características e hipótese são configuráveis em todos os aspectos da sua versão desacoplada e podem escrever em disco os resultados parciais de suas operações para verificação externa ao Athena.

Para determinar o correto funcionamento deste sistema, é necessário testar, baseado na entrada, a saída de cada passo, para um conjunto de eventos. Para tal, utilizou-se uma base de dados com eventos simulados tipo  $Z \rightarrow e^- + e^-$  que estava disponível no formato adequado. Uma configuração de anelamento utilizando 100 anéis conforme descritos anteriormente, e uma rede neural com 100 entradas, 8 neurônios escondidos e um neurônio na camada de saída. Cerca de 850 RoI's foram selecionadas pela simulação do LVL1 indicando objetos a serem averiguados no LVL2. Roda-se o sistema Athena, configurado para utilizar os algoritmos de anelamento e detecção neural extraíndo-se 3 tipos de dados:

1. O valores de energia e posicionamento das células utilizadas para o cálculo dos anéis, num formato compatível com o padrão do *NeuralRinger*;

2. Uma base de dados de anéis já processados e normalizados, em formato XML;
3. Uma base de dados com as saídas do processamento neural, também no formato XML.

Com base na versão desacoplada (*standalone*, não integrada ao ambiente Athena ou ao Sistema de Filtragem) do *NeuralRinger*, executam-se os mesmos processos:

1. Baseado nos valores originais das células provida pelo Athena, calculam-se a soma de anéis;
2. Baseado na saída dos anéis provida pelo Athena, calculam-se a saída da rede neural.

A Figura 7.6 mostra as diferenças, RoI a RoI, nos cenários 1 e 2 descritos anteriormente. Como é possível observar no primeiro histograma desta figura, os resultados são idênticos com uma precisão melhor que  $1 \times 10^{-7}$  no caso do Teste 1. Esta pequena diferença existe pois os valores relativos às energias das células, escritos em disco para a análise pelo *NeuralRinger*, são truncados na décima casa decimal, por opção de projeto. A propagação deste erro através da extração de anéis induz um erro relativo aparentemente maior na saída deste processo.

Para o Teste 2, no histograma na parte inferior da Figura 7.6, o resultados são idênticos para uma precisão melhor que  $2 \times 10^{-6}$ . Este erro existe pois os valores relativos às energias dos anéis, escritos em disco para a análise pelo *NeuralRinger*, são truncados a partir da sexta casa decimal.

Para completar a análise, roda-se o processo completo de extração e classificação neural, baseando-se nos valores das células, comparando a saída deste procedimento à saída provida pelo sistema neural acoplado ao ambiente Athena. A Figura 7.7 mostra um gráfico de correlação das saídas dos dois sistemas para as configurações de anelamento e rede neural descritas anteriormente. Neste caso, onde não há escrita em disco entre os dois passos do processamento, o erro observado é zero, entre os dois sistemas. Estes resultados, portanto, indicam que a precisão de escrita é suficiente para a reprodução desacoplada e que o sistema migrado funciona adequadamente.

O segundo passo é a confirmação dos tempos de operação destacados ao longo do texto. Neste caso, deseja-se averiguar se a migração não acarretou algum impacto ao

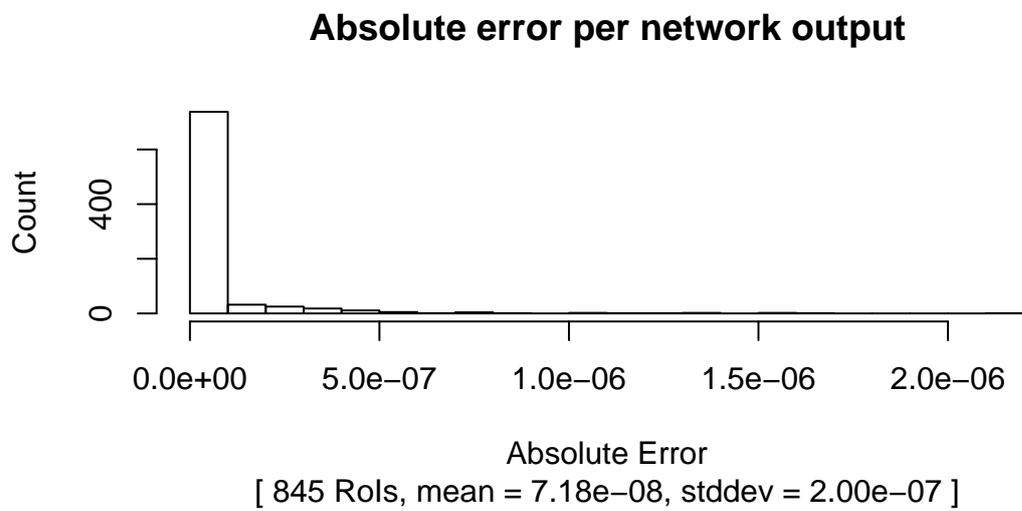
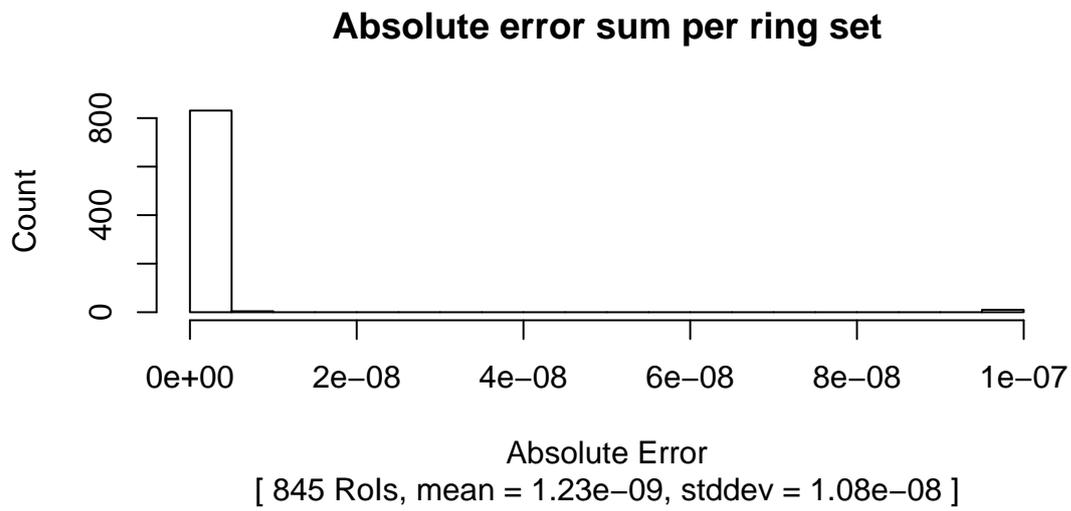


Figura 7.6: Erros entre os processos de extração de características e hipótese neural entre o Athena e o *NeuralRinger* rodando em modo desacoplado.

### Full chain Athena vs. NeuralRinger output

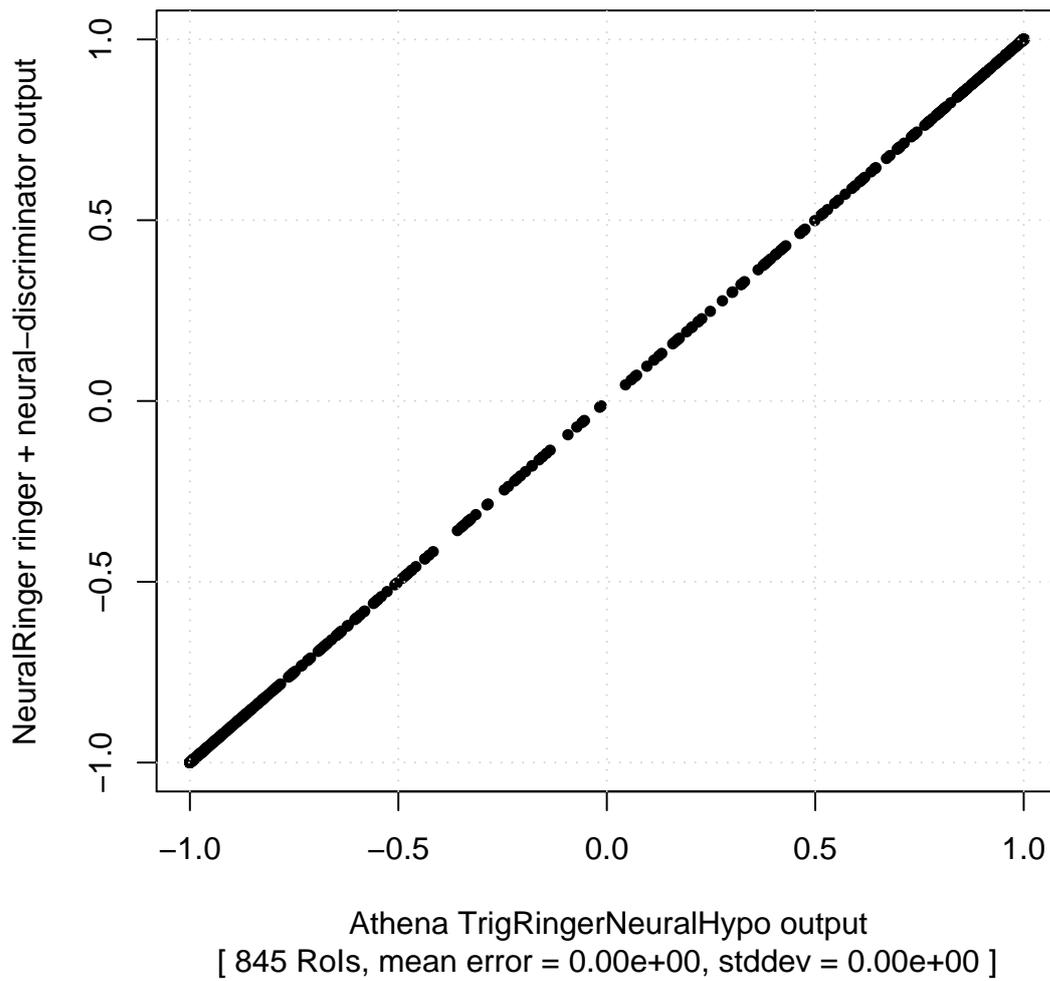


Figura 7.7: Gráfico de correlação entre a saída final do Athena e de uma versão do *NeuralRinger* rodando em modo desacoplado.

desempenho do sistema, uma vez que uma infraestrutura maior de suporte à execução está sendo utilizada. Ademais, uma vez que se estará realizando acesso remoto aos dados e calibração, processos que não são executados no ambiente desacoplado, deseja-se estimar o impacto que estas novas atividades causariam ao emprego do anelamento e discriminação neural no Sistema de Filtragem. Por outro lado, obtém-se com este estudo um conjunto de valores comparativos mais justos a outros métodos de detecção, tais como o sistema proposto pelo T2Calo+EGammaHypo. Isto acontece pois estaremos interfaceando com as diversas partes da infraestrutura de processamento do Sistema de Filtragem da mesma forma. Por questões de praticidade ligadas à disponibilidade dos dados e quantidade de memória disponível, escolhe-se a máquina baseada no processador AMD Opteron 275 utilizada anteriormente, para realização dos testes.

**Serviço de Temporização do Athena:** O serviço de temporização do Athena (no pacote *TrigTimeSvc*) utiliza a função `gettimeofday` [83] para executar a operação de medida de tempo entre dois pontos quaisquer dentro do código. A utilização de um relógio em tempo-real é benéfica quando se deseja medir tempos de processamento com uma precisão menor que 10 milissegundos (fornecida pela função `clock`).

Um dos inconvenientes da função `gettimeofday` é que o valor obtido em uma chamada é uma medida absoluta do tempo. Como um processo Unix está sujeito a interrupções de sua execução (dado o compartilhamento da máquina por outros processos e agentes), o tempo parado também será contado, eventualmente. Ademais, máquinas modernas possuem normalmente mais de um processador. É portanto provável que processos que estejam parados, neste contexto, sejam mais rapidamente agendados. Desta forma, o tempo final observado será uma mistura de tempos de execução reais, somados aos tempos de parada e re-agendamento. Estes tempos não são normalmente significativos para longos processos, mas podem facilmente afetar a percepção de processos mais curtos. Para evitar que estes tipos de evento ocorram com frequência durante a medida de tempo, ajustou-se a prioridade de execução do processo Athena de forma a minimizar o tempo que o processo estaria parado ou em transição de processador.

A Figura 7.8 mostra 7 histogramas correspondentes às diversas fases do processa-

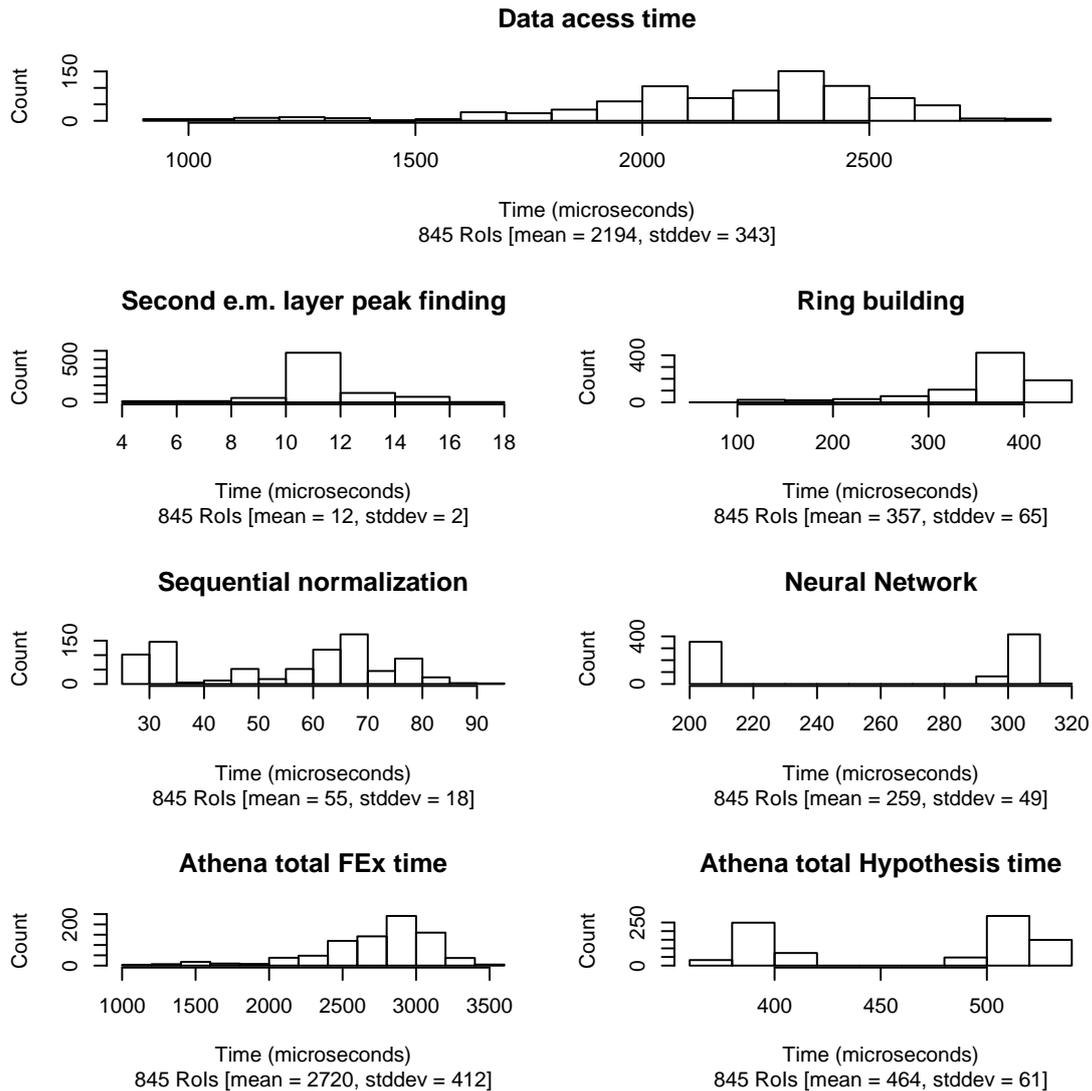


Figura 7.8: Tempos de execução do *NeuralRinger* funcionando de forma integrada ao ambiente Athena.

mento do *NeuralRinger* integrado ao ambiente Athena. Na parte superior da figura, observa-se o tempo de acesso aos dados. Este tempo de acesso é partilhado por todos os algoritmos de calorimetria, e apresenta neste caso uma média de 2,2 milissegundos por RoI. No ambiente Athena, este valor inclui o tempo de formatação dos dados no padrão dos detetores do ATLAS. Os quatro histogramas que seguem representam histogramas equivalentes aos da Figura 7.4, onde é possível observar os tempos individuais de cada uma das fases do processo de extração de características e classificação neural. Os dois histogramas no final da figura mostram os tempos de processamento totais, por algoritmo.



Removendo-se a contribuição do tempo de acesso aos dados, o processo de extração leva em média  $2720 - 2194 = 526$  microssegundos. O tempo médio de classificação baseado em uma rede neural leva 259 microssegundos. Desta forma, o tempo total, comparável àquele da Figura 7.1, é de  $526 + 259 = 785$  microssegundos. Este valor representa um acréscimo de cerca de 56% do tempo total calculado anteriormente. Uma parte do tempo adicional é gasto na criação de estruturas de dados que sejam transportáveis entre os algoritmos, passo necessário no ambiente Athena e que foi suprimido na execução desacoplada. Uma outra parte advém da utilização da infraestrutura de acesso a dados e relatório de erros disponível no ambiente Athena, que é mais lenta que no ambiente desacoplado.

O tempo total de processamento do algoritmo integrado é de cerca de 3,2 milissegundos. Comparando-se os tempos relativos de processamento para diversas configurações de anéis e detetores neurais das Figuras 7.1, 7.2 e 7.3 observa-se uma relação aproximadamente constante de cerca de 10 a 15% de diferença entre as plataformas AMD Opteron-275 e Intel Xeon, dependendo do número de anéis utilizados. Escalando o valor de tempo observado considerando-se a diferença máxima de 15%, chega-se a um valor de  $\approx 3,8$  milissegundos. Este valor é comparável ao tempo total de processamento do T2Calo, na Tabela 7.1, de 3,64 milissegundos. Portanto, com este exercício, demonstra-se a viabilidade deste sistema de detecção para operação *online*.

A Figura 7.9 ilustra a dependência dos tempos parciais de processamento com a localização da região de interesse, por  $\eta$ . A Figura 7.10 mostra a relação com a variável  $\phi$ . Naturalmente, o processo é tendencioso com relação à variável  $\eta$ , devido às variações da granularidade e independente da localização em  $\phi$ .

