

**Kehittämistutkimus:
CERNin CMS-kokeen avoimen
tutkimusdatan hyödyntäminen
fysiikan laboratoriotyössä
yliopistotasolla**

Pro gradu -tutkielma, 10.2.2019

Tekijä:

PAAVO RIKKILÄ

Ohjaajat:

JAN SARÉN

JYVÄSKYLÄN YLIOPISTON FYSIIKAN LAITOS

KATI LASSILA-PERINI

FYSIIKAN TUTKIMUSLAITOS (HIP)



JYVÄSKYLÄN YLIOPISTO
FYSIIKAN LAITOS

Tiivistelmä

Rikkilä, Paavo

Kehittämistutkimus: CERNin CMS-kokeen avoimen tutkimusdatan hyödyntäminen fysiikan laboratoriotyössä yliopistotasolla

Pro gradu -tutkielma

Fysiikan laitos, Jyväskylän yliopisto, 2019, 131 sivua.

Tässä kehittämistutkimuksessa kehitettiin hiukkasfysiikan avointa tutkimusdataa hyödyntävä tietokoneella tehtävä laboratoriotyö fysiikan yliopisto-opetukseen. Tutkimuksella tavoiteltiin kolmenlaista tietoa: työn kehittämisesä kohdattavia haasteita ja valintoja, työn testatun version onnistumista fysiikan kokeellisena työnä sekä toimivan, avointa tutkimusdataa hyödyntävän työn kehittämisen suosituksia.

Teoreettisessa ongelma-analyysissa määritettiin kehittämisen tavoitteet, tarpeet ja haasteet, sekä nimettiin kehittämiseen liittyviä haasteita ja rajoitteita. Ensimmäisessä kehittämissvaiheessa laadittiin laboratoriotyön ensimmäinen versio. Kehittämissessä kohdattiin haasteita suuren datatiedoston käsittelyssä ja työn teknisessä toteutuksessa sekä valintoja ohjelmointiin ja sisältöön liittyen.

Toisessa ongelma-analyysissa työtä testattiin luonnontieteiden yliopisto-opiskelijoilla. Testauksista kerättiin aineistoa haastatteluiden, havainnoinnin ja kyselylomakkeiden avulla. Aineistoa analysoitiin teorialähtöisen sisällönanalyysin ja sisällön erittelyn keinoin. Analyysin perusteella työn ensimmäinen versio täytti tyydyttävästi tai melko hyvin neljä viidestä kokeellisten töiden tavoitekatgorioista, ja jätti yhden kategorian täyttämättä.

Toisessa kehittämissvaiheessa työtä kehitettiin tulosten perusteella. Kehittämistuotoksena saatiin Jyväskylän yliopiston fysiikan laitoksella käyttöön otettava laboratoriotyö, jossa tehdään yksinkertainen data-analyysi avoimella hiukkasfysiikan tutkimusdatalla. Kehittämistuotoksen lisäksi muodostettiin kehittämisen viitekehys tuleville opetusinterventioille. Viitekehysten perusteella onnistuneen, hiukkasfysiikan avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämiseen liittyviä suosituksia ovat mm. varautuminen kompromisseihin ohjelmoinnin määrässä, hiukkasfysiikan teorian ja oikeiden analyysikeinojen sisällyttäminen työhön sekä työn toteutus pareittain tai ohjaajan avustamana tehtäväksi. Lisäksi todettiin Python-kielen olevan teknisiltä ominaisuuksiltaan riittävä hiukkasfysiikan datan käsittelyyn.

Avainsanat: kehittämistutkimus, hiukkasfysiikka, CERN, avoin data, laboratoriotyö

Abstract

Rikkilä, Paavo

Design Research: Using the Open Data of CERN's CMS Experiment at Physics Practical Work at University Level

Master's thesis

Department of Physics, University of Jyväskylä, 2019, 131 pages.

In this design research a computer-based physics laboratory experiment using particle physics open data was designed for teaching physics at university. Three kinds of knowledge were aimed at: challenges and choices that will be faced during the process, information about how the tested product succeeded as a physics practical work and guidelines related to designing an experiment that uses open data.

In the theoretical problem analysis goals, needs and challenges of the design were defined, and challenges and limitations were named. In the first design procedure the first version of the laboratory experiment was created. During the procedure there were challenges about handling the large data file and related to the technical implementation of the experiment, and choices related to programming and content of the experiment.

In the second problem analysis the experiment was tested with university students of natural sciences. Material from the tests were collected with interviews, observation and questionnaires. Material was analysed with theory-based content analysis. Based on the analysis, the first version of the experiment reached satisfyingly or quite well the four out of five objectives of practical work, and left one unreached.

In the second design procedure the experiment was designed further based on results. The design solution was a laboratory experiment where a simple data-analysis with open particle physics data is done, and which will be implemented in Department of Physics, University of Jyväskylä. Furthermore a design framework for future interventions was formed. Based on the framework, guidelines related to designing a successful laboratory experiment using particle physics open data are e.g. preparing for compromises on the amount of programming, including particle physics theory and real analysis methods to the experiment and implementation of the experiment on the way that it could be done with another student or with the help of an instructor. Also the Python programming language's technical features were stated to be sufficient for handling particle physics data.

Keywords: design research, particle physics, CERN, open data, practical work

Esipuhe

Olen kiitollinen siitä, että sain tehdä pro gradu -tutkielman mielenkiintoisesta aiheesta sekä luoda jotakin konkreettista. Prosessi oli hyvin opettavainen ja siihen osallistui monia ihmisiä, joita haluan kiittää.

Kiitokset ohjaajilleni Jan Sarénille ja Kati Lassila-Perinille kommentteista, ajatuksista ja tuesta. Työnsitte aina sopivissa kohdissa oikeaan suuntaan. Jouni Viirille kiitos aineistonkeruun vinkeistä. TIM-alustan kehittäjätiimiä kiitän nopeista ja kattavista vastauksista teknisten haasteiden kanssa. Fysiikan tutkimuslaitokselle (HIP) kiitos harjoittelupaikasta ja CERN-seikkailuista sekä Jyväskylän yliopiston fysiikan laitokselle mainiosta opiskeluajasta. Tahdon myös kiittää perhettäni ja ystäviäni kannustuksesta. Ilonalle kiitokset kirityksestä ja onnittelut gradukilpailun rehellisestä voitosta.

