Uniwersytet Jagielloński



# Rozróżnianie elektronów i $\pi^0$ w ciekłoargonowej komorze projekcji czasowej przy użyciu sieci neuronowych

Tomasz Wąchała

## Praca Magisterska

wykonana w Zakładzie XIII Instytutu Fizyki Jądrowej im. H. Niewodniczańskiego w Krakowie

> pod kierunkiem dr Michała Markiewicza

Kraków, czerwiec 2005

#### Streszczenie

W pracy dokonano analizy danych Monte-Carlo za pomocą oprogramowania z wstępnej fazy projektu eksperymentu T2K. Badano efektywność rozpoznawania kaskad elektromagnetycznych pochodzących od elektronów i pionów za pomocą algorytmów sieci neuronowych. Rozróżnianie  $e^-$  i  $\pi^0$  było przeprowadzone z użyciem informacji o stratach energii na jonizację na pierwszych kilku drutach, a następnie poprawiono wyniki analizy dodając informację o topologii kaskad. Przeprowadzono także analizę z uwzględnieniem położenia wierzchołka pierwotnego oddziaływania.

# Spis treści

W	stęp		<b>2</b>
1	Kas	kady elektromagnetyczne	6
	1.1	Ogólna charakterystyka kaskad elektromagnetycznych	6
	1.2	Oddziaływanie fotonów i elektronów z materią	$\overline{7}$
		1.2.1 Promieniowanie hamowania (Bremsstrahlung)	$\overline{7}$
		1.2.2 Produkcja par elektron - pozyton	7
		1.2.3 Straty energii na jonizację	8
		1.2.4 Efekt Comptona	10
	1.3	Prosty model rozwoju kaskady elektromagnetycznej	11
	1.4	Cechy rozwoju podłużnego i poprzecznego kaskad elektromagne-	
		tycznych	12
<b>2</b>	$\operatorname{Ciel}$	kłoargonowa komora projekcji czasowej	16
	2.1	Zasada działania ciekłoargonowej komory projekcji czasowej	16
	2.2	Zastosowanie TPC w eksperymentach neutrinowych z długa baza	
		pomiarowa	18
		2.2.1 Wstep	18
		2.2.2 Eksperyment ICARUS	20
		2.2.3 Projekt T2K	22
3	Siec	i Neuronowe	26
	3.1	Wstep	26
	3.2	Budowa i działanie sieci neuronowych	27
		3.2.1 Neuron	27
		3.2.2 Sieć neuronowa	29
		3.2.3 Perceptron wielowarstwowy	30
	3.3	Nadzorowane uczenie sieci	31
		3.3.1 Wstep	31
		3.3.2 Algorytm wstęcznej propagacji delta	32
		3.3.3 Inne algorytmy uczenia	35
		g - <i>j</i>	

	3.4	Sieci neuronowe w fizyce wysokich energii	36					
		3.4.1 Klasyfikacja za pomocą sieci neuronowych	36					
		3.4.2 Krzywa czystość-wydajność	37					
		3.4.3 Modyfikacje czystości dla małej liczby przypadków $\ . \ . \ .$	39					
4	Roz	różnianie elektronów i mezonów $\pi^0$	41					
	4.1	Wstęp	41					
	4.2	Przygotowanie danych Monte-Carlo	42					
	4.3	Rozróżnianie kaskad w oparciu o straty energii na jonizację na pierw-						
		szych drutach	46					
	4.4	Analiza przy użyciu zmiennych poprzecznych	57					
	4.5	Wykorzystanie informacji o położeniu wierzchołka oddziaływania . $.$	67					
<b>5</b>	Dys	skusja Wyników	74					
Po	Podsumowanie 77							

# Wstęp

Fakt istnienia neutrina został zapostulowany przez Wolfganga Pauliego w 1930 roku [18]. Wprowadzenie nowej czastki miało wyjaśnić problem ciagłego widma elektronów w rozpadach  $\beta$ . Rozpady, którymi zajmował się Pauli miały postać:  $n \to p + e^- + \bar{\nu}_e$  oraz  $p \to n + e^+ + \nu_e$ . W swoim słynnym liście Pauli zapostulował istnienie cząstki elektrycznie obojętnej, o spinie równym 1/2, którą nazwał neutronem. Po odkryciu przez Jamesa Chadwicka w 1932 roku prawdziwego neutronu, zmieniono nazwę hipotetycznej cząstki Pauliego na neutrino. Enrico Fermi, który zaproponował nową nazwę stworzył teorię oddziaływań słabych, która pozwoliła stwierdzić jak hipotetyczna cząstka Pauliego będzie oddziaływać z materią. Okazało się, że neutrino o energii 2.5 MeV musiałoby przelecieć  $2.5 \times 10^{20} cm$  wody aby oddziałało. Wielu fizyków, w tym nawet sam Pauli, wobec powyższego faktu uważało, że neutrino już na zawsze pozostanie cząstką, której istnienia nie będzie można zweryfikować. Jednak w 1956 roku F. Reines i C.L. Cowan Jr dokonali pierwszej obserwacji neutrina elektronowego  $\nu_e$  [18]. W ich eksperymencie użyto neutrin pochodzących z reaktora jądrowego w Savannah River, a detektor eksperymentu składał się z 400 litrów wody z rozpuszczonym w niej chlorkiem kadmu. Schemat poszukiwania oddziaływań neutrin opierał się na identyfikacji obu cząstek w stanie końcowym w reakcji:  $\bar{\nu}_e + p \rightarrow n + e^+$ . W 1995 roku połowa Nagrody Nobla przyznana została Reinesowi, między innymi za eksperymentalny dowód na istnienie antyneutrina elektronowego.

Po odkryciu mionu przez Andersona, badania wskazywały, że mion rozpada się na elektron i towarzyszące mu dwa neutrina. Powstało pytanie, czy oba neutrina pochodzące z reakcji są tego samego rodzaju. W 1962 roku eksperyment akceleratorowy w Brookhaven rozwiązał ten problem. Za odkrycie neutrina mionowego  $\nu_{\mu}$  w tym eksperymencie Lederman, Schwartz i Steinberger otrzymali w 1988 roku Nagrodę Nobla.

Naturalnym następstwem odkrycia leptonu  $\tau$  w SLACu było pytanie o kolejny rodzaj neutrina  $\nu_{\tau}$  towarzyszącego taonowi. W roku 2000 eksperyment DONUT ogłosił obserwację czterech przypadków oddziaływań takich neutrin.

Na podstawie pomiaru szerokości rezonansu  $Z^0$ , którego dokonano na akceleratorze LEP w CERNie pod Genewą w roku 1989, pokazano, że istnieją tylko trzy rodzaje neutrin:  $\nu_e, \nu_\mu, \nu_\tau$ .

Po fotonach, neutrina są najczęściej występującymi cząstkami we Wszechświecie. Rozpiętość energetyczna neutrin docierających do Ziemi z kosmosu sięga ponad dwudziestu rzędów wielkości [17]. Najniższą energię mają tzw. neutrina reliktowe - powstałe podczas Wielkiego Wybuchu. Ich energia to około  $10^{-4}eV$ , a gęstość w przestrzeni kosmicznej to 300 na  $cm^3$ . Drugą grupę stanowią neutrina słoneczne, których dociera do powierzchni Ziemi najwięcej bo  $7 \times 10^{10}$  na  $cm^2$  na sekundę. Powstawanie neutrin w Słońcu jest tłumaczone przez ugruntowany na przestrzeni lat Standardowy Model Słońca, nad którym John N. Bahcall pracował od 1962 roku. Wśród neutrin słonecznych dominują neutrina o średniej energii poniżej 0.42 MeV, powstające w procesie syntezy pp<sup>-1</sup>. Maksymalna energia neutrin słonecznych to 20 MeV.

Kolejnym źródłem neutrin są wybuchy gwiazd zwanych Supernową. Średnia energia takich neutrin to 10-20 MeV. Supernowe emitują bardzo duże ilości neutrin w bardzo krótkim okresie czasu. Na przykład w roku 1987, w impulsie trwającym 12 sekund, eksperyment Kamiokande zaobserwował 12 przypadków neutrin pochodzących z wybuchu Supernowej 1987A. Niestety zjawiska wybuchu Supernowych są rzadkie - szacuje się, że częstotliwość ich występowania w naszej galaktyce, to raz na sto lat [18].

Bardzo szeroką grupą w zakresie energii są tzw. neutrina atmosferyczne. Powstają one w wyniku oddziaływania promieniowania kosmicznego z ziemską atmosferą, głównie z rozpadów mionów i pionów. Średnia energia neutrin atmosferycznych to 0.4 GeV. Pochodzenie neutrin najwyższych energii (powyżej TeV) jest nadal zagadką. Sugeruje się, że powstają one w źródłach punktowych w naszej galaktyce lub poza nią. Badanie tych neutrin jest wazne z punktu widzenia współczesnej astrofizyki. Astrofizycy przypuszczają, że mogą one pomóc odpowiedzieć na pytanie skąd bierze się wysokoenergetyczne promieniowanie kosmiczne docierające do Ziemi.

Ziemskie źródła neutrin to głównie procesy rozszczepienia jąder naturalnych pierwiastków promieniotwórczych. Strumień tych neutrin wynosi  $6 \times 10^6/cm^2 \cdot sek$ a średnia energia - 2.5 MeV. Wśród neutrin wytwarzanych na Ziemi są także antyneutrina reaktorowe oraz neutrina akceleratorowe. Ilość antyneutrin powstających w elektrowni jądrowej o mocy 3 GW, to około  $6 \times 10^{20}$  na sekundę. Reaktory jądrowe są także wykorzystywane jako źródła neutrin w ekperymentach badających ich oddziaływania. Neutrina akceleratorowe powstają w wyniku zderzeń wiązki protonów z tarczą. W zderzeniach takich powstają mezony  $\pi$ , których rozpady są źródłem neutrin  $\nu_{\mu}$ . Zaletą neutrin powstających w akceleratorach jest możliwość kontroli energii, składu i kierunku wiązki.

Od chwili zapostulowania istnienia neutrin przez Pauliego, odgrywają one ważną

<sup>&</sup>lt;sup>1</sup>tj. w jednej z jego reakcji:  $pp \rightarrow D + e^+ + \nu_e$ 

rolę w wielu dziedzinach fizyki. Właściwości neutrin mogą stanowić klucz do rozwiązania wielu problemów z zakresu budowy i ewolucji Wszechświata oraz fizyki wysokich energii.

Ponieważ neutrina są stabilnymi cząstkami neutralnymi ich detekcja jest możliwa jedynie przez pomiar produktów reakcji, które wywołują. Obecnie prowadzone eksperymenty neutrinowe skoncentrowane są na badaniach oscylacji neutrin [17]. Dokładniejsze omówienie tej tematyki znajduje się w rozdziale 2.2.1.

Detektory neutrin mają na celu rejestrację oddziaływań neutrin, poprzez ich oddziaływanie z materią. Identyfikacja rodzaju neutrina, biorącego udział w reakcji z wymianą prądów naładowanych (*Charged Current*), odbywa się na podstawie analizy stanu końcowego oddziaływania. Tło dla powyższych zdarzeń stanowią, między innymi, oddziaływania z wymianą prądów neutralnych (*Neutral Current*), w których w stanie końcowym pojawiają się w detektorze neutralne mezony  $\pi^0$ , rozpadające się na parę fotonów:  $\pi^0 \to \gamma \gamma^2$ . W przypadkach CC, w stanie końcowym, pojawia się naładowany lepton. Jeżeli leptonem tym jest elektron lub taon (rozpadający się, po przebyciu znikomej drogi, na elektron i dwa neutrina), w detektorze pojawia się kaskada elektromagnetyczna. Podobną sygnaturę ma oddziaływanie NC, gdzie cząstki  $\gamma$  z rozpadu  $\pi^0$  inicjują kaskady elektromagnetyczne. Identyfikacja cząstki ( $e^-$  lub  $\pi^0$ ), która zainicjowała kaskadę, odbywa się poprzez analizę topologiczną i kinematyczną przypadku.

Niniejsza praca poświęcona jest zagadnieniu rozróżniania elektronów i mezonów  $\pi^0$  obecnych w oddziaływaniach CC i NC w ciekłym argonie. Badane przypadki to dane Monte-Carlo wygenerowane za pomocą oprogramowania dla ciekłoargonowego detektora T2K bazującego na środowisku *Geant4*. W pracy zostało przeanalizowanych kilka sposobów, które mogłyby podnieść jakość rozpoznawania. Analiza danych została przeprowadzona w Zakładzie XIII Instytutu Fizyki Jądrowej w Krakowie.

 $<sup>^2 {\</sup>rm Prawdopodobieństwo zajścia rozpadu } \pi^0 \to \gamma\gamma$ w stosunku do innych rozpadów to  $98.798 \pm 0.032\%$ 

# Rozdział 1

## Kaskady elektromagnetyczne

### 1.1 Ogólna charakterystyka kaskad elektromagnetycznych

Kaskady elektromagnetyczne (*electromagnetic showers*), to obiekty, które powstają w wyniku lawinowej produkcji cząstek wtórnych, odbywającej się poprzez odddziaływania elektromagnetyczne. Początek takiej "lawinie" cząstek daje elektron (pozyton) lub foton, który poprzez oddziaływanie z atomami danego ośrodka powoduje powstawanie cząstek wtórnych [2].

Dla energii powyżej 10 MeV, elektrony oraz pozytony tracą swoją energię w większości poprzez promieniowanie hamowania (Rysunek 1.1) [2]. Emitowane w tym procesie fotony niosą znaczną część początkowej energii elektronu. Jeśli rozważamy natomiast fotony o tych samych energiach, to dominującym procesem oddziaływania z materią jest tutaj produkcja par  $e^+e^-$ , w wyniku której otrzymujemy następne elektrony lub fotony. W kaskadzie elektromagnetycznej cząstki wtórne produkowane są więc poprzez konwersję  $\gamma \rightarrow e^+e^-$  oraz promieniowanie hamowania.

Produkcja cząstek wtórnych w kaskadzie trwa dopóki energia fotonów nie spadnie poniżej progu na produkcję par, a dla elektronów zaczną dominować inne procesy strat energii niż Bremstrahlung.

Aby lepiej zrozumieć zachowanie kaskad elektromagnetycznych konieczne jest poznanie kilku procesów oddziaływania cząstek z materią. W dalszej części tego rozdziału zostały przedstawione te procesy, które są kluczowe w opisie kaskad.

### 1.2 Oddziaływanie fotonów i elektronów z materią

#### 1.2.1 Promieniowanie hamowania (Bremsstrahlung)

Promieniowanie hamowania jest jednym z dwóch podstawowych procesów zachodzących w kaskadach elektromagnetycznych [2, 3]. Bremsstrahlung to fotony, które są emitowane przez elektron (lub cząstkę naładowaną posiadającą odpowiednio wysoką energię) spowolniony w polu kulobowskim jądra atomowego (Rysunek 1.1). Straty energii związane z tym procesem mogą być opisane wzorem [2]:

$$-\frac{dE}{dx} = \frac{E}{X_0} \tag{1.1}$$

gdzie E to energia elektronu, a  $X_0$  to droga radiacyjna - wielkość charakterystyczna dla danego ośrodka, która określa średnią odległość, po przebyciu której elektronowi pozostaje średnio 1/e jego początkowej energii.



Rysunek 1.1: Promieniowanie hamowania (*Bremsstrahlung*) emitowane przez elektron w polu kulombowskim jąder

Straty energii na promieniowanie hamowania dla elektronu są proporcjonalne do energii (Wzór 1.1), więc energia wypromieniowanych fotonów będzie tym większa, im większa będzie energia oddziaływującego elektronu.

#### 1.2.2 Produkcja par elektron - pozyton

Drugim ważnym dla kaskad elektromagnetycznych procesem jest produkcja par  $e^+e^-$  przez foton w reakcji  $\gamma \rightarrow e^+e^-$  [2]. Dla zajścia tego procesu konieczna jest obecność pola kulombowskiego jądra atomowego oraz przekroczenie progu energii przez foton. Próg ten wynosi [2]:

$$E_{\gamma} \ge 2m_e c^2 \tag{1.2}$$

gdzie  $m_e$ -masa elektronu, a c - prędkość światła.

Całkowity przekrój czynny na ten proces ma postać [2]:

$$\sigma_{e^+e^-} = \frac{7}{9} \frac{A}{N_A} \cdot \frac{1}{X_0}$$
(1.3)

gdzie A - liczba masowa,  $N_A$  - liczba Avogadro.

Analizując Równanie 1.3 można zauważyć, że w procesie produkcji par elektronpozyton, podobnie jak w promieniowaniu hamowania, ważną rolę odgrywa droga radiacyjna  $X_0$ . Wielkość tą można również zdefiniować jako 7/9 średniej drogi na produkcję pary  $e^+e^-$  przez wysokoenergetyczny foton.