### 7.1.2 *NeuralRinger* e o sistema de aquisição

O último passo da integração é a demonstração da viabilidade do sistema de filtragem em um protótipo do sistema de aquisição, junto a outros algoritmos de filtragem. Neste caso, inicialmente validamos o funcionamento do algoritmo no contexto de uma configuração mais complexa baseada na aplicação AthenaMT [39], que simula, ainda que de modo *offline*, o ambiente de aquisição tal qual será encontrado pelos algoritmos do HLT *online*. Este é um passo simples, dado que se reutiliza

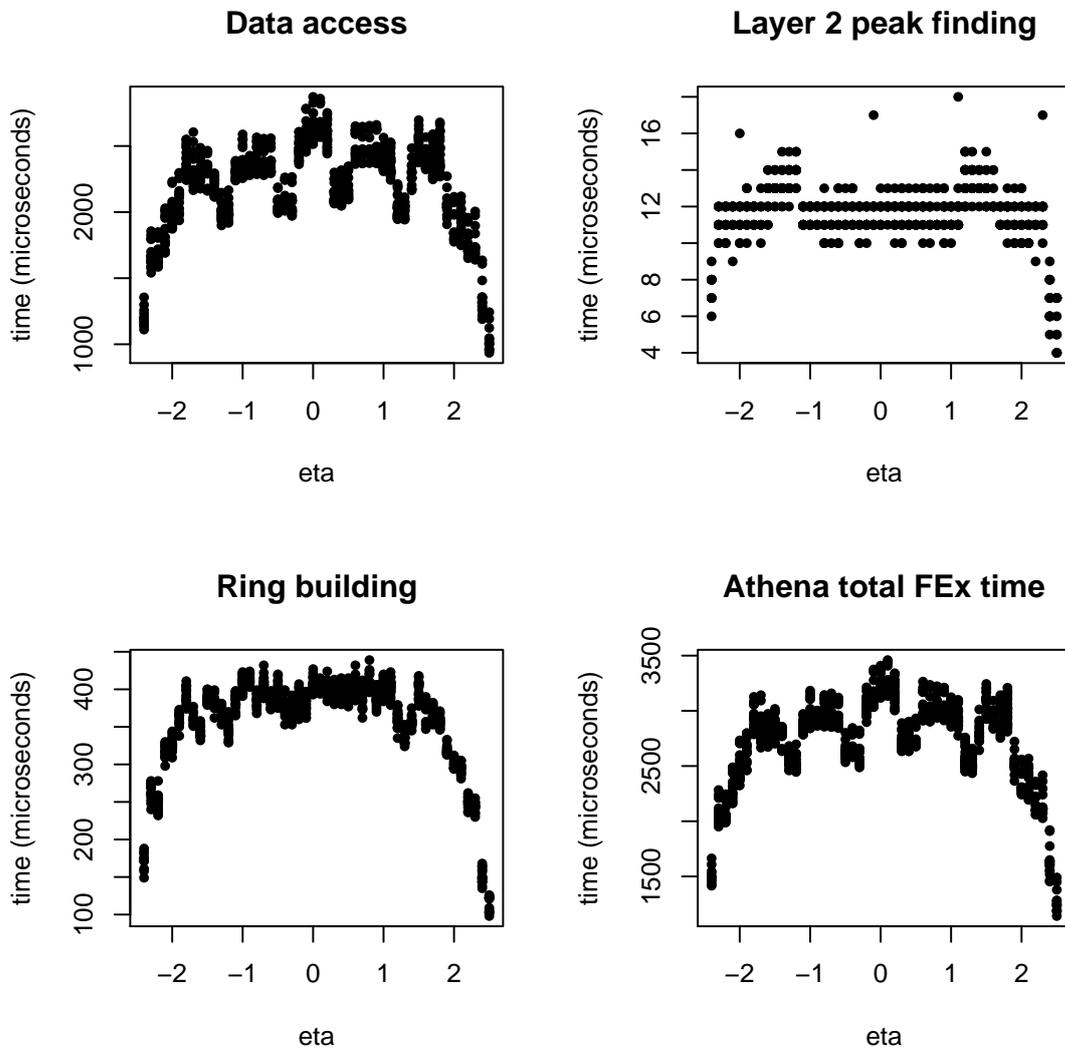


Figura 7.9: Relação entre o posicionamento da RoI em  $\eta$  e os tempos de processamento do *NeuralRinger* em diversas das fases da extração de características no ambiente Athena.

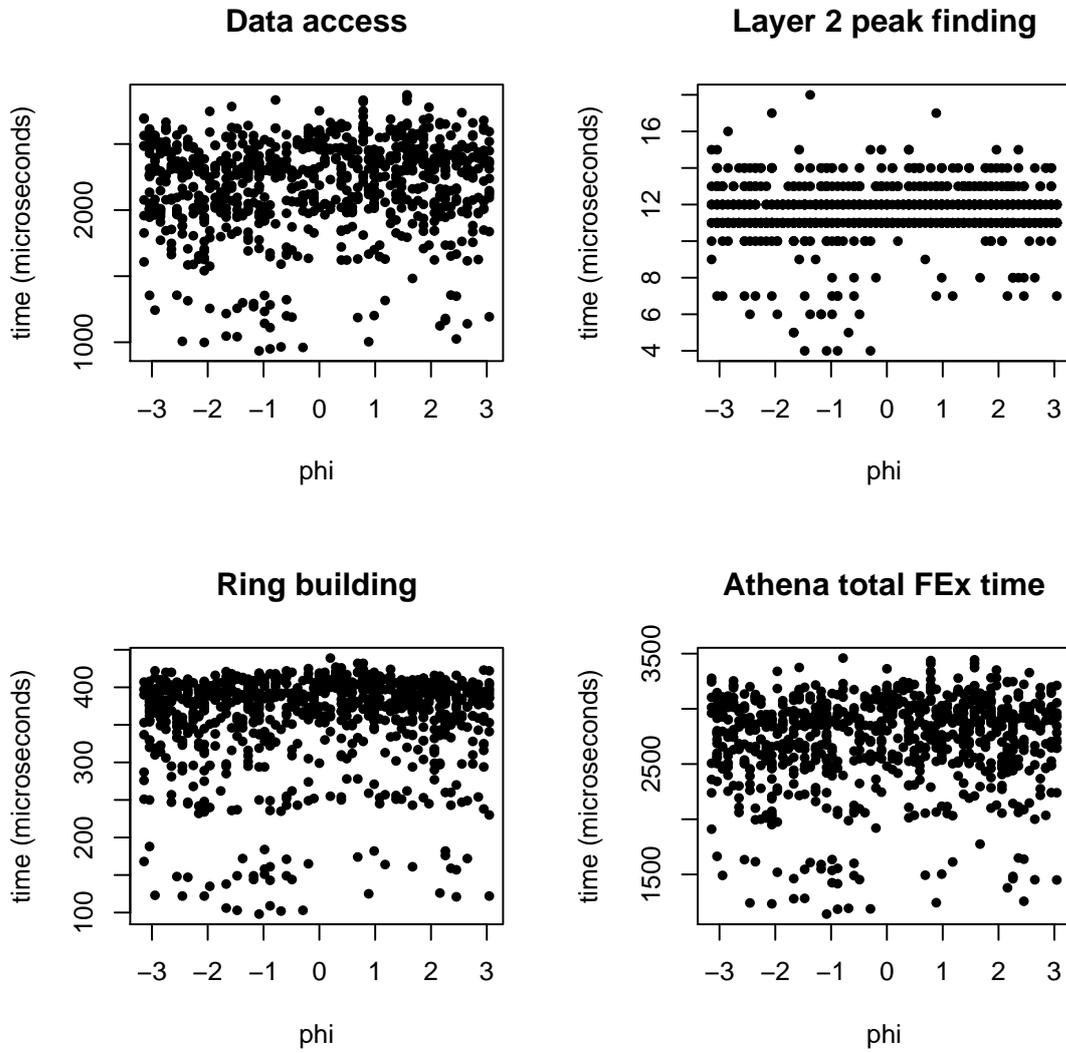


Figura 7.10: Relação entre o posicionamento da RoI em  $\phi$  e os tempos de processamento do *NeuralRinger* em diversas das fases da extração de características no ambiente Athena.

o mesmo conjunto de bibliotecas usadas para o ambiente Athena, sem necessitar de recompilação. A vantagem é que o sistema já encontra-se depurado e a única verificação a ser feita é se a utilização dos novos algoritmos não introduz erros para os demais.

Executou-se a aplicação AthenaMT incluindo uma configuração para a detecção de objetos tipo e.m. (T2Calo e algoritmos de detecção de traços) junto ao sistema baseado no anelamento e detecção neural. Observou-se o consumo de memória enquanto executava-se a aplicação para que possíveis vazamentos possam ser rastreados. Não foram observados problemas em nenhum aspecto.

O passo final é a configuração de uma bancada de testes para executar os algoritmos dentro do ambiente do Sistema de Filtragem do ATLAS. Para tal, utilizaram-se 7 máquinas conectadas via *gigabit* ethernet e uma máquina adicional para controlar o sistema remotamente. A Figura 7.11 esquematiza o sistema utilizado para o teste e o mapeamento das diversas aplicações nas máquinas. Esta bancada é, naturalmente, uma versão bastante reduzida do sistema que estará disponível no experimento. Para emular as condições de funcionamento do ATLAS, carregam-se dados de elétrons simulados nos nós nomeados ROS-\* e L2SV-\*. Desta forma, para as L2PU's, as condições de operação são idênticas àquelas do sistema final. A bancada consiste de 8 L2PU's, 8 Emuladores do Sistema de Leitura do detetor (ROS) e 1 L2SV. Estes sistemas estão conectados através de uma chave *Gigabit ethernet* e conectados ao sistema controlador através da rede do CERN.

O sistema de filtragem e aquisição do ATLAS provê um conjunto de ferramentas que podem ser utilizadas para monitorar a execução do sistema. A Figura 7.12 mostra uma captura de tela na máquina controladora, mostrando algumas destas ferramentas em atuação durante o teste. A luz verde na maior das janelas indica que o sistema está operante e nenhum erro foi detetado. O número de eventos processado, no total, pelas oito unidades de processamento é de aproximadamente meio milhão. O valor está marcado em vermelho na parte inferior esquerda da figura. A taxa de funcionamento do sistema parece estável em cerca de 245 Hz, como indica o gráfico na parte inferior direita da figura. Na parte superior direita, uma janela do sistema de monitoração *online* mostra um histograma, preenchido durante a execução do sistema e atualizado durante a operação com o tempo de

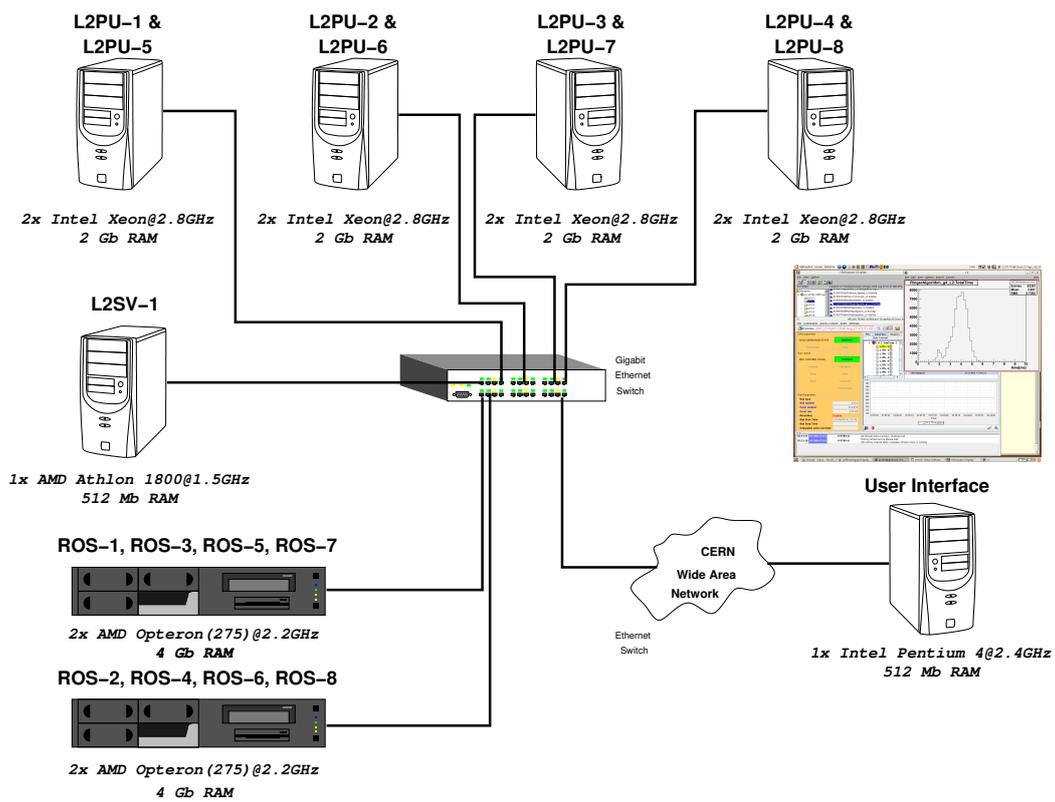


Figura 7.11: Esquema de bancada de testes para a verificação do funcionamento do NeuralRinger dentro do ambiente de aquisição de dados do ATLAS.

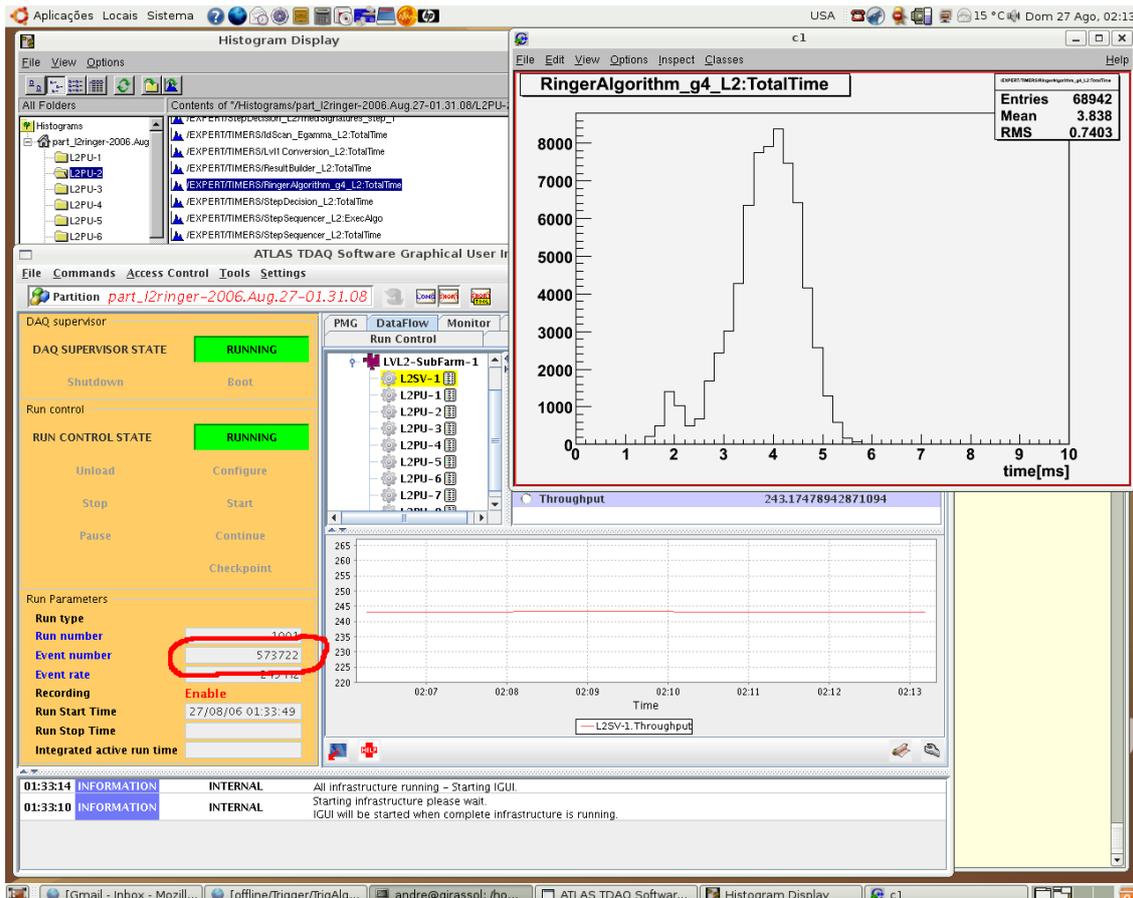


Figura 7.12: Captura de tela mostrando a operação do sistema de extração de características baseado em anéis rodando dentro de uma banca de testes do sistema de filtragem e aquisição do ATLAS.

processamento de um dos algoritmos de anelamento rodando na L2PU número 2. O tempo de processamento médio deste algoritmo nesta L2PU é de cerca de 3,8 ms. Este tempo inclui o tempo de acesso aos dados através da rede *Gigabit ethernet* levando-se em conta a configuração da bancada.

## 7.2 Implementação em um DSP

As bibliotecas de anelamento e processamento neural, desenvolvidas no contexto do *NeuralRinger* (veja Apêndice C), são extremamente portáteis e podem ser utilizadas em plataformas domésticas ou especialistas. Nesta seção, descrevemos uma implementação do *NeuralRinger*, baseada em DSP's (do inglês *Digital Signal Processors*), que pode ser utilizada como plataforma de desenvolvimento de sistemas de

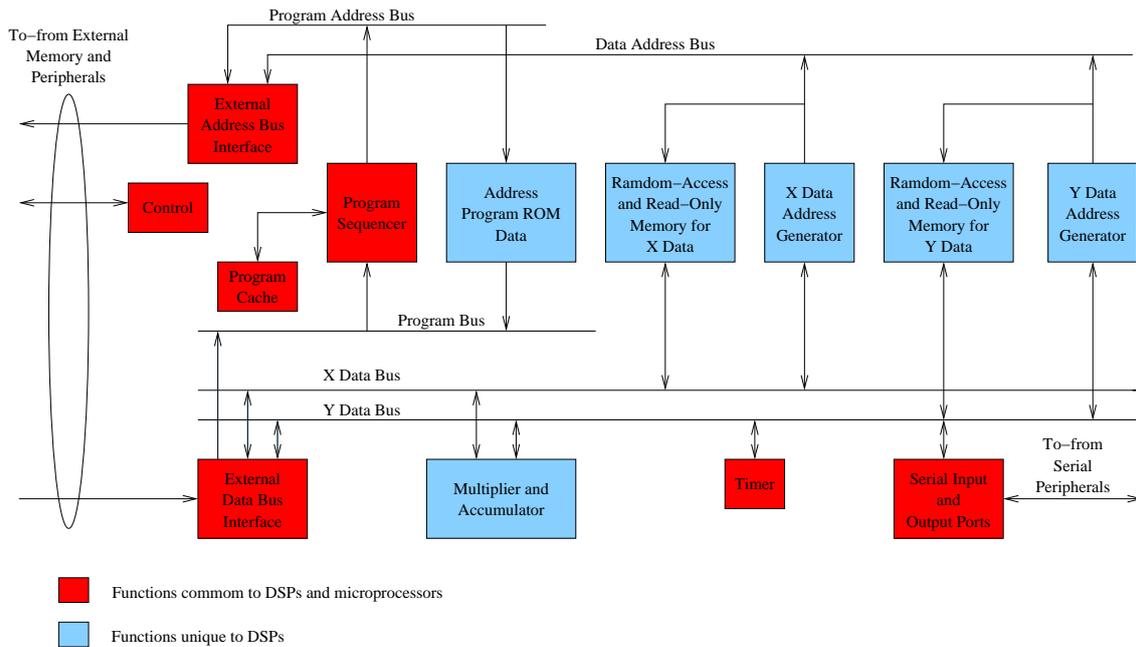


Figura 7.13: Visão geral da estrutura interna de um DSP.

anelamento e detetores neurais para operarem no Sistema de Filtragem do ATLAS.

DSP's são dispositivos que implementam, de forma bastante eficaz, operações vetoriais como somas e multiplicações seguidas de acumulação, *buffers* circulares sendo bastante eficientes em processos iterativos, como restrições de operação em tempo-real [84]. A Figura 7.13 mostra algumas das diferenças fundamentais entre processadores de uso genérico (abordados até aqui) e DSP's. Enquanto processadores genéricos possuem apenas uma unidade lógica aritmética, com apenas um barramento para transportar instruções e dados [85], DSP's contêm, adicionalmente, um sistema de multiplicação implementado em *hardware*, conectado diretamente à memória interna através de múltiplos barramentos. Ademais, DSP's são normalmente dotados de dispositivos para entrada e saída dedicados, adicionados para se obter acesso eficiente aos dados e de forma a manter as unidades computacionais focadas somente no processamento lógico e aritmético. Em específico para este estudo, o DSP utilizado foi um SHARC ADSP-21160 da *Analog Devices*, operando em ponto-flutuante, com 100 MHz de *clock*, 4 Mbits de memória interna e unidades aritméticas duplicadas para operação em modo SIMD (do inglês *Single-Instruction, Multiple Data*) [86].

Desta forma, a parte central de computação do anelamento e detecção neural

do *NeuralRinger* foi extraída, formando uma sub-estrutura de código diretamente focada no Sistema de Filtragem. Este código foi, sem maiores modificações, recompilado utilizando-se as ferramentas providas pelo fabricante do DSP e operado dentro desta plataforma.

Para a separação elétron/jato baseada em 100 anéis de deposição energética, esta plataforma atinge um tempo de processamento de  $4,692 \pm 1,108$  milissegundos por evento, que é compatível com os tempos de processamento obtidos para execução dentro da pilha de *software* do experimento, em máquinas hipoteticamente 30 vezes mais rápidas. A Figura 7.14 mostra o tempo cumulativo (integral do histograma de distribuição dos tempos de processamento) para este sistema, dividido em sub-fases do processamento. É possível notar que a geração dos anéis seja, computacionalmente, a fase mais importante do processamento, consumindo mais de 90% do tempo total. Isto se deve ao grande número de condições que rege a geração de anéis, nos afastando do propósito operacional desta plataforma: multiplicações e acumulações vetoriais, precedidas do pré-carregamento em, normalmente grandes *pipelines*. A fase de determinação do pico de deposição energética é a segunda fase que mais consome tempo de processamento, por motivos idênticos àqueles do anelamento. A normalização, juntamente com a detecção por métodos neurais são as partes mais rápidas. A rede neural leva apenas, em média,  $10,43 \pm 0,47$  microssegundos por evento processado.

