

# Razvoj algoritma za identifikaciju hadronskih mlazova nastalih raspadom ultrarelativističkih Higgsovih bozona u dva b kvarka korištenjem nadograđenog CMS piksel detektora

---

Roguljić, Matej

Master's thesis / Diplomski rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:217:254326>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-10**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

Matej Roguljić

Razvoj algoritma za identifikaciju hadronskih  
mlazova nastalih raspadom ultrarelativističkih  
Higgsovih bozona u dva b kvarka korištenjem  
nadograđenog CMS piksel detektora

Diplomski rad

Zagreb, 2017.

SVEUČILIŠTE U ZAGREBU  
PRIRODOSLOVNO-MATEMATIČKI FAKULTET  
FIZIČKI ODSJEK

INTEGRIRANI PREDDIPLOMSKI I DIPLOMSKI SVEUČILIŠNI STUDIJ  
FIZIKA; SMJER ISTRAŽIVAČKI

**Matej Roguljić**

Diplomski rad

**Razvoj algoritma za identifikaciju  
hadronskih mlazova nastalih  
raspadom ultrarelativističkih  
Higgsovih bozona u dva b kvarka**

Voditelj diplomskog rada: Dr. sc. Dinko Ferenček

Ocjena diplomskog rada: \_\_\_\_\_

Povjerenstvo: 1. \_\_\_\_\_

2. \_\_\_\_\_

3. \_\_\_\_\_

Datum polaganja: \_\_\_\_\_

Zagreb, 2017.

Zahvaljujem se CMS grupi Instituta Ruđer Bošković, a posebno mentoru dr. sc. Dinku Ferenčeku. Hvala obitelji i prijateljima čija je potpora indirektno doprinijela izradi diplomskog rada.

## Sažetak

U prvoj polovici 2017. godine izvršena je nadogradnja CMS piksel detektora sa Phase 0 na Phase 1 konfiguraciju. U ovom diplomskom radu ispitan je učinak nadogradnje piksel detektora na algoritam za dvostruko  $b$  označavanje. Učinak je određen usporedbom performansi algoritma na simuliranim podacima prije i nakon nadogradnje i rezultati pokazuju da se performanse algoritma poboljšaju korištenjem Phase 1 detektora. Također je izvršen i novi trening algoritma na Phase 1 podacima čime je dobiveno dodatno poboljšanje performansi. Usporedbom Phase 1 i Phase 0 treninga na odgovarajućim podacima u rasponu transverzalnog impulsa hadronskih mlazova od 300 do 1000 GeV pokazano je da Phase 1 algoritam ima bolje performanse na svim radnim točkama. Relativno poboljšanje iznosi oko 10% na radnim točkama visoke čistoće signala što se polako smanjuje smanjenjem čistoće. Utvrđeno je da poboljšanju najviše doprinosi bolja mogućnost raspoznavanja  $b$  hadronskih mlazova od hadronskih mlazova nastalih iz lakih kvarkova i gluona. Ispitana je ovisnost algoritma za  $b$  označavanje o transverzalnom impulsu hadronskog mlaza i značajne se degradacije performansi pojavljuju tek od 1000 GeV. Dodatno poboljšanje algoritma dobiveno je korištenjem bolje konfiguracije parametara metode stabla odlučivanja koja se koristi tijekom treniranja.

Ključne riječi: CMS piksel detektor, Higgsov bozon, ultrarelativistički režim,  $b$  označavanje, dvostruko  $b$  označavanje

# Development of an algorithm for identification of boosted Higgs bosons decaying in two b quarks using the upgraded CMS pixel detector

## Abstract

The CMS pixel detector was upgraded from Phase 0 to Phase 1 configuration in the first half of 2017. The effect of the upgrade on the double b tagging algorithm is studied in this master's thesis. It is determined by comparing the performance of the algorithm on the simulated data before and after the upgrade. Results show that the Phase 1 upgrade improves the performance of the double b tagging algorithm. Retraining of the algorithm using Phase 1 data is also performed to get additional improvement. Comparison of Phase 1 and Phase 0 training in the 300–1000 GeV transverse momentum range shows that Phase 1 training performs better on the whole range of working points. Relative improvement starts at about 10% for the high signal purity working points and slowly decreases with decreasing signal purity. It is determined that the upgrade mostly improves discrimination between b hadron jets and light-flavour jets. The stability of the algorithm with respect to the jet transverse momentum is confirmed and significant degradation of performance starts only above 1000 GeV. Further improvement of the algorithm is achieved by using a more optimal set of parameters in the training of the boosted decision tree.

Keywords: CMS pixel detector, Higgs boson, ultrarelativistic limit, b tagging, double b tagging

# Sadržaj

<b>1</b>	<b>Uvod</b>	<b>1</b>
<b>2</b>	<b>Eksperimentalni postav</b>	<b>3</b>
2.1	Veliki hadronski sudarivač . . . . .	3
2.2	Detektor CMS . . . . .	5
2.2.1	Sustav za detekciju tragova nabijenih čestica . . . . .	6
2.2.2	Kalorimetri . . . . .	8
2.2.3	Magnetska zavojnica . . . . .	9
2.2.4	Sustav za detekciju miona . . . . .	10
2.2.5	Nadogradnja piksel detektora . . . . .	11
<b>3</b>	<b>Rekonstrukcija događaja u CMS-u</b>	<b>12</b>
3.1	Particle-flow algoritam . . . . .	12
3.2	PUPPI algoritam . . . . .	14
3.3	Rekonstrukcija hadronskih mlazova . . . . .	16
<b>4</b>	<b>Algoritam za dvostruko b označavanje</b>	<b>19</b>
4.1	Metoda strojnog učenja . . . . .	23
4.2	Simulacija i selekcija događaja . . . . .	27
4.3	Treniranje algoritma za dvostruko b označavanje . . . . .	27
<b>5</b>	<b>Zaključak</b>	<b>35</b>
	<b>Dodaci</b>	<b>37</b>
<b>A</b>	<b>Konfiguracija parametara BDT metode</b>	<b>37</b>
	<b>Literatura</b>	<b>38</b>

# 1 Uvod

Higgsov bozon elementarna je čestica Standardnog modela čije je postojanje predviđeno još u 60-im godinama 20. stoljeća i pitanje njegovog postojanja bilo je otvoreno dugo vremena [1] [2]. Ime nosi po Peteru Higgsu, jednom od 6 znanstvenika koji su u svojim radovima predstavili Higgsov mehanizam koji sugerira postojanje takve čestice. U Standardnom je modelu Higgsov bozon nestabilna bozonska čestica koja nema spin, električni ni bojni naboj. 2012. godine objavljen je pronalazak čestice mase oko 125 GeV [3] [4] koja se slaže s predviđanjima Standardnog modela unutar preciznosti dosadašnjih mjerenja, ali dodatna su mjerenja potrebna da bi se utvrdilo radi li se doista o Higgsovom bozonu i posjeduje li svojstva predviđena Standardnim modelom. Higgsov je bozon prvi put opažen u kanalu raspada u dva fotona i u kanalu raspada u dva Z bozona koji se zatim svaki raspadaju u dva fermiona. Primjetimo da su oba kanala raspada bozonska. Iako Standardni model predviđa vezanje Higgsovog bozona na fermione pomoću Yukawinog vezanja, do sada nije potvrđen nijedan raspad Higgsovog bozona u fermionskom kanalu sa sigurnošću od 5 sigma. Pogledom na predviđene omjere grananja vidimo da se Higgsov bozon najčešće raspada u par b kvark-antikvark što su fermioni, dok su vjerojatnosti raspada u prvo potvrđene kanale relativno male (2 promila za raspad u dva fotona i 3% za raspad u dva Z bozona, nasuprot 58% za raspad u dva b kvarka). Da bismo objasnili zašto još nije opažen raspad Higgsovog bozona u dva b kvarka unatoč relativno visokom omjeru grananja, prvo moramo objasniti pojmove pozadine i signala.

Nestabilne se čestice u eksperimentu ne mogu izravno opaziti (osim ako ne žive dovoljno dugo poput miona) nego detektiramo produkte njihovog raspada. Pogledajmo što bi to značilo za raspad Higgsovog bozona u dva b kvarka.

Recimo da se prilikom sudara snopova protona stvori Higgsov bozon i neka se on raspadne u dva b kvarka koji su također nestabilne čestice tako da se i oni nakon nekog vremena raspadnu (prije nego se raspadnu, unutar detektora stignu preletjeti nekoliko milimetara). Poznato je da zbog zatočenja boje kvarkovi ne mogu biti izolirani te stoga dolazi do hadronizacije. Hadronizacija je proces formiranja hadrona iz kvarkova i gluona gdje se nastali kvarkovi kombiniraju sa spontano stvorenim kvarkovima i antikvarkovima iz vakuuma. Uzak, usmjeren stožac čestica koji nastaje hadronizacijom naziva se hadronski mlaz. Stoga se raspad Higgsovog bozona u dva b kvarka u



detektoru vidi kao dva hadronska mlaza čija je invarijantna masa bliska masi Higgsovog bozona. Međutim, postoje mnogi drugi procesi čiji je detektorski "potpis" isti i teško je razlikovati koji su detektirani događaji došli od raspada Higgsovog bozona, a koji ne. Sve detektirane (ili simulirane) događaje koji su došli od raspada Higgsovog bozona u dva b kvarka zovemo signalnim događajima dok sve događaje koji nisu došli od tog raspada, ali u detektoru izgledaju poput njih, zovemo pozadinskim događajima.

I tu dolazimo do odgovora na pitanje zašto još nismo opazili fermionski raspad Higgsovog bozona sa sigurnošću od pet sigma. Omjer signala i pozadine u fermionskim je kanalima puno manji nego u bozonskim kanalima, odnosno, bozonski su kanali puno čišći pa je potrebno puno manje podataka skupiti da bi se došlo do željene signifikantnosti.

Pozadina se može ukloniti tako da primijenimo selekciju na podatke i tako odaberemo podskup signalnih događaja, ali s dobrim odabirom odbacimo relativno više pozadinskih događaja. Zato se u ovom radu razvija upravo algoritam za identifikaciju ultrarelativističkih Higgsovih bozona. Stavljanjem uvjeta da transverzalni moment Higgsovog bozona bude najmanje 250 GeV prihvaćamo tek oko 5% signala, ali odbacujemo mnogo više pozadine te nam kanal postaje čišći. Također postoje i algoritmi koji, do neke mjere, uspijevaju razlikovati signalne događaje od pozadinskih i razvojem takvih algoritama dobivamo sve bolje diskriminatorske sposobnosti i u ovom se radu upravo radi na razvoju jednom takvog algoritma.

## 2 Eksperimentalni postav

### 2.1 Veliki hadronski sudarivač

CERN, Europska Organizacija za Nuklearna Istraživanja, međunarodna je znanstvena organizacija osnovana 1954. godine sa sjedištem u Ženevi, neposredno uz granicu Švicarske i Francuske i trenutno broji 22 zemlje članice. Glavni je zadatak CERN-a pružiti akcelerator čestica i popratnu infrastrukturu za istraživanje fizike elementarnih čestica. Prvi akcelerator čestica na CERN-u, Sinkrociklotron, dosežao je energije do 600 MeV. Svaki je idući akcelerator mogao doseći sve veće energije i zadnji dodatak akceleratorskom kompleksu CERN-a je Veliki hadronski sudarivač (eng. Large Hadron Collider, LHC) - trenutno najveći i najsnažniji akcelerator čestica. Najveći jer je opseg LHC-a 27 kilometara, a najsnažniji jer je dizajniran da postigne energije od 14 TeV u sustavu centra mase dva sudarajuća protona. Treba napomenuti da ta energija još nije dosegnuta i trenutno se sudari protona odvijaju na 13 TeV. Smješten je u tunelu u kojem se prethodno nalazio LEP (Large Electron-Positron), sudarivač elektrona i pozitrona energije do 209 GeV. Dubina tunela varira između 50 i 175 metara.

Zanimljivo je istaknuti da se u kompleksu LHC-a može vidjeti povijest akceleratora na CERN-u jer se neki prethodni akceleratori koriste za ubrzavanje protona prije nego li se oni injektiraju u sam LHC. Ciklus pokretanja LHC-a započinje puštanjem vodikovog plina iz boce kojem se tada "skidaju" elektroni i zatim se protoni dovode u LINAC 2 (linearni akcelerator 2) gdje se ubrzavaju do 50 MeV. Sljedeći stupanj ubrzavanja odvija se u proton-sinkrototronskom pojačivaču (PSB, Proton Synchrotron Booster) gdje se ubrzavaju do 1.4 GeV i prosljeđuju na daljnje ubrzavanje proton sinkrotronu (PS) do energije 26 GeV. Zadnji stupanj ubrzanja prije LHC-a je super protonski sinkrotron (SPS) u kojem protoni postižu energiju od 450 GeV nakon čega se prosljeđuju u LHC gdje se akceleriraju i potom sudaraju u četiri točke duž prstena LHC-a u kojima se snopovi protona presijecaju.

Osim protona, unutar LHC-a se nekad ubrzavaju i ioni olova. Procedura njihovog ubrzavanja je slična kao i kod ubrzavanja protona. Ioni se dobivaju isparavanjem olova iz vrlo čistog uzorka olova zagrijanog na oko 800 stupnjeva Celzijevih. Dobi-vena para se ionizira u više stupnjeva ionizacije te se na kraju uglavnom dobiju  $Pb^{54+}$  ioni koji se najprije ubrzavaju u Linac3 (linearni akcelerator 3) ubrzivaču do 4.5

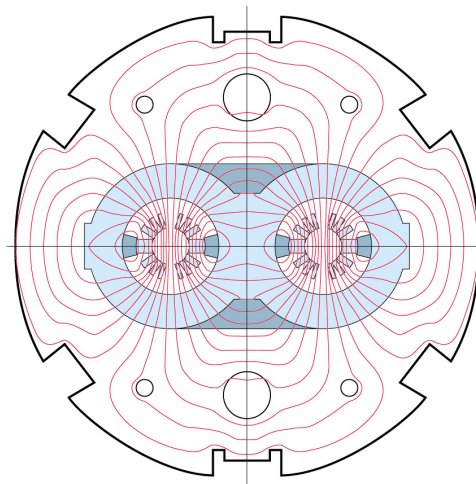
MeV po nukleonu, a nakon toga u Ionskom prstenu za niske energije (Low Energy Ion Ring, LEIR) gdje se ubrzavaju do 72 MeV po nukleonu. Tada ih PS ubrzava do 5.9 GeV po nukleonu i šalje u SPS gdje se konačno skidaju preostali elektroni te nam preostaje samo jezgra, odnosno  $Pb^{82+}$ . SPS zatim akcelerira ione na 177 GeV po nukleonu i šalje u LHC gdje se postiže energija od 2.56 TeV po nukleonu.

U LHC tunelu nalaze se dvije vakuumske cijevi, svaka promjera 6.3 centimetra, i kroz njih prolaze dva snopa protona u suprotnim smjerovima. Snopovi protona nisu kontinuirani nego su protoni raspodjeljeni na 2808 paketića u svakoj zruci. Jedan paketić protona na početku ubrzavanja sadrži  $10^{11}$  protona. Snopovi se presijecaju na četiri lokacije koje odgovaraju lokacijama četiri velika detektora - ATLAS, CMS, LHCb i ALICE i tamo dolazi do sudaranja paketića protona svakih 25 ns. Da bi se protoni kretali po željenim putanjama, koristi se sustav snažnih supravodljivih magneta.

Najbrojniji među njima su glavni dipolni magneti čiji je poprečni presjek prikazan

### **LHC dipole**

---



Computed magnetic flux map at  $B_0=10$  Tesla

CERN AC - HE110 - 10/10/95

Slika 2.1: Poprečni presjek dipolnog magneta i silnice magnetskog toka.

