

Universidade de São Paulo
Instituto de Física

Estudo da reação de breakup ${}^4\text{He}({}^{17}\text{F}, {}^{16}\text{O}+\text{p}){}^4\text{He}$
usando o alvo ativo pAT-TPC: uma abordagem
usando técnicas de Machine Learning

Guilherme Ferrari Fortino

Orientador: Prof. Dr. Valdir Guimarães _____

Coorientador: Dr. Juan Carlos Zamora Cardona _____

Dissertação de mestrado apresentada ao Instituto de
Física da Universidade de São Paulo, como requisito
parcial para a obtenção do título de Mestre(a) em Ciências.

Banca Examinadora:

Prof(a). Dr(a). Nome do(a) Professor(a) - Orientador (instituição de trabalho)

Prof(a). Dr(a). Nome do(a) Professor(a) (instituição de trabalho)

Prof(a). Dr(a). Nome do(a) Professor(a) (instituição de trabalho)

São Paulo
2022

Sumário

1	Introdução	4
2	O experimento	7
2.1	Produção do feixe secundário ^{17}F usando o sistema TWINSOL	7
2.2	O alvo ativo pAT-TPC	10
3	Desenvolvimento de ferramentas de <i>machine learning</i> para analise de dados	16
3.1	Tipos de redes neurais	17
3.2	Estrutura da rede neural	20
3.3	Sistemas de <i>machine learning</i>	22
4	Reconstrução de nuvens de pontos a partir de algoritmos de <i>machine learning</i>	25
4.1	Construção do banco de dados para as redes neurais	26
4.1.1	Estimativa do fundo	27
4.1.2	Deconvolução do sinal	30
4.2	Análise dos pulsos com <i>machine learning</i>	32
4.2.1	Rede neural para o fundo	32
4.2.2	Rede neural para a deconvolução	35
4.2.3	Identificação de picos	38
4.2.4	Acoplando as redes neurais	41
5	Análise das nuvens de pontos	44
5.1	Identificação de trajetórias	44
5.2	Abordagens alternativas	50
5.2.1	Identificação de eventos com <i>machine learning</i>	50
5.2.2	Identificação de <i>outliers</i>	53
6	Resultados	55
6.1	O vértice da reação	55

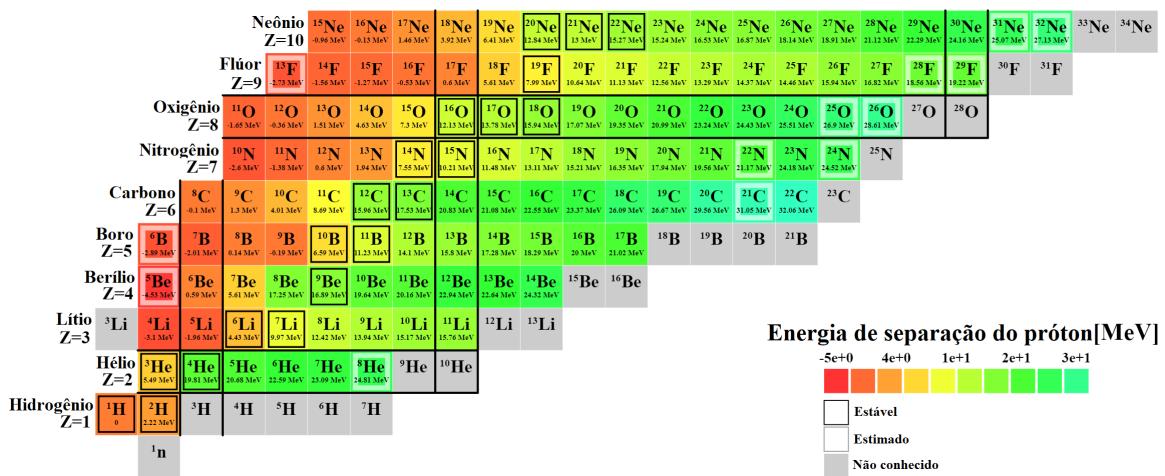
6.1.1	Reconstrução do vértice da reação	55
6.1.2	Reconstrução da energia das partículas	57
6.1.3	Determinação dos ângulos de reação	58
6.2	Identificação do canal de reação de breakup	59
6.3	Construção das distribuições angulares	65
7	Conclusão	67
A	Aplicações de machine learning na física nuclear	69
A.1	Análise de espectros para identificação de partículas (<i>particle identification</i> , PID)	69
A.2	Estimativa de raios e massas nucleares	70
A.3	Decaimento β e processo r	72
A.4	Alvos ativos	73

Capítulo 1

Introdução

Um dos principais objetivos do estudo em física nuclear é entender a estrutura do núcleo. Apesar do sucesso do modelo de camadas em explicar as estruturas de núcleos estáveis, os núcleos instáveis ou exóticos, ricos ou pobres em nêutrons, continuam sendo um grande desafio para nossa compreensão.

Os núcleos leves radioativos são de grande interesse para a astrofísica nuclear [1], onde processos de captura rápida de nêutrons e prótons (que ocorrem em eventos explosivos de novas ou supernovas) envolvem os núcleos radioativos longe da linha de estabilidade. Conhecer as propriedades desses núcleos, como massa, probabilidades de decaimento, seções de choque de captura e de breakup é fundamental para o entendimento da vários eventos astrofísicos [1, 2]. A figura 1.1 mostra parte da tabela de nuclídeos, onde é mostrado núcleos longe da linha de estabilidade, com a informação da energia de separação do próton (S_p).



Para investigar a estrutura de núcleos ricos em prótons ou em nêutrons, é possível usar reações nuclear como espalhamento elástico, transferência e breakup [4]. A reação de breakup é a reação nuclear tal que o projétil ao interagir com o alvo, se dissocia em seus constituintes [4, 5]. A reação estudada nessa dissertação é a reação de breakup ${}^4\text{He}({}^{17}\text{F}, {}^{16}\text{O}+\text{p}){}^4\text{He}$, onde o ${}^{17}\text{F}$ é um núcleo próton-halo que pode ser descrito como um núcleo de ${}^{16}\text{O}$ mais um próton fracamente ligado ($S_p = 600$ keV) [5].

Para a investigação de reação nuclear de breakup, é importante identificar todas as partículas que se quebraram, necessitando de um sistema de detecção sofisticado, como o alvo ativo. O alvo ativo é um sistema detector que funciona tanto quanto meio detector quanto alvo de reações nucleares. Com o alvo ativo, é possível determinar as trajetórias das partículas presentes antes e depois de reações nucleares, sendo de grande interesse para medir reações de breakup. Na experiência, o alvo ativo utilizado foi o gás ${}^4\text{He}$.

Nos últimos anos, câmaras de projeção no tempo com alvos ativos, *active-target time projection chambers* (ATTPCs), se tornaram importantes e relevantes para o estudo de núcleos exóticos em física nuclear [6]. Diversos projetos de TPCs foram desenvolvidos em diferentes laboratórios de física nuclear pelo mundo [7, 8, 9]. O alvo ativo permite o uso de um gás específico que serve tanto quanto alvo quanto meio detector. Isso permite a detecção de múltiplas reações nucleares ao mesmo tempo, com a capacidade de medidas completas da cinemática das reações [6, 10, 7]. Com a cinemática completa é possível identificar o mecanismo de reação e obter as seções de choque.

O sistema de detecção de TPCs, usualmente, é baseado no *Micro-MEsh GAseous Structures* (MICROMEGAS) [11] e no Gas Electron Multiplier (GEM) [12], que possui alta granularidade do plano detector, da ordem de 10^3 a 10^4 canais [6]. Experimentos que envolvem o TPC produzem enormes quantidades, onde, por exemplo, uma semana de experimento com o *Active Target - Target Projection Chamber* (AT-TPC) produz cerca de 10 TB de dados [7, 13]. Devido à grande quantidade de dados produzidos, a análise precisa ser dividida em diversas etapas, de modo que a eficiência e tempo computacional têm papel fundamental no processo. Diversos algoritmos foram desenvolvidos para a análise de diferentes etapas dos dados [6, 14], porém o uso de algoritmos de *machine learning* pode trazer vantagens significativas para a análise. Um dos objetivos desse trabalho foi o de desenvolver técnicas para a análise desses dados produzidos.

Este trabalho abordou a análise do experimento feito com o *prototype Active Target - Time Projection Chamber* (pAT-TPC) [13] usando técnicas de *machine learning*. Os algoritmos de *machine learning* viabilizaram análises mais rápidas em comparação à algoritmos mais tradicionais. O trabalhou, até a data em que a dissertação foi escrita, resultou nos artigos encontrados nas Refs. [14, 6].

A dissertação está esquematizada da seguinte forma: no capítulo 2 está feita a descrição do experimento; no capítulo 3 está feita uma breve descrição do que é *machine learning* com exemplos de aplicações em física nuclear; O capítulo 4 foi dedicado para a análise dos sinais (pulsos) gerados no experimento; No capítulo 5 está descrita a análise das nuvens de pontos reconstruídas a partir dos sinais; No capítulo 6 foi feita a construção das distribuições angulares; Por fim, no capítulo 7 está a conclusão do trabalho.

Capítulo 2

O experimento

Esse trabalho corresponde a analise de dados obtidos da experiência onde foi medida a reação de breakup para o $^{17}\text{F} + ^4\text{He}$. A experiência foi realizada em outubro de 2019 na University of Notre Dame, Estados Unidos. O feixe radioativo de ^{17}F foi obtido como sistema de produção chamado TWINSOL (TWIN SOlenids) [15] e conduzido até o alvo ativo pAT-TPC (prototype Active Target - Time Projection Chamber) [13]. Neste capítulo está descrito como o feixe radioativo ^{17}F foi obtido e como os produtos da reação de breakup foram medidos.

2.1 Produção do feixe secundário ^{17}F usando o sistema TWINSOL

O feixe radioativo ^{17}F foi produzido em voo usando o sistema TWINSOL [16, 17]. Um feixe primário de ^{16}O foi acelerado a uma energia de 70 MeV e intensidade da ordem de $200 \mu\text{A}$ ($10^{11} \sim 10^{12}$ partículas por segundo) e foi conduzido até a câmara de produção. Nessa câmara está localizada o alvo de produção que consiste de uma célula gasosa, preenchida com gás de deutério à uma pressão de 1 atm. A partir da reação do feixe primário (^{16}O) com o alvo de produção (deutério), partículas como ^{17}F e ^{17}O são produzidas a partir de reações nucleares de transferência [16, 15]. Essas partículas são selecionadas com um sistema de duplo solenoides [15, 18]. A figura 2.1 mostra o desenho do sistema TWINSOL acoplado com o pAT-TPC.

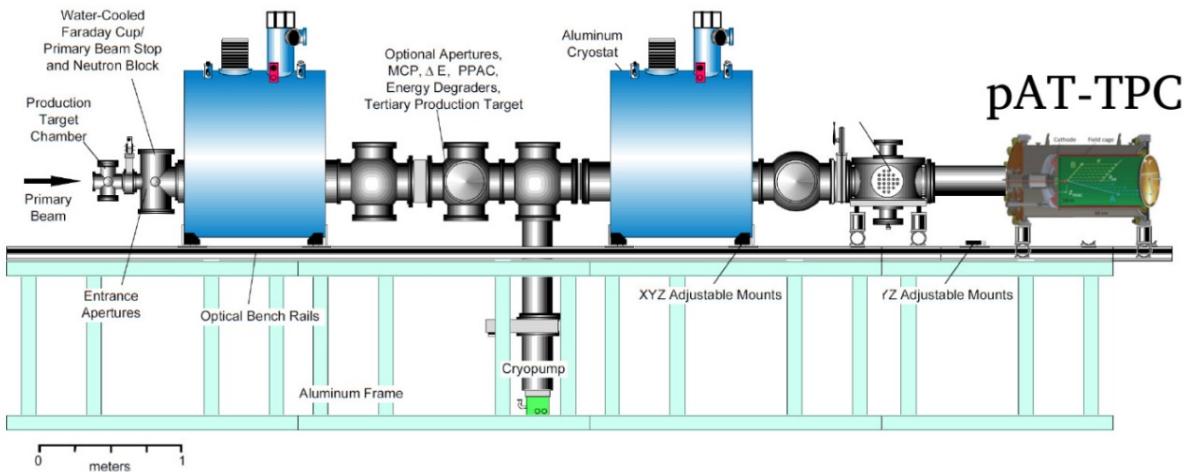


Figura 2.1: Desenho do TWINSOL à esquerda e do pAT-TPC à direita. O feixe estável de ^{16}O entra à esquerda do TWINSOL, para então ser produzido o feixe secundário ^{17}F que irá ser conduzido até o alvo ativo pAT-TPC. Todo o sistema está localizado na University of Notre Dame.

O TWINSOL é um sistema de produção de feixes radioativos em voo que possui dois solenoides supercondutores alinhados que são usados para produzir, coletar, transportar, focar e analisar feixes estáveis e radioativos. O sistema se baseia na seleção de partículas a partir da sua rigidez magnética ($B\rho$) [15, 18, 19]. Cada solenoide possui 30 cm de raio interno e 1 m de comprimento [15]. O fato de ser um solenoide finito faz com que surjam efeitos de borda na componente radial do campo magnético do solenoide, cujo efeito faz com que o solenoide seja capaz de focalizar partículas. Para entender melhor o efeito de borda no campo magnético, e consequentemente o funcionamento do TWINSOL, simulações computacionais usando a biblioteca GEANT4 [20] foram feitas usando a geometria do sistema “irmão” do TWINSOL, o Radioactive Ion Beams in Brasil (RIBRAS), que também possui dois solenoides supercondutores alinhados [18, 21]. A figura 2.2 mostra a geometria usada na simulação, onde os solenoides são de cor verde focalizando as partículas de cor laranja em um ponto do plano focal em vermelho. O campo magnético usado na simulação é função da posição do eixo, de cada solenoide, e está mostrado na figura 2.3.



Figura 2.2: Simulação computacional do sistema RIBRAS, onde as partículas carregadas em laranja surgem do ponto à direita da figura e passam pelos dois solenoides em verde, para então serem focalizada em um ponto do plano focal em vermelho. A parte em preto dos solenoides corresponde aos limites físicos da bobina [18].

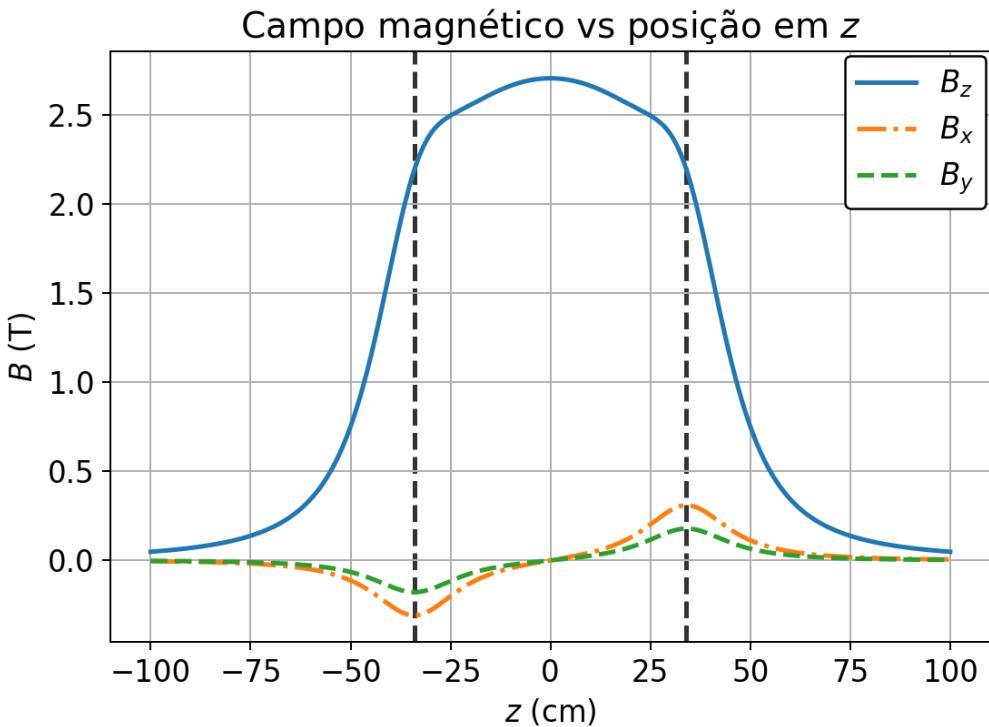


Figura 2.3: Valor do campo magnético B em tesla em função da posição em z em centímetro da bobina. A linha vertical tracejada preta indica o limite físico da bobina. O campo foi calculado à uma distância de 8 cm do eixo do solenoide. É possível ver claramente o efeito de borda que há em um solenoide finito [22].

A trajetória das partículas dentro do solenoide são helicoidais devido à força de Lorentz e possuem uma determinada frequência de ciclotrônio [18]. Além disso, cada solenoide é uma lente grossa usada para focalizar os feixes. Em uma aproximação de um solenoide como uma lente grossa, o foco depende da rigidez magnética da partícula através da relação [16, 19]:

$$\frac{1}{f} = \frac{B_z^2}{(B\rho)^2}, \quad (2.1)$$

onde f é o ponto focal, B_z a componente z do campo magnético, e $B\rho$ é dado por:

$$B\rho = \frac{mv}{q} = \frac{\sqrt{2mE}}{q}, \quad (2.2)$$

onde E é a energia, m sua massa e q seu estado de carga.

No experimento, os campos magnéticos dos solenoides foram ajustados para focalizar o feixe de ^{17}F dentro do pAT-TPC, produzindo ^{17}F com energia de 51 MeV e intensidade da ordem de $10^2 \sim 10^3$ partículas por segundo. No entanto, mesmo para partículas diferentes, o $B\rho$ pode ser muito próximo ou igual. Isso faz com que não seja possível obter um feixe de ^{17}F com 100% de pureza, e sim um coquetel de partículas [19]. O coquetel de partículas produzido possui 54% de ^{17}F , 41% de ^{16}O e cerca de 5% de ^{17}O . A figura 2.4 mostra o espectro biparamétrico de identificação de partículas obtido durante o experimento, onde é possível identificar as partículas que estão presentes no feixe (coquetel de partículas). Por fim, o feixe produzido pelo TWINSOL foi conduzido até o pAT-TPC.

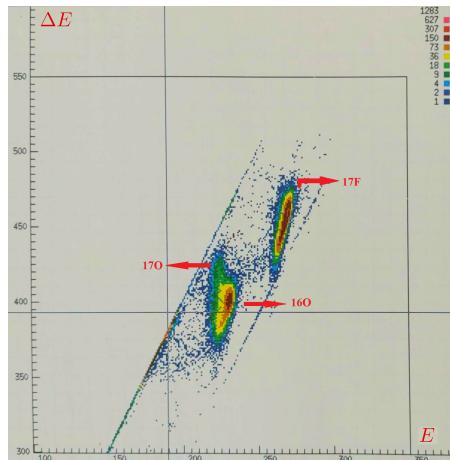


Figura 2.4: Espectro biparamétrico ΔE - E de identificação de partículas. Nele é possível identificar que o feixe possui a presença de ^{17}F , ^{16}O e uma pequena parte de ^{17}O .

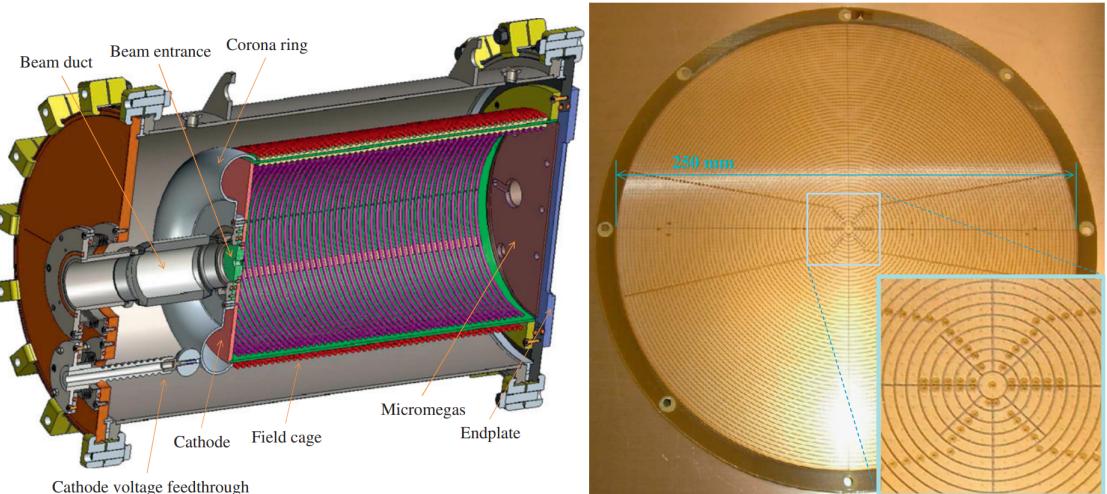
O alvo ativo foi utilizado como detector e como alvo. A seguir está descrito o funcionamento desse equipamento.

2.2 O alvo ativo pAT-TPC

Alvos ativos vêm sendo cada vez mais utilizados em experimentos para investigações de reações nucleares [6], dentre os diversos laboratórios que utilizam alvos ativos pode-

mos citar o TexAT [9] e o MAIKo [8].

A figura 2.5a mostra o desenho esquemático do pAT-TPC, que possui o alvo ativo utilizado no experimento. O detector possui uma cela cilíndrica de 50 cm de comprimento e 28 cm de diâmetro, onde o seu eixo é alinhado com o eixo do feixe [13], que passa pela câmara de íons até o duto central. Durante a passagem pela câmara de íons e pela janela (feita de PPTA - p-phenylene terephthalamide - com $12 \mu\text{m}$ de espessura) do pATTPC, o feixe ^{17}F perde energia e entra na câmara com cerca de 34.76 MeV. Nesse experimento, a câmara foi preenchida com ^4He gasoso puro à uma pressão de 350 Torr que serve tanto para alvo de reações nucleares, quanto para a própria medição e detecção dos produtos da reação [13, 23]. Tanto o feixe quanto partículas originadas da reação ionizam o gás e os elétrons que surgem dessa ionização foram conduzidos por um campo elétrico de 1 kV/cm perpendicular ao eixo da câmara até o plano detector (*pad plane*), o *micromegas* [11], mostrado na figura 2.5b.



(a) Visão transversal do pAT-TPC. O gás é preenchido dentro da cela que possui um campo elétrico perpendicular ao plano do micromegas, à direita da figura. O feixe incide na câmara entrando pelo duto de feixe à esquerda da figura.

(b) Foto do micromegas. O detector é multi-pixelado com uma maior densidade no centro, parte destacada na imagem. O *pad* central tem diâmetro de 5 mm enquanto que as faixas coaxiais possuem passo de 2mm [7, 10].

Figura 2.5: Figura esquemática do pAT-TPC e o detector micromegas[13].

O micromegas é um dispositivo de amplificação de elétrons, que consiste em um plano detector com 2048 canais (*pads*) triangulares com eletrônica independente, que usa o *Generic Electronics for TPCs* (GET) [12]. Detalhes sobre a eletrônica podem ser encontrados nas Refs. [12, 10]. O formato triangular dos canais tem como objetivo maximizar a resolução espacial do detector [7]. Cada canal possui uma posição (x, y) fixa e a terceira coordenada z será determinada a partir do tempo de deriva dos elétrons no gás [13, 23, 7,

10]. Isso só é possível pois a velocidade de deriva (drift) dos elétrons é constante [24], portanto a posição em z da partícula é diretamente proporcional ao tempo de voo. Esse princípio que deu origem ao nome de *Time Projection Chamber*, pois o evento é projetado no tempo de deriva dos elétrons no gás. A equação 2.3 (equação de Langevin) descreve o movimento de um elétron com massa m e carga e é descrito por [24]

$$m \frac{d\vec{v}}{dt} = e (\vec{E} + \vec{v} \times \vec{B}) - \frac{m}{\tau} \vec{v}, \quad (2.3)$$

onde \vec{E} é o vetor campo elétrico, \vec{B} o vetor campo magnético, \vec{v} é o vetor de velocidade do elétron e τ é o tempo de colisão médio, que depende das propriedades termodinâmicas do gás. No caso deste experimento, \vec{B} é zero e a solução estacionária para a velocidade de drift do elétron é

$$\vec{v} = \frac{\tau}{m} e \vec{E}. \quad (2.4)$$

A velocidade de deriva depende das propriedades termodinâmicas do gás (temperatura, pressão) e também de sua condutividade elétrica [24]. Isso significa que a calibração da velocidade não depende só do campo elétrico, mas também das propriedades do gás dentro do alvo ativo [13, 24]. Para acharmos a coordenada z , basta integrar a equação 2.4 para obter

$$z = \frac{\tau}{m} e \vec{E} (t - t_0), \quad (2.5)$$

onde no tempo $t_0 = 0$ o elétron está no plano do detector ($z = 0$).

O pAT-TPC conta com uma camada extra de thick gems acoplada ao detector micromegas. Thick gems usam do fato de que, no momento em que o elétron passa para uma região de campo elétrico ordens de grandeza maior que de sua origem, ocorre a ionização secundária (quando o elétron ioniza o gás). Isso provoca o que é chamado de avalanche de elétrons, amplificando a intensidade do sinal recebido [12]. A figura 2.6 mostra a esquematização do micromegas, onde na eletrônica de saída é produzido um sinal em função do tempo.

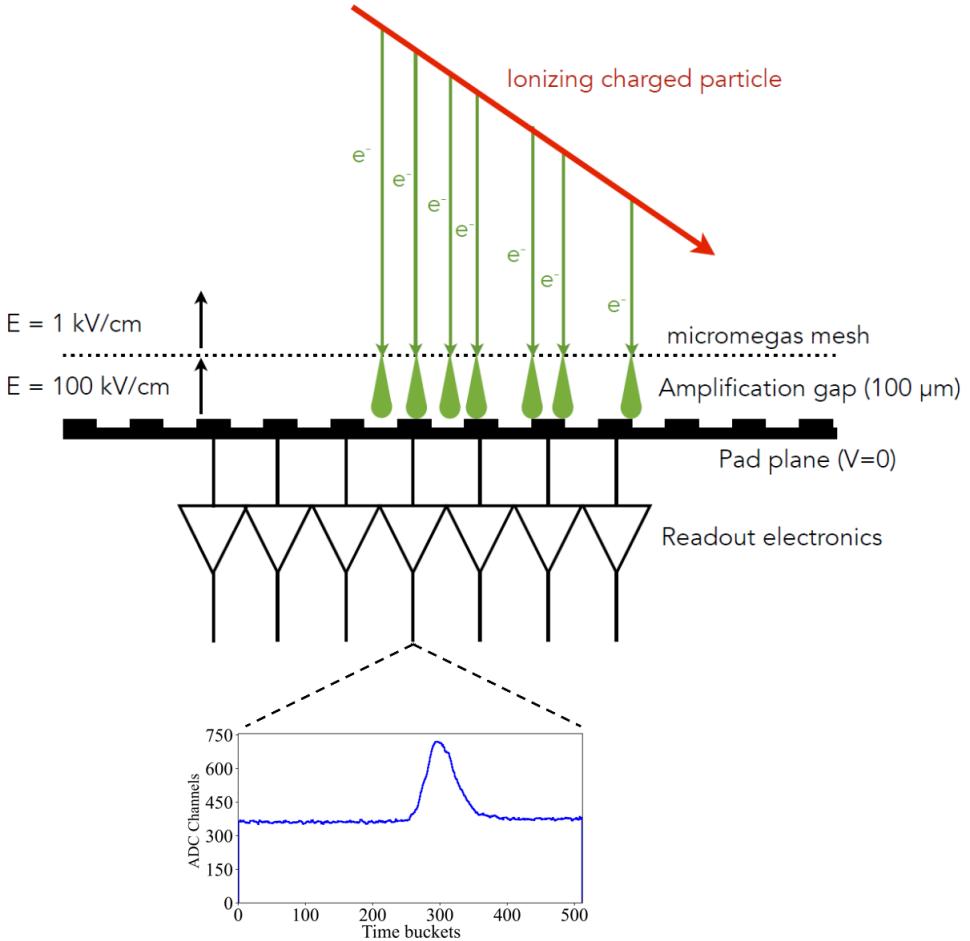


Figura 2.6: Plano do micromegas com a esquematização das thick gems. Os elétrons quando passam para um campo elétrico mais intenso ionizam o gás, produzindo ainda mais elétrons (evento chamado de avalanche de elétrons). Cada canal da eletrônica de saída produz um pulso como mostrado na parte de baixo da figura.