Dla niskich energii fotonu ( $E_{\gamma} \ll 50 MeV$ ), podział energii między wyprodukowanymi: elektronem i pozytonem jest symetryczny. Asymetrie w podziale energii stają się większe, przy wyższych energiach ( $E_{\gamma} > 1 GeV$ ) [2].

#### 1.2.3 Straty energii na jonizację

Straty energii na jonizację to bardzo ważny proces w rejestracji cząstek w detektorach fizyki wysokich energii. Jest on kluczowy również w ciekłoargonowych komorach projekcji czasowej, dlatego został przedstawiony w tym rozdziale.

Cząstka naładowana oddziaływuje z materią poprzez oddziaływania elektromagnetyczne wybijając elektrony z atomów ośrodka i powodując ich jonizację. Jonizacja to oderwanie elektronów od atomów i powstanie ładunku dodatniego oraz ujemnego - elektronów jonizacji (Rysunek 1.2).



Rysunek 1.2: Mechanizm jonizacji ośrodka przez cząstkę naładowaną. Cząstka naładowana wybija elektrony z atomów ośrodka i powoduje powstanie ładunku dodatniego (zjonizowane atomy) i ujemnego (wybite elektrony)

Za pomocą przyłożonego pola elektrycznego można zebrać powstały w wyniku jonizacji ładunek i w ten sposób zarejestrować cząstkę w detektorze. Straty energii na jonizację na jednostkę długości ośrodka rządzone są formułą Bethe - Blocha [2] [3]:

$$-\frac{dE}{dx} = \kappa z^2 \cdot \frac{Z}{A} \cdot \frac{1}{\beta^2} \left[ \frac{1}{2} ln \frac{2m_e c^2 \gamma^2 \beta^2}{I^2} E_{kin}{}^{max} - \eta \right]$$
(1.4)

gdzie:

$$I = 10eV \cdot Z \tag{1.5}$$

to potencjał jonizacji ośrodka,

$$\kappa = 4\pi N r_e^2 m_e c^2 \tag{1.6}$$

 $\eta$ - parametr ekranowania,  $\beta = v/c, z$ - ładunek cząstki, Z- liczba atomowa i masowa ośrodka,  $m_e, r_e$ - masa i promień electronu, N- liczba Avogadro,  ${E_{kin}}^{max}$ - maksymalny transfer energii do elektronu w spoczynku.



Rysunek 1.3: Średnie straty energii na jonizację znormalizowane do strat energii w minimum jonizacji dla mionów, pionów, kaonów i protonów w ciekłym argonie [21].

W obszarze niskich energii cząstki jonizującej, przechodzącej przez dany ośrodek, straty energii na jonizację  $\frac{dE}{dx}$  zdominowane są przez czynnik  $\frac{1}{\beta^2}$  (Rysunek 1.3). Maleją one wraz z energią, aż do osiągnięcia minimum - tzw. *minimum jonizacji*. Dla wyższych energii następuje kolejny wzrost ze względu na dominujący wkład od członu  $ln\gamma$  (*wzrost relatywistyczny, logarytmiczny*). Dla bardzo wysokich energii następuje ostatecznie efekt wysycenia oraz plateau, zwane *plateau Fermiego* [2].

Ważną wielkością charakteryzującą dany ośrodek jest *energia krytyczna* [3]. Jest to wartość energii, dla której straty energii na jonizację i na promieniowanie hamowania są równe, czyli:

$$\frac{dE}{dx}(E_C)_{Brems} = \frac{dE}{dx}(E_C)_{Ion} \tag{1.7}$$

#### 1.2.4 Efekt Comptona

W poprzednich częściach tego rozdziału opisane zostały zjawiska, dzięki którym zachodzi produkcja cząstek wtórnych w kaskadzie elektromagnetycznej. W tej części przedstawiony zostanie proces, który dla niższych energii (rzędu 10 MeV) często zakłóca identyfikację cząstek (Rozdział 4.2).

Efektem Comptona, nazywamy rozpraszanie fotonu na quasi-swobodnym elektronie w atomie (Rysunek 1.4).



Rysunek 1.4: Schemat efektu Comptona. Foton w stanie początkowym rozprasza się na elektronie w atomie, zmieniając swój kierunek i energię oraz wybijając elektron z atomu ośrodka.

W zjawisku tym foton ulega rozproszeniu na elektronie w atomie i zmienia swoją energię oraz kierunek, w którym się porusza. Elektron, na którym foton uległ rozproszeniu zostaje wybity z atomu ośrodka. Energię rozproszonego fotonu można obliczyć korzystając ze wzoru [2]:

$$E'_{\gamma} = E_{\gamma} \frac{1}{1 + \varepsilon (1 - \cos\theta_{\gamma})} \tag{1.8}$$

gdzie  $\varepsilon$  - transfer energii,  $\theta_{\gamma}$  - kąt rozproszenia (kąt między kierunkiem pierwotnego fotonu  $\gamma$  i fotonu rozproszonego  $\gamma'$ ).

### 1.3 Prosty model rozwoju kaskady elektromagnetycznej

Główne cechy produkcji cząstek w kaskadach elektromagnetycznych można opisać używając bardzo prostego modelu [2], bazującego na założeniu że elektrony w kaskadach tracą swoją energię głównie na promieniowanie hamowania, a fotony na produkcję par elektron - pozyton (Rysunek 1.5).



Rysunek 1.5: Schemat powstawania kaskady elektromagnetycznej. Elektron inicijuje kaskadę emitując foton promieniowania hamowania na drodze t=1. Fotony odziaływują z materią produkując pary  $e^+e^-$ . t - odległość mierzona w jednostkach drogi radiacyjnej.

Zarys tego modelu [2], oraz główne parametry fizyczne kaskady elektromagnetycznej, jakie można obliczyć za jego pomocą, zostały przedstawione poniżej.

Foton o energii  $E_0$  rozpoczyna kaskadę poprzez produkcję pary  $e^+e^-$  po przebyciu drogi radiacyjnej  $X_0$ . Jeśli założy się, że energia w każdym kroku jest dzielona symetrycznie pomiędzy wyprodukowane cząstki, to na głębokości t dostajemy [2]:

$$N(t) = 2^t \tag{1.9}$$

cząstek o energii:

$$E(t) = \frac{E_0}{N(t)} = E_0 2^{-t} \tag{1.10}$$

gdzie t - odległość mierzona w jednostkach drogi radiacyjnej  $X_0$ . Produkcja cząstek wtórnych ma miejsce do momentu, gdy energia elektronów spadnie poniżej energii krytycznej  $E_C$  (Równanie 1.7) :

$$E_C = E_0 2^{-t_{max}} (1.11)$$

Powyżej progu  $t > t_{max}$  ( $E < E_C$ ) zachodzi tylko absorbcja cząstek.

Położenie maksimum kaskady, rozumianego jako położenie, dla którego liczba wygenerowanych cząstek na jednostkę długości jest największa, można obliczyć korzystając z Równania 1.11:

$$t_{max} = \frac{\ln(E_0/E_C)}{\ln 2} \propto \ln E_0 \tag{1.12}$$

Natomiast 95% cząstek kaskady znajduje się w przedziale  $[0, t_{95\%}]$ , gdzie:

$$t_{95\%} \approx t_{max} + 0.08Z + 9.6 \tag{1.13}$$

Całkowita liczba cząstek w kaskadzie wynosi:

$$N_{TOT} = \sum_{t=0}^{t=t_{max}} N(t) = \sum 2^t = 2^{t_{max}+1} - 1 \approx 2^{t_{max}+1}$$
(1.14)

i jest proporcjonalna do początkowej energii  $E_0$ , ponieważ:

$$N_{TOT} = 2 \cdot 2^{t_{max}} = 2 \frac{E_0}{E_C} \propto E_0$$
 (1.15)

Za pomocą tego prostego modelu można opisać głównie cechy jakościowe kaskady elektromagnetycznej. Jeśli jednak potrzebne są wyniki ilościowe zazwyczaj używa się bardziej skomplikowanego modelu opisu kaskady [21]. Jego elementy zostały przedstawione w następnym rozdziale.

## 1.4 Cechy rozwoju podłużnego i poprzecznego kaskad elektromagnetycznych

W tej części rozdziału omówiona zostanie ewolucja kaskad elektromagnetycznych. Omówione będą cechy ich rozwoju podłużnego (strat energii) oraz poprzecznego (kształt). W opisie zachowania kaskad elektromagnetycznych wygodnie jest używać następujących zmiennych [21, 23]:

$$t = \frac{x}{X_0} ,$$
  

$$y = \frac{E}{E_C} ,$$
(1.16)

czyli odległość mierzyć w jednostkach drogi radiacyjnej  $X_0$  (rozdział 1.2.1), a energię w jednostkach energii krytycznej  $E_C$  (Równanie 1.7).

Bardziej realistyczny niż przedstawiony w poprzednim rozdziale model rozwoju kaskady zakłada, że jej rozwój podłużny może być przybliżony poprzez równanie [23]:

$$\frac{dE}{dt} = E_0 b \frac{(bt)^{a-1} e^{-bt}}{\Gamma(a)}$$
(1.17)

Natomiast położenie maksimum kaskady wynosi [23]:

$$t_{max} = \frac{a-1}{b} = 1.0 \times (lny + C_j) \tag{1.18}$$

gdzie

$$C_{e^-} = -0.5$$
,  
 $C_{\gamma} = +0.5$  (1.19)

to stałe zależne od cząstki, która zainicjowała kaskadę, a, b - inne stałe.

W Równaniu 1.17 widac zależność strat energii na jednostkę długości od czynnika  $e^{-bt}$ , który powoduje zanikanie kaskady elektromagnetycznej wraz z jej rozwojem (Rysunek 1.6). Ze Wzoru 1.18 wynika, że położenie maksimum kaskady zależy od rodzaju cząstki, która ją zainicjowała ( $C_{e^-}$  i  $C_{\gamma}$  są różne).

Rozwój poprzeczny kaskady elektromagnetycznej odbywa się głównie poprzez rozpraszanie wielokrotne [2]. Na początku kaskada ma kształt wąskiego rdzenia, który rozszerza się wraz z jej rozwojem (Rysunek 1.6). Kształt kaskady w zmiennych poprzecznych może być przybliżony przez funkcję [23]:

$$f(r) = \frac{2rR^2}{(r^2 + R^2)^2} \tag{1.20}$$

gdzie r - odległość od osi kaskady, R - fenomenologiczna funkcja zależna od  $x/X_0$  oraz lnE. Powszechnie stosowaną jednostką w opisie rozwoju poprzecznego kaskad jest promień Moliere'a wyrażony wzorem [3]:

$$R_m = \frac{21MeV}{E_C} X_0[g/cm^2]$$
(1.21)



Rysunek 1.6: Rozwój poprzeczny i podłużny kaskady elektromagnetycznej. Odległości mierzone są w jednostkach drogi radiacyjnej  $X_0$ . Kaskada elektromagnetyczna zanika eksponencjalnie wraz z jej rozwojem podłużnym. Posiada natomiast kształt wąskiego rdzenia, który rozszerza się wraz z jej rozwojem poprzecznym [3].

Ważną informacją wynikającą z analizy kaskad elektromagnetycznych jest fakt, że w jednorodnym kalorymetrze 95% energii kaskady zawiera się w cylindrze o promieniu  $2R_m$  wokół jej osi [3]. Informacja ta często jest bardzo pomocna w rozróżnianiu kaskad elektromagnetycznych i hadronowych.

# Rozdział 2

# Ciekłoargonowa komora projekcji czasowej

## 2.1 Zasada działania ciekłoargonowej komory projekcji czasowej

Komora projekcji czasowej (ang. *Time Projection Chamber*), to typ detektora cząstek zyskujący ogromną popularność w fizyce wysokich energii. Zasada działania komory TPC [4] została przedstawiona na Rysunku 2.1. Cząstka naładowana przechodząc przez ośrodek, powoduje jego jonizację wzdłuż toru przelotu. Elektrony jonizacji dryfują do drutów anodowych pod wpływem przyłożonego pola elektrycznego, wytworzonego za pomocą dwóch elektrod. Anoda składa się z kilku płaszczyzn drutów ustawionych w stosunku do siebie pod różnymi kątami. Informacja o liczbie elektronów, które dotarły do drutów anodowych oraz czas dryfu elektronów, pozwalają jednoznacznie zrekonstruować zdarzenia w trzech wymiarach przestrzennych.

W eksperymencie T2K, podobnie jak w eksperymencie ICARUS, zdecydowano się na zastosowanie komory TPC wypełnionej ciekłym argonem. Jej znakomite właściwości pozwalają na badanie bardzo rzadkich procesów jakimi są niewątpliwie oddziaływania neutrin, czy poszukiwanie rozpadu protonu. Ciekłoargonowa komora TPC posiada wysoką granulację oraz jest bardzo precyzyjna w rekonstrukcji torów cząstek. Z tego względu, ciekłoargonowy detektor eksperymentu ICARUS nazywany jest "elektroniczną komorą pęcherzykową" <sup>3</sup>.

Zastosowanie ciekłoargonowej komory TPC pozwala na budowę detektora o dużej masie potrzebnej w badaniach neutrin. Popularność ciekłego argonu jako

<sup>&</sup>lt;sup>3</sup>Komora pęcherzykowa - hermetyczna komora wypełniona przegrzaną cieczą. Jony powstałe w wyniku jonizacji wzdłuż toru przelotu cząstki powodują wrzenie cieczy w tych punktach. Powstałe w ten sposób bąbelki można sfotografować przy odpowiednim oświetleniu.



Rysunek 2.1: Zasada działania komory TPC. Przechodząca cząstka powoduje powstanie elektronów jonizacji wzdłuż toru przelotu, które rejestrowane są na anodzie dzięki przyłożonemu polu elektrycznemu.

materiału wypełniającego komorę ma swoje źródła głównie w:

- właściwościach fizycznych argon jest gazem szlachetnym, ma dużą liczbę atomową Z, od której zależy przekrój czynny na oddziaływania neutrin; wzbudzenia atomowe, spowodowane przejściem cząstki, prowadzą do emisji szybkiego światła scyntylacyjnego (5ns), które może być użyte jako wyzwalanie dla elektroniki odczytu (tzw. samowyzwalanie).
- wsparciu ze strony przemysłowej w produkcji, przechowywaniu i czyszczeniu ciekłego argonu istnieje szereg firm zajmujących się produkcją ciekłego argonu na skalę przemysłową.

Podstawowe parametry ciekłego argonu zostały przedstawione w Tablicy 2.1.

ρ	$T_w$	$X_0$	$R_M$	dE/dx	Wion	$\mu$	Z/A	$\lambda_I$	$E_c$
$[g/cm^3]$	[°C]	[cm]	[cm]	[MeV/cm]	eV	$[\mathrm{cm}^2/\mathrm{Vs}]$	[-]	[cm]	[MeV]
1.394	-185.7	14.0	10.2	2.10	23.6	500	0.45	80	29.1

Tablica 2.1: Wielkości charakteryzujące ciekły argon:  $\rho$  – gęstość,  $T_w$  – temperatura wrzenia,  $X_0$  – droga radiacyjna,  $R_M$  – promień Moliere'a, dE/dx – straty energii na jednostkę długości dla cząstek w minimum jonizacji,  $W_{ion}$  – energia potrzebna na kreację pary elektron-jon,  $\mu$  – mobilność elektronów, Z – liczba atomowa, A – liczba masowa,  $\lambda_I$  – droga oddziaływania jądrowego,  $E_c$  – energia krytyczna. Wielkości  $X_0$ ,  $E_c$ ,  $R_M$  zostały omówione w rozdziałe 1.

## 2.2 Zastosowanie TPC w eksperymentach neutrinowych z długą bazą pomiarową

#### 2.2.1 Wstęp

Ciekłoargonowa komora TPC znalazła zastosowanie w badaniach prowadzonych w eksperymencie ICARUS oraz w projekcie detektora dla przyszłego eksperymentu T2K. Są to eksperymenty neutrinowe z tzw. długą bazą pomiarową, których głównym zadaniem jest pomiar parametrów oscylacji neutrin.

Efekt oscylacji neutrin został poraz pierwszy stwierdzony eksperymentalnie przez eksperyment SuperKamiokande w roku 1998 [18]. Było to bardzo ważne odkrycie w fizyce cząstek elementarnych - świadczy o tym fakt, że praca, w której ogłoszono to zjawisko, jest najczęściej cytowaną pracą w dziedzinie fizyki wysokich energii.

Oscylacje to efekt kwantowo-mechaniczny. Zjawisko to może zajść, gdy mamy do czynienia z cząstkami swobodnymi, które różnią się masą. Obserwowane w przyrodzie neutrina:  $\nu_e, \nu_\mu, \nu_\tau$  to stany kwantowe (zapachowe), które są kombinacjami liniowymi stanów masowych  $\nu_1, \nu_2, \nu_3$ . Na skutek faktu, że ewolucja czasowa poszczególnych stanów masowych jest różna, dochodzi do przejść pomiędzy stanami zapachowymi (mieszania). Ponieważ neutrina oscylują, masa conajmniej jednego z nich musi być różna od zera.