O tempo total de processamento encontrado neste exercício pode ser reduzido considerando-se DSP's que operem com *clocks* de frequências mais altas. A Figura 7.15 mostra os tempos totais de processamento que poderiam ser atingidos com DSP's mais velozes da mesma família do sistema atualmente empregado (SHARC). O tempo de execução nesta figura é obtido, simplesmente, dividindo-se o número de instruções executadas no algoritmo (uma vez que DSP's desta família executam instruções em apenas um ciclo do *clock*) pela frequência do novo DSP. Como é possível ver, com um DSP de 400 MHz, o tempo total de processamento seria de apenas cerca de 1 ms.



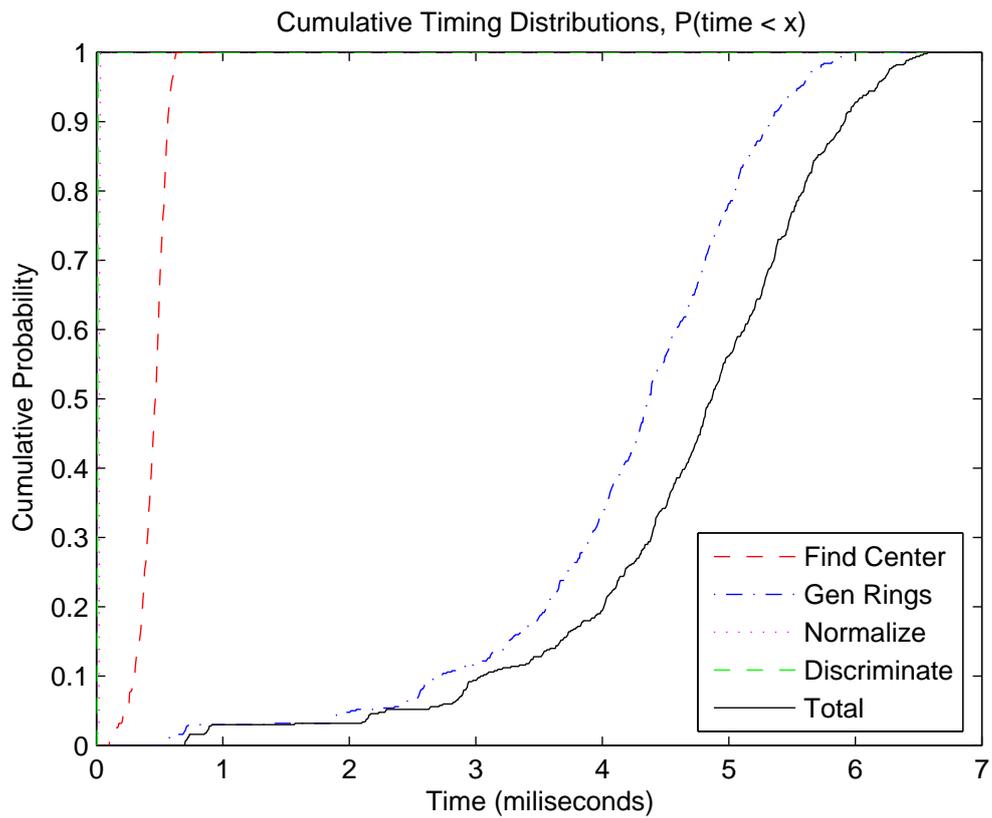


Figura 7.14: Função de distribuição cumulativa para os tempos de processamento, subdivididos por fase de processamento, para implementação do *NeuralRinger* em um DSP SHARC 21160 da *Analog Devices*.

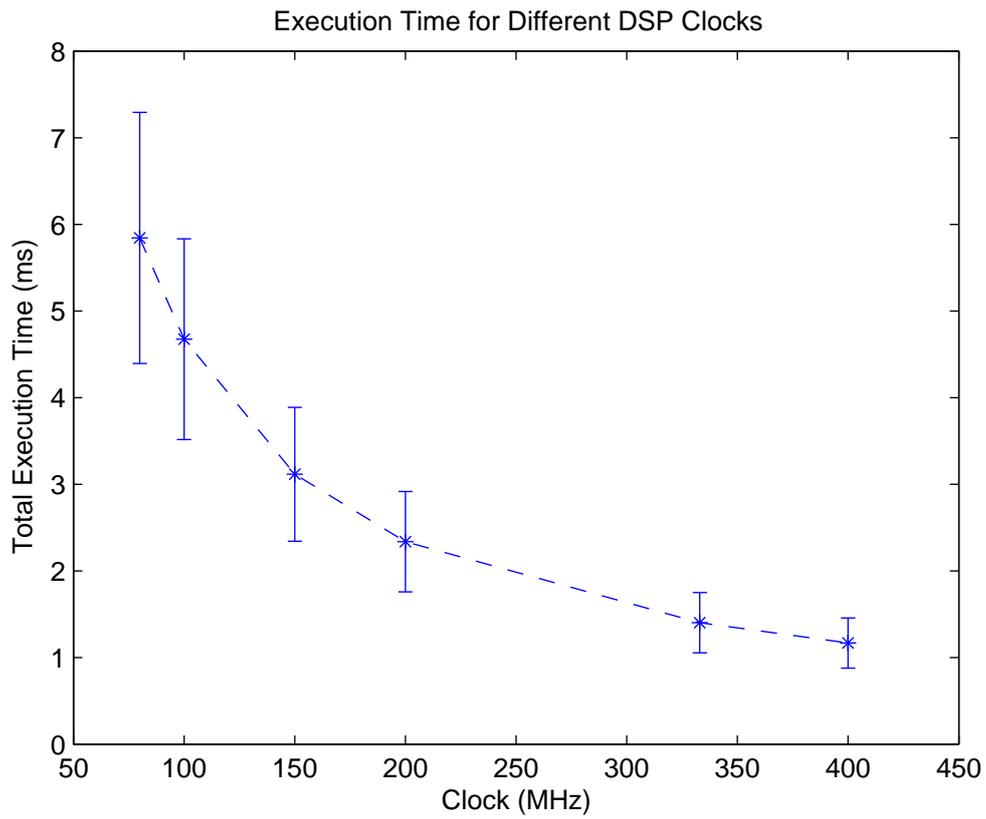


Figura 7.15: Tempo total de execução para o anelador e detetor neural para diferentes DSP's da família SHARC, da *Analog Devices*.

# Capítulo 8

## Conclusões

O Sistema de Filtragem do experimento ATLAS, no CERN, oferece um ambiente desafiador a seus sistemas de detecção. A taxa de eventos na entrada do LVL2 é bastante alta, na ordem de 100 mil por segundo. O sinal de interesse é um evento raro, apresentando-se escondido em uma massa de eventos ordinários também bastante alta, na ordem de 1 para  $10^9$ . O sistema é construído de forma altamente distribuída. Os dados utilizados na detecção são transferidos através de uma rede às unidades de processamento que, por sua vez, devem rapidamente (10 ms) responder com um sinal de rejeição, na maioria dos casos, ou aceitação.

A detecção elétron/jato representa um dos estudos mais importantes de decaimento da rara Física do bóson de Higgs. O processo de discriminação de elétrons é conduzido, inicialmente, por uma confirmação baseada em calorimetria. Os dados de cada região de interesse (RoI) destacada pelo Primeiro Nível de Filtragem são (lentamente) descarregados e tratados nas Unidades de Processamento do LVL2, deixando pouco tempo de processamento para a operação dos algoritmos de detecção propriamente ditos. Os dados de uma RoI estão organizados de forma altamente granular e segmentada, favorecendo a detecção através da análise do perfil de deposição energética da partícula ao longo de sua trajetória. A granularidade, segmento a segmento, varia ao longo dos eixos do detetor, podendo cruzar regiões totalmente desprovidas de células de detecção. Cada RoI, em média, é composta de mais de 1.000 células.

A Unidade de Processamento do LVL2 foi desenvolvida tendo por base a maximização da utilização dos recursos computacionais pelos algoritmos de detecção,

funcionando em múltiplas tarefas de processamento. Uma interface de “cola” foi criada entre o ambiente *online*, onde eventos fluem pela rede do Sistema de Filtragem e o ambiente *offline*, onde os algoritmos são desenvolvidos e testados. Essa unidade foi testada com relação a sua robustez e confiabilidade, tendo passado aos diversos critérios necessários a seu emprego no Sistema de Filtragem. Testes apontam que seja possível um modelo de desenvolvimento de algoritmos baseado nestes preceitos: criação *offline*, aplicação *online*. Neste ambiente, os diversos algoritmos que devem ser executados no Sistema de Filtragem são desenvolvidos de forma desacoplada e re-acoplados durante a execução combinada do sistema.

A execução dentro dos Altos Níveis de Filtragem é modelada sempre distinguindo-se a fase de Extração de Características (Fase 1) e Hipótese (Fase 2), que tem sua execução agendada, uma após a outra, complementando-se, mas guardando sua independência. O algoritmo T2Calo é provido atualmente como implementação de referência para a Fase 1. Esse sistema altamente depurado ao longo de muitos anos de desenvolvimento compacta a dimensão de entrada ao detetor elétron/jato em apenas 4 variáveis altamente discriminantes. As variáveis desse sistema não são linearmente separáveis, mas apresentam bom grau de classificação quando utilizam-se classificadores baseados em cortes uni-dimensionais sucessivos (sinal 91, 85%/10, 19% falso alarme ou 2,55 kHz) ou lineares (sinal 90, 77%/9, 79% falso alarme ou 2,45 kHz). A adição de mais variáveis à saída do T2Calo tende a melhorar a qualidade de detecção que pode ser obtida com esse método de compactação.

A compactação por anéis de deposição energética é uma implementação alternativa para o processo de Extração de Características (Fase 1) que aproveita-se do perfil de deposição energética das partículas que interagem com calorímetros. Esta técnica evoluiu com o experimento, suas modificações e a melhoria de sua simulação, mantendo-se simples e eficaz. Desenvolvemos uma implementação de referência para a extração baseada em anéis que é suficientemente rápida (podendo ser executada em até 125  $\mu$ s de forma desacoplada), resolvendo os problemas de variação de granularidade, dados faltantes e carência de sensores nas regiões de *crack* de forma igualmente eficiente e flexível. Ao passo que simplifica a Fase 1 do processamento das RoI's de candidatos a elétrons, essa técnica introduz um conjunto de variáveis (25 vezes) maior que aquele provido pelo T2Calo.

A técnica de compactação por anéis é desenvolvida baseando-se no pico de deposição energética obtido no LVL2 e, portanto, é independente da precisão de detecção do LVL1. O critério de normalização empregado varia com a energia do objeto analisado, apresentando-se muito eficaz em praticamente todo o espectro de energia considerado, como nos mostra a Figura 6.53.

A segunda fase do processamento de uma RoI é a detecção do objeto analisado, sendo realizada após a compactação. A implementação de referência do ATLAS é baseada em uma análise de cortes uni-dimensionais (EGammaHypo), de difícil (e longa) calibração, requerendo a intervenção especialista freqüentemente. Em contrapartida, se propõe a utilização de Redes Neurais Artificiais (RNA's) como forma de se melhorar a eficiência de detecção de elétrons no LVL2. O emprego de classificadores, que tentam encontrar correlações de ordens mais elevadas, é necessária nesse ambiente uma vez que:

1. O problema se apresente de uma forma linearmente inseparável;
2. O nível de correlação entre as 4 variáveis utilizadas para detecção é baixo.

Redes neurais podem ser empregadas tanto acopladas às saídas do T2Calo quanto às somas em anéis proposta, permitindo maiores níveis de eficiência na detecção de elétrons em ambos os casos. Quando associadas à compactação provida pelo T2Calo aos dados de uma RoI, RNA's podem melhorar o desempenho de classificação comparado a sistemas lineares de classificação, atingindo 92,38% de detecção contra 9,05% de falso-alarme (2,26 kHz). Considerando-se as demais saídas disponíveis neste algoritmo de compactação, é possível atingir níveis ainda mais altos: 94,79% de detecção contra apenas 7,93% de falso-alarme (1,98 kHz). Um estudo da ortogonalidade deste conjunto de 14 variáveis (Análise de Componentes Principais) revela que não há correlação direta que seja importante ao sistema de detecção.

Quando associadas ao compactador utilizando anéis de deposição energética, RNA's conseguem melhorar a relação sinal-ruído substancialmente, podendo atingir uma eficiência de classificação de 96,89% contra apenas 2,93% de falso-alarme (0,73 kHz). Comparativamente à dupla T2Calo/EGammaHypo, estes números representam um decréscimo de 3 vezes na taxa de jatos que seria aprovada ao demais níveis de filtragem. Um conjunto de análises apoia os resultados de eficiência, de-

monstrando que a utilização desta versão desenvolvida da soma em anéis à classificadores neurais, está livre de tendências com relação à localização do objeto ou seu valor energético, apresentando-se igualmente eficaz em uma multitude de intervalos.

Uma Análise de Relevância baseada no poder de classificação dos detetores (produto SP), ao invés do desvio no EMQ, é também conduzida. Esta análise pode ser utilizada como uma ferramenta de poda, para gerar classificadores mais compactos ou como um simples teste de robustez para sistemas de detecção. De acordo com uma comparação a sua análoga baseada no EMQ, apresentou-se mais poderosa na escolha de variáveis que, pela poda, minimizam o impacto na classificação.

Em alguns casos, esta análise também indica elementos que estejam prejudicando o poder classificatório de um detetor. Resultados (marginalmente) melhores na classificação, na ordem de alguns pontos percentuais, puderam ser obtidos baseando-se em podas conduzidas através da relevância de discriminação.

Finalmente, demonstrou-se que a metodologia proposta para o Sistema de Filtragem (desenvolvimento *offline*, aplicação *online*) funciona adequadamente, caso o desenvolvedor respeite critérios como tempo de execução e uso de recursos computacionais. Pacotes que implementam as Fase 1 e 2 de detecção elétron/jato, baseada em anéis e RNA's foram migrados para dentro do ambiente Athena. Enquanto, de forma desacoplada, é possível executar a detecção em tempos que variam de 100 a 500 microssegundos, dependendo da configuração dos anéis e das plataformas de execução, quando incorporados ao pesado ambiente de desenvolvimento do ATLAS, as mesmas instruções são executadas em cerca de 4 milissegundos.

Estes tempos de processamento são compatíveis com outros sistemas atualmente empregados no LVL2, inclusive com a dupla T2Calo/EGammaHypo, que consegue processar RoIs em um tempo equivalente, aproximadamente 3,6 milissegundos. A implementação baseada na infraestrutura do Sistema de Filtragem pode ser executada tanto *offline* quanto *online*, totalmente conectada ao sistema de leitura do detetor e de controle do experimento.

Uma implementação do sistema de anelamento e detecção neural, baseada em DSP's SHARC da Analog Devices, é derivada das mesmas bibliotecas como plataforma de desenvolvimento adicional. Esta implementação indica a portabilidade do algoritmo, que pode ser executado, igualmente, em cerca de 4 milissegundos den-

tro desta plataforma. Nesta implementação observa-se que a fase do processamento que mais leva tempo é a fase de extração dos anéis. Isto acontece pois este processamento é altamente condicional e, desta forma, não é possível tirar proveito da capacidade computacional (vetorial) desta plataforma. O tempo de processamento poderia ser melhorado utilizando-se DSP's mais rápidos, como indica a Figura 7.15, para se obter valores totais próximos a 1 milissegundo.

## 8.1 Linhas de Pesquisa

A Análise de Componentes Principais (ACP), aplicada ao problema da separação elétron/jato através dos dados de uma RoI ou através da informação de anéis, foi abordada em uma tese de doutorado [87], com bons resultados. Este sistema propõe um compactador (Fase 1) baseado na correlação entre os valores de energia de cada célula de uma RoI ou, ainda, entre seus anéis de deposição energética. Resultados indicam que a utilização de uma grande quantidade de sensores dificulta a análise estatística devido ao grande número de graus de liberdade na formulação do compactador. A ACP, quando extraída diretamente dos anéis de deposição energética, mostra-se mais eficiente e de simples aplicação.

A Análise de Componentes Independentes (ACI) representa uma extensão da Análise de Componentes Principais [88]. Desta vez, no lugar de se procurar uma transformação linear que descorrelacione os dados no novo espaço, deseja-se computar a transformação linear na qual as variáveis do espaço transformado sejam independentes. Esta é uma premissa muito mais forte que a simples descorrelação.

Uma das pré-condições para a ACI é de que o ruído introduzido pelo sistema no sinal de análise seja branco e gaussiano. Esta é, no entanto, uma hipótese bastante razoável em experimentos de Física de Altas Energias [11, 12]. A ACI pode ser conduzida utilizando-se diversos métodos que deverão ser avaliados com relação a sua velocidade e qualidade. Espera-se que esta técnica possa relevar informações no padrão de interação das partículas com os calorímetros do ATLAS, que estão mascarados nos anéis de deposição energética, aumentando, assim, a qualidade da discriminação conduzida no LVL2.

O conjunto de componentes extraídas não poderá ser organizado, *a priori*, como

acontece no caso da Análise de Componentes Principais, já que não há uma correlação evidente entre os coeficientes da transformação e sua importância. Neste caso, outros métodos, como o método da relevância de discriminação introduzido neste trabalho, poderão ser explorados para que se atinja a compactação dos dados de entrada.

Na mesma linha de pesquisa, a utilização de uma ACP não-linear pode também revelar correlações não-lineares que não puderam ser observadas neste trabalho ou em [87].

Métodos de treinamento mais sofisticados podem ser também introduzidos no sistema atual de detecção (Fase 2), aumentando a qualidade do treinamento, tempo de convergência e a robustez do sistema de discriminação final. Em especial, técnicas mais modernas, como redes neurais quânticas [89] ou menos exploradas neste contexto, como máquinas de vetor de suporte, [75] podem ser pensadas como alternativas interessantes, oferecendo melhor robustez ao ruído e a casos de empilhamento (*pile-up*).

A compactação baseada na soma em anéis de deposição energética pode ser explorada de forma a ser maximizada. Novas combinações devem ser exploradas tendo em vista a quantidade de dados que devam ser lidas do sistema de leitura do detetor e o fator de compactação propriamente dito. Especificamente, a técnica de normalização dos dados poderia ser ajustada tendo em vista os resultados da análise de relevância.

A detecção baseada em anéis pode também ser utilizada para análise e reconstrução *offline*. O sistema de detecção desenvolvido neste trabalho deve ser readaptado para trabalhar neste ambiente. Comparativos diretos com as técnicas atuais de reconstrução poderão ser feitos. Métodos de análise estocástica são bastante difundidos e utilizados atualmente nessa sub-área da Física de Altas Energias.

Finalmente, seria possível expandir a utilização da técnica de anelamento para outras assinaturas interessantes, filtráveis por uma análise calorimétrica, tais como  $H \rightarrow \gamma + \gamma$  ou, ainda, na separação  $\gamma/\pi^0$ , tanto *online* como *offline*. Para aumentar a velocidade do processamento, seria possível segmentá-lo, executando a filtragem baseada em camadas mais discriminantes (tais como a segunda e.m.) primeiramente, e seguindo-se para as demais. Desta forma, somente RoI's de mais difícil detecção



sobreviveriam até o final do processamento.