O sinal é discretizado no tempo levando em conta a velocidade de deriva dos elétrons, dividindo em 512 canais o tempo que o elétron leva para percorrer toda câmara do TPC [10, 13]. A velocidade de deriva do elétron no gás de ${}^4\text{He}$ puro nesse experimento variou a cada run (conjunto de eventos para serem analisados). A figura 2.7 mostra a velocidade de deriva obtida para cada run ($\sim 1 \text{ h}$) do experimento.

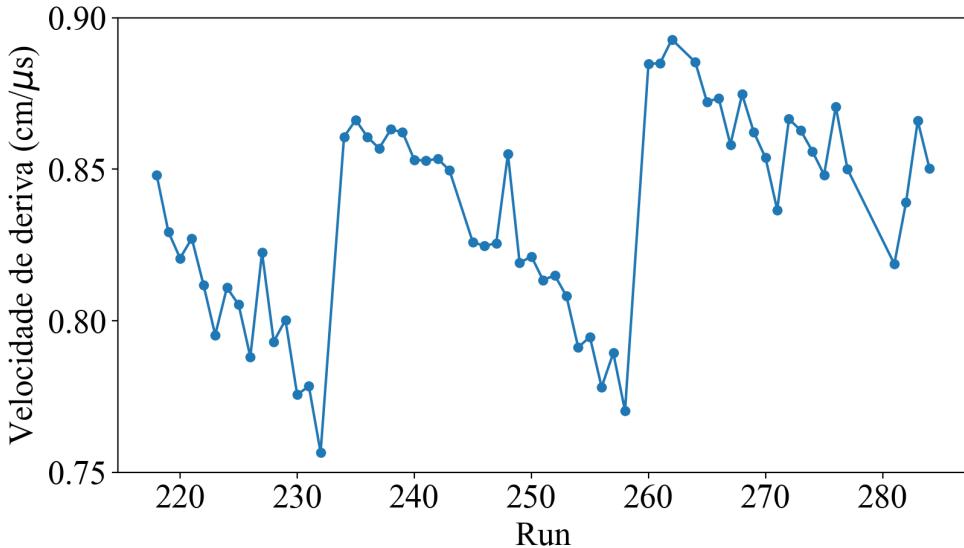


Figura 2.7: Gráfico de velocidade de deriva dos elétrons a cada run. O tempo de duração de cada run foi de aproximadamente uma hora.

Pela figura 2.7 é perceptível que a velocidade de deriva dos elétrons no gás teve uma variação por volta do 10% durante o experimento. A queda na velocidade ocorre, por exemplo, por problemas relacionados à impurezas no gás devido à filtragens de ar dentro do gás (ou outgassing) de alguns materiais dentro do alvo. Foi necessário corrigir os dados experimentais run por run a partir desses valores de velocidade de deriva.

Para a run 218, a velocidade de deriva dos elétrons foi aproximadamente de $0.85 \text{ cm}/\mu\text{s}$. Isso significa que os elétrons produzidos próximo na janela de entrada do detector percorrem os 50 cm da câmara em cerca de $59 \mu\text{s}$. Dividindo esse tempo pelos 512 canais tem-se que cada canal (time bucket) possui cerca de 115 ns de largura.

Cada interação de uma partícula carregada com o gás é detectada como um pulso eletrônico, onde o centroide corresponde ao tempo de deriva. A carga acumulada Q dessa interação é a área do pulso associado ao centroide. Cada centroide então representa um ponto no espaço (x, y, t, Q) . Um exemplo de evento reconstruído está na figura 2.8.

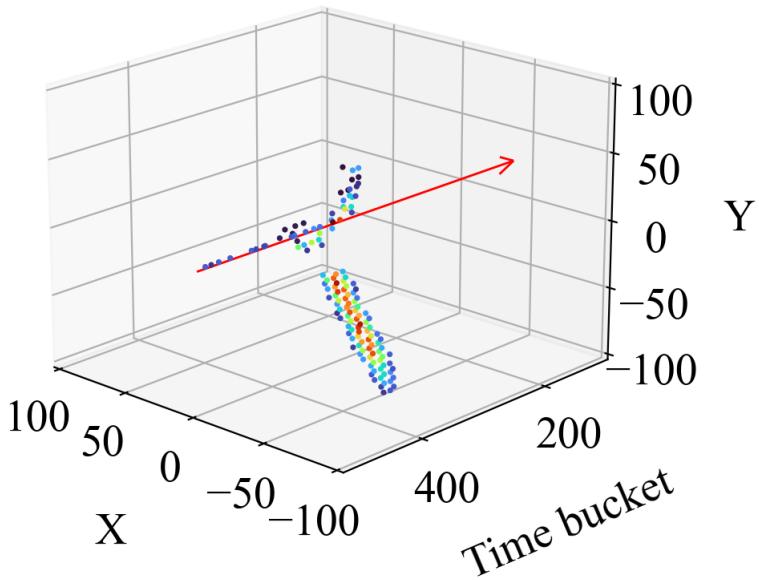


Figura 2.8: Evento reconstruído a partir da análise dos pulsos gerados pelo micromegas. A cor representa a carga integrada de cada ponto de interação com o gás.

Para reconstituir eventos como o da figura 2.8, foram analisados cerca de 300 pulsos. No entanto, o número de evento reconstruídos é da ordem de milhões, portanto a quantidade de sinais que precisam ser analisados é muito grande, em comparação com experimentos em física nuclear com apenas alguns canais de detecção, o que gera a necessidade de desenvolvimento de algoritmos com alta eficiência em tempo para que a análise não demande muito tempo. Para a análise completa do experimento foram seguidas as seguintes etapas:

- Análise dos pulsos de cada interação das partículas com o gás. Isso envolve remover o fundo, localizar os picos e obter os tempos e carga integrada de cada caso;
- Reconstruir eventos em 3D (nuvens de pontos) a partir da análise de sinais. As nuvens de pontos precisam ser analisadas com algoritmos de reconhecimento de padrões que permitem ajustar as trajetórias das partículas em 3D;
- Reconstruir a cinemática das partículas com as trajetórias e energia depositada no gás. Isto permite obter as distribuições angulares.

A análise dos pulsos, reconstituição de eventos, reconstrução da cinemática, gráficos das distribuições angulares e resultados são apresentadas nos próximos capítulos.

Capítulo 3

Desenvolvimento de ferramentas de *machine learning* para análise de dados

Para que possamos analisar a grande quantidade de dados gerados em experimentos com alvo ativo podemos utilizar técnicas de *machine learning*. Nesse capítulo está explicada a metodologia usada para a implementação dessas ferramentas. O apêndice A mostra algumas aplicações de *machine learning* na física nuclear.

Machine learning é a área de estudo que desenvolve algoritmos para que eles possam aprender com os dados, sem serem explicitamente programados para isso [25]. Supõe-se que uma rede neural imite um sistema biológico, em que os neurônios interajam enviando sinais na forma de funções matemáticas entre as camadas. Isso inspirou o uso do modelo matemático simples de uma função linear nos parâmetros para um neurônio artificial [26]:

$$y = f \left(\sum_{i=1}^n \omega_i x_i + b_i \right) = f(z), \quad (3.1)$$

onde y é a saída do neurônio, que corresponde à função de ativação f que depende da soma ponderada, onde o peso é ω_i , das entradas x_i dos outros n neurônios. O termo b_i corresponde ao parâmetro *bias*. A ideia é fazer um neurônio receber a informação de todos os outros neurônios da camada anterior, fazendo uma média ponderada (onde o peso que será estimado pelo algoritmo de *machine learning*) e somando com um termo independente (*bias*, que também é estimado). Os parâmetros ω_i e b_i serão estimados através de um determinado procedimento, chamado de minimização (o treino da rede neural).

3.1 Tipos de redes neurais

Uma rede neural artificial, *Artificial Neural Network* (ANN), é um modelo computacional que consiste de camadas de neurônios. ANNs foram desenvolvidas para o estudo de inteligência artificial[25, 27]. ANNs consistem principalmente numa camada de entrada (*input layer*), uma camada de saída (*output layer*) e eventuais camadas entre essas duas, chamadas de camadas ocultas (*hidden layers*). Os tipos mais comuns são:

Feed-Forward Neural Networks

A *Feed-forward neural networks* (FFNN) é a primeira e mais simples rede neural desenvolvida [28, 29]. Nessa rede a informação se move apenas para frente através de camadas (da camada de entrada até a camada de saída). A figura 3.1 mostra uma representação de rede, onde os neurônios são representados por círculos, enquanto que as linhas mostram as conexões entre os neurônios. Cada neurônio recebe informação de todos os neurônios da camada anterior, portanto a rede é chamada de totalmente conectada, *fully-connected* (FC), FFNN.

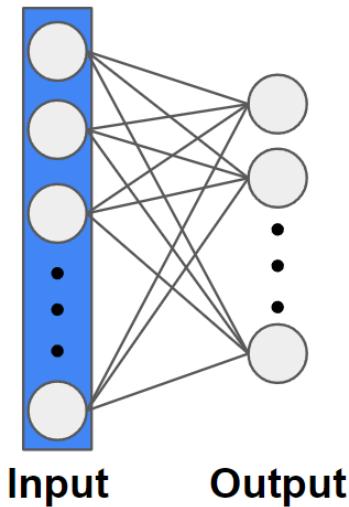


Figura 3.1: Exemplo de FFNN. A camada de entrada na esquerda propaga a informação para a direita (camada de saída). Todos os neurônios entre camadas estão conectados entre si.

Convolutional Neural Network

Uma variante da FFNN é a chamada de rede neural convolucional, *convolutional neural network* (CNN). Do ponto de vista matemático sobre convoluções, a convolução descrita como $(f * g)(t)$ de uma função $f(t)$ e outra $g(t)$ é definida como:

$$(f * g)(t) \equiv \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau. \quad (3.2)$$

Para o caso discreto, com g sendo uma função resposta finita de tamanho $2M$, temos

$$(f * g)[n] = \sum_{m=-M}^{M} f[n-m]g[m]. \quad (3.3)$$

Convoluçãoções são invariantes sobre rotação e translação, portanto são muito utilizadas para processamento de sinais e imagens [30]. Para a convolução discreta se escolhe um filtro que irá atuar no vetor desejado. Para ilustrar o que significa isso, no caso discreto e unidimensional, a figura 3.2 mostra o processo de convolução de um vetor de tamanho 9 e um filtro de tamanho 3.

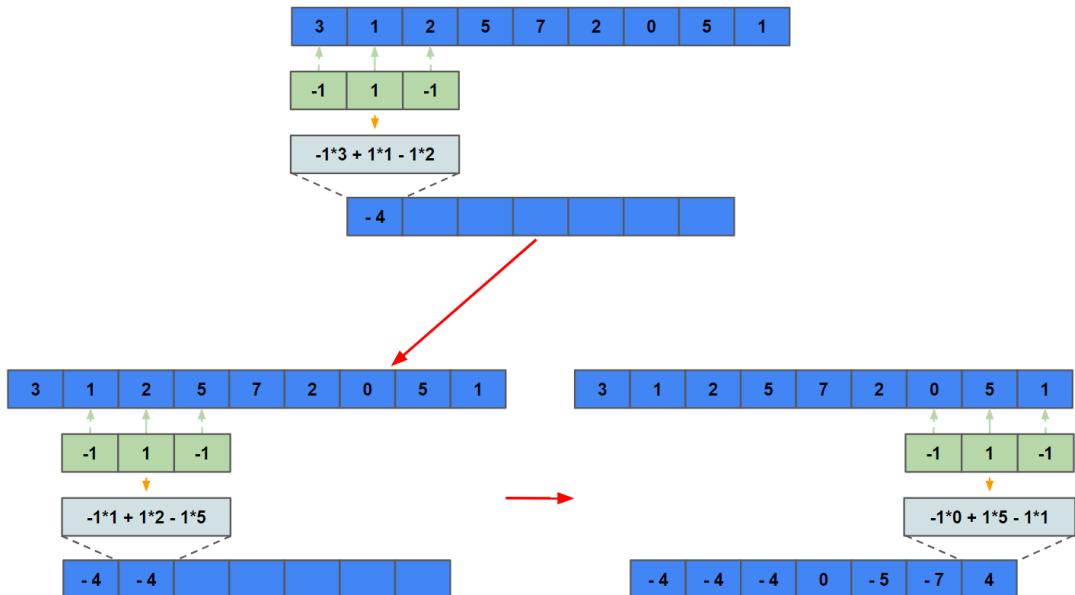


Figura 3.2: Processo de convolução entre sinal azul em cima e o filtro em verde, resultando no sinal azul embaixo. A multiplicação é feita ponto a ponto e está indicada na caixa azul-clara.

Percebe-se que o sinal resultante tem dimensão menor que o sinal original. O filtro (também chamado de *kernel*) atua em um ponto que possua vizinhos o suficiente para o restante do filtro poder fazer a multiplicação ponto a ponto. Esse tipo de convolução tem o chamado emparelhamento válido (*valid padding*). O tamanho n_2 resultante do vetor de saída é

$$n_2 = n_1 - m + 1, \quad (3.4)$$

onde n_1 é o tamanho do vetor de entrada e m o tamanho do filtro (*kernel size*). Para que o vetor de saída tenha o mesmo tamanho do vetor de entrada, são acrescentados zeros

em torno da entrada, de forma que a saída tenha o mesmo tamanho da entrada. Esse é o chamado emparelhamento igual (*same padding*). A figura 3.3 ilustra esse processo.

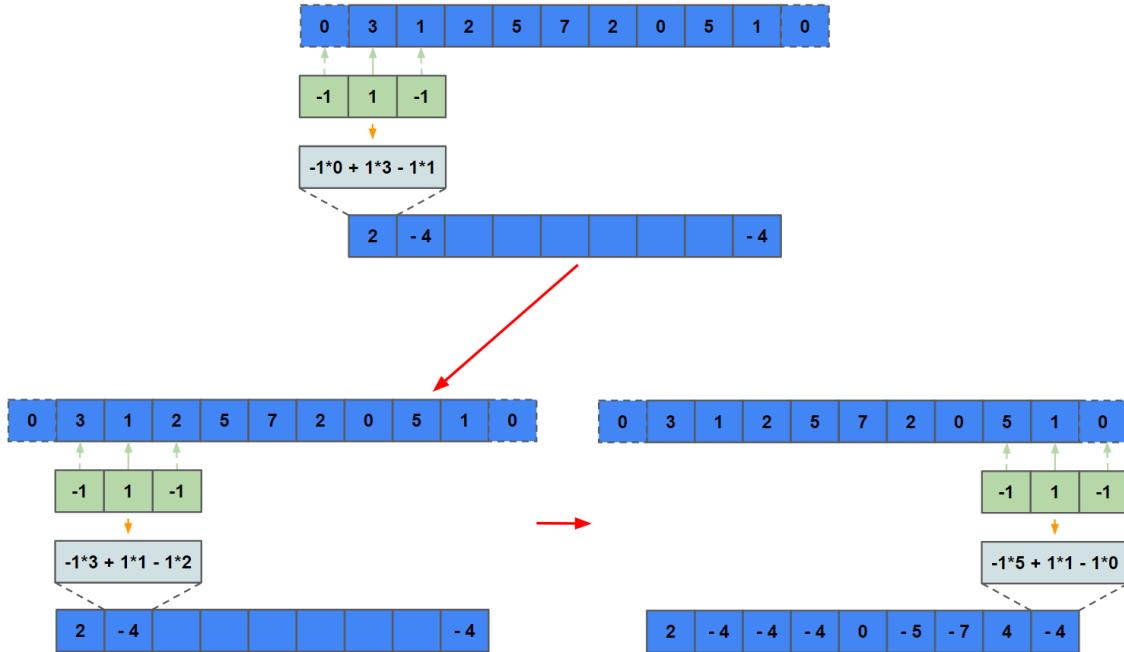


Figura 3.3: Processo de convolução entre o sinal azul em cima e o filtro em verde, resultando no sinal azul embaixo. Agora são acrescentados zeros no inicio e no final do vetor para que o vetor saída tenha o mesmo tamanho do vetor de entrada (nesse caso 9).

Como estamos no contexto de *machine learning* (inteligência artificial), *a priori* não sabemos quais os valores dos filtros que devem ser aplicados, apenas seus tamanhos e como agem. A ideia é estimar os valores do filtro (através do treino da rede neural) que deve ser aplicado para se obter o resultado desejado.

Cada filtro aplicado gera um mapa característico (*feature map*), que é o resultado da atuação do filtro em um vetor. Usualmente, em uma CNN se escolhe o tamanho do filtro, *padding* (*valid* ou *same*) e quantos filtros serão aplicados (para saber quantos *feature maps* serão gerados). Como temos vários mapas gerados por cada filtro, isso acarreta em um aumento de dimensionalidade. Para filtrar/selecionar os mapas é usado um critério, como por exemplo selecionar valores máximos dos mapas gerados dada uma janela de atuação (quantos mapas serão comparados para selecionar o máximo valor). O *Max-Pooling* faz isso, selecionando valores máximos para uma determinada quantidade de mapas sendo comparados (*pool size*).

Existem outros tipos de redes neurais que não estão discutidas aqui, porém podem ser encontradas nas referências [31, 32].

3.2 Estrutura da rede neural

Para a construção de uma rede neural (nesse caso em específico de uma rede neural supervisionada, que está discutida mais para frente), é preciso definir sua estrutura. A estrutura da rede neural possui camadas e cada camada pode possuir uma função de ativação. No geral, tanto o *input* quanto o *output* possuem dimensão fixa. A rede é treinada minimizando uma função custo, otimizando os parâmetros da rede através de um otimizador.

Para cada camada da arquitetura devemos escolher sua função de ativação. Tanto FNNNs quanto CNNs podem possuir funções de ativação (função f da equação 3.1). Dentre muitas funções de ativação podemos citar a *Rectified Linear Units* (ReLU)[33], sigmoide [34], linear e tangente hiperbólica[35]. A figura 3.4 mostra os gráficos dessas funções de ativação, onde no eixo x é o argumento e no eixo y o resultado da função.

Funções de ativação

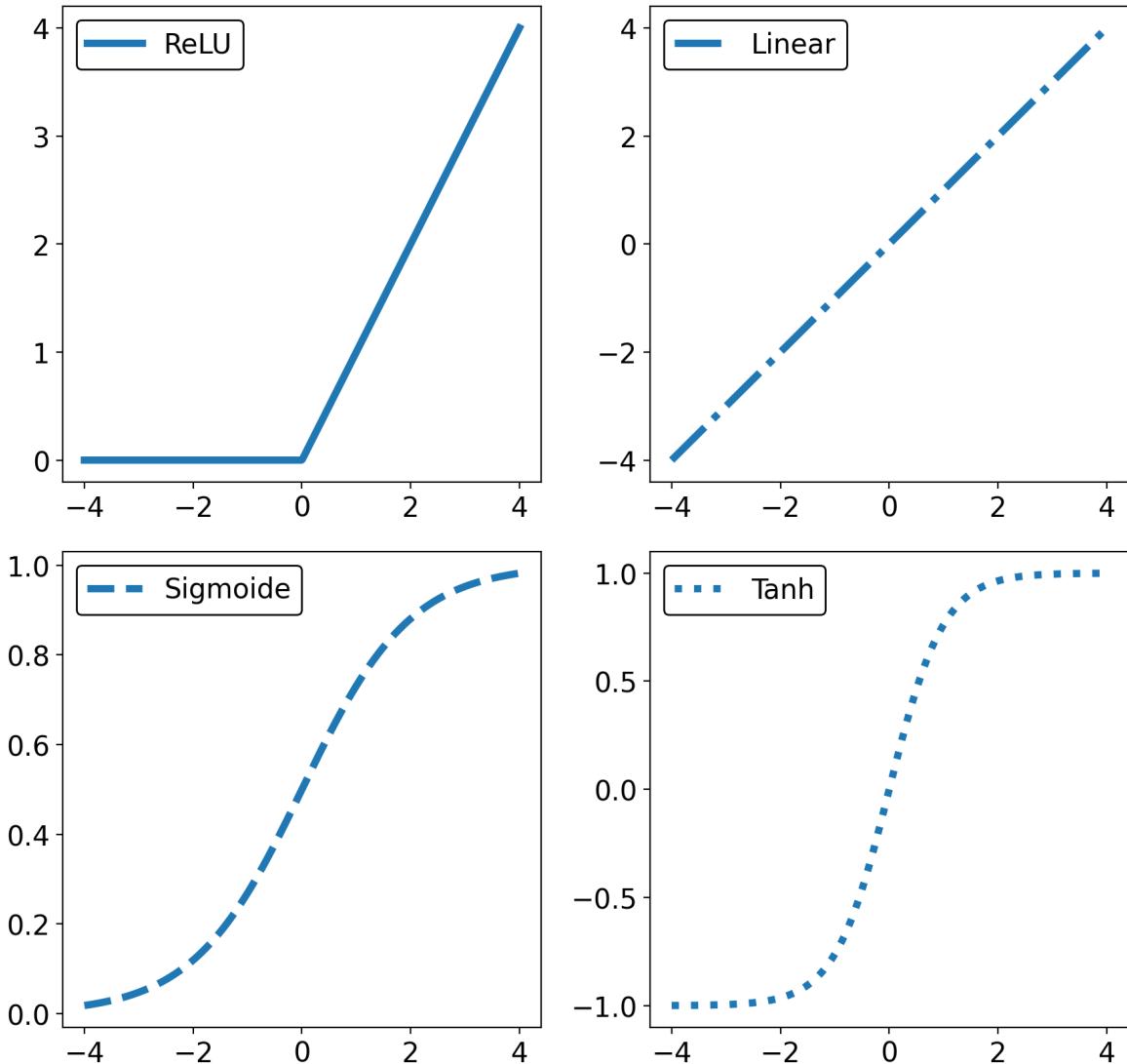


Figura 3.4: Funções de ativação e seus respectivos gráficos.

O próximo passo é definir a função custo (também chamada de *loss*) e o otimizador. A função custo tem o papel de retornar valores altos para previsões erradas e valores baixos para previsões corretas. Por exemplo, se queremos treinar uma rede neural para classificação binária (que prevê duas saídas possíveis), devemos usar a função custo chamada de *binary cross-entropy* dada por[36]

$$C(p(y_i)) = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)), \quad (3.5)$$

onde y_i é o rótulo (*label*), $p(y_i)$ é a probabilidade do ponto y_i ser 1 e N é o número de pontos. O objetivo da rede neural é achar o mínimo da função $C(p(y_i))$, o que implica

diretamente na melhor solução para o conjunto de dados. Isso é feito pelo método de retropropagação do erro (*backpropagation* [37]) por um otimizador. Outros exemplos de *loss* são o erro quadrático médio ou *categorical cross-entropy* [38].

O otimizador tem o objetivo de otimizar os parâmetros presentes na rede neural, buscando o mínimo global da função custo, o que nem sempre acontece, pois a minimização pode parar em um mínimo local da função. Existem diversos otimizadores, como por exemplo o *Stochastic Gradient Descent* (SGD), ADAM, ADAMAX [39], entre outros [40]. Para o SGD, temos que a atualização de parâmetros é dada por

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} C(\theta), \quad (3.6)$$

onde θ_j é o parâmetro a ser atualizado, α é a *learning rate* e $C(\theta)$ é a *loss* que depende dos parâmetros θ .

Para enfim treinar a rede neural, se escolhe o *batch size*, que é o tamanho de amostras que será usada para o treino, por iteração em cada rodada de treino (*epoch*). Por exemplo, se usamos 1000 dados para o treino, e o *batch size* é 500, cada *epoch* terá duas iterações. No geral, se usam *batch sizes* pequenos, pois o consumo de memória é mais eficiente.

Para avaliação do modelo se usam dados de validação, que servem para verificar o comportamento da rede neural que está sendo treinada. Esse conjunto de dados usados não é usado para o treino, são usados apenas para verificar possíveis problemas como o *overfit*. O *overfit* ocorre quando a rede neural começa a se adequar perfeitamente aos dados de treino, perdendo a capacidade de previsão em dados que não estão sendo usados no treino pela rede neural.

Além dos dados de validação, podemos escolher métricas que auxiliam a visualização do comportamento da rede neural durante o treino e nos retornam informações importantes sobre sua qualidade. Exemplos importantes de métricas são: acurácia binária, erro médio absoluto e acurácia categórica [41]. Por exemplo, caso seja necessário verificar se uma rede neural está fazendo previsões certas em um problema cuja classificação é binária, então a métrica deve ser a acurácia binária. Tudo depende do objetivo da rede neural.

3.3 Sistemas de *machine learning*

Podemos dividir os sistemas de *machine learning* em quatro tipos:

Aprendizado Supervisionado

Aprendizado supervisionado é quando fornecemos para a rede neural um conjunto de dados para o treino com a solução desejada (chamados de *labels*). Um uso típico é para problemas de classificação [25]. Por exemplo, classificação de imagens (identificação de figuras), previsão de valores numéricos etc. Exemplos de algoritmos supervisionados são:

- *k-Nearest Neighbors* [42]
- Regressão linear
- *Support Vector Machines* (SVMs)
- *Decision Trees and Random Forests*
- Redes neurais

Aprendizado não supervisionado

Aprendizado não supervisionado é quando fornecemos o conjunto de dados para o treino, porém sem solução. A ideia é aprender sem supervisão. Um problema comum, por exemplo, é quando queremos identificar *clusters* em um conjunto de dados (*clustering*) [43, 44].

Aprendizado semi supervisionado

Aprendizado supervisionado é quando apenas parte do conjunto de dados para o treino possui *labels*. Isso é comum quando se obtém conjuntos de dados diferentes e apenas parte deles foi classificado [45].

Aprendizado por reforço

Aprendizado por reforço é quando um sistema, chamado de *agente* nesse contexto, aprende através do ambiente, realizando ações que maximizam sua recompensa. Por exemplo, caso o sistema realize uma ação incorreta, ele recebe uma penalidade, fazendo com que procure outra maneira de realizar a ação, dessa vez de maneira correta, para poder ganhar uma recompensa [46]. Esse tipo de sistema é muito usado, por exemplo, em automatização robótica, como carros que pilotam sozinhos, robôs que aprendem a andar etc [47].