Jeśli założy się, że mamy do czynienia z wiązką  $\nu_{\alpha}$  o średniej energii E, oraz że możliwe są tylko przejścia neutrin  $\nu_{\alpha} \le \nu_{\beta}$  to prawdopodobieństwo takiej oscylacji <sup>4</sup> ma postać [17]:

$$P(\nu_{\alpha} \to \nu_{\beta}) = \sin^2 2\theta \sin^2 \left( 1.27 \Delta m^2 \frac{L}{E} \right)$$
(2.1)

 $<sup>^4\</sup>mathrm{Sa}$ to oscylacje próżniowe. Opis oscylacji neutrin w materii może być dużo bardziej skomplikowany.

gdzie:  $\Delta m^2$  - różnica kwadratów mas neutrin,  $\theta$  - kąt mieszania, L - odległość od źródła  $\nu_{\alpha}$  do detektora [19]. Kąt mieszania oraz różnica kwadratów mas to parametry teoretyczne, natomiast L i E to parametry doświadczalne oscylacji. Najlepszą dokładność pomiaru  $\Delta m^2$  uzyskuje się dobierając L oraz E w ten sposób, aby  $E/L \approx \Delta m^2$ . Z tego powodu, planując eksperyment dobiera się, przedstawionym wyżej sposobem, energię wiązki neutrin oraz odległość od źródła do detektora, tak aby trafić jak najbliżej aktualnie przewidywanej wartości  $\Delta m^2$ .

Wśrod neutrin, dla których zaobserwowano oscylacje, były neutrina atmosferyczne. Neutrina atmosferyczne powstają głownie z rozpadów pionów i mionów w atmosferze. W wyniku bombardowania atmosfery przez promieniowanie kosmiczne powstają hadrony, z których najwięcej jest tych najlżejszych - mezonów  $\pi$ . Pion o odpowiednio niskiej energii ulega rozpadowi, którego produktem jest lepton  $\mu$ . Miony oraz piony rozpadają się w następujący sposób:  $\pi^+ \to \mu^+ \nu_{\mu}, \ \mu^+ \to e^+ \nu_e \bar{\nu}_{\mu}$ lub:  $\pi^- \to \mu^- \bar{\nu}_{\mu}, \ \mu^- \to e^- \bar{\nu}_e \nu_{\mu}$ . Wobec tego stosunek liczby neutrin  $\nu_{\mu}/\nu_e$  powinien być równy dwa, dla energii poniżej 3 GeV <sup>5</sup>. Tymczasem w ekperymentach Kamiokande, IMB i SoudanII [18] zmierzony stosunek był bliski jedynce. Dane z eksperymentu SuperKamiokande pozwoliły na analizę niedoboru neutrin mionowych i potwierdziły ostatecznie oscylacje  $\nu_{\mu} \to \nu_e$  w sektorze neutrin atmosferycznych.

Rozpoczęty w 1968 roku radiochemiczny ekperyment prowadzony przez Davisa w kopalni Homestake, po 30 latach zbierania danych ogłosił, że strumień neutrin słonecznych jest około trzy razy mniejszy niż wynikałoby to ze Standardowego Modelu Słońca, co sugerowało oscylacje neutrin także w sektorze słonecznym. Eksperymenty SuperKamiokande a później SNO potwierdziły te sugestie. Pomiary SNO wykazały, że niedobór neutrin dotyczy tylko  $\nu_e$ , podczas gdy całkowity strumień wszystkich trzech rodzajów neutrin jest zgodny z modelem. Wobec powyższych obserwacji został przyjęty fakt, że w sektorze słonecznym zachodzą oscylacje:  $\nu_e \rightarrow \nu_{\mu,\tau}$  [18].

Eksperymenty neutrinowe, których celem jest badanie oscylacji neutrin najczęściej realizowane są przy użyciu tzw. długiej bazy pomiarowej [17]. Wykorzystuję się wspomniany już wyżej fakt, że maksimum oscylacji zachodzi dla  $E/L \approx \Delta m^2$ . Źródłem wiązki neutrin w eksperymentach z długą bazą pomiarową są zazwyczaj akceleratory cząstek ale obecne są także eksperymenty wykorzystujące reaktory jądrowe. Zazwyczaj na drodze wiązki umieszczone są dwa detektory: pierwszy położony blisko a drugi daleko od źródła neutrin. Poprzez porównanie liczby przypadków dla bliskiego i dalekiego detektora oraz widm energetycznych wiązki, możliwe jest stwierdzenie, czy wystąpiły oscylacje.

Pierwszym ekperymentem z długą bazą pomiarową był K2K, gdzie neutrina  $\nu_{\mu}$  pochodzące z akceleratora KEK wysyłano do oddalonego o 250 kilometrów

 $<sup>^5</sup>$ Dla wyższych energii mionów współczynnik ten jest większy niż 2, ponieważ nie wszystkie miony zdążą się rozpaść przed dotarciem do detektora.

wodnego detektora Czerenkowa - SuperKamiokande. K2K udało się zmierzyć parametry oscylacji neutrin atmosferycznych  $\nu_{\mu} \rightarrow \nu_{\tau}$ . Użycie detektora SuperKamiokande jest planowane także w planowanym eksperymencie neutrinowym drugiej generacji - T2K [6], którego dokładniejsze omówienie znajduje się w Rozdziale 2.2.3.

#### 2.2.2 Eksperyment ICARUS

Detektor eksperymentu ICARUS (Imaging Cosmic And Rare Underground Signals) znajduje się w największym na świecie podziemnym laboratorium w Gran Sasso we Włoszech (1400 metrów pod powierzchnią Ziemi). Laboratorium to jest położone w pobliżu tunelu autostrady Rzym - Teramo. Docelowo detektor eksperymentu ma posiadać masę 1800 ton (T1800) i składać się z dwóch mniejszych modułów - T600 oraz T1200 [4, 5].

Moduł T600 składa się z dwóch kriostatów, zawierających po 300 ton ciekłego argonu każdy. Kriostaty to prostopadłościany o wymiarach  $3.5 \times 4 \times 20$  metrów (Rysunek 2.2). W każdym z nich znajdują się dwie komory TPC.



Rysunek 2.2: Detektor T600. Na rysunku widać dwa kriostaty, z których każdy przedzielony jest na dwie części katodą [5].

Pole elektryczne uzyskiwane jest za pomocą katody, umieszczonej w środku kriostatu oraz anody. Przy każdej ze ścian znajdują się po trzy płaszczyzny drutów anodowych, które dają informację o współrzędnych - *Induction1, Induction2* 

oraz Collection. Płaszczyzny Induction rejestrują ładunek wyindukowany przez elektrony jonizacji. Płaszczyzna Collection zbiera docierające do niej elektrony. Najbliżej katody położona jest płaszczyzna z drutami poziomymi, natomiast dwie pozostałe płaszczyzny mają druty nachylone pod kątem  $\pm 60^{\circ}$  do poziomu. Różnica potencjałów między katodą i anodą wynosi 75kV.

Dryfujące elektrony indukują prąd na drutach komory, który jest następnie wzmacniany za pomocą elektroniki i przetworzony do postaci cyfrowej przez przetwornik *FADC*<sup>5</sup>. Ponieważ elektronika odczytu mierzy sygnał na drutach w funkcji czasu (moduł *TDC*<sup>6</sup>) z próbkowaniem równym 400 ns, obrazowanie może być traktowane jako ciągłe. Dane cyfrowe są następnie przetwarzane na dwuwymiarowy obraz w skali szarości (Rysunek 2.3), gdzie jedną współrzędną jest numer drutu, a drugą czas dryfu. Dwuwymiarowe obrazowanie przeprowadzane jest dla każdej płaszczyzny. Wykorzystanie informacji zarejestrowanej przez dwie płaszczyzny daje możliwość rekonstrukcji prostych przypadków. Bardziej złożone zdarzenia wymagają wykorzystania wszystkich trzech płaszczyzn. Użycie trzech płaszczyzn pozwala na bardzo dobrą rekonstrukcję przestrzenną - najmniejszy element trójwymiarowego obrazu zdarzenia (komórka) ma wymiary  $3 \times 3 \times 0.6$  mm.



Rysunek 2.3: Obrazowanie zarejestrowanego zdarzenia w detektorze T600 - widok w płaszczyźnie Collection. Rekonstrukcja w dwu wymiarach wykorzystuje informacje o numerze drutu, na którym zarejestrowano jonizację oraz o czasie dryfu elektronów jonizacji do drutów.

T600 został zbudowany w latach 1997-2001 i przeszedł pomyślnie naziemne testy w Pavii w lecie 2001 roku. W testach tych zarejestrowano około 30000 przypadków, głównie oddziaływań promieni kosmicznych [4].

 $<sup>^5\</sup>mathrm{FADC}$  (Fast Analog to Digital Converter) - konwerter przekształcający amplitudę sygnału na dane cyfrowe

<sup>&</sup>lt;sup>6</sup>TDC (Time to Digital Conventer) - konwenter zamieniający zmierzony czas na dane cyfrowe

W programie fizycznym eksperymentu ICARUS są poszukiwania rozpadu protonu, pomiary oscylacji neutrin słonecznych, neutrin atmosferycznych oraz akceleratorowych, oraz neutrin pochodzących z wybuchów Supernowych. Źrodłem neutrin akceleratorowych w eksperymencie ICARUS będzie wiązka CNGS (Cern Neutrinos to Gran Sasso) wysyłana z instytutu CERN pod Genewą. Neutrina aby dotrzeć do laboratorium w Gran Sasso będą musiały pokonać drogę równą 732 kilometry. Jest to wiązka wysokoenergetycznych neutrin mionowych o średniej energii 17 GeV, a program jej badania jest nastawiony na zaobserwowanie oddziaływań neutrin taonowych i elektronowych powstałych w wyniku oscylacji neutrin mionowych [18].

#### 2.2.3 Projekt T2K

Projekt T2K (Tokai to Kamioka) to eksperyment neutrinowy drugiej generacji, którego główną częścią ma być istniejący już detektor SuperKamiokande [6]. Celem T2K jest użycie wiązki protonów o energii 50 GeV, pochodzącej z synchrotronu, do produkcji wiązki neutrin mionowych skierowanych w stronę detektora SuperKamiokande, a następnie rejestracja ich oddziaływań.

Wiązka neutrin o średniej energii 1 GeV ma być wysyłana z Jparc w Jaeri na wschodnim wybrzeżu Japoni. Odległość między Jparc i SuperKamiokande wynosi 295 kilometrów. Pomiędzy Jparc i SuperKamiokande planowane są cztery stanowiska z detektorami umieszczone w odległości: 140 metrów, 280 metrów, 2 kilometry i 295 kilometrów od źródła neutrin. Pierwsze stanowisko ma dawać informację o kierunku wiązki, drugie będzie mierzyć widmo energetyczne neutrin potrzebne w pomiarze oscylacji. Trzeci detektor ma na celu zmniejszenie błędów systematycznych wynikających z ekstrapolacji spektrów neutrin z bliskiego detektora (280 m) do dalekiego (295 km). Redukcja powyższych błędów może zostać osiągnięta poprzez pomiary dokonane w odległości dwóch kilometrów od Jparc, gdzie widma energetyczne neutrin będą podobne do tych z dalekiego detektora SuperKamiokande. Ostatni, oddalony o 295 kilometrów od źródła neutrin detektor, to wodny detektor Czerenkowa o masie 50 kiloton, dzięki któremu możliwe są pomiary parametrów oscylacji [6].

W odległości dwóch kilometrów od Jaeri ma znajdować się wodny detektor Czerenkowa o całkowitej masie 1000 ton, komora mionowa do pomiaru pędu mionów, oraz ciekłoargonowa komora TPC o masie 100 ton, która pozwala na badanie oddziaływań neutrin niskich energii (Rysunek 2.4).

Stutonowy detektor ciekłoargonowy mieści się w cylindrze o długości 5 metrów i średnicy 6.6 metra. Część detektora, w której znajduje się komora TPC, to wypełniony ciekłym argonem prostopadłościan o masie 150 ton (Rysunek 2.5).

Pole elekryczne uzyskiwane jest dzięki dwóm elektrodom: katodzie i anodzie,



Rysunek 2.4: Bliski detektor T2K znajdujący się dwa kilometry od źródła neutrin. Widoczne na rysunku: ciekłoargonowa komora TPC, wodny detektor Czerenkowa oraz detektor mionowy [6].

pomiędzy którymi wytwarzane jest jednorodne pole elektryczne. W środku prostopadłościanu znajduje się katoda, której kształt zależy od kształtu dodatkowej tarczy umieszczonej wewnątrz detektora (*inner target*) [6]. Dwa rozpatrywane warianty to (Rysunek 2.6):

- 1. Tarcza w kształcie cylindra wypełnionego lodem lub $CO_2$ w stanie stałym i pojedyńcza katoda.
- 2. Tarcza w kształcie płaszczyzny utworzonej z ułożonych równoleg<br/>le do siebie rur wypełnionych lodem lub $CO_2$ w stanie stałym. Ta op<br/>cja wymaga dwóch katod po obydwu stronach tarczy.

Na zewnętrznych, długich ścianach komory znajdują się dwie (lub trzy) płaszczyzny drutów anodowych (Induction i Collection) - Tablica 2.2.

Wiele rozwiązań technicznych, dotyczących zarówno budowy detektora, jak i elektroniki odczytu, a także oprogramowanie do wizualizacji i rekonstrukcji zdarzeń, bazują na doświadczeniu wypracowanym przez kolaborację eksperyment ICA-RUS [6].

Eksperyment T2K został podzielony na dwa etapy:

1. Precyzyjne pomiary oscylacji neutrin  $\nu_{\mu} \rightarrow \nu_{\tau}$ , oraz poszukiwania oscylacji  $\nu_{\mu} \rightarrow \nu_{e}$ , pochodzących z wiązki o mocy 0.75 MW przy użyciu detektora SuperKamiokande (50 kton).



Rysunek 2.5: Ciekłoargonowa komora TPC w eksperymencie T2K. Na rysunku widoczny jest cylinder, w którym znajduje się właściwy detektor - prostopadłościan wypełniony ciekłym argonem [6].



Rysunek 2.6: Zastosowanie dodatkowej tarczy neutrinowej w TPC eksperymentu T2K. Dwa warianty kształtu tarczy: a) cylinder wypełniony lodem lub  $C0_2$  w stanie stałym, b) płaszczyzna utworzona z ułożonych do siebie równolegle rur wypełnionych podobnie jak w przypadku cylindra [6].

Długość	5 m
$\operatorname{Szerokość}$	$4.5 \mathrm{m}$
Wysokość	$4.5 \mathrm{m}$
Całkowita masa	140  ton
Liczba komór	2
Liczba płaszczyzn drutowych w komorze	2(3)
Orientacja drutów w stosunku do poziomu	$\pm 45^{\circ} (90^{\circ})$
Odległość między drutami	$3 \mathrm{mm}$
Odległość między płaszczyznami drutowymi	$3 \mathrm{mm}$
Średnica drutu	$150~\mu m$

Tablica 2.2: Główne parametry geometryczne ciekło<br/>argonowej komory TPC w eksperymencie T2K

2. Pomiary z użyciem wiązki o mocy 4 MW i detektora o masie 1 Megatony (HyperKamiokande).

Rozpoczęcie pierwszego etapu pomiarów jest planowane na rok 2009.

# Rozdział 3

# Sieci Neuronowe

### 3.1 Wstęp

Nazwą *Sieci Neuronowe* określa się najczęściej układy realizujące pseudo-równoległe przetwarzanie informacji, składające się z połączonych ze sobą elementów, nazwanych poprzez analogię ze swoimi biologicznymi odpowiednikami - *neuronami*.

Mózg człowieka jest ciągle najpotężniejszym obecnie istniejącym układem przetwarzającym informacje w czasie rzeczywistym, dlatego jest on tutaj niedościgłym wzorcem.

Składa się on z ogromnej liczby prostych elementów (około  $10^{10}$  neuronów). Jest to układ przetwarzający równolegle, a informacja zakodowana jest w *połączeniach synaptycznych* (~  $10^{13}$  synaps) [11]. Główne zalety mózgu to:

- duża wydajność
- male rozmiary
- odporność na szumy i uszkodzenia
- zdolność uczenia i adaptacji, uogólnianie.

Sieci neuronowe można zaliczyć do rozwiązań nazywanych AI - Artifical Intelligence (sztuczna inteligencja). Są to realizacje (sprzętowe lub programowe) urządzeń o właściwociach podobnych do ludzkiego mózgu i próbujących naśladować inteligencję człowieka. Stanowią one dziedzinę będącą na pograniczu statystyki, fizyki i neurobiologii. Wszystkie te dziedziny mają swój wkład zarówno w rozwój algorytmów, jak również pozwalają lepiej opisać działanie sieci. Przykładami takiego wpływu, wywodzącego się z fizyki mogą być: algorytm uczenia zwany maszyną Boltzmana [11] oparty na metodach mechaniki statystycznej, czy opis sieci Hopfielda <sup>7</sup> [11] pochodzący z teorii szkieł spinowych.