# Referências Bibliográficas

- [1] A. Anjos. *Sistema de classificação baseado em uma máquina com sistema distribuído*. Projeto final, Departamento de Eletrônica/UFRJ, setembro 1997.
- [2] A. Anjos. *Sistema neuronal rápido de decisão baseado em calorimetria de altas energias*. Dissertação de mestrado, COPPE/UFRJ, abril 2001.
- [3] internet: <http://particleadventure.org>. Acessado em 29 de junho de 2004.
- [4] Halliday, Resnick, Walker. *Fundamentos da Física*. Livros Técnicos e Científicos Editora, 4a. edição, 1995.
- [5] The DZERO experiment. internet: <http://www-d0.fnal.gov/index.html>. Acessado em 29 de junho de 2004.
- [6] The compact muon solenoid. internet: <http://cmsinfo.cern.ch/Welcome.html>. Acessado em 29 de junho de 2004.
- [7] The ATLAS Collaboration. ATLAS: Technical propose for a general purpose pp experiment at the large hadron collider at CERN. Relatório técnico, CERN, 1994.
- [8] C.N. Booth. Aspects of modern physics. internet: <http://www.shef.ac.uk/-physics/teaching/phy311>. Acessado em 29 de junho de 2004.
- [9] R.K. Bock, A.Vascileu. *The Particle Detector BriefBook*. Springer, 1999.
- [10] R. Wigmans. Advances in hadron calorimetry. *Annual Review of Nuclear and Particle Science*, 41, páginas 133–185, dezembro 1991.
- [11] Glenn Knoll. *Radiation Detection and Measurement*. John Wiley and Sons, 2a. edição, 1989.

- [12] Willian Leo. *Techniques for Nuclear and Particle Physics Experiments - A How-to Approach*. Springer-Verlag, 1987.
- [13] V. Brigljevic, G. Bruno, E. Cano *et al.* The CMS High Level Trigger. Em *IEEE Nuclear Science Symposium*, outubro 2003.
- [14] Y. Hu. The Central Track Trigger System of D0 Experiment. Em *IEEE Nuclear Science Symposium*, outubro 2003.
- [15] B. Angstadt, G. Brooijmans, D. Chapin *et al.* Ethernet-based data acquisition system for the d0 experiment at fermilab. Em *IEEE Nuclear Science Symposium*, outubro 2003.
- [16] ATLAS HLT/DAQ/DCS Group. ATLAS high-level trigger, data acquisition and controls. Relatório técnico ATLAS TDR 016, CERN, outubro 2003.
- [17] Gordon Watts. Review of triggering. Em *IEEE Nuclear Science Symposium*, outubro 2003.
- [18] The european centre for particle physics. internet: <http://www.cern.ch>. Acessado em 29 de junho de 2004.
- [19] internet: <http://lhc.web.cern.ch/lhc>. Acessado em 07 de janeiro de 2007.
- [20] A Large Ion Collider Experiment at CERN LHC. internet: <http://alice.web.cern.ch/Alice/AliceNew>. Acessado em 29 de junho de 2004.
- [21] The large hadron collider beauty experiment. internet: <http://lhcb.web.cern.ch/lhcb>. Acessado em 29 de junho de 2004.
- [22] A Toroidal LHC ApparatuS. internet: <http://atlas.web.cern.ch/Atlas>. Acessado em 29 de junho de 2004.
- [23] ATLAS Inner Detector Group. Inner detector. Relatório técnico CERN/LHCC/97-16, CERN/LHCC/97-17, CERN, abril 1997.
- [24] ATLAS Collaboration. ATLAS: Calorimeter performance. Relatório técnico CERN/LHCC 96-40, CERN, janeiro 1997.

- [25] ATLAS LARG Unit. The liquid argon calorimeter. Relatório técnico CERN/LHCC/96-41, CERN, 1996.
- [26] ATLAS Tile Calorimeter Collaboration. Tile calorimeter. Relatório técnico, CERN, dezembro 1996.
- [27] J.M. Seixas, A.S. Cerqueira, L.P. Calôba. A high-performance active adder. *IEEE International Conference on Electronics, Circuits and Systems*, páginas 377–380, 1998. Lisboa, Portugal.
- [28] ATLAS Muon Spectrometer Group. Muon spectrometer. Relatório técnico CERN/LHCC/97-22, CERN, junho 1997.
- [29] ATLAS Level-1 Trigger Group. The first level trigger. Relatório técnico CERN/LHCC/98-14, CERN, 1998.
- [30] C. Haerberli, A. Anjos *et al.* The DataFlow System of the ATLAS Trigger and DAQ. Em *Computing for High Energy Physics*, La Jolla, Califórnia, EUA, março 2003.
- [31] H.P. Beck, A. Anjos *et al.* ATLAS-TDAQ datacollection software. Em *Real-Time 2003*, La Jolla, Califórnia, EUA, março 2003.
- [32] H.P. Beck, A. Anjos *et al.* The base-line DataFlow system of the ATLAS Trigger and DAQ. *IEEE Transactions on Nuclear Science*, 51(3), junho 2004.
- [33] S. Gadomski, A. Anjos. Experience with multi-threaded C++ applications in the ATLAS dataflow software. Em *Computing for High Energy Physics*, La Jolla, Califórnia, EUA, março 2003.
- [34] The GNU C Compiler Suite. internet: <http://www.gnu.org/software/gcc>. Acessado em 29 de junho de 2004.
- [35] The GNU C++ Standard Library. internet: <http://gcc.gnu.org/onlinedocs/libstdc++/documentation.html>. Acessado em 29 de junho de 2004.
- [36] A. Anjos. The event format user requirements document. CERN/ATLAS Note ATL-DQ-EN-0005, CERN, maio 2003. internet: <https://edms.cern.ch/file/383859/1/eformat-urd-1-0.pdf>, acessado em 29 de junho de 2004.

- [37] A. Anjos. The event format analysis and design. CERN/ATLAS Note ATL-DQ-EN-0006, CERN, maio 2003. internet: <https://edms.cern.ch/file/384821/-1/eformat-ooad-0-1.pdf>, acessado em 29 de junho de 2004.
- [38] A. Anjos *et al.* The Second Level Trigger of the ATLAS Experiment at CERN's LHC. *IEEE Transaction on Nuclear Science*, 51(3), junho 2004.
- [39] W. Wiedenmann, A. Anjos *et al.* Studies for a common Event Selection Software: from LVL2 to Offline reconstruction. *IEEE Transaction on Nuclear Science*, 51(3), junho 2004. Este trabalho também foi apresentado no Simpósio de Ciências Nucleares 2003 em outubro de 2003.
- [40] J.T.M. Baines, S. González, R.K. Mommsen, A. Radu, T.G. Shears, S. Sivoklokov, M. Wielers. First study of the LVL2-EF boundary in the high-pT e/gamma high-level trigger. ATLAS Note ATL-DAQ-2000-045, CERN, 2000.
- [41] internet: <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/OO/architecture/index.html>. Acessado em 29 de junho de 2004.
- [42] Athena: The ATLAS common framework. internet: <https://twiki.cern.ch/twiki/bin/view/Atlas/WorkBookAthenaFramework>. Acessado em 20 de novembro de 2006.
- [43] A. Bogaerts, A. Anjos, S. González, W. Wiedenmann. Guidelines for offline preparation and testing of LVL2 code. Disponível em: <http://cern.ch/rabello/work/cern/hlt/l2-coding/index.html>. Acessado em 20 de novembro de 2006.
- [44] R. Wigmans. *Calorimetry: Energy Measurement in Particle Physics*. Oxford, 2000.
- [45] M.A. Abolins. Triggering and calorimetry. *Calorimetry in High Energy Physics*, páginas 145–159, 2000.
- [46] T. Monteiro. Recent developments in the CMS Calorimeter Trigger. *Calorimetry in High Energy Physics*, páginas 605–610, 1999.
- [47] D. Froidevaux, M. Nesi, P. Nevski. Electron Identification at LHC. *Calorimetry in High Energy Physics*, páginas 343–356, 1992.

- [48] S. Gonzalez, T. Hansl-Kozanecka, M. Wielers. Selection of high-pT electromagnetic clusters by the level-2 trigger of ATLAS. ATLAS-Trigger DAQ Note 99, CERN, novembro 1999.
- [49] N. Ellis. Level-1 and Level-2 Triggering at LHC. *Calorimetry in High Energy Physics*, páginas 471–487, 1992.
- [50] Jan Guida. The D0 Calorimeter Trigger. *Calorimetry in High Energy Physics*, páginas 465–470, 1992.
- [51] M. Palutan. The KLOE calorimeter trigger. *Calorimetry in High Energy Physics*, páginas 169–176, 2000.
- [52] D. Damazio. Página wiki do t2calo. internet: <https://uimon.cern.ch/twiki/bin/view/Atlas/TrigT2Calo>. Acessado em 07 de janeiro de 2007.
- [53] M. Wielers. Samples for e/g trigger studies. internet: <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/DC/DC1/DC1%5F1/mediumStat/eg/-eg%5Ftrigger.html>. Acessado em 07 de janeiro de 2007.
- [54] J.F. Laporte. Validation samples. internet: <http://atlas.web.cern.ch/Atlas/GROUPS/SOFTWARE/DC/DC1/DC1%5F1/validation/-validation%5Fsamples.html>. Acessado em 07 de janeiro de 2007.
- [55] H.L. Van Trees. *Detection, Estimation and Modulation*. John Wiley & Sons, 1966.
- [56] B. Denby. A bat’s ear view of neural nets in physics. *Nuclear Instruments and Methods in Physics Research A*, 389, páginas 8–11, 1997.
- [57] B. Denby. Trends in physics data analysis algorithms. *Nuclear Instruments and Methods in Physics Research A*, 534, páginas 343–347, 2004.
- [58] A. Wilk. Analysis of the electron/pion separation capability with real size ALICE TRD prototypes using a neural network algorithm. *Nuclear Instruments and Methods in Physics Research A*, 563, páginas 314–316, 2006.

- [59] J.C. Prévotet, B. Denby, P. Garda, B. Granado, C. Kiesling. Moving NN Triggers to Level-1 at LHC Rates. *Nuclear Instruments and Methods in Physics Research A*, 502, páginas 511–512, 2003.
- [60] J.K. Köhne *et al.* Realization of a second level neural network trigger for the H1 experiment at HERA. *Nuclear Instruments and Methods in Physics Research A*, 389, páginas 128–133, 1997.
- [61] J. Varela, N. Leonardo. Study of a neural approach for lower level e/gamma calorimeter trigger in CMS. Relatório técnico, CERN, 1998.
- [62] C. Peterson. Track finding with neural networks. *Nuclear Instruments and Methods in Physics Research A*, 279, páginas 537–545, 1988.
- [63] J. Zimmermann, C. Kiesling. Statistical learning methods in high-energy and astrophysics analysis. *Nuclear Instruments and Methods in Physics Research A*, 534, páginas 204–210, 2004.
- [64] P.C. Bhat. Search for the top quark at D0 using multivariate methods. *AIP*, páginas 308–319, julho 1996.
- [65] K.H. Becks, P. Buschmann, J. DREES, U. Müller, H. Wahlen. Selection of w-pair-production in DELPHI with feed-forward neural networks. *VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, páginas 80–82, 2001.
- [66] S. Tentindo-Repond, P.C. Bhat, H.B. Prosper. Neural networks for higgs physics. *VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, páginas 98–100, 2001.
- [67] L. Dudko. Use of neural networks in a search for single top quark production at D0. *VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, 2001.
- [68] C. S. Lindsey, B. Denby, H. Haggerty, K. Johns. Real time track finding in a drift chamber with a VLSI neural network. *Nuclear Instruments and Methods in Physics Research A*, 317, páginas 346–356, 1992.

- [69] D. Chakraborty. Particle identification at D0 using neural networks. *VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, páginas 92–94, 2001.
- [70] M.R. Vassali, J.M. Seixas. Principal component analysis for neural electron/jet discrimination on highly segmented calorimeters. *VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, outubro 2000.
- [71] P.V.M. da Silva, J.M. Seixas, J. Seixas. A hybrid training method for neural energy estimation in calorimetry. *VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, páginas 86–88, 2001.
- [72] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics (antigo Annals of Eugenics)*, 7, páginas 179–188, 1936.
- [73] B. Stroustrup. *The C++ Programming Language*. Addison-Wesley, 3a. edição, 1997.
- [74] G. Booch. *Object Oriented Analysis and Design with Applications*. Addison-Wesley, 2a. edição, 1993.
- [75] Simon Haykin. *Neural Networks: A comprehensive foundation*. Prentice Hall, 2a. edição, 1999.
- [76] M. Riedmiller, H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *III IEEE International Conference on Neural Networks*, páginas 586–591, 1993.
- [77] internet: <http://www-ra.informatik.uni-tuebingen.de/SNNS>. Acessado no dia 1 de maio de 2006.
- [78] A. Gruber, J. Möck, J. Fröchtenicht, D. Goldner, C. Kiesling, T. Klober, H. Kolanoski, T. Krämerkämper, P. Ribarics. Performance of backpropagation networks in the second level trigger of the H1-experiment. *New Computing Techniques in Physics Research*, 1994.



- [79] A. Anjos, J. Seixas. Mapeamento em anéis para uma separação neuronal elétron-jato usando calorímetros multi-camadas e multi-segmentados. *XXI Encontro Nacional de Física de Partículas e Campos*, outubro 2000.
- [80] D. Damazio. Tile calorimeter fast bytestream conversion. LVL2 Calorimeter Egamma Code Review. internet:<http://indico.cern.ch/materialDisplay.py?confId=a062327>, acessado no dia 15 de agosto de 2006.
- [81] N.G. Unel *et al.* Studies with ATLAS Trigger and Data Acquisition pre-series setup. Em *Computing for High-Energy Physics*, Tata Institute of Fundamental Research, Mumbai, India, 2006.
- [82] internet: <http://linux.cern.ch>. Acessado no dia 18 de agosto de 2006.
- [83] The GNU C Library. internet: <http://www.gnu.org/software/libc/manual>. Acessado em 28 de agosto de 2004.
- [84] J.H. McClellan, R.W. Schafer, M.A. Yoder. *DSP First: A Multimedia Approach*. Prentice Hall, 1998.
- [85] J.G. Ackenhusen. *Real-Time Signal Processing*. Prentice Hall, 1999.
- [86] Analog Devices. *ADSP-21160: SHARC DSP Hardware Reference*, 2a. edição, maio 2002.
- [87] H.P. Lima Jr. *Monitoração Online e Processamento Distribuído para um Ambiente com Alta Taxa de Eventos e Fina Segmentação*. Tese de doutorado, COPPE/UFRJ, 2006.
- [88] Aapo Hyvärinen, Juha Karhunen, Erkki Oja. *Independent Component Analysis*. Wiley Interscience, 2001.
- [89] T. Menneer. *Quantum Artificial Neural Networks*. Tese de doutorado, University of Exeter, 1998.
- [90] R.O. Duda, P.E. Hart, D.H. Stork. *Pattern Classification*. Wiley Interscience, 2a. edição, 2000.

- [91] B. Widrow, S.D. Stearns. *Adaptive Signal Processing*. Prentice Hall, 1985.
- [92] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 1a. edição, 1996.
- [93] F.Rosenblatt. On the convergence of reinforcement procedures in simple perceptrons. Relatório técnico, Cornell Aeronautical Laboratory, 1960.
- [94] internet: <http://www.w3.org/XML>. Acessado no dia 12 de agosto de 2006.
- [95] internet: <http://xml.apache.org/xerces-c>. Acessado no dia 1 de maio de 2006.
- [96] internet: <http://xmlsoft.org>. Acessado no dia 1 de maio de 2006.
- [97] internet: <http://www.gnu.org/software/gsl>. Acessado no dia 12 de agosto de 2006.
- [98] internet: <http://www.graphviz.org>. Acessado no dia 12 de agosto de 2006.
- [99] A. Anjos *et al.* Deployment of the ATLAS High-Level Trigger. *IEEE Transactions on Nuclear Science*, 53, páginas 2144–2149, agosto 2006. Volume 4, Parte 2.
- [100] H. von der Schmitt, A. Anjos *et al.* A configuration system for the ATLAS trigger. *Journal of Instrumentation*, maio 2006.
- [101] A. Anjos, R. Torres, J.M. Seixas, B. Ferreira, T. Xavier. Neural triggering system operating on high resolution calorimetry information. *Nuclear Instruments and Methods in Physics Research*, 559, páginas 134–138, abril 2006.
- [102] D. Burckhart-Chromek, A. Anjos *et al.* Testing on a Large Scale: running the ATLAS Data Acquisition and High Level Trigger Software on 700 PC Nodes. *Computing In High Energy and Nuclear Physics*, fevereiro 2006.
- [103] A. Anjos, R. Torres, J.M. Seixas, B. Ferreira, T. Xavier. Discriminação Neural de Elétrons no Segundo Nível de Trigger do ATLAS. *XXVI - Encontro Nacional de Física de Partículas e Campos*, 2005.
- [104] A. Anjos, R. Torres, J.M. Seixas, B. Ferreira, T. Xavier, D. Damazio. Otimização do Sistema de Trigger do Segundo Nível do ATLAS Baseado em

- Calorimetria. *XXVI - Encontro Nacional de Física de Partículas e Campos*, 2005.
- [105] A. Gesualdi-Mello, A. Anjos *et al.* Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC. *14th IEEE - NPSS Real Time Conference 2005 Nuclear Plasma Sciences Society*, junho 2005.
- [106] A. Anjos *et al.* Configuration of the ATLAS trigger. *14th IEEE - NPSS Real Time Conference 2005 Nuclear Plasma Sciences Society*, junho 2005.
- [107] A. Anjos, J.M. Seixas. Os Filtros de Alto Nível do Experimento ATLAS. *XXV - Encontro Nacional de Física de Partículas e Campos*, 2004.
- [108] C. Schiavi, S. Armstrong, A. Anjos *et al.* Implementation and Performance of the High Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC. *Computing in High Energy Physics and Nuclear Physics*, setembro 2004.
- [109] S. Armstrong, A. Anjos *et al.* Algorithms for the ATLAS high-level trigger. *IEEE Transactions on Nuclear Science*, 51, páginas 367–374, junho 2004. Volume 3.
- [110] C. Haeberli, A. Anjos *et al.* ATLAS TDAQ data collection software. *IEEE Transactions on Nuclear Science*, 51, páginas 585–590, junho 2004. Volume 3.
- [111] P. Casado, A. Anjos *et al.* Architecture of the ATLAS online physics-selection software at LHC. Em *8th ICATPP Conference Proceedings*, Villa Erba, Como, Itália, outubro 2003.
- [112] A. Anjos, J. Seixas. Neural particle discrimination for triggering interesting physics channels with calorimetry data. *Nuclear Instruments And Methods In Physics Research A - Accelerators, Spectrometers, Detectors And Associated Equipment*, 502(2), páginas 713–715, 2003.
- [113] M. Grothe, A. Anjos *et al.* Architecture of the ATLAS high level trigger event selection software. Em *Computing for High-Energy Physics*, La Jolla, Califórnia, EUA, março 2003.

- [114] A. Anjos, J.M. Seixas. Classificação Elétron-Jato para o Segundo Nível de Filtragem do Experimento ATLAS Usando Redes Neurais Especialistas. *XXI - Encontro Nacional de Física de Partículas e Campos*, 2004.
- [115] A. Anjos, J.M. Seixas. Integrando Plataformas e Algoritmos para o Segundo Nível de Trigger do Experimento ATLAS. *XX - Encontro Nacional de Física de Partículas e Campos*, 1999.
- [116] J.M. Seixas, A. Anjos, A.C.H. Dantas, B.P. Kastrup, L.P. Calôba. Um Protótipo do Sistema de Validação de Nível 2 para as Condições do LHC. *XIX Encontro Nacional de Física de Partículas e Campos*, páginas 32–33, dezembro 1998.
- [117] J.M. Seixas, A. Anjos, L.P. Calôba, C.B. Prado, J.C.R. Aguiar, A.C.H. Dantas. Neural classifiers implemented in a transputer based parallel machine. *International Meeting on Vector and Parallel Processing (VECPAR)*, 1998.