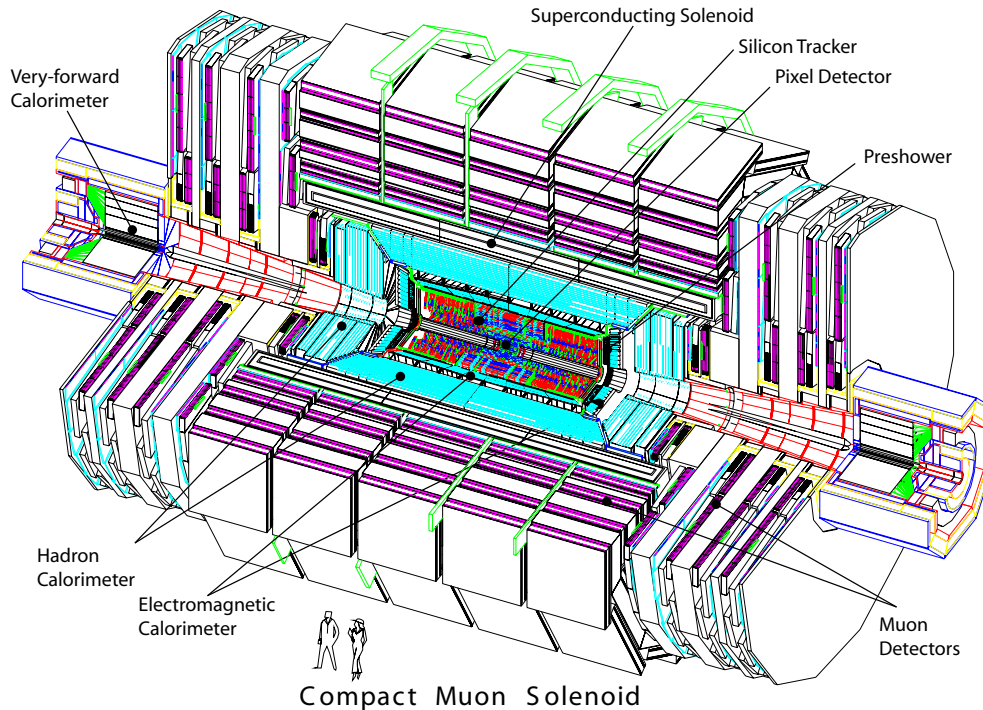
na Slici 2.1 i njih ima 1232. Njihova je glavna svrha zakretati protone tako da slijede kružnu putanju. Dipolni su magneti dizajnirani da postignu magnetsko polje do 8.3 T unutar vakuumskih cijevi. Napravljeni su od niti niobij-titanija koje su supravodljive na temperaturi ispod 10 K. Dipoli se hlade helijem i drže na temperaturi od 1.9 K što

omogućuje protok struje od 11850 A koja je potrebna za magnetsko polje od 8.3 T. Ako bismo magnete držali na temperaturi od 4.5 K, najveća moguća struja bi bila oko 8500 A što bi odgovaralo polju od 6 T.

Osim zakretanja, snopove je potrebno i fokusirati jer se pozitivno nabijeni protoni međusobno odbijaju i snop protona bi se rasipao ukoliko ga ne bismo fokusirali. Tomu služe 392 kvadrupolna magneta. Kvadrupolni magneti fokusiraju snop u jednoj ravnini dok lagano defokusiraju u drugoj pa tako razlikujemo horizontalne i vertikalne kvadrupolne magnete ovisno o tome u kojoj ravnini fokusiraju snop. Zbog tog svojstva se oni naizmjenično koriste te je ukupni efekt taj da se snop fokusira. Koriste se još i druge vrste magneta (sa više polova) koji se koriste za korigiranje drugih efekata poput utjecaja Zemljine gravitacijske sile na snop protona ili činjenice da se energije protona u zraci blago razlikuju. Na LHC-u se nalaze četiri velika detektora navedena u prošlom potpoglavlju koji se međusobno razlikuju po izvedbi i svrsi. LHCb (Large Hadron Collider beauty experiment) se bavi istraživanjem sitnih razlika u ponašanju materije i antimaterije proučavanjem b kvarkova. ALICE (A Large Ion Collider Experiment) je dizajniran za proučavanje jako-interagirajuće materije na ekstremno velikim gustoćama energije gdje se formira kvark-gluon plazma. ATLAS (A Toroidal LHC ApparatuS) i CMS (Compact Muon Solenoid) su slični detektori generalne namjene i njihova je svrha potraga za Higgsovim bozonom, ali i fizikom izvan Standardnog modela. Detaljnije ćemo opisati CMS jer će nam pomoći u shvaćanju rada algoritma za b označavanje.

## **2.2 Detektor CMS**

CMS (Compact Muon Solenoid) je engleska kratica za kompaktni mionski solenoid. Mase detektora CMS [5] je 14 000 tona i oblikom je sličan cilindru duljine 21 metra i promjera 15 metara kao što je prikazano na Slici 2.2. Izgrađen je tako da vrlo precizno može detektirati mione i sadrži snažnu magnetnu zavojnicu (solenoid) koja proizvodi polje od 3.8 T. To je ujedno i najveći supravodljivi magnet do sada izgrađen. Poprečni presjek CMS-a prikazan je na Slici 2.3. U samom centru detektora događaju se sudari protona. Detektorski sustav za detekciju tragova nabijenih čestica nalazi se najbliže točki sudara. Nakon njega dolaze elektromagnetski kalorimetar i zatim hadronski kalorimetar. Sve navedeno nalazi se u homogenom magnetskom polju od

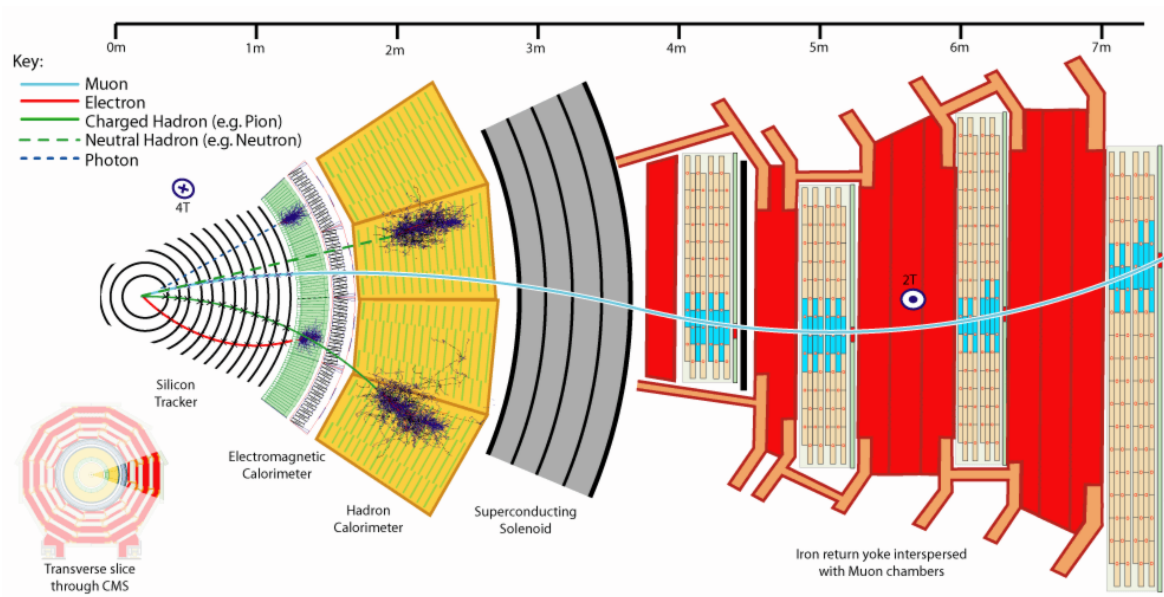


Slika 2.2: Skica detektora CMS s označenim glavnim detektorskim podsustavima [6]

3.8 T koje proizvodi supravodljiva zavojnica koja ujedno i obavija "unutarnje" slojeve detektora. Izvan zavojnice nalaze se mionske komore. Poprečni presjek CMS-a nalazi se na Slici 2.3. Opisani dio naziva se "barrel" (bačva) i on čini glavninu detektora, dok se krajnji dijelovi zajedno nazivaju "endcaps" (čep). Redoslijed detektorskih podsustava u endcapu je isti kao i u barrelu uz izuzetak magnetske zavojnice koje nema u endcapu.

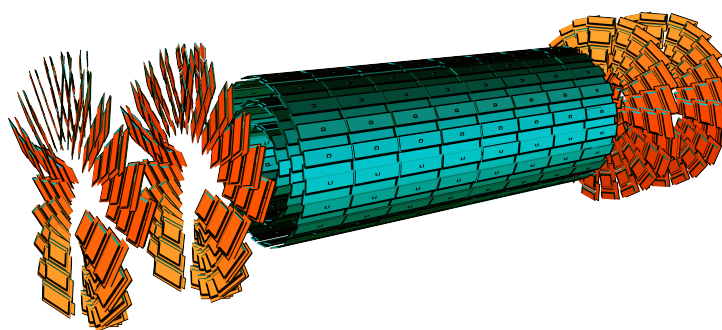
### 2.2.1 Sustav za detekciju tragova nabijenih čestica

Sustav za detekciju tragova nabijenih čestica može se podijeliti u dvije kategorije u ovisnosti o tome koliko se daleko nalazimo od područja gdje se događaju sudari. Što smo bliže točki sudara, to će biti i veći tok čestica pa tako moramo koristiti različite izvedbe detektora na različitim udaljenostima od točke sudara. Najbliže točki sudara nalazi se piksel detektor koji se sastoji od 65 milijuna silicijskih piksela veličine  $\approx 100 \times 150 \mu\text{m}^2$ . U barrelu su poslagani u tri koncentrična plašta cilindra duž osi snopa protona na udaljenostima od 4, 7 i 11 cm, a u endcapu su poslagani u diskove kako je prikazano na Slici 2.4. Nabijene čestice prolaskom kroz poluvodički senzor piksel detektor izbijaju elektrone iz atoma silicija i time stvaraju elektron-šupljina parove. Električnim poljem pokupe se izbijeni elektroni te se dobije mali električni puls



Slika 2.3: Poprečni presjek detektora CMS. Na slici su prikazane putanje i interakcije raznih vrsta čestica u detektoru.

koji se dalje pojačava i registrira. Skupljeni podaci tada se koriste za rekonstrukciju putanja čestica. Podaci iz piksel detektora najbitniji su za određivanje položaja iz kojeg su različite čestice došle.



Slika 2.4: Prikaz piksel detektora unutar CMS-a.

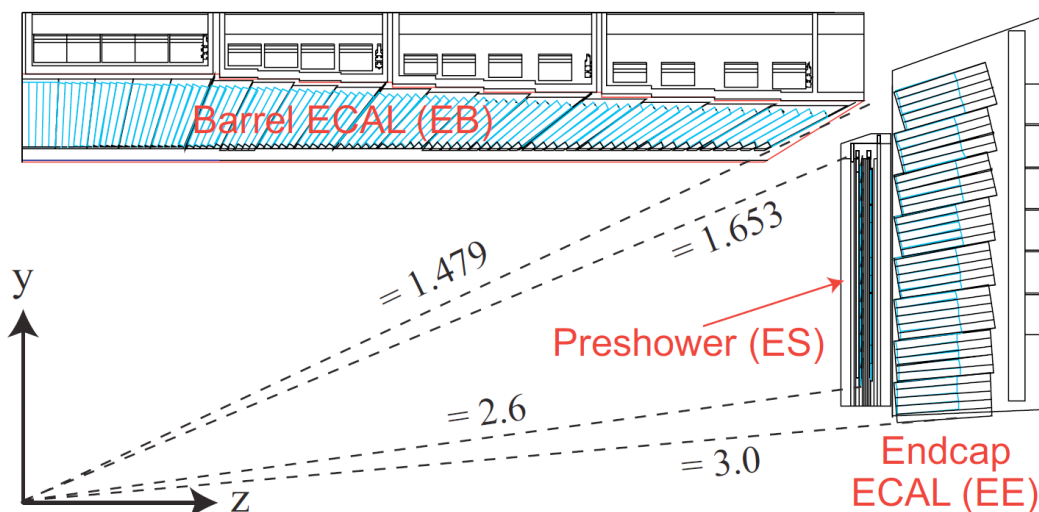
Silicijske trake koriste se na većim udaljenostima od točke sudara jer je tamo manji tok čestica pa nije potrebna vrlo visoka rezolucija kao kod piksel detektora. Tako čestice, nakon tri sloja piksel detektora, prolaze kroz deset koncentričnih slojeva silicijskih detektorskih traka u barrelu dok se u endcapu nalaze četiri sloja. Princip za

detektiranje čestica isti je kao i kod piksel detektora te se prikupljeni podaci zajedno koriste za određivanje putanje nabijenih čestica iz čije se zakrivljenosti određuje impuls čestica.

### 2.2.2 Kalorimetri

Nakon sustava za detektiranje tragova čestica nalaze se sustavi za mjerenje energija čestica. Bliže točki sudara nalazi se elektromagnetski kalorimetar (ECAL) koji služi za mjerenje energije fotona i elektrona i te se čestice ujedno i zaustave u tom dijelu detektora. ECAL je izgrađen od prozirnih kristala olovnog volframata ( $\text{PbWO}_4$ ). Prolaskom elektrona ili fotona kroz kristal, kristal zasvijetli i ukupna energija nastale svjetlosti proporcionalna je energiji upadne čestice. Količina producirane svjetlosti snažno ovisi o temperaturi te stoga postoji sustav za hlađenje koji ih drži na konstantnoj temperaturi (unutar 0.1 K). Također, količina svjetlosti je mala te je na pozadinu svakog kristala zalijepljen po jedan fotodetektor koji pretvara svjetlost u električni signal koji se zatim pojačava. U barrelu se nalazi 61200 kristala, a u endcapu 15000. Kristali se lagano oštećuju zbog visoke radijacije, ali ih se dovodi na sobnu temperaturu kada se ne događaju sudari i tada se sami "popravljaju" ukoliko nisu značajnije oštećeni. Ispred endcap dijela elektromagnetskog kalorimetra nalazi se "Preshower" (PS) koji ima istu ulogu kao i glavni dio kalorimetra, ali ima puno finiju granularnost (detektorske trake u PS-u su 2 mm široke, a ECAL kristali 3 cm). Kut između dva fotona nastala raspadom piona u području endcapa može biti dovoljno mali da bi se oni u ECAL kristalima detektirali kao jedan foton visoke energije. Zadaća PS-a upravo je ta da spriječi tu pogrešku jer se takvi fotoni, zbog puno bolje prostorne rezolucije, u PS-u zaista vide kao dva fotona. PS je izgrađen od dva sloja olova iza kojih se nalaze silicijski senzori slični onima korištenim u detektoru tragova. Upadajući foton prolaskom kroz sloj olova stvori elektromagnetski pljusak koji se detektira u silicijskim sensorima. Iz te dvije točke se ekstrapolira putanja fotona te se pogleda u kojem je ECAL kristalu ostavio svoju energiju koja se pribroji energiji izgubljenoj u PS-u. Na Slici 2.5 može se vidjeti odnos ECAL-a i PS-a.

Hadronski kalorimetar (HCAL) nalazi se iza elektromagnetskog i služi za određivanje energije hadrona. Hermetički je zatvoren, odnosno, pokriva sve moguće putanje čestica tako da ukoliko se utvrdi neravnoteža u energiji i impulsu, može se zaključiti da je riječ o "nevidljivoj" čestici za taj dio detektora. Zbog toga je sastavljen od na-



Slika 2.5: Prostorni položaj elektromagnetskog kalorimetra u CMS-u. Snop protona giba se duž z osi. Brojevima su označeni pseudorapiditeti.

izmjeničnih slojeva apsorbera i scintilatora. Hadron prolaskom kroz apsorber stvara hadronski pljusak čestica čiji jedan dio dođe do scintilatorskog sloja gdje se zabilježi energija tih čestica. Iz energije ostavljene u svakom sloju kroz koji je hadron prošao može se zaključiti kolika je bila ukupna energija hadrona. Treba istaknuti da se u hadronskom kalorimetru mjere nabijeni i neutralni hadroni. Potrebne su velike količine materijala da se hadroni zaustave pa je i sam hadronski kalorimetar relativno velik i težak. Za njegovu izgradnju korištene su čahure ruske artiljerije iz Drugog svjetskog rata jer su bile potrebne velike količine kvalitetne mjedi. Nakon što je postignut dogovor, preko milijun komada artiljerije je bilo razoružano i čahure su bile otopljene i iskorištene za izgradnju hadronskog kalorimetra.