Sisältö

Tiivistelmä	3
Abstract	5
Esipuhe	7
1 Johdanto	11
2 Kehittämistutkimus	13
2.1 Kehittämistutkimus tutkimusmenetelmänä	13
2.2 Tutkimuksen toteutus ja tutkimuskysymykset	18
2.3 Kehittämistutkimuksen luotettavuus ja haasteet	21
3 Teoreettinen ongelma-analyysi	23
3.1 Hiukkasfysiikka	23
3.2 Kokeellinen hiukkasfysiikka CERNissä	25
3.2.1 CERN	25
3.2.2 CMS-ilmaisim	27
3.2.3 Hiukkasten epäsuora havaitseminen	29
3.3 Avoin data	33
3.4 TIM-alusta	36
3.5 Hiukkasfysiikka Jyväskylän yliopistossa	36
3.6 CERNin avoimen datan opetuskäyttö	37
3.7 Fysiikan käytännön töiden tavoitteet	40
3.8 Teoreettisen ongelma-analyysin yhteenveto	41
4 Ensimmäinen kehittämisvaihe	45
4.1 TIM-alustaan tutustuminen	45
4.2 Työn rakenteen luominen	48
4.2.1 Teoriaosuus	49
4.2.2 Datan hakeminen	50
4.2.3 Invariantin massan laskeminen	53
4.2.4 Histogrammin laatiminen	55
4.2.5 Funktion sovittaminen histogrammiin	57
4.2.6 Histogrammin analysointi	61
4.2.7 Koko datatiedoston histogrammi	62
4.3 Ensimmäisen kehittämisvaiheen yhteenveto	65

5	Toinen ongelma-analyysi	67
5.1	Aineistonkeruumenetelmät	67
5.1.1	Haastattelu	67
5.1.2	Havainnointi	69
5.1.3	Kyselylomake	70
5.2	Kuvailu ensimmäisestä testauksesta	71
5.3	Kuvailu toisesta testauksesta	74
5.3.1	Testauksen valmistelu	74
5.3.2	Pohjaluento	76
5.3.3	Työn tekeminen	77
5.3.4	Verkkokyselylomake	81
5.3.5	Haastattelut	82
5.4	Sisällönanalyysi ja sisällön erittely	83
5.5	Analyysin tulokset	87
5.5.1	Esitietolomake	87
5.5.2	Fysiikan kokeellisten töiden tavoitteet	90
5.5.3	Työssä kohdatut haasteet	97
5.5.4	Työssä kehitettävät asiat	99
5.5.5	Työssä onnistuneet asiat	102
5.6	Toisen ongelma-analyysin yhteenveto	102
6	Toinen kehittämisvaihe	109
7	Kehittämistuotos ja luotu kehittämisen viitekehys	123
8	Johtopäätökset ja pohdinta	127
8.1	Tutkimuskysymyksiin vastaaminen	127
8.2	Tutkimuksen eettisyys ja luotettavuus	129
8.3	Suosituksat jatkotutkimuksille	131
	Lähteet	133
A	Laboratoriotyön Jyväskylässä testattu versio	A-1
B	Ensimmäisten testaaajien informointi- ja esitietolomake	B-1
C	Toisten testaaajien informointi- ja esitietolomake	C-1
D	Pohjaluento	D-1
E	Työn osiot ”Funktion sovittaminen histogrammiin” ja ”Histogrammin analysointi” toisen kehittämisvaiheen jälkeen	E-1
F	Työhön toisessa kehittämisvaiheessa lisätty osio	F-1

1 Johdanto

Tämä fysiikan pedagoginen pro gradu -tutkielma käsittelee kehittämistutkimusta, jossa kehitetään CERNin CMS-kokeen avointa tutkimusdataa hyödyntävä laboratoriotyö fysiikan yliopisto-opetukseen. CERN, Euroopan hiukkasfysiikan tutkimuskeskus, on julkistanut keräämäänsä hiukkasfysiikan tutkimusdataa vapaaseen yleiseen käyttöön [1]. Varsinaisen tutkimuksen lisäksi data soveltuu esimerkiksi hiukkasfysiikan opetukseen [2].

Kehittämistutkimus on monitahoinen tutkimusmenetelmä, jonka tavoitteena on suunnitella ja kehittää ratkaisuja todellisista opetustilanteista nouseviin haasteisiin tai mahdollisuuksiin [3] [4]. Kehittämistutkimus sekoittaa kokeellista opetuksen tutkimista teoriavetoiseen oppimisympäristöjen suunnitteluun [5]. Kehittämistutkimuksella pyritään yleistettävyyteen sekä käytettävien ja yleistettävien teorioiden luomiseen [4].

Tässä kehittämistutkimuksessa kehitetään Jyväskylän yliopiston fysiikan laitoksen modernin fysiikan kurssille hiukkasfysiikan avointa dataa hyödyntävä laboratoriotyö. Työtä testataan luonnontieteiden yliopisto-opiskelijoilla. Testauksista kerätään aineistoa haastatteluiden, havainnoinnin ja kyselylomakkeiden avulla. Aineistoa analysoidaan teorialähtöisen sisällönanalyysin ja sisällön erittelyn keinoin. Analyysin tulosten perusteella arvioidaan, miten kehitetyn työn testattu versio täytti Nivalaisen, Asikaisen ja Hirvosen [6] vuonna 2013 määrittelemät fysiikan kokeellisten töiden tavoitekategoriat. Arvioinnin perusteella työtä kehitetään edelleen. Toinen kehittämisvaihe johtaa valmiiseen kehittämistuotokseen.

Tutkimuksen kehittämistuotos on tietokoneella tehtävä laboratoriotyö, joka otetaan käyttöön Jyväskylän yliopiston fysiikan laitoksella. Työ tehdään Jyväskylän yliopiston tietotekniikan laitoksen TIM-alustalla. Työn tavoitteena on yksinkertaisen data-analyysin tekeminen aidolla, CMS-ilmaisimen keräämällä tutkimusdatalla. Työssä tehtävänä on todeta Z^0 -bosonin esiintyminen sekä selvittää Z^0 :n massa. Yliopistolle laadittavasta työstä julkaistaan myös englanninkielinen versio vapaaseen yleiseen käyttöön.

Kehittämistuotoksen lisäksi tutkimuksessa luodaan kehittämisen viitekehys tuleville hiukkasfysiikan avoimen datan opetusinterventioille. Viitekehyksessä määritellään tutkimuksen tulosten perusteella suosituksia onnistuneen, hiukkasfysiikan avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämiseen. Lisäksi tutkimuksen kehittämispöytäkirjan kuvailusta saadaan tietoa siitä, mitä haasteita ja valintoja kohdataan, kun kehitetään hiukkasfysiikan avointa tutkimusdataa hyödyntävää laboratoriotyötä.

Tutkimuksen raportti noudattaa kehittämistutkimuksen teorian mukaista rakennetta. Raportin alussa tutustutaan kehittämistutkimuksen teoriataustaan ja tutkimuksen toteutukseen. Menetelmän esittelyä seuraa kuvaus kehittämisprosessin eri vaiheista kronologisessa järjestyksessä. Tämän jälkeen esitellään saatu kehittämistulos. Raportin lopussa käsitellään tutkimustehtävien täyttymistä sekä pohditaan saatuja tuloksia.

2 Kehittämistutkimus

Tässä luvussa tarkastellaan *kehittämistutkimusta*, jota käytetään tutkimuksen tutkimusmenetelmänä. Luvussa käsitellään menetelmän teoriataustaa, käytännön toteutusta sekä luotettavuutta ja haasteita. Tutkimus toteutetaan Edelsonin [4] kehittämistutkimuksen määrittelyä soveltaen.