# Apêndice A

## As Coordenadas do ATLAS

O sistema de coordenadas usado em experimentos com feixes não é, em muitos casos, o sistema polar. É um sistema adequado ao formato cilíndrico dos detectores dispostos ao redor do ponto de impacto, ou seja, um sistema que *acompanha* a direção dos feixes de partículas provenientes da colisão. As coordenadas empregadas são  $\eta$ ,  $\phi$  e  $z$  em contraposição às coordenadas cartesianas  $x$ ,  $y$  e  $z$ .  $\eta$  e  $\phi$  seguem uma transformação não-linear de  $x$ ,  $y$  e  $z$ :

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

$$\theta = \text{atan}\left(\frac{x}{z}\right) \quad (\text{A.2})$$

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (\text{A.3})$$

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

Os valores dados das variáveis  $\eta$  e  $\phi$  são apenas para referência do leitor. A variável  $\phi$ , como é possível ver no canto direito da parte superior da figura, possui uma região em que dois valores são possíveis: 0 e  $2\pi$ . Esta área é chamada de região

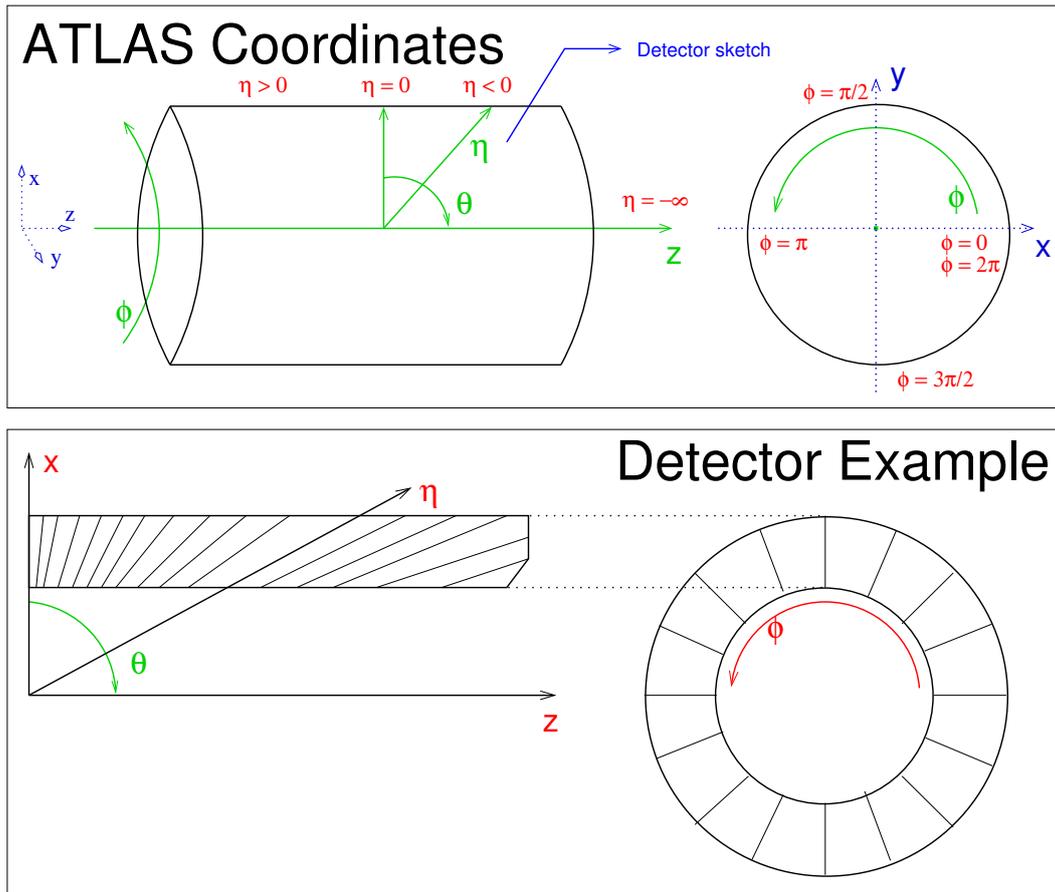


Figura A.1: O sistema de coordenadas do ATLAS.

*wrap-around*. Cálculos utilizando esta variável devem atentar para este fato. Os detetores são simétricos, com relação a  $\phi$ . A construção dos dispositivos é feita em gomos.

Nota-se que quando alcança o eixo  $z$ ,  $\eta = \infty$ , isto significa que objetos com valores grandes em  $\eta$  representam colisões onde as partículas do feixe apenas se desviaram. É comum utilizarem-se detetores com baixa resolução quando  $\eta > 3$ .

Na parte inferior da figura A.1, é possível ver um exemplo de como um detetor genérico é segmentado, acompanhando as coordenadas  $\eta$  e  $\phi$ , tanto para o barril, quanto para uma tampa.

# Apêndice B

## Introdução ao processamento neural

O treinamento de redes neurais teve seu início através do desenvolvimento de técnicas para o treinamento de sistemas mais simples, como detetores lineares. O discriminador de Fisher [72] ou a Análise de Discriminação Linear descreve um algoritmo para que se maximize a capacidade discriminante de um corte num plano com  $N$  dimensões, que separa duas classes de dados. O resultado do discriminante é ótimo no caso dos dados apresentarem uma distribuição gaussiana e as matrizes de covariância, para ambas as classes em separado, sejam idênticas<sup>1</sup>. Muitas vezes, na prática, os valores de média e a covariância não estão disponíveis e devem ser estimados, o que normalmente leva a falhas no cálculo do ponto ótimo de discriminação. Dentre as técnicas de estimação da média e covariância, pode-se destacar a estimação por máxima verossimilhança ou por máxima probabilidade *a posteriori* [90]. O discriminante de Fisher maximiza, através de uma transformação linear, a distância entre as classes que se deseja detetar, ao mesmo tempo diminuindo a distância de elementos intraclasse.

De posse dos valores de média e covariância das classes, o plano de separação seria definido da seguinte forma:

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_0 - \vec{\mu}_1) \tag{B.1}$$

---

<sup>1</sup>A Análise de Discriminação Quadrática, no entanto, demonstra que esta característica pode ser relaxada considerando-se que seja sempre possível projetar, através de uma transformação linear, o espaço de entrada em um outro espaço onde as matrizes de covariância sejam iguais e portanto recaindo no caso simples.

O operador  $\Sigma$  na Equação (B.1) representa a covariância ou *variância cruzada* das observações do universo de entrada e  $\mu$  as médias das duas classes de eventos que se deseja separar. Um problema que se segue é da inversibilidade de  $\Sigma$ , que dependerá de quão bem o conjunto de amostras representa as classes a serem discriminadas. Por exemplo, se  $\Sigma$  possui combinações lineares dos dados disponíveis, o posto desta matriz será inferior ao número de amostras e, portanto, a matriz não será inversível. Embora existam maneiras de superar o problema, é possível fazer uso de outras técnicas derivadas deste sistema primário para que se maximize a discriminação das classes de eventos dado um conjunto de amostras.

Por causa das dificuldades discutidas, muitas técnicas iterativas surgiram para que seja possível a definição de um plano ótimo de separação linear a partir de amostras de dados reais. Dentre elas, o algoritmo do Mínimo Médio Quadrático [91] (do inglês *Least Mean Square* ou LMS) é um dos mais utilizados. Este método pode ser também implementado por meio de uma rede neural totalmente conectada e sem realimentação (veja Seção 6.2 para uma discussão mais detalhada), com as seguintes ressalvas:

- Há somente um neurônio conectando todas as entradas com a saída da rede;
- A função de ativação deste neurônio ( $\phi(\cdot)$ ) é a função identidade  $f(x) = x$ .

Para o treinamento, definem-se alvos para as duas classes de eventos e a função de erro:

$$E(\vec{w}) = \frac{1}{2}e^2(n) \quad (\text{B.2})$$

Nesta equação,  $\vec{w}$  é o vetor de pesos que define o plano de separação linear e  $e(n)$  é o erro na saída da rede com relação ao alvo para a classe escolhida na amostra  $n$ , ou seja  $e(n) = d(n) - y(n)$  ( $d(n)$  é o alvo). Desta forma, derivando-se esta função de erro com relação aos pesos sinápticos  $\vec{w}$ , mostra-se que o gradiente de  $E$  será:

$$\frac{\partial E(\vec{w})}{\partial \vec{w}} = -x(n)e(n) \quad (\text{B.3})$$

Nesta equação,  $x(n)$  representa a  $n$ -ésima entrada que leva o neurônio linear a



uma saída  $y(n)$ . E, com este resultado, define-se a fórmula de treinamento do LMS:

$$\hat{w}(n+1) = \hat{w}(n) + \alpha x(n)e(n) \quad (\text{B.4})$$

Nesta equação,  $\alpha$  representa a taxa de aprendizagem, um parâmetro para a suavização da trajetória do treinamento no espaço da função de erro  $E(\vec{w})$ . A Equação (B.4) representa a fórmula clássica do treinamento de um classificador LMS, indicando um método bem-definido de atualização dos pesos da rede para que se convirja a um erro mínimo. Em implementações realísticas de um classificador LMS, é possível agrupar a correção dos pesos  $\hat{w}$  em bateladas ou épocas, de forma que se suavize a migração do sistema para o mínimo global. No treinamento em épocas, os pesos sinápticos são corrigidos tendo por base a média dos erros para todos os eventos de uma época, ao invés da atualização instantânea proposta pela Equação (B.4).

Como é possível inspecionar diretamente na Equação (B.4), os pesos são atualizados segundo um produto interno dos erros multiplicados pelas respectivas entradas, uma época do treinamento. No caso de existirem grandes diferenças de magnitude entre as variáveis de entrada de um sistema LMS, é possível que o processo de treinamento se torne tendencioso em função destas variáveis [92]. Para eliminar o risco de tendência no treinamento devido à magnitude das componentes da entrada, é também prática que cada entrada seja subtraída da sua média e dividida pela sua variância.

Finalmente, a Figura B.1 mostra o diagrama de blocos de um discriminador LMS. O valor da saída  $y$  é utilizado para definir a R.O.C. do discriminador e escolher, ao invés de quatro (como no caso do EGammaHypo), apenas um corte que maximize a capacidade discriminante do sistema. Alvos ( $t$ ) para cada uma das classes são escolhidos, e o sinal de erro é utilizado para corrigir os pesos sinápticos até que o sistema convirja para o erro mínimo.

## B.1 Processamento Neural

A Figura B.2 contém uma modelagem de um neurônio genérico. Analogamente ao sistema LMS descrito na Seção 6.1.5, os elementos processadores de uma rede

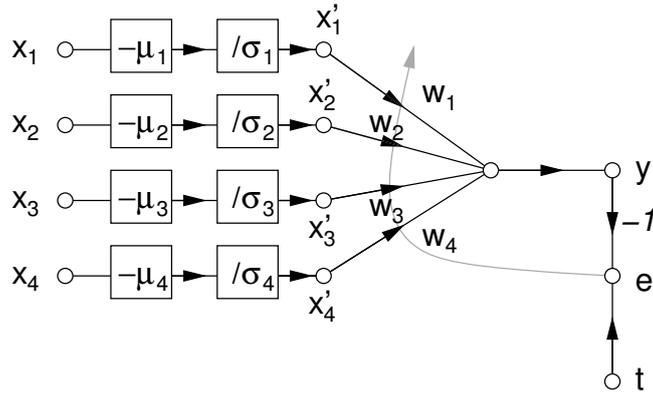


Figura B.1: O diagrama de fluxo do discriminador LMS que será empregado na discriminação elétron-jato.

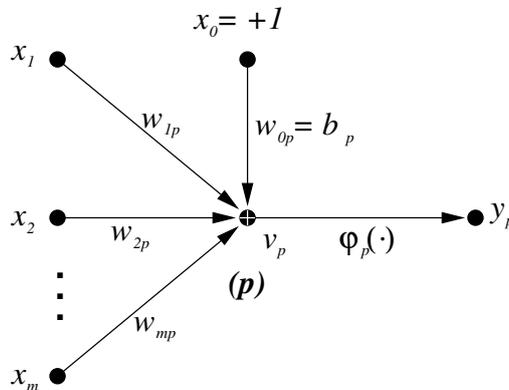


Figura B.2: Grafo de fluxo de sinal de um neurônio artificial.

neuronal, chamados neurônios ou percéptrons, são ativados por um conjunto de entradas  $\vec{x}$ , somadas de acordo com um sistema linear de ponderação  $\vec{w}$ . O diferencial entre o sistema linear proposto anteriormente e um neurônio está na função de ativação que conduz o campo induzido  $v_p = \vec{x}^T * \vec{w}_p$  à saída. Enquanto que no caso do LMS utiliza-se a função identidade, i.e.,  $y_p = v_p$ , no caso de percéptrons, faz-se uso de uma função não-linear, como por exemplo a função tangente hiperbólica:

$$\tanh(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}} = \frac{e^{2v} - 1}{e^{2v} + 1} \quad (\text{B.5})$$

Ou a função logística (simplificada):

$$P(v) = \frac{1}{1 + e^{-v}} \quad (\text{B.6})$$

A razão da escolha de uma função não-linear para a ativação de um campo

induzido pode ser qualificado a partir da seguinte constatação: se o sistema que se deseja classificar apresenta um comportamento gaussiano, para ambas as classes, i.e., os momentos de ordem superior a 2 para ambas as classes são todos iguais a zero, o classificador ótimo (bayesiano) resume-se a um sistema linear [75]. Naturalmente nota-se que é sempre possível *aproximar* ou modelar um sistema não-gaussiano em um desta espécie, tendo-se, por conseqüência, os erros relativos a esta aproximação. De fato, isto foi realizado na Seção 6.1.5, quando escolheu-se utilizar um classificador linear para separar as quatro variáveis definidas pelo T2Calo.

Se o sistema que se deseja resolver apresenta um comportamento não-gaussiano, o classificador ótimo não pode ser representado por um classificador linear. Neste caso, utiliza-se uma função de ativação não-linear para que se aproxime, de alguma forma, o comportamento não-linear do sistema de interesse. As funções nas Equações (B.5) e (B.6) são normalmente utilizadas por apresentarem amplitude limitada e possuírem derivada trivialmente calculável. A razão de procurar-se funções com derivadas simples (e suaves) ficará mais clara adiante, quando se definir o algoritmo de treinamento. A amplitude limitada pode rapidamente ser verificada como uma característica de interesse, observando-se que o treinamento de um sistema que trabalha por aprendizado é muitas vezes executado com a retro-alimentação de erros. Nesse caso, para evitar oscilações e pontos de ressonância, um sistema cuja saída seja limitada apresenta natural vantagem se comparado a outro análogo [75].

Este estudo está focado na utilização de RNA's utilizando percéptrons em múltiplas camadas (do inglês *Multi-layer perceptrons* ou MLP), completamente conectadas, sem realimentação e com treinamento baseado tanto na retropropagação de erros simples (do inglês *back-propagation*) quanto na retropropagação resiliente (do inglês *resilient back-propagation*). Introduziremos o processo de treinamento destes sistemas a seguir.

### **B.1.1 Treinamento por retro-propagação de erros simples**

Para sistemas com apenas uma camada neural, observa-se que, ao buscar-se o ponto de mínimo da superfície de erro (quadrático), naturalmente otimiza-se o conjunto de pesos de tal forma que o sistema consiga mapear a entrada nos alvos de saída. A única restrição do método é que a função de ativação do campo induzido do

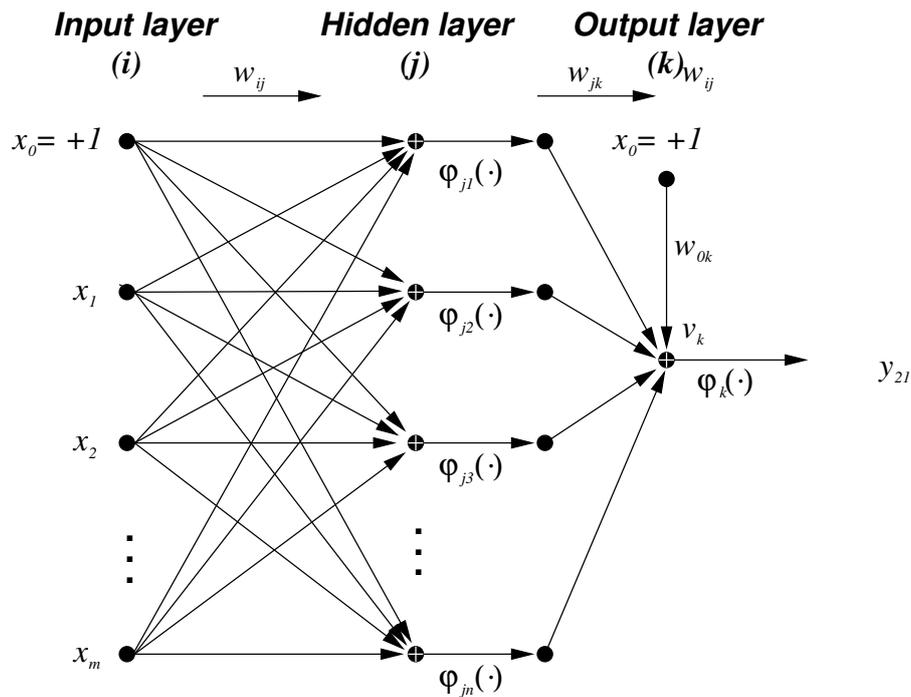


Figura B.3: Modelagem de uma rede MLP, totalmente conectada e sem retropropagação de sinal.

neurônio  $\varphi(\cdot)$  seja diferenciável. [93]. O problema do treinamento de uma rede MLP é um pouco mais complexo. Um grafo de fluxo de uma rede MLP completamente conectada pode ser visto na Figura B.3. Neste caso considera-se que a rede possui apenas uma camada escondida e apenas um neurônio de saída, o que se identifica intimamente com os casos de uso que abordaremos. Ainda sim, seria possível considerar casos com um número maior de neurônios de saída ou com mais camadas escondidas, generalizando os casos exemplificados aqui.

Para o sistema em questão, o mecanismo de treinamento para o neurônio de saída é idêntico ao treinamento de um perceptron simples, e pode ser facilmente definido da seguinte forma:

$$\text{Definindo o erro como } \mathcal{E}(n) = \frac{1}{2}e_k^2(n) \quad (\text{B.7})$$

$$\text{se observa que } \frac{\partial \mathcal{E}(n)}{\partial w_{jk}(n)} = \frac{\partial \mathcal{E}(n)}{\partial e_k(n)} \frac{\partial e_k(n)}{\partial y_k(n)} \frac{\partial y_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial w_{jk}(n)} \quad (\text{B.8})$$

$$\text{então } \frac{\partial \mathcal{E}(n)}{\partial w_{jk}(n)} = e_k(n)(-1)\varphi'_k(v_k(n))y_k(n) \quad (\text{B.9})$$

$$\text{resumindo, } \Delta w_{jk}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{jk}(n)} = \eta e_k(n)\varphi'_k(v_k(n))y_k(n) \quad (\text{B.10})$$

Normalmente, define-se o gradiente local como:

$$\delta_k(n) = -e_k(n)\varphi'_k(v_k(n)) \quad (\text{B.11})$$

$$\text{e dessa forma, } \Delta w_{jk}(n) = \eta \delta_k(n)y_k(n) \quad (\text{B.12})$$

Este sistema de equações segue o princípio definido anteriormente para o LMS, mas sem assumir nada sobre a função de ativação  $\varphi(\cdot)$ . Nota-se que para uma função de ativação onde  $y_k(n) = v_k(n)$ , recai-se no algoritmo de treinamento do LMS:  $\Delta w_{jk}(n) = \eta e_j(n)y_j(n)$ . Dessa forma, é possível considerar o LMS como um caso especial de uma rede neural com apenas uma camada e cuja função de ativação é a função identidade. Nestas equações,  $n$  representa a época ou batelada de treinamento.

O segundo caso de interesse acontece quando o neurônio  $j$  está localizado em uma camada oculta da rede. Neste caso, não existe uma resposta desejada (alvo) específica para aquele neurônio. Desta forma, tentar-se-á definir o sinal de erro de um neurônio escondido recursivamente, através da retro-propagação do erro na saída final da rede em direção ao neurônio desejado. Iniciamos, intuitivamente, com a definição do gradiente local do neurônio escondido  $j$ :

$$\delta_j(n) = -\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \varphi'_j(v_j(n)) \quad (\text{B.13})$$

Tendo em conta que para o caso específico em análise:

$$\mathcal{E}(n) = \frac{1}{2}e_k^2(n) \quad (\text{B.14})$$

$$\text{conclui-se que } \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = e_k \frac{\partial e_k(n)}{\partial y_j(n)} = e_k \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \quad (\text{B.15})$$

A primeira derivada parcial,  $\partial e_k(n)/\partial v_k(n)$ , já foi calculada para o caso do neurônio de saída, na passagem da Equação (B.8) para a Equação (B.9). Utiliza-se a mesma analogia aqui. Para o cálculo da segunda derivada parcial, pela Figura B.3 deduz-se que:

$$v_k(n) = \sum_j w_{jk}(n)y_j(n) \quad (\text{B.16})$$

$$\text{e, portanto: } \frac{\partial v_k(n)}{\partial y_j(n)} = w_{jk}(n) \quad (\text{B.17})$$

Assim sendo, o gradiente local do neurônio escondido  $j$  assim se define:

$$\delta_j(n) = \varphi'_j(v_j(n))\delta_k(n)w_{jk}(n) \quad (\text{B.18})$$

Substituindo  $\delta_j(n)$  na Equação (B.12), chega-se a fórmula de treinamento do neurônio escondido. No caso onde há muitas saídas na rede, a Equação (B.18) é trivialmente redefinida da seguinte forma:

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n)w_{jk}(n) \quad (\text{B.19})$$

No caso em que existem múltiplas camadas escondidas, propaga-se recursivamente o erro em direção à camada de entrada, aplicando-se a Equação (B.19).