### 2.2.3 Magnetska zavojnica

Jedan od glavnih značajki CMS-a supravodljivi je magnet težak 12000 tona koji obavlja dosad navedene detektorske sustave. Magnet je oblika zavojnice dužine 13 m i promjera 6m. Dizajnom je predviđeno da se unutar zavojnice može postići magnetsko polje od 4 T, ali prilikom rada CMS-a postiže se 3.8 T radi produljenja vremena života magnetu. Struja potrebna za polje takve jakosti je 18160 A. Magnetsko polje van zavojnice suprotnog je smjera, manje je homogeno i slabije jačine. Ukupna energija pohranjena u magnetskom polju iznosi oko 2.3 GJ što je dovoljno za rastaliti 18 tona zlata.



#### 2.2.4 Sustav za detekciju miona

Detekcija miona, kao što i samo značenje imena Compact Muon Solenoid govori, vrlo je važan zadatak detektora CMS jer se mnoge potencijalno nove čestice raspadaju u mione. Npr. kanal raspada Higgsovog bozona u dva Z bozona može rezultirati konačnim stanjem s četiri miona. Mioni ne interagiraju jakim silom i puno su teži od elektrona pa relativno lako prolaze kroz kalorimetre i izlaze iz detektora. Da bi se ipak detektirali, izgrađen je posljednji detektorski podsustav za detekciju miona do kojeg uglavnom prodiru i ostavljaju tragove samo mioni. Sastoji se od četiri mionske postaje gdje mion prolazeći ostavlja tragove. U mionskim postajama koriste se mionske driftne komore u barrelu, katodne komore u endcapu i komore s pločastim otpornicima koja se koristi uz dvije navedene komore u barrelu i endcapu. Sustav se sastoji od 250 driftnih cijevi (DT), 540 komora s katodnim trakama (CSC) i 610 komora s pločastim otpornicima (RPC).

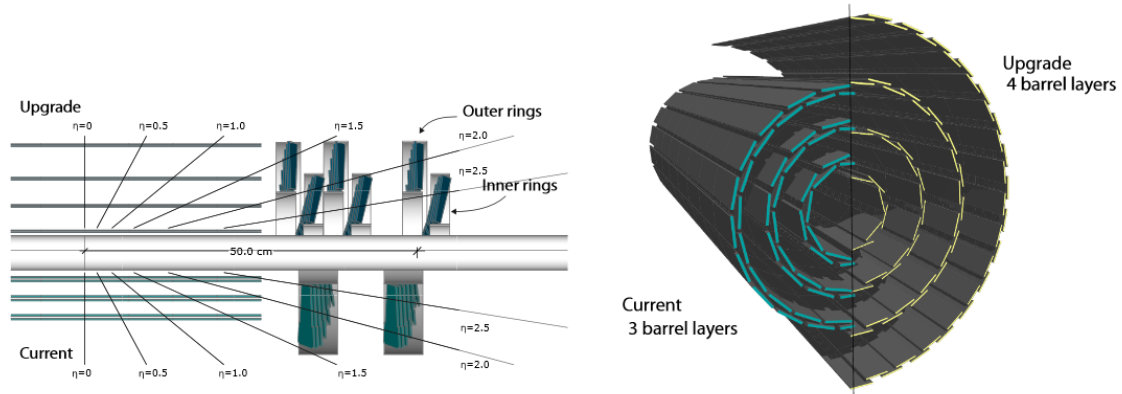
Driftne cijevi i katodne komore rade na sličan način. CSC komore sastoje se od niza pozitivnih žica (anoda) koje se križaju sa negativno nabijenim (katodama) bakrenim trakama. Mion koji prolazi kroz komoru ionizirat će atome plina koji se nalazi u komori te će nastali elektroni krenuti prema anodi i stvoriti lavinu elektrona koji će se registrirati u anodi kao puls naboja. Pozitivno ionizirani atomi krenut će prema katodama i također će se registrirati električni puls. Anode i katode okomito su postavljene jedne na druge što nam omogućuje određivanje položaja miona koji prolazi. DC komore su zapravo cijevi širine 4 centimetra ispunjene plinom i u kojima se nalazi žica. Prolaskom miona dolazi do lavine elektrona te se po obliku signala sa svake mionske komore dobije putanja miona.

RPC komore su napravljene od dvije paralelne ploče (anode i katode) od plastičnog materijala vrlo visoke otpornosti. Između njih se nalazi plin koji se ionizira prilikom prolaska miona i kao i kod prošlih detektora, stvori se lavina elektrona, ali ovdje naboje ne pokupe elektrode nego vanjske metalne pločice i to nakon točno određenog vremena. Iz oblika signala ostavljenim u komori, brzo se može izračunati impuls miona.

Navedeni se detektori također koriste kao "trigger" što znači da se podaci skupljeni u njima koriste prilikom odlučivanja hoće li se sudar protona zabilježiti ili ne. Detektor koji se koristi u triggeru mora biti brz i upravo to svojstvo imaju ovi detektori.

## 2.2.5 Nadogradnja piksel detektora

Na kraju 2016. godine i veći dio prve polovice 2017. godine odvijala se produljena godišnja tehnička pauza LHC-a tijekom koje nije bilo sudara te se radilo na popravcima dijelova oštećenih tijekom godine i nadogradnji nekih dijelova. Na CMS-u je, uz nekoliko manjih nadogradnji, obavljena nadogradnja piksel detektora. Usporedba prvotnog i novog detektora prikazana je na Slici 2.6.



Slika 2.6: Usporedba Phase 0 i Phase 1 konfiguracije piksel detektora.

Nova verzija piksel detektora ima četiri koncentrična sloja u barrelu umjesto sadašnjih tri, a endcap dio tri diska umjesto dva. Ova nadogradnja poboljšat će detekciju tragova čestica i moći će se preciznije odrediti točke odakle su došle čestice što će se pokazati bitnim upravo za b hadronske mlazove.

## 3 Rekonstrukcija događaja u CMS-u

### 3.1 *Particle-flow algoritam*

Pomoću opisanih detektorskih sustava prilikom sudara dobije se mnoštvo signala koji se mogu iskoristiti za rekonstrukciju čestica da bismo zatim mogli dobiti uvid u to što se točno dogodilo u kojem događaju. Za to je zadužen particle-flow algoritam [7]. Particle-flow rekonstrukcija teži identificiranju svih stabilnih čestica u događaju — elektrona, miona, fotona, nabijenih i neutralnih hadrona koji se identificiraju kombinacijom podataka iz svih poddetektorskih sustava CMS-a. Osim određivanja vrste čestice, određuje joj se i četveroimpuls. Rekonstruirane čestice tada se mogu koristiti za izgradnju hadronskih mlazova, određivanje nedostajuće energije (čime se može deducirati postojanje neutrina ili drugih nevidljivih čestica u događaju), itd.

Fotoni se rekonstruiraju pomoću podataka iz elektromagnetskog kalorimetra. Za razlikovanje fotona od elektrona u ECAL-u koristi se prisutnost ili odsutnost traga koji bi bio konzistentan s deponiranom energijom. Povećana granularnost pre-shower dijela ECAL-a omogućuje prepoznavanje raspada čestica u dva izrazito kolimirana fotona koji bi se u  $\text{PbWO}_4$  kristalima pogrešno identificirali kao jedan foton. ECAL se zajedno sa sustavom za detekciju tragova koristi za identifikaciju elektrona koja je problematična zbog česte emisije Bremsstrahlung fotona. Zbog toga se koristi poseban algoritam za tragove elektrona i pridruživanje Bremsstrahlung fotona elektronima kako se ne bi energija pogrešno pridružila drugim česticama, čime se ujedno i postiže bolja energetska rezolucija pri rekonstrukciji elektrona.

Nabijeni i neutralni hadroni većinu svoje energije ostavljaju u hadronskom kalorimetru koji ima znatno manju granularnost od ECAL-a. Identifikacija nabijenih hadrona postiže se pridruživanjem traga nabijene čestice čiji je položaj konzistentan s položajem ćelije u kojoj je energija deponirana. Neutralni se hadron u ćeliji hadronskog kalorimetra identificira kao razlika energije izmjerene u ćeliji i energije nabijenih hadrona čiji tragovi vode do te ćelije. Putanja nabijenih hadrona i procjena njihove energije određuje se iz sustava za detekciju tragova nabijenih čestica.

Tragovi iz unutrašnjeg sustava za detekciju tragova i podaci iz mionskih komora koriste se za rekonstrukciju miona koja je visoke efikasnosti i čistoće. Zaključak o prisutnosti neutrina i drugih slabo interagirajućih čestica donosi se gledajući iznos nedostajuće transverzalne energije (iznos vektorske sume transverznog impulsa svih

čestica u događaju).

Čestica će općenito ostavljati svoj potpis u nekoliko poddetektora te se svi ti particle-flow elementi moraju sjediniti za potpunu rekonstrukciju svake čestice. Algoritam za sparivanje elemenata tvori blokove koje se tada uzima kao potpuni opis čestice. Blokovi uglavnom sadrže od jedan do tri elementa i svakoj vezi elemenata pridružena je "udaljenost" koja govori o kvaliteti njihove povezanosti. Veza između tragova nabijenih čestica i kalorimetarskog klastera energije se uspostavlja tako da se ekstrapolira putanja čestice koristeći podatke iz sustava za detekciju tragova čestica. Ako ekstrapolirana putanja prolazi kroz klaster energije u kalorimetru, ta se energija pridruži tragu čestice. Veličina klastera može se povećati za veličinu jedne kalorimetarske ćelije. Povećanjem veličine klastera uzimamo u obzir rupe između dvije kalorimetarske ćelije ili kalorimetarskih modula kao i raspršenje čestica niskih impulsa. Udaljenost ove veze je udaljenost ekstrapolirane putanje i pozicija klastera u prostoru  $(\eta, \phi)$ . Kod elektrona se također moraju u obzir uzeti i Bremsstrahlung fotoni te se između svakog sloja detektora za detekciju nabijenih čestica ekstrapolira tangenta na putanju elektrona kojem se pridruži energija ako tangenta prolazi kroz klaster energije koji se također može povećati za veličinu jedne kalorimetarske ćelije. Veza između kalorimetarskih klastera (npr. ECAL-a i HCAL-a) uspostavlja se ako su granice klastera koji se nalazi u dijelu s boljom granulacijom (u ovom slučaju ECAL) unutar granica klastera koji se nalazi u kalorimetru s lošijom granulacijom (u ovom slučaju HCAL), a udaljenost veze se računa kao udaljenost dvaju klastera u  $(\eta, \phi)$  prostoru. Putanja miona određuje se koristeći sustav za detekciju nabijenih čestica i mionske komore. Veza između tih elemenata se uspostavlja ako  $\chi^2$  dobiven prilagodbom traga miona mjerenjima u detektoru ima dovoljno nisku vrijednost što je ujedno i udaljenost elemenata. Ako se jednom tragu u mionskim komorama može pridružiti više tragova iz sustava za detekciju tragova nabijenih čestica tada se bira onaj trag koji ima najmanju vrijednost  $\chi^2$ . Ovakav se mion naziva globalni mion.

Dobiveni blokovi zatim prolaze dodatne provjere u točno određenom redosljedu nakon čega se producira lista "particle flow" čestica s izmjerenim kinematičkim svojstvima koje se mogu dalje obrađivati i koristiti u analizama.

### 3.2 PUPPI algoritam

Prilikom sudara dvaju snopova protona dolazi do više proton-proton sudara, ali u većini je tih sudara prijenos impulsa malen te stoga nisu "zanimljivi". Takve događaje nije poželjno zabilježiti jer bi se nepotrebno trošio prostor za pohranu podataka te se u tu svrhu koristi detektorski sustav za okidanje, "trigger". Trigger je sustav koji može brzo donijeti odluku radi li se u zanimljivom događaju ili ne. Međutim, u pohranjenom "zanimljivom" događaju bit će zabilježeni i dodatni proton-proton sudari prisutni u tom događaju. Ovi neželjeni dodatni sudari nazivaju se "pileup" i predstavljaju problem kod rekonstrukcije hadronskih mlazova. "Pileup" čestice mogu se slučajno naći u blizini čestica hadronskog mlaza i tada ih algoritmi za rekonstrukciju mlazova mogu svrstati kao dio mlaza što će izobličiti njegova fizikalna svojstva. PUPPI algoritam [8] [9] razvijen je radi uklanjanja doprinosa neželjenih dodatnih čestica iz podataka.

Prije objašnjavanja načina rada PUPPI algoritma, potrebno je objasniti što je primarni verteks. Podaci o tragovima nabijenih čestica koriste se za rekonstruiranje točaka odakle su te čestice došle i iz toga se mogu odrediti točke proton-proton sudara. Točka u kojoj se dogodio sudar koji nas zanima naziva se primarni verteks. Dakle, za nabijene čestice možemo odrediti jesu li došle iz primarnog verteksa ili ne pa za one koje nisu možemo reći da su to pileup čestice.

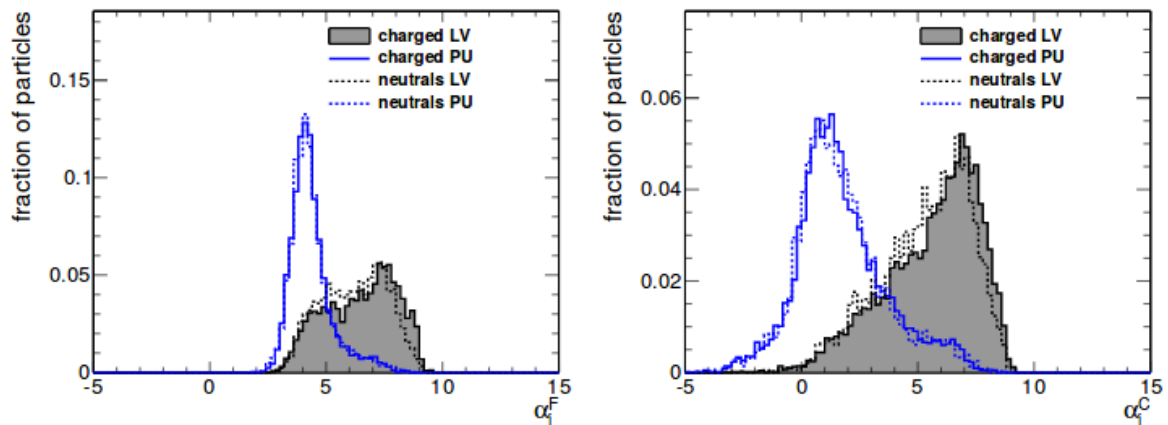
PUPPI algoritam za svaku česticu (neutralnu i nabijenu) izračuna veličinu  $\alpha$  koja se za česticu  $i$  definira kao:

$$\alpha_i = \log \sum_{j \in \text{Event}} \xi_{ij} \times \Theta(R_{\min} \leq \Delta R_{ij} \leq R_0), \quad (3.1)$$

gdje je  $\xi_{ij} = \frac{p_{Tj}}{\Delta R_{ij}}$  omjer transverzalnog impulsa čestice  $j$  mjenog u GeV-ima i  $\Delta R_{ij}$  udaljenost čestica  $i$  i  $j$  u prostoru pseudorapiditeta i azimutalnog kuta. Heaviside step funkcijom ( $\Theta$ ) zahtjevamo da je ta udaljenost između dva parametra uobičajeno postavljena na  $R_{\min} = 0.02$  i  $R_{\max} = 0.3$ . Iz jednadžbe 3.1 se vidi da će čestica imati veći  $\alpha$  što je više čestica u njenoj blizini. Prilikom hadronizacije, stvori se mnoštvo čestica unutar uskog stošca pa će čestice koje pripadaju hadronskom mlazu uglavnom i imati mnogo čestica u svojoj blizini. Uz to, događaji koji dolaze iz primarnog verteksa su energičniji te će čestice uglavnom imati veći transverzni impuls. Čestice iz "pileup" događaja će se naći u blizini drugih čestice tek slučajno i imat će manji

transverzni impuls te će stoga čestice iz primarnog verteksa uglavnom imati više vrijednosti  $\alpha$  nego čestice iz pileupa. Postoje dva skupa čestica ( $j$ ) preko kojih možemo vršiti sumiranje. Prvi skup je skup nabijenih čestica koje dolaze iz primarnog verteksa, a drugi je skup svih čestica u događaju. Prvi nam je skup dostupan samo u centralnom dijelu detektora pa se tada veličina označava sa  $\alpha^C$  ("central"), a tamo gdje nemamo dostupne podatke o tragovima nabijenih čestica (prednji dio detektora) preostaje nam sumirati po svim česticama  $\alpha^F$  ("forward").