Dzisiejsze, tak szerokie zainteresowanie sieciami neuronowymi, wynika przede wszystkim z poszukiwań bardziej wydajnych i niezawodnych urządzeń służących do przetwarzania informacji. Również ze względu na swój nieliniowy charakter oraz możliwość odtworzenia bardzo złożonych funkcji, stanowią one duże wyzwanie dla obecnych metod modelowania nieliniowego.

Zastosowania sieci neuronowych to między innymi [11]:

- Aproksymacja (interpolacja, ekstrapolacja) ze znajomości wartości funkcji w dyskretnych punktach:  $\{x_i, f(x_i)\}$  odtworzyć wartość funkcji w dowolnym punkcie  $\{x_j, f(x_j)\}$ .
- Przewidywanie (predykcja) znając wartość funkcji f w punktach  $\{x_{n-k}, x_{n-k+1}, ..., x_n\}$  odtworzyć  $f(x_{n+1})$ .
- Klasyfikacja zaszeregowanie wektora danych wejściowych do jednej z kilku klas.
- Pamięć asocjacyjna podanie wzorca na wejściu powoduje pojawienie się skojarzonego z nim wzorca na wyjściu.
- Sterowanie sterowanie urządzeniami przemysłowymi (roboty).

### 3.2 Budowa i działanie sieci neuronowych

#### 3.2.1 Neuron

Podstawową jednostką, z której zbudowany jest mózg jest neuron (Rysunek 3.1).

Sygnał do neuronów jest doprowadzany za pomocą *dendrytów*, których neuron może posiadać bardzo dużo. Synapsa, to "furtka" neuronu, która może zmienić moc sygnału napływającego poprzez jego wejście - dendryt. Jądro neuronu, to "centrum obliczeniowe", w którym zachodzą procesy kluczowe dla funkcjonowania komórki nerwowej. Sygnał wyprowadzany jest przez *akson*, za pomocą którego neuron powiadamia świat zewnętrzny o swojej odpowiedzi na dane wejściowe. Neuron posiada tylko jeden akson.

Schemat sztucznego neuronu został opracowany w 1943 roku przez McCullocha i Pittsa [14] w oparciu o budowę biologicznej komórki nerwowej (Rysunek 3.2).

 $<sup>^7\</sup>mathrm{Sie\acute{c}}$  Hopfielda - jeden z rodzajów sieci rekurencyjnych używany często do rozpoznawania obrazów.



Rysunek 3.1: Budowa biologicznego neuronu. Dendryty doprowadzają sygnał do jądra neuronu. Moc doprowadzanego sygnału jest regulowana przez synapsy. W jądrze neuronu następuje przetwarzanie sygnału, który jest wyprowadzany na zewnątrz za pomocą aksonu [25].



Rysunek 3.2: Schemat sztucznego neuronu. Na rysunku widoczne są: wejścia neuronu  $x_1, ..., x_p$ , współczynniki wagowe  $w_{k_1}, ..., w_{k_p}$ , sumator  $\Sigma$  oraz element aktywujący  $\varphi$ . Widoczna jest analogia pomiędzy budową sztucznego neuronu i neuronu biologicznego: wejścia  $x_i$  - odpowiadają dendytom, wagi  $w_i$  - synapsom, natomiast wyjście oznaczone jako  $y_k$ - aksonowi.

Na wejście neuronu podawane są sygnały wejściowe  $\{x_i\}$ , które następnie są mnożone przez odpowiednie współczynniki wag  $\{w_{ki}\}$  (opowiedniki połączeń synaptycznych w biologicznym neuronie). Następnie, ważone sygnały są sumowane [8]:

$$u_k = \sum_i w_{ki} x_i \tag{3.1}$$

Na podstawie wartości  $u_k$  wyznacza się aktywność neuronu [11]:

$$y_k = \varphi(u_k - \theta_k) \tag{3.2}$$

gdzie $\theta_k$ to wartość progu, <br/>a $\varphi$ to tzw.  $funkcja \ aktywacji,$ która może mieć postać

• funkcji skoku jednostkowego:

$$\varphi(u) = sgn(u) \tag{3.3}$$

• funkcji liniowej:

$$\varphi(u) = au + b \tag{3.4}$$

• funkcji sigmoidalnej:

$$\varphi(u) = \frac{1}{1 + e^{-\beta u}} \tag{3.5}$$

$$\varphi(u) = tanh(\beta u) \tag{3.6}$$

Szczególnym przypadkiem neuronu jest *neuron stochastyczny*, którego funkcja aktywacji to prawdopodobieństwo, że na wyjściu pojawi się wartość 1. Funkcja aktywacji informuje nas dla jakich wartości ważonej sumy sygnałów wejściowych na wyjściu neuronu pojawi się sygnał.

Wybór funkcji aktywacji (Rysunek 3.3) zależy od charakterystyki danego problemu oraz od architektury sieci.

#### 3.2.2 Sieć neuronowa

Poprzez odpowiednie połączenie wielu neuronów otrzymujemy tzw. warstwy, które następnie możemy połączyć w sieć (Rysunek 3.4). W sieci neuronowej możemy rozróżnić trzy rodzaje warstw neuronów":

1. warstwa wejściowa



Rysunek 3.3: Funkcje aktywacji neuronu [25].

- 2. warstwy ukryte
- 3. warstwa wyjściowa

W zapisie architektury sieci wielowarstwowych stosuje się nomenklaturę w której liczba neuronów w kolejnych warstwach jest oddzielona myślnikiem. Przykładowo, zapis 2-4-1 oznacza więc sieć o dwóch neuronach w warstwie wejściowej, czterech neuronach w warstwie ukrytej i jednym neuronie w warstwie wyjściowej. Sieci dzielimy również ze względu na kierunek przesyłania informacji [14]:

- Sieci jednokierunkowe są to sieci, w których informacja przesyłana jest tylko w jednym kierunku od warstwy wejściowej do wyjściowej (Rysunek 3.4).
- Sieci rekurencyjne wartości wyjściowe tej sieci podawane są z powrotem na wejścia z pewnym opóźnieniem czasowym (Rysunek 3.5).

#### 3.2.3 Perceptron wielowarstwowy

Największym zainteresowaniem cieszą się perceptrony (MLP - MultiLayer Perceptron), czyli sieci wielowarstwowe, jednokierunkowe, posiadające conajmniej jedną warstwę ukrytą neuronów (Rysunek 3.4). W perceptronie nie ma połączeń między neuronami tej samej warstwy, a połączenia między warstwami zachowują uporząd-kowanie, czyli połączone są w następującej kolejności: warstwa wejściowa, warstwy ukryte i warstwa wyjściowa. Najczęściej, dla perceptronu używa się jako funkcji aktywacji funkcji skoku jednostkowego (Równanie 3.3) a dla neuronów w ostatniej warstwie funkcji liniowej (Równanie 3.4).



Rysunek 3.4: Sieć wielowarstwowa jednokierunkowa. Na rysunku widoczne są od lewej: warstwa wejściowa, warstwa neuronów ukrytych i warstwa wyjściowa. Kierunek przepływu informacji został zaznaczony na rysunku czerwoną strzałką.



Rysunek 3.5: Sieć wielowarstwowa rekurencyjna. Sygnał na wyjściu podawany jest z powrotem na wejścia neuronów sieci.

### 3.3 Nadzorowane uczenie sieci

### 3.3.1 Wstęp

Jedną z najważniejszych cech, które przyczyniły się do popularności sieci neuronowych jest umiejętność uczenia się i uogólniania nabytej wiedzy. W procesie uczenia pierwszą czynnością jest przygotowanie dwóch ciągów danych: uczącego i weryfikującego (testującego). Ciąg uczący powinien w miarę dokładnie charakteryzować określony problem. Jednorazowa porcja danych nazywa się wektorem uczącym. W skład wektora uczącego wchodzą: wektor (wzorzec) wejściowy - czyli dane, które podawane są na wejścia sieci oraz wektor (wzorzec) wyjściowy - dane, jakie sieć powinna wygenerować na wyjściach.

Dany jest ciąg uczący: wzorców wejściowych:  $\{x^{\mu}\} = \{x_1^{\mu}, x_2^{\mu}, ..., x_k^{\mu}\}$  oraz wzorców wyjściowych:  $\{\varsigma^{\mu}\} = \{\varsigma_1^{\mu}, ..., \varsigma_p^{\mu}\}$ , gdzie: k - ilość wejść, p - ilość wyjść,  $\mu$  - numer wzorca ( $\mu = 1, 2, ..., L$ ). O(x) to odpowiedź sieci na podany wzorzec x. Uczenie sieci polega na podaniu na wejście sieci wzorców wejściowych i wyjściowych w celu zminimalizowaniu funkcji błędu. Definicja funkcji błędu (celu) ma postać błędu średniokwadratowego [11]:

$$E(w) = \sum_{\mu=1}^{L} \sum_{i=1}^{p} \left(\varsigma_{i}^{\ \mu} - O_{i}(x^{\mu})\right)^{2}$$
(3.7)

Proces uczenia można traktować jako proces minimalizacji funkcji wielu zmiennych w przestrzeni N+1 wymiarowej, gdzie N wymiarów to współczynniki wagowe sieci, natomiast N+1-szy wymiar stanowi wartość funkcji błędu E(w) dla danego zestawu wag [12]. W ten sposób każdy ze stanów sieci podczas procesu uczenia może być traktowany jako punkt na N-wymiarowej hiperpłaszczyźnie, a N+1-szy wymiar jako wysokość ponad tą hiperpłaszczyzną. Wszystkie możliwe wartości współczynników wagowych oraz funkcji E (jest ich nieskończenie wiele) tworzą tzw. hiperpowierzchnię błędu. Każdy krok w procesie uczenia to jeden punkt na wspomnianej hiperpowierzchni. Celem, do którego dązy się ucząc sieć jest osiągnięcie minimum globalnego powierzchni błędu E(w).

Minimalizacja funkcji wielu zmiennych, to znany problem z analizy numerycznej inaczej zwana *optymalizacją*, dla którego został opracowany szereg metod rozwiązania. Z tego powodu istnieje także wiele sposobów minimalizacji funkcji błędu E(w). Wybór metody minimalizacji zależy od charakterystyki danego problemu.

#### 3.3.2 Algorytm wstecznej propagacji delta

W procesie uczenia wejścia sieci aktywowane są wektorem wejściowym, a sygnały przetwarzane są przez wszystkie neurony. Wektor wejściowy wybiera się zazwyczaj losowo z ciągu uczącego, przy czym każdy wektor może być wylosowany tylko raz. Losowo generowane są także początkowe wartości współczynników wa-gowych. Po przetworzeniu wzorca wejściowego, porównywane są wartości otrzymane:  $O^{\mu}$  z wartościami oczekiwanymi:  $\varsigma^{\mu}$ , a następnie obliczane są błędy na każdym wyjściu sieci.

W celu obliczenia błędu neuronów we wszystkich warstwach, propaguje się błędy w odwrotnej kolejności - czyli od warstwy wyjściowej do wejściowej. Jest to *algorytm propagacji wstecznej* (Rysunek 3.6).

Błąd neuronu oblicza się w następujący sposób [11]:


Rysunek 3.6: Uczenie metodą propagacji wstecznej. Czarną strzałką zaznaczono kierunek przepływu informacji, a strzałką czerwoną kierunek propagacji błędu.

• w warstwie wyjściowej:

$$\delta_{iN} = \varphi'_{iN}(\varsigma_{iN} - O_{iN}) \tag{3.8}$$

gdzie N - ilość warstw sieci.

• w pozostałych warstwach:

$$\delta_{ij} = \varphi'_{ij} \sum_{i} (\delta_{ij+1} w_{ij+1}) \tag{3.9}$$

Jak wynika z Wzoru 3.9 błąd  $\delta$  na neuronie (i, j) zależy od wartości pochodnych funkcji aktywacji każdego neuronu warstwy poprzedniej (o numerze j+1) i wartości wag na połączeniach.

Na podstawie błędów oblicza się korektę wag w każdym neuronie w celu zmniejszenia błędu odpowiedzi [11]:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \tag{3.10}$$

gdzie

$$\frac{\partial E}{\partial w_{ab}^n} = \delta_a^n O_b^{k+1} \tag{3.11}$$

gdzie a,b - konkretne neurony, n - numer warstwy.

Nowe wartości wag w(t+1) oblicza się w następujący sposób [11]:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$
(3.12)



Rysunek 3.7: Wpływ współczynnika  $\eta$  na proces uczenia. Przebieg procesu uczenia polega na znalezieniu minimum na hiperpowierzchni błedu. Dłuższe strzałki to większy współczynnik uczenia (dłuższy krok) - wtedy możliwe jest ominięcie minimum globalnego funkcji błędu. Krótsze strzałki to optymalny dobór współczynnika uczenia - wartość, przy której minimum globalne nie może być ominięte oraz czas uczenia jest możliwie najkrótszy.

gdzie t - aktualny krok uczenia. Procedurę powtarza się do momentu, gdy błąd sieci będzie mniejszy niż pewna założona wartość. Wtedy na wejście podaje się kolejny wektor wejściowy i powtarza czynności. Przetworzenie całego ciągu uczącego nazywa się *epoką*. Dla całej epoki oblicza się następnie błąd, którego spadek poniżej dopuszczalnej wartości powoduje zakończenie procesu uczenia.

Współczynnik  $\eta$  to tzw. współczynnik uczenia, który wpływa na szybkość procesu uczenia. Zbyt duża jego wartość wiąże się ze zbyt dużą wartością kroku przy korekcji wag, co może prowadzić do pominięcia minimum funkcji błędu. Zbyt mała wartość może spowodować utknięcie w minimum lokalnym (Rysunek 3.7). Najlepszym rozwiązaniem jest dynamiczna zmiana współczynnika  $\eta$  w zależności od postępu procesu uczenia, czyli odległości od poszukiwanego minimum globalnego funkcji błędu.

Aby jeszcze bardziej usprawnić uczenie sieci dodaje się do korekty wag człon bezwładności [11]:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} + \alpha \Delta w_{ij}(t-1)$$
(3.13)

który ma uchronić od gwałtownych zmian wartości wag.  $\alpha$  to tzw. współczynnik momentum. Dodanie członu bezwładności powoduje zależność zmiany wag w bieżącym cyklu od zmian jakie wystąpiły w cyklu poprzednim, co pomaga w opuszczeniu płytkich minimów lokalnych i poprawia efektywność uczenia w przedziałach, gdzie funkcja 3.7 jest monotoniczna.

#### 3.3.3 Inne algorytmy uczenia

Klasyczna metoda propagacji wstecznej bazuje na *metodzie największego spadku*. Jej schemat wygląda następująco:

- wyliczamy gradient  $\nabla E(w)$
- minimalizujemy funkcję E(w) w kierunku gradientu czyli szukamy minimum funkcji  $f(\lambda) = E(\lambda \nabla E + w)$
- robimy krok  $\lambda \nabla E$ .

Ma ona jednak trzy poważne wady: utyka w płytkich minimach lokalnych oraz wymaga ciągłej zmiany współczynnika uczenia i jest wolno zbieżna. Problemy te zostały częściowo rozwiązane poprzez zastosowanie innych odmian metod gradientowych lub przeszukiwania strategicznego. Najczęściej stosowane algorytmy to:

#### 1. Metoda gradientów sprzężonych

Algorytm polega na minimalizacji funkcji błędu w ten sposób, aby nie zaprzepaścić osiągniętego w uprzednich krokach kierunku optymalizacji [12]. Można to osiągnąć utrzymując w następnych krokach drugą pochodną wyznaczoną w tym kierunku na poziomie zerowym. Aby utrzymać zerową drugą pochodną minimalizacja E(w) przebiega w kierunku [11]:

$$d(t) = -\nabla E(w(t)) + \beta d(t-1)$$
(3.14)

gdzie:

$$\beta = \frac{(\nabla E(w(t)) - \nabla E(w(t-1)))\nabla E(w(t))}{\nabla^2 E(w(t-1))}$$
(3.15)

#### 2. Quickprop

Ustalenie kierunku wzdłuż wagi  $w_{ij}$  i minimalizacja funkcji  $E(w_{ij}) = E(w + w_{ij})$  przybliżonej parabolą [11].

#### 3. Resillent Backpropagation (Rprop)

Jest to prosty algorytm zaproponowany przez Riedmillara i Brauna [16]. Uwzględnia się w nim przy zmianie wag jedynie znak składowej gradientu, pomijając jej wartość. Współczynnik uczenia jest dobierany w każdym cyklu dla każdej wagi indywidualnie: jeśli w dwóch kolejnych krokach znak gra-dientu jest taki sam, współczynnik  $\eta$  jest zwiększany, natomiast zostaje zmniejszany, jeśli znak ten jest różny.