Kehittämistutkimukseen perehdytään sekä Pernaan [7] toimittaman suomenkielisen kokoomateoksen *Kehittämistutkimus opetuslalla* että kehittämistutkimuksesta kertovien tieteellisten artikkelien kautta.

2.1 Kehittämistutkimus tutkimusmenetelmänä

Artikkelissaan *Kehittämistutkimus tutkimusmenetelmänä* Perna [3, s. 12] kertoo kehittämistutkimuksesta seuraavasti: ”Kehittämistutkimus on monitahoinen tutkimusmenetelmä, jolle ei voida esittää yhtä yksiselitteistä määritelmää.” Kehittämistutkimuksen tavoitteena on ”– – suunnitella ja kehittää ratkaisuja todellisista ope-
tustilanteista nouseviin haasteisiin.” [3, takakansi]

Kehittämistutkimusta kutsutaan suomeksi myös *design-tutkimukseksi*. Englanniksi kehittämistutkimuksesta käytetään mm. termejä *design research* ja *design-based research*. Lisäksi 1990-luvun alun alussa menetelmästä käytettiin nimeä *design experiment*. [3]

Design-Based Research Collective [5] määrittelee kehittämistutkimuksen tutkimukseksi, joka sekoittaa kokeellista opetuksen tutkimusta teoriavetoiseen oppimisympäristöjen suunnitteluun. Heidän mielestään kehittämistutkimus on tärkeä menetelmä tutkittaessa opetuksellisten innovaatioiden toimivuutta käytännössä. [5, s. 5]

Kehittämistutkimuksen synty ajoittuu vuoteen 1992, jolloin julkaistiin esimerkiksi Brownin [8] ja Collinsin [9] tutkimusartikkelit ensimmäisten menetelmästä julkais-
tujen artikkelien joukossa [3]. Brown kuvailee kirjoituksessaan pyrkimystään luokkatilojen monitasoiseen uudistamiseen, sekä niihin liittyvien käsitteiden uudelleen-
käsitteellistämiseen. Lisäksi Brown pohtii monimutkaisen luokkatilan tutkimiseen liittyviä teoreettisia ja menetelmällisiä ongelmia. Hän kertoo kehittämistutkijana yrittävänsä kehittää innovatiivisia oppimisympäristöjä ja samalla pyrkivänsä tutki-
maan niitä kokeellisesti. Hän pyrkii myös perustamaan interventionsa¹ teoreettisille

¹Interventio voidaan käsittää viittauksena kaikkiin kokonaisuuksiin, joita voidaan suunnitella ja kehittää [10].

kuvailuille, jotka määrittelevät miksi interventiot toimivat. Tämä tekee interventioista luotettavia ja toistettavia. [8, s. 141–144]

Collinsin [9] artikkeli käsittelee tavoitetta rakentaa systemaattisempi menetelmä kehittämistutkimusten ohjaamiseksi. Collins tavoittelee myös teorian kehittämistä ohjaamaan tulevaisuuden innovaatioiden toteutusta. Collinsin mukaan opetuksen kehittämistutkimuksen tulisi olla suunnitelmallinen tiede (design science). Hän vertaa kehittämistutkimusta mm. ilmailuun, jossa tavoitteena on tutkia miten erilaiset mallit vaikuttavat esimerkiksi nosteeseen ja ohjattavuuteen. Samalla tapaa opetuksen kehittämistutkimuksen tulisi määritellä miten erilaiset oppimisympäristöjen suunnitelmat vaikuttavat esimerkiksi oppimiseen, yhteistyöhön ja motivaatioon. [9, s. 14–15]

Edelson [4] kertoo artikkelissaan *Design Research: What We Learn When We Engage in Design*, että kehittämistutkimuksessa kehittämistä pidetään teorioiden kehittämisen ja jalostamisen strategiana. Kehittämistutkimus mahdollistaa ainutlaatuisen oppimisen, tuottaa käytännöllisiä ja välittömästi sovellettavia oppeja sekä sitouttaa tutkijoita kasvatuksellisten käytäntöjen parantamiseen. Menetelmän tärkeä piirre on Edelsonin mielestä se, että kehittämistutkimus häivyttää rajaa kehittämisen ja tutkimuksen välillä. Lisäksi kehittäminen mahdollistaa tutkijoiden ymmärryksen kasvamisen. [4, s. 105–106]

Tämä tutkimus toteutetaan Edelsonin kehittämistutkimuksen määritelmää soveltaen. Seuraavaksi tutustutaan tarkemmin Edelsonin määritelmiin kehittämistutkimukseen liittyen.

Kehittämistutkimus alkaa usein ongelmasta tai mahdollisuudesta, johon halutaan vastata jollakin idealla, joka usein täsmentyy kehittämisprosessin edetessä [4, s. 109]. Kehittämisprosessin Edelson määrittelee sarjaksi päätöksiä, joilla tasapainotellaan tutkimuksen tavoitteiden ja rajoitteiden välillä. Jokaisessa kehittämisessä tulee pohdittua, ei välttämättä suoraan tai formaalisti muotoiltuna, seuraavia lopputulokseen vaikuttavia päätöksiä: [4, s. 108]

- Mihin tarpeisiin ja mahdollisuuksiin kehittäminen vastaa?
- Miten kehittämisprosessi etenee?
- Minkä muodon kehittämisen tulos saa?

Edelson nimeää nämä päätökset (järjestyksessä) *ongelma-analyysiksi, kehittämisprosessiksi* sekä *kehittämistuotokseksi*.

Ongelma-analyysi määrittelee kehittämisen tavoitteet, tarpeet tai mahdollisuudet, joihin tutkimuksella pyritään vastaamaan. Lisäksi ongelma-analyysiin kuuluu kehittämiseen liittyvien haasteiden, rajoitteiden ja mahdollisuuksien nimeäminen. Tyypillisesti ongelma-analyysi tulee myös muuttumaan kehittämisen edetessä. [4, s. 109]

Kehittämisprosessi määrittelee kehittämisen rakentamiseen osalliset prosessit ja ihmiset. Tutkijoiden täytyy usein kehittää sarja prosesseja, joiden tulee vastata mm.

sunnittelun, kehityksen, toteutuksen ja arvioinnin tarpeisiin. Tutkijan tulee myös koota joukko oleellisia asiantuntijoita. Yksittäisissä tutkimuksissa tavoitteet ja rajoitteet määrittelevät tarvittavat prosessit ja asiantuntemuksen. [4, 108–109]

Kehittämistuotos kuvailee lopputuloksena saatua mallia. Kehittämistuotos on tutkijan työn ansiosta syntynyt tulos. Tuotosta muodostaessaan tutkijat usein hajottavat monimutkaisen kehittämisiongelman helposti käsiteltäviin komponentteihin. Myös kehittämistuotos voi kehittyä kehittämisprosessin edetessä, kun tutkijat saavat lisää tietoa kehittämiskontekstista. [4, s. 109]