Koristeći nabijene čestice iz centralnog dijela detektora (za njih nam je poznato jesu li došle iz primarnog verteksa ili ne) izračunaju se sljedeće raspodjele:  $\alpha_{LV}^C, \alpha_{PU}^C, \alpha_{LV}^F$  i  $\alpha_{PU}^F$  gdje LV označava raspodjelu za čestice iz primarnog verteksa, a PU za čestice iz pileupa. Nas zanimaju ove raspodjele za neutralne čestice i one se dobiju tako da pretpostavimo da su iste kao i za nabijene. Na Slici 3.1 nalaze se raspodjele dobivene Monte Carlo simulacijama na kojima se može vidjeti da je su distribucije za neutralne i nabijene čestice uistinu slične.



Slika 3.1: Distribucije  $\alpha^F$  i  $\alpha^C$  za različite vrste čestica. Punim su linijama označene nabijene čestice, a isprekidanim neutralne. LV označava čestice koje su došle iz primarnog verteksa, a PU one koje nisu.

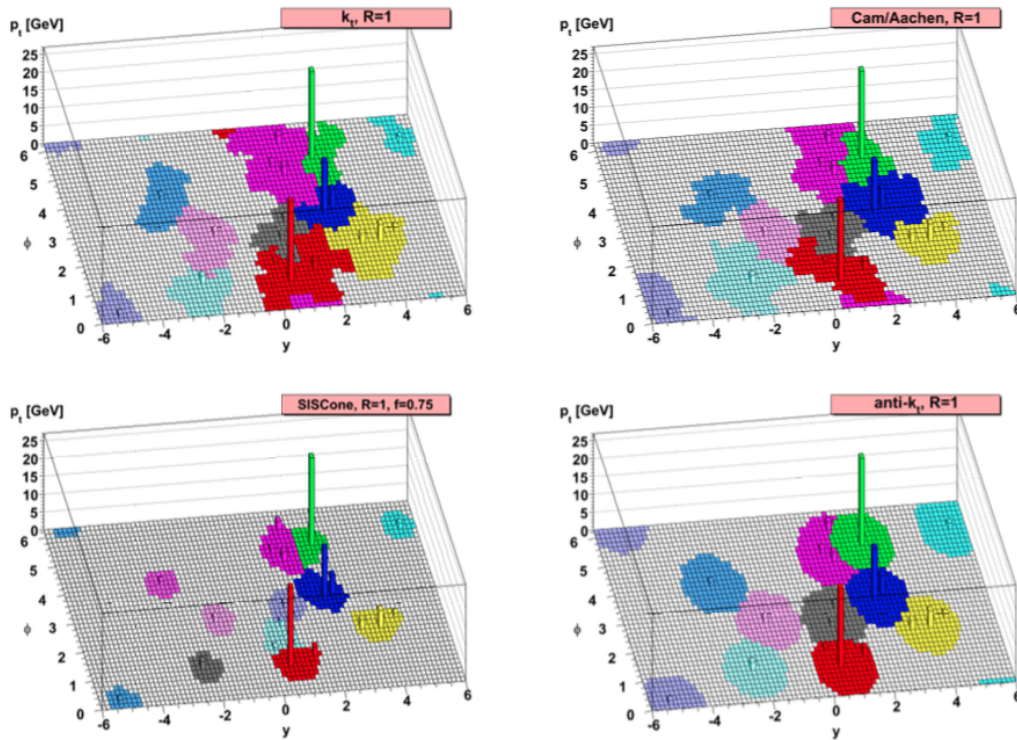
Svojstva dobivenih distribucija (medijan i RMS) tada se koriste za pridruživanje težina česticama. Težina se na česticu primjeni tako da se njen četveroimpuls pomnoži s težinom. Tako će nabijene čestice dobiti vrijednost 0 ili 1 ovisno o tome dolaze li iz pileup-a ili iz primarnog verteksa dok će neutralne čestice imati težine između te dvije vrijednosti.

### 3.3 Rekonstrukcija hadronskih mlazova

Particle-flow čestice koriste se za izgradnju hadronskih mlazova. Rekonstrukcija hadronskih mlazova može se izvršiti na više načina pa stoga postoji više algoritama za tu namjenu. 1990. godine postignut je dogovor o tome koja svojstva algoritmi za rekonstrukciju hadronskih mlazova trebaju zadovoljavati [10]. Bitna su svojstva infracrvena i kolinearna sigurnost. Ta svojstva zahtjevaju da se skup hadronskih mlazova neće promijeniti dodavanjem kolinearnog raspada ili nisko-energetske emisije u događaj.

Razlikujemo dvije vrste algoritama za rekonstrukciju hadronskih mlazova [11]. Prvu vrstu čine konusni algoritmi. Kod njih se pretpostavlja da će čestice hadronskog mlaza biti unutar stošca u prostoru azimutalnog kuta i rapiditeta, ali oni uglavnom nisu infracrveno sigurni. Popularnije verzije konusnih algoritama su iterativni konusni algoritam (Iterative cone - IC) i besjemeni konusni algoritam (Seedless Infra-red Sade Cone - SIScone). Prvi nije kolinearno siguran, ali drugi je i infracrveno i kolinearno siguran i jedina mu je slabost nemogućnost razlučivanja bliskih mlazova.

Druga vrsta algoritama su algoritmi sekvencijalnog grupiranja gdje se prilikom grupiranja čestica osim njihovog položaja u  $y - \phi$  prostoru u obzir uzima i transverzni impuls,  $p_T$  (dio impulsa čestice okomit na protonske snopove). Ova vrsta algoritama infracrveno je sigurna, ali prije se nije koristila zbog sporijeg vremena izvršavanja na računalima. Jednom kada se riješio problem brzine izvršavanja, algoritmi ove vrste postali su preferirani za korištenje. Kod ove vrste algoritama računaju se dvije "udaljenosti". Prva je udaljenost između dviju čestica  $d_{ij} = \min(p_{Ti}^a, p_{Tj}^a) \times \frac{R_{ij}^2}{R}$ , gdje je  $a$  eksponent čija se vrijednost mijenja ovisno o točno kojem se algoritmu sekvencijalnog grupiranja radi,  $R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$  udaljenost u prostoru  $y - \phi$  i  $R$  parametar radijusa koji određuje veličinu mlaza i obično je između 0.4 i 1.5. Druga udaljenost je  $d_{iB} = p_{Ti}^a$  što je transverzni impuls  $i$ -te čestice (udaljenost u prostoru impulsa između snopa protona i čestice  $i$ ). Algoritam tada traži najmanju vrijednost među  $d_{ij}$  ili  $d_{iB}$  među svim česticama. Ukoliko je najmanja vrijednost jedna od  $d_{ij}$ , tada se te čestice  $i$  i  $j$  spoje u jednu sumirajući njihove četveroimpulse. Ukoliko je  $d_{iB}$  minimum,  $i$  se proglašuje mlazom i uklanja se iz liste "čestica". Ovaj se proces ponavlja sve dok se sve čestice ne pridruže mlazovima čija je međusobna udaljenost ( $R_{ij}$ ) veća od  $R$ , i to se naziva inkluzivnim grupiranjem, ili se proces ponavlja dok se ne nađe određeni broj, i to se tada naziva ekskluzivnim grupiranjem. Algoritmi



Slika 3.2: Usporedba algoritama za rekonstrukciju hadronskih mlazova. Rekonstrukcija je izvršena na istim podacima, ali različitim algoritmima. Gore lijevo je prikazan  $k_T$ , a dolje desno anti- $k_T$  algoritam. Gore desno Cambridge/Aachen i dolje lijevo besjemeni konusni algoritam. Vidimo da anti- $k_T$  daje kružne mlazove, ali ukoliko se dva mlaza nađu jedan pored drugoga, mlaz s većim impulsom može "pojesti" manjeg do neke mjere. Preuzeto iz [11]

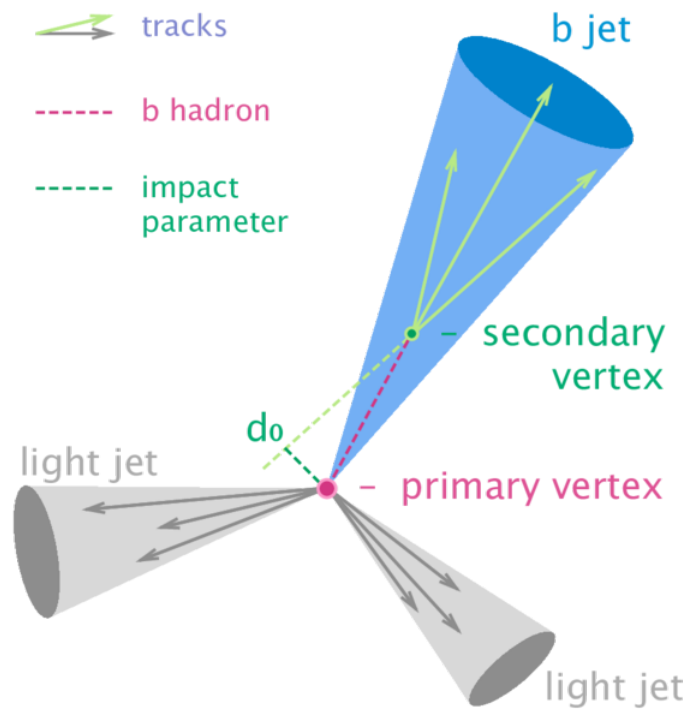
ove vrste međusobno se razlikuju po odabiru parametra  $a$ .  $k_T$  algoritam uzima vrijednost  $a = 2$  i pogledom na formulu kojom se računa  $d_{ij}$  vidimo da se preferiraju prvo grupirati mekše čestice što dovodi do toga da je podložan "pileup" događajima. Budući da se tek na kraju "spajaju" strukture s većim transverznim impulsom, ovaj je algoritam dobar za proučavanje podstrukture mlazova ("odvrtimo" unazad algoritam par koraka i možemo vidjeti hoće li se jedan mlaz razdvojiti na dva manja i sl.). anti- $k_T$  algoritam ima  $a = -2$  i, suprotno od  $k_T$  algoritma, grupira prvo tvrde čestice pa je manje podložan pileup-u, ali upravo zbog toga teško razlučuje podstrukturu jer pri kraju grupiranja dodaje meke čestice na periferiji mlaza. Prednost anti- $k_T$  nad drugima je ta što rekonstruirani mlazovi imaju kružni oblik. Algoritam koji postavlja  $a = 0$  naziva se Cambridge/Aachen algoritam i po ponašanju je sličan  $k_T$  jer je dobar za prepoznavanje podstrukture mlazova. Na Slici 3.2 mogu se vidjeti rezultati rekonstrukcije mlazova napravljene različitim algoritmima na istim podacima. Iz liste particle-flow čestica, prije rekonstrukcije mlaza, moraju se ukoniti neželjene



čestice koje su se slučajno našle u blizini drugih čestica hadronskog mlaza jer bi njih algoritam za rekonstrukciju pridružio mlazu. U tu se svrhu koristi PUPPI algoritam opisan u prethodnom potpoglavlju. PUPPI se koristi tek od nadogradnje piksel detektora, a prije njega je korištena CHS [9] (Charged Hadron Subtraction) metoda za čišćenje mlazova koja se pokazala inferiornijom u odnosu na PUPPI pri visokim luminozitetima.

## 4 Algoritam za dvostruko b označavanje

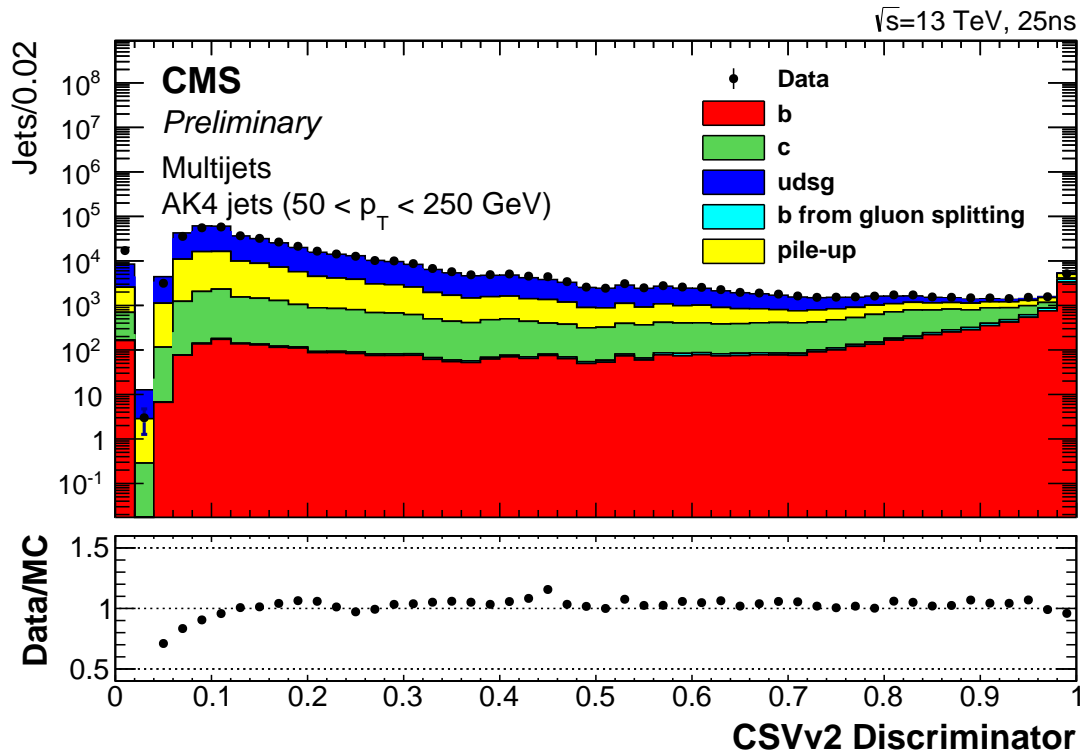
Spomenuli smo da se Higgsov bozon najčešće raspada u dva b kvarka od kojih nastanu dva b hadronska mlaza. Općenito, kod hadronskih mlazova nije moguće nedvosmisleno odrediti od kojeg je okusa kvarka potekao. Srećom, postoje neka svojstva po kojima se upravo b mlazovi razlikuju od ostalih. Možda najvažnije je to da hadroni koji nastaju prilikom hadronizacije b kvarka imaju relativno dugo vrijeme života,  $\tau \approx 1.5 \cdot 10^{-12}$  s. Ovo dugo vrijeme života omogućuje b hadronima da proputuju nekoliko milimetara prije nego li se raspadnu što se očituje u pojavi sekundarnog verteksa i tragova odmaknutih od primarnog verteksa kao što je prikazano na Slici 4.1



Slika 4.1: Prikaz b hadronskog mlaza. Na slici se može vidjeti stvaranje sekundarnog verteksa raspadom b mezona čija je udaljenost od primarnog verteksa mjerljiva i iznosi otprilike nekoliko milimetara. Ovo također dovodi do pojave odmaknutih tragova s povećanim parametrom sudara na slici označenim sa  $d_0$  za jednu od čestica.