#### 4. Metody Monte Carlo

Generujemy losowo nowy zestaw wag  $w^*$  i jeśli  $E(w^*) < E(w)$ , to  $w = w^*$ 



Rysunek 3.8: Sieć neuronowa jako klasyfikator w fizyce wysokich energii. Na wejście sieci podawane są parametry charakteryzujące sygnał oraz tło. Wyjście sieci neuronowej reprezentuje przynależność danego przypadku do klasy sygnału (odpowiedź równa 1) lub tła (odpowiedź równa 0).

#### 5. Symulowane wyżarzanie

Strategia postępowania jest podobna do metody Monte-Carlo, z tym zastrzeżeniem, że jeśli  $E(w^*) > E(w)$ , akceptujemy nowy zestaw wag z prawdopodobieństwem:

$$P(T) \propto exp(-\frac{\Delta E}{T})$$
 (3.16)

gdzie T - temperatura, której wysoką wartość ustalamy na początku, a następnie sukcesywnie obniżamy [11].

# 3.4 Sieci neuronowe w fizyce wysokich energii

#### 3.4.1 Klasyfikacja za pomocą sieci neuronowych

Sieci neuronowe w fizyce wysokich energii stosowane są głównie jako klasyfikatory - czyli ich zadaniem jest przyporządkowanie danego wektora  $\mathbf{x}$  do jednej z M klas:  $K_1, K_2, ..., K_M$ :

$$[x_1, x_2, ..., x_n] \longrightarrow \mathbf{KLASYFIKATOR} \longrightarrow K_i$$

Najczęsciej spotykanym problemem jest zagadnienie sygnał-tło, gdzie ilość klas M = 2. Klasa  $K_1$  to sygnał, natomiast klasa  $K_2$  - tło (Rysunek 3.8).

W fizyce wysokich energii często bada się bardzo rzadkie zjawiska (rzędu kilku zdarzeń na rok), jak chociażby analiza oddziaływań neutrin, czy łamanie symetrii CP. W analizie danych z detektora podczas badania rzadkich zjawisk kluczową sprawą jest wydobycie interesujących przypadków z bardzo dużej ilości zdarzeń, w których przeważają zjawiska mało interesujące z punktu widzenia eksperymentu. Klasyfikacja typu sygnał-tło polega na rozróżnieniu interesujących przypadków (sygnał) od tych mniej ważnych (tło). Przykładem bardzo rzadkiego procesu fizycznego może być produkcja bozonu Higgsa podczas zderzenia proton-proton<sup>8</sup>. Do wykrycia rzędu 10<sup>3</sup> cząstek Higgsa na rok, potrzeba aby zdarzenie występowało  $10^{34}$  razy na sekundę a prawdopodobieństwo innych procesów jest wyższe o czynnik  $10^{13}$ . W ciągu roku zostanie wygenerowane  $10^9$  na sekundę zdarzeń będących tłem. Do oddzielenia interesujących zdarzeń od tła używa sie tzw. *dyskryminatorów* - urządzeń elektronicznych, w których z powodzeniem można zastosować sieci neuronowe, ze względu na ich szybkość i równoległe działanie.

W praktyce klasyfikacja sygnał-tło za pomocą sieci neuronowych polega na znalezieniu parametrów charakteryzujących sygnał oraz tło a następnie nauczeniu sieci na pewnej ilości przypadków, podając na jej wejścia wartości tychże parametrów. Jeśli zbiór uczący w wystarczający sposób charakteryzował globalne cechy problemu, sieć powinna dawać poprawne wyniki klasyfikacji także innych przypadków. Sieci neuronowe są używane z powodzeniem zarówno przy obróbce danych w czasie rzeczywistym, jak i w analizie "off-line", gdzie szybkość działania nie ma już tak wielkiego znaczenia. Kluczową rolę odgrywają tutaj nieliniowe właściwości sieci neuronowych oraz możliwości bardzo efektywnej, wielowymiarowej analizy złożonych funkcji.

#### 3.4.2 Krzywa czystość-wydajność

Wykorzystanie tzw. krzywej czystość-wydajność (*purity-efficiency plot*) jest jednym ze sposobów weryfikacji jakości klasyfikacji [15]. Jest to metoda najczęściej stosowana w przypadkach, gdy mamy do czynienia z problemem dwuklasowym, gdzie naszym celem jest jak najbardziej wydajne oddzielenie przypadków sygnału od tła (Rysunek 3.9).

Zmienne, które definiuje się na potrzeby takiej analizy to [15]:

• Czystość próbki (purity):

$$czystosc = 100\% \cdot \frac{N_{syg}(Wyjscie)}{N_{tlo}(Wyjscie) + N_{syg}(Wyjscie)}$$
(3.17)

• Wydajność (efficiency):

<sup>&</sup>lt;sup>8</sup>Eksperymentem, który dedykowany jest poszukiwaniu bozonu Higgsa w zderzeniach protonproton jest eksperyment ATLAS na akceleratorze LHC, którego start planowany jest na 2007 rok.



Rysunek 3.9: Klasyfikacja sygnał-tło. Zdarzenia powyżej progu traktowane są jako sygnał, poniżej jako tło.

$$wydajnosc = 100\% \cdot \frac{N_{syg}(Wyjscie)}{N_{syg}(Wejscie)}$$
(3.18)

gdzie  $N_{syg}, N_{tlo}$ , to ilości sygnału i tła w danym zbiorze. Wejscie to zbiór danych, które klasyfikujemy, Wyjscie to zbiór z odpowiedziami sieci powyżej zadanego progu.

Zmieniając wartość progu dyskryminacji uzyskujemy różne wartości czystości i wydajności, które układają się w tzw. krzywą czystość-wydajność. Używanie tej krzywej jest bardzo wygodnym sposobem oceny jakości klasyfikacji. Jeśli krzywa czystość-wydajność położona jest wyżej na wykresie - jakość klasyfikacji sieci neuronowej jest lepsza (Rysunek 3.10) [15].

Główny przyczynek do niepewności pomiarowych przy analizie typu czystośćwydajność pochodzi z błędu czystości, który definiuje się jako [15]:

$$\sigma_{czystosc} = \sqrt{\sigma_{N_{syg}(Wyjscie)}^2 \left(\frac{\partial czystosc}{\partial N_{syg}(Wyjscie)}\right)^2 + \sigma_{N_{tlo}(Wyjscie)}^2 \left(\frac{\partial czystosc}{\partial N_{tlo}(Wyjscie)}\right)^2}$$
(3.19)

gdzie  $\sigma_{N_{syg}(Wyjscie)}$  i  $\sigma_{N_{tlo}(Wyjscie)}$  to fluktuacje ilości sygnału i tła powyżej ustalonego progu:

$$\sigma_{N_{syg}(Wyjscie)} = \sqrt{N_{syg}(Wyjscie)}$$

$$\sigma_{N_{tlo}(Wyjscie)} = \sqrt{N_{tlo}(Wyjscie)}$$
(3.20)

Użycie w analizie błędu czystości jest wygodnym kryterium oceny, czy zbiór przypadków, na którym sieć jest uczona, ma wystarczającą liczebność. Często zdarza się sytuacja, że zbiór uczący nie reprezentuje globalnych własności całego problemu. Powodem tego jest często zbyt mała liczba przypadków w próbce uczącej niewystarczająca aby wypełnić w dostateczny sposób wielowymiarową przestrzeń parametrów wejściowych. Niebezpieczeństwo jest większe jeśli liczba parametrów wejściowych jest duża - potrzebna jest wtedy bardzo duża liczebność zbioru uczącego. Problem słabej generalizacji sieci <sup>9</sup> jest widoczny wyraźnie jeśli dokona się porównania krzywych czystość-wydajność dla zbioru uczącego i dowolnego zbioru testowego. Jeśli krzywe dla zbioru testowego i uczącego nie pokrywają się w granicach błędów, oznacza to, że liczebność zbioru uczącego jest niewystarczająca.

#### 3.4.3 Modyfikacje czystości dla małej liczby przypadków

W przypadku niskich ilości sygnału w próbce potrzebne są dodatkowe modyfikacje w technice klasyfikacji i uczenia sieci neuronowej. Utrzymanie proporcji sygnał-tło w próbce na poziomie rzeczywistych wartości wymagałoby użycia bardzo dużego zbioru uczącego. Zbiór ten musiałby zawierać ilość sygnału wystarczającą aby wypełnić wielowymiarową przestrzeń i kilka razy większą liczbę przypadków tła. Wygodnym sposobem ominięcią tego problemu jest użycie porównywalnej ilości sygnału i tła w zbiorze uczącym i przewidywanie wyników klasyfikacji dla rzeczywistych proporcji sygnał-tło. W tym celu modyfikuje się względną liczbę sygnału i tła w zbiorze przeskalowywując  $N_{syg}$  i  $N_{tlo}$ , a następnie obliczając czystość próbki w następujący sposób [15]:

$$czystosc = 100\% \cdot a \cdot \frac{N_{syg}(Wyjscie)}{b \cdot N_{tlo}(Wyjscie) + a \cdot N_{syg}(Wyjscie)}$$
(3.21)

$$a = \frac{real}{testing} ,$$
  
$$b = \frac{1 - real}{1 - testing}$$
(3.22)

gdzie *testing* - proporcja sygnału w zbiorze wejściowym, *real* - rzeczywista wartość proporcji sygnał/tło. Na Rysunku 3.10 został przedstawiony przykład zastosowania powyższej modyfikacji.

W dwóch pierwszych częściach niniejszej pracy zostały przedstawione podstawy fizyczne kaskad elektromagnetycznych oraz krótka charakterystyka eksperymentów neutrinowych, w których rejestruje się kaskady jako efekty oddziaływań cząstek,

<sup>&</sup>lt;sup>9</sup>Generalizacja - proces uogólniania. Sieć, której zdolności generalizacji są dobre, nauczona na pewnym zbiorze uczącym, daje poprawne wyniki na dowolnym zbiorze danych testowych



Rysunek 3.10: Krzywa czystość-wydajność dla 55% sygnału (elektronów) w próbce oraz przeskalowana do 12% sygnału. Dla 12% sygnału sieć wykazuje gorszą jakość klasyfikacji - krzywa czystość-wydajność położona jest niżej na wykresie niż dla 55%.

będących w stanie końcowym procesów z udziałem neutrin. W części trzeciej scharakteryzowane zostały sieci neuronowe - jedne z najbardziej efektywnych narzędzi do klasyfikacji oddziaływań w fizyce wysokich energii, używane także do rozróżniania kaskad elektromagnetycznych. W ostatniej części pracy przedstawio-na zostanie analiza danych Monte-Carlo dla ciekłoargonowego detektora eksperymentu T2K, wykonana za pomocą sieci neuronowych.

# Rozdział 4

# Rozróżnianie elektronów i mezonów $\pi^0$ w ciekłoargonowym detektorze eksperymentu T2K - analiza danych Monte-Carlo

### 4.1 Wstęp

Jednym z celów eksperymentu T2K jest pomiar parametrów oscylacji neutrin mionowych ( $\nu_{\mu} \rightarrow \nu_{e}$ ). Przy użyciu ciekłoargonowego detektora TPC można dokonać analizy wiązki neutrinowej, której skład ilościowy pozwoli wyznaczyć poszukiwane parametry.

Głównym celem rozróżniania elektronów i pionów w detektorze jest odróżnienie od siebie dwóch reakcji:

•  $\nu_e$  z wymianą prądów naładowanych (CC):

$$\nu_e + n \to e^- + p \tag{4.1}$$

•  $\nu_{\mu}$  z wymianą prądów neutralnych (NC):

$$\nu_{\mu} + n \to \nu_{\mu} + \pi^0 + X \tag{4.2}$$

które w kanale wyjściowym dają kaskady elektromagnetyczne. Znalezienie cech charakterystycznych, które pozwolą rozróżnić kaskadę zainicjowaną przez lepton  $e^-$  od kaskad pochodzących od fotonów, z rozpadu  $\pi^0$ , jest tematem tej części pracy.

Zapach neutrina i typ oddziaływania	Liczba oddziaływań
$\nu_e (CC)$	3704 (12%)
$\nu_{\mu} (\text{NC})$	26253 (88%)
Suma	29957~(100%)

Tablica 4.1: Przewidywane liczby oddziaływań zarejestrowane przez ciekłoargonowy detektor eksperymentu T2K, o masie 100 ton, w jednostkach  $10^{10}/cm^2/10^{21}p.o.t.$ 

Niniejsza analiza została przeprowadzona na danych Mote-Carlo dla ciekłoargonowej komory TPC eksperymentu T2K. Przewidywane ilości oddziaływań CC i NC dla stutonowego detektora zostały przedstawione w Tablicy 4.1 [6].

W dalszej części pracy oddziaływania elektronów zostały nazwane sygnałem, natomiast mezonów  $\pi^0$  tłem. Ponieważ przewidywana w eksperymencie zawartość sygnału w zbiorze danych, zawierających oba wyżej wymienione typy oddziaływań wynosi 12%, w analizie za każdym razem przedstawiane są wyniki klasyfikacji przeskalowane <sup>6</sup> do takiej ilości sygnału w próbce testowej.

#### Przygotowanie danych Monte-Carlo 4.2

Dane do analizy zostały przygotowane za pomocą generatora Monte-Carlo opartego o środowisko Geant4 - G4T2K w wersji 1.3. Geometria detektora zaimplementowana w programie została przedstawiona na Rysunku 4.1.

Przyjęty został tutaj wariant detektora z cylindrem wypełnionym zamrożoną wodą w roli tarczy neutrinowej [6]. Druty anodowe umieszczone są w płaszczyźnie równoległej do osi z i płaszczyzny x = 0, w pobliżu ścian długich detektora. W każdej z płaszczyzn drutowych mamy do czynienia z 2414 drutami. Obraz przypadku w danej płaszczyźnie ma wymiary  $2414 \times 5302$  punktów, gdzie ostatnia liczba to ilość próbek czasowych, które odpowiadają czasowi dryfu elektronów jonizacji do drutów anodowych (Rysunek 4.2).

Wygenerowane cząstki są monoenergetyczne i mają energie 1 GeV. Przypadki zostały wygenerowane bez tła - czyli bez dodatkowych cząstek w stanie początkowym oraz bez szumu elektroniki. Rysunek 4.3 przedstawia dwa typowe zdarzenia dla: elektronu (4.3a) i pionu (4.3b), zobrazowane na płaszczyźnie Collection, wygenerowane dla detektora T2K.

Cząstki skierowane zostały wzdłuż osi z pod kątami:  $+45^{\circ}$  i  $-45^{\circ}$  w stosunku do drutów płaszczyzn Induction i Collection (Rysunek 4.4). Jest to ten sam kierunek,

<sup>&</sup>lt;sup>6</sup>Idea skalowania została wyjaśniona w Rozdziale 3.4.3



Rysunek 4.1: Geometria ciekłoargonowego detektora T2K zaimplementowana w programie G4T2K oraz wygenerowana kaskada elektromagnetyczna pochodząca od pionu o energii 1 GeV.



Rysunek 4.2: Przykład zobrazowanego przypadku w detektorze T2K. Wymiary obrazu to 2414 drutów oraz 5302 próbek czasowych.

jaki będzie miała wiązka, w stosunku do detektora w eskperymencie T2K.



Rysunek 4.3: Przykłady kaskad elektromagnetycznych wygenerowanych przy użyciu programu G4T2K, pochodzących od elektronu (a) i pionu (b) o energii 1GeV, w ciekłoargonowym detektorze T2K. Obrazowanie w płaszczyźnie Collection



Rysunek 4.4: Położenie kierunku generowanych cząstek w stosunku do kierunku drutów anodowych.

Do rekonstrukcji zdarzeń zostało użyte oprogramowanie, będące rozwinięciem pakietu Qscan, stworzonego przez kolaborację ICARUS, a w szczególności jego części: biblioteka Fullreco w wersji 1.3 oraz odpowiednio zmodyfikowany program Qbatch, który służy do rekonstrukcji zdarzeń w trybie wsadowym.

Poniżej została przedstawiona anglojęzyczna terminologia, którą stosuje się w rekonstrukcji zdarzeń. Będzie ona używana w dalszej części niniejszej analizy.

- 1. Hit fragment toru, który posiada jednoznacznie określone współrzędne w danej płaszczyźnie drutowej - czyli numer drutu i numer próbki czasowej (czas dryfu), lub współrzędne w przestrzeni, oraz amplitudę zarejestrowanego sygnału.
- 2. Klaster (Cluster) zbiór kolejno po sobie następujących hitów

Rekonstrukcja przypadków polega na znalezieniu za pomocą odpowiednich algorytmów, mając do dyspozycji dane Monte-Carlo (lub rzeczywiste dane z detektora), wszystkich hitów, oraz torów cząstek w płaszczyznach drutowych i w przestrzeni.