Com essa breve descrição sobre *machine learning* e exemplos do seu uso em problemas de física nuclear (apêndica A), pode-se seguir adiante e entender seu uso dentro deste trabalho, que está feito nos próximos capítulos.

Capítulo 4

Reconstrução de nuvens de pontos a partir de algoritmos de *machine learning*

Esse capítulo descreve o procedimento usado para criar as nuvens de pontos (*point-clouds*) a partir dos pulsos gerados por cada pixel do micromegas, usando algoritmos de *machine learning* supervisionado.

Algoritmos baseados em CNNs vem sendo usados com relativo sucesso para o processamento e análise de sinais [6], como por exemplo para discriminação de pulsos [48]. CNNs possuem a capacidade de fazer ajustes multidimensionais e aprender padrões complexos. Além disso, CNNs são mais eficientes em tempo para a análise de grandes quantidades de dados em comparação com algoritmos comuns [6].

A quantidade de pulsos gerados nesse experimento é da ordem de centenas de milhões. Para a análise completa dos pulsos, são necessárias diferentes etapas que envolvem processos que muitas vezes precisam ser refeitos por causa de algum erro no processo de análise. Notar algum erro no processo, ou ter que mudar algum parâmetro da análise, significaria ter que analisar novamente os dados, sendo um processo muito custoso em tempo.

O uso de algoritmos de CNNs tem o objetivo de diminuir o tempo consumido para a análise dos pulsos. As redes neurais criadas são treinadas uma única vez e podem ser utilizadas de modo separado e/ou acoplado [6], como está mostrado na sequencia do capítulo.

A análise dos dados experimentais foi dividida em três etapas: correção do fundo, deconvolução do sinal e identificação de picos. As redes neurais criadas foram supervisionadas, o que significa que foi preciso determinar a entrada (*input*) e a saída (*output*)

das redes. Para isso foi feito um grande banco de dados, para utilizar no treinamento das redes neurais, que se beneficiam da grande quantidade de dados usada para o treino [25].

Na seção 4.1, está mostrado como o banco de dados para o treino das redes neurais desenvolvidas foi criado, e a construção das redes neurais está na seção 4.2.

4.1 Construção do banco de dados para as redes neurais

Cada um dos pixels do detector micromegas possui uma eletrônica independente que gera pulso digitais como os mostrados na 4.1. Centenas de sinais coo essas são produzidas em cada um dos eventos.

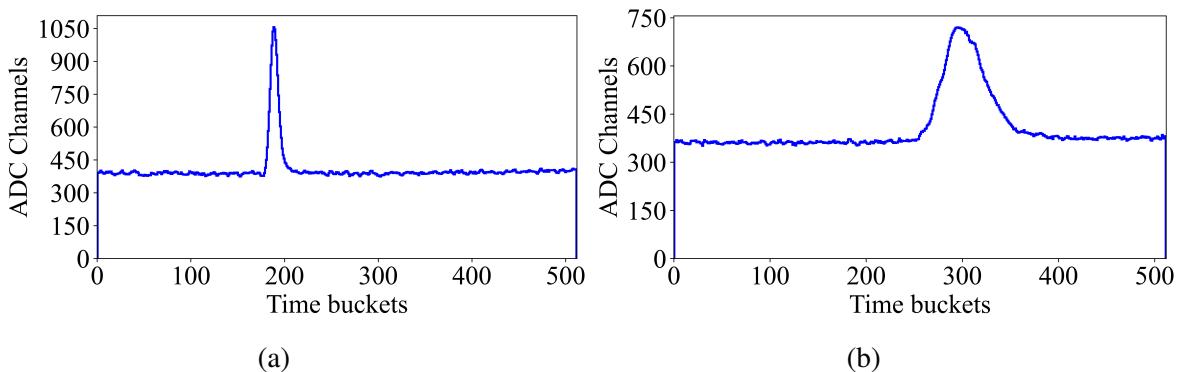


Figura 4.1: Exemplos de sinais produzidos pelos canais do detector. Em 4.1a o sinal possui apenas um pulso, enquanto em 4.1b há vários pulsos em sobreposição, formando um único pulso com largura maior que em 4.1a.

No eixo x , cada um dos 512 *time buckets* possui largura de 195 ns. No eixo y tem-se a carga acumulada no detector para cada *time bucket*. Na figura 4.1a há um sinal com um pedestal (fundo ou *baseline*) com altura entre 300 e 450, e um pulso estreito em cima. Como mostrado na seção 2, os elétrons que surgiram da ionização do gás foram conduzidos perpendicularmente pelo campo elétrico até o detector. A interação da partícula com o gás é evidenciada justamente pelo pulso presente em 4.1a. Cada pixel i do detector está em uma posição (x_i, y_i) , o centroide de cada gaussiana fornece a coordenada em t (*time bucket*) para então ser convertida na posição em coordenada z do ponto de interação da partícula com o gás, e a energia depositada Q é obtida da área do pulso sem fundo (gaussiana com centroide t).

Para a figura 4.1a tem-se apenas um pulso estreito, o que corresponde à um feixe incidindo paralelamente àquele canal do plano detector (perpendicular ao campo elétrico), pois a projeção da interação da partícula com gás no tempo é uma distribuição estreita. No caso da figura 4.1b, há uma distribuição ampla do sinal do tempo, o que corresponde

ao feixe incidindo perpendicularmente ao canal do plano detector. A ilustração desse processo está na figura 4.2, que mostra o processo da passagem de uma partícula carregada e como o sinal é gerado a partir disso.

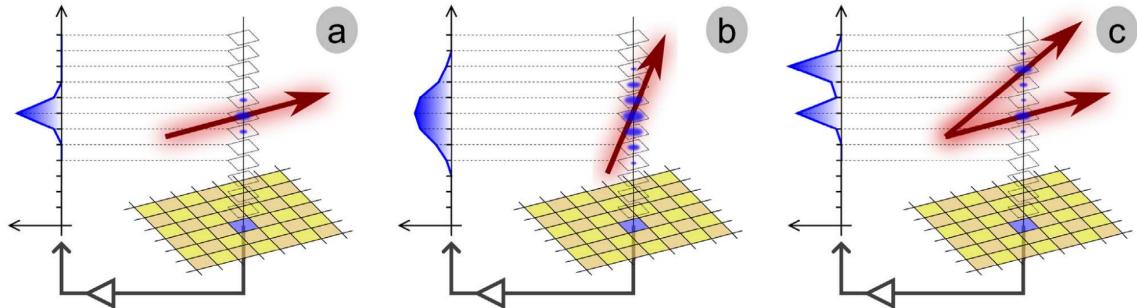


Figura 4.2: Ilustração que mostra a variação no formato da carga coletada a partir da passagem de uma partícula carregada dentro do TPC, onde o plano do detector está embaixo. No lado esquerdo de cada imagem, a distribuição do sinal coletado por um único pad (escuro) do plano de coleta é mostrado (o canal eletrônico de leitura é representado pela seta cinza em negrito). No caso de uma trajetória quase horizontal em relação ao plano do detector (a), o sinal é uma distribuição estreita, enquanto para uma trajetória próxima a uma direção vertical (ou perpendicular) em relação ao detector (b), a distribuição deve ser muito mais ampla (vários pontos de interação da partícula com o gás devem ser extraídos desse sinal). A última imagem ilustra o caso em mais de uma trajetória de partículas contribui para o sinal [12].

Para analisar os pulsos deve-se primeiro remover o fundo (pedestal ou *baseline*) dos sinais. O sinal de fundo é complexo e pode variar por canal e também por evento. Desde flutuações causadas pelo circuito eletrônico até efeitos sistemáticos gerados pela memória de *buffer* circular alteram o sinal, podendo o tornar não analítico [6, 12]. Com o sinal sem o fundo, se faz a deconvolução para determinar todos os centroides e cargas acumuladas dos pontos, para a reconstrução da nuvem de pontos. Todo esse processo foi realizado nesse trabalho com algoritmos de *machine learning* supervisionado[6]. Para isso, foi criado um banco de dados que serviu de *output* e/ou *input* para o treino das redes neurais.

A criação dos dados para o treino das redes neural está mostrado na seção 4.1.1. A criação das redes neurais está mostrada na seção 4.2.

4.1.1 Estimativa do fundo

Essa subseção descreve a estimativa do fundo de cada, para formar o banco de dados da rede neural que estima o fundo descrita na subseção 4.2.1.

A primeira tentativa de estimar o sinal sem o fundo é usando transformada de Fourier e um filtro passa-baixa, que é a função resposta do detector fornecida na Ref. [12]. Seja

$f(t)$ uma função qualquer, sua transformada de Fourier é dada por

$$\hat{f}(\nu) = \mathcal{F}[f(t)] = \int_{-\infty}^{\infty} f(t)e^{-2\pi i \nu t} dt. \quad (4.1)$$

Primeiro calculamos a transformada de Fourier $\hat{f}(\nu)$ do sinal, em seguida multiplicamos pela função resposta do detector $h(\nu)$ dada por [12]

$$h(\nu) = A * \exp(\nu\tau) (\nu\tau)^3 \sin(\nu\tau), \quad (4.2)$$

onde A está relacionado com o ganho de amplificação e τ é o tempo de pico (*peaking time*), que é o tempo de modelagem da cadeia de amplificação [12]. Do teorema da convolução, temos que [49]

$$\mathcal{F}^{-1}[\hat{f}(\nu)\hat{g}(\nu)] = (f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau, \quad (4.3)$$

onde $(f * g)(t)$ é a convolução entre $f(t)$ e $g(t)$. Multiplicar o sinal transformado por $h(\nu)$ e depois inverter a transformação é o mesmo que convoluir o sinal original com a transformação inversa de $h(\nu)$, o que resulta no sinal sem o fundo [10, 12]. Resultados desse procedimento estão na figura 4.3.

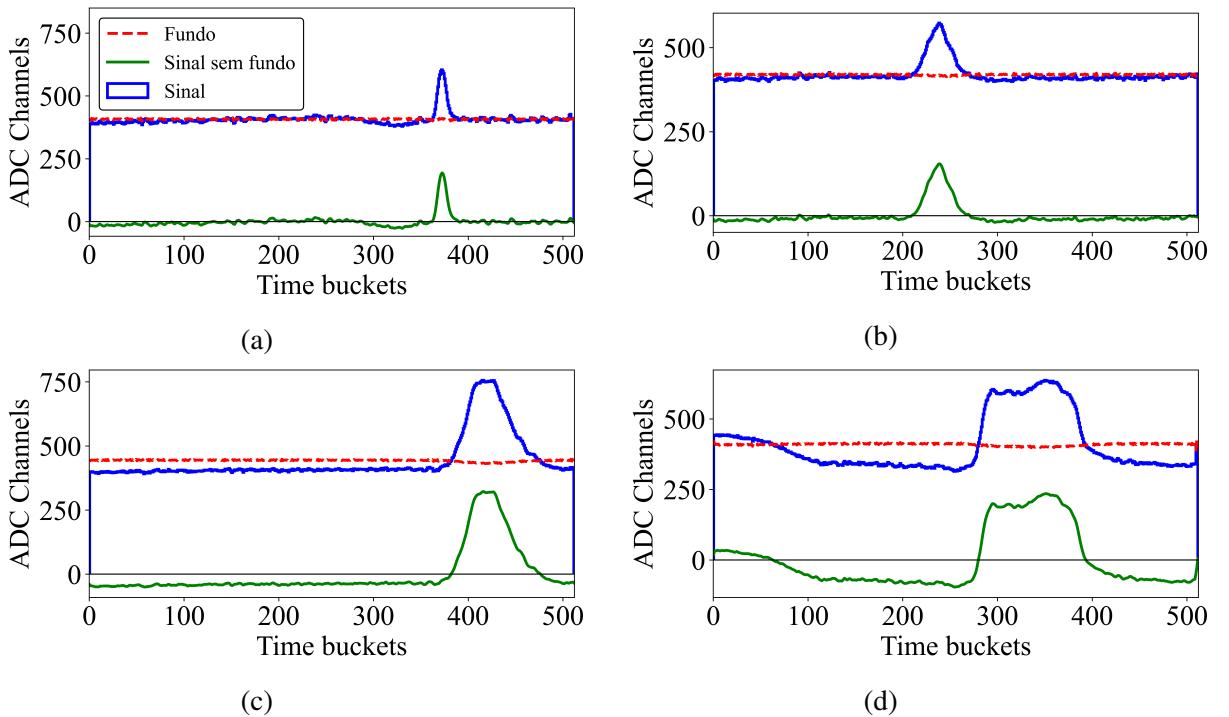


Figura 4.3: Histogramas com as respectivas *baselines* (linhas tracejadas) estimadas pelo método da convolução. O espectro resultante (sem o fundo) está em verde.

Fica claro que visualmente, por exemplo na figura 4.3d, que o filtro utilizado não é a melhor função resposta do detector. Poderia-se estimar essa função resposta empiricamente, porém os canais auxiliares chamados de *Fixed Pattern Noise* (FPN) [12] usados para esta estimativa não foram armazenados. Portanto, a estimativa do fundo foi feita sinal por sinal[6, 12]. Para isso, o fundo foi determinado usando o algoritmo *background removal* da biblioteca *TSpectrum* do *ROOT* [50]. A função tem a capacidade de separar o fundo dos picos presentes no espectro [51, 52, 53]. Exemplos de estimativa do fundo estão na figura 4.4.

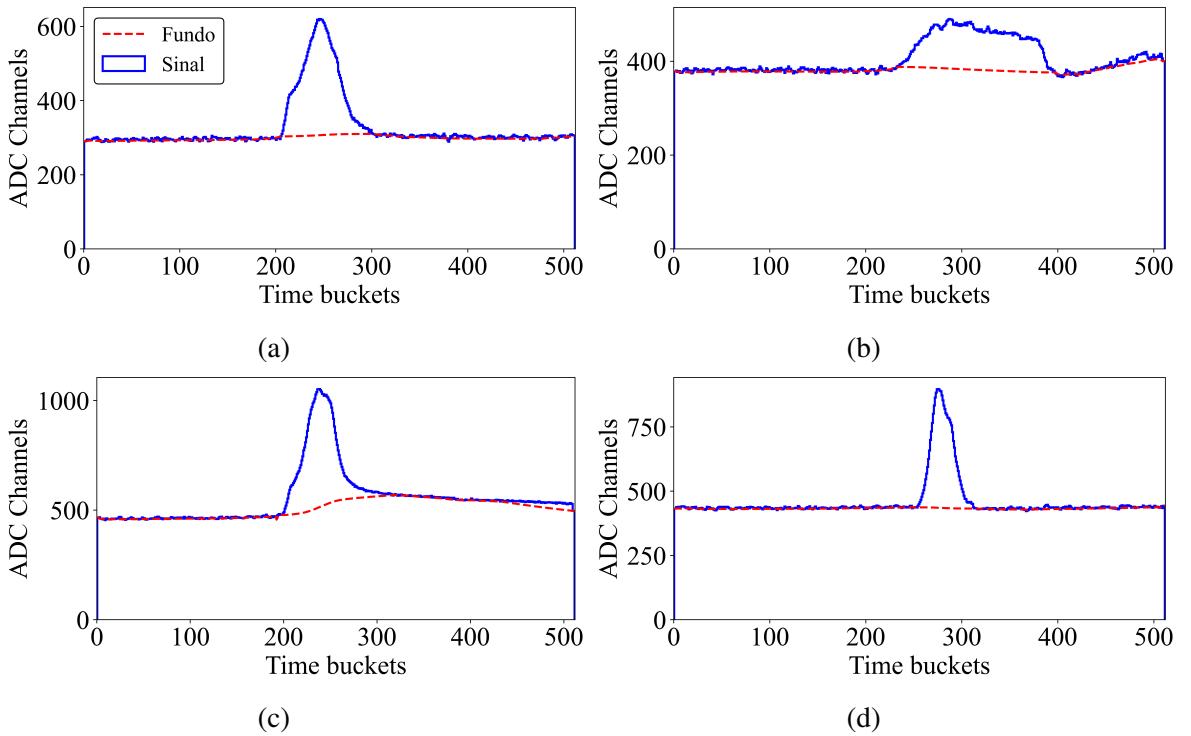


Figura 4.4: Histogramas com as respectivas *baselines* (linhas tracejadas) calculadas pelo *TSpectrum*.

Os resultados do cálculo do fundo de cada sinal foram armazenados e usados para a etapa de deconvolução do sinal, que deve ser feita com o sinal sem o fundo. Para diminuir as flutuações numéricas após a remoção do fundo, o valor mínimo do sinal sem o fundo é zero. Sem o fundo podemos buscar por todos os picos e suas cargas correspondentes no sinal. Não podemos identificar diretamente todos os picos pois muitos deles estão em sobreposição. Para isso foi feita a deconvolução do sinal, descrita na subseção 4.1.2.

4.1.2 Deconvolução do sinal

Para aumentar a resolução dos picos foi usado o algoritmo *gold deconvolution* presente na biblioteca *TSpectrum* do *ROOT* [54]. O algoritmo tem como objetivo fazer a deconvolução do espectro, gerando uma função (nesse caso um sinal) resposta de acordo com o sigma esperado para os pulsos. O sinal resposta corresponde ao espectro com as gaussianas não sobrepostas. Isso significa que foi necessário determinar qual o valor de sigma dos pulsos para buscar a função resposta.

A largura dos pulsos é o mesmo de um sinal que possui apenas um pico. Ou seja, a largura dos pulsos foi determinada fazendo a análise de sinais que possuem apenas 1 pico, fazendo um ajuste pelo método dos mínimos quadrados (MMQ) de uma gaussiana. Para buscar espectros com apenas um pico, foi usado o algoritmo de identificação de picos *peak_finder*, presente na biblioteca do *scipy* [55], e para o ajuste da gaussiana foi usada o pacote *lmfit* [56]. O valor de sigma encontrado foi de 4.09 (17) *time buckets*.

A largura escolhida foi ligeiramente maior pois, verificando empiricamente, em alguns casos o algoritmo separava o que deveria ser uma única gaussiana em duas. O valor de sigma usado na deconvolução foi de 4.30 *time buckets*. Foi determinado também o número de iterações do algoritmo de deconvolução. O número de iterações escolhido foi de 700, menos que isso o algoritmo não estava separando totalmente picos sobrepostos. O limiar para a escolha de um ponto como um pico foi definido como ter altura maior que 20% do valor máximo do sinal. Resultados da deconvolução estão na figura 4.5.

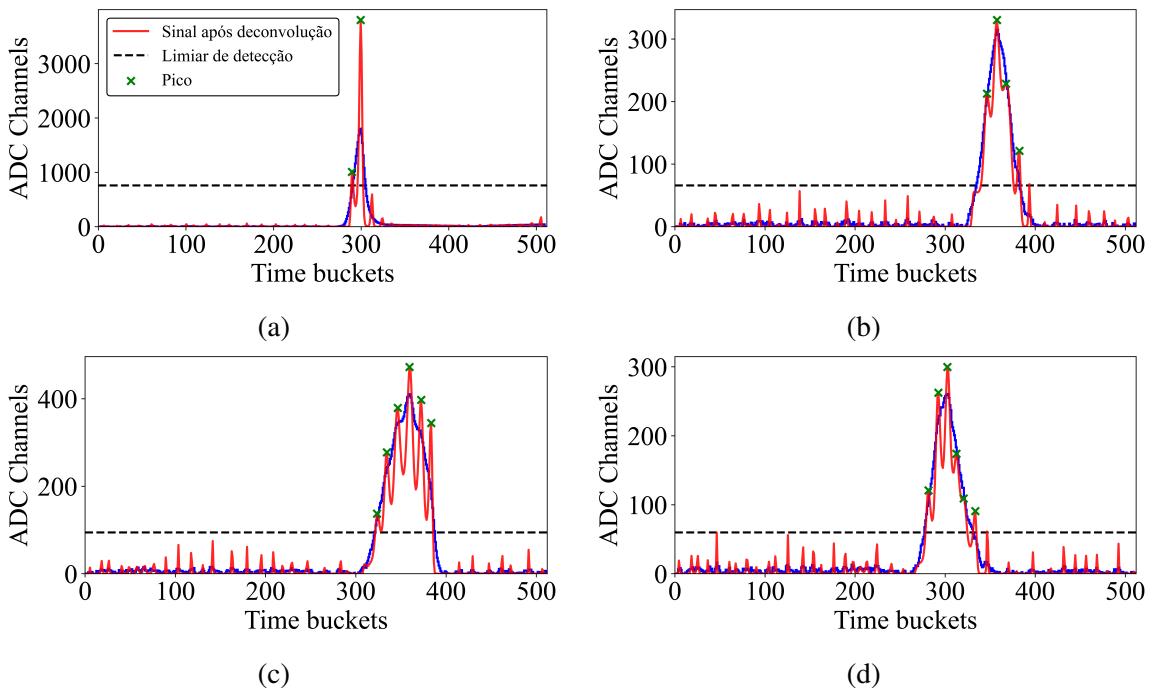


Figura 4.5: Histogramas sem as *baselines* antes (em azul) e depois da deconvolução (em vermelho). Os picos (em verde) e o limiar (linha tracejada preta) de identificação também estão indicados.

O algoritmo de deconvolução também retorna a posição dos centroides encontrados, que indica a localização de um pico. Os mesmos picos podem ser obtidos com o *peak_finder*, com a vantagem de que o algoritmo possui muitos parâmetros diferentes para calibração, melhorando a identificação em comparação com os picos identificados pelo algoritmo do TSpectrum. A execução de 200.000 sinais, desde a estimativa e remoção do fundo, até a identificação dos centroides, demora cerca de 23.25 minutos, usando o processador Ryzen 5 3600X.

Para determinar a carga acumulada Q de cada ponto, é necessário calcular a área do centroide do pico identificado. A área do sinal antes e depois da deconvolução é a mesma, mesmo para a região dos pulsos, portanto pode-se analisar diretamente o sinal após a deconvolução. Para achar a área do pulso, foi calculada a largura dos pulsos após a deconvolução, para determinar a área como uma simples integral gaussiana. O sigma dos pulsos após a deconvolução é $\sigma_{dd} = 1.1543$ (44) *time buckets*. Com isso foi calculada a carga acumulada para cada ponto descoberto do evento. A carga acumulada Q para cada ponto i é dada por:

$$Q = \int_{-\infty}^{\infty} A e^{-(t' - t_i)^2 / 2\sigma_{dd}^2} dt' = A |\sigma_{dd}| \sqrt{2\pi}, \quad (4.4)$$

onde A é a amplitude do ponto com centroide t_i e desvio padrão apó a deconvolução

σ_{dd} .

Com o banco de dados para a deconvolução e também para a identificação de picos, basta criar as redes neurais, que serão descritas na seção 4.2.

4.2 Análise dos pulsos com *machine learning*

Com *machine learning* temos a possibilidade de criar algoritmos de alta complexidade sem definir operações explícitas. Usando os resultados das seções anteriores foram desenvolvidas três redes neurais, com o objetivo de: estimar o fundo (subseção 4.2.1), fazer a deconvolução (4.2.2) e por fim identificar os picos (subseção 4.2.3).

4.2.1 Rede neural para o fundo

O objetivo foi criar uma rede neural que reproduza o comportamento do algoritmo *background removal* que estima o fundo do sinal, que foi discutido na seção 4.1.1, tentando reproduzir resultados muito similares. A rede neural é supervisionada, onde os dados de entrada são os sinais brutos e as saídas devem ser os fundos de cada sinal. A arquitetura está na figura 4.6.

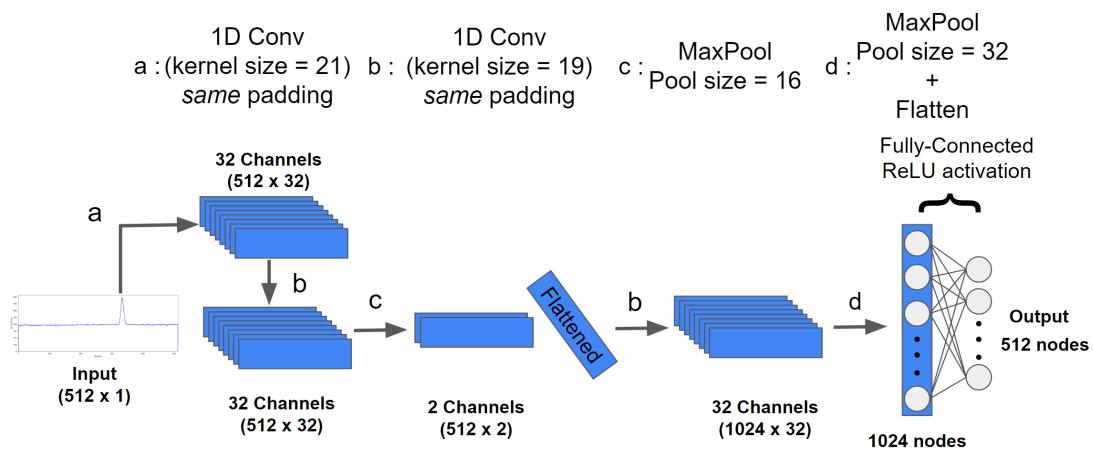


Figura 4.6: Arquitetura da rede neural que faz a inferência do fundo. O vetor de entrada deve ter dimensionalidade 512×1 . Todas as partes com convolução não possuem o parâmetro *bias*.

A entrada da rede é o sinal cru com dimensionalidade 521×1 . Há duas convoluções seguidas (passagens *a* e *b*) com *padding same*, seguida de uma camada com *Max pooling*. Os dois canais restantes sofrem uma planificação (ou *flat*), diminuindo sua dimensionalidade, para então passar por mais uma convolução com *padding same* e filtros de tamanho 19 seguido de uma camada *Max pooling* e uma camada *Fully Connected* com função de

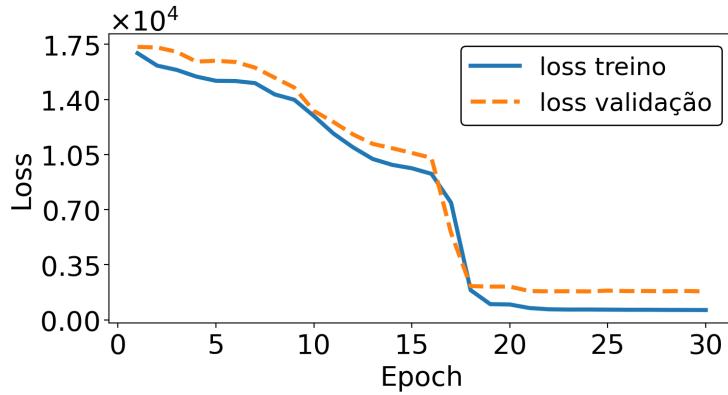
ativação *ReLU*. Toda a rede neural foi construída usando o TensorFlow 2 e possui um total de 545.536 parâmetros, todos treináveis. O tamanho dos filtros das convoluções levam em conta a largura do pulso, sendo no mínimo maior que a largura, a fim de que cada *kernel* atue em um pulso completo na convolução [6].

A camada final com função de ativação *ReLU* garante o valor mínimo de saída em 0 e, principalmente, pelo fato de não causar problemas à minimização do gradiente [57]. Foram testadas diversas combinações e a mostrada na figura 4.6 é a que obteve os melhores resultados [6].