## Convergência

O algoritmo de retropropagação usa uma estimativa do gradiente da superfície de erro no espaço dos pesos. Por esta razão, este sistema possui uma inerente natureza estocástica e tende a oscilar ao redor da direção de convergência ótima, ao mínimo da função de erro. Isso acontece pois é difícil antever se inclinações demasiado bruscas ou suaves em uma das direções da superfície de erro não influenciarão excessivamente o deslocamento do sistema. Muitas vezes, a utilização de um amortecimento (ou *momento*) durante o treinamento pode melhorar a resposta do sistema. Utiliza-se esta técnica para suavizar a migração das redes em direção ao mínimo da superfície de erro. Outra técnica é a normalização dos dados de entrada da rede, de forma que se evite que a diferença de magnitude e variabilidade das componentes de entrada cause tendências no treinamento neural.

Um problema recorrente no treinamento de redes neurais são mínimos locais, que podem fazer com que o sistema fique “preso” em uma região que não represente o mínimo global da superfície de erro. Esta característica não é somente mais uma consequência da utilização da estimativa instantânea para definir a direção de movimentação dos pesos, mas muitas vezes ocorre por dispor-se de uma quantidade limitada de eventos que representem o fenômeno que se deseja mapear.

Para que se assegure de que o sistema convirja sempre a um patamar equivalente de mínimo, deve se realizar um número de experimentos com os mesmos parâmetros de treinamento e teste para todos os casos de estudo. Em cada teste inicializam-se os pesos sinápticos a partir de um ponto diferente da superfície de erro. Desta forma, será possível detetar e avaliar se o problema em estudo estará sujeito a mínimos locais.

### **B.1.2 Treinamento por retro-propagação de erros resiliente**

Embora a retro-propagação de erros clássica seja um dos mais amplamente utilizados algoritmos para o treinamento de redes neurais, ela apresenta problemas fundamentais ligados a sua implementação, dentre os quais a difícil escolha dos parâmetros de treinamento. Esta parametrização é fortemente dependente do problema abordado e requer normalmente um certo empirismo para seu ajuste. A técnica de retro-propagação de erros resiliente (originalmente descrita em [76]), descreve uma forma de atualizar os pesos sinápticos de uma rede neural sem retro-alimentação que praticamente remove a necessidade de parametrização do treinamento. Neste método de treinamento, a atualização dos pesos sinápticos é feita de forma independente do sinal de erro retro-propagado, dependendo somente da direção do gradiente. Com esta técnica, o processo de atualização dos pesos sinápticos torna-se independente do valor do erro retro-propagado, o que, segundo os autores, melhora a dinâmica de treinamento. O trecho de pseudo-código abaixo descreve o algoritmo:

```
Para todos os pesos {  
  se (GradErro(t-1) * GradErro(t) > 0) então {  
    Taxa(t) = mínimo(Taxa(t-1) * Incentivo, TaxaMáxima)  
    Atualização(t) = -sinal(GradErro(t)) * Taxa(t)  
  }  
}
```

```

de outra forma, se (GradErro(t-1) * GradErro(t) < 0) então {
    Taxa(t) = máximo(Taxa(t-1) * Repressão, TaxaMínima)
    Atualização(t) = -Atualização(t-1)
    GradErro(t) = 0
}
de outra forma, se (GradErro(t-1) * GradErro(t) = 0) então {
    Atualização(t) = -sinal(GradErro(t)) * Taxa(t-1)
}
Peso(t+1) = Peso(t) + Atualização(t)
}

```

A atualização funciona da seguinte forma:

1. Se o gradiente do erro mantiver o mesmo sinal entre duas atualizações consecutivas da taxa de aprendizado, significa que se encontra em uma descida na superfície de erro e, para aumentar a velocidade, multiplica-se a taxa de aprendizado ( $Taxa(t)$ ) por um fator de incentivo. O valor do fator de incentivo é normalmente igual a 1,2, funcionando bem para vários tipos de problemas;
2. Se o gradiente do erro muda de sinal entre duas atualizações, então perde-se o ponto de mínimo e deve-se:
  - (a) Retroceder a sinapse do valor do último ajuste (já que foi demasiado);
  - (b) Reprimir a taxa de aprendizado, fazendo com que o próximo passo seja menos longo, para atingir-se o mínimo. O valor de repressão é normalmente fixo, igual a 0,5;
  - (c) Força-se o valor do gradiente a ser zero, de tal forma que na próxima atualização, o sistema recaia no último caso, onde a atualização do peso não terá um incentivo;
3. Se o gradiente do erro for zero, então é porque o último passo falhou o mínimo (zero forçado) e portanto deve-se simplesmente avançar, mantendo-se a mesma taxa de aprendizagem.

Para evitar *overflows* e *underflows*, institui-se um mínimo e um máximo para a taxa de aprendizado, em  $10^{-6}$  e 50, respectivamente. A escolha de fatores de



incentivo e repressão dependem, naturalmente, da superfície de erro na qual se está trabalhando. Superfícies com mais mínimos locais podem requerer valores de incentivo maiores.

### B.1.3 Funções de ativação

A função de ativação  $\varphi(\cdot)$  deve ser diferenciável em toda a extensão do domínio de interesse. No entanto, para simplificar o cálculo computacional dos gradientes locais, é habitual a escolha de funções que possuam um cálculo trivial de suas derivadas. Como exemplo, foram citadas as funções tangente hiperbólica (Equação B.5) e a função logística simplificada (Equação B.6). No caso da função logística:

$$\varphi(z) = \frac{1}{1 + e^{-z}} \quad (\text{B.20})$$

$$\text{daí } \varphi'(z) = \frac{e^{-z}}{[1 + e^{-z}]^2} \quad (\text{B.21})$$

Levando-se em consideração que para um neurônio genérico  $y = \varphi(v)$ , então:

$$\varphi'(v) = \frac{e^{-v}}{[1 + e^{-v}]^2} = \frac{1 + e^{-v}}{[1 + e^{-v}]^2} - \frac{1}{[1 + e^{-v}]^2} = \varphi(v)[1 - \varphi(v)] = y(1 - y) \quad (\text{B.22})$$

No caso da tangente hiperbólica, equivalentemente:

$$\varphi(z) = \tanh(z) \quad (\text{B.23})$$

$$\varphi'(z) = (\text{sech})^2(z) = 1 - \tanh^2(z) \quad (\text{B.24})$$

$$\varphi'(v(n)) = 1 - \varphi^2(z) = 1 - y^2(n) \quad (\text{B.25})$$

Estas duas funções de ativação permitem um cálculo absolutamente trivial do gradiente local, e por esta razão são computacionalmente bastante eficientes. No contexto deste trabalho utilizaremos a função tangente hiperbólica como função de ativação dos percéptrons das redes estudadas.

# Apêndice C

## *NeuralRinger*: Projeto e implementação

O projeto do *NeuralRinger* está dividido em quatro sub-pacotes que executam funções bem-definidas para o interfaceamento, configuração e execução de redes neurais. A Figura C.1 contém um diagrama UML (do inglês, *Unified Modelling Language*, [74]) nomeando e mostrando a relação de dependência entre estes pacotes.

**SYS:** O sub-pacote `sys` está abaixo na cadeia de dependências e não depende de nenhum outro. Este pacote inclui ferramentas para a manipulação de arquivos no formato XML (do inglês *eXtensible Markup Language*, [94]) e um sistema básico para o relatório de erros. A escolha da linguagem XML, como formato de troca de dados e valores de configuração, vem da grande quantidade de interpretadores (*parsers*) disponíveis livremente na *Internet*, já que é um padrão bastante difundido. O suporte a XML disponível atualmente inclui, mas não se limita a, verificações de sintaxe automatizadas (via *XML schemas*) e transformações para que se adicione informação ou seja possível a visualização, de forma adequada, do conteúdo de uma base de dados XML.

A Figura C.2 traz uma visão geral dos componentes e seu relacionamento dentro deste pacote. Na parte central do diagrama, encontra-se o tipo `Reporter` que define uma interface para o relatório de erros. Objetos deste tipo contêm uma implementação específica do sistema de relatório de erros. Inicialmente, somente a implementação baseada em arquivos-padrão tais como `std::cout` e `std::cerr` [35]

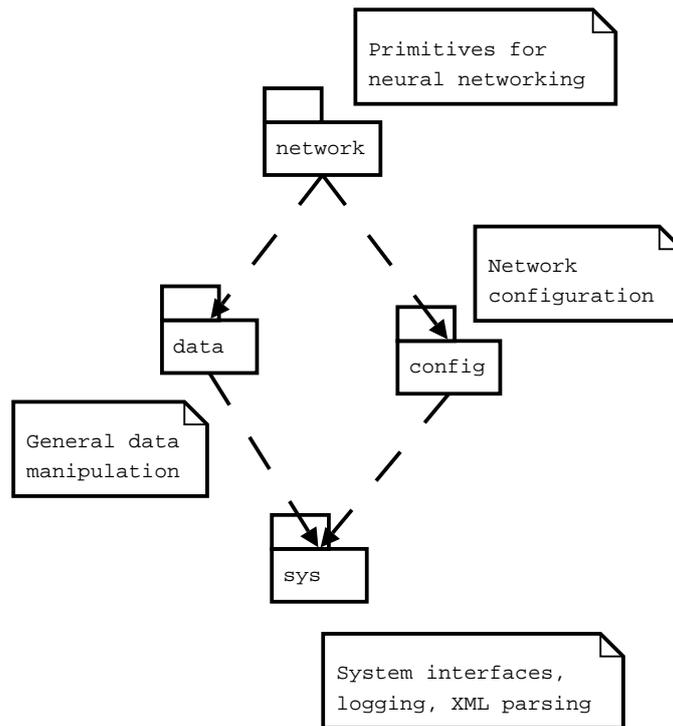


Figura C.1: Diagrama de blocos mostrando a relação entre os pacotes do *NeuralRinger*.

foi desenvolvida.

À direita observa-se a modelagem para o sistema de interpretadores XML. Uma interface (abstrata) é utilizada como base para implementações específicas baseadas em interpretadores habitualmente encontrados nos sistemas operacionais correntes. Especificamente, implementações para o sistema Xerces C++ [95], do grupo Apache e libxml2 [96], do grupo GNOME, estão presentes na implementação atual. O interpretador XML utiliza-se do sistema de relatório de erros para relatar problemas na leitura ou escrita de arquivos de forma unificada. Uma caixa de ferramentas baseada no tipo abstrato `XMLProcessor` provê um conjunto de primitivas para facilitar a escrita e a leitura de parâmetros e trechos de texto. No evento de erros fatais, uma exceção baseada no tipo `Exception` será lançada pela parte afetada do código.

O pacote `sys` também fornece um sistema para a interpretação de opções de linha de comando, baseado no interpretador `popt`. Este tipo imbuete um sistema de atribuição automática e verificação de valores, conectando as opções na linha de comando diretamente com o contexto onde as opções serão utilizadas dentro do programa. Seguindo a filosofia dos interpretadores XML, este tipo também faz uso

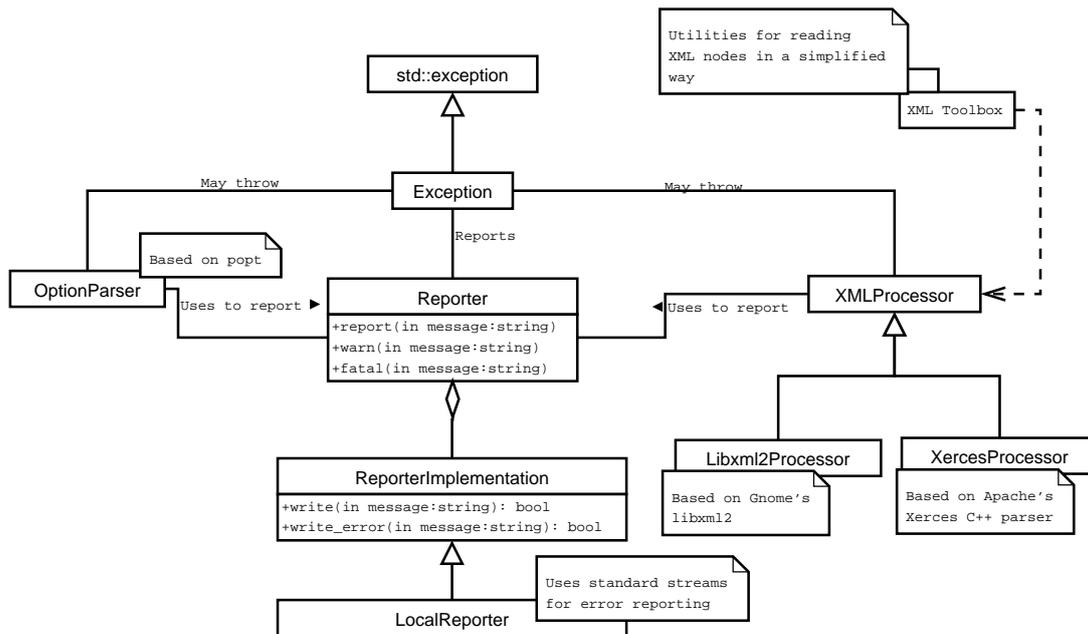


Figura C.2: Diagrama UML mostrando as relações dos componentes do pacote `sys`.

do sistema central de relatório de erros para informar problemas ao usuário e poderá, no caso de problemas sérios, também lançar exceções de operação.

**DATA:** O pacote `data` contém as primitivas para a manipulação dos dados, prévia e posteriormente ao processamento neural. Este sub-sistema utiliza-se das primitivas definidas no pacote `sys` para executar a reportagem de erros, lançamento de exceções ou a leitura de arquivos XML. A implementação das primitivas neste pacote utiliza elementos otimizados da biblioteca `GSL` (*GNU Scientific Library*, [97]) para a manipulação de valores contidos em matrizes e vetores. A Figura C.3 mostra um diagrama indicando a relação das classes dentro do pacote `data`.

O elemento central deste pacote chama-se `PatternSet`. Ele representa um conjunto de vetores ou padrões formando uma matriz de dados. Esta matriz é representada internamente por um objeto tipo matriz `GSL` (`gsl_matrix`). Para o armazenamento dos dados, utilizou-se a seguinte estratégia: cada evento do fenômeno observado será guardado em uma linha da matriz. Neste caso, cada coluna representará uma característica (ou *Feature* no inglês) do dado. A biblioteca `GSL` permite acesso rápido tanto aos padrões (linhas) como características (colunas) da matriz de dados.

Baseando-se nas primitivas de acesso ao tipo matriz `GSL`, a classe base `PatternSet`

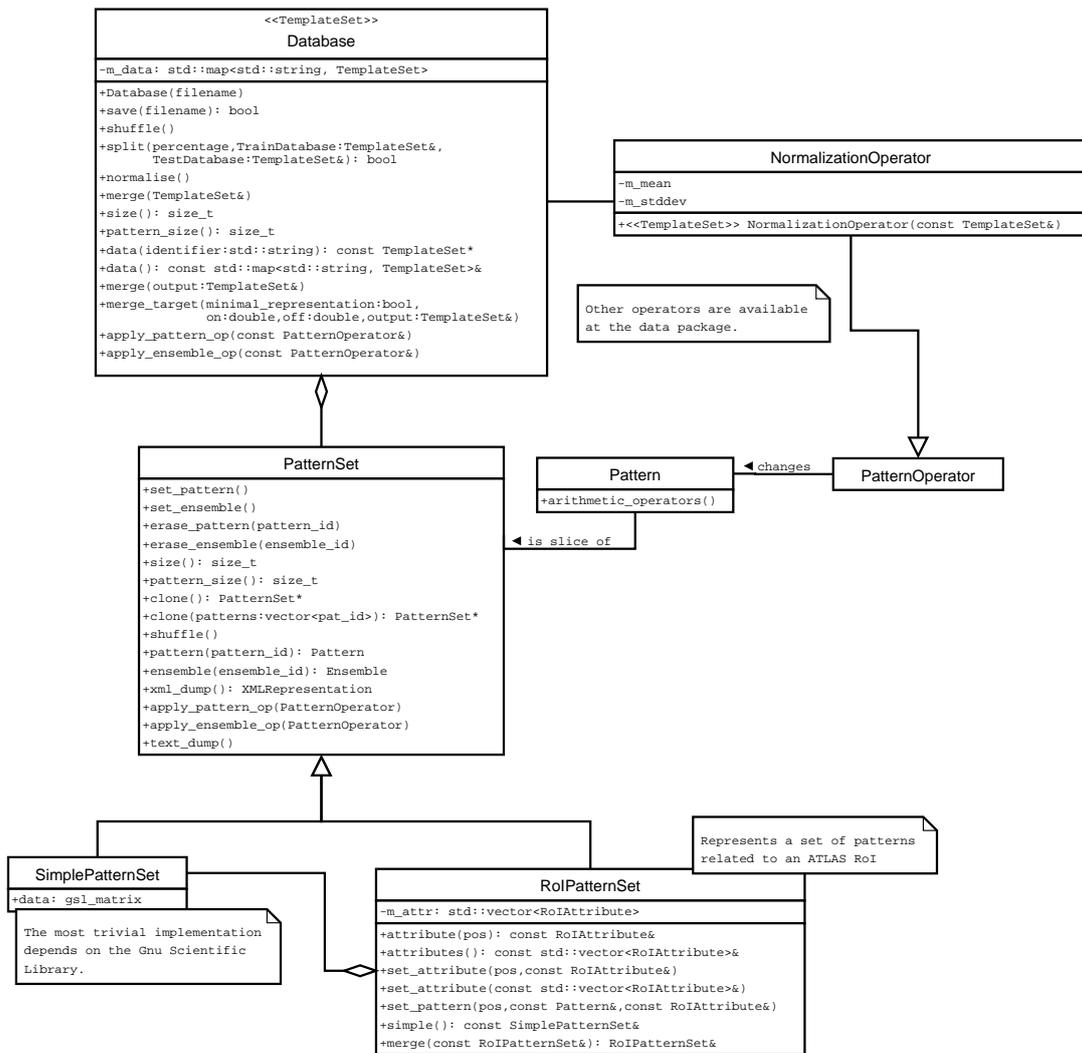


Figura C.3: Diagrama UML mostrando as relações dos componentes do pacote data.

desenvolve métodos típicos utilizados durante a manipulação dos dados, tais como acesso simplificado, leitura e escrita em disco, mistura dos padrões, manipulação de campos específicos e cópia. O formato de escrita em disco escolhido utiliza o padrão XML. O banco de dados básico contém um cabeçalho com as informações do autor e dos dados, indicando a data de criação e a última atualização. Dentro do arquivo, os dados podem ser divididos em classes. Cada classe é dividida por sua vez em padrões, que contêm um número opcional de atributos.

Para o caso específico do processamento de dados de RoIs, uma classe é derivada da implementação de base de um `PatternSet`, chamada `RoIPatternSet`. Esta classe implementa, adicionalmente à classe de base, métodos para a salvaguarda, leitura e manipulação de campos especificamente relacionados a uma RoI, assim como os identificadores do primeiro nível de filtragem e os valores das coordenadas  $\eta$  e  $\phi$ . Desta forma, é possível correlacionar os resultados obtidos com outros testes ou com relação ao posicionamento da região de interesse. Exemplos de conjuntos de padrões são elétrons ou jatos.

Uma coleção de um ou mais conjuntos de padrões é chamada `Database`, ou banco de dados. Um banco de dados reúne os diferentes conjuntos de padrões que se deseja processar através de um mapa. Para cada conjunto de padrões, o usuário pode atribuir um nome, mapeando os dados propriamente ditos a um valor legível por uma outra pessoa. Usando um objeto da classe `Database`, é possível manipular os conjuntos de padrões de uma forma unificada, por exemplo, quando deseja-se extrair a média ou desvio padrão global dos conjuntos de padrões de um problema.