Osim dugog vremena života, b hadroni imaju relativno veliku masu (5 GeV) kao i multiplicitet tragova (4–5), velike poluleptopnske omjere grananje — za elektrone i mione 20% svaki. Također, karakterizira ih tvrda funkcija fragmentacije što znači da veliki udio originalnog impulsa b kvarka odnosi b hadron.

Koristeći ova svojstva b hadrona, razvijeni su algoritmi za b označavanje koje svakom hadronskom mlazu pridruže vrijednost diskriminatora (obično između -1 i 1

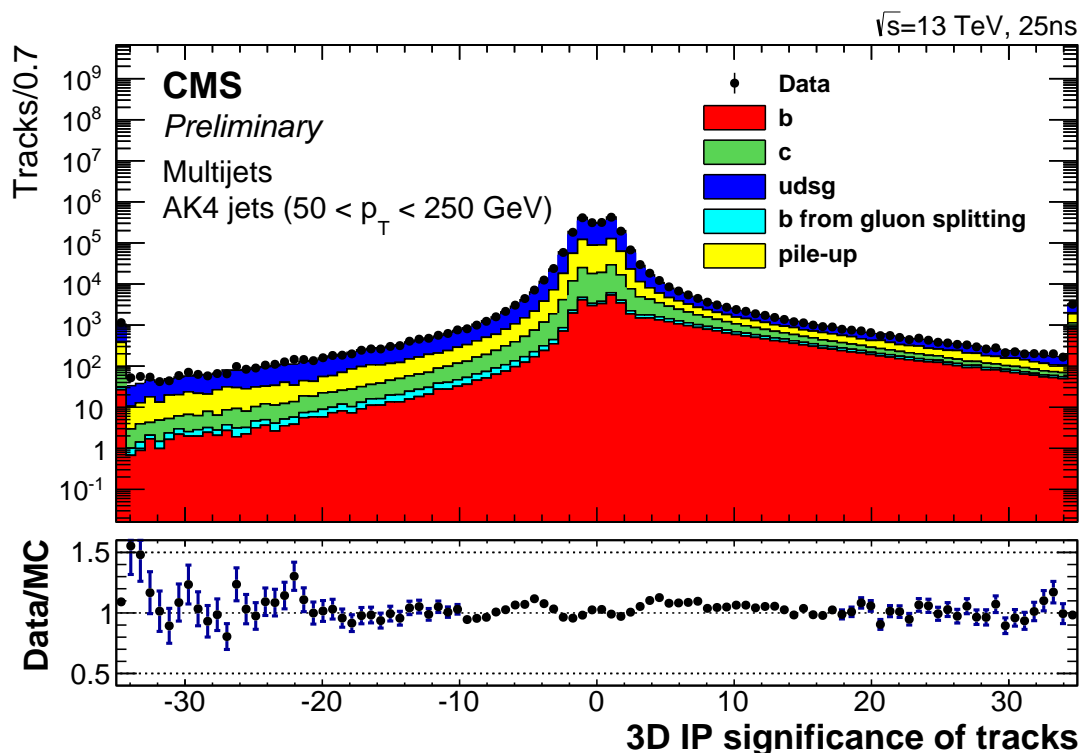


Slika 4.2: Raspodjela CSVv2 diskriminatora za b hadronske mlazove i različite vrste pozadinskih mlazova. Vidljivo je da algoritam pridružuje veće vrijednosti diskriminatora b mlazovima. Preuzeto iz [12]

ili 0 i 1) koja govori o vjerojatnosti da je hadronski mlaz došao od b kvarka. Postoji više algoritama za b označavanje koji se koriste unutar CMS-a i razlikuju se po tome što koriste za određivanje diskriminatora [12]. Npr. jedna vrsta koristi tragove pridružene hadronskom mlazu, druga koristi rekonstruirane vertekse dok treća kombinira obje vrste podataka. Jedan od takvih diskriminatora je CSVv2 [12] čija je raspodjela za b kvarkove i ostale pozadinske procese prikazana na Slici 4.2. Primjer varijable koja se koristi u CSVv2 je signifikantnost 3D parametra sudara koja se za tragove definira kao omjer parametra sudara traga i procjenjene greške parametra sudara. Tragovi pridruženi b kvarkovima teže imati nešto veću vrijednost signifikantnosti 3D parametra sudara kao što je i prikazano na Slici 4.3. Uz spomenute signifikantnosti 3D parametra sudara tragova u mlazu, koriste se još i:

- Signifikantnost udaljenosti preleta u ravnini okomitoj na snop protona
- Masa sekundarnog verteksa
- Broj tragova pridruženih verteksu

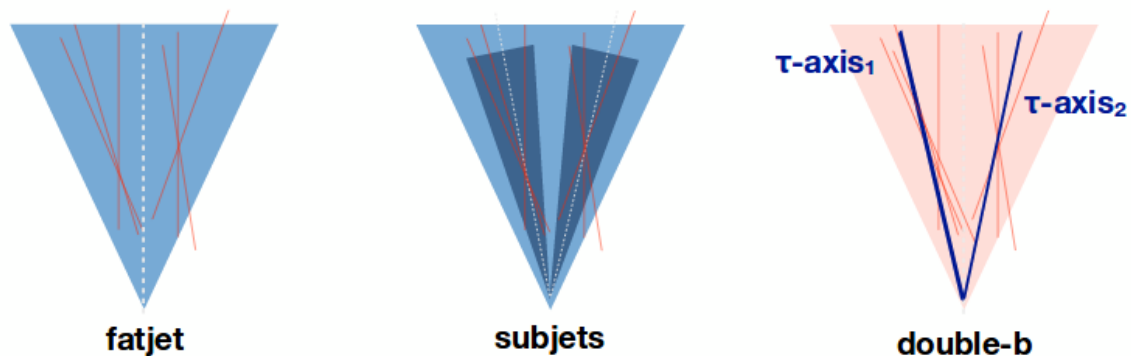
- Omjer energije tragova pridruženih verteksu i energije svih tragova u hadronskom mlazu
- Pseudorapiditeti tragova sekundarnog verteksa u odnosu na os hadronskog mlaza
- Signifikantnost parametra sudara u ravnini okomitoj na snop protona prvog traga koji podiže invarijantnu masu tragova pridruženih mlazu iznad 1.5 GeV (tragovi se poredaju od najveće do najmanje signifikantnosti parametra sudara i nakon svakog dodavanja traga, računa se invarijantna masa skupa tragova)
- Broj tragova u hadronskom mlazu
- Udaljenost u prostoru polarnog kuta i pseudorapiditeta između vektora smjera sekundarnog verteksa i osi mlaza
- Broj sekundarnih verteksa



Slika 4.3: Raspodjela signifikantnosti 3D parametra sudara za tragove pridružene b hadronskim mlazovima i tragove pridružene ostalim pozadinskim mlazovima. Preuzeto iz [12]

Algoritam za prepoznavanje b hadronskih mlazova koji su došli iz raspada ultra-relativističkog Higgsovog bozona donekle će biti specifičan zbog činjenice da se radi

o algoritmu za dvostruko b označavanje (dva mlaza), ali i zbog posebne topologije. Gledano iz sustava centra mase, b kvarkovi nastali raspadom Higgsovog bozona imat će impulse istog iznosa, ali suprotnog smjera. Međutim, u laboratorijskom sustavu to ne mora biti tako. Zbog zakona očuvanja količine gibanja, b kvarkovi imat će jedan dio impulsa u istom smjeru te će povećanjem impulsa Higgsovog bozona biti kolimiraniji. Posljedica toga je da se hadronski b mlazovi približe jedan drugome, a u ekstremnim slučajevima se počnu i preklapati kako je i prikazano na Slici 4.4. Budući da se u ovom radu proučavaju raspadni ultrarelativističkih Higgsovih bozona, to će svojstvo biti izraženo te će se zbog preklapanja mlazova malo izmijeniti način rekonstruiranja mlazova. Takvi se mlazovi rekonstruiraju AK8 metodom, odnosno anti- $k_T$  algoritmom radijusa 0.8 i nazivaju se Fat hadronski mlazovi ("Fat Jets"). Ovaj se radijus odnosi na veličinu stošca hadronskog mlaza i za obične se mlazove uglavnom koristi vrijednost  $R=0.4$ .

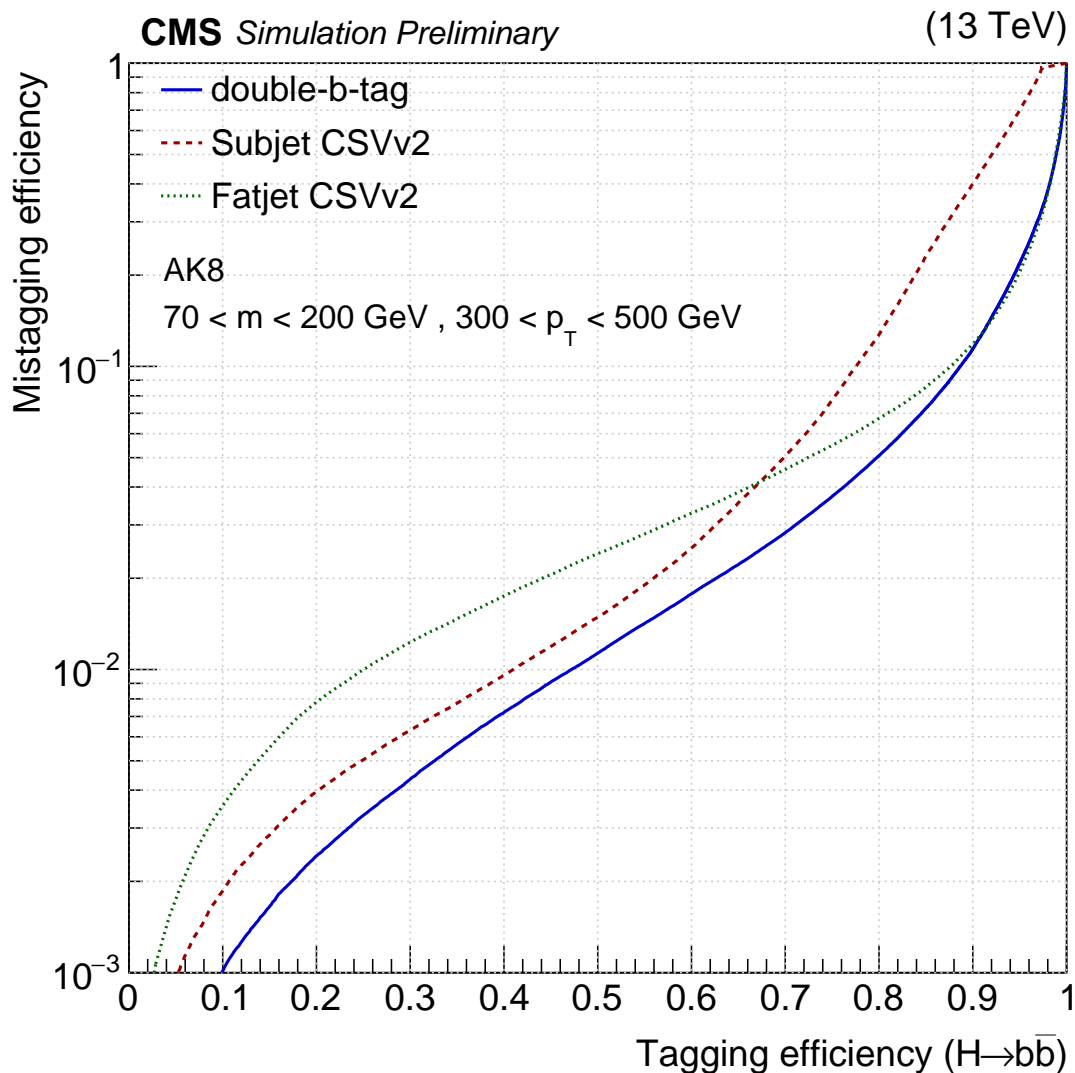


Slika 4.4: Konstrukcija fat hadronskog mlaza. Osi podmlazova koji ga čine nazivaju se  $\tau$  osima.

Algoritam za dvostruko b označavanje nije primjenjiv samo za raspad Higgsovih bozona u dva b kvarka, nego se on može koristiti za ultrarelativističke režime raspada čestica mase bliske masi Higgsovog bozona koje se raspadaju u dva b kvarka, npr. raspad Z bozona u dva b kvarka. Potrage za fizikom izvan Standardnog modela na sve većim energijama nužno će dovesti do produciranja čestica koje imaju puno veći impuls od njihove mase što može rezultirati već opisanom topologijom fat hadronskog mlaza gdje bi algoritam za dvostruko označavanje mogao biti primjenjiv. Ova se fleksibilnost postiže tako da se izbjegne postojanje jake ovisnosti performansi algoritma o masi i transverzalnom impulsu fat mlazova što se osigura biranjem uzo-

raka za treniranje algoritma koji imaju podjednak broj događaja u cijelom rasponu željenog transverzalnog impulsa.

Pokazalo se da algoritam za dvostruko b označavanje ima veću diskriminatorsku moć od običnih algoritama za b označavanje (Slika 4.5) što je i motivacija za daljnji razvoj algoritma za dvostruko b označavanje.



Slika 4.5: Usporedba performansi algoritma za dvostruko b označavanje, minimuma CSVv2 diskriminatora za dva podmlaza i CSVv2 verziji algoritma za fat hadronske mlazove. Preuzeto iz [13]

#### 4.1 Metoda strojnog učenja

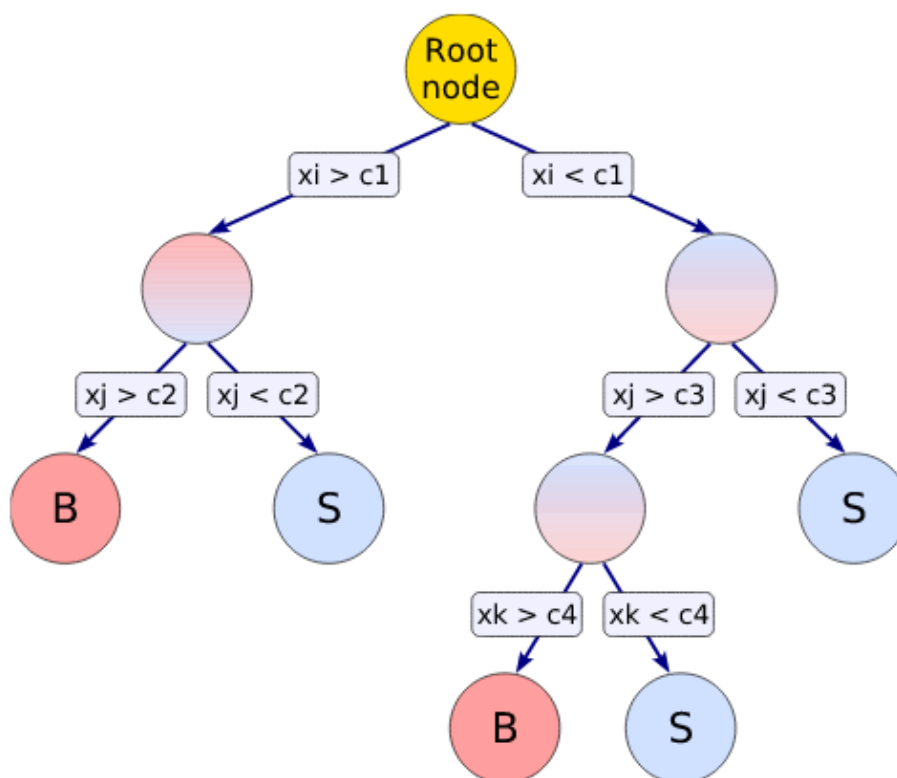
TMVA (Toolkit for Multivariate Analysis) programski paket pruža mogućnost korištenja mnogih metoda strojnog učenja za probleme klasifikacije i/ili regresije. Dizajniran je

upravo za potrebe fizike visokih energija, ali se može koristiti i u druge svrhe. Algoritam za b označavanje razvijat ćemo koristeći metodu *boosted* stabla odlučivanja (Boosted Decision Tree, u tekstu će se koristiti kratica BDT). Kao i sve druge metode, BDT pokušava naučiti razlikovati različite kategorije (u našem slučaju signal i pozadinu) proučavanjem podataka proslijeđenih metodi. To postiže donošenjem raznih odluka koje dijele podatke na jednu ili drugu vrstu. Ovaj se slučaj naziva klasifikacijskim stablom odlučivanja, a ukoliko bismo pokušali npr. predvidjeti neku veličinu u događaju ovisno o tome kolike su vrijednosti ostalih varijabli onda bi se radilo o regresijskom stablu odlučivanja. Slika 4.6 prikazuje jedan primjer stabla odlučivanja i vidimo da se uzastopnim donošenjem odluka prikaz širi, tj. grana po čemu je metoda i dobila ime.