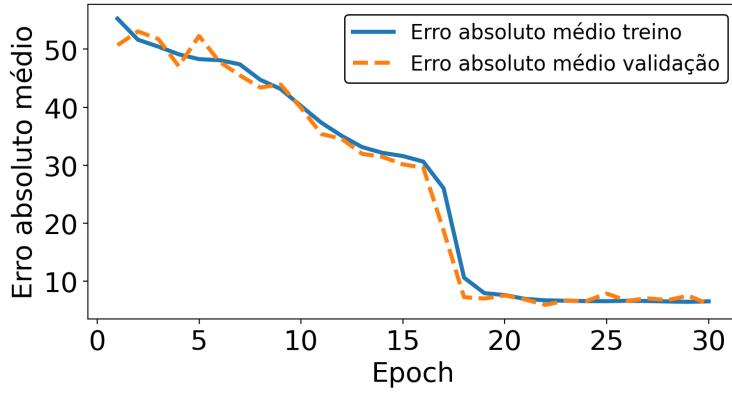
Para o treino foram usados 160.000 sinais para treino e 40.000 para validação. O *loss* foi escolhido como sendo o erro quadrático médio (equação 4.5, o otimizador foi o *ADAMAX* [39], com *learning rate* de 0.0005, e métrica para avaliação foi o erro médio absoluto, dado por

$$E = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i|, \quad (4.5)$$

onde E é o erro absoluto médio, N é o número de pontos e x_i o ponto da saída da rede para ser comparado com o ponto original \hat{x}_i . Foram 30 *epochs* e o *batch-size* foi 8. O treino foi realizado no Google Colaboratory [58] usando a GPU (*graphics processing unit*) NVIDIA Tesla P100 e durou cerca de 34 minutos. Os resultados do treino estão na figura 4.7.



(a) *Loss* dos dados de treino (linha contínua) e dos dados de validação (linha tracejada) em função da *epoch* no treino da rede dada pela figura 4.6.



(b) Erro absoluto médio dos dados de treino (linha contínua) e dos dados de validação (linha tracejada) em função da *epoch* no treino da rede dada pela figura 4.6.

Figura 4.7: Resultados do treino da rede neural dada pela figura 4.6. A rede atingiu seu melhor resultado a partir da *epoch* 20 aproximadamente, quando começa um platô no *loss*.

A arquitetura da figura 4.6 (assim como as próximas desse capítulo) foi determinada de forma empírica. Uma arquitetura com menos passagens e/ou menos parâmetros fornece resultados menos adequados em comparação com a arquitetura apresentada. No caso de mais parâmetros e /ou passagens (consequentemente com aumento no tempo de execução), a rede neural não demonstrou melhora substancial.

Exemplos de resultados de previsões da rede neural estão na figura 4.8. A previsão do fundo possui um erro absoluto nos dados de treino de 6.5315 ADC Channels e nos dados de validação de 6.0783 ADC Channels. O sinal cru é subtraído do fundo, colocando o valor mínimo da subtração em 0. Comparando o erro médio absoluto de 200.000 sinais sem o respectivo fundo (resultante do algoritmo do TSpectrum) com o resultado da rede neural obtemos 4.5 ADC Channels.

Rede neurais convolucionais têm a vantagem de usarem poucas variáveis e serem facilmente paralelizadas em sua execução [25]. Uma vantagem de redes neurais é o seu

tempo de execução. Empiricamente a rede neural pode processar 200.000 sinais em apenas 8s (ou 25.000 sinais por segundo), sendo muito eficiente em tempo.

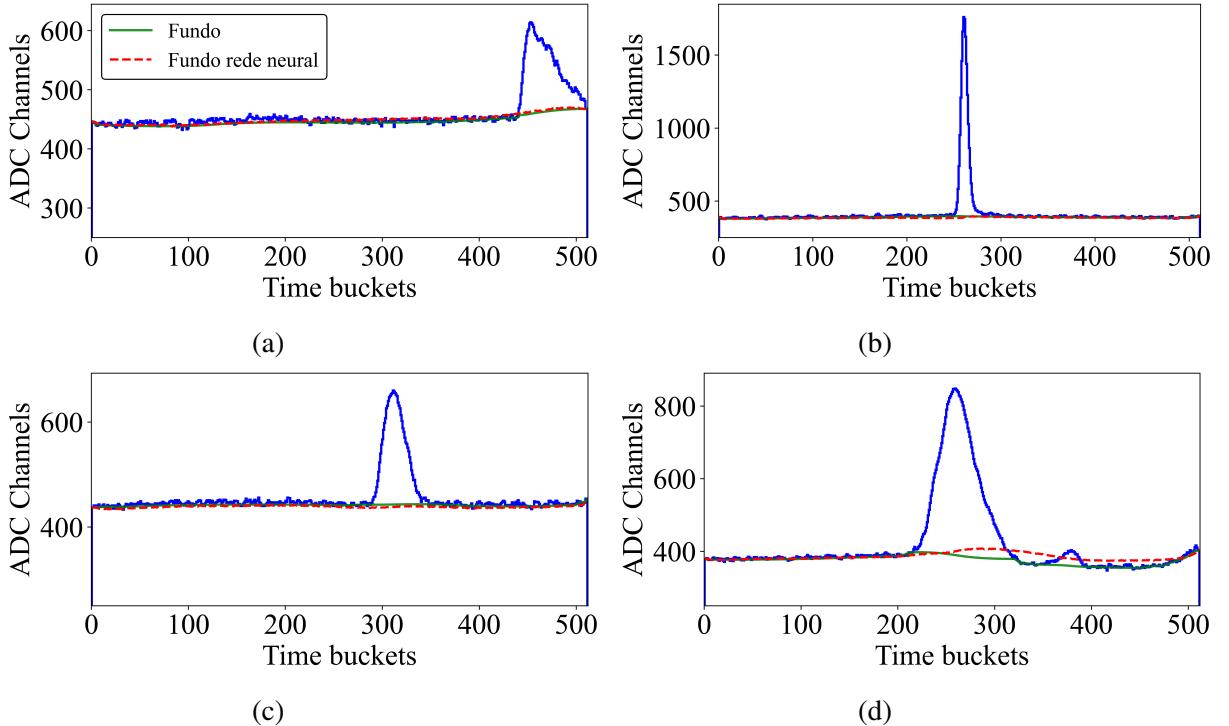


Figura 4.8: Exemplos da rede neural dada pela figura 4.6 em comparação com a saída do *TSpectrum*.

Nos exemplos mostrados nas figuras 4.8a, 4.8b e 4.8a os fundos dos sinais possuem grande flutuação e a rede neural se mostrou eficaz na previsão. No exemplo 4.8d a *baseline* do sinal é muito complexa de se determinar, pois o sinal, aproximadamente do canal 300 a 500, varia em cerca de 50 unidades em y. Apesar da rede neural determinar o fundo acima do fundo original, ao subtrair o espectro do fundo e colocar o valor mínimo em 0, o pulso presente entre os canais 200 e 300 é praticamente inalterado.

Com os resultados obtidos pela rede neural que calcula o sinal de fundo, o próximo passo foi criar a rede neural que faz a deconvolução do espectro sem a *baseline*.

4.2.2 Rede neural para a deconvolução

A mesma abordagem da rede neural anterior foi usada, que é fazer uma sequencia de convoluções e por fim uma camada *fully connected* com função de ativação *ReLU*, pois precisamos ter o valor mínimo do espectro em 0. Os filtros das convoluções precisam ter tamanho mínimo de duas vezes o sigma das gaussianas para atuarem sobre cada pulso do espectro. A entrada da rede é o sinal com o fundo subtraído e com mínimo em 0. A

saída é o sinal apóis deconvolução dada pelo algoritmo *gold deconvolution* na biblioteca *TSpectrum* do ROOT, já mostrado na subseção 4.1.2. A figura 4.9 mostra a arquitetura da rede de deconvolução.

A rede é a sequência de duas convoluções com 32 filtros, *valid padding* e *kernels* de tamanho 19 e 17, respectivamente, seguida de uma camada *Max pooling* com *pool size* igual à 16. No final há o *flat* na camada para seguir com uma camada *fully connected* com função de ativação *ReLU*. O *valid padding* se mostrou mais eficiente para a convergência da rede. Toda a rede foi construída usando o TensorFlow 2, possuindo 508.000 parâmetros treináveis [6].

Assim como na rede anterior, foram usados 160.000 sinais para treino e 40.000 para validação. O *loss* foi escolhido como sendo o erro quadrático médio, o otimizador foi o *ADAM*, com *learning rate* de 0.0005 porém com o parâmetro *clipnorm* igual a 0.45. A métrica para avaliação foi o erro médio absoluto. Foram 75 *epochs* e o *batch-size* foi 8. Os resultados do treino estão na figura 4.10.

Alterar a norma do gradiente (usar o parâmetro *clipnorm* = 0.45) significa que, caso a norma do vetor do gradiente exceda 0.45, então o valor da norma é reajustado para o limiar (*threshold*) escolhido (0.45) [6]. Isso faz com que não ocorra problemas comuns como o gradiente sumir [57, 39], um dos problemas que teve que ser resolvido nessa rede.

O treino foi realizado no Google Colaboratory [58] usando a GPU NVIDIA Tesla P100 e demorou cerca de 54 minutos. Os resultados do treino estão na figura 4.7, onde eles indicam que, visualmente, a rede neural consegue distinguir muito bem diferentes centroides presentes no pulso. Empiricamente, a rede é capaz de executar 200.000 sinais em 5.4 segundos (ou 37.000 sinais por segundo).

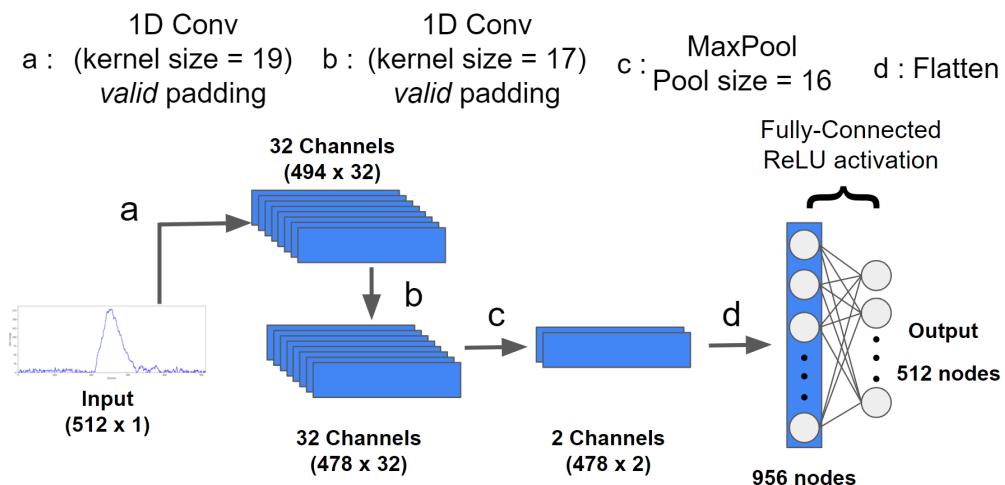
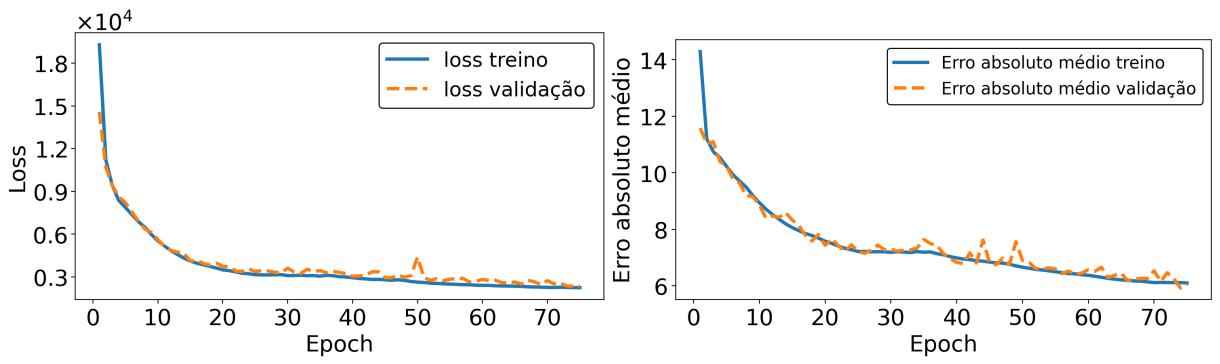


Figura 4.9: Arquitetura da rede neural que faz a inferência da deconvolução do espectro. O vetor de entrada deve ter dimensionalidade 512 x 1. Todas as partes com convolução não possuem o parâmetro *bias*.



(a) *Loss* dos dados de treino (linha contínua) e dos dados de validação (linha tracejada) em função da *epoch* no treino da rede dada pela figura 4.9.

(b) Erro absoluto médio dos dados de treino (linha contínua) e dos dados de validação (linha tracejada) em função da *epoch* no treino da rede dada pela figura 4.9.

Figura 4.10: Resultados do treino da rede neural dada pela figura 4.6.

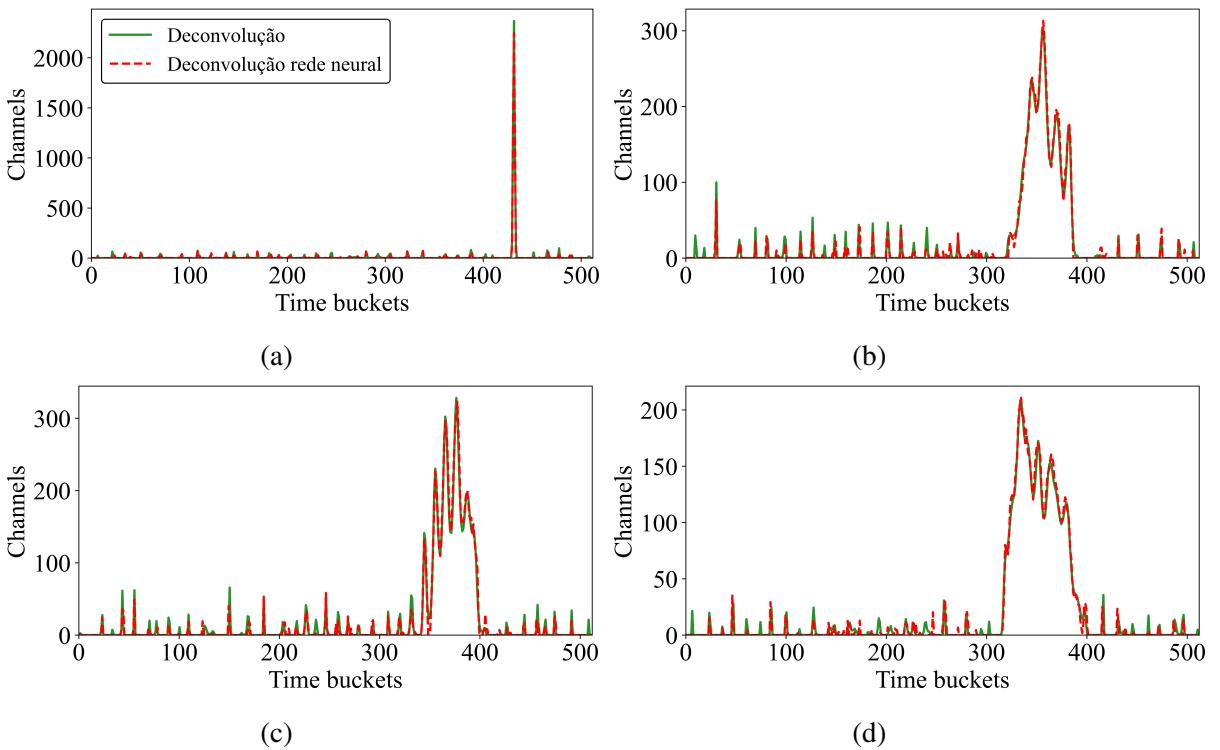


Figura 4.11: Exemplos de deconvolução da rede neural dada pela figura 4.6.

Com a rede neural para a deconvolução feita, a última etapa é a construção de uma rede neural para a identificação de picos, mostrada na 4.2.3.

4.2.3 Identificação de picos

A última etapa dessa análise com *machine learning* foi analisar possíveis soluções para a identificação de picos. Esse é um problema muito complexo, pois há uma variação muito grande na quantidade de picos por sinal e também há um desbalanço muito grande na quantidade de pontos comuns (aqueles que não são picos) e pontos que são picos. Por exemplo, caso haja um sinal que possui apenas um pico, deve-se identificar uma posição, dar o valor de saída como 1, por exemplo, dentre 512 pontos, onde 511 terão o valor de saída como 0. Caso a rede determine que todos os pontos são não picos, ainda assim a acurácia binária seria maior que 99%. Isso é conhecido como desbalanço de classe[59].

Para corrigir esse desbalanço, foram acrescentados pontos simetricamente em torno do pulso, de forma que a somatória das amplitudes das regiões é equivalente a carga Q acumulada no ponto. Isso faz com que não seja preciso identificar um único ponto, mas sim uma região em torno do pico. Com isso podemos nos basear na ideia de segmentar o sinal para destacar regiões de interesse[60]. Segmentar significa ter uma rede neural com a saída com o mesmo tamanho do vetor de entrada (512) e saída com valores entre 0 e 1, onde 1 indica uma região com um pico e 0 não. A figura 4.12 mostra um exemplo de um sinal após a deconvolução onde há os picos identificados com o algoritmo *peak_finder* e os pontos acrescentados simetricamente em torno dos picos para representar as regiões dos pulsos.

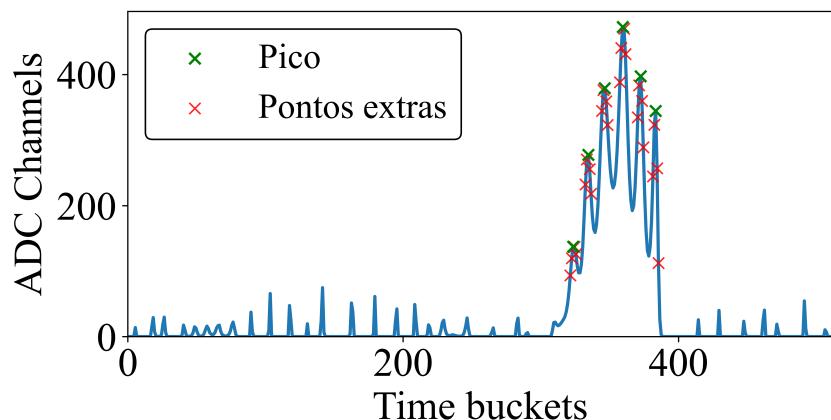


Figura 4.12: Sinal após a deconvolução que mostra o pico identificado mais os pontos adicionais que irão facilitar o trabalho da rede neural (evitar o desbalanço de classe). Foram acrescentados 2 pontos à esquerda e à direita.

A rede construída é a sequência de uma convolução com *kernel* de tamanho 13 e *same padding* seguida de *Max-Pooling* e uma *fully-connected* com função de ativação sigmoide, possuindo um total de 263.104 parâmetros treináveis. Para o treino foram usados sinais

que possuíam entre 1 e 6 picos, resultantes da saída do algoritmo *peak_finder* do *SciPy*, o que resultou em 120.024 de dados para o treino e 30.006 para validação. A escolha pelos picos identificados pelo *peak_finder* ao invés do algoritmo de identificação do TSpectrum é pela maior flexibilidade de ajuste fino do algoritmo, tornando a identificação de picos muito melhor[6]. A função custo escolhida foi a *binary cross-entropy* (dada pela equação 3.5), o otimizador o *ADAM* com *learning rate* de 0.001. A métrica utilizada foi a acurácia binária. O treino também foi realizado por uma GPU NVIDIA Tesla P100 e durou cerca de 8 minutos com 12 *epochs*. A arquitetura da rede está na figura 4.13 e os resultados do treino estão na figura 4.14.

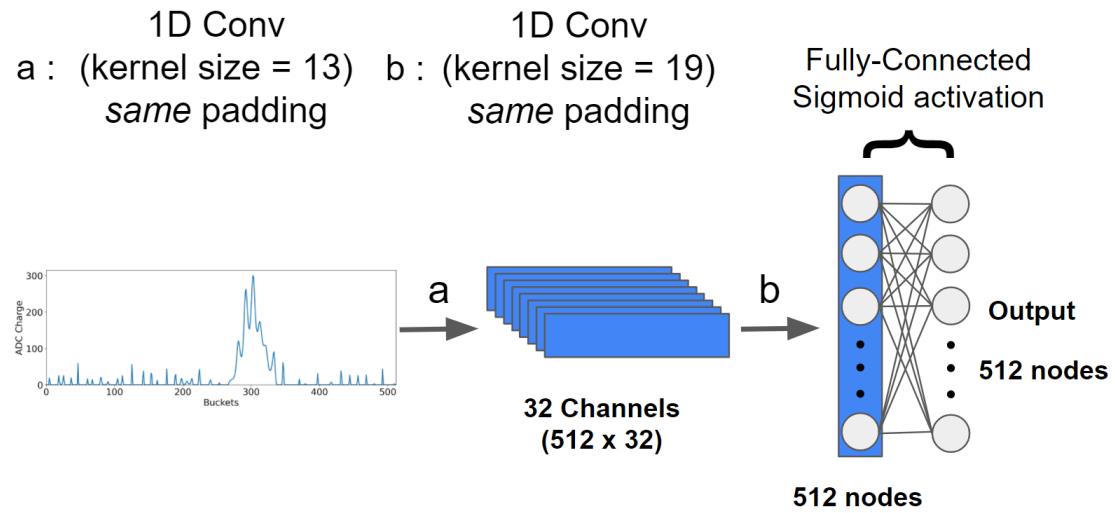


Figura 4.13: Arquitetura da rede neural que faz o recorte das regiões com picos. O vetor de entrada deve ter dimensionalidade 512 x 1.

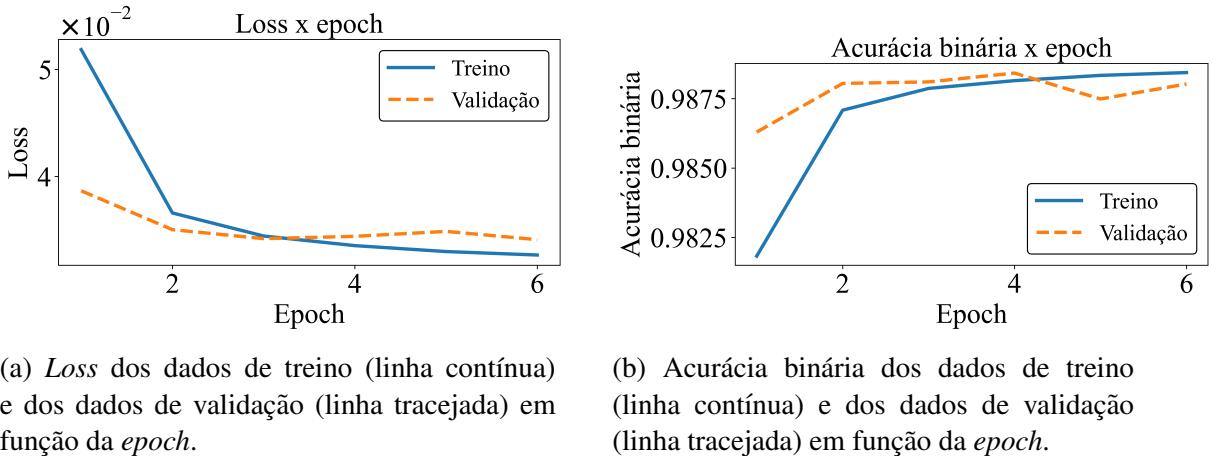


Figura 4.14: Resultados do treino da rede neural dada pela figura 4.13.

A saída da rede neural é um vetor de tamanho 512 com valores entre 0 e 1, onde valores maiores que 0.5 são considerados como pertencentes à um pulso. Com as regiões

identificadas podemos fazer uma média ponderada com o espectro de entrada para achar o centroide. O tempo de processamento da rede neural é de 150.030 sinais em cerca de 4.11 segundos (aproximadamente 36.500 sinais por segundo). Para determinar os picos a partir da saída da rede neural, o tempo é de aproximadamente 4.3 segundos, onde o algoritmo pode ser ainda mais rápido se for paralelizado. Resultados para picos identificados pela rede neural em comparação com o algoritmo *peak-finder* estão na figura 4.15.

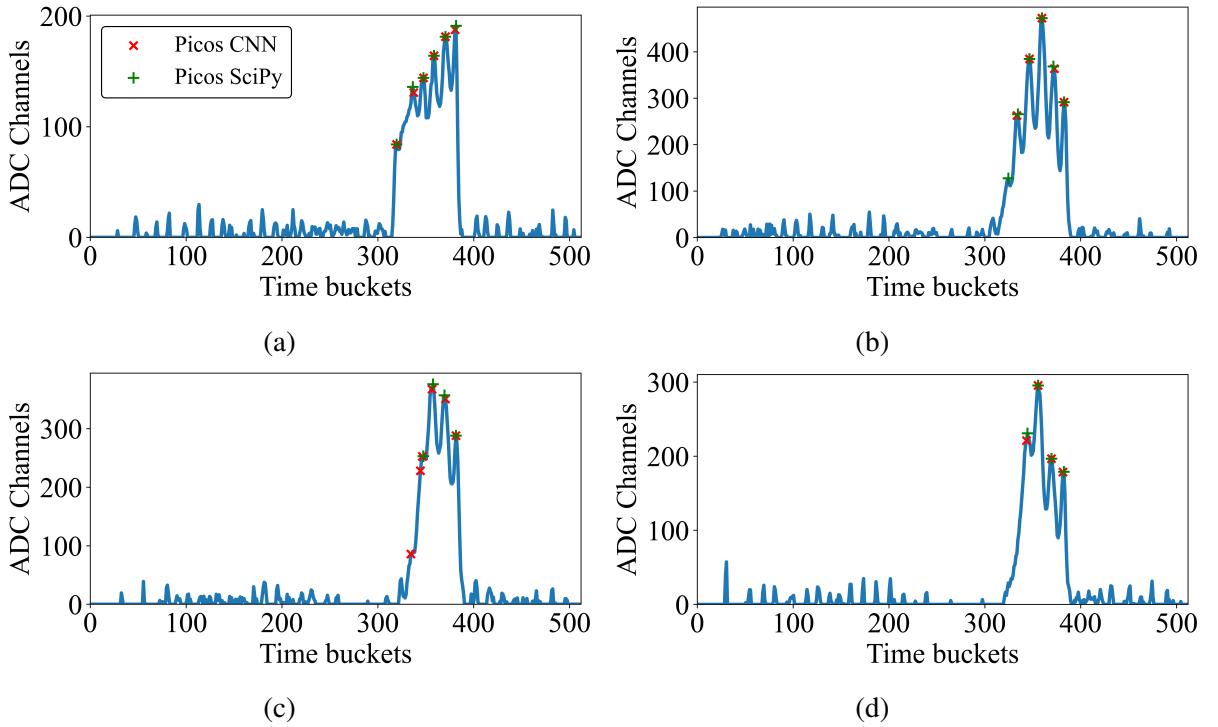


Figura 4.15: Exemplos de identificação de picos usando a rede neural, em comparação com a identificação feita pelo algoritmo presente no SciPy, mostrada na figura 4.13. Os centroides identificados pela rede neural estão em vermelho, e os centroides identificados pelo SciPy estão em verde. Em azul está o espectro sem fundo após a deconvolução, resultante

Para determinar carga acumulada Q_i de cada ponto associado à cada centroide i , foi usada a relação

$$Q_i = \sum_{i=2}^{i+2} f(t_i) \alpha_\sigma, \quad (4.6)$$

onde $f(t_i)$ é a amplitude do espectro no *time bucket* t da posição i e α_σ é uma constante real para calibrar o valor da área em função do sigma dos pulsos, que foi determinado empiricamente como $\alpha_\sigma = 1.2$.