Podczas przygotowywania danych zostały wygenerowane zbliżone ilości elektronów i mezonów  $\pi^0$ . Charakterystyka całości wygenerowanych przypadków została przedstawiona w Tablicy 4.2.

	Dane
Całkowita liczba przypadków	10325~(100%)
Liczba elektronów	5555~(53.8%)
Liczba $\pi^0$	4770 (46.2%)
Energia [GeV]	1

Tablica 4.2: Charakterystyka wygenerowanych danych Monte-Carlo.

Do analizy za pomocą sieci neuronowych została użyta biblioteka MLP będąca częscią pakietu ROOT. Zaimplementowano w niej obsługę perceptronu wielowarstwowego, wraz z metodami uczenia opisanymi w Rozdziale 3.

Dane, podawane na wejścia sieci w procesie uczenia, dzielone są na dwie próbki: uczącą i testową. Zakończenie procesu uczenia następuje po dobranej przez użytkownika liczbie epok. Zbiór uczący został przygotowany w ten sposób, aby wraz z wartościami wejściowymi były podawane sieci także docelowe wartości wyjściowe, potrzebne przy procesie uczenia nadzorowanego (Rozdział 3). W Tablicy 4.3 zostały przedstawione podstawowe parametry wybranych zbiorów: uczącego i testującego.

	Zbior uczący	Zbiór testowy
Całkowita liczba przypadków	6000 (100%)	4325 (100%)
Liczba elektronów	3163 (52.7%)	2392~(55%)
Liczba $\pi^0$	2837(47.3%)	1933~(45%)
Energia [GeV]	1	1

Tablica 4.3: Charakterystyka zbioru uczącego i testowego.

Błędy czystości oraz przeskalowania w niniejszej analizie obliczane były zgodnie ze Wzorami 3.19 oraz 3.21.

# 4.3 Rozróżnianie kaskad w oparciu o straty energii na jonizację na pierwszych drutach

W celu znalezienia cech różniących kaskady elektromagnetyczne pochodzące od elektronów i pionów potrzebne jest wydobycie informacji o jonizacji ośrodka jeszcze przed rozpoczęciem kaskady. Zanim rozpocznie się kaskada, sygnał pochodzący od pary  $e^+e^-$ , powstałej w wyniku konwersji fotonu z rozpadu  $\pi^0 \to \gamma\gamma$ , powinien być dwa razy większy niż sygnał pochodzący od pojedyńczego elektronu. Fakt ten jest motywacją do użycia informacji o stratach energii na jonizację na kilku pierwszych drutach do rozpoznawania elektronów i pionów.

Za początek kaskady był przyjmowany pierwszy sygnał jonizacyjny, który pojawił się wzdłuż kierunku osi z, a który należał do toru posiadającego conajmniej 15 hitów. Taki wybór eliminuje większość krótkich torów pochodzących od elektronów Comptonowskich (praca [1] oraz Rozdział 1). Drut, na którym wspomniany wyżej sygnał został zarejestrowany, to drut numer jeden. Numery drutów liczone są kolejno od tego punktu wzdłuż toru przelotu cząstki.

Rozkłady strat energii na jonizację na kolejnych sześciu drutach w płaszczyźnie Collection, zostały przedstawione na Rysunku 4.5.

Ponieważ foton pochodzący z rozpadu  $\pi^0$  może ulec konwersji na parę  $e^+e^$ pomiędzy drutami numer 1 i 2, część jonizacji może być zarejestrowana przez pierwszy drut. Fakt ten jest źródłem rozmytego rozkładu dE/dx na pierwszym drucie dla pionu. Także na drugim drucie można zauważyć dużo przypadków niskoenergetycznych  $\pi^0$ , które mogą być spowodowane częsciowo błędną rekonstrukcją zdarzenia. Mogą one zakłócić identyfikację. Z tych właśnie powodów dwa pierwsze druty zostały wykluczone z analizy (Rysunek 4.5).

Można zauważyć, że rozkłady strat energii na jonizację coraz bardziej się nakładają wraz ze wzrostem numeru drutu (Rysunek 4.6), dlatego użycie w analizie zbyt dużej liczby drutów także jest niewskazane.

W celu dobrania optymalnej liczby drutów, które należy wziąc pod uwagę przy analizie, przetestowanych zostało dziewięć kolejnych drutów (od 3 do 11). Testy zostały przeprowadzone za pomocą prostej sieci o architekturze 1-2-1, na której wejście podawane były średnie z wartości strat energii na jonizację na odpowiedniej liczbie drutów w płaszczyźnie Collection:

$$\left\langle \frac{dE}{dx} \right\rangle_N = \frac{1}{N} \sum_{i=3}^N \left( \frac{dE}{dx} \right)_i \tag{4.3}$$



Rysunek 4.5: Rozkłady strat energii na jonizację na sześciu kolejnych drutach dla elektronu (linia ciągła) i pionu (linia przerywana) w płaszczyźnie Collection



Rysunek 4.6: Rozkłady strat energii na dalszych drutach dla elektronu (linia ciągła) i pionu (linia przerywana) w płaszczyźnie Collection. Na wykresie widać spadek separacji krzywych wraz ze wzrostem numeru drutu.

gdzie  $\left(\frac{dE}{dx}\right)_i$  to straty energii zarejestrowane na drucie numer *i*, *N* ilość drutów.

Sygnał na neuronie wyjściowym tej sieci symbolizuje rodzaj cząstki: 1 - odpowiada elektronowi, 0 - mezonowi  $\pi^0$ . Cykl uczenia wynosił 200 epok. Krzywe czystość-wydajność dla tej analizy zostały przedstawione na Rysunku 4.7.



Rysunek 4.7: Krzywe czystość-wydajność dla różnej liczby drutów (N) wziętej pod uwagę przy obliczaniu  $\langle dE/dx \rangle$ . Użyto sieci 1-2-1 (200 epok). Dolna część rysunku przedstawia zbliżenie interesującego obszaru. Najwyżej położona na wykresie jest krzywa dla trzech drutów.

Na wykresie, który pokazuje obszar wydajności powyżej 85% widać wyraźnie, że najwyżej położona jest krzywa dla N = 3 drutów, użytych przy obliczaniu średniej < dE/dx >.

Zatem optymalna zmienna wejściowa ma postać:

$$\left\langle \frac{dE}{dx} \right\rangle_3 = \frac{1}{3} \sum_{i=3}^5 \left( \frac{dE}{dx} \right)_i \tag{4.4}$$

Dla powyższej zmiennej uzyskane wyniki zostały przedstawione w Tablicy 4.4.

Wydajność [%]	Czystość [%]	Błąd czystości [%]
94.06	83.61	0.71
90.05	95.99	0.41

Tablica 4.4: Wyniki klasyfikacji za pomocą sieci 1-2-1 z użyciem średnich strat na jonizację z drutów numer 3,4,5 w płaszczyźnie Collection, przy 55% udziale elektronów w próbce testowej.

Osiągnięty wynik: 95.99% czystości próbki, przy 90.05% wydajności jest podobny do rezultatów osiągniętych w pracy [1], gdzie jako metod klasyfikacji użyto prostych cięć na  $\langle dE/dx \rangle$  oraz metody logarytmicznej funkcji podobieństwa (*lo*glikehood method). Autorzy osiągnęli tam 96.3% wydajności odrzucania mezonów  $\pi^0$  (odpowiadającej czystości) przy 90% wydajności rozpoznawania elektronów. Różne są natomiast liczby drutów, przy użyciu których uzyskano te wyniki. W niniejszej analizie wykorzystano trzy z nich, podczas, gdy w pracy [1] zostało ich użytych osiem. Widać jednak (Rysunek 4.7), że dla wydajności rzędu 90%, różnice pomiędzy krzywymi dla różnej ilości drutów, są minimalne ( $\sim$  kilku procent). Oprócz 90% poziomu wydajności, w dalszej analizie badano także jakość klasyfikacji dla 94%, aby sprawdzić czy przy wyższej wartości wydajności można osiągnąć wyniki przynajmniej porównywalne do rezultatów pracy [1]. Zarówno wysoka wartość wydajności, która odpowiada liczbie poprawnie rozpoznanych elektronów, jak i czystości, odpowiadającej procentowej zawartości elektronów w próbce sklasyfikowanej przez sieć jako sygnał, jest istotna dla jakości klasyfikacji. Najbardziej pożądaną byłaby zatem sytuacja, gdy przy podniesionej wydajności otrzymalibyśmy podobną czystość.

Kolejną możliwością poprawy jakości klasyfikacji zdarzeń jest wykorzystanie wycałkowanych przebiegów sygnału pochodzących z drutów płaszczyzny Induction. Wyżej wspomniane całki oznaczone zostały jako *sygnal* (Rysunek 4.8).

Do dalszej analizy wybrano trzy kolejne druty z płaszczyzny Induction (3-5), wykorzystując wynik jaki uzyskano dla płaszczyzny Collection (Rysunek 4.9).

Rysunek 4.10 pokazuje krzywe czystość-wydajność uzyskane przy użyciu sieci 1-2-1, gdzie na wejście podawana jest tylko średnia wartość  $\langle dE/dx \rangle$  w płaszczyźnie Collection, oraz dla przypadku, gdy klasyfikatorem jest sieć 2-2-1,



Rysunek 4.8: Rozkłady wycałkowanych przebiegów sygnału na pierwszych drutach (od 3 do 6) dla elektronu (linia ciągła) i pionu (linia przerywana) w płaszczyźnie Induction



Rysunek 4.9: Rozkłady średnich strat energii na jonizację na trzech drutach dla elektronu (linia ciągła) i pionu (linia przerywana) w płaszczyźnie Collection oraz średniego sygnału na odpowiadających im trzech drutach płaszczyzny Induction. Druty brane pod uwagę podczas obliczania  $\langle dE/dx \rangle_{col}$  i  $\langle sygnal \rangle_{ind}$  miały numer od 3 do 5.

a na wejście podawana jest także druga zmienna - średnia wartość sygnału w płaszczyźnie Induction.



Rysunek 4.10: Krzywe czystość-wydajność dla sieci 1 - 2 - 1, oraz 2 - 2 - 1, na których wejścia podawano odpowiednio: tylko sygnały z płaszczyzny Collection, oraz sygnały z płaszczyzny Collection i Induction. Dolna część rysunku przedstawia zbliżenie interesującego obszaru. Na wykresie wyraźnie widać poprawę jakości klasyfikacji po użyciu drugiej zmiennej (krzywa jest położona wyżej na wykresie).

Widoczna jest wyraźna poprawa jakości klasyfikacji po dodaniu drugiej zmiennej wejściowej. Wyniki liczbowe zostały przedstawione w Tablicy 4.5.

Po dodaniu informacji o średnim sygnale w płaszczyźnie Induction, czystość próbki uległa poprawie z 83.61% do 88% przy tej samej wydajności rzędu 94%. Konfiguracja 2-2-1 okazała się także optymalna pod względem liczby neuronów w warstwie ukrytej. Zwiększenie liczby neuronów w tej warstwie nie prowadzi do

Zmienne wejściowe	Wydajność [%]	Czystość [%]	Błąd czystości [%]
$dE/dx >_{col}$	94.06	83.61	0.71
$\overline{\langle dE/dx \rangle_{col} + \langle sygnal \rangle_{ind}}$	94.11	88.0	0.64

Tablica 4.5: Wyniki klasyfikacji za pomocą sieci 1 - 2 - 1 (tylko  $\langle dE/dx \rangle_{col}$  jako zmienna wejściowa), oraz 2 - 2 - 1 ( $\langle dE/dx \rangle_{col}$  oraz  $\langle sygnal \rangle_{ind}$  jako zmienne wejściowe).

poprawy jakości klasyfikacji. Krzywe dla różnej ilości neuronów w warstwie ukrytej pokrywają się w granicach błędów (Rysunek 4.11).



Rysunek 4.11: Krzywe czystość-wydajność dla różnej liczby neuronów (N) w warstwie ukrytej dla sieci 2 - x - 1. Nie ma widocznej poprawy jakości klasyfikacji po dodaniu dodatkowych neronów w warstwie ukrytej - krzywe pokrywają się w granicach błędów.

Rozkład odpowiedzi sieci na zbiór danych testowych został przedstawiony na Rysunku 4.12.

Z analizy powyższego rozkładu wynika, że pojawiają się elektrony oraz mezony  $\pi^0$ , które zostały błędnie sklasyfikowane a odpowiedź sieci w ich przypadku jest zupełnie przeciwna niż powinna być, tzn. dla elektronów jest bliska zeru, a dla  $\pi^0$  jedynce. Fakt ten świadczy o tym, że informacja o stratach energii na jonizację w płaszczyźnie Collection oraz o sygnale z płaszczyzny Induction, nie są wystarczające do jednoznacznej identyfikacji cząstek. Potrzebne jest zastosowanie dodatkowych kryteriów aby poprawić zarówno wydajność rozpoznawania jak i czystość próbki. Zachowanie krzywych czystość-wydajność dla zbioru uczącego i testowego



Rysunek 4.12: Rozkład odpowiedzi sieci 2 - 2 - 1. Widoczne są elektrony, dla których odpowiedź sieci jest bliska zeru, oraz mezony  $\pi^0$  dla których sieć dała odpowiedź bliską jedynce.

(Rysunek 4.13) świadczy o tym, że ilość wygenerowanych przypadków w zbiorze testowym i uczącym była wystarczająca - wykresy dla zbioru uczącego i testowego pokrywają się w granicach błędów w badanym obszarze (do 94% wydajności).



Rysunek 4.13: Wykres czystość-wydajność dla sieci 2-2-1. Krzywe dla zbioru uczącego i testowego pokrywają się w granicach błędów dla wydajności mniejszej od 95%.

Krzywe czystość-wydajność obrazujące jakość klasyfikacji dla sieci2-2-1po

przeskalowaniu ilości sygnału w próbce do 12% (Rozdział 3.4.3) zostały przedstawione na Rysunku 4.14.



Rysunek 4.14: Krzywe czystość-wydajność dla zbioru testowego dla 55% sygnału w próbce oraz przeskalowane do ilości sygnału przewidywanej w eksperymencie T2K równej 12%.

Wyniki analizy, w której wykorzystano informację o stratach energii na jonizację w płaszczyźnie Collection oraz o sygnale w płaszczyźnie Induction na drutach od 3 do 5, dla 55% udziału elektronów w próbce, oraz przeskalowane do 12%, zostały przedstawione w Tablicy 4.6.

Ilość elektronów w próbce	Architektura sieci	Wydajność [%]	Czystość [%]
55%	2-2-1	94.11	$88.0 \pm 0.64$
12%	2-2-1	94.11	$44.69 \pm 1.50$

Tablica 4.6: Wyniki klasyfikacji zdarzeń za pomocą sieci 2-2-1, uzyskane przy użyciu informacji o stratach energii na jonizację na pierwszych drutach w płaszczyźnie Collection i średniej wartości sygnału w płaszczyźnie Induction. Do znalezienia wartości czystości dla ilości sygnału równej 12%, użyto przeskalowania omówionego dokładniej w Rozdziale 3.4.3

Dla 55% udziału elektronów w próbce około 94.11 % elektronów zostało poprawnie rozpoznanych. Powyżej ustalonego progu dyskryminacji znalazło się 88%  $e^-$ , natomiast około 12% mezonów  $\pi^0$  zostało źle sklasyfikowanych jako elektrony. Po przeskalowaniu do liczby oddziaływań, jaka jest spodziewana w ciekłoargonowym detektorze T2K (12% sygnału w próbce), uzyskujemy już bardzo duży spadek czystości do 44.69%.

Aby porównać osiągnięte wyniki z rezultatami osiągniętymi w pracy [1], w Tablicy 4.7 przedstawiono także wyniki dla 90% wydajności.

Wydajność [%]	Czystość [%]	Błąd czystości [%]
94.11	88.0	0.64
90.01	97.16	0.35

Tablica 4.7: Ostateczne wyniki klasyfikacji zdarzeń za pomocą sieci 2-2-1, uzyskane przy użyciu informacji o stratach energii na jonizację na trzech kolejnych drutach w płaszczyźnie Collection i średniej wartości sygnału w płaszczyźnie Induction.

Widoczna jest wyraźna poprawa w stosunku do analizy przeprowadzonej przez autorów noty [1]. Osiągnięto tutaj 97.16% czystości próbki przy 90.01% wydajności, podczas gdy we wspomnianej wyżej pracy uzyskano 96.3% czystości przy tej samej wartości wydajności.

Podsumowując należy zauważyć, że wyniki analizy przeprowadzonej w tej części pracy stanowią optymistyczną prognozę do dalszych badań. Pokazują one, że dodając dodatkowe parametry charakteryzujące przypadki oddziaływań elektronów i mezonów  $\pi^0$  (zmienna  $\langle sygnal \rangle_{ind}$ ) można poprawić jakość ich klasyfikacji.