Edelson korostaa, että hän ei pyri edellä mainituilla osa-alueilla määrittelemään kehittämisen prosessiteoriaa. Ennemmin kehittämisen määrittely näiden osa-alueiden kautta on tapa luonnehtia kehittämisen tilaa missä vaiheessa kehitystä tahansa. Eri osa-alueiden ei tarvitse olla konkreettisesti olemassa, vaan ne voivat esiintyä epäsuorasti tutkijoiden toimissa. [4, s. 109]

Kehittämistutkimuksessa pyritään yleistettävyyteen. Tämä erottaa sen tavallisesta kehittämisestä. Tavallisessa kehittämisessä tavoitteena on luoda onnistunut kehittämistuotos. Kehittämistutkimuksessa tämän lisäksi tavoitellaan käytettävien ja yleistettävien teorioiden muodostamista. [4, s. 112]

Jokainen edellä mainituista osa-alueista tuottaa erilaisia teorioita. Edelson nimeää em. osa-alueiden tuottamat teoriat (järjestyksessä) *luokkateorioiksi* (domain theories), *kehittämisen metodologioiksi* (design methodologies) sekä *kehittämisen viitekeh्यiksi* (design frameworks) (suomennokset: Rikkilä). [4, s. 112–116]

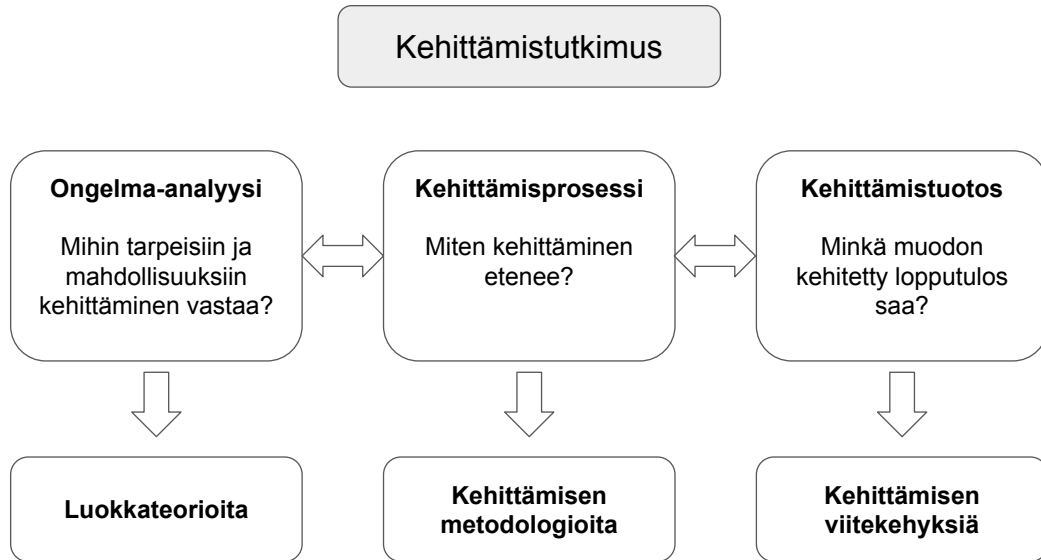
Luokkateoriat ovat ongelma-analyysin osien yleistyksiä. Luokkateoriat eivät rajoitu vain kehittämiseen itseensä, vaan ovat yleisiä teorioita. Ne ovat kuvailevia, eivät ohjailevia teorioita. Kehittämistutkimuksesta voi nousta kahdenlaisia luokkateorioita: kontekstisidonnaisia teorioita sekä lopputulosta kuvaavia teorioita. Kontekstisidonnainen teoria kartoittaa kehittämiskontekstiin liittyviä haasteita ja mahdollisuuksia. Lopputulosta kuvaava teoria taas luonnehtii joukon lopputuloksia, jotka liittyvät johonkin interventioon. Edelsonin mukaan ongelma-analyysissä tulee joko suoraan tai epäsuoraan nimetä halutut kehittämisen lopputulokset sekä keinot niiden saavuttamiseksi. [4, s. 113]

Kehittämisen metodologiat ovat yleisiä kehittämisen toimintamalleja. Ne kuvailevat prosesseja, tarvittavia asiantuntijuuksia sekä näitä asiantuntijuuksia edustavien yksilöiden rooleja. Kehittämisen metodologiat ovat ohjailevia teorioita. [4, s. 115]

Kehittämisen viitekehыkset taas ovat yleistettyjä kehittämiseratkaisuja. Ne kuvailevat, mitä ominaisuuksia kehitetyllä tuotoksella täytyy olla, jotta sillä voitaisiin saavuttaa tietyt tavoitteet tietyssä kontekstissa. Kehittämisen viitekehыkset ovat myös ohjailevia teorioita. [4, s. 114]

Edellä kuvailtu kehittämistutkimuksen rakenne Edelsonin mukaan on tiivistetty kuvioon 1. Kuvioista ilmenee, että ongelma-analyysi, kehittämiserakke ja kehittämiserakke ovat vuorovaikutuksessa toistensa kanssa. Lisäksi kuvio havainnollistaa eri

teorioita, joita osa-alueet mahdollisesti tuottavat.



Kuvio 1. Kehittämistutkimuksen rakenne ja mahdolliset tuotettavat teoriat Edelsonin [4] mukaan. Ongelma-analyysi voi tuottaa luokkateorioita, kehittämisprosessi kehittämisen metodologioita ja kehittämistuotos kehittämisen viitekehyksiä. Kuvio on muokattu versio Pernaan [3, s. 16] alkuperäisestä kaaviosta.

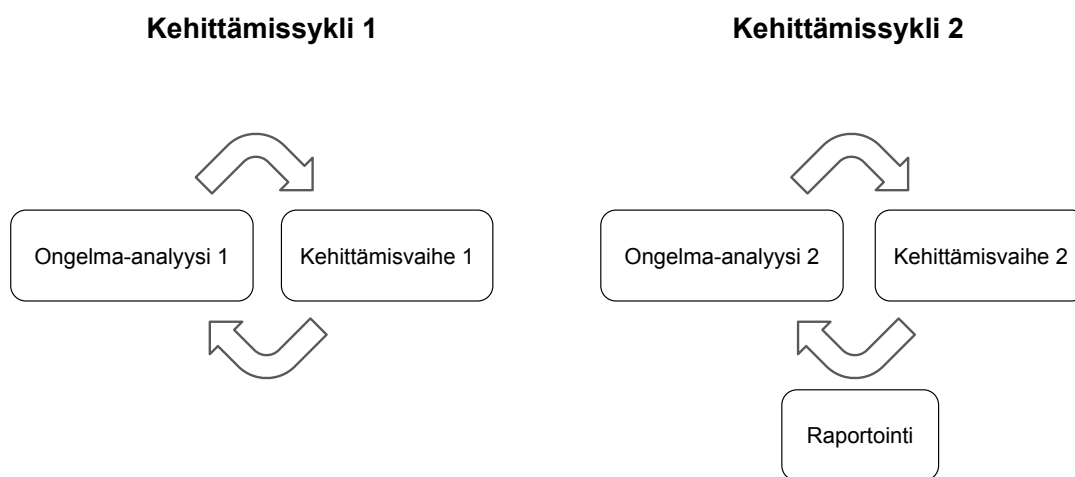
Hyvä kehittämistutkimus täyttää Design-Based Research Collectiven [5, s. 5] mukaan seuraavat viisi tunnusmerkkiä:

- Oppimisympäristöjen suunnittelun ja oppimisen teorioiden keskeiset tavoitteet on kiedottu yhteen.
- Kehitys ja tutkimus toteutetaan jatkuvina sykleinä. Syklit sisältävät suunnittelua, soveltamista, analysointia ja uudelleensuunnittelua.
- Tutkimus johtaa jaettavaan teorioihin.
- Tutkimus selittää, miten suunnitelmat toimivat aidoissa ympäristöissä.
- Selitys tukeutuu menetelmiin, joilla voidaan dokumentoida ja yhdistää suunnitteluprosessit lopputulokseen.

Syklisyys määrittelee käytännön rakenteen kehittämistutkimukselle. Edelsonin [4] mukaan yksittäinen kehittämissykli sisältää suunnittelua ja soveltamista. Jokainen soveltamisvaihe tarjoaa mahdollisuuden kerätä tietoa tulevia suunnitteluvaiheita varten. Näitä syklejä toistamalla muodostuu syklinen prosessi kehittämistutkimukselle. [4, s. 106]

Akselan ja Pernaan [11, s. 185–186] mukaan pro gradu -tutkielmassa toteutetaan yleensä yksi tai kaksi kehittämissykliä. Kuitenkin yhden syklin tutkielmista on mahdollista saada Helsingin yliopiston kemian opettajankoulutusyksikössä korkeintaan arvosana hyvä, cum laude approbatur, sillä kehittämistutkimuksessa syklisyys voidaan mieltää yhdeksi tutkimuksen laadun merkiksi [5, s. 5].

Akselan ja Pernaan ehdottama, kahden kehittämissyklin pro gradu -tutkielman rakennetta mukaileva rakenne tälle tutkimukselle on esitetty kuviossa 2 [11, s. 186].



Kuvio 2. Tutkimuksen kehittämissykli. Tutkimus alkaa ensimmäisellä ongelma-analyysillä, josta siirrytään ensimmäiseen kehittämisvaiheeseen. Toisessa syklissä tutkimus jatkuu toisella ongelma-analyysillä, joka johtaa toiseen kehittämisvaiheeseen. Nuolet syklien vaiheiden välillä kuvaavat sitä, että vaiheet ovat vuorovaikutuksessa keskenään.

Collinsin, Josephin ja Bielaczycin [12] mukaan kehittämistutkimuksen raportointi vaatii tavanomaisesta kokeellisesta tutkimuksesta poikkeavan rakenteen. He ehdottavat, että raportin tulisi sisältää seuraavat viisi osa-aluetta [12, s. 38–39]:

- kehittämisen tavoitteet ja osatekijät
- toteuttamisessa käytetyt valinnat
- kuvailut jokaisesta vaiheesta
- löydetty tulokset
- yhteenveto onnistumisista, rajoitteista ja epäonnistumisista

Aksela ja Perna suosittelivat kokemustensa pohjalta, että kehittämistutkimus kannattaa raportoida kronologisen etenemisen mukaan. Esimerkkinä he antavat seuraa-

van rakenteen: [11, s. 190–193]

1. Johdanto
2. Kehittämistutkimus
3. Teoreettinen ongelma-analyysi
4. Kehittämispöytäkirja
5. Kehittämistuotos
6. Jatkokehittäminen
7. Johtopäätökset ja pohdinta

Liitteet

Tämä tutkimus mukailee yllä esitettyä Akselan ja Pernaan esimerkkirakennetta. Johdannon ja tutkimusmenetelmän jälkeen kuvaillaan kronologisessa järjestyksessä kehittämissuorituksen eri vaiheet. Lopuksi esitellään kehittämissuoritus, kerrotaan jatkokehittämisen mahdollisuudesta sekä esitetään tutkimuksen johtopäätöksiä ja pohdintaa.

Seuraavassa alaluvussa kuvaillaan tarkemmin, millä tavalla kehittämissuoritusta tehtiin tässä tutkimuksessa.

2.2 Tutkimuksen toteutus ja tutkimuskysymykset

Idea tälle kehittämissuoritukselle alkoi muodostua Jyväskylässä 25.11.2016 pöytäkirjassa *Particle Physics Day* -tapahtumassa. Particle Physics Day on vuosittainen tapahtuma, jossa Suomen hiukkasfysiikot kokoontuvat erilaisten esitelmien sekä Suomen fyysikkoseuran hiukkasfysiikan jaoston tapaamisen merkeissä. Osallistuin päivän ohjelmaan pitämällä lyhyen esitelmän Euroopan hiukkasfysiikan tutkimuskeskus CERNin CMS-kokeen avoimesta datasta ja sen opetuskäytöstä. Olin perehtynyt hiukkasfysiikan avoimen datan opetuskäyttöön kesällä 2016 CERNissä tekemässäni kesäharjoittelussa. Particle Physics Dayn esitelmässä mainitsin, että yliopistotasolla hiukkasfysiikan avointa dataa voitaisiin käyttää esimerkiksi laboratoriotöiden yhteydessä. [13]

Päivän ohjelman jälkeen Jyväskylän yliopiston fysiikan laitoksen yliopistontutkija Sami Räsänen kertoi innostuneensa laboratoriotyöstä, jossa käytettäisiin oikeaa CERNin tutkimusdataa. Muutamaa kuukautta myöhemmin 31.1.2017 hän ehdotti sähköpostilla, että fysiikan laitoksen opetussuunnitelman vaihtumisen yhteydessä uudistuvalla modernin fysiikan kurssille voisi laatia tällaisen laboratoriotyön. Tapasimme aiheeseen liittyen yhdessä yliopistonlehtorien Sakari Juutisen ja Jussi Maunukselan kanssa. Tutkimuksen tekohetkellä Juutinen vastasi fysiikan laitoksen laboratoriotöistä ja Maunuksela opinto-ohjauksesta.

Tapaamisessa sovimme, että tulen laatimaan kurssille avointa hiukkasfysiikan tutkimusdataa hyödyntävän laboratoriotyön. Lisäksi heräsi ajatus toteuttaa pro gradu -tutkielma aiheesta. Kiinnostuin ideasta ja ehdotin aihetta Fysiikan tutkimuslaitoksen (HIP) puolen ohjaajalleni Kati Lassila-Perinille sekä Jyväskylän yliopiston fysiikan laitoksen ohjaajalleni Jan Sarénille. Kummatkin antoivat hyväksyntänsä aiheelle.