Com as três redes neurais criadas, é necessário verificar o acoplamento das redes, a fim

de analisar a qualidade dos algoritmos e verificar os tempos de execução, para comparar com os algoritmos discutidos nas subseções 4.1.1 e 4.1.2.

4.2.4 Acoplando as redes neurais

Usando novamente o TensorFlow 2 pode-se carregar as redes neurais discutidas nas subseções 4.2.1, 4.2.2 e 4.2.3, já treinadas e usar como se fosse uma única rede neural. Acoplando as três arquiteturas, temos a nova arquitetura mostrada na figura 4.16. O *input* da rede é o sinal cru o *output* da rede neural é o um vetor de tamanho 1024, que possui a segmentação e o sinal após a deconvolução, pois as duas informações são necessárias para determinar os centroides.

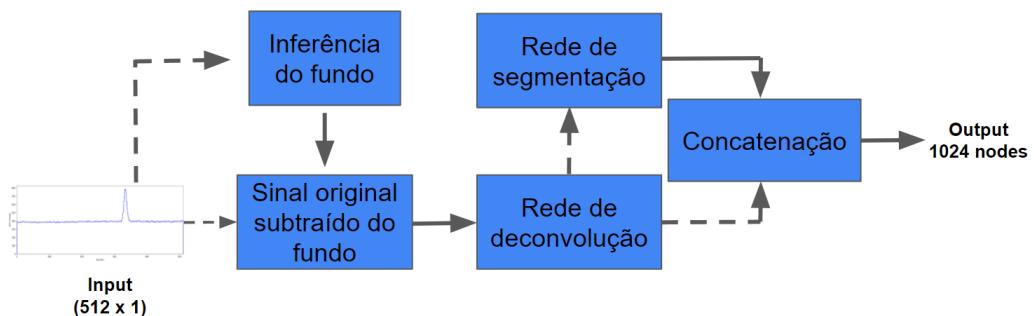


Figura 4.16: Arquitetura da rede neural que faz a inferência da *baseline*, em seguida faz a deconvolução do espectro sem o fundo e por fim faz a segmentação do sinal. O resultado da segmentação e da deconvolução são concatenados na parte final da rede neural. O vetor de entrada deve ter dimensionalidade 512 x 1.

A rede é apenas a sequencia das redes anteriores, ou seja, o espectro passa pelo cálculo da *baseline*, então o espectro original é subtraído dessa *baseline* (colocando o valor mínimo em 0) para passar pela etapa da deconvolução e em seguida o sinal é segmentado. Na etapa final, o resultado da segmentação e da deconvolução são concatenados em um único vetor, para assim determinar os centroides. Pelo fato de cada parte ser treinada de modo separado não há necessidade de treinar a rede unificada, apenas carregar as variáveis das redes neurais treinadas de cada parte. A rede unificada possui 1.316.640 de parâmetros.

Foi comparada a saída da rede acoplada para o espectro após a deconvolução, com a saída de referência, que é o espectro após a deconvolução mostrado na subseção 4.1.2. Podemos usar novamente o erro médio absoluto para fins de comparação. O erro médio absoluto, para os 200.000 sinais, é de 7.45 ADC Channels. Com relação à incerteza, ela foi estimada como o desvio padrão da diferença de cada *time bucket* do sinal tido

como referência e o sinal após a rede neural acoplada. A incerteza é da ordem de 6% da amplitude do sinal no ponto.

Com relação a eficiência em tempo, a rede neural da figura 4.16 processa 200.000 sinais em 12 segundos, usando a GPU NVIDIA Tesla P100. Somando esse tempo com a determinação dos centroides, que é de aproximadamente 4.3 segundos, o tempo total para processar 200 mil sinais é de cerca de 16.3 segundos, cerca de 90 vezes mais rápido que os métodos mostrados na seção 4.1. Exemplos da reconstrução das nuvens de pontos usando a rede neural da figura 4.16 são apresentados na figura 4.17d.

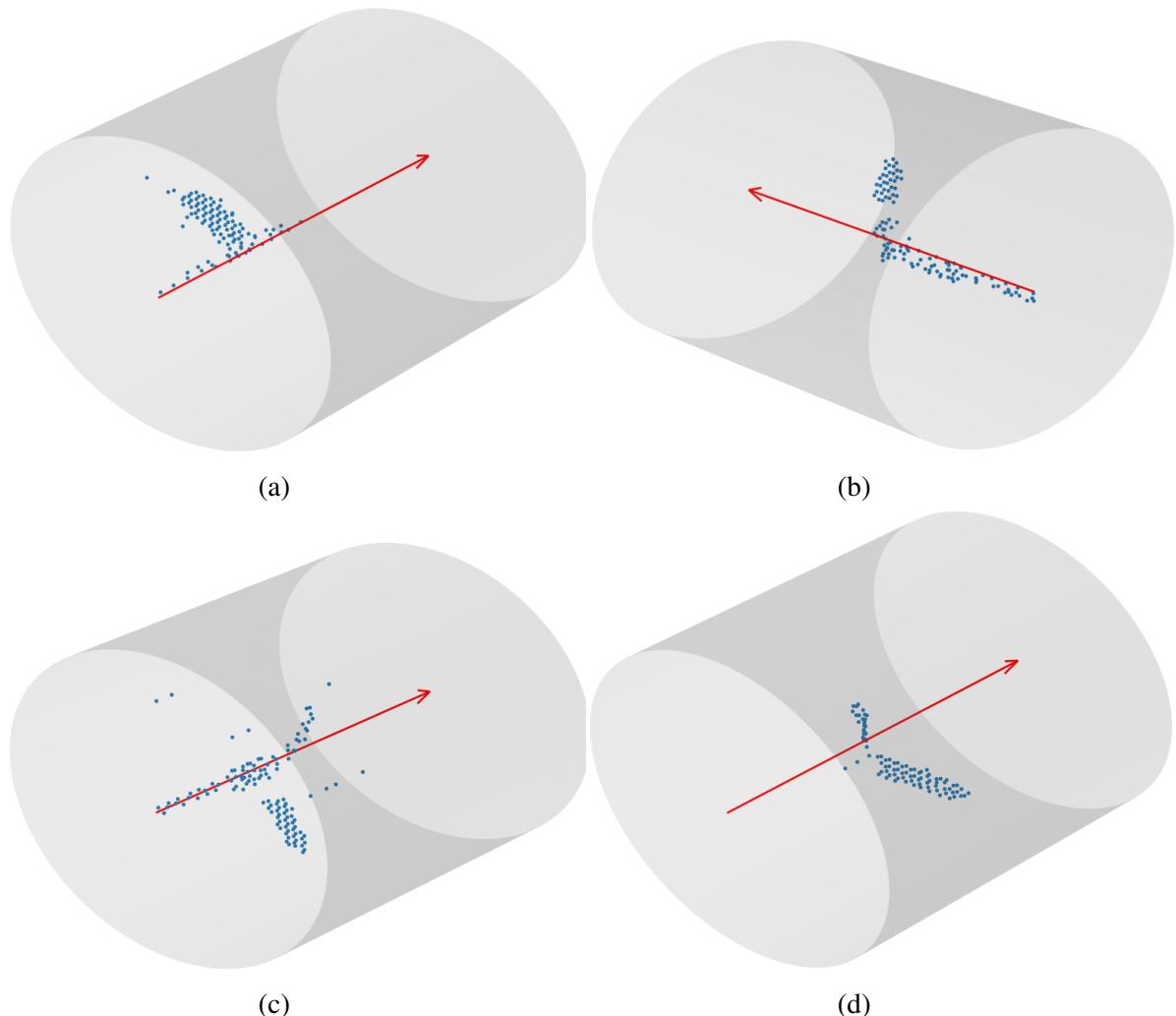


Figura 4.17: Exemplos de eventos reconstruídos através da análise dos sinais com *machine learning*. A seta vermelha indica o sentido do feixe.

Com as redes neurais desenvolvidas, o consumo de tempo para o processamento dos pulsos diminuiu em cerca de 90 vezes em relação à algoritmos mais tradicionais mostrados na seção 4.1, abrindo possibilidades para a análise em tempo real de um experimento com alvo ativo usando grande granularidade (milhares de canais). Com as nuvens de

pontos reconstruídas, pode-se extrair propriedades físicas dos eventos, processo que será descrito no capítulo 5.

Capítulo 5

Análise das nuvens de pontos

As nuvens de pontos representam eventos de reações nucleares dentro do alvo ativo. O objetivo desta etapa, do ponto de vista computacional, foi de identificar *clusters* (aglomerados) tridimensionais, que no caso do pAT-TPC sem a presença de campo magnético, possuem estruturas em forma de linha reta. Já do ponto de vista físico, o objetivo é, com as trajetórias identificadas, extrair as informações físicas (energia E , momento \vec{p} e comprimento L). Com isso, foi possível distinguir cada trajetória identificada, e determinar o vértice de reação (entre uma partícula originada da reação nuclear entre o feixe e o gás). Após a identificação, foram selecionados os eventos correspondentes a uma reação nuclear.

Assim como no capítulo 4, esse capítulo descreve o uso de algoritmos de *machine learning* para a análise, desta vez para as nuvens de pontos. Os algoritmos usados foram tanto não supervisionados (algoritmos de *clustering*) e supervisionados (redes neurais), com diferentes objetivos.

Na seção 5.1 está mostrado o processo de identificação de trajetórias (*tracking*) com algoritmo de *machine learning* não supervisionados. Na seção 5.2.1 estão mostradas alternativas com outros algoritmos de *machine learning*, porém que não foram utilizadas para a análise dos dados, mas abrem um grande leque de possibilidades para estudos futuros.

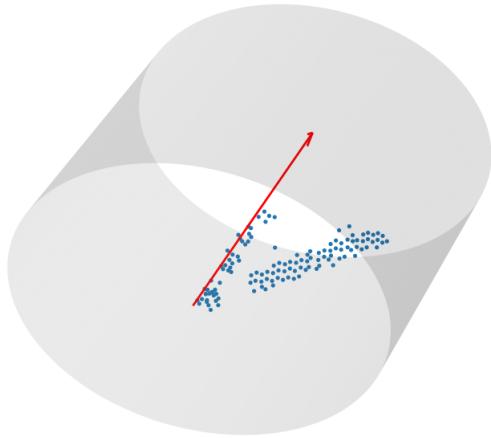
5.1 Identificação de trajetórias

Uma vez reconstruídas as nuvens de pontos de cada evento, é necessário classificar cada uma das trajetórias (tracks) das partículas dentro do gás. A figura 5.1a mostra um exemplo de uma nuvem de pontos. A classificação de tracks pode ser realizada a partir de algoritmos de *machine learning* não supervisionados. Nessa figura é notável ao olho

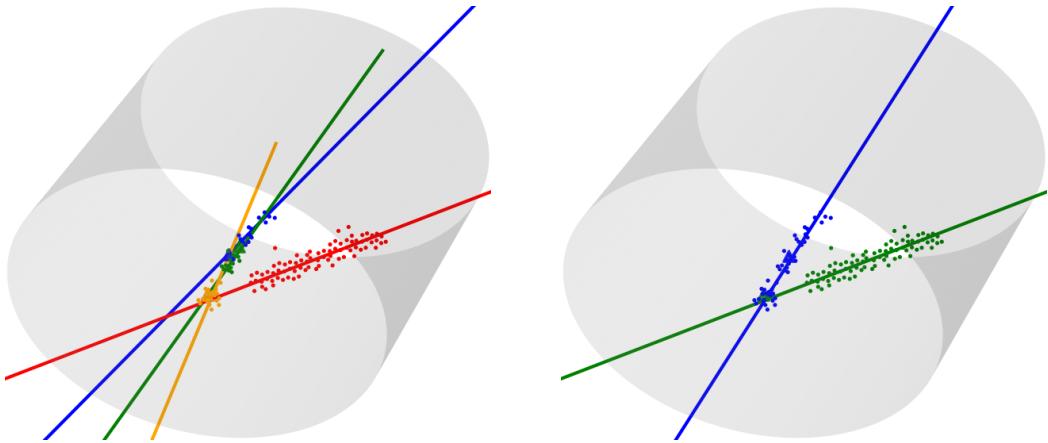
humano que os pontos formam estruturas de aglomerados de pontos. Pode-se associar uma reta a essas estruturas que corresponderiam a trajetória das partículas dentro do alvo gasoso. Essa retas foram identificadas com algoritmos de *clustering*, pelo fato da estrutura aparente dos dados e, também, por não existir um tamanho fixo das nuvens de pontos. O número de pontos pode variar de 30 até cerca de 200, não considerando pontos que não pertencem à nenhum aglomerado (também chamado de *outliers*).

Para diminuir a quantidade de *outliers*, foram usados dois filtros. O primeiro filtro exclui pontos baseado na sua carga, colocando um limiar onde pontos com carga $Q < 110$ são descartados. O segundo filtro, chamado de *outlier removal*, elimina pontos considerados *outliers* globais, de modo que, caso um ponto não possua um número mínimo de vizinhos $n_{or} = 4$ em um raio de distância tridimensional $d_{or} = 12$ mm, então ele é descartado. O *outlier removal* está presente na biblioteca Open3D[61] no Python e funciona excluindo pontos muito isolados uns dos outros.

Existem diversas opções de algoritmos de clustering, como o *Density-based spatial clustering of applications with noise* (DBSCAN) [62], porém a escolha deve levar em conta a performance em tempo, por isso o algoritmo usado foi o *Hierarchical DBSCAN* (HDBSCAN) [63, 64]. O algoritmo tem como entrada a nuvem de pontos e parâmetros esperados para a densidade dos *clusters* presentes no conjunto de pontos. O algoritmo retorna os diferentes *clusters* identificados. Os *clusters* então passam por um ajuste por mínimos quadrados, para determinar o versor e um ponto arbitrário que determinam a reta tridimensional. A figura 5.1b mostra o resultado da aplicação do HDBSCAN na nuvem de pontos mostrada na figura 5.1a. As nuvens de pontos reconstruídas não possuem, em certos casos, densidade de pontos o suficiente para resultar em 100% de acurácia do evento.



(a) Exemplo de evento analisado. Os pontos em azul são das partículas detectadas pelo TPC, a seta vermelha indicando o sentido do feixe e o TPC está representado pelo cilindro cinza.



(b) Evento com as identificações sem a correção. As três retas de cores amarela, verde e azul são de um único *cluster*.

(c) Evento corrigido. Agora o evento possui as duas retas corretas, a azul e a verde.

Figura 5.1: Sequência de análise de um evento. Em 5.1a temos o evento que é recebido para ser analisado, em 5.1b temos o mesmo evento após o HDBSCAN (antes da correção) e 5.1c mostra depois da correção. As cores das retas são arbitrárias e servem apenas para a diferenciação.

O HDBSCAN apresenta falhas no seu resultado, como visto na figura 5.1b, em que, por exemplo, um único *cluster* acaba sendo dividido em três *clusters* muito próximos. A próxima etapa foi de correção da saída do *clustering*. A ideia é comparar, dois a dois, todos os *clusters* resultantes do algoritmo e, se satisfazer um determinado critério, unificar os dois clusters [14]. Do ponto de vista computacional, esse problema é abordado avaliando a semelhança entre dois *clusters* usando métricas, como a distância de Jaccard [65] e o coeficiente de silhueta [66]. A correção feita se dá em duas etapas: primeiro comparando os versores entre duas retas e depois verificando se a condição da equação 5.1 é satisfeita.

Caso a diferença absoluta entre os ângulos com relação ao versor $(0, 0, 1)$ (direção do

feixe de ^{17}F) seja menor que 9° (determinado empiricamente), então as duas retas serão combinadas se obedecerem a condição dada por

$$\sum_{i=0}^{N_1} \frac{d_{i2}}{N_1} < \alpha d_{min}, \quad (5.1)$$

onde N_1 é o número de pontos da reta 1, d_{i2} a distância do ponto i da reta 1 em relação à reta 2, α é um parâmetro com valor a ser escolhido e d_{min} é a distância mínima do ponto a reta. Os valores foram determinados empiricamente, tais que $\alpha = 1.75$ e $d_{min} = 15$ mm. A figura 5.1c mostra o resultado da correção baseada nesses critérios no resultado anterior mostrado na figura 5.1b.

Após a correção, é necessário classificar cada reta como sendo ou o feixe, ou uma partícula originada de uma reação nuclear. O feixe incide na câmara com um ângulo muito pequeno com relação ao versor $(0, 0, 1)$, que é o versor de incidência do feixe. Além disso, mesmo se o ângulo for pequeno, a reta do feixe cruza o plano da janela do TPC próximo do ponto mais provável da entrada o feixe. O ponto mais provável foi calculado usando a posição média da projeção dos pontos de um conjunto de eventos no plano x - y . Disso obtemos que a posição inicial mais provável do feixe é tal que $x_f = -3.4$ (6.7) mm e $y_f = -0.9$ (6.3) mm. A incerteza é alta pois o feixe incide de modo bem distribuído da abertura da janela.

Portanto, se o ângulo entre o versor \hat{v}_i de uma reta r_i for menor que 5° (determinado de modo empírico novamente) e a distância d entre o ponto P_i que intercepta o plano e o ponto $(x_f, y_f, 0)$ for menor que 15 mm (pouco mais que duas vezes a incerteza de cada ponto), então a reta foi considerada como o feixe do evento. No caso de não satisfazer essas condições, então ela foi classificada como uma possível partícula originada da reação do feixe com o gás.

Importante notar que há eventos que não possuem projétil ou ejétil (^{17}F por exemplo), como mostrado na figura 5.2. Neste caso, foi necessário assumir as propriedades da reta mais provável para o feixe, ou seja, precisa passar pelo ponto $(x_f, y_f, 0)$ e ter versor $(0, 0, 1)$.

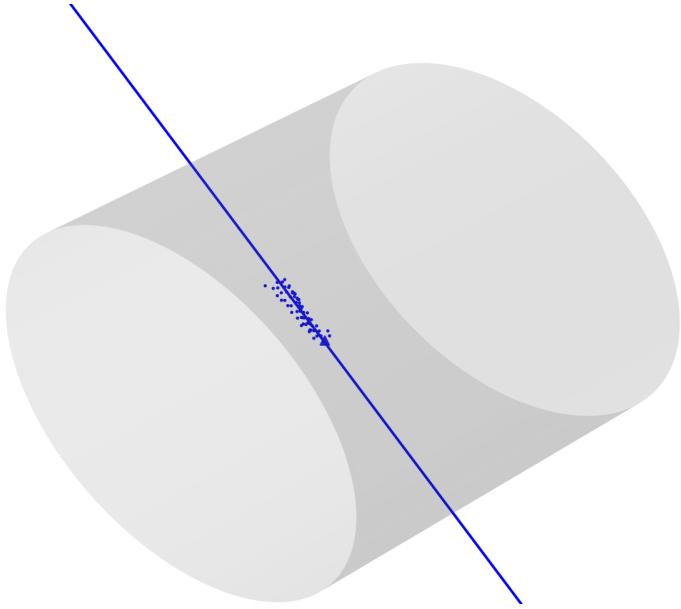


Figura 5.2: Evento em que não foi identificado o feixe, apenas a partícula espalhada. O triângulo azul é o local calculo do vértice de reação dado pela equação 5.6.

Para completar a cinemática do evento, foi necessário calcular o vértice de reação, para cada evento, entre o track da partícula e a direção do feixe. O vértice de reação é o ponto médio do segmento de reta que conecta a reta a ser analisada (partícula) e o feixe, no ponto de menor distância entre as retas. Ele permitiu fazer correções em etapas futuras e diferenciar o começo e o fim de uma trajetória.

Para deduzir a equação para o vértice de reação, primeiro temos as seguintes equações das retas \vec{P}_1 e \vec{P}_2 como vetores:

$$\begin{aligned}\vec{P}_1 &= \vec{A}_1 + \vec{V}_1 * t_1 \\ \vec{P}_2 &= \vec{A}_2 + \vec{V}_2 * t_2,\end{aligned}\tag{5.2}$$

onde \vec{A}_1 e \vec{A}_2 são pontos arbitrários que pertencem as retas 1 e 2, respectivamente, \vec{V}_1 e \vec{V}_2 são os versores, t_1 e t_2 são os hiperparâmetros das retas.

A reta que conecta a menor distância possui versor

$$\vec{V}_c = \frac{\vec{V}_1 \times \vec{V}_2}{|\vec{V}_1 \times \vec{V}_2|}.\tag{5.3}$$

Podemos então construir uma reta \vec{P}_3 que conecta \vec{P}_1 e \vec{P}_2 . Essa reta deve começar no ponto de menor distância da reta 1 e terminar no ponto de menor distância da reta 2. Ou seja, tem-se o seguinte sistema linear:

$$\vec{A}_2 + \vec{V}_2 * \tilde{t}_2 = \vec{A}_1 + \vec{V}_1 * \tilde{t}_1 + \vec{V}_c * \tilde{t}_3.$$

Rearranjando temos que

$$\vec{V}_1 * \tilde{t}_1 - \vec{V}_2 * \tilde{t}_2 + \vec{V}_c * \tilde{t}_3 = \vec{A}_2 - \vec{A}_1, \quad (5.4)$$

onde \tilde{t}_1 , \tilde{t}_2 e \tilde{t}_3 são os hiperparâmetros a serem determinados. Caso \vec{V}_1 seja paralelo à \vec{V}_2 , então não há solução (não há vértice de reação, ou seja, não há uma reação nuclear em comum entre a partícula e o feixe analisado). Com a solução do sistema pode-se obter os pontos de menor distância nas duas retas:

$$\begin{aligned} \vec{P}_1 &= \vec{A}_1 + \vec{V}_1 * \tilde{t}_1 \\ \vec{P}_2 &= \vec{A}_2 + \vec{V}_2 * \tilde{t}_2. \end{aligned} \quad (5.5)$$

Com isso, é possível determinar que o vértice de reação \vec{V}_r é dado por

$$\vec{V}_r = \frac{1}{2}(\vec{P}_1 + \vec{P}_2). \quad (5.6)$$

Com isso pode-se definir a distância de máxima aproximação d_{max} das retas, dada pela equação 5.7.

$$d_{max} = |\vec{P}_1 - \vec{P}_2|. \quad (5.7)$$

Da equação 5.7 foi estabelecido um limite superior para d_{max} tal que valores maiores que esse limite indicam uma reação nuclear em comum entre a partícula e o feixe. O valor foi determinado empiricamente e foi definido como $d_{max}^{sup} = 25$ mm. Trajetórias cuja distância máxima de aproximação excedia d_{max}^{sup} , então a trajetória é descartada. A última condição para garantir que houve uma reação nuclear, é garantir que o vértice de reação está dentro da câmara, cujos limites são $|x| < 140$ mm, $|y| < 140$ mm e $|t| < 512$.

Para completar essa etapa, foi necessário determinar a energia E da trajetória e o comprimento L da trajetória. A energia E é definida como

$$E = \sum_{i=1}^N Q_i, \quad (5.8)$$

onde Q_i é a carga acumulada do i-ésimo ponto de um conjunto com N pontos. O comprimento é definido como

$$L = \max d_{i,\vec{V}_r}, \quad (5.9)$$

onde d_{i,\vec{V}_r} é distância entre o ponto i pertencente a trajetória até o vértice de reação \vec{V}_r .

Essa não foi a única abordagem utilizada no trabalho. Na seção 5.2 foram mostradas alternativas para se buscar, com as nuvens de pontos crua, eventos específicos que ocorreram no experimento, sem ter que previamente buscar trajetórias e também outros usos de *machine learning* para a análise.

5.2 Abordagens alternativas

O objetivo da análise das nuvens de pontos foi buscar eventos que possuíam, de modo geral, três trajetórias (uma sendo o feixe, e as outras duas de partículas que surgiram da reação nuclear do feixe com o gás). Nem todos os eventos puderam ser analisados, devido à falta de trajetórias, o que significa que foi utilizado tempo para a busca desses eventos que poderiam ser descartados.

5.2.1 Identificação de eventos com *machine learning*

Para evitar o processo de *clustering* de nuvens de pontos que não possuem eventos de interesse (como o *breakup*), é possível usar redes neurais supervisionadas capazes de processar nuvens de pontos tridimensionais para selecionar eventos de interesse. A rede neural usada para esse processo chama-se PointNet [67] e a arquitetura está na figura 5.3.

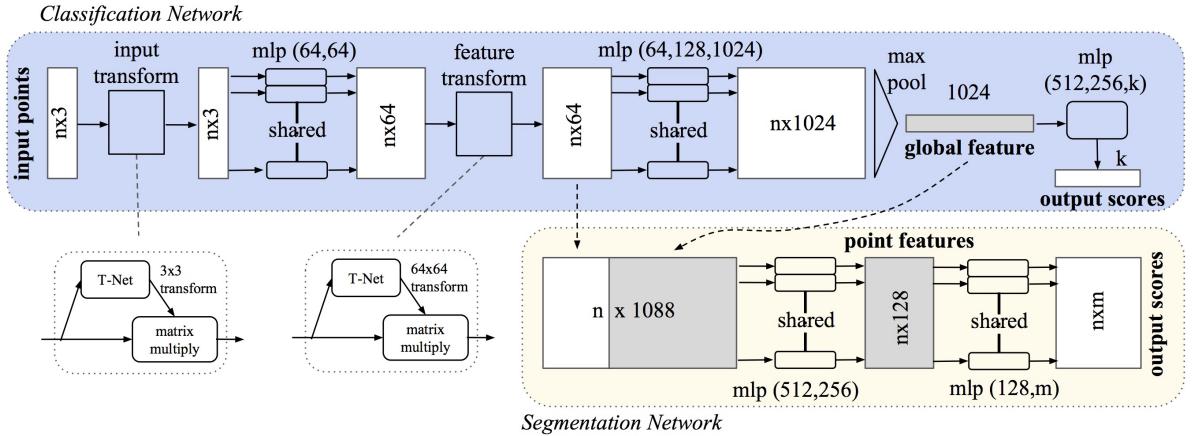


Figura 5.3: Arquitetura da PointNet. A rede de classificação tem o *input* com n pontos com 3 coordenadas, onde são aplicadas sequências de transformação que são agregadas por uma camada de max pooling. O *output* é a classificação para k classes possíveis. A rede de segmentação é uma extensão da rede de classificação, classificando ponto a ponto a nuvem de pontos, em m classes possíveis. Mais detalhes sobre a arquitetura podem ser encontrados na Ref. [67]

A PointNet é uma rede neural capaz de processar *pointclouds*, entendendo que a estrutura dos dados é invariante à troca de pontos e invariante sobre transformações (rotação e translação) [68], pode ser usada tanto para classificação quanto segmentação semântica de nuvens de pontos [67]. A classificação é no para a *pointcloud* completa, já a segmentação semântica é a classificação ponto a ponto da nuvem de pontos [67].