## 4.4 Analiza przy użyciu zmiennych poprzecznych

Wyniki klasyfikacji uzyskane w poprzednim rozdziale są zadowalające, jednak nadal pozostaje 12% mezonów  $\pi^0$ , które nie zostały prawidłowo rozpoznane. Ta część pracy będzie dotyczyć prób poprawienia wydajności klasyfikacji, poprzez zastosowanie zmiennych wynikających z analizy topologicznej kaskad, czyli ich kształtu oraz rozwoju poprzecznego i podłużnego.

Analiza topologiczna kaskad została dokonana pod kątem znalezienia różnic między kaskadami pochodzącymi od  $e^-$  i od  $\pi^0$ . W oparciu o dane zrekonstruowane za pomocą programu Qbatch zostały zdefiniowane zmienne, które charakteryzują topologię kaskad. Podczas pracy zostało przebadanych szereg parametrów, które miały charakteryzować kaskady pochodzące od elektronów i mezonów  $\pi^0$ . W wyniku analizy ich rozkładów zostały wybrane odpowiednie zmienne, jako wielkości pozwalające poprawić jakość klasyfikacji (rozkłady tych wielkości dawały możliwie najlepsze rozróznienie dla  $e^-$  i  $\pi^0$ ). Wykorzystano tu wcześniejsze prace [2, 3, 6, 21], jak również kilka oryginalnych pomysłów:

$$\langle W \rangle = \frac{W_{ind} + W_{col}}{2}$$
 (4.5)

$$N_{tot} = N_{ind} + N_{col} \tag{4.6}$$

$$\langle L \rangle_{tr} = \frac{L_{tr}(Ind) + L_{tr}(Col)}{2} \tag{4.7}$$

$$<\alpha>=\sum_{i=1}^{N_{hit}}\alpha_i \cdot \left(\frac{dE}{dx}\right)_i$$
(4.8)

$$\langle r \rangle = \sum_{i=1}^{N_{hit}} r_i \cdot \left(\frac{dE}{dx}\right)_i$$

$$(4.9)$$

$$<\alpha>_{max} = max\{<\alpha>_{A_{\beta}}\}_{\beta=0}^{180^{o}}$$

$$(4.10)$$

gdzie  $W_{ind}$ ,  $W_{col}$  - szerokości kaskady w płaszczyźnie odpowiednio Induction i Collection (w jednostkach próbek czasowych);  $N_{ind}$ ,  $N_{col}$  - liczba hitów w płaszczyźnie Induction i Collection;  $L_{tr}(Ind)$ ,  $L_{tr}(Col)$  - długość zrekonstruowanego toru o największej liczbie hitów dla odpowiednich płaszczyzn (w jednostkach liczby drutów);  $\left(\frac{dE}{dx}\right)_i$  - strata energii na jonizację dla hitu numer i ( $i = 1, 2, ..., N_{hit}$ );  $\alpha_i$  kąt pomiędzy osią kaskady a prostą łączącą jej początek z danym hitem numer i (Rysunek 4.15);  $r_i$  - odległość hitu od osi kaskady (Rysunek 4.15);  $< \alpha >_{A_{\beta}}$  średni kąt w płasczyźnie  $A_{\beta}$  nachylonej pod kątem  $\beta$  do osi OX (Rysunek 4.16).



Rysunek 4.15: Parametry używane przy obliczaniu zmiennych  $\langle r \rangle$ ,  $\langle \alpha \rangle$ .

 $\langle W \rangle$ ,  $N_{tot}$  i  $\langle L \rangle_{tr}$ , to zmienne wykorzystujące tylko informację z dwuwymiarowego obrazowania przypadku w każdej z płaszczyzn. Średnią szerokość kaskady  $\langle W \rangle$  uzyskano poprzez uśrednienie szerokości w płaszczyźnie Induction i Collection, gdzie szerokości  $W_{ind}$  i  $W_{col}$  są obliczane w kierunku prostopadłym do osi kaskady.  $\langle L \rangle_{tr}$  to, uśredniona po obydwu płaszczyznach drutowych, długość toru o największej ilości hitów. Długości  $L_{tr}(Ind)$  i  $L_{tr}(Col)$  są obliczane jako odległość pomiędzy początkowym i końcowym hitem, w kierunku równoległym do osi kaskady, toru o największej liczbie hitów w odpowiedniej płaszczyźnie.



Rysunek 4.16: Zasada tworzenia zmiennej  $\langle \alpha \rangle_{max}$ . Przy obliczaniu średniego kąta  $\langle \alpha \rangle_{A_{\beta}}$  (analogicznie do zmiennej opisanej Równaniem 4.8) brane są pod uwagę tylko hity w płaszczyźnie  $A_{\beta}$  (zaznaczonej na niebiesko), która jest nachylona pod kątem  $\beta$  do osi OX.

Do analizy zostały użyte zmienne wykorzystujące rekonstrukcję przestrzenną przypadków. Zostały one omówione poniżej. Średni promień kaskady  $\langle r \rangle$  oblicza się, poprzez wyśredniowanie odległości wszystkich hitów od osi kaskady (w centymetrach), ważonych przez straty energii na jonizację (w MeV/cm). Podobnie ma się sytuacja dla średniego kąta rozwarcia kaskady  $\langle \alpha \rangle$ . Średniowanie przebiega po wszystkich hitach, a średniuje się kąty między osią kaskady a prostą przechodzącą przez hit i początek kaskady. Kąty te dodatkowo są jeszcze ważone przez straty energii na jonizację w danym punkcie.

 $<\alpha>_{max}$  to maksymalny z średnich kątów  $<\alpha>_{A_{\beta}}$ , obliczanych w płaszczyznach  $A_{\beta}$  nachylonych pod różnymi kątami  $\beta$  do osi OX. Średni kąt  $<\alpha>_{A_{\beta}}$ oblicza się podobnie do  $<\alpha>$  (Wzór 4.8), uwzględniając tylko hity leżące w płaszczyźnie  $A_{\beta}$ .

Na Rysunkach 4.17 oraz 4.18 zostały przedstawione rozkłady zmiennych wyrażonych Wzorami 4.5 - 4.10 dla zbioru uczącego.

Podczas rozpadu mezonu  $\pi^0$  produkowane są dwa fotony  $\gamma$ . Jeśli energia obydwu fotonów jest dostatecznie duża - powstają dwie kaskady. Z tego powodu sumaryczna szerokość całego przypadku powinna być większa dla  $\pi^0$  niż dla elektronu (Rysunek 4.17a). Liczba zrekonstruowanych torów  $N_{tot}$  dla pionu z tego samego powodu powinna być większa niż dla elektronu (Rysunek 4.17b). Ponieważ w przypadku elektronu cała jego energia jest zużyta na zapoczątkowanie jednej kaskady, długość toru o największej liczbie hitów  $\langle L \rangle$  powinna być większa niż dla pionu (Rysunek 4.17c). Większa szerokość kaskady pochodzącej od mezonu  $\pi^0$  jest także widoczna w pozostałych zmiennych opisujących topologię kaskady ( $\langle r \rangle$  i  $\langle \alpha \rangle$ ).

Może się zdarzyć, że płaszczyzna rozpadu  $\pi^0 \rightarrow \gamma \gamma$  będzie ułożona w stosunku do płaszczyzn drutowych w ten sposób, że zarejestrowana kaskada nie będzie wy-



Rysunek 4.17: Zmienne charakteryzujące topologię kaskad elektromagnetycznych, w których zaobserwowano różnice między  $e^-$  i  $\pi^0$ : a) szerokość kaskady w jednostkach próbek czasowych, b) całkowita liczba zrekonstruowanych torów, c) długość toru o największej liczbie hitów w jednostkach liczby drutów



Rysunek 4.18: Zmienne charakteryzujące topologię kaskad elektromagnetycznych, w których zaobserwowano różnice między  $e^-$  i  $\pi^0$ : a) średni promień kaskady, b) średni kąt rozwarcia kaskady, c) maksymalny ze średnich kątów w różnych płasz-czyznach (szczegóły w dalszej części rozdziału).

kazywać większej szerokości  $\langle W \rangle$  lub informacja ta zostanie zgubiona podczas średniowania w zmiennych  $\langle r \rangle$  i  $\langle \alpha \rangle$ . Z tego powodu została stworzona zmienna  $\langle \alpha \rangle_{max}$ . Podczas obliczania tego parametru znajduje się maksimum rozkładu średnich kątów rozwarcia w płaszczyznach nachylonych do osi OX pod kątami 0 – 180°. W ten sposób znajdywana jest płaszczyzna, w której średni kąt kaskady jest największy, będąca płaszczyzną rozpadu.

W celu zweryfikowania przydatności poszczególnych zmiennych przeprowadzono analizę polegającą na klasyfikacji zdarzeń za pomocą sieci x-2-1 dodając kolejno zmienne i sprawdzając jakość klasyfiacji. Wyniki tej analizy zostały przedstawione na Rysunku 4.19. Można zauważyć, że najwyżej położona jest krzywa czystośćwydajność uzyskana przy wykorzystaniu siedmiu zmiennych - czyli zostały użyte wszystkie parametry oprócz  $\langle \alpha \rangle_{max}$ . Zaskakującym jest fakt, że dodanie ostatniej zmiennej  $\langle \alpha \rangle_{max}$ , która zgodnie z Rysunkiem 4.18c powinna poprawić zdolność rozróżniania  $e^-/\pi^0$ , pogarsza jakość klasyfikacji. Przyczyna takiego zachowania nie została wykryta.

Aby sprawdzić czy dodanie neuronów w warstwie ukrytej, przy liczbie zmiennych równej siedem, poprawi jakość klasyfiacji, przeprowadzona została analiza, w której badano zachowanie krzywych czystość-wydajność dla sieci 7 - x - 1 dla różnych wartości x. Wyniki analizy zostały przedstawione na Rysunku 4.20.

Jakość klasyfikacji nie uległa poprawie w wyniku dodania neuronów do warstwy ukrytej sieci - krzywe czystość-wydajność pokrywają się w granicach błędów.

Rozkład odpowiedzi sieci 7-2-1 na zbiorze danych testowych został przedstawiony na Rysunku 4.21.

Jak widać, pozostała pewna mała część elektronów, dla których odpowiedź sieci była bliska zeru, co sugerowałoby że są to mezony  $\pi^0$ . Natomiast została poprawiona wyraźnie czystość próbki - liczba pionów, dla których odpowiedź sieci miała wartość bliską jedynce jest znikoma.

Krzywa czystość-wydajność dla zbioru uczącego i testowego przedstawiona na Rysunku 4.22 pokazuje, że liczba przypadków w zbiorze testowym i uczącym była wystarczająca, a zbiór uczący reprezentował dostatecznie globalne własności klasyfikowanych danych - krzywe dla zbioru uczącego i testowego pokrywają się w granicach błędów w całym zakresie wydajności.

Sieć 7 - 2 - 1 daje również dobre wyniki po przeskalowaniu ilości sygnału w próbce do przewidywanej dla eksperymentu T2K wartości 12% (Rysunek 4.23 oraz Tablica 4.8).

Nawet przy 12% udziale elektronów w próbce testowej osiągnięto czystość 60.2% przy wydajności 94.11%, co daje istotną poprawę w stosunku do wyników przedstawionych w Tablicy 4.7.

Ostateczne wyniki analizy, przy wykorzystaniu informacji o stratach energii na jonizację na pierwszych drutach oraz informacji o topologii kaskady, zostały



Rysunek 4.19: Krzywe czystość-wydajność uzyskane dzięki zastosowaniu dodatkowych zmiennych wejściowych, charakteryzujących topologię kaskady (sieć x-2-1). Znak "+" oznacza dodawanie kolejnych zmiennych wejściowych. Dolna część rysunku przedstawia zbliżenie interesującego obszaru. Najlepsza jakość klasyfikacji została uzyskana dla siedmiu zmiennych wejściowych, bez użycia  $\langle \alpha \rangle_{max}$ .



Rysunek 4.20: Krzywe czystość-wydajność dla różnej liczby neuronów (N) w warstwie ukrytej dla sieci 7 - x - 1. Nie zaobserwowano poprawy jakości klasyfikacji po dodaniu kolejnych neuronów w warstwie ukrytej - krzywe pokrywają się w granicach błędów.



Rysunek 4.21: Rozkład odpowiedzi sieci 7–2–1. Na wykresie widoczna jest część elektronów, które zostały niepoprawnie sklasyfikowane jako mezony  $\pi^0$  - obecność małego piku w okolicy zera.



Rysunek 4.22: Krzywa czystość-wydajność dla sieci 7-2-1. Zbiór uczący reprezentuje globalne własności klasyfikowanych danych - krzywe dla zbioru uczącego (training) i testowego (testing) pokrywają się w granicach błędów.



Rysunek 4.23: Krzywe czystość-wydajność dla zbioru testowego dla 55% i 12% sygnału w próbce (sieć 7-2-1).

Udział elektronów w próbce	Wydajność [%]	Czystość [%]	Błąd czystości [%]
55%	94.11	93.21	0.51
12%	94.11	60.20	1.94

Tablica 4.8: Wyniki klasyfikacji za pomocą sieci7-2-1.Użyte zmienne wejściowe opisane są Wzorami 4.5 - 4.10.

przedstawione w Tablicy 4.9.

Zmienne wejściowe	Architektura	Wydajność [%]	Czystość [%]
$< dE/dx >_{col}$			
$+ < sygnal >_{ind}$	2 - 2 - 1	94.11	$88.0 \pm 0.64$
$< dE/dx >_{col}$			
$+ < sygnal >_{ind}$	7 - 2 - 1	94.11	$93.21 \pm 0.51$
+ topologia kaskady			

Tablica 4.9: Wyniki klasyfikacji za pomocą sieci x - 2 - 1. Dla sieci 7 - 2 - 1 użyto zmiennych charakteryzujących topologię kaskady elektromagnetycznej (Wzory 4.5-4.8).

W porównaniu do poprzedniej analizy, gdzie używane były tylko zmienne  $\langle dE/dx \rangle_{col}$  i  $\langle sygnal \rangle_{ind}$ , nastąpił znaczny wzrost czystości próbki przy tej samej wydajności rozpoznawania (o 5.21%). Otrzymane wyniki potwierdzają przypuszczenia, jakie były sformułowane przy analizie Wykresu 4.21, gdzie widać wyraźnie, że liczba mezonów  $\pi^0$ , dla których odpowiedź sieci jest bliska jedynce jest znikomo mała. Odzwierciedleniem tego faktu jest wysoka wartość czystości: 93.21% oraz jej mniejsza niepewność: 0.51 %.

W Tablicy 4.10 przedstawiono także wyniki tej części analizy dla 90% wydajności w celu porównania ich z pracą [1].

Wydajność [%]	Czystość [%]	Błąd czystości [%]
94.11	93.21	0.51
89.97	97.73	0.32

Tablica 4.10: Ostateczne wyniki klasyfikacji zdarzeń za pomocą sieci 7-2-1, uzyskane przy użyciu informacji o stratach energii na jonizację na pierwszych drutach w płaszczyźnie Collection i sygnału w płaszczyźnie Induction oraz zmiennych opisujących topologię powstałej kaskady elektromagnetycznej.

Wyniki w powyższej tabeli wskazują, że dla 90% wydajności uzyskano poprawę w stosunku do pracy [1]. Wartość czystości podniosła się z 96.3% do 97.73% dzięki zastosowaniu dodatkowych zmiennych:  $\langle sygnal \rangle_{ind}$  oraz parametrów opisujących topologię kaskady zainicjowanej przez elektron lub foton.

# 4.5 Wykorzystanie informacji o położeniu wierzchołka oddziaływania

W celu pokazania, jak wysoką czystość próbki można uzyskać, przy założeniu znajomości położenia wierzchołka pierwotnego oddziaływania, została przeprowadzona analiza za pomocą sieci neuronowej o ośmiu neuronach wejściowych. Na wejścia siedmiu neuronów były podawane zmienne wykorzystane w poprzednim rozdziale, natomiast na wejście ósmego neuronu dodana została informacja o odległości pierwszej zarejestrowanej jonizacji  $x_{jon}$  od wierzchołka pierwotnego. Ponieważ wygenerowane dane Monte-Carlo zawierały tylko zdarzenia oddziaływujących elektronów i mezonów  $\pi^0$  w ciekłym argonie, położenie wierchołka oddziaływania pierwotnego oznaczało tutaj punkt w którym cząstka została wstrzelona do detektora.



Rysunek 4.24: Położenie pierwszych jonizacji dla elektronu i  $\pi^0$  oddziałujących w detektorze (oznaczone strzałkami), które powstają w różnych odległościach od wierzchołka oddziaływania pierwotnego (zaznaczonego na rysunku dużą kropką). Elektron przelatując przez ciekły argon daje natychmiastową jonizację, mezon  $\pi^0$  rozpada się na dwa fotony, które po przebyciu pewnej drogi konwertują w pary  $e^+e^-$ , dające jonizację ośrodka.