Operadores (`PatternOperators` no diagrama), são elementos que podem modificar um padrão e são usados para extrair a média, variância ou aplicar normalizações segundo diferentes critérios nos dados disponíveis. Na figura, mostra-se o mais relevante dos operadores na biblioteca, chamado *NormalizationOperator*, que é inicializado a partir de uma base de dados e pode, quando aplicado a um padrão, remover a média e desvio padrão da base de dados. Este operador é utilizado em todos os testes neste trabalho como forma de normalização dos dados de entrada de classificadores LMS ou Redes Neurais. Quando um operador é aplicado a um padrão, primitivas da biblioteca GSL são utilizadas para que a eficiência do processo seja maximizada.

**NETWORK:** Este pacote contém a definição de neurônios, sinapses e redes formadas a partir destes elementos. A Figura C.4 apresenta um diagrama UML dos tipos neste sub-sistema. No topo deste diagrama está o tipo **Network** (Rede), que representa, de forma abstrata, um conjunto de neurônios (tipo **Neuron**) interconectados via objetos do tipo sinapse (tipo **Synapse**). Uma rede pode ser executada utilizando-se um dos métodos `run()` e treinada usando-se uma das variantes de `train()`. De fato, objetos do tipo **Network** não contêm nenhuma informação direta sobre a topologia da rede que executam, funcionando apenas como agentes delegadores de responsabilidade. O sistema funciona da seguinte forma:

- **Para executar a rede:** Neste modo, o usuário irá iniciar o processo aplicando à entrada da rede um padrão ou conjunto de padrões. O objeto do tipo **Network** irá, então, para cada coluna da matriz de dados de entrada, chamar o método `run()` dos neurônios de entrada, correlacionando os dados do usuário com as posições relativas do neurônio da camada de entrada. Opcionalmente, se existirem neurônios de polarização (**BiasNeurons**), o mesmo será feito para cada um destes neurônios. Cada neurônio de entrada e polarização é responsável por propagar seu sinal aos neurônios ao qual está conectado, através das sinapses (objetos do tipo **Synapse**). As sinapses irão multiplicar o valor de suas entradas pelo seu próprio peso e propagar o sinal ao neurônio destino. Este último, após receber todas as entradas relativas às sinapses que estão conectadas a si, executará o mesmo procedimento até que o sinal chegue ao neurônio de saída da rede.
- **Para treinar a rede:** O sistema de treinamento funciona de modo equivalente. Uma vez calculado o sinal de erro, ou a *lição* a partir da técnica que convier, este sinal é injetado ao neurônio de saída, que retropropagará o sinal até o neurônio de entrada.

Os tipos de neurônio são divididos nas seguintes classes:

- **Entrada (**InputNeuron**):** são neurônios que recebem a entrada da rede, opcionalmente aplicando um fator de normalização aos dados e propagando o sinal para neurônios escondidos ou de saída. Como estão no início da rede, estes neurônios não podem retropropagar sinais recebidos em suas saídas para fins

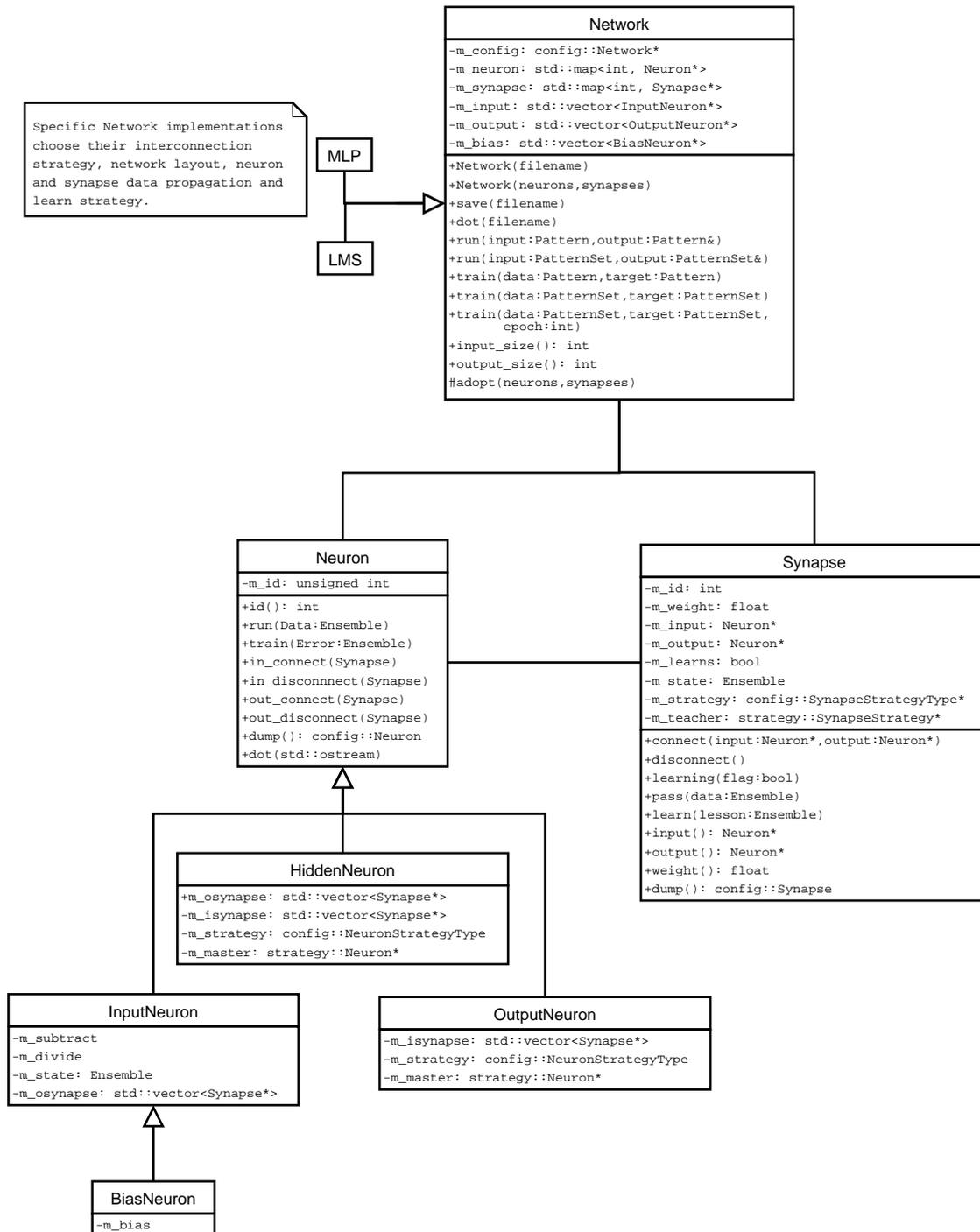


Figura C.4: Diagrama UML mostrando as relações dos componentes do pacote network.



de treinamento. Uma vez que atingem estes neurônios, os sinais de erro retro-propagados são apagados da memória e o passo de treinamento é considerado como executado;

- Polarização (`BiasNeuron`): este tipo herda suas características da classe `InputNeuron`, mas não permite que o usuário defina uma entrada arbitrária que será propagada a rede. No lugar, é possível definir um valor fixo (normalmente +1), que é alimentado, via sinapses de conexão, aos neurônios de saída ou escondidos;
- Escondido (`HiddenNeuron`): são neurônios que estão entre os neurônios de entrada e/ou polarização, conectando a entrada à saída. São opcionais, já que é possível desenvolver redes sem neurônios desta classe. Neurônios escondidos transmitem os dados recebidos em suas entradas a seus neurônios de saída, aplicando uma operação de soma aos valores de entrada, seguida, opcionalmente de uma função de ativação configurável;
- Saída (`OutputNeuron`): são os neurônios que estão na saída da rede. Seus valores constituem a saída do sistema neural. Assim como os neurônios escondidos, podem aplicar uma função de ativação ao campo induzido produzido somando-se os dados de entrada.

A execução da rede e treinamento das sinapses é regido por dois parâmetros que estão atrelados a estes elementos: as estratégias. Estratégias para neurônios (objetos da classe `NeuronStrategy`) definem de que forma os neurônios propagaram os sinais de erro quando são injetados na saída rede, para correção dos valores sinápticos. No momento, há somente um tipo de estratégia possível: a retropropagação de erros. Três funções de ativação estão implementadas: a função linear (utilizada para redes tipo LMS), a função logística simplificada, definida na Equação (B.6) e a função tangente hiperbólica, definida na Equação (B.5).

Métodos para salvar e restaurar uma rede a partir de um arquivo estão implementados. A salvaguarda é implementada através de um processo iterativo de delegação de responsabilidades, em que cada neurônio e sinapse envolvidos na rede se auto-descrevem no arquivo de saída. O formato escolhido é XML, uma vez que as bibliotecas de base do pacote `sys` já implementam as primitivas de leitura e escrita

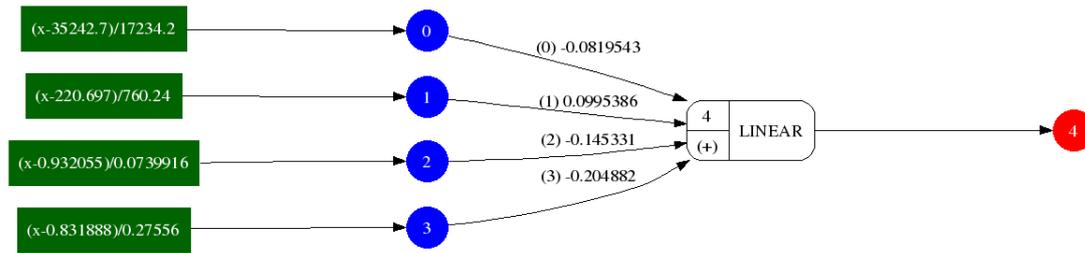


Figura C.5: Exemplo de um gráfico de fluxo produzido pelo método `dot()` para uma rede LMS, tal qual utilizada para os testes na Seção 6.1.5.

neste formato. Um segundo método (chamado `dot()`) está disponível para objetos do tipo rede. Este método implementa a escrita em disco de uma descrição da rede no formato `dot` [98], que permite a visualização dos pesos e interconexões da rede de forma gráfica. Exemplos de uma rede LMS e uma rede MLP encontram-se nas Figuras C.5 e C.6, respectivamente. As caixas na parte esquerda da figura representam o fator de normalização aplicado às entradas. Os círculos à esquerda, seguindo o caminho de propagação do sinal representam os neurônios de entrada. Cada neurônio possui uma numeração única, que o distingue dos outros. Este número corresponde ao identificador usado dentro das aplicações do *NeuralRinger*.

Neurônios escondidos e aqueles pertencentes à camada de saída são representados por uma caixa com as bordas arredondadas. Nesta caixa, o identificador do neurônio está representado na parte superior esquerda. Na parte inferior desta região, abaixo do identificador, um sinal de adição (+) indica que as entradas deste neurônio serão somadas antes da aplicação da função de ativação, definida na parte esquerda. Nós de polarização são mostrados junto ao neurônio a que estão conectados. Os diversos neurônios são conectados por retas que apontam na direção da propagação do sinal de entrada da rede. A saída da rede é mostrada usando-se um círculo e localizada na extrema direita da figura. O número dentro deste círculo corresponde ao identificador do neurônio de saída e casa com o número indicado no retângulo que precede o círculo.

**CONFIG:** Este pacote implementa a leitura e a escrita de parâmetros de redes a um arquivo XML. É um pacote que depende da infraestrutura provida pelo pacote `sys` e provê suporte à configuração das redes definidas no pacote `network`.

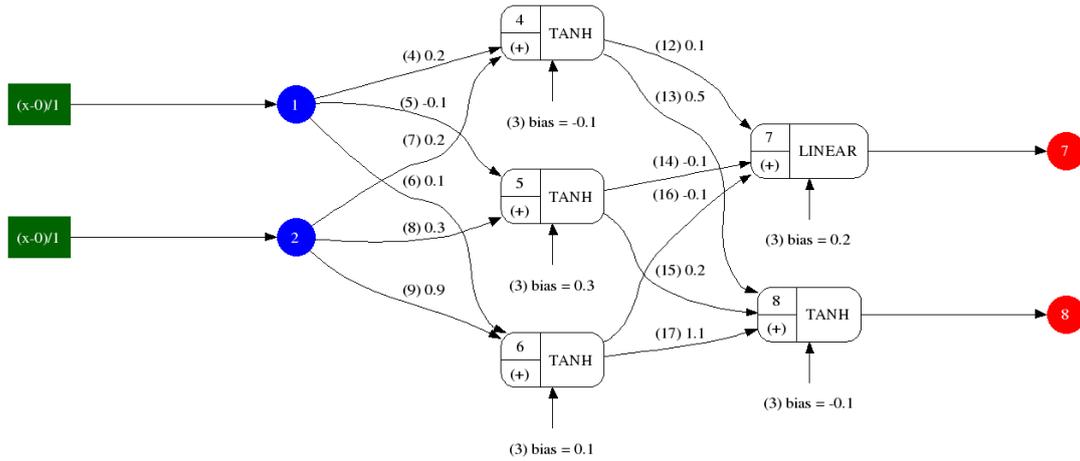


Figura C.6: Exemplo de uma gráfico de fluxo produzido pelo método `dot()` para uma rede MLP.

**Outros pacotes:** O *NeuralRinger* contém, ainda, 3 outros sub-projetos que apoiam sua funcionalidade:

- **lv11:** contém um simulador simplificado da filtragem baseada em calorimetria executada no LVL1. Este pacote pode ser utilizado para filtrar, dentre as regiões de interesse disponíveis, aquelas que seriam aprovadas pelo LVL1. Ele pode ser usado no lugar de uma simulação completa do Athena para separar dados para uma análise menos criteriosa. Para o filtro de referência, é preferível que o usuário utilize o programa Athena;
- **roiformat:** define o formato de intercâmbio entre o sistema de calorimetria implementado no Athena e o *NeuralRinger*, possibilitando que os dados disponíveis no primeiro possam ser facilmente escritos em disco e lidos por versões do *NeuralRinger* não acomodadas ao Athena. Isso normalmente diminui o tempo de desenvolvimento e teste, e é aconselhável;
- **rbuild:** este pacote contém primitivas para a construções de anéis baseados nos dados brutos das RoI disponíveis. Este método de detecção é uma alternativa mais eficaz ao processamento utilizando as características definidas pelo T2Calo. Este sistema será introduzido mais adiante.

## Aplicações:

Um conjunto de aplicativos foi escrito utilizando o sistema de pacotes definido na seção anterior. Estes aplicativos podem ser executados independentemente do sistema Athena, o que simplifica o desenvolvimento e treinamento das redes neurais, assim como a depuração de eventuais problemas no código. Estes são os programas disponíveis:

- **eta-filter**: Este programa pode sub-dividir uma base de dados contendo RoIs, de forma a classificar os dados por segmentos em  $\eta$ . O programa é iniciado com um arquivo contendo as RoIs no formato XML padrão e com um valor para  $|\Delta_\eta|$ . Para cada sub-intervalo de tamanho  $|\Delta_\eta|$  dentro do intervalo  $[-2, 5, +2, 5[$ , um arquivo de dados é gerado. Estes arquivos conterão dados para RoIs nos respectivos intervalos em  $\eta$ . O processamento leva em consideração a localização da RoI determinada pelo LVL1, que está disponível nas bases de dado XML;
- **getroi**: Separa RoIs de uma base de dados no formato `roiformat` para inspeção por um programa externo;
- **lms-train**: Implementa um sistema completo de treinamento de uma rede LMS. O sistema é configurável quanto aos parâmetros de treinamento, permitindo a especificação da taxa de treinamento, tamanho da época, critério de parada, tempo de parada abrupta e os nomes dos arquivos de entrada e saída;
- **lvl1-filter**: Este programa implementa um filtro baseado em uma simulação simplificada do LVL1. Ele pode ser utilizado para filtrar, dentre as regiões de interesse disponíveis, aquelas que seriam aprovadas pelo LVL1. O usuário poderá especificar os diversos limites de corte para a filtragem do primeiro nível:
  - Limite hadrônico;
  - Limite eletromagnético e
  - Isolamento eletromagnético;

- **merge**: Este programa pode aglutinar múltiplas bases de dados XML em um único arquivo de saída;
- **mlp-relevance**: Este programa avalia a relevância das características de entrada de uma rede (LMS ou MLP), tanto considerando-se o valor MSE quanto o produto SP originais da rede;
- **mlp-run**: Roda uma rede (LMS ou MLP) e escreve uma base de dados de saída com os resultados;
- **mlp-train**: Implementa um sistema completo de treinamento de uma rede MLP. De forma análoga ao programa `lms-train`, esta aplicação é configurável quanto aos parâmetros de treinamento, permitindo a especificação da taxa de treinamento, tamanho da época, critério de parada, tempo de parada abrupta e os nomes dos arquivos de entrada e saída;
- **relevance-filter**: Este programa pode remover uma ou mais colunas de uma matriz de dados, baseando-se na análise da relevância e um parâmetro de corte. Com este programa também é possível reduzir o espaço de entrada de um discriminador, minimizando as perdas na classificação, como será mostrado mais à frente;
- **ringer**: Este programa pode calcular, levando-se em consideração um padrão de normalização, os anéis de energia que serão utilizados nesse trabalho. A saída deste programa é uma base de dados no formato XML padrão do *NeuralRinger*;
- **ringer-run**: Este programa executa a extração e a discriminação de elétrons e jatos baseado na técnica do anelamento, descrita no Capítulo 6. Ele pode ser utilizado para o cálculo do desempenho do algoritmo de filtragem *offline* e permite uma otimização da execução antes da implementação no ambiente Athena;
- **xml2dot**: Escreve a representação de uma rede no formato `dot` para que seja possível a visualização tal qual no exemplo das Figuras C.5 e C.6;

- **xml2text**: Este programa converte uma base de dados no formato XML padrão do *NeuralRinger* em uma representação no formato texto, que pode ser lido por programas que não suportem XML.

Cada um dos programas responde ao comando `-help`, que imprime na tela o conjunto de opções e valores padrões disponíveis. Um conjunto de *scripts* de análise encontra-se no pacote, permitindo ao usuário a automação do treinamento e a análise dos resultados considerando-se figuras de mérito padrão para a análise baseada em calorimetria no ATLAS.

# Apêndice D

## Publicações

Durante a execução deste estudo, um conjunto de trabalhos foi publicado, com a participação do autor, na forma de artigos e notas técnicas. Este apêndice traz uma descrição deste material em ordem cronológica reversa. Os artigos mais recentes incluem a atuação durante o estudo que culmina nesta tese de doutoramento, enquanto que os mais antigos (anteriores ao ano de 2001), foram publicados durante a elaboração da dissertação do mestrado e da tese de graduação.

**Agosto de 2006:** *Deployment of the ATLAS High-Level Trigger* Publicado na *IEEE Transactions on Nuclear Science*. Referência [99].

O teste combinado com feixe do ATLAS na segunda metade de 2004 contou com o primeiro emprego dos Filtros de Alto-Nível do ATLAS (HLT). Este artigo revê a experiência adquirida durante este exercício, o estado atual e melhorias que deverão ser executadas antes do início do experimento. O artigo inclui discussões sobre o sistema de fluxo de dados, a integração de algoritmos de seleção, procedimentos de teste, distribuição de *software* e instalação.

**Mai de 2006:** *A configuration system for the ATLAS trigger* Publicado no *Journal of Instrumentation*. Referência [100].

Este artigo descreve o sistema a ser empregado para a configuração dos Filtros de Alto Nível do experimento ATLAS. Tal sistema provê ao

Sistema de Filtragem todos os parâmetros necessários à tomada de decisões e para a salvaguarda de seu histórico. O mesmo sistema reconfigura a reconstrução de eventos, as simulações de Monte Carlo e a análise de dados. Ferramentas para acessar e manipular este sistema também são discutidas.

**Abril de 2006: *Neural triggering system operating on high resolution calorimetry information*** Publicado na *Nuclear Instruments and Methods in Physics Research*. Referência [101].