Para o uso da PointNet para classificação, foi criado um banco de dados para o treino da rede neural que tem como *input* a nuvem de pontos e como *output* o número de trajetórias presentes no evento. Um dos pontos negativos da PointNet é seu tamanho fixo nos dados de entrada [67], pois no caso dos dados desse trabalho, o número de pontos por evento não é fixo. Para resolver isso, foi determinado um tamanho fixo $n = 300$ pontos para todas as nuvens de pontos. Eventos com mais pontos que n são descartados, já eventos com menos pontos eram preenchidos com pontos repetidos da mesma nuvem de pontos, para completar até o tamanho n . Eventos com menos que 100 pontos também foram descartados.

Para determinar o número de trajetórias nos eventos para o treino da rede neural, foi utilizado o estimador robusto que chamei de prototype-RANSAC (RANdom SAMple Consensus), que é uma variação do RANSAC [69, 14]. A escolha dele no lugar do HDBSCAN é pela melhor acurácia na identificação de trajetórias, apesar de ser cerca de 4 vezes mais lento. As nuvens de pontos (sem o acréscimo de pontos para a PointNet) ainda passam pelos dois filtros e critérios de correção mostrados na seção 5.1. O algo-

ritmo 1 mostra o funcionamento do p-RANSAC. O algoritmo seleciona dois pontos de modo aleatório (*Random Sampling*) e determina o versor \hat{v} e o ponto P_b que descrevem a única reta r que passa pelos dois pontos. A reta é selecionada caso tenha um número mínimo de pontos $N_{min} = 24$ que pertencem à reta (chamado de *inliers*) e tenha o mínimo (com relação aos outros conjunto de pontos) da estimativa C dada por [14]

$$C = \sum_{i=0}^N \frac{d_i^2}{N}, \quad (5.10)$$

onde N é o número total de pontos de uma reta e d_i é a distância do i-ésimo ponto à reta. O número de iterações do algoritmo foi determinado como sendo 700.

Algoritmo 1: p-RANSAC

Dados: pointcloud, N , d_{min} , N_{min}

- 1 **para** cada iteração $i = 1, 2, \dots, N$ **faz**
 - 2 Seleciona dois pontos da *pointcloud* de modo aleatório (*Random Sampling*);
 - 3 Estima versor v e um ponto P_b que passe pela reta r formada pelos dois pontos;
 - 4 **para** cada ponto P **faz**
 - 5 Calcula a distância d do ponto à reta r ;
 - 6 **se** $d < d_{min}$ **então**
 - 7 Guarda P como pertencente à r ;
 - 8 **se** Número de pontos de $r > N_{min}$ **então**
 - 9 Guarda v , P_b e C ;
 - 10 Ordena as retas do menor para o maior C ;
 - 11 **para** cada reta r ordenada **faz**
 - 12 **se** Número de pontos de $r > N_{min}$ **então**
 - 13 Guarda v , P_b e pontos $P \in r$;
 - 14 **retorna** Retas r selecionadas na última etapa;
-

Para o *output* foram escolhidos eventos que possuem de 0 até 5 trajetórias. Com o *output* construído, a rede para classificação foi treinada, onde a função custo foi a *categorical cross entropy* [38] e o otimizador o ADAM [39]. A métrica utilizada foi a acurácia categórica e foi de 70% quando concluído o treino. Exemplo do resultado da aplicação da rede neural em nuvens de pontos está na figura 5.4.

Trajetórias = 3. Resultado PointNet = 3

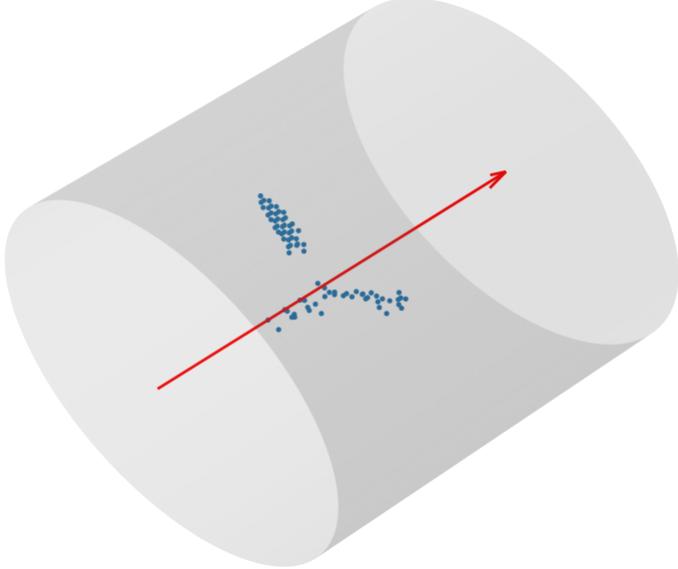


Figura 5.4: Nuvem de pontos que possui 3 trajetórias para serem identificadas e a rede neural de classificação calculou que haviam 3 trajetórias, o que indica um resultado correto do algoritmo.

Eventos com 4 ou 5 trajetórias ocorrem com uma frequência muito menor (menos que 5% do total de dados) que eventos de 1 e 2 trajetórias (cerca de 85% do total de dados), caracterizando o banco de dados com o problema de desbalanço de classe, o que pode ter prejudicado o treino para identificar eventos mais incomuns. No dados do experimento desse trabalho não foi necessário utilizar essa rede neural, pois a etapa de identificação de trajetórias é rápida em comparação com a mesma etapa usado dados gerados nos experimentos com o AT-TPC [7, 6].

5.2.2 Identificação de *outliers*

Um possível uso para a rede de segmentação semântica é para classificar pontos como *inliers* ou *outliers*, semelhante ao uso do algoritmo *outlier removal* da biblioteca Open3D [61]. A diferença é que com a PointNet pode-se incluir *outliers* locais [68], não apenas os globais para serem identificados.

Para o *output* da rede neural, os pontos classificados como *outliers* globais ou locais possuem valor 0 e pontos que são *inliers* (pertencem à alguma trajetória) possuem valor 1. A rede de segmentação foi treinada, com a função custo sendo a *binary cross entropy* e a métrica a acurácia binária. A acurácia da rede foi de aproximadamente 93%, se mostrando uma boa alternativa para eliminação de ruído nos eventos. A figura 5.5 mostra o resultado da aplicação da rede de segmentação em uma nuvem de pontos.

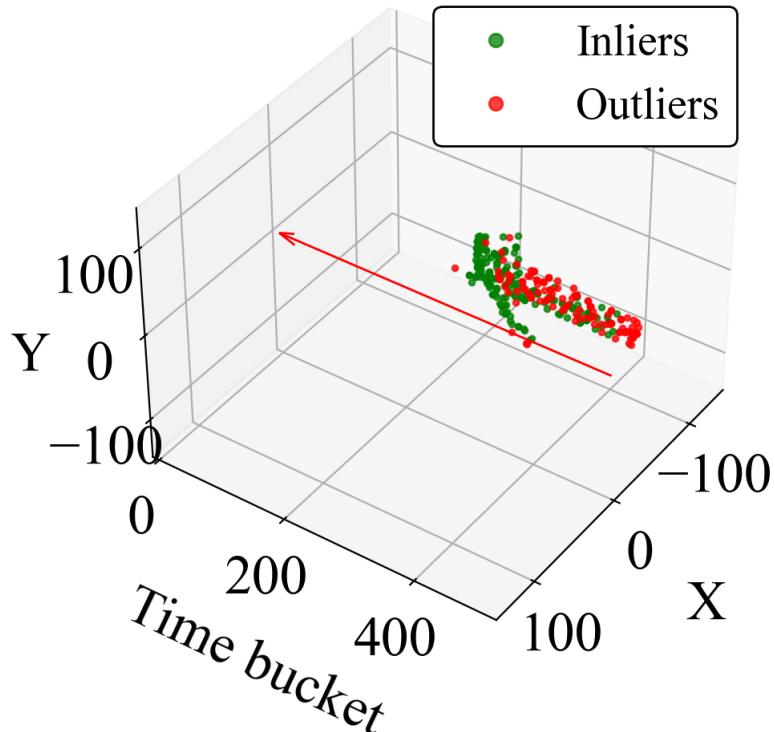


Figura 5.5: Identificação de outliers com a rede neural de segmentação. A rede neural foi capaz de identificar os *outliers* desse evento com 95% de acurácia.

Existem outras redes neurais que são capazes de lidar com *pointclouds*, como a PointNet++ [70] e Dynamic Graph CNN [71]. Para a abordagem de identificação de trajetórias de forma direta, uma opção é utilizar a rede neural ContrastNet [72], pois a rede é capaz de fazer *clustering* nos dados, retornando os diferentes *clusters* existentes nos dados. Num futuro, estamos pensando em utilizar esses tipos de ferramentas para realizar *clustering* em nuvens de pontos em outros experimentos usando múltiplos canais de reação. No entanto, esse projeto esta além dos objetivos desta dissertação.

A classificação e ajuste de trajetórias apresentados no presente capítulo permitem obter informação da cinemática das reações nucleares e extrair as respectivas distribuições angulares. Essa análise é discutida no próximo capítulo.

Capítulo 6

Resultados

Esse capítulo mostra como foram construídas as distribuições angulares da reação de breakup do ^{17}F a partir da análise das nuvens de pontos, como foi apresentado no capítulo 5.

6.1 O vértice da reação

O processo de reconstrução de reações nucleares no alvo ativo é complexo e envolve várias etapas como: identificação de trajetórias e multiplicidade, determinação do vértice de reação, identificação de partículas e reconstrução da energia e ângulos da reação. As subseções a seguir descrevem essas etapas.

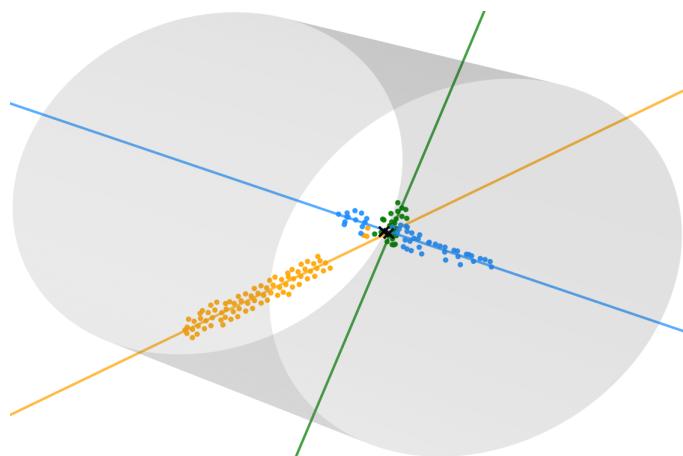
6.1.1 Reconstrução do vértice da reação

Como foi descrito no capítulo 5, para analisar as nuvens de pontos (reações nucleares) é necessário identificar os aglomerados é necessário identificar as trajetórias das partículas dentro do gás que formam linhas retas originadas num determinado vértice. Isso é feito com estimadores robustos, como o algoritmo p-RANSAC (algoritmo 1, apresentado na subseção 5.2.1 e também na referência [14]), que é capaz de identificar os diferentes aglomerados mesmo na presença de *outliers* [14].

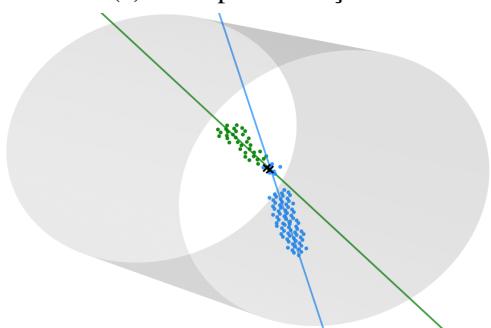
Após a identificação e ajuste das trajetórias, é necessário determinar o vértice da reação para cada evento. O vértice é o ponto no espaço no qual ocorre a reação nuclear e pode ser determinado pela extrapolação das trajetórias na região central do alvo ativo. Para isso, é necessário obter o ponto de interseção (ou mais próximo) entre duas retas, como mostra a equação 5.6. O feixe incide na parte central pelo eixo axial do alvo, região que possui sinais saturados (devido à passagem do feixe). Portanto, para calcular

o vértice de reação, é necessário extrapolar as trajetórias na região central. A partir do vértice, então, é possível calcular o comprimento L da trajetória (dado pela equação 5.9), mesmo com a ausência de informação na região central.

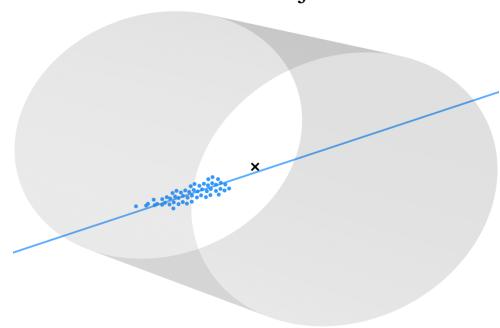
Para alguns eventos foi possível observar a trajetórias do feixe e partículas produto da reação saindo de um mesmo vértice (ver figura 6.1a). Neste caso, o vértice corresponde ao ponto médio da interseção dessas trajetórias. Porém, na maioria dos casos não foi possível observar a trajetória do feixe por causa do baixo ganho nos pixels centrais do detector. Por exemplo, a figura 6.1b mostra só dois produtos de reação saindo do mesmo vértice. Neste caso, o vértice pode ser reconstruído pela extração das retas ajustadas às trajetórias das partículas. Por fim, o caso mais complicado é quando apenas foi identificada uma trajetória (ver figura 6.1c). Para reconstruir o vértice com apenas uma trajetória, é necessário assumir uma direção para a partícula do feixe. Assim como foi mencionado no capítulo 5, se assume que o feixe tem direção $(0,0,1)$ e ponto de corte $(-3.4,-0.9,0)$ que foi calculado usando a posição média da projeção dos pontos de um conjunto de eventos no plano x - y . Portanto, o vértice corresponde ao ponto mais próximo da trajetória partícula com a trajetória assumida para o feixe.



(a) Exemplo de reação nuclear onde foram identificadas três trajetórias.



(b) Exemplo de reação nuclear onde foram identificadas duas trajetórias.



(c) Exemplo de reação nuclear onde foi identificada apenas uma trajetória.

Figura 6.1: Exemplos de reações nucleares com as trajetórias identificadas. O vértice de reação é indicado por um X, em preto. As cores servem apenas para a diferenciação.

6.1.2 Reconstrução da energia das partículas

Numa primeira aproximação, a energia da trajetória pode ser estimada a partir da equação 5.8, onde são somadas todas as cargas acumuladas dos pontos da trajetória. Porém a eficiência de detecção depende da energia depositada pela partícula no gás. Da mesma forma, alguns pontos das trajetórias das partículas podem não ser identificados devido ao processamento eletrônico ou mesmo na reconstrução das nuvens de pontos. Outra forma de reconstruir a energia da partícula é a partir do comprimento da trajetória usando o respetivo poder de freamento (*stopping power*). O *stopping power* depende principalmente da carga da partícula incidente e das propriedades do gás, como a pressão e temperatura. O *stopping power* foi calculado usando o programa LISE++[73] e os respectivos comprimentos das trajetórias associadas a partículas como ^{16}O , alfas e protones num gás de ^4He à uma pressão de 350 Torr.

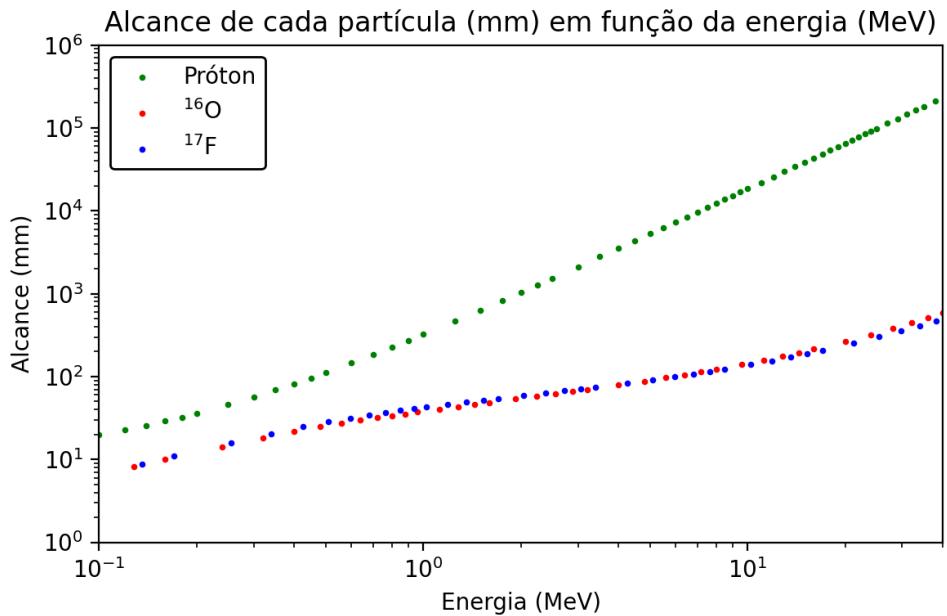


Figura 6.2: Alcance (mm) em função da energia (MeV) para o ^{17}F , ^{16}O e próton. A grande diferença está no próton que tem um alcance muito maior que os outros núcleos.

Portanto, cada tipo de partícula tem um alcance diferente dentro do mesmo gás que depende da energia incidente. Por exemplo, para determinar a energia cinética do ^{17}F antes da reação nuclear, pode se usar a distância entre a janela do detector e o vértice da reação. Para isso, o espaço no eixo z foi discretizado, a fim de calcular a energia na posição da reação. O feixe, ao entrar no gás, possui energia inicial $E_i = 34.76$ MeV, e perde energia ao atravessar o gás. A energia de reação E_r é dada por

$$E_r = E_i - E_{loss}, \quad (6.1)$$

onde E_{loss} é a energia do ^{17}F depositada no gás que é calculada a partir da distância L até o vértice da reação. O comprimento L é transformado em energia cinética usando uma tabela de *stopping power*. Similarmente, a energia de outras partículas (i.e. alfa), que param completamente no gás, é calculado com o alcance da partícula (relativo ao vértice) e a sua respetiva tabela de *stopping power*.

6.1.3 Determinação dos ângulos de reação

Para determinar os ângulos azimutal e axial (em coordenadas esféricas), é necessário fazer os ajustes tridimensionais dos aglomerados identificados (usando estimadores robustos ou algoritmos de *clustering*, como discutido no capítulo 5).

O ângulo de reação, para cada trajetória identificada, depende da determinação do

vessor de cada uma das trajetórias e também do vessor do feixe. A partir das propriedades geométricas das trajetórias (versores das retas), são calculados os ângulos azimuthal e axial com relação à direção em que o feixe incide na câmara (veja que a direção do feixe não necessariamente coincide com o eixo simétrico do TPC).

Para cada trajetória (reta) i , o ângulo axial ϕ_i é dado por

$$\phi_i = \text{atan2} \left(\frac{y_i}{x_i} \right), \quad (6.2)$$

onde atan2 é o arco-tangente calculado entre $-\pi$ e π , y_i e x_i são componentes x e y do vessor da reta i . Para o ângulo azimuthal θ , ele é calculado a partir do produto interno entre o vessor da reta e o vessor do feixe, ou seja:

$$\theta = \arccos \left(\frac{\vec{V}_i \cdot \vec{V}_f}{|\vec{V}_i| |\vec{V}_f|} \right), \quad (6.3)$$

onde \vec{V}_i é o vessor da reta i e \vec{V}_f é o vessor do feixe.

6.2 Identificação do canal de reação de breakup

Com o alvo ativo, é possível medir diversos canais de reação num mesmo experimento. Isso fica evidente ao olhar para a figura 6.3, que mostra um histograma bidimensional do comprimento da trajetória da partícula em função do ângulo de espalhamento (axial) no referencial de laboratório. Eventos de espalhamento em ângulos menores que 5 graus e reações na janela de entrada foram previamente removidos neste histograma.

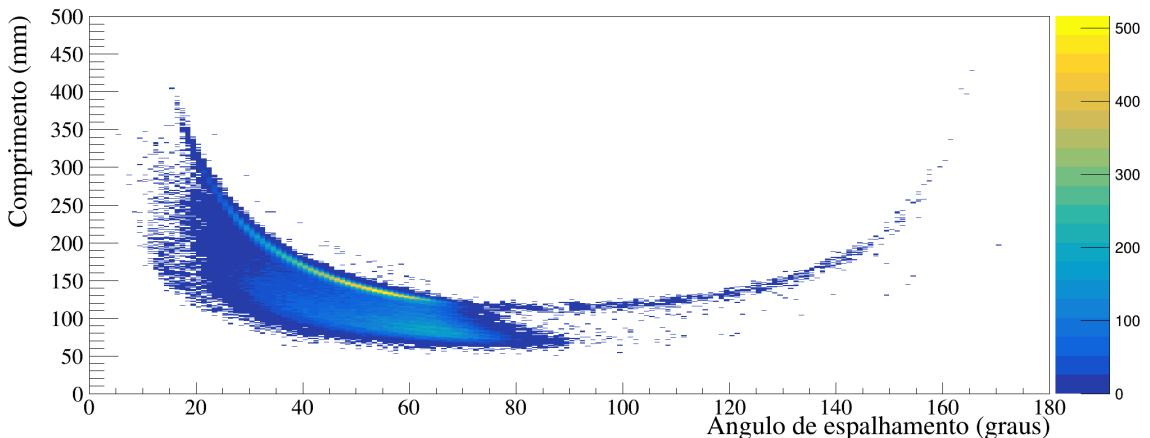


Figura 6.3: Histograma de comprimento de *track* no eixo y e ângulo de espalhamento no eixo x .

Pela figura 6.3, é possível perceber dois aspectos importantes sobre os dados. O

primeiro é que o canal de reação elástico e inelástico é predominante no espectro. Esse é um dos motivos pela assimetria com relação ao ângulo de espalhamento. Devido a que o experimento foi realizado em cinemática inversa, o ângulo de espalhamento da partícula de recuo depende do valor do Q da reação. Por exemplo, partículas alfa espalhadas elasticamente são detectadas em ângulos dianteiros menores que 90 graus. Os ângulos maiores que 90 graus neste caso correspondem a reações com $Q > 0$, por exemplo a reação (α, p) . O segundo aspecto é a ausência de eventos na parte inferior da figura 6.3. Essa ausência ocorre devido ao baixo ganho da região central do detector micromegas que limita a detecção de tracks menores do que 60 mm. Da mesma forma, a ausência de eventos na parte superior do histograma corresponde a limitação geométrica do TPC que possui um raio de 150 mm e comprimento de 500 mm.

A fim de obter o canal de breakup $\alpha(^{17}\text{F}, ^{16}\text{O} + \text{p})$, é necessário identificar os eventos com prótons. A figura 6.4 mostra um exemplo de breakup de ^{17}F em ^4He observado no experimento.

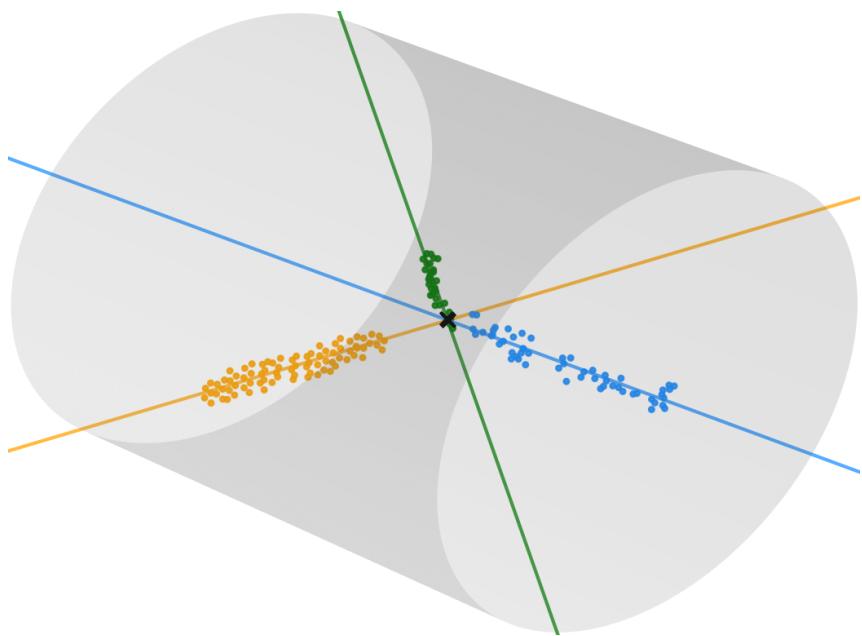


Figura 6.4: Evento onde há a trajetória originada por um próton (trajetória em amarelo), onde percebe-se que seu comprimento vai até o limite geométrico da câmara (cilindro cinza transparente). O feixe está em azul e o ^{16}O originou a trajetória em verde.