Tässä kehittämistutkimuksessa kehitän CERNin CMS-kokeen avointa tutkimusdataa hyödyntävän laboratoriotyön Jyväskylän yliopiston fysiikan laitoksen uudistuneelle modernin fysiikan kurssille. CERN, Euroopan hiukkasfysiikan tutkimuskeskus, on julkistanut keräämäänsä hiukkasfysiikan tutkimusdataa vapaaseen yleiseen käyttöön [1]. Dataa voidaan käyttää esimerkiksi eritasoisten hiukkasfysiikan data-analyysien toteuttamiseen. Lisäksi data soveltuu hiukkasfysiikan opetukseen [2].

Olen testannut kehitettyä työtä luonnontieteiden yliopisto-opiskelijoilla. Testausten perusteella arvioin, miten työn ensimmäinen versio täytti Nivalaisen, Asikaisen ja Hirvosen [6] määrittelemät fysiikan käytännön töiden tavoitekategoriat. Arvioinnin jälkeen olen kehittänyt työtä edelleen.

Kehittäminen on johtanut tietokoneella TIM-alustalla tehtävään laboratoriotyöhön, joka on tutkimuksen kehittämistuotos. Työssä tavoitteena on tehdä yksinkertainen data-analyysi aidolla, CMS-ilmaisimen keräämällä tutkimusdatalla. Analyysin perusteella tehtävänä on todeta Z^0 -bosonin esiintyminen sekä selvittää Z^0 :n massa. Työ on otettu käyttöön syksyllä 2018 ensimmäistä kertaa järjestetyllä Jyväskylän yliopiston fysiikan laitoksen uudistetulla modernin fysiikan kurssilla.

Tutkimuksessa selvitän myös, mitä haasteita ja valintoja kohdataan, kun kehitetään hiukkasfysiikan avointa dataa hyödyntävää laboratoriotyötä. Lisäksi määritellen kehittämisen viitekehysten tuleville hiukkasfysiikan avoimen datan opetusintentioneille. Viitekehyksessä esitän suosituksia onnistuneen, hiukkasfysiikan avointa tutkimusdataa hyödyntävän laboratoriotyöhön kehittämiseen.

Listattuna tutkimuksen tutkimuskysymykset ovat seuraavat:

1. Mitä haasteita ja valintoja kohdataan CERNin avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämisessä?
2. Miten kehitetyn työn testattu versio täytti Nivalaisen ym. [6] fysiikan kokeellisten töiden tavoitekategoriat?
3. Minkälaisia asioita on hyvä ottaa huomioon, kun halutaan kehittää onnistunut hiukkasfysiikan avointa tutkimusdataa hyödyntävä laboratoriotyö?

Kehittämistutkimukset noudattavat syklistä rakennetta. Tämän tutkimuksen kehittämissykliä ovat kuvattuna kuviossa 2. Seuraavaksi tarkastellaan tarkemmin, mitä osia syklien eri vaiheet sisältävät.

Ensimmäinen ongelma-analyysi sisältää tässä tutkimuksessa tutustumisen hiukkasfysiikan teoriaan, avoimen datan määritelmään sekä käytettävään sähköiseen TIM-

alustaan. Lisäksi ensimmäisessä ongelma-analyysissä tarkastellaan hiukkasfysiikan opetusta Jyväskylän yliopistossa sekä aikaisempia sovelluksia, joissa hiukkasfysiikan tutkimusdataa on hyödynnetty opetuksessa. Ongelma-analyysin lopuksi tutustutaan Nivalaisen ym. [6] määrittelemiin fysiikan käytännön töiden tavoitekategorioihin. Ensimmäinen ongelma-analyysi on kuvailtu luvussa 3.

Ensimmäinen kehittämisvaihe koostuu laboratoriotyön konkreettisesta luomisesta ja kehittämisestä TIM-alustalla. Loin ensimmäisessä kehittämisvaiheessa työn omiin kokemuksiini, asiantuntijoiden apuun sekä teoreettiseen ongelma-analyysiin perustuen. Luvussa 4 kuvaillaan kehittämisprosessin kulku sekä perustelut erilaisille prosessin aikana tehdyille valinnoille.

Toinen kehittämissykli alkaa toisella ongelma-analyysillä, joka on kuvailtu luvussa 5. Siinä olen testannut ensimmäisessä kehittämisvaiheessa tuotettua kehittämistuotosta kahdessa eri osassa. Ensimmäisessä osassa testaajat tekivät kehitetyn laboratoriotyön pareittain kommentoiden ääneen tekemisiään. Tämän jälkeen haastattelin testaajia pareittain. Ensimmäisessä vaiheessa testaajina toimivat kuusi matemaattis-luonnontieteellisten alojen yliopisto-opiskelijaa, jotka olivat testausten aikana CERNissä kesäharjoittelussa.

Ennen toista testausta tein työhön pienimuotoisen kehittämisen ensimmäisten testaajien havainnointien perusteella. Tässä vaiheessa kehitin vain pieniä osia työstä enkä esimerkiksi ottanut ensimmäisten testaajien haastatteluja huomioon. Toisessa testauksen osassa työtä testasi Jyväskylän yliopiston fysiikan laitoksen opiskelijoista koostuva testiryhmä. Testiryhmä osallistui ennen testaamista työn aihetta käsittelevälle luennolle. Tällä pyrin autenttiseen tilanteeseen, jossa työ toteutettaisiin osana kurssia. Luennon jälkeen testaajat tekivät työn samalla tapaa kuin työ tullaan toteuttamaan kurssillakin. Havainnoin testitilannetta tilaisuuden aikana kirjatulla huomioilla sekä ruutukaappausvideoilla. Työn tekemisen jälkeen testaajat täyttivät kyselylomakkeen, jossa kysyttiin työn tekemiseen ja toimivuuteen liittyviä asioita. Lisäksi haastattelin kutakin testaajaa testaamiseen liittyen. Haastatteluilla keräsin tietoa seuraavaa kehittämisvaihetta varten.

Aineiston keräämisen jälkeen analysoin aineistoa sisällönanalyysin ja sisällön erittelyn keinoin. Analyysin tuloksilla pyrin selvittämään, miten työn ensimmäinen versio täytti Nivalaisen ym. [6] fysiikan töiden tavoitekategoriat. Lisäksi tavoittelin tietoa työssä kohdatuista haasteista, onnistuneista asioista sekä kehittämistä vaativista asioista.

Toisen ongelma-analyysin jälkeen seuraa toinen kehittämisvaihe, jossa kehitin työtä edelleen testauksessa ilmenneiden huomioiden ja aineiston analyysin pohjalta. Toinen kehittämisvaihe on kuvailtu luvussa 6. Tämän kehittämisen jälkeen työ otettiin käyttöön Jyväskylän yliopiston fysiikan laitokselle. Lisäksi käännsin työn englanniksi vapaaseen yleiseen käyttöön ja raportoin tutkimuksesta. Tämä dokumentti toimii tutkimuksen raporttina.