Proces "rasta" stabla odlučivanja počinjemo s prosljeđivanjem podataka metodi. Podaci se dijele na one korištene u treningu i one korištene u evaluaciji metode. BDT tijekom treninga koristi odgovarajući skup podataka za "učenje" o ispravnim podjelama na signal i pozadinu. Nakon toga se dobiveno stablo odlučivanja testira na podacima koji nisu korišteni u treningu (set za evaluaciju) te se mogu iščitati njegove performanse.

Prva se odluka donese tako da se pogledaju sve varijable i odabere se podjela pomoću jedne varijable koja najbolje razdvaja signal od pozadine. Kvaliteta raspodjele može se procijeniti raznim kriterijima. Češće korišteni su Gini Index,  $G = p(1 - p)$  gdje je  $p$  udio signala u obuhvaćenom području, pogreška u klasifikaciji,  $K = 1 - \max(p, 1 - p)$  (gdje je  $p$  definiran kao i kod Gini Indexa) i drugi. Odlike svih kriterija su da postižu maksimum za maksimum pomiješanosti signala i pozadine ( $p = 0.5$ ) i poprimaju vrijednost nula ako se obuhvaćeni događaji sastoje samo od jedne vrste. Dakle, traži se podjela u jednoj od varijabli koja će minimizirati kriterij podjele i nakon toga podijeliti podatke u dva skupa.

Nakon prve podjele algoritam pokuša svaki skup opet podijeliti tako da minimizira kriterij podjele. Na taj način "raste" stablo odlučivanja. Ako bismo neometano pustili donošenje odluka, došli bi do toga da savršeno možemo razlikovati signal od pozadine, ali ako bismo testirali takvo stablo odlučivanja na skupu za evaluaciju, vidjeli bismo da su performanse loše jer je stablo odlučivanja naučilo specifičnosti samog skupa podataka za trening. Može se reći da je "zapamtilo" fluktuacije. Takva se greška naziva "overtraining" (pretreniranje) i da bi se ona spriječila može se koristiti



Slika 4.6: Primjer stabla odlučivanja. Serijom uzastopnih binarnih odluka odabiremo pravokutnike u hiperprostoru svih varijabli koje klasificiramo kao signal ili pozadinu.

nekoliko metoda. Jedan način je taj da se ograniči dubina stabla odlučivanja, drugi je da se pusti do relativno velike dubine, ali se onda "podrežu" zadnje, najmanje značajne odluke ("pruning").

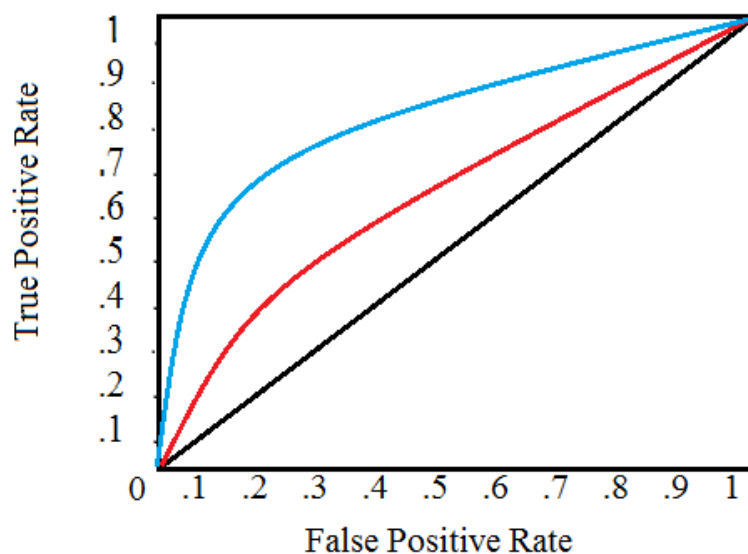
Metoda koja se pokazala kao najboljom za taj problem naziva se "boosting". Kod te se metode izgrađuje više stabala (šuma) ograničene dubine. Nakon svakog izgrađenog stabla, različiti događaji dobiju različite težine i to na način da događaji koji su pogrešno klasificirani dobiju veće težine, a ostali događaji se tada reskaliraju kako bi ukupan zbroj težina ostao isti. Tada se za konačnu odluku o nekom testnom događaju koristi prosjek odluka svih stabala. Ovaj se pristup u praksi pokazao najboljim.

Prednost BDT metode je što je jednostavnija za implementaciju od ostalih metoda za analizu više varijabli, ali ta jednostavnost utječe na to da su teoretski najbolje performanse BDT metode lošije u usporedbi s kompliciranijima poput neuralnih mreža. U praksi se ipak zna često dogoditi da BDT metoda da najbolje performanse jer nisu ispunjeni optimalni uvjeti za ostale. Na primjer, premali broj događaja za treniranje ili nije pronađena najbolja konfiguracije neuralne mreže. Također, BDT je neosjetljiv na dodavanje nevažnih varijabli (jednostavno neće raditi podjelu na temelju njih) što



može predstavljati problem kod drugih metoda.

Važan dio razvoja algoritma je evaluacija njegovih performansi. Performanse se uglavnom prikazuju ROC (receiver operating characteristic) krivuljom. Krenemo od skupa događaja za evaluaciju i podijelimo ga na signal i pozadinu. Svakom će događaju BDT pridružiti neku vrijednost diskriminatora (između -1 i 1) i što je ona veća, to je veća vjerojatnost da je događaj signal. Postavimo neku vrijednost diskriminatora koja će nam poslužiti kao uvjet za selekciju događaja (radna točka) i pogledamo koliko signalnih, a koliko pozadinskih događaja ispunjava taj uvjet. Zatim te brojeve podijelimo s ukupnim brojem signalnih, odnosno pozadinskih događaja da bismo dobili koliki udio koje vrste događaja zadovoljava taj uvjet. Dobivena dva broja su  $x, y$  koordinate jedne točke ROC krivulje. Mijenjanjem vrijednosti diskriminatora koja se koristi za selekciju i ponavljanjem ovog postupka dobiju se različite točke koje se tada mogu prikazati kao neprekidna krivulja. Primjer ROC krivulje može se vidjeti na Slici 4.7



Slika 4.7: Primjer ROC krivulje. Na  $y$  osi se nalazi udio signala koji prolazi selekciju na određenoj radnoj točki, a na  $x$  osi pozadina koja prolazi selekciju. Idealna ROC krivulja nalazila bi se u gornjem lijevom kutu pa se može zaključiti da diskriminator koji je prikazan plavom linijom ima bolje performanse od onog prikazanog crvenom. Crnom krivuljom prikazane su performanse koje se dobiju kada algoritam nije u stanju razlikovati signal od pozadine.

## 4.2 Simulacija i selekcija događaja

Prije razvoja samog algoritma potrebno je pripremiti Monte Carlo uzorke raspada Higgsovog bozona u dva  $b$  kvarka i pripadajuću pozadinu. Signalne su uzorke činili Monte Carlo simulacije raspada teških rezonancija (gravitona) u dva Higgsova bozona sa različitim pretpostavljenim masama gravitona (od 1000 do 4500 GeV) kako bi se pokrio širok prostor transverznih impulsa hadronskih mlazova. Pozadinu čine QCD hadronski mlazovi. Signalni su događaji simulirani MG5\_aMC@NLO generatorom događaja, a partonski pljusak i hadronizacija signalnih uzoraka napravljeni su pomoću Pythia8 generatora [14] [15]. Generiranje i hadronizacija QCD događaja provedeni su Pythia8 generatorom. Za simulaciju interakcija čestica s detektorom upotrijebljen je Geant4 [16].

Producirani podaci se zatim procesiraju kako bi poprimili oblik pogodan za treniranje algoritma nakon čega se prosljeđuju skripti za treniranje algoritma koja pokušava što bolje razlučiti signalne od pozadinskih mlazova koristeći Boosted Decision Tree metodu iz TMVA paketa. U treniranju algoritma korišteni su hadronski mlazovi koji zadovoljavaju sljedeće uvjete:

- "Pruned" masa [17] hadronskog mlaza veća od  $m_{\min} = 50$  GeV i manja od  $m_{\max} = 200$  GeV
- Transverzalni impuls mlaza veći od  $p_{T,\min} = 250$  GeV (ultrarelativistička granica)
- Apsolutna vrijednost pseudorapiditeta  $|\eta| < 1.5$

## 4.3 Treniranje algoritma za dvostruko $b$ označavanje

U algoritmima za dvostruko  $b$  označavanje standardno se koristi 27 varijabli koje su se pokazale najboljima za prepoznavanje hadronskih mlazova nastalih raspadom ultrarelativističkih Higgsovih bozona u dva  $b$  kvarka [13]. Neke se varijable odnose na svojstva ukupnog hadronskog mlaza, a neke na njegovu podstrukturu. Varijable koje se koriste su:

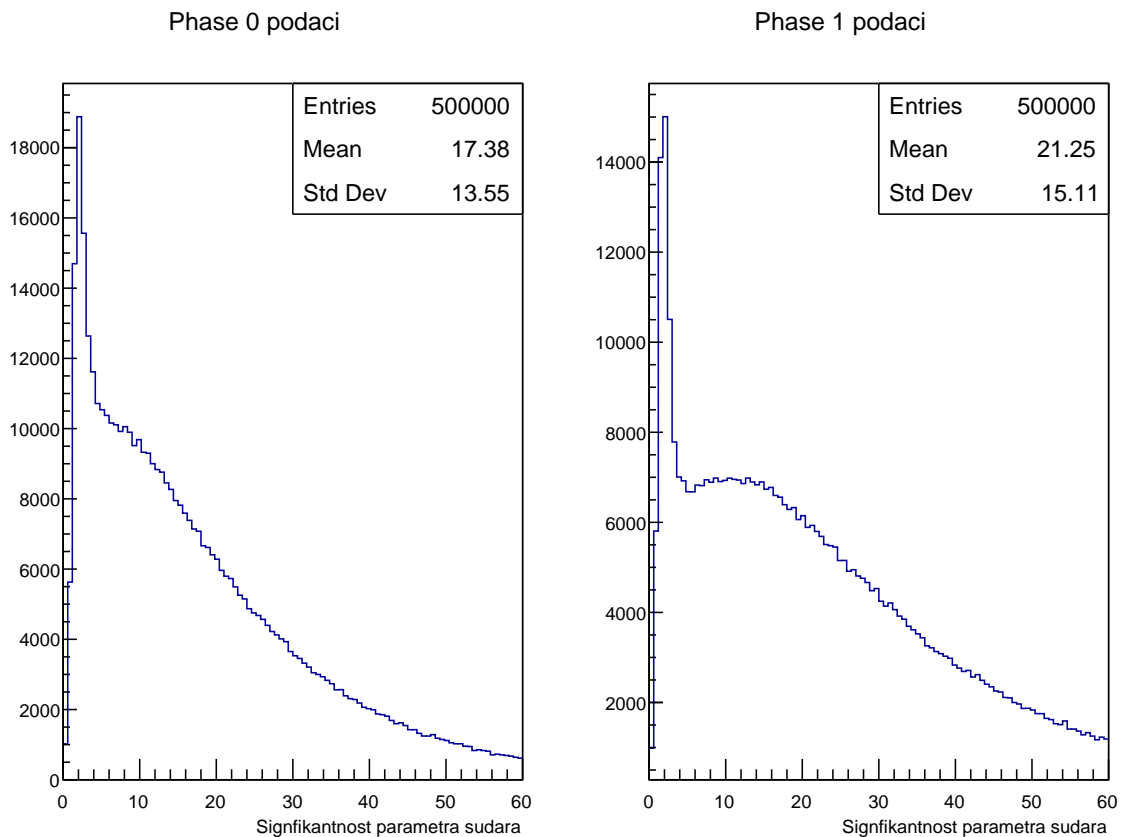
- Četiri najveće signifikantnosti parametra sudara za tragove koji pripadaju fat hadronskom mlazu poredane od veće signifikantnosti prema manjoj. Signifi-

kantnost parametra sudara definira se kao omjer parametra sudara i njegove nepouzdanosti.

- Za svaku os podmlazova ( $\tau$  os [18]) dvije najveće signifikantnosti parametra sudara.
- Signifikantnosti parametra sudara u ravnini okomitoj na snop protona prvih dviju tragova koji podižu invarijantnu masu sekundarnih verteksa iznad 5.2 GeV (prag b kvarka)
- Signifikantnost parametra sudara u ravnini okomitoj na snop protona prvog traga koji podižu invarijantnu masu sekundarnih verteksa iznad 1.5 GeV (prag c kvarka)
- Broj sekundarnih verteksa pridruženih hadronskom mlazu
- Signifikantnost udaljenosti između primarnog i sekundarnog verteksa, tj. duljine preleta u ravnini okomitoj na snop protona za sekundarni verteks s najmanjom nepouzdanošću trodimenzionalne duljine preleta; za obje  $\tau$  osi.
- $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$  između sekundarnog verteksa s najmanjom nepouzdanošću trodimenzionalne duljine preleta i pripadajuće  $\tau$  osi; za obje  $\tau$  osi
- Najmanja tri relativna pseudorapiditeta ( $\eta_{\text{rel}}$ ) tragova koji sačinjavaju sekundarne vertekse u odnosu na njihovu  $\tau$  os; za obje  $\tau$  osi
- Ukupna masa sekundarnih verteksa pridruženih  $\tau$  osi; za obje  $\tau$  osi
- Omjer energija sekundarnih verteksa pridruženih  $\tau$  osi i energije svih tragova iz primarnog verteksa pridruženih  $\tau$  osi; za obje  $\tau$  osi
- z-omjer;  $z = \Delta R(SV_0, SV_1) \cdot \frac{p_{T,SV_1}}{m(SV_0, SV_1)}$  gdje su  $SV_0$  i  $SV_1$  sekundarni verteksi s najmanjom nepouzdanošću trodimenzionalne duljine preleta. Varijabla je konstruirana za razlikovanje mlazova koji dolazi od konverzije gluona u par b kvarkova i mlazova koji dolaze od raspada teške rezonancije (poput Higgsovog bozona)

Prvi zadatak je ispitati performanse algoritma za dvostruko b označavanje prije i nakon nadogradnje piksel detektora. Postav CMS-a prije nadogradnje piksel detektora naziva se "Phase 0" (P0), a nakon nadogradnje "Phase 1" (P1). Budući da se