A seleção de eventos é feita aplicando vários gates (condições lógicas) para identificar os eventos com prótons. Isto pode ser feito num histograma de carga do track em função de seu comprimento. A relação entre essas duas variáveis se dá por meio do stopping power, onde o termo dominante da equação de Bethe Block [74] em baixas energias pode

ser aproximado a

$$\frac{dE}{dx} \propto -\frac{4\pi Z^2 e^4 N_e}{m_e v^2}, \quad (6.4)$$

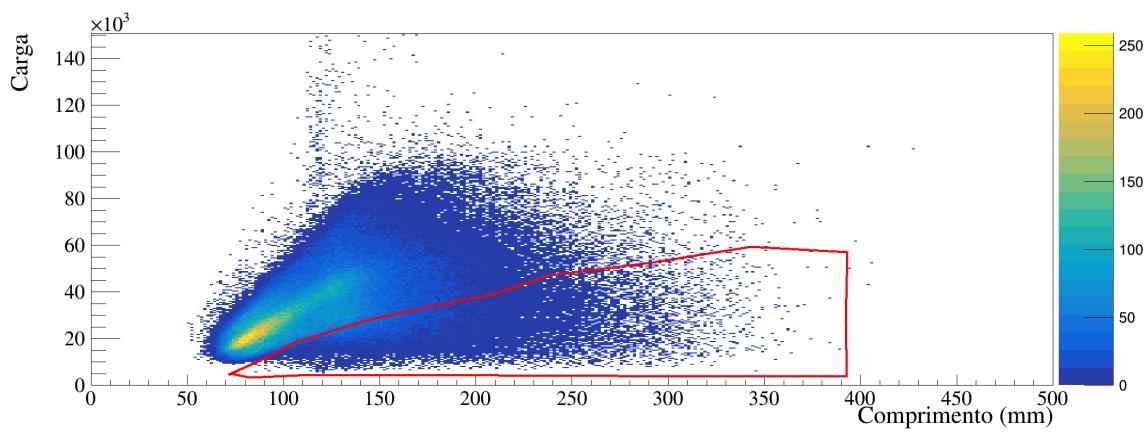
onde E é a energia da partícula com carga ze (e é a carga do elétron) e velocidade v ao atravessar um gás com N_e elétrons por unidade de volume. Sabendo que $2E = Mv^2$, onde M é a massa da partícula, segue que

$$\int_0^E |2E'dE'| \propto \int_0^L \frac{4\pi Z^2 e^4 N_e M}{m_e} dx, \quad (6.5)$$

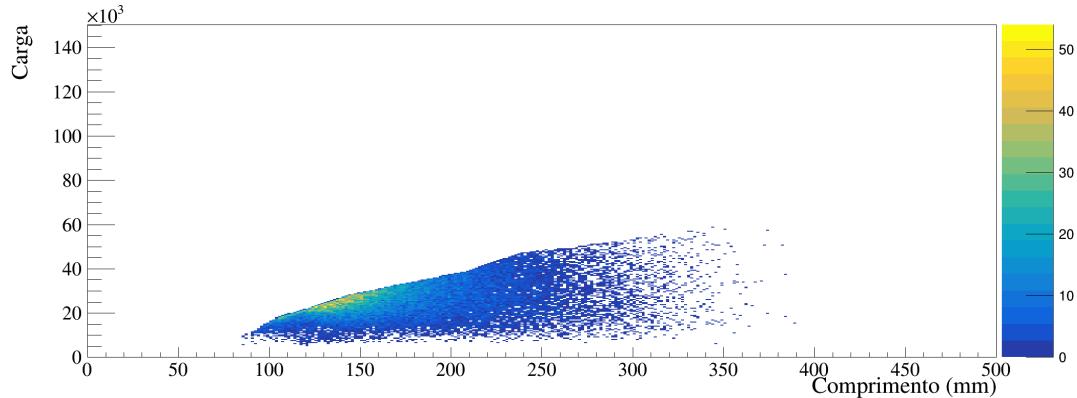
integral que resulta em

$$E^2 \propto (LM) \frac{4\pi Z^2 e^4 N_e}{m_e}. \quad (6.6)$$

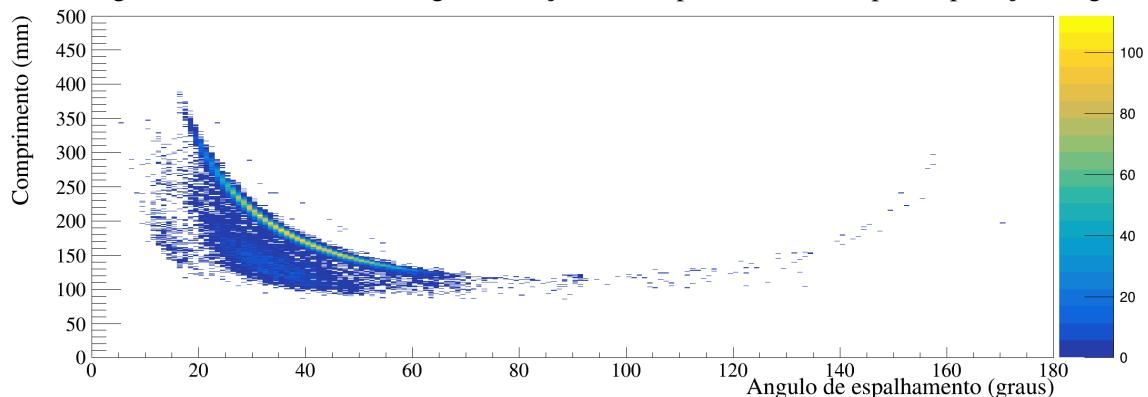
Portanto, existe uma relação do quadrado da energia (ou carga depositada) com o comprimento da trajetória. O termo M mostra que quanto maior o Z da partícula, maior é a carga depositada. Dessa forma, os prótons podem ser identificados num histograma de carga em função do comprimento do track (proporcional à energia depositada) como a região inferior à faixa das alfas, assim como mostrado na figura 6.5a. O resultado da aplicação do gate para o espectro de comprimento em função do ângulo de espalhamento (ver figura 6.3) está na figura 6.5c.



(a) Histograma bidimensional da carga em função do comprimento (mm). A linha vermelha indica o gate, onde pontos (trajetórias) fora da região foram descartados.



(b) Histograma bidimensional da carga em função do comprimento (mm) após a aplicação do gate.



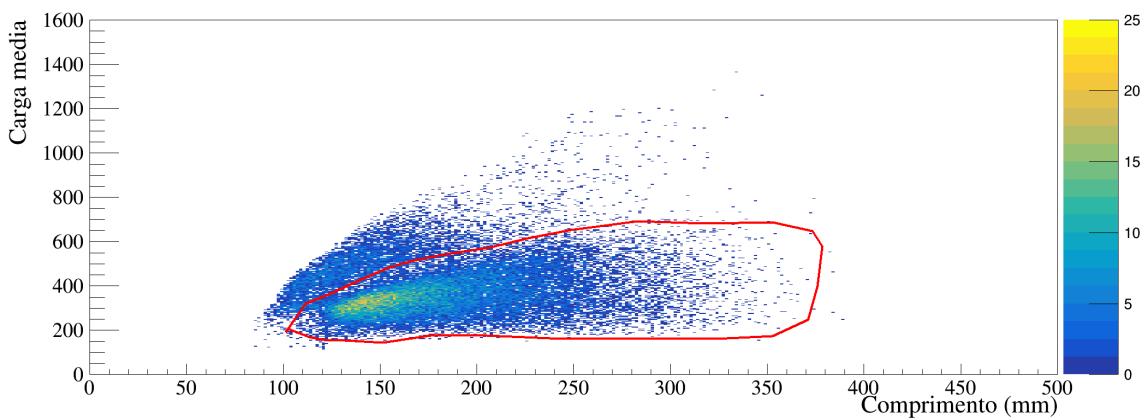
(c) Histograma bidimensional do comprimento (mm) em função do ângulo de espalhamento (graus) no referencial do laboratório.

Figura 6.5: Processo de aplicação do primeiro gate para identificar as trajetórias de reação de breakup. O resultado é o espectro da figura 6.5c.

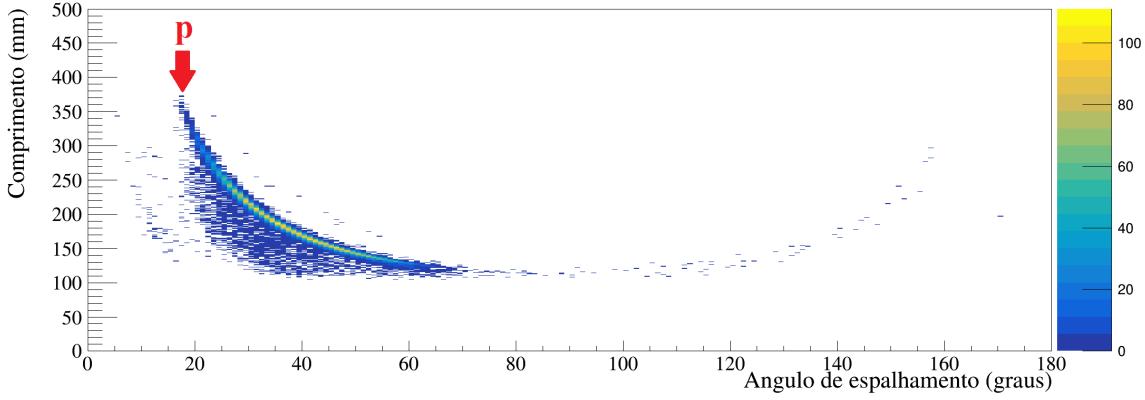
Como pode-se perceber, o gate removeu quase todos os eventos de espalhamento elástico e inelástico de alfas. Somente sobraram duas faixas que correspondem às partículas que não param completamente no gás e partículas espalhadas em ângulos menores que 20 graus. A primeira faixa corresponderia principalmente aos prótons e alfas, pois com ape-

nas 1 MeV de energia cinética atravessam o gás (ver figura 6.2). A segunda faixa seriam as partículas resíduo pesadas (i.e. ^{16}O), devido a que elas estão na região central do TPC onde o ganho é baixo e só partículas que depositam muita energia podem produzir um sinal nessa região.

O gate anterior precisa ainda ser refinado pois ainda é possível ter alguma contribuição de partículas alfa no espectro. Devido a que as partículas alfa depositam maior energia num menor comprimento do que os prótons, a carga média permite visualizar essa diferença. A carga média é definida como a carga total depositada pela partícula dividida pelo número total de pontos no track. O espectro apresentado na figura 6.6a, apresenta a carga média em função do comprimento do track usando o gate da figura 6.6. Pode-se perceber que tem dois aglomerados, um na parte inferior com maior densidade (prótons) e outro na parte superior, logo acima da linha vermelha na imagem, onde há partículas alfas, onde que suas trajetórias possuem uma carga média maior que a dos prótons. Ao excluir trajetórias fora da região delimitada pela linha vermelha na figura 6.6a, é possível identificar os prótons no espectro de comprimento em função do ângulo de espalhamento, mostrado na figura 6.6b.



(a) Histograma bidimensional da carga média em função do comprimento (mm). A linha vermelha indica o segundo gate, onde eventos com trajetórias fora da região foram descartados.



(b) Histograma bidimensional do comprimento (mm) em função do ângulo de espalhamento (graus) após a aplicação do gate mostrado na figura 6.6a. Com esse gate, é possível identificar as trajetórias do próton, indicadas na figura.

Figura 6.6: Processo de aplicação do segundo gate para identificar os prótons. O resultado é um espectro da figura 6.6b.

Para identificar eventos de breakup (ou o canal de breakup), podendo identificar agora os prótons, basta aplicar a condição tal que sejam identificadas duas trajetórias coincidentes (tais que não sejam a do feixe), ou seja, que possuem o mesmo vértice de reação. Aplicando tais condições para identificar eventos com duas trajetórias (coincidências) com o mesmo vértice de reação, eliminando o espalhando elástico e inelástico do espectro, tem-se o espectro do comprimento em função do ângulo de espalhamento para eventos com duas trajetórias coincidentes, mostrado na figura 6.7.

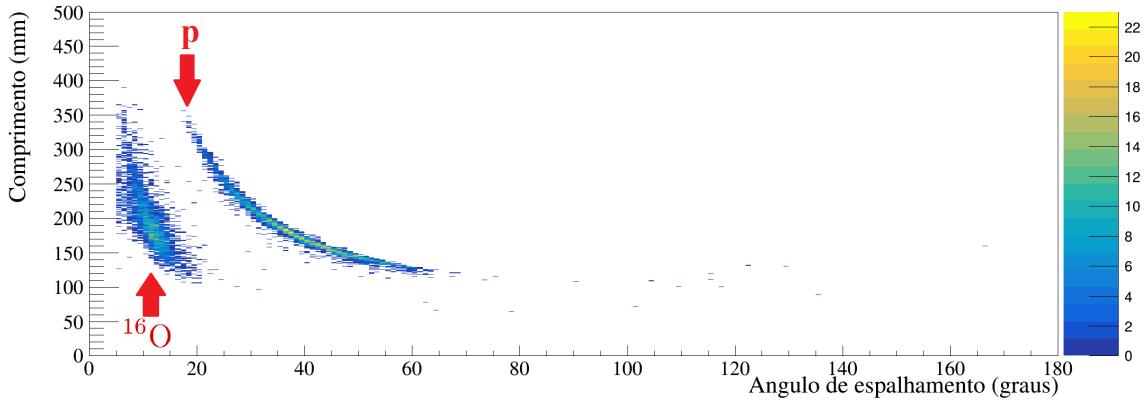


Figura 6.7: Histograma bidimensional do comprimento (mm) em função do ângulo de espalhamento (graus) para eventos com coincidência. Nele é possível identificar os prótons (trajetórias que não param no gás) e o ^{16}O , que possui trajetórias com ângulos mais direcioneiros e comprimentos menores.

Identificados o canal de breakup nos eventos, foi possível construir as distribuições angulares da reação de breakup do ^{17}F em ^4He .

6.3 Construção das distribuições angulares

Identificado o canal de breakup, foi possível construir as distribuições angulares. Como mostrado anteriormente, não é possível identificar o ^{16}O em coincidência com o próton em todos os eventos (o que caracteriza o breakup exclusivo, quando há a identificação de ambas as partículas do breakup). Portanto, as distribuições angulares para o breakup exclusivo foram construídas para eventos com coincidência (^{16}O e próton identificado). Para o breakup inclusivo, foram considerados apenas eventos com identificação apenas do próton.

Com o alvo ativo é possível medir reações em diversos ângulos e energias. Portanto, é preciso primeiro discretizar (colocar em canais) os ângulos de espalhamento θ_p (ângulos de espalhamento dos prótons). Cada canal θ_{bin} de θ_p possui largura de 4° . A seção de choque diferencial $\frac{d\sigma}{d\Omega}(\theta_p)$ é dada por

$$\frac{d\sigma}{d\Omega}(\theta_p) = \frac{1}{\Delta\Omega(\theta_p)L}, \quad (6.7)$$

onde L é a luminosidade (fator de normalização) e $\Delta\Omega(\theta_p)$ é o ângulo sólido calculado em θ_p dado por

$$\Delta\Omega(\theta_p) = 2\pi \int_{\theta_p}^{\theta_p + \theta_{bin}} \sin(\theta) d\theta. \quad (6.8)$$

A normalização deve ser obtida a partir da análise do canal elástico e inelástico do experimento. Até o momento da escrita dessa dissertação, ainda está sendo trabalhado o valor da normalização, portanto os valores apresentados das distribuições angulares estarão em unidades arbitrárias (u.a.).

A energia do feixe também foi discretizada. Como a energia do feixe depende da posição em z do vértice de reação, ela foi discretizada a cada 10 mm, o que resulta em canais de 1.2 MeV de largura. A partir desses canais, foi calculada a energia do feixe. Algumas das distribuições angulares estão mostradas na figura 6.8.

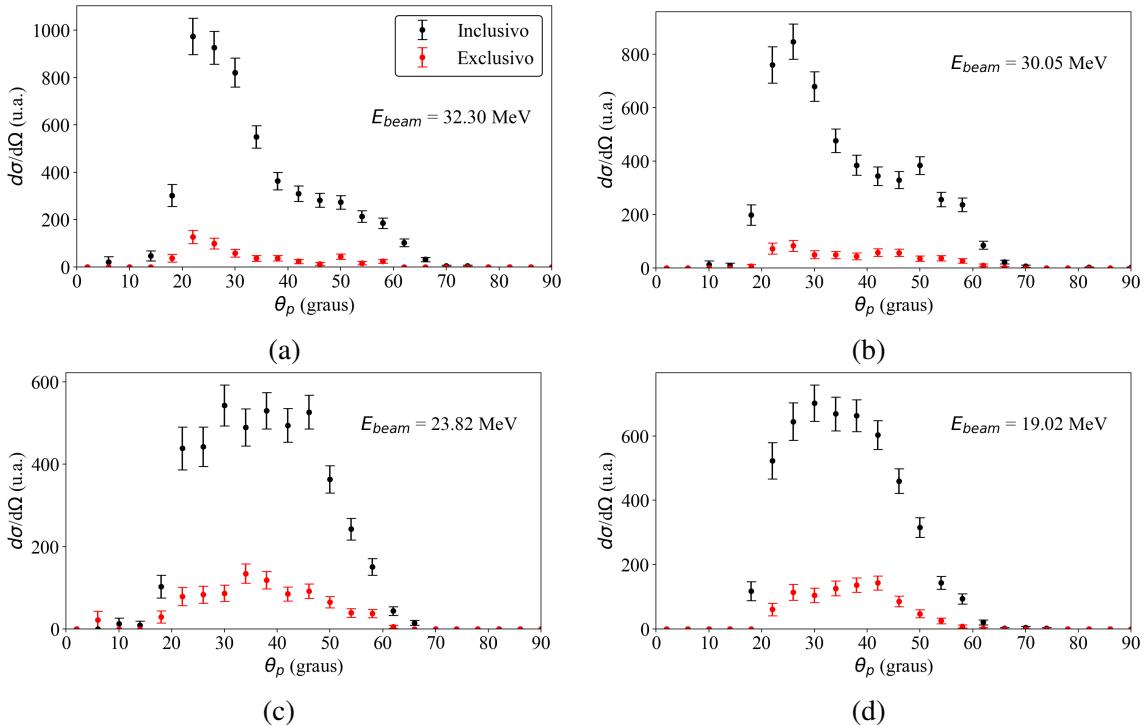


Figura 6.8: Distribuições angulares de breakup do ^{17}F em ^4He em função do ângulo de espalhamento do próton θ_p no referencial do laboratório.

Os resultados mostrados na fig 6.8 são parciais, porém mostram uma forte dependência da energia do feixe, como já discutido na Ref. [4], onde está discutido o breakup do ^{17}F em ^{58}Ni e em ^{208}Pb [5].

Capítulo 7

Conclusão

Este trabalho abordou o uso de técnicas de machine learning para análise de reações nucleares no alvo ativo pAT-TPC, especificamente a reação nuclear de breakup ${}^4\text{He}({}^{17}\text{F}, {}^{16}\text{O}+\text{p}){}^4\text{He}$. No trabalho está descrita a parte experimental, desde a produção do feixe secundário ${}^{17}\text{F}$ até o uso do alvo ativo. Além da parte experimental, as ferramentas que foram utilizadas na análise estão descritas, tanto a parte que envolvem machine learning bem quanto à matemática envolvida em certos algoritmos. As distribuições angulares foram obtidas e seus resultados comparados com resultados de experimentos envolvendo o breakup do ${}^{17}\text{F}$ em ${}^{58}\text{Ni}$ e em ${}^{208}\text{Pb}$.

Para a compreensão da análise feita, a parte matemática e computacional foi brevemente desenvolvida. Para a primeira parte da análise (análise dos pulsos), foram desenvolvidas três redes neurais supervisionadas (uma para a remoção do sinal de fundo, uma para a deconvolução do sinal e outra para identificação de picos) com o objetivo de reconstruir as nuvens de pontos que representam os eventos que ocorreram dentro do alvo ativo. As redes neurais criadas demonstraram bons resultados. O consumo de tempo do processamento dos pulsos com as redes neurais é até 90 vezes menor em comparação com algoritmos tradicionais. O trabalho que envolveu o desenvolvimento desses algoritmos de machine learning resultou no artigo que pode ser encontrado na Ref. [6].

Com as nuvens de pontos construídas, algoritmos de machine learning não supervisionado e estimadores robustos (o estudo de estimadores robustos resultado no artigo encontrado na Ref. [14]) foram usados para identificação de trajetórias. Com as trajetórias, propriedades importantes como o ângulo de reação e comprimento foram determinadas. Também foram mostrados algoritmos de machine learning (PointNet e ClusterNet) com possíveis utilizações para diferentes problemas, como a identificação de outliers e clustering de nuvens de pontos.

Por fim, com as trajetórias identificadas e suas propriedades, foi possível identificar o

canal de reação de breakup, a partir de condições lógicas aplicados em diferentes espectros. Identificado o canal de breakup, as distribuições angulares foram obtidas e resultados parciais foram comentados.

Em conclusão, pode-se dizer que a abordagem com machine learning usada na análise resultou em resultados muito úteis para a análise de experimentos com alvos ativos. Com os resultados obtidos, o próximo passo é utilizá-los em novos experimentos, bem como explorar abordagens diferentes citadas ao longo do trabalho.

Apêndice A

Aplicações de machine learning na física nuclear

Em física nuclear, o uso de técnicas de machine learning tem se mostrado cada vez mais importante. Na continuação são citados alguns exemplos.

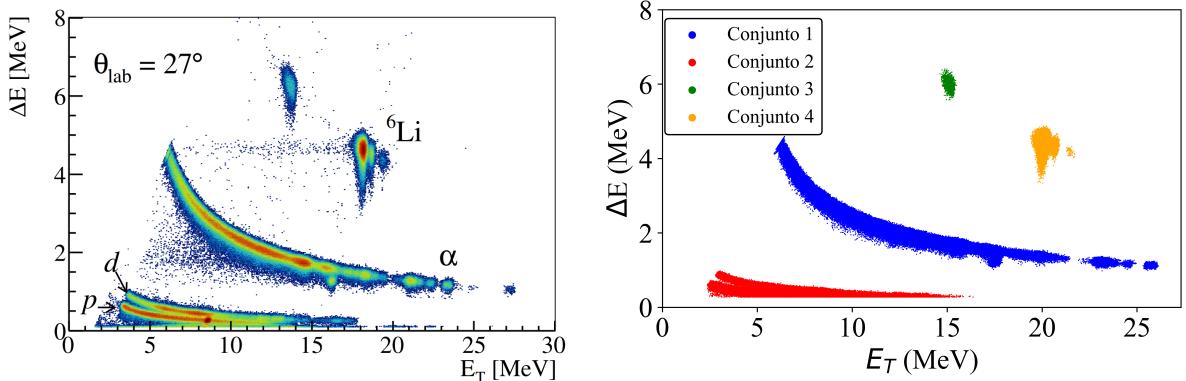
A.1 Análise de espectros para identificação de partículas (*particle identification, PID*)

O uso de técnicas de aprendizado não supervisionado pode ser usado para a identificar de partículas em espectros $\Delta E - E$. Neste tipo de espectros, as partículas são identificadas e separadas por regiões que facilmente podem ser identificadas visualmente [75]. No entanto, a identificação visual desses espectros com os sistemas de detecção com milhares de canais é inviável de forma manual. Assim, ferramentas de machine learning (ou inteligencia artificial) são de grande relevância para esse tipo de analise.

Os resultados apresentados nessa seção foram obtidos na etapa inicial desta dissertação. Para o presente estudo, o espectro biparamétrico ($\Delta E-E$) mostrado na Figura A.1a foi analisado. Os dados correspondem a reação de espalhamento e transferência no sistema ${}^6\text{Li} + {}^{12}\text{C}$.

O objetivo é identificar conjuntos (ou *clusters*) diferentes a partir do espectro. É claro para o olho humano que existem conjuntos diferentes, e eles podem ser identificados usando algoritmos de aprendizado não supervisionado, como por exemplo o *density-based spatial clustering of applications with noise* (DBSCAN) [62]. O DBSCAN consegue identificar *clusters* apenas com as informações características de densidade existente pelos conjuntos que existem nos dados. O resultado da aplicação do algoritmo está na figura

A.1b, onde é mostrado que o espectro agora está separado por diferentes conjuntos.



(a) Espectro biparamétrico $\Delta E \times E$ (energia no referencial do centro de massa).

(b) Espectro biparamétrico $\Delta E \times E$ (energia no referencial do centro de massa) com os conjuntos identificados com o uso do DBSCAN.

Figura A.1: Espectros biparamétricos $\Delta E \times E$, em A.1a o espectro cru com uma pré marcação das partículas, e em A.1b os conjuntos identificados usando *machine learning*.

Usando o conjunto 1, por exemplo, é possível criar o histograma de identificação de partículas dado pela figura A.2.

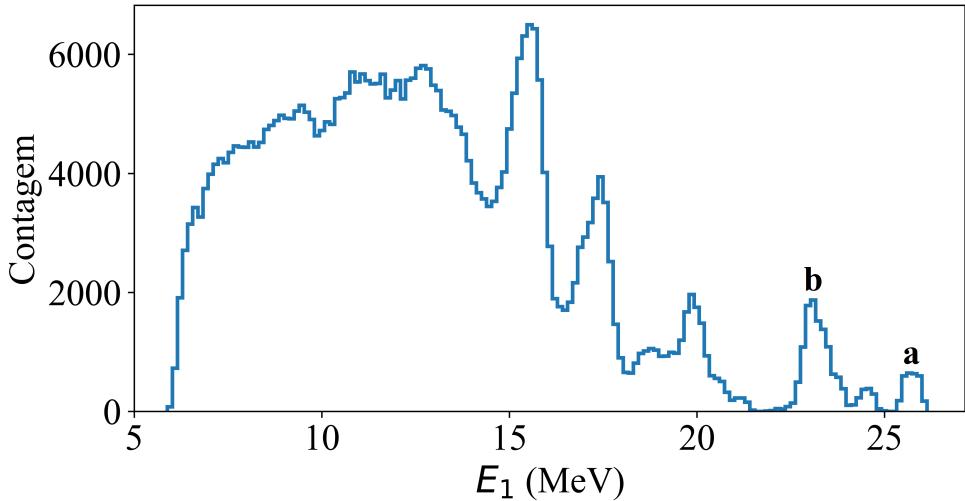


Figura A.2: Histograma de identificação de partículas, a partir do conjunto 1 identificado na figura A.1b. Em **a** temos o primeiro pico que corresponde ao estremo fundamental do ^{14}N e em **b** temos o primeiro estado excitado do ^{14}N .

A.2 Estimativa de raios e massas nucleares

Frequentemente é necessário calcular com alta precisão observáveis que ainda não foram medidos, para contribuir com dados já existentes. É possível estimar propriedades

de núcleos usando modelos e dados experimentais já existentes em conjunto com técnicas de *machine learning*.

É possível usar redes neurais que façam previsão de massa nucleares com informações de massa já disponíveis. Isso foi feito usando o algoritmo *Light Gradient Boosting Machine* (LightGBM) [76]. O algoritmo é uma rede neural cujo aprendizado é supervisionado e minimiza o erro entre a energia de ligação teórica e a energia de ligação experimental dos núcleos (essa diferença é chamada de resíduo), usando 10 quantidades físicas como dados de entrada (*input*) [77]. O desvio quadrático médio para a massa dos núcleos é de 0.234 ± 0.022 MeV, valor melhor que muitos modelos físicos de massa nuclear. A figura A.3 mostra os resíduos calculados entre a energia de ligação calculada pelo rede neural e a energia de ligação determinada experimentalmente.

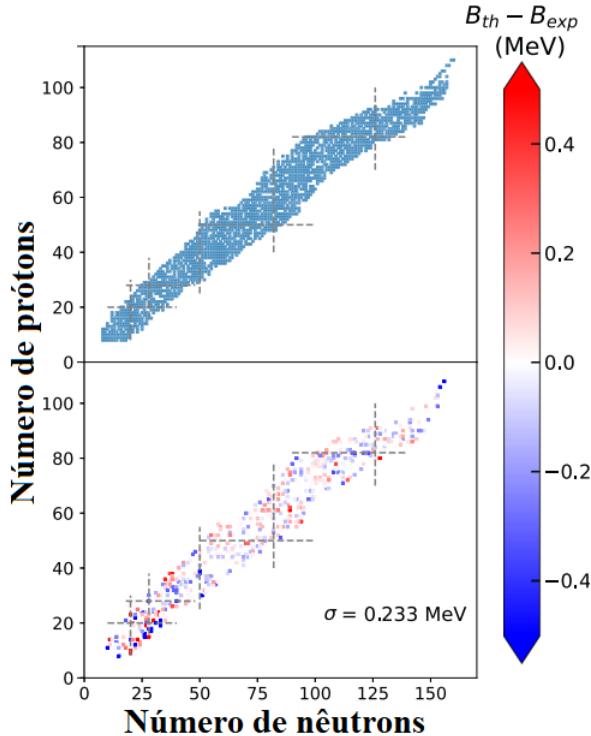


Figura A.3: Painel acima mostra a localização dos núcleos usados para o treino da rede neural. No painel de baixo temos o erro entre a previsão e o valor experimental da energia de ligação do núcleo para os núcleos usados como dados de validação. σ é o erro quadrático médio da rede neural.

Outro exemplo de uso é para a estimativa de raio de carga nuclear. Usando uma rede neural (chamada de *Bayesian neural network extended liquid drop* - BNN-ELD) FC com duas variáveis de entrada, o número de prótons Z e o número de massa A , com a condição de que $Z \geq 20$ e $A \geq 40$. A rede é supervisionada e usa 722 núcleos para dados de treino e 98 núcleos como validação. Resultados para o desvio quadrático médio do raio de carga

são de cerca de 0.02 fm para os 820 núcleos utilizados. A figura A.4 mostra a diferença entre o raio de carga calculado teoricamente e o determinado experimentalmente para isótopos do chumbo, usando diferentes neurais e modelos teóricos. Mais detalhes podem ser encontrados na Ref. [78].