Elektron daje prawie natychmiastową jonizację, w odróżnieniu od mezonu  $\pi^0$ ,

który rozpada się po przebyciu znikomej drogi na dwa fotony, które konwertują na pary  $e^+e^-$ ,na średniej odległości równej drodze radiacyjnej w ciekłym argonie  $(X_0=14 \text{ cm})$ . Dopiero elektron lub pozyton, powstały w wyniku konwersji jednego z fotonów, daje sygnał jonizacyjny (Rysunek 4.24). Z tego właśnie powodu na wykresie obrazującym odległość pierwszej jonizacji od wierzchołka pierwotnego  $x_{jon}$ , obserwuje się wysoki pik w okolicy zera dla elektronów oraz płaski, zanikający eksponencjalnie rozkład dla  $\pi^0$  (Rysunek 4.25).



Rysunek 4.25: Odległość pierwszej jonizacji od wierzchołka pierwotnego dla elektronu i mezonu  $\pi^0$ . Na rysunku widoczny jest wysoki pik dla elektronu w okolicy zera (natychmiastowa jonizacja), natomiast dla pionu widoczny jest zanik eksponencjalny.

Porównanie jakości klasyfikacji uzyskanej przy użyciu sieci 7-2-1 z poprzedniego rozdziału, z wynikami osiągniętymi za pomocą sieci o architekturze 8-2-1, z użyciem zmiennej  $x_{ion}$ , zostało przeprowadzone na Rysunku 4.26.

Jak widać wyraźna poprawa jakości klasyfikacji została uzyskana dzięki dodaniu informacji o położeniu wierzchołka oddziaływania pierwotnego - krzywa czystość-wydajność jest położona wyżej niż dla klasyfikacji za pomocą sieci 7-2-1.

Aby uzyskać optymalną wartość wydajności i czystości przebadany został ponownie wpływ dodawania neuronów w warstwie ukrytej na jakość klasyfikacji. Znaleziono nową konfigurację, dla której krzywa czystość-wydajność jest położona najwyżej na wykresie. Jest to sieć o trzech neuronach w warstwie ukrytej (Rysunek 4.27).

Rozkład odpowiedzi sieci 8-3-1, dla zbioru danych testowych został przedstawiony na Rysunku 4.28, gdzie widoczna jest znacząca poprawa trafności odpowiedzi sieci na dane testowe, uzyskana dzięki dodaniu dodatkowej zmiennej  $x_{jon}$ .


Rysunek 4.26: Porównanie jakości klasyfikacji dokonanej przy użyciu informacji o stratach energii na jonizację i topologii kaskad oraz po dodaniu zmiennej  $x_{jon}$ , określającej położenie wierzchołka oddziaływania pierwotnego. Dolna część rysunku przedstawia zbliżenie interesującego obszaru. Widoczna jest wyraźna poprawa po uwzględnieniu dodatkowej informacji (krzywa położona wyżej na wykresie).



Rysunek 4.27: Krzywe czystość-wydajność dla różnej liczby neuronów (N) w warstwie ukrytej dla sieci 8 - x - 1. Optymalną jakość klasyfikacji uzyskuje się dla N=3.

Liczba elektronów dla których sieć dała odpowiedź bliską zeru oraz ilość mezonów  $\pi^0$ , dla których odpowiedź jest bliska jedynce, jest dużo mniejsza niż na Rysunkach 4.12 oraz 4.21. Jest to zwiastunem wysokich wartości zarówno wydajności rozpoznawania  $e^-$ , jak i czystości próbki.

Krzywe czystość-wydajność dla zbioru uczącego i testowego pokrywają się w granicach błędów czystości, co potwierdza dobre własności globalizacji sieci 8-3-1 (Rysunek 4.29).

Rysunek 4.30 przedstawia przebieg krzywych czystość-wydajność uzyskanych dla sieci 8-3-1 dla różnych zawartości sygnału (elektronów) w próbce. Widać, że krzywa dla 12% elektronów w próbce nie różni się tutaj tak bardzo od przypadku dla 55%, jak dla poprzednich częci analizy.

Ostateczne wyniki klasyfikacji za pomocą sieci o architekturze 8-3-1, z użyciem informacji o średnich stratach energii na jonizację, topologii kaskad elektromagnetycznych oraz położeniu wierzchołka oddziaływania pierwotnego zostały przedstawione w Tablicy 4.11.

Udział elektronów w próbce	Wydajność [%]	Czystość [%]	Błąd czystości [%]
55%	94.06	99.34	0.17
12%	94.06	94.30	1.39

CT 11' 4 1 1	<b>TT7</b> •1 •	1.1 0.1	••		• •	0	0	1
l'ablica /LLL	M/m n l z 1	- Lagy to Lag	C11 79	nomoco	C10C1	× _	_ 3 _	
Tabilla 4.11.		niasviina	ui za	DOMOCA	SICUL	0 -	- 0 -	т.
	· · · /	·- •/	.,					



Rysunek 4.28: Rozkład odpowiedzi sieci 8-3-1. Sieć daje odpowiedzi bliskie zeru dla bardzo małej liczby elektronów oraz bliskie jedynce także dla znikomej ilości mezonów  $\pi 0$ .



Rysunek 4.29: Krzywa czystość-wydajność dla sieci 8-3-1 uzyskane dla zbioru uczącego (training) oraz testowego (testing) - pokrywają się w granicach błędów.

Rezultaty potwierdzają przypuszczenia wynikające z analizy rozkładu odpowiedzi sieci. Uzyskano bardzo wysoką wartość czystości próbki elektronów wynoszącą 99.34% przy niskim błedzie równym 0.17%. Wartości te zostały otrzymane dla wydajności rozpoznawania elektronów równej 94.06%. Jest to bardzo dobry wynik tylko 0.66% pionów (tła) nie zostało odrzuconych.

Wydajność [%]	Czystość [%]	Błąd czystości [%]
94.06	99.34	0.17
89.80	99.77	0.10

Tablica 4.12: Ostateczne wyniki klasyfikacji zdarzeń za pomocą sieci 8 – 3 – 1, uzyskane przy użyciu informacji o stratach energii na jonizację na pierwszych drutach w płaszczyźnie Collection i średniej wartości sygnału w płaszczyźnie Induction, zmiennych opisujących topologię powstałej kaskady elektromagnetycznej oraz zmiennej  $x_{ion}$ .

Wyniki klasyfikacji dla przypadku 90% wydajności wykazują znaczną poprawę w stosunku do pracy [1]. Wartość czystości osiągnęła wartość 99.77%, przy bardzo niskiej niepewności: 0.1%, podczas gdy autorzy noty [1] osiągnęli 96.3% czystości.

Rezultaty osiągnięte w tej części analizy są istotnie lepsze od wcześniejszych. Należy jednak zwrócić uwagę na fakt, że jeśli mamy do czynienia z rzeczywistymi danymi z detektora, dokładna znajomość położenia wierzchołka pierwotnego jest sytuacją mocno wyidealizowaną. W rzeczywistych warunkach wartość  $x_{jon}$  będzie



Rysunek 4.30: Krzywe czystość-wydajność dla zbioru testowego dla 55% i 12% sygnału (elektronów) w próbce, po wykorzystaniu zmiennej  $x_{jon}$  (sieć 8 - 3 - 1).

obarczona dodatkowym błędem. Może się też okazać, że określenie wartości  $x_{jon}$ nie będzie w ogóle możliwe, chociażby ze względu na obecność w interesującym nas obszarze innych cząstek niż elektron lub  $\pi^0$ .

## Rozdział 5 Dyskusja Wyników

Analiza danych Monte-Carlo za pomocą sieci neuronowych została podzielona na trzy części. W każdej z tych części badana była jakość klasyfikacji elektronów i mezonów  $\pi^0$ . W części pierwszej użyto między innymi informacji o średnich stratach energii na jonizację na kilku kolejnych drutach. W drugiej części dodane zostały informacje o kształcie kaskad. Część trzecia to analiza, przeprowadzona przy założeniu znajomości położenia wierzchołka oddziaływania pierwotnego. Rezultaty badań dla każdej z trzech części zostały przedstawione w Tablicy 5.1, oraz na Rysunku 5.1.

Użyta informacja	Wydajność [%]	Czystość [%]
	94.11	$88\pm0.64$
$< dE/dx >_{col} + < sygnal >_{ind}$		
	90.01	$97.16 \pm 0.35$
	94.11	$93.21 \pm 0.51$
$< dE/dx >_{col} + < sygnal >_{ind}$		
+ Topologia kaskad	89.97	$97.73 \pm 0.32$
$\overline{ < dE/dx >_{col} + < sygnal >_{ind} }$	94.06	$99.34 \pm 0.17$
+ Topologia kaskad		
$+$ Odległość pierwszej jonizacji $x_{jon}$	89.80	$99.77 \pm 0.10$

Tablica 5.1: Porównanie wyników klasyfikacji za pomocą sieci neuronowych dla wszystkich trzech części analizy.

Z wyników przedstawionych w Tablicy 5.1 wynika, że największy postęp w poprawie jakości klasyfikacji uzyskano dla wysokich wartości wydajności równej około 94%. W części pierwszej osiągnięto czystość próbki równą 88% przy wydajności 94.11%. Rezultaty takie uzyskano korzystając jedynie z informacji o średnich



Rysunek 5.1: Krzywe czystość-wydajność dla trzech różnych metod klasyfikacji. Dolna część rysunku przedstawia zbliżenie interesującego obszaru.

stratach energii na jonizację  $\langle dE/dx \rangle_{col}$ , oraz sygnału  $\langle sygnal \rangle_{ind}$  na trzech kolejnych drutach (od 3 do 5). Wyniki zostały poprawione po uwzględnieniu różnic w topologii kaskad elektromagnetycznych zainicjowanych przez elektron i mezon  $\pi^0$ . Otrzymano bardzo dobre wyniki dzięki zastosowaniu zmiennych, które charakteryzowały kształt kaskad. Wartość czystości, jaką osiągnięto to 93.21 ± 0.51% przy wydajności 94.11%.

Dla wydajności 90% nie uzyskano już tak wyraźnej poprawy czystości jak w przypadku 94%. Używając zmiennych  $\langle dE/dx \rangle_{col}$ , oraz  $\langle sygnal \rangle_{ind}$  uzyskano wartość czystości równą 97.16%. Dodanie zmiennych charakteryzujących topologię kaskad poprawiło ten wynik do 97.73%. Porównując ten wynik z rezultatami pracy [1], gdzie otrzymano 96.3% czystości, należy stwierdzić, że dla niższych wartości wydajności rozpoznawania elektronów trudniej jest uzyskać znaczący zysk czystości. Na wykresach czystość - wydajność objawia się to faktem, że dla wydajności 90% i niższych krzywe zaczynają się pokrywać w granicach błędów.

We wspomnianej wyżej pracy [1], gdzie do klasyfikacji zdarzeń stosowano zupełnie inne metody niż w niniejszej analizie, a jako zmienną dyskryminującą użyto tylko informacji o średniej  $\langle dE/dx \rangle$  w płaszczyźnie Collection, autorzy zwracają uwagę, że jakość klasyfikacji  $e^-/\pi^0$  można znacznie poprawić stosując oprócz  $\langle dE/dx \rangle$  także informację o położeniu wierzchołka pierwotnego. Oczekiwana przez nich wartość wydajności rozpoznawania pionów (w niniejszej pracy odpowiada ona czystości) to 99.8%. Aby sprawdzić powyższą hipotezę przeprowadzono analizę z użyciem informacji o odległości pierwszej jonizacji od wierzchołka pierwotnego oddziaływania. Dla wysokiej wydajności (94%) osiągnięto bardzo dobry wynik: 99.34 ± 0.17% czystości próbki, który jest bardzo bliski przewidywanej wartości. Natomiast wartość czystości uzyskana dla wydajności 90%, równa 99.77±0.10%, pokrywa się dokładnie z oczekiwaniami autorów. Można więc stwierdzić, że ich przewidywania były słuszne i zostały potwierdzone w niniejszej pracy.

## Podsumowanie

W przedstawionej pracy badano wydajność rozróźniania elektronów i mezonów  $\pi^0$ , mającego na celu rozpoznanie typu oddziaływania pierwotnego neutrina (CC lub NC) w ciekłoargonowej komorze TPC. Klasyfikacja została przeprowadzona na danych Monte-Carlo dla ciekłoargonowego detektora eksperymentu T2K za pomocą sieci neuronowych. Analiza została dokonana zarówno przy skorzystaniu z informacji o średnich stratach na jonizację na pierwszych drutach anodowych, które zarejestrowały sygnał, jak i przy użyciu zmiennych charakteryzujących kształt kaskady elektromagnetycznej zainicjowanej przez elektron lub foton z rozpadu mezonu  $\pi^0$ . Zbadano także sytuację, w której znane jest położenie wierzchołka pierwotnego oddziaływania neutrina. Uzyskane wyniki dają poprawę jakości klasyfikacji, w stosunku do rezultatów osiągniętych w pracy [1]. Dzięki zastosowaniu bardzo wydajnej metody klasyfikacji, jaką są sieci neuronowe, jak również zmiennych charakteryzujących własności geometryczne kaskad elektromagnetycznych, wzrosła wyraźnie jakość rozróżniania elektronów i mezonów  $\pi^0$ .

Zastosowana w niniejszej pracy klasyfikacja  $e^-/\pi^0$  może stanowić cenne narzędzie w analizie rzeczywistych danych z eksperymentów neutrinowych z zastosowaniem ciekłoargonowych komór TPC, takich jak T2K czy ICARUS.

## Podziękowania

Chciałbym podziękować mojemu promotorowi dr Michałowi Markiewiczowi za pomoc i opiekę naukową w trakcie pisania pracy. Szczególnie serdeczne podziękowania składam na ręce prof. Agnieszki Zalewskiej za jej bezcenne uwagi merytoryczne oraz wskazówki techniczne. Dziękuję także Andrzejowi Szelcowi za to, że służył mi pomocą w rozwiązywaniu problemów natury komputerowej. Na zakończenie dziękuję mojej narzeczonej za to, że wytrzymała ze mną podczas trudnych chwil w trakcie pisania niniejszej pracy.

## Bibliografia

- [1] Y.Ge, P.Sala, A.Rubbia, " $e/\pi^0$  separation in ICARUS liquid argon time projection chamber", ICARUS-TM/03-05
- [2] C.Gruppen, "Physics of particle detection", ICFA Instrumentation School, 1999
- [3] C.Joram, "Particle detectors", CERN Academic Training, 1997/98
- [4] A.Zalewska, "The ICARUS experiment at Gran Sasso underground laboratory", Kraków, 2004
- [5] ICARUS Collaboration, "The ICARUS EXPERIMENT- A Second-Generation proton decay experiment and Neutrino Observatory at the Gran Sasso laboratory", ICARUS-TM/2001-03
- [6] "The T2K-LAr Project", Draft March 2, 2005
- [7] http://neutrinooscillation.org
- [8] S.Osowski, "Sieci neuronowe", Warszawa 1994
- [9] S.Osowski, "Sieci neuronowe w ujęciu algorytmicznym", Warszawa 1996
- [10] R.Tadeusiewicz, "Sieci neuronowe", Warszawa 1993
- [11] http://alphas.if.uj.edu.pl/7Epbialas/NN/wyklad.ps
- [12] http://aneksy.pwn.pl/efw/?id=569
- [13] http://www.neural-networks-at-your-fingertips.com
- [14] http://nrn.prv.pl
- [15] R.Sulej i inni, "Search for CC tau neutrino interactions in CNGS beam using neural network", ICARUS-TM/04-13

- [16] R.Adamczak, "Ontogeniczne sieci neuronowe w zastosowaniu do klasyfikacji danych medycznych", Praca doktorska pod kierunkiem prof. Włodzisława Ducha, Katedra metod komputerowych Uniwersytetu Mikołaja Kopernika, Toruń 2001
- [17] A.Zalewska, "Topical questions in experimental neutrino physics", Ustroń, 2003
- [18] http://aneksy.pwn.pl/efw/?id=549
- [19] A.Bueno, S.Navas, "Random Forest and Neural Network classification algorithms applied to particle identification", ICARUS-TM/05-01
- [20] A.Bueno, A.Martinez de la Ossa, S.Navas, "Statistical pattern recognition application to numu->nutau oscillation search at CNGS", ICARUS-TM/04-03
- [21] A.Bueno, S.Navas, "Particle ID in ICARUS", ICARUS-TM/04-09
- [22] J.Brau, "Elementary particle phenomenology", Physics 663 Detectors, Spring 2004
- [23] A.Ferrari, P.Sala, "Treating high energy showers", Roma 1998
- [24] S.Eidelman i inni, Phys. Lett. B 592, 1(2004)
- [25] http://edward\_ch.republika.pl