2.3 Kehittämistutkimuksen luotettavuus ja haasteet

Edelsonin [4] mukaan yksi yleensä mainittu kehittämistutkimukseen liittyvä huoli on sen tulosten suhteellinen luotettavuus perinteiseen kokeelliseen tutkimukseen verrattuna. Kehittämistutkimus ei johda tuloksiin, joilla on tilastollisesti määritetty luotettavuus. [4, s. 117]

Toisaalta Edelsonin mukaan kehittämistutkimuksen tavoite eroaa perinteisestä kokeellisesta tutkimuksesta, joten niitä ei tulisi arvioida samoilla normeilla. Kehittämistutkimuksen tavoitteena on tuottaa uusia, käytännöllisiä teorioita, joten sen arvioinnin mittareina ovat uutuus ja käytettävyys. Lisäksi kehittämistutkimuksen vahvuus on eri kuin perinteisen kokeellisen tutkimuksen. Perinteisen kokeellisen tutkimuksen vahvuus on tilastollisessa otannassa, kun taas kehittämistutkimuksen kautta kehitettyjen teorioiden vahvuutena on niiden selittävä voima sekä tutkimuksen mukaisiin erityisiin kokemuksiin perustuminen. [4, s. 117–118]

Yleensä tutkimusmenetelmien luotettavuutta käsitellään validiteetin ja reliabiliteetin kautta. Nämä käsitteet ovat peräisin määrällisen tutkimuksen perinteistä. Koska kehittämistutkimus voi sisältää sekä laadullisia että määrällisiä osioita, eivät validiteetti ja reliabiliteetti pelkästään riitä kuvailemaan kehittämistutkimuksen luotettavuutta. Pernaa [3, s. 20] mainitseekin kehittämistutkimuksen luotettavuusanalyysin haasteelliseksi. [14]

Laadullisen tutkimuksen luotettavuutta voidaan arvioida esimerkiksi Lincolnin ja Guban [15] määrittelemien neljän kriteerin kautta. Kriteerien alkuperäiset nimet ovat *credibility*, *transferability*, *dependability* ja *confirmability* [15]. Kriteerien suomenoksissa esiintyy erilaisia variaatioita [14, s. 161–163]. Esimerkiksi Pernaa [3, s. 20] käyttää termejä *uskottavuus*, *siirrettävyys*, *luotettavuus* ja *varmuus* sekä *vahvistettavuus*.

Pernaa [3, s. 20] esittää, että kehittämistutkimuksen luotettavuutta voidaan arvioida peilaamalla Design-Based Research Collectiven [5, s. 5] määrittelemiä laadukkaan kehittämistutkimuksen kriteereitä (ks. luku 2.1) edellä mainittuun Lincolnin ja Guban luokitteluun. Pernaa muodostaa seuraavan kriteeristön:

- ”Kehittämisen tulee olla kokonaisvaltaista, jolloin kehittämistuloksena saadaan sekä ohjaavia malleja ja teorioita että kuvailevia teorioita (uskottavuus ja siirrettävyys).”
- ”Kehittämisen tulee edetä sykleittäin ja sisältää jatkuvaa kehittämistä ja arviointia (uskottavuus, luotettavuus ja vahvistettavuus).”
- ”Kehittämisessä tulee pyrkiä teorioihin, jotka ovat siirrettävissä kentälle opettajien tai muiden opetusalan ammattilaisten käyttöön (siirrettävyys).”
- ”Kehittämisprosessiin tulee sisältyä testaamista autenttisissa olosuhteissa (siirrettävyys, luotettavuus ja vahvistettavuus).”

- ”Kehittämistutkimuksen kaikki syklit tulee dokumentoida tarkasti (luotettavuus ja vahvistettavuus).” [3, s. 20, suora lainaus listasta]

Kehittämistutkimuksen luotettavuutta voidaan vahvistaa hyödyntämällä triangulaatiota [3, s. 20]. Triangulaatiolla tarkoitetaan yksinkertaisten erilaisten metodien, tutkijoiden, tiedonlähteiden tai teorioiden yhdistämistä. Triangulaatiosta voidaan erottaa neljä eri päätyyppiä, jotka ovat *tutkimusaineistoon liittyvä*, *tutkijaan liittyvä*, *teoriaan liittyvä* ja *metodinen* triangulaatio. Tässä tutkimuksessa näistä käytetään metodista triangulaatiota, jolla tarkoitetaan useiden eri metodien, tutkimusstrategioiden tai tiedonhankintamenetelmien käyttöä. Triangulaatiota hyödynnettiin keräämällä aineistoa sekä haastatteluilla että kyselylomakkeella. [14, s. 168]

3 Teoreettinen ongelma-analyysi

Tässä luvussa tarkastellaan tutkimuksen ensimmäistä vaihetta, teoreettista ongelma-analyysia. Ongelma-analyysissa määritellään tavoitteet, tarpeet ja mahdollisuudet, joihin kehittämällä pyritään vastaamaan. Lisäksi nimetään kehittämisen yhteyteen liittyviä haasteita ja rajoitteita.

Luvun alussa perehdytään hiukkasfysiikan teoriaan, kokeelliseen hiukkasfysiikkaan CERNissä sekä avoimen datan käsitteeseen. Taustatietoina tarkastellaan TIM-alustaa, hiukkasfysiikan opetusta Jyväskylän yliopistossa sekä aiempia sovelluksia ja oppimateriaaleja, joissa on käytetty CERNin avointa tutkimusdataa opetustarkoituksessa. Lisäksi esitellään Nivalaisen ym. [6] määrittelemät viisi fysiikan kokeellisten töiden tavoitekategoriaa.