Este artigo apresenta um sistema de discriminação elétron/jato para operar no Segundo Nível de Filtragem do experimento ATLAS. De forma a manipular a grande dimensionalidade de dados, as RoIs são organizadas na forma de anéis concêntricos de energia e, desta forma, tanto compressão como um melhor desempenho podem ser obtidos. A informação dos anéis é alimentada em uma rede neural sem realimentação (*feedforward*). Esta implementação resulta em uma detecção de elétrons em 97% contra um falso alarme de apenas 3%. A cadeia completa de discriminação é executável em menos de 500  $\mu$ s.

**Fevereiro de 2006: *Testing on a Large Scale : running the ATLAS Data Acquisition and High Level Trigger Software on 700 PC Nodes*** Apresentado no *Computing In High Energy and Nuclear Physics*. Referência [102].

O Sistema de Filtragem e Aquisição de Dados do ATLAS (TDAQ) terá à sua disposição, inicialmente, 2000 PCs assumindo a função de controlar, ler e filtrar os dados dos eventos do experimento. Este grande número de nós de processamento será adquirido somente em 2007. Em julho de 2005, no entanto, o CERN disponibilizou o uso de seu sistema de processamento *batch* durante 5 semanas para testes em larga escala dos elementos do TDAQ. No total, 700 nós de processamento foram utilizados no exercício. Este artigo discute os diversos aspectos abordados durante o teste, enfatizando um conjunto de melhorias que devam ser executadas antes da operação final.



**Outubro de 2005: Discriminação Neural de Elétrons no Segundo Nível de Trigger do ATLAS** Apresentado no XXVI Encontro Nacional de Física de Partículas e Campos. Referência [103].

Este trabalho apresenta um discriminador neural para o segundo nível de filtragem do ATLAS, atuando no problema de separação elétron/jato baseado em informações de calorimetria. Para reduzir a alta dimensionalidade dos dados de entrada, as regiões de interesse (RoI) identificadas no primeiro nível são organizadas em anéis concêntricos de deposição energética. Este tipo de pré-processamento dos dados permite eficiente compactação dos sinais e alcança elevada capacidade de identificar elétrons. Atualmente, esse sistema vem sendo portado para o ambiente de emulação do sistema de filtragem ATHENA, de modo a se obter uma avaliação realística de seu desempenho. O ambiente tem o objetivo de simular o comportamento do sistema de filtragem, ajudando, desta forma, no desenvolvimento e validação dos algoritmos. Em caráter comparativo, o sistema proposto foi também implementado usando a tecnologia DSP.

**Outubro de 2005: Otimização do Sistema de Trigger do Segundo Nível do ATLAS Baseado em Calorimetria** Também apresentado no XXVI Encontro Nacional de Física de Partículas e Campos. Referência [104].

Este trabalho apresenta um discriminador neural que opera sobre as quantidades calculadas pelo algoritmo T2Calo, responsável pela detecção elétron/jato no Segundo Nível de Filtragem do experimento ATLAS. Este sistema de detecção melhora a eficiência de detecção em quase 10 pontos percentuais, mantendo um nível de desempenho compatível com as restrições operacionais do sistema de filtragem.

**Julho de 2005: *Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC.*** Apresentado na *14th IEEE - NPSS Real Time Conference 2005 Nuclear Plasma Sciences Society*. Referência [105].

Os Altos Níveis de Filtragem do experimento ATLAS estão descritos em *software*, ao contrário de seu Primeiro Nível, que é baseado em *hardware*. Estes sistemas deverão reduzir a taxa de eventos de entrada de 75 kHz na saída do Primeiro Nível de Filtragem para apenas 200 Hz na saída do Filtro de Eventos. Os eventos selecionados serão finalmente guardados definitivamente em mídia permanente. Este artigo descreve uma visão geral de tal sistema. O desempenho é avaliado segundo simulações de Monte Carlo, tendo sido parcialmente demonstrado durante a bancada de testes com feixe em 2004. A eficiência para os canais com o sinal de interesse, a taxa esperada de seleção, o mecanismo de preparação de dados e os tempos de execução são discutidos. Alguns casos da Física de interesse que será analisada pelo experimento são detalhados.

**Julho de 2005: *Configuration of the ATLAS trigger*** Apresentado no *14th IEEE - NPSS Real Time Conference 2005 Nuclear Plasma Sciences Society*. Referência [106].

Este artigo discute um protótipo do sistema de configuração para os Altos Níveis de Filtragem do experimento ATLAS (HLT). Tal sistema configurará os algoritmos de filtragem no HLT para a detecção de eventos importantes para o experimento. Apresenta-se uma visão geral do HLT e a implementação do protótipo. Possíveis melhorias e tendências são discutidas no final do artigo.

**Outubro de 2004: Os Filtros de Alto Nível do Experimento ATLAS** Apresentado no XXV Encontro Nacional de Física de Partículas e Campos. Referência [107].

O Experimento ATLAS conta com um sistema de Filtragem bastante complexo e dividido em 4 grandes sub-sistemas:

- O Primeiro Nível de Filtragem que realiza os primeiros passos da seleção de eventos na cadeia de filtragem;
- O *Software Online*, responsável pelas áreas de controle e operação do sistema;

- O Sistema de Fluxo de Dados (*Dataflow*) que coordena a transmissão e armazenamento dos dados do detetor do experimento;
- Os Filtros de Alto Nível, que implementam os algoritmos de discriminação, representando o topo da cadeia de seleção de eventos no ATLAS.

Para aumentar a portabilidade entre os algoritmos de filtragem desenvolvidos por toda a comunidade do experimento, os desenvolvedores dos Filtros de Alto Nível (ou simplesmente HLT; *High-Level Triggers*) propuseram a reutilização do ambiente de programação *online* Athena dentro do sistema que operará em tempo real. Para tal, o HLT utiliza as ferramentas propostas pelo sub-sistema de Fluxo de Dados para coordenar as operações da transferência de informação para dentro e para fora dos nós de processamento sistema. Dentre outras restrições, o produto final deverá ser suficientemente rápido e operável em tarefas concorrentes em máquinas (SMP) com vários processadores rodando Linux. Neste trabalho apresentamos alguns dos problemas e soluções encontrados pelo grupo no desenvolvimento e teste do conjunto de bibliotecas que compõe o HLT.

**Setembro de 2004: *Implementation and Performance of the High Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC*** Apresentado na conferência *Computing in High Energy Physics and Nuclear Physics*. Referência [108].

Os Altos Níveis de Filtragem do experimento ATLAS (HLT) serão descritos em *software* e executados em PCs comercialmente disponíveis. Sua implementação é baseada na arquitetura padrão, orientada à objetos, disponível para análise *offline* no experimento. Este artigo traz uma visão geral da implementação atual da seleção de elétrons e fótons no HLT. Os níveis de desempenho desta implementação são avaliados tendo por base simulações de Monte Carlo, comparando-se a eficiência para o sinal de interesse, a taxa de eventos esperada, tempos de pré-processamento dos dados e para a discriminação em si. Alguns exemplos

de física são discutidos, assim como a experiência de utilização deste sistema na bancada de testes combinada com feixe em 2004.

**Agosto de 2004:** *The Second Level Trigger of the ATLAS Experiment at CERN's LHC* Publicado na *IEEE Transactions on Nuclear Science*. Referência [38].

O Sistema de Filtragem do ATLAS reduz a taxa de eventos interessantes a ser gravada para análise *offline* em 3 níveis sucessivos de filtragem, passando de 40 MHz iniciais a 100 kHz, 2 kHz e, finalmente, 200 Hz. Os Filtros de Alto Nível e o Sistema de Aquisição de Dados são projetados para se beneficiarem do mercado de computação e interconexão doméstica (PC's sob Ethernet) para atingir este nível de desempenho. Neste artigo, discutimos o Segundo Nível de Filtragem (LVL2) e apresentam-se resultados de medida de desempenho.

**Junho de 2004:** *Algorithms for the ATLAS high-level trigger* Publicado na *IEEE Transactions on Nuclear Science*. Referência [109].

Seguindo um rigoroso processo de projeto e análise em *software*, uma arquitetura baseada em objetos foi desenvolvida para deduzir o Segundo e Terceiro Níveis de Filtragem do experimento ATLAS, no LHC. Os componentes funcionais neste sistema, responsáveis por gerar as decisões, são algoritmos operando dentro desta arquitetura de *software*. Aspectos relevantes deste sistema são revistos neste artigo, tal como exemplos de algoritmos específicos e seu desempenho em cortes “verticais” de várias estratégias de seleção.

**Junho de 2004:** *The base-line DataFlow system of the ATLAS Trigger and DAQ* Publicado na revista *IEEE Transactions on Nuclear Science*. Referência [32]

As linhas de base do projeto e implementação do Sistema de Fluxo de Dados do ATLAS são descritas. Os componentes principais, suas

interações, bandas-passantes e taxas são discutidos assim como seu desempenho em um protótipo com 10% da dimensão final do experimento. Este sistema é uma combinação de componentes com projeto dedicado e aplicações multi-tarefas implementadas em C++ e rodando em PCs num ambiente baseado no Sistema Operacional Linux. Os nós de processamento são conectados por uma rede Gigabit-Ethernet padrão.

**Junho de 2004:** *ATLAS TDAQ data collection software* Publicado na revista *IEEE Transactions on Nuclear Science*. Referência [110].

O Sistema de Coleta de Dados (DC) é um componente do Sistema de Filtragem e Aquisição de dados do experimento ATLAS, no LHC. Ele é responsável pela movimentação de dados do Sistema de Leitura do Detetor (*Readout System*, ROS) para o Segundo Nível de Filtragem e para o Filtro de Eventos. Esta funcionalidade é distribuída em várias aplicações rodando sobre uma plataforma Linux, interconectada por Gigabit-Ethernet. Para o projeto e implementação destas aplicações, um enfoque comum foi adotado. Tal enfoque leva a uma implementação comum do *software* de DC, provendo uma infraestrutura de serviços de base a todas as aplicações deste sub-sistema.

**Outubro de 2003:** *Architecture of the ATLAS Online Physics-Selection Software at the LHC*. Apresentado no *8<sup>o</sup> ICATPP*. Referência [111].

A filtragem de eventos no experimento ATLAS é organizada em dois níveis distintos: o Segundo Nível de Filtragem e o Filtro de Eventos. Um enfoque unificado para selecionar eventos em ambos os níveis foi escolhido. Desta forma, um conjunto de rotinas de base foi projetada para maximizar o compartilhamento das interfaces e componentes *offline*, ainda que mantendo uma flexibilidade suficientemente grande para atender aos requisitos operacionais do Sistema de Filtragem, notavelmente aqueles relacionados ao desempenho e robustez. Este artigo descreve a arquitetura e o projeto do sistema de seleção de eventos e mostra como esta implementação é compatível com os desafios do experimento.

**Outubro de 2003: *Studies for a common Event Selection Software: from LVL2 to Offline reconstruction*** Publicado na revista *IEEE Transactions on Nuclear Science*. Referência [39].

A função primária dos Filtros de Alto Nível do experimento ATLAS será implementada em dois níveis independentes de detecção operando em componentes de *software*. Enquanto este enfoque provê um conjunto de interfaces de programação unificado ao sistema de seleção de eventos, também impõe requisitos bastante estritos nos componentes *offline* que devem ser re-utilizados. A decisão do Segundo Nível de Filtragem do ATLAS deve, tipicamente, ser atingida no espaço temporal de 10 milissegundos e conta com paralelismo em múltiplas tarefas de processamento. De forma a atender a estas restrições, foram desenvolvidos protótipos que incorporam elementos do Sistema de Fluxo de Dados do ATLAS, dos Filtros de Alto Nível e do Sistema de Processamento *Offline*. De forma a implementar um ambiente homogêneo de programação para os Filtros de Alto Nível, um controlador para o sistema de execução do Segundo Nível de Filtragem foi desenvolvido. Tanto com fatias do sistema de seleção baseado no canal elétron/fóton quanto fatias de seleção de múons, demonstra-se que o desempenho necessário poderá ser atingido se os componentes *offline* são cuidadosamente projetados e otimizados para que também sejam aplicáveis aos Altos Níveis de Filtragem do experimento.

**Agosto de 2003: *Neural Particle Discrimination for Triggering Interesting Physics Channels with Calorimetry Data*** . Publicado na revista *Nuclear Instruments And Methods In Physics Research A - Accelerators, Spectrometers, Detectors And Associated Equipament*. Referência [112].

Experimentos em Física de Altas Energias usam sistemas de validação *online* (filtros) para distingüir a Física conhecida de eventos interessantes que devam ser analisados. Este artigo introduz um esquema de filtração para processadores com alta taxa de entrada, baseado em redes neurais artificiais. A técnica é aplicada ao problema de discriminação

elétron/jato, presente no Segundo Nível de Filtragem do experimento ATLAS, sendo construído no CERN, Suíça. A solução proposta possui um melhor desempenho que aquela proposta atualmente no CERN, tanto no caráter discriminatório quanto em velocidade, tornando-se um algoritmo candidato para implementação no experimento.

**Março de 2003: *ATLAS-TDAQ DataCollection Software*** . Publicado na revista *IEEE Transactions on Nuclear Sciences*. Referência [31].

O Sistema Coleção de Dados é um sub-sistema do Sistema de Filtragem e Aquisição de Dados do experimento ATLAS. Ele é responsável pela movimentação dos dados de cada evento, a partir do Sistema de Leitura até o Segundo Nível de Filtragem e ao Filtro de Eventos. Esta funcionalidade é distribuída por vários componentes de *software* operando sob Linux em PC's interconectados por Gigabit Ethernet. Para o projeto e implementação destas aplicações, um enfoque comum foi adotado. Este enfoque nos guiou ao projeto e implementação de um conjunto de interfaces compartilhadas ao sistema de Coleção de Dados, provendo uma suíte de serviços comuns.

**Março de 2003: *The DataFlow System of the ATLAS Trigger and DAQ*** . Apresentado no congresso *Computing for High-Energy Physics' 2003*. Referência [30].

Este artigo apresenta o projeto e implementação de um protótipo do Sistema de Fluxo de Dados do experimento ATLAS. Sua decomposição funcional é descrita e medidas de desempenho para cada componente individual são mostradas.

**Março de 2003: *Architecture of the ATLAS High Level Trigger Event Selection Software*** . Apresentado no congresso *Computing for High-Energy Physics' 2003*. Referência [113]

Os Altos Níveis de Filtragem do experimento ATLAS consistem de dois passos: o Segundo Nível e um Filtro de Eventos. Ambos serão

implementados em *software*, operando, em sua maior parte, sob componentes domésticos. O modelo da seleção de eventos em ambos também é unificado, de forma que um conjunto de interfaces comuns foi projetado para maximizar a coerência, ainda que mantendo bons níveis de flexibilidade para garantir a interoperabilidade aos dois níveis. Este enfoque é ainda estendido para permitir que o mesmo conjunto de programas e bibliotecas opere *offline*, facilitando o desenvolvimento. Este artigo descreve a arquitetura e o projeto deste sistema.

**Março de 2003: *Experience with multi-threaded C++ applications in the ATLAS DataFlow Software*** Apresentado na conferência *Computing for High-Energy Physics' 2003*. Referência [33].

O Sistema de Fluxo de Dados do experimento ATLAS é responsável pela recepção, estocagem e subsequente movimentação tanto parcial quanto completa dos dados de um evento para os Altos Níveis de Filtragem. O projeto deste conjunto de bibliotecas e programas é baseado em uma metodologia orientada à objetos e sua implementação baseia-se no uso de Tarefas (*threads*) POSIX e na Biblioteca de Modelos-Padrão do C++ (*Standard Template Library*). Este artigo apresenta nossa experiência com Linux, Tarefas POSIX e a Biblioteca de Modelos-Padrão do C++ no ambiente *online* do Sistema de Fluxo de Dados do ATLAS.

**Julho de 2002: *ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report*** Este documento descreve o projeto e construção do Sistema de Filtragem do experimento ATLAS em si, contendo todos os detalhes, medidas de desempenho e operabilidade. Foi publicado no CERN. Referência [16].

**Outubro de 2001: *Redes Neurais Especialistas para a Separação Elétron-Jato usando Calorímetros Multi-camadas e Multi-segmentados*** Apresentado no XX Encontro Nacional de Física de Partículas e Campos. Referência [114].

O experimento ATLAS estará operacional no ano de 2006. O objetivo principal deste experimento é a detecção do bóson de Higgs, usando



entre outros tipos de detetores, calorímetros. Um dos canais de detecção mais importantes no experimento é o de elétrons com alta energia transversa, representando de 30 a 40% do total das assinaturas a serem analisadas pelo Sistema de Filtragem. Jatos (de partículas) confundem-se comumente com elétrons pela forma que interagem com os calorímetros. Neste trabalho, apresentamos um sistema de discriminação elétron-jato baseado em redes neurais especialistas, utilizando os dados dos calorímetros do ATLAS. Este sistema, depois de treinado, compacta o espaço de variáveis de entrada (células dos calorímetros) em um subespaço que mantém os aspectos necessários para uma detecção eficiente de elétrons. Os resultados apresentados se mostram melhores que os resultados obtidos usando-se técnicas desenvolvidas no CERN, com o mesmo objetivo.

**Outubro de 2000: Mapeamento em Anéis para uma Separação Neuronal Elétron-Jato usando Calorímetros Multi-camadas e Multi-segmentados** Apresentado no XIX Encontro Nacional de Física de Partículas e Campos. Referência [79].

Propõe-se neste trabalho que a análise conduzida no Segundo Nível de Filtragem do Experimento ATLAS, no CERN, seja feita por meio de processamento neural sobre a região de interesse previamente destacada pelo Primeiro Nível nos Calorímetros. Por depender do posicionamento da RoI no detetor, os números de camadas, granularidades e profundidades das células do calorímetro são desconhecidos até o momento da chegada do evento ao sistema de análise. Ainda assim, estima-se que o número de células para análise estará em torno de 1000 por RoI. As eficiências de separação obtidas, tempos de execução e uma comparação com a eficiência de outros métodos empregados para a mesma atividade são discutidas.

**Outubro de 1999: Integrando Plataformas e Algoritmos para o Segundo Nível de Trigger do Experimento ATLAS** Apresentado no XVIII Encontro Nacional de Física de Partículas e Campos. Referência [115].

Este artigo contém um sumário do trabalho realizado no âmbito dos estudos de portabilidade da infraestrutura do fluxo de dados do Sistema de Filtragem, originalmente escritos em C e operando em Sistemas Operacionais comerciais para uma implementação orientada a objetos baseada em C++ rodando sobre Linux. Ele discute as vantagens deste enfoque, tanto em termos de manutenibilidade quanto do custo final de projeto.

**Outubro de 1998: Um Protótipo do Sistema de Validação de Nível 2 para as Condições do LHC** Apresentado no XVII Encontro Nacional de Física de Partículas e Campos. Referência [116].

O experimento ATLAS pretende comprovar a existência do bóson de Higgs. Para tal, um grande sistema de detecção e aquisição vem sendo projetado. O sistema aquisição tem a função de separar em tempo real interações originárias do decaimento de um Higgs de física ordinária.

A filtragem de eventos no sistema de aquisição é concebida em três níveis, de complexidade crescente e velocidade decrescente. O segundo nível pretende utilizar redes de computadores pessoais (PC's) interconectados por rápidos sistemas de rede. A escolha de fabricantes, sistemas operacionais e algoritmos de processamento ainda não foi feita, mas esforços em prol desta decisão vêm sendo realizados. Neste trabalho desenvolve-se uma fração do filtro de segundo nível utilizando-se de processamento paralelo, redes neurais artificiais e DSP's.

**Setembro de 1998: *Neural Classifiers Implemented In A Transputer Based Parallel Machine*** Apresentado no *International Meeting on Vector and Parallel Processing (VECPAR)*. Referência [117].

O projeto de um protótipo para o Segundo Nível de Filtragem para as condições do LHC é realizado através de uma implementação em uma máquina paralela. Este computador contém 16 nós de processamento baseados em *transputers* e usa processadores digitais de sinais (*DSP's*) como unidades de co-processamento para otimizar o processamento de

dados. Um ambiente de desenvolvimento em C é usado para obter das aplicações seu máximo desempenho. A implementação é baseada na informação provida pelos quatro sistemas de detecção do experimento ATLAS em duas fases: extração de características e decisão global. A extração de características para calorímetros e decisão global são desempenhadas por redes neurais artificiais. Os parâmetros de pré-processamento e da rede neural são mantidos em memória, e a função de ativação da rede é implementada usando uma tabela de procura. Dados simulados para o Segundo Nível de Filtragem são utilizados para a avaliação do desempenho deste sistema.