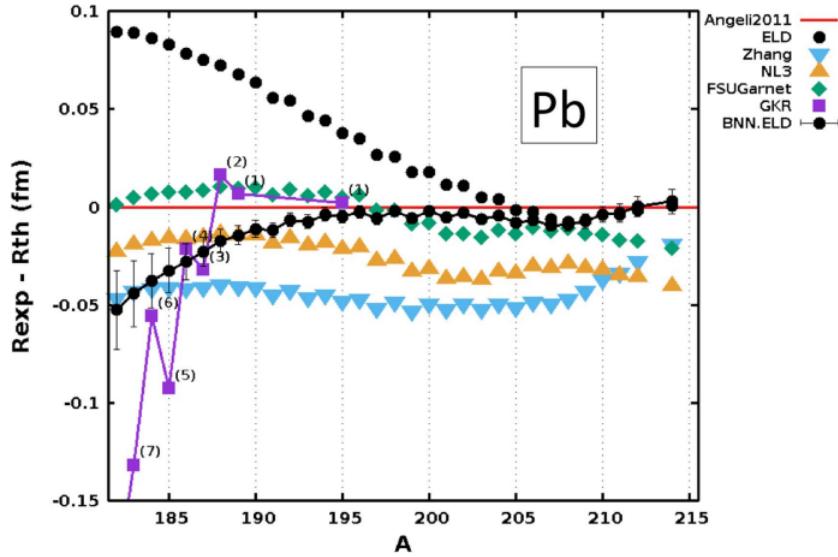


Figura A.4: Previsões para o raio de carga para isótopos do chumbo ($Z = 82$) para diferentes modelos teóricos ou que usam redes neurais. A previsão que contém barras de erro é a que foi brevemente descrita no texto.

A.3 Decaimento β e processo r

O decaimento β é fundamental para entender a origem dos elementos pesados. Prever o tempo de meia vida do decaimento β é de grande importância para simulações do processo r (captura rápida de nêutrons). Com redes neurais é possível fazer previsões que levam em conta a física do problema, como visto na Ref. [79]. O modelo de inteligência artificial leva em conta a teoria de Fermi para o decaimento beta, onde a função f que é minimizada (função custo) é dada por

$$f = \log_{10}(T_{1/2}^a/T_{1/2}^b), \quad (\text{A.1})$$

onde $T_{1/2}^a$ é o tempo de meia vida do decaimento beta medido experimentalmente e $T_{1/2}^b$ é o tempo de meia vida do decaimento beta determinado teoricamente. As variáveis de entrada são o número de prótons Z , o número de nêutrons N , a energia total do decaimento beta e o parâmetro de paridade $\delta = 1, 0, -1$, para núcleos par-par, ímpar-par e ímpar-ímpar, respectivamente.

A figura A.5 mostra a previsão do tempo de meia vida ($T_{1/2}$) do decaimento β (em segundos) para isótonos com número de nêutrons $N = 126$ usando redes neurais.

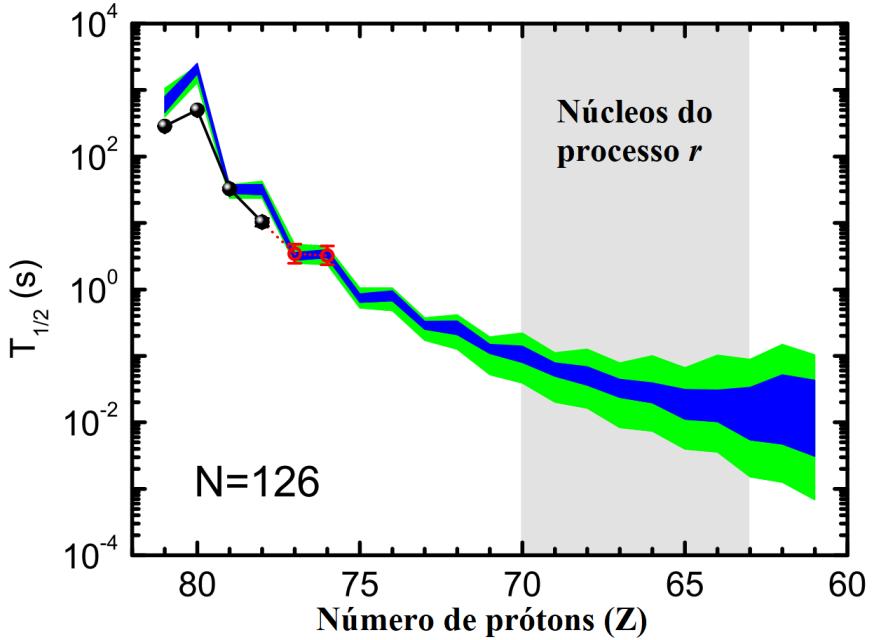


Figura A.5: Meias-vidas de decaimento β para isótonos com $N = 126$. A região hachurada verde mostra as previsões de uma rede neural. A região hachurada em azul mostra os resultados da mesma rede neural, porém seus dados de aprendizado são estendidos para incluir três meias-vidas extras de decaimento β para cada isótopo (indicado por círculos abertos) em direção à *drip-line* de nêutrons [79].

A.4 Alvos ativos

Experimentos com alvos ativos geram enormes quantidades de dados. Uma semana de experimento pode gerar até 10Tb de dados [80], o que gera a necessidade do uso de algoritmos de *machine learning* para diminuir o consumo de tempo da análise desses dados.

A análise dos dados envolve a reconstrução tridimensional dos eventos e posteriormente a reconstrução da cinemática, com a identificação de partículas e de reações nucleares. O uso de técnicas de *machine learning* pode diminuir o consumo de tempo dessas etapas.

O uso de CNNs em imagens feitas a partir das projeções de eventos pode ser útil para a classificação de eventos, sem a necessidade de reconstruir a cinemática em uma etapa anterior. A figura A.6 mostra exemplos de projeções de trajetórias de partículas dentro do alvo ativo, onde a partir da projeção no plano xy é possível identificar a partícula que a originou.

Primeiro, com o objetivo de classificar as projeções entre próton ou carbono, é possível construir uma rede neural para a classificação binária. Para isso foi usada uma arquitetura com camadas convolucionais seguidas de *max-pooling* e por fim uma camada FC [80]. A função a ser minimizada é a dada pela equação 3.5 e a métrica para entender o comportamento da rede neural é a acurácia binária.

Caso o objetivo seja classificar entre três ou mais possibilidades (como por exemplo próton, carbono e outros), então a classificação não será binária, passará a ser categórica, pois será necessário classificar a imagem em alguma categoria. A função custo passará a ser a *categorical cross-entropy* e a métrica será a acurácia categórica.

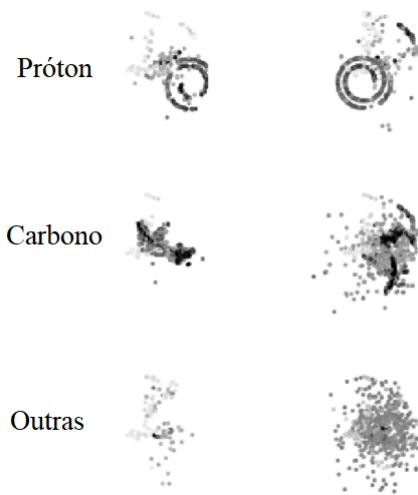


Figura A.6: Projeções no plano xy de partículas dentro do alvo ativo, onde quanto mais escura for a cor, mais carga tem o ponto, e quanto mais clara a cor, menos carga tem o ponto. O objetivo da rede neural pode ser classificar corretamente se a imagem à direita corresponde à uma trajetória de um próton, carbono ou outra partícula, ou pode ser uma rede neural para classificação binária caso seja necessário classificar apenas entre próton ou carbono [80].

A rede criada para a classificação binária (entre próton e carbono) atingiu cerca de 90% de acurácia. Somado ao fato de que redes neurais são muito eficientes em tempo, isso mostra uma possibilidade de que seja possível utilizar redes neurais como essa para o processamento em tempo real dos dados, com pouca supervisão humana durante o processo [80].

Outro problema que pode ser resolvido com técnicas de *machine learning* é a remoção de ruído de eventos. Para isso, pode ser usado o algoritmo não supervisionado *Hough transform* [81], que é capaz de identificar padrões de interesse nos dados e evitar regiões ou pontos que não seguem o padrão desejado.

Referências

- [1] D.W. Bardayan et al. “Determining the $^{14}\text{O}(\alpha,\text{p})^{17}\text{F}$ astrophysical rate from Measurements at TwinSol”. In: *Physics Procedia* 90 (2017). Conference on the Application of Accelerators in Research and Industry, CAARI 2016, 30 October – 4 November 2016, Ft. Worth, TX, USA, pp. 415–420. ISSN: 1875-3892. DOI: <https://doi.org/10.1016/j.phpro.2017.09.043>. URL: <https://www.sciencedirect.com/science/article/pii/S187538921730202X>.
- [2] George Abud Scotton. “Produção de feixes radioativos ricos em nêutrons na região de massa $A = 10$ a 15 .” MA thesis. University of Sao Paulo, Brazil, 2017.
- [3] *The colourful nuclide chart beta*. URL: <https://people.physics.anu.edu.au/~ecs103/chart-beta/>.
- [4] J. J. Kolata, V. Guimarães, and E. F. Aguilera. “Elastic scattering, fusion, and breakup of light exotic nuclei”. In: *The European Physical Journal A* 52.5 (2016), p. 123. ISSN: 1434-601X. DOI: [10.1140/epja/i2016-16123-1](https://doi.org/10.1140/epja/i2016-16123-1). URL: <https://doi.org/10.1140/epja/i2016-16123-1>.
- [5] Y. Kucuk and A. M. Moro. “Exclusive breakup of ^{17}F on ^{58}Ni and ^{208}Pb within the continuum-discretized coupled-channels method”. In: *Phys. Rev. C* 86 (3 2012), p. 034601. DOI: [10.1103/PhysRevC.86.034601](https://doi.org/10.1103/PhysRevC.86.034601). URL: <https://link.aps.org/doi/10.1103/PhysRevC.86.034601>.
- [6] G.F. Fortino et al. “Digital signal analysis based on convolutional neural networks for active target time projection chambers”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1031 (2022), p. 166497. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2022.166497>. URL: <https://www.sciencedirect.com/science/article/pii/S016890022200119X>.
- [7] J. Bradt et al. “Commissioning of the Active-Target Time Projection Chamber”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators*,

- Spectrometers, Detectors and Associated Equipment* 875 (2017), pp. 65–79. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2017.09.013>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900217309683>.
- [8] T. Furuno et al. “Performance test of the MAIKo active target”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 908 (2018), pp. 215–224. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2018.08.042>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900218309951>.
 - [9] E. Koshchiy et al. “Texas Active Target (TexAT) detector for experiments with rare isotope beams”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 957 (2020), p. 163398. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2020.163398>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900220300073>.
 - [10] J. W. Bradt. “Measurement of isobaric analogue resonances of ^{47}Ar with the Active-Target Time Projection Chamber”. PhD thesis. Michigan State University, USA, 2017. URL: https://publications.nscl.msu.edu/thesis/%20Brandt_2017_5279.pdf.
 - [11] Y. Giomataris et al. “MICROMEGAS: a high-granularity position-sensitive gaseous detector for high particle-flux environments”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 376.1 (1996), pp. 29–35. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(96\)00175-1](https://doi.org/10.1016/0168-9002(96)00175-1). URL: <https://www.sciencedirect.com/science/article/pii/0168900296001751>.
 - [12] J. Giovinazzo et al. “GET electronics samples data analysis”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 840 (2016), pp. 15–27. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2016.09.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900216309408>.
 - [13] D. Suzuki et al. “Prototype AT-TPC: Toward a new generation active target time projection chamber for radioactive beam experiments”. In: *Nuclear Instruments*

and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment 691 (Nov. 2012), p. 39. DOI: 10.1016/j.nima.2012.06.050.

- [14] J.C. Zamora and G.F. Fortino. “Tracking algorithms for TPCs using consensus-based robust estimators”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2020), p. 164899. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2020.164899>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900220312961>.
- [15] F.D. Becchetti et al. “The TwinSol low-energy radioactive nuclear beam apparatus: status and recent results”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 505.1 (2003). Proceedings of the tenth Symposium on Radiation Measurements and Applications, pp. 377 –380. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01101-X](https://doi.org/10.1016/S0168-9002(03)01101-X). URL: <http://www.sciencedirect.com/science/article/pii/S016890020301101X>.
- [16] J.J. Kolata et al. “A radioactive beam facility using a large superconducting solenoid”. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 40-41 (1989), pp. 503–506. ISSN: 0168-583X. DOI: [https://doi.org/10.1016/0168-583X\(89\)91032-X](https://doi.org/10.1016/0168-583X(89)91032-X). URL: <https://www.sciencedirect.com/science/article/pii/0168583X8991032X>.
- [17] Panagiotis Gastis et al. “Improved phenomenological description of equilibrium charge state distributions for Ni, Co, and Cu ions in Mo based on new experimental data at 2 MeV/u”. In: (Sept. 2015).
- [18] L. E. Tamayose et al. “Simulation of the RIBRAS Facility with GEANT4”. In: *Brazilian Journal of Physics* 52.3 (2022), p. 89. ISSN: 1678-4448. DOI: 10.1007/s13538-022-01090-y. URL: <https://doi.org/10.1007/s13538-022-01090-y>.
- [19] J.C. Zamora. “Estudo do espalhamento elástico dos isótopos ${}^7\text{Be}$, ${}^9\text{Be}$ e ${}^{10}\text{Be}$ em alvo de ${}^{12}\text{C}$ ”. MA thesis. University of Sao Paulo, Brazil, 2011. DOI: 10.11606/D.43.2011.tde-30092011-132427.

- [20] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [21] R. Lichtenthäler et al. “Radioactive ion beams in Brasil (RIBRAS)”. en. In: *Brazilian Journal of Physics* 33 (June 2003), pp. 294 –296. ISSN: 0103-9733. URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-97332003000200025&nrm=iso.
- [22] Sérgio R. Muniz, Vanderlei S. Bagnato, and M. Bhattacharya. “Analysis of off-axis solenoid fields using the magnetic scalar potential: An application to a Zeeman-slower for cold atoms”. In: *American Journal of Physics* 83.6 (2015), pp. 513–517. DOI: 10.1119/1.4906516. URL: <https://doi.org/10.1119%2F1.4906516>.
- [23] J.S. Randhawa et al. “Beam-induced space-charge effects in time projection chambers in low-energy nuclear physics experiments”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 948 (2019), p. 162830. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2019.162830>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900219312641>.
- [24] Thomas Lohse and Werner Witzeling. “The Time Projection Chamber”. In: *Instrumentation in High Energy Physics*, pp. 81–155. DOI: 10.1142/9789814360333_0002. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789814360333_0002. URL: https://www.worldscientific.com/doi/abs/10.1142/9789814360333_0002.
- [25] A. Geron. *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*. 2nd ed. O’Reilly, 2019. ISBN: 9781492032649.
- [26] *Machine Learning and Data Analysis for Nuclear Physics*. 2020. URL: <https://ectstar.fbk.eu/node/4472>.
- [27] Maryam M. Najafabadi et al. “Deep learning applications and challenges in big data analytics”. In: *Journal of Big Data* 2.1 (2015), p. 1. ISSN: 2196-1115. DOI:

10.1186/s40537-014-0007-7. URL: <https://doi.org/10.1186/s40537-014-0007-7>.

- [28] Morten Hjorth-Jensen. *Data Analysis and Machine Learning: Neural networks, from the simple perceptron to deep learning*. 2020. URL: <https://nucleartalent.github.io/MachineLearningECT/doc/pub/Day4/pdf/Day4.pdf> (visited on 08/2020).
- [29] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York, 2016. ISBN: 9781493938438. URL: <https://books.google.com.br/books?id=kOXDtAEACAAJ>.
- [30] Steven B. Damelin and Willard Miller Jr. *The Mathematics of Signal Processing*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2011. DOI: 10.1017/CBO9781139003896.
- [31] K.-L Du and M.N.s Swamy. “Radial Basis Function Networks”. In: Dec. 2014, pp. 299–335. ISBN: 978-1-4471-5570-6. DOI: 10.1007/978-1-4471-5571-3_10.
- [32] Alex Sherstinsky. “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network”. In: *Physica D: Nonlinear Phenomena* 404 (2020), p. 132306. ISSN: 0167-2789. DOI: <https://doi.org/10.1016/j.physd.2019.132306>. URL: <https://www.sciencedirect.com/science/article/pii/S0167278919305974>.
- [33] Abien Fred Agarap. “Deep Learning using Rectified Linear Units (ReLU)”. In: *arXiv e-prints*, arXiv:1803.08375 (2018), arXiv:1803.08375. arXiv: 1803.08375 [cs.NE].
- [34] Moraga C. Han J. “The influence of the sigmoid function parameters on the speed of backpropagation learning”. In: *Lecture Notes in Computer Science* 930 (1995). DOI: https://doi.org/10.1007/3-540-59497-3_175.
- [35] Tomasz Szandała. *Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks*. Oct. 2020. DOI: https://doi.org/10.1007/978-981-15-5495-7_11.
- [36] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [37] Hecht-Nielsen. “Theory of the backpropagation neural network”. In: *International 1989 Joint Conference on Neural Networks*. 1989, 593–605 vol.1. DOI: 10.1109/IJCNN.1989.118638.

- [38] Douglas Kline and Victor Berardi. “Revisiting squared-error and cross-entropy functions for training neural network classifiers”. In: *Neural Computing and Applications* 14 (Dec. 2005), pp. 310–318. DOI: 10.1007/s00521-005-0467-y.
- [39] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: (2014). DOI: 10.48550/ARXIV.1412.6980. URL: <https://arxiv.org/abs/1412.6980>.
- [40] Sebastian Ruder. “An overview of gradient descent optimization algorithms”. In: *ArXiv* abs/1609.04747 (2016).
- [41] Mohammad Hossin and Sulaiman M.N. “A Review on Evaluation Metrics for Data Classification Evaluations”. In: *International Journal of Data Mining & Knowledge Management Process* 5 (Mar. 2015), pp. 01–11. DOI: 10.5121/ijdkp.2015.5201.
- [42] N. S. Altman. “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression”. In: *The American Statistician* 46.3 (1992), pp. 175–185. DOI: 10.1080/00031305.1992.10475879. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/00031305.1992.10475879>. URL: <https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879>.
- [43] Angela Serra and Roberto Tagliaferri. “Unsupervised Learning: Clustering”. In: Jan. 2018. ISBN: 9780128096338. DOI: 10.1016/B978-0-12-809633-8.20487-1.
- [44] Kristina Sinaga and Miin-Shen Yang. “Unsupervised K-Means Clustering Algorithm”. In: *IEEE Access* PP (Apr. 2020), pp. 1–1. DOI: 10.1109/ACCESS.2020.2988796.
- [45] Y Reddy, Viswanath Pulabaigari, and Eswara B. “Semi-supervised learning: a brief review”. In: *International Journal of Engineering & Technology* 7 (Feb. 2018), p. 81. DOI: 10.14419/ijet.v7i1.8.9977.
- [46] Michael L Littman Leslie Pack Kaelbling and Andrew W Moore. “Reinforcement learning: A survey”. In: *Journal of artificial intelligence research* (1996), 237–285. DOI: <https://doi.org/10.1613/jair.301>.
- [47] Kim D et al. “Review of machine learning methods in soft robotics”. In: *PLoS ONE* (2021). DOI: <https://doi.org/10.1371/journal.pone.0246102>.

- [48] P. Holl et al. “Deep learning based pulse shape discrimination for germanium detectors”. In: *The European Physical Journal C* 79.6 (2019), p. 450. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6869-2. URL: <https://doi.org/10.1140/epjc/s10052-019-6869-2>.
- [49] Edmundo Capelas de Oliveira and José Emilío Maiorino. *Introdução aos métodos da matemática aplicada*. 3rd ed. Editora UNICAMP, 2010. ISBN: 9788526809062.
- [50] *ROOT Data Analysis Framework*. 2021. URL: <https://root.cern.ch/>.
- [51] C.G. Ryan et al. “SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications”. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 34.3 (1988), pp. 396–402. ISSN: 0168-583X. DOI: [https://doi.org/10.1016/0168-583X\(88\)90063-8](https://doi.org/10.1016/0168-583X(88)90063-8). URL: <https://www.sciencedirect.com/science/article/pii/0168583X88900638>.
- [52] Miroslav Morháč et al. “Background elimination methods for multidimensional coincidence γ -ray spectra”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 401.1 (1997), pp. 113–132. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(97\)01023-1](https://doi.org/10.1016/S0168-9002(97)01023-1). URL: <https://www.sciencedirect.com/science/article/pii/S0168900297010231>.
- [53] Donald D. Burgess and Richard J. Tervo. “Background estimation for gamma-ray spectrometry”. In: *Nuclear Instruments and Methods in Physics Research* 214.2 (1983), pp. 431–434. ISSN: 0167-5087. DOI: [https://doi.org/10.1016/0167-5087\(83\)90612-9](https://doi.org/10.1016/0167-5087(83)90612-9). URL: <https://www.sciencedirect.com/science/article/pii/0167508783906129>.
- [54] Miroslav Morháč et al. “Identification of peaks in multidimensional coincidence γ -ray spectra”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 443.1 (2000), pp. 108–125. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(99\)01005-0](https://doi.org/10.1016/S0168-9002(99)01005-0). URL: <https://www.sciencedirect.com/science/article/pii/S0168900299010050>.
- [55] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

- [56] Matt Newville et al. *lmfit/lmfit-py 1.0.2*. Version 1.0.2. Feb. 2021. DOI: 10.5281/zenodo.4516651. URL: <https://doi.org/10.5281/zenodo.4516651>.
- [57] Xavier Glorot and Y. Bengio. “Understanding the difficulty of training deep feed-forward neural networks”. In: *Journal of Machine Learning Research - Proceedings Track 9* (Jan. 2010), pp. 249–256.
- [58] Ekaba Bisong. “Google Colaboratory”. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Berkeley, CA: Apress, 2019, pp. 59–64. ISBN: 978-1-4842-4470-8. DOI: 10.1007/978-1-4842-4470-8_7. URL: https://doi.org/10.1007/978-1-4842-4470-8_7.
- [59] Andrea Dal Pozzolo et al. “Calibrating Probability with Undersampling for Unbalanced Classification”. In: Dec. 2015. DOI: 10.1109/SSCI.2015.33.
- [60] Ashraf A Aly, Safaai Bin Deris, and Nazar Zaki. “Research review for digital image segmentation techniques”. In: *Int. J. Comput. Sci. Inf. Technol. Res.* 3.5 (2011), p. 99.
- [61] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. “Open3D: A Modern Library for 3D Data Processing”. In: *arXiv:1801.09847* (2018).
- [62] Erich Schubert et al. “DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN”. In: *ACM Trans. Database Syst.* 42.3 (2017). ISSN: 0362-5915. DOI: 10.1145/3068335. URL: <https://doi.org/10.1145/3068335>.
- [63] Leland McInnes, John Healy, and Steve Astels. “hdbscan: Hierarchical density based clustering”. In: *The Journal of Open Source Software* 2.11 (2017), p. 205.
- [64] Leland McInnes and John Healy. “Accelerated Hierarchical Density Based Clustering”. In: *Data Mining Workshops (ICDMW), 2017 IEEE International Conference on*. IEEE. 2017, pp. 33–42.
- [65] Sam Fletcher and Md Islam. “Comparing sets of patterns with the Jaccard index”. In: *Australasian Journal of Information Systems* 22 (Mar. 2018). DOI: 10.3127/ajis.v22i0.1538.

- [66] Peter J. Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53 –65. ISSN: 0377-0427. DOI: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). URL: <http://www.sciencedirect.com/science/article/pii/0377042787901257>.
- [67] Charles R Qi et al. “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation”. In: *arXiv preprint arXiv:1612.00593* (2016).
- [68] JMaosheng Tang. “Semantic Classification of 3D Point Cloud by Random Fortest-based Feature Learning”. MA thesis. Technische Universität Berlin, Germany, 2019.
- [69] Yassid Ayyad et al. “Novel particle tracking algorithm based on the Random Sample Consensus Model for the Active Target Time Projection Chamber (AT-TPC)”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 880 (2018), pp. 166 – 173. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2017.10.090>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900217311798>.
- [70] Charles R Qi et al. “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”. In: *arXiv preprint arXiv:1706.02413* (2017).
- [71] Yue Wang et al. *Dynamic Graph CNN for Learning on Point Clouds*. 2018. DOI: 10.48550/ARXIV.1801.07829. URL: <https://arxiv.org/abs/1801.07829>.
- [72] Ling Zhang and Zhigang Zhu. *Unsupervised Feature Learning for Point Cloud by Contrasting and Clustering With Graph Convolutional Neural Network*. 2019. DOI: 10.48550/ARXIV.1904.12359. URL: <https://arxiv.org/abs/1904.12359>.
- [73] MP Kuchera et al. “LISE++ Software Updates and Future Plans”. In: *Journal of Physics: Conference Series* 664.7 (2015), p. 072029. DOI: 10.1088/1742-6596/664/7/072029. URL: <https://doi.org/10.1088/1742-6596/664/7/072029>.
- [74] William R Leo. “Techniques for nuclear and particle physics experiments”. In: *Nucl Instrum Methods Phys Res* 834 (1988), p. 290.

- [75] Stefano Carboni et al. “Particle identification using the (DELTA)E-E technique and pulse shape discrimination with the silicon detectors of the FAZIA project”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 664 (Feb. 2012), 251–263. DOI: [10.1016/j.nima.2011.10.061](https://doi.org/10.1016/j.nima.2011.10.061).
- [76] Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., 2017, 3149–3157. ISBN: 9781510860964.
- [77] Ze-Peng Gao et al. “Machine learning the nuclear mass”. In: *Nuclear Science and Techniques* 32.10 (2021), p. 109. ISSN: 2210-3147. DOI: [10.1007/s41365-021-00956-1](https://doi.org/10.1007/s41365-021-00956-1). URL: <https://doi.org/10.1007/s41365-021-00956-1>.
- [78] R Utama, Wei-Chia Chen, and J Piekarewicz. “Nuclear charge radii: density functional theory meets Bayesian neural networks”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.11 (2016), p. 114002. DOI: [10.1088/0954-3899/43/11/114002](https://doi.org/10.1088/0954-3899/43/11/114002). URL: <https://doi.org/10.1088/0954-3899/43/11/114002>.
- [79] Z. M. Niu et al. “Predictions of nuclear β -decay half-lives with machine learning and their impact on r-process nucleosynthesis”. In: *Phys. Rev. C* 99 (6 2019), p. 064307. DOI: [10.1103/PhysRevC.99.064307](https://doi.org/10.1103/PhysRevC.99.064307). URL: <https://link.aps.org/doi/10.1103/PhysRevC.99.064307>.
- [80] M.P. Kuchera et al. “Machine learning methods for track classification in the AT-TPC”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 940 (2019), pp. 156–167. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2019.05.097>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900219308046>.
- [81] Christoph Dalitz, Tilman Schramke, and Manuel Jeltsch. “Iterative Hough Transform for Line Detection in 3D Point Clouds”. In: *Image Processing On Line* 7 (2017). <https://doi.org/10.5201/ipl.2017.208>, pp. 184–196.