3.1 Hiukkasfysiikka

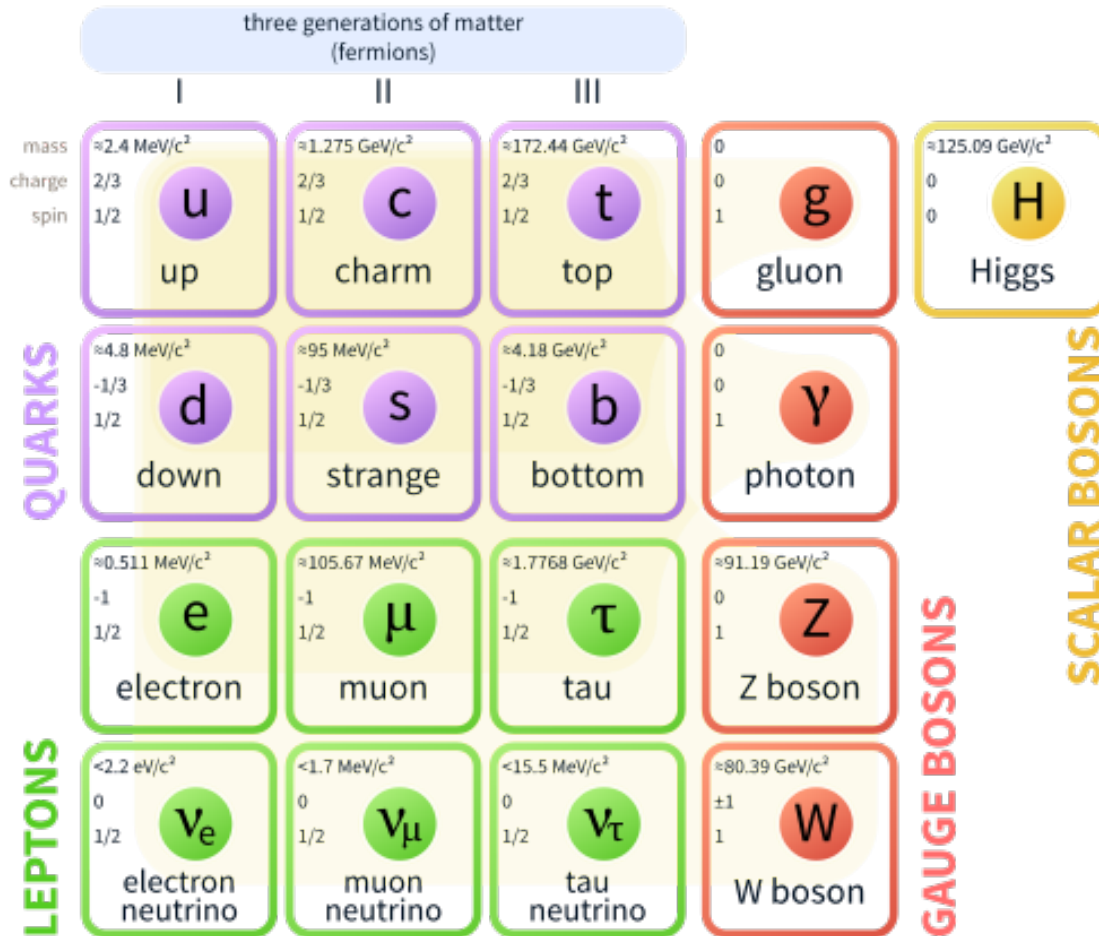
Hiukkasfysiikka on yksi fysiikan tutkimuksen osa-alueista. Martin ja Shaw [16, s. 1] määrittelevät teoksensa *Particle Physics* alussa hiukkasfysiikan seuraavasti: ”Particle Physics is the study of the fundamental constituents of matter and the forces between them.” Hiukkasfysiikassa siis tutkitaan aineen perusrakennneosasia ja niiden välisiä voimia.

Griffiths [17] mainitsee teoksessaan *Introduction to Elementary Particles* hiukkasfysiikasta seuraavasti: ”Elementary particle physics addresses the question, ’What is matter made of?’ at the most fundamental level – which is to say, on the smallest scale of size.” Griffithsin mukaan hiukkasfysiikka siis käsittelee pienimmässä mahdollisessa kokoluokassa sitä, mistä materia on tehty.

Yleisesti perusrakenneosasia kutsutaan alkeishiukkasiksi, eli hiukkasiksi joilla ei ole sisäistä rakennetta tai viritystiloja [16, s. 1]. Alkeishiukkaset eivät enää koostu muista hiukkasista.

Mitä nämä alkeishiukkaset sitten ovat? Tähän ei ole yksiselitteistä vastausta, sillä kyseessä on yksi hiukkasfysiikan ratkaisemattomista kysymyksistä. Tällä hetkellä voimassa oleva hiukkasfysiikan teoria nimeltään *standardimalli* (standard model) pyrkii selittämään kaikki hiukkasfysiikan ilmiöt malliin kirjattujen alkeishiukkasten ominaisuuksien ja vuorovaikutusten avulla [16, s. 1]. Standardimalli ei kuitenkaan ole vielä täydellinen teoria, vaan yhä keskeneräinen. Hiukkasfysiikan standardimalli on esitetty kuviossa 3.

Standard Model of Elementary Particles



Kuvio 3. Hiukkasfysiikan standardimalli. Malli sisältää tunnetut alkeishiukkas-
set ja välittäjäbosonit, sekä näiden luokittelut. Kuva Wikimedia-sivustolta [18].

Standardimallin mukaan kaikki materia koostuu alkeishiukkasista, jotka esiintyvät kahdessa eri luokassa: leptoneina ja kvarkkeina. Kummassakin luokassa on kuusi eri hiukkasta. Lisäksi standardimalliin kuuluvat ns. välittäjäbosonit. Nämä mahdollistavat olemassaolon osalle luonnon perusvuorovaikutuksista. Perusvuorovaikutuksia ovat sähkömagneettinen vuorovaikutus, heikko vuorovaikutus, vahva vuorovaikutus ja gravitaatio. Standardimalli ei kuitenkaan selitä näistä gravitaatiota. Martin ja Shaw [16, s. 2] eivät edes laske gravitaatiota perusvuorovaikutukseksi, sillä sen vuorovaikutus alkeishiukkasten kanssa on niin pieni verrattuna muihin kolmeen vuorovaikutukseen. [19]

Kehitettävässä laboratoriotyössä käsitellään tarkemmin välittäjäbosoneista Z^0 -bosonia sekä alkeishiukkasista myoneja. Z^0 -bosonit (jatkossa lyhennetyllä merkinnällä Z) ovat raskaita ja epästabiileja välittäjäbosoneita. Varaukseltaan Z on

neutraali. Z -bosonin massa on n. $91,2 \text{ GeV}/c^2$ ja elinikä hyvin lyhyt, n. $2,6 \times 10^{-25} \text{ s}$. Ensimmäisen kerran Z -bosoni havaittiin CERNissä vuonna 1983 Super Proton Synchrotron –protoni-antiprotonitörmäytinillä. [16, s. 183–188]

Myonit (μ^-) ovat varattuja leptoneita. Myoneiden varauksen suuruus on $-e$, eli negatiivinen alkeisvaraus. Myonin massa on n. $105,7 \text{ MeV}/c^2$. Myonia vastaavaa antihiukkasta kutsutaan antimyoniiksi (μ^+). [16, s. 22] [20]

Hiukkasfysiikan tutkimus jakautuu teoreettiseen ja kokeelliseen tutkimukseen. Tässä tutkimuksessa perehdytään hiukkasfysiikan kokeelliseen puoleen, sillä kehitettävissä laboratoriotyössä käytetään kokeellista hiukkasfysiikan tutkimusdataa. Kokeellista hiukkasfysiikan tutkimusta tehdään tutkimuskeskuksissa eri puolilla maailmaa. Tutkimuskeskuksia ovat mm. CERN Sveitsissä, Fermilab USA:ssa, DESY Saksassa sekä KEK Japanissa. Kehitettävässä laboratoriotyössä käytetään CERNin CMS-ilmaisimen keräämää tutkimusdataa.

3.2 Kokeellinen hiukkasfysiikka CERNissä