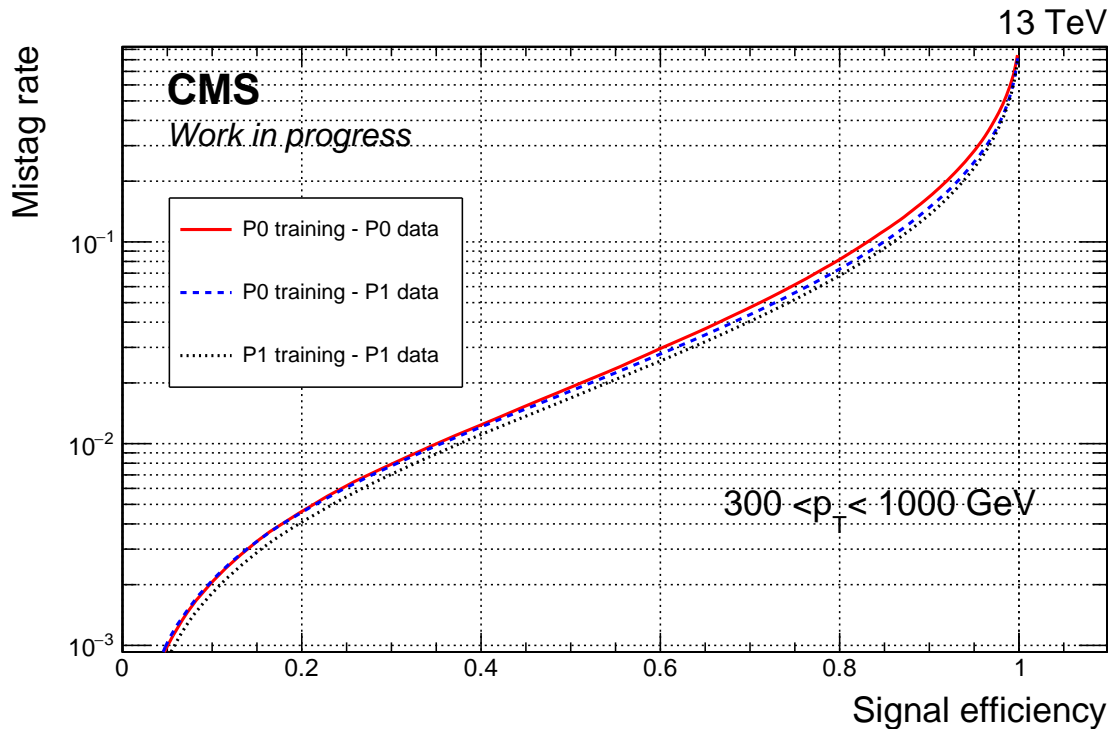
nadograđuje sustav za detekciju tragova nabijenih čestica, i to dio najbliži sudarima, očekivano je da će se tragovi mjeriti s većom preciznošću. Zbog toga očekujemo da će nadogradnja piksel detektora imati najviše utjecaja na signifikatnosti parametra sudara prikazanih na Slici 4.8.



Slika 4.8: Usporedba signifikantnosti parametra sudara prije (lijevo) i nakon (desno) nadogradnje piksel detektora. Dodavanjem još jednog sloja piksela preciznije se izmjere pozicije primarnog i sekundarnog verteksa što dovodi do preciznijeg određivanja parametra sudara što se vidi kao povećanje srednje vrijednosti signifikantnosti. Histogrami su napravljeni na signalnim uzorcima.

Preostaje provjeriti koji će utjecaj nadogradnja piksel detektora imati na učinkovitost algoritma za b označavanje. U svrhu toga uspoređuju se tri slučaja. Prvo se gleda učinkovitost algoritma u Phase 0 konfiguraciji, a zatim se isti algoritam primijeni na Phase 1 podatke. Time gledamo koliko (ako ikakvo) poboljšanje dobijemo preciznijim mjerenjima. Na kraju se trenira algoritam na Phase 1 podacima. Na taj način nastojimo vidjeti koliko poboljšanje dobijemo korištenjem preciznijeg mjerenja u treniranju.

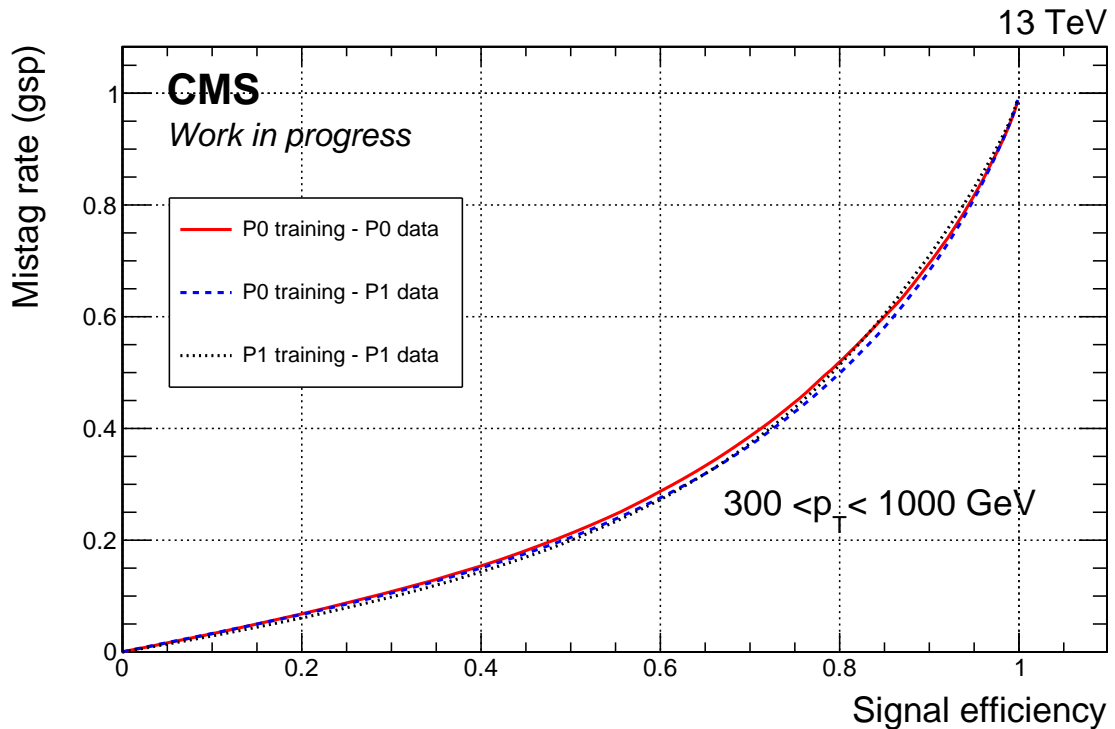
Sa Slike 4.9 vidljivo je da su performanse Phase 0 algoritma nešto bolje ukoliko se primijene na Phase 1 podatke i da se ponovnim treniranjem algoritma, ali na Phase



Slika 4.9: Usporedba algoritma za b označavanje u raznim konfiguracijama u rasponu transversalnog impulsa od 300 to 1000 GeV. Na x osi se nalazi učinkovitost prihvaćanja signala, a na y osi prihvaćanje pozadine. Osi su obrnute u odnosu na Sliku 4.7 pa su stoga idealne performanse u donjem desnom kutu.

1 podacima, performanse dodatno poboljšaju. Razlika u algoritmima ovisi o radnoj točki koju gledamo. Za primjer možemo fiksirati propuštanje pozadine na 3%. Phase 0 treniranje bi pri tome prihvaćalo oko 57% signala što se poveća na 60% Phase 1 treniranjem što je relativno poboljšanje za 5%. Zapravo se na Slici 4.9 može uočiti da je apsolutno poboljšanje algoritma gotovo svugdje oko 2 do 3%. Iz toga se može zaključiti da će se Phase 1 treniranjem najviše dobiti u radnim točkama visoke čistoće gdje je relativno poboljšanje prihvaćanja signalnih događaja oko 10%. Također se postavlja pitanje može li se Phase 1 treniranje iskoristiti na Phase 0 podacima. Poželjno je da se može jer bi tada Phase 1 treniranje moglo samo zamijeniti postojeće i izbjeglo bi se postojanje dvije verzije algoritma za dvostruko b označavanje u službenom CMS softveru. Ispostavilo se da Phase 1 verzija algoritma čak daje neznatno bolje performanse na Phase 0 podacima nego Phase 0 treniranje.

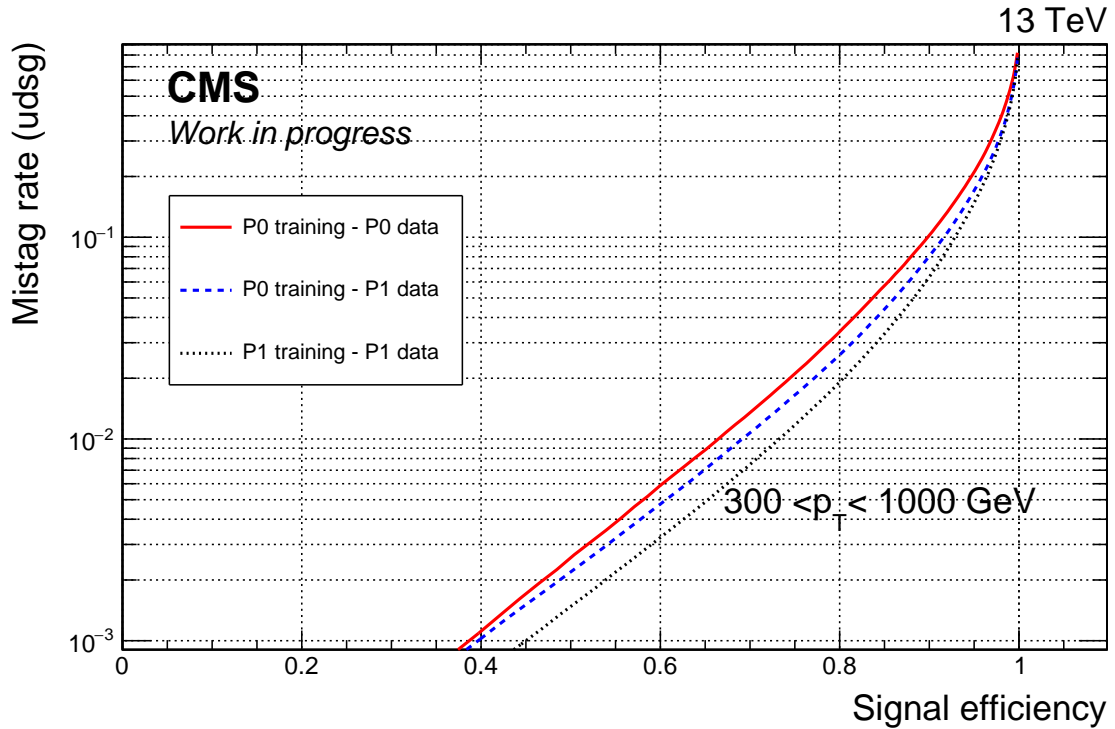
QCD pozadina za raspad Higgsovog bozona u dva b kvarka može se razložiti na pozadinu koja dolazi od konverzije gluona u par b kvarkova ("gluon splitting") i pozadinu lakših kvarkova. Korisno je pogledati performanse algoritma za pojedine pozadine. Sa Slika 4.10 i 4.11 vidi se da je konverzija gluona puno problematičnija



Slika 4.10: ROC krivulja algoritma za dvostruko b označavanje gdje je konverzija gluona (gluon splitting - gsp) u par b kvarkova stavljena za pozadinu. Na y-osi koristi se linearna skala za razliku od ostalih prikazanih ROC krivulja

od lakih kvarkova. Također se vidi da se nadogradnjom piksel detektora značajno poboljšaju mogućnosti diskriminacije lakih kvarkova od signala dok to nije slučaj za konverziju gluona. Ono što iznenađuje kod konverzije gluona je da se performanse Phase 1 treniranjem neznatno poboljšaju za male propusnosti, ali pogoršaju za visoku propusnost signala (i pozadine) što je pomalo neočekivano. Phase 1 podaci s Phase 0 treniranjem daju bolje performanse na širokom rasponu efikasnosti signala (nešto su lošije performanse na visokim efikasnostima signala) pa bi se moglo očekivati da će novo treniranje dati još bolje performanse, ali to bi moglo biti pogrešno. Naime, prilikom treniranja se uzima cijela pozadina, a ne samo konverzija gluona. Moguće je da je prilikom Phase 1 treniranja algoritam donio neke odluke koje nisu dobre za razlučivanje signala i gluonske pozadine, ali su dobre za razlikovanje signala od pozadine općenito. Da je treniranje bilo izvršeno na podacima koji su za pozadinu imali samo konverziju gluona, algoritam bi (vrlo) vjerojatno dao bolje performanse za taj tip pozadine, ali ne na ukupnoj pozadini. Iz ovoga se može zaključiti da nadogradnja piksel detektora posebice pomaže boljem razlikovanju b hadrona od lakih hadrona, ali ne pomaže u istoj mjeri razlikovanju različitih konfiguracija sa b hadro-

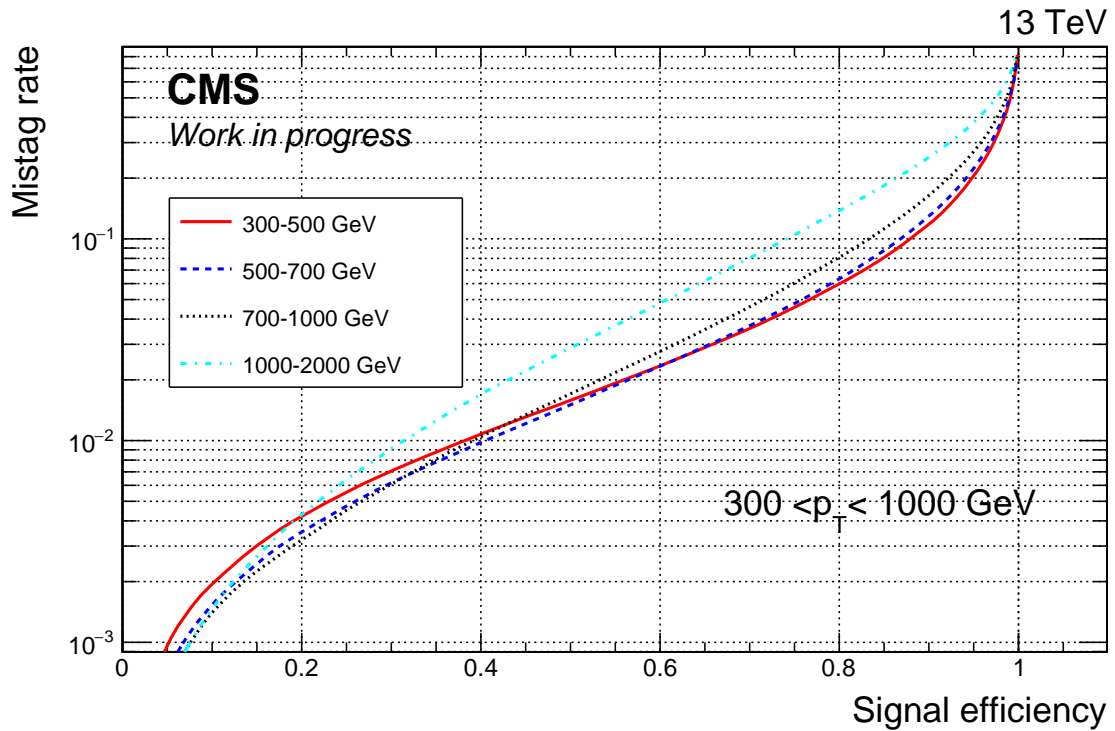
nima. Imajući na umu da je treniranje u oba scenarija vrlo slično, dobiveni rezultati upućuju na to da skup ulaznih varijabli ili neka opcija u BDT konfiguraciji nije optimalna za Phase 1 scenarij.



Slika 4.11: ROC krivulja algoritma za dvostruko b označavanje sa lakim kvarkovima (udsg) kao pozadinom.

Potrebno je još i ispitati stabilnost algoritma s obzirom na transverzni impuls hadronskih mlazova. Poželjno je da algoritam nema snažnu ovisnost o impulsu mlazova da bi bio upotrebljiv na širokom rasponu transverznih impulsa. Dobiveni se rezultati mogu pogledati na Slici 4.12. Performanse algoritma su relativno stabilne. Najbolje ponašanje se dobije za raspon od 500 do 700 GeV, a tek nešto lošije na rasponu 300 do 500 GeV. Razlog degradacije performansi algoritma na nižim impulsima može biti to što se produkti raspada Higgsovog bozona koji je imao niži impuls nalaze bliže rubovima fat hadronskog mlaza. Za visoke transverzne impulse počinju se lošije razlučivati podmlazovi u fat mlazu jer se počinju preklapati u većoj mjeri i uz to se smanjuje efikasnost rekonstrukcije tragova što također dovodi do degradacije algoritma koja je izražena na vrlo visokim impulsima (1000 GeV pa nadalje) što je bilo i očekivano.

BDT metoda ima nekoliko parametara čija se lista s objašnjenjima može pronaći u službenoj TMVA dokumentaciji [19]. Optimalne postavke parametara različite

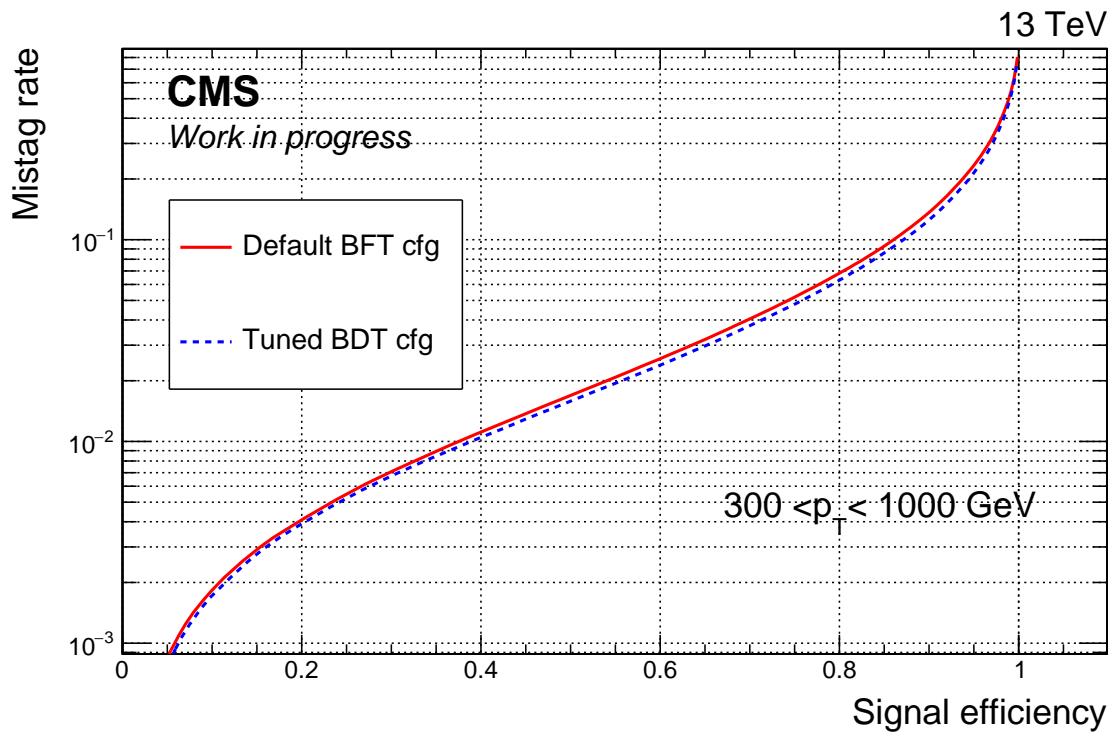


Slika 4.12: Usporedbe performansi algoritma za dvostruko b označavanje za različite raspone transverzalnog impulsa.

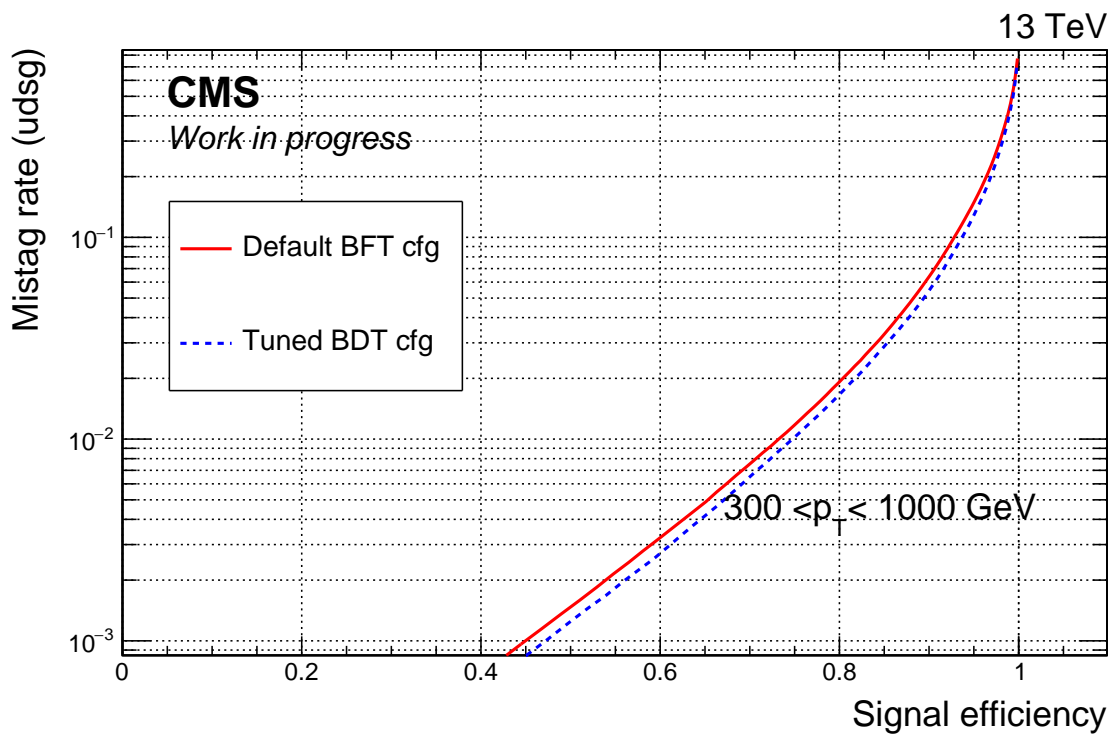
su za različite klasifikacijske probleme pa se stoga probalo vidjeti hoće li se rezultati poboljšati mijenjanjem parametara. Pojedinačnim mijenjanjem parametara nisu dobivena primjetna poboljšanja performansi osim kod povećanja broja stabala odlučivanja koji se koriste u BDT metodi. Simultanim mijenjanjem više parametara, Dodatak A, uspješno je dobiveno poboljšanje, prikazano na Slici 4.13.

Poboljšanje za pojedinu vrstu pozadine prikazano je na Slikama 4.14 i 4.15 s kojih se vidi da smo dobili poboljšanje za obje vrste pozadine. Neočekivan rezultat da trening na Phase 0 podacima (Slika 4.15) daje bolje rezultate za procese konverzije gluona prisutan je i nakon optimizacije parametara, ali je to područje znatno smanjeno.

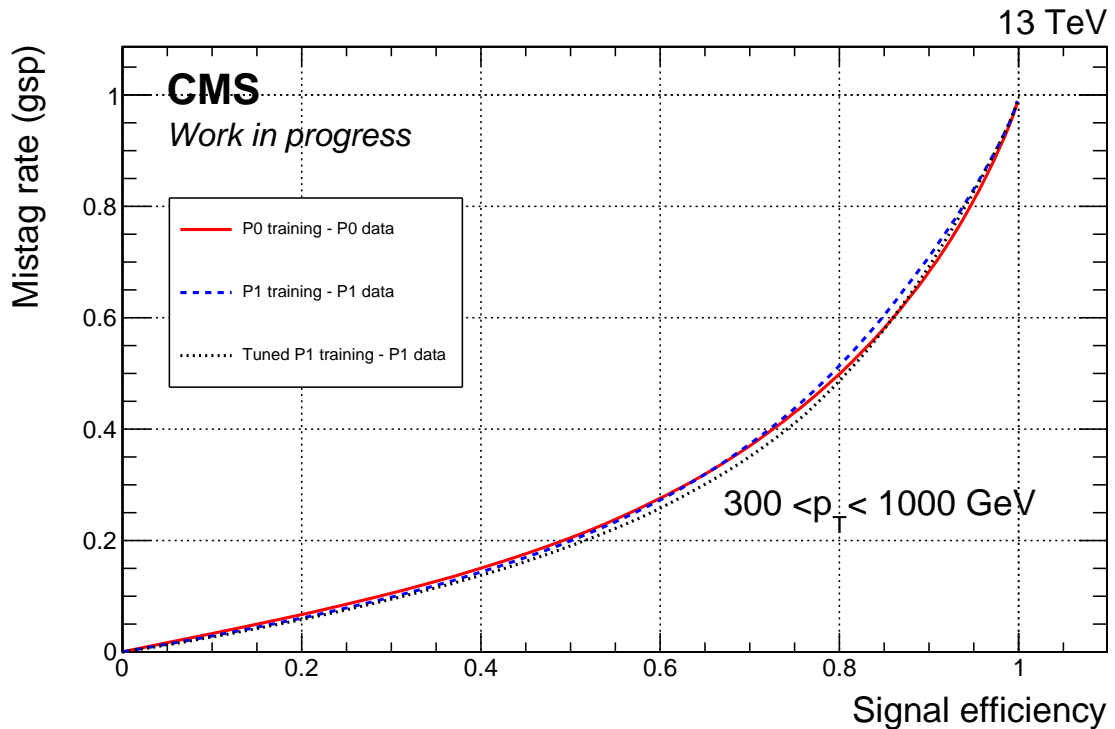




Slika 4.13: Poboljšanje performansi algoritma dobiveno optimiziranjem parametara BDT metode.



Slika 4.14: Usporedbi performansi algoritma prije i nakon optimizacije parametara metode BDT gdje pozadinu čine laki kvarkovi.



Slika 4.15: Usporedbi performansi algoritma prije i nakon optimizacije parametara metode BDT gdje pozadinu čine b kvarkovi koji su potekli od konverzije gluona. Prikazane su i performanse algoritma treniranog na Phase 0 podacima, ali primjenjenog na Phase 1 podacima i u jednom dijelu pokazuje bolje performanse nego algoritmi trenirani na Phase 1 podacima.

## 5 Zaključak

Glavni je cilj rada bio provjeriti ponašanje algoritma za dvostruko b označavanje nakon nadogradnje CMS piksel detektora. Za provjeru je bilo potrebno producirati Monte Carlo simulirane događaje raspada ultrarelativističkog Higgsovog bozona u dva b kvarka kao i pripadajuću QCD pozadinu i to u verzijama sa i bez nadograđenog piksel detektora. Ti su se podaci tada koristili za treniranje samog algoritma za dvostruko b označavanje korištenjem metode *boosted* stabla odlučivanja. Za treniranje je korišteno 27 varijabli korištenih u ranijim algoritmima za b označavanje. Usporedbom performansi algoritama utvrđeno je da nadogradnja piksel detektora poboljšava performanse za otprilike 5%. Iznos poboljšanja ovisi o izboru radne točke pa tako na radnim točkama niske propusnosti relativno poboljšanje doseže i do 10%, ali pada s povećanjem propusnosti algoritma. Također je utvrđeno da poboljšanje dolazi od boljeg razlučivanja b hadrona od lakih hadrona te da nadogradnja piksel detektora u manjoj mjeri poboljšava diskriminaciju različitih konfiguracija s b hadronima što

u konačnici ograničava ukupno poboljšanje. Testirano je i ponašanje algoritma s obzirom na transverzni impuls mlazova i degradacija algoritma postaje značajna tek na vrlo visokim transverznim impulsima, od 1000 GeV pa nadalje zbog smanjene efikasnosti rekonstrukcije tragova i značajnijeg preklapanja dvaju b hadronskih mlazova. Pronađena je i konfiguracija parametara BDT metode koja daje bolje rezultate od uobičajeno korištene.

Dodatno je poboljšanje traženo u promjeni načina rekonstrukcije verteksa čiji je učinak bio nedovoljno velik za uključivanje u algoritam. Za daljnju bi se optimizaciju mogla promijeniti sama metoda koja se koristi tijekom treniranja jer se u ranijim istraživanjima ispostavilo da dobro podešene duboke neuralne mreže mogu nadmašiti metodu stabla odlučivanja.

# Dodaci

## Dodatak A Konfiguracija parametara BDT metode

Parametri BDT metode korišteni u ovom radu navedeni su ispod. Značenje svakog parametra mogu se pogledati u TMVA korisničkom priručniku [19].

- Broj stabala odluke (NTrees)=300
- Minimalna veličina čvora (MinNodeSize)=2.5 %
- Maksimalna dubina pojedinog stabla (MaxDepth)=4
- Boost Type="Grad"
- Brzina učenja "Grad" metode (Shrinkage)=0.2
- Kriterij donošenja odluka (SeparationType)=MisClassificationError
- Broj točaka za podjelu raspona varijabli (nCuts)=20
- Metoda "podrezivanja" stabala (PruneMethod)=CostComplexity
- Snaga "podrezivanja" (PruneStrength)=2

Poboljšanje performansi algoritma za dvostruko b označavanje dobiveno je ovim promjenama:

- Broj stabala odluke=2000
- Maksimalna dubina pojedinog stabla=5
- Brzina učenja "Grad" metode=0.05
- Broj točaka za podjelu raspona varijabli=100

## Literatura

- [1] P. Higgs "Broken symmetries, massless particles and gauge fields". // Physics Letters 12.2, str. 132-133.
- [2] P. Higgs "Broken Symmetries and the Masses of Gauge Bosons". // Physical Review Letters 13.16, str. 508-509.
- [3] Chatrchyan, Serguei et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". // Physics Letters B716, str. 30-61.
- [4] Aad, Georges et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". // Physics Letters B716, str. 1-29.
- [5] Chatrchyan, S. et al. "The CMS Experiment at the CERN LHC". // Journal of Instrumentation, 2008. vol 3
- [6] CMS Physics TDR, Volume I: CERN-LHCC-2006-001, 2 February 2006
- [7] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET; CERN - CMS kolaboracija, travanj 2009., CMS-PAS-PFT-09-001
- [8] D. Bertolini, "Pileup Per Particle Identification", JHEP 1410 (2014) 59, doi: 10.1007/JHEP10(2014)059
- [9] "Pileup Removal Algorithms", CERN - CMS kolaboracija, 2014, CMS-PAS-JME-14-001"
- [10] Salam G.: "Towards jetography", arXiv:0906.1833
- [11] R. Atkin, "Review of jet reconstruction algorithms", J. Phys. Conf. Ser. 645 (2015), no. 1, 012008, doi:10.1088/1742-6596/645/1/012008.
- [12] "Identification of b quark jets at the CMS Experiment in the LHC Run 2", CERN - CMS kolaboracija, 2016, CMS-PAS-BTV-15-001
- [13] "Identification of double-b quark jets in boosted event topologies", CERN - CMS kolaboracija, 2016, CMS-PAS-BTV-15-002

- [14] Torbjörn Sjöstrand et al: "A Brief Introduction to PYTHIA 8.1", doi: 10.1016/j.cpc.2008.01.036
- [15] J. Alwall et al: "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", JHEP07(2014)079, doi: 10.1007/JHEP07(2014)079
- [16] S. Agostinelli et al: "Geant4-A simulation toolkit", Nucl. Instr. Meth. A506 (2003) 250
- [17] S. Ellis, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", Phys.Rev.D81:094023,2010
- [18] J. Thaller, "Maximizing Boosted Top Identification by Minimizing N-subjettiness", JHEP 1202 (2012) 093
- [19] TMVA User Guide (10.6.2017.), <http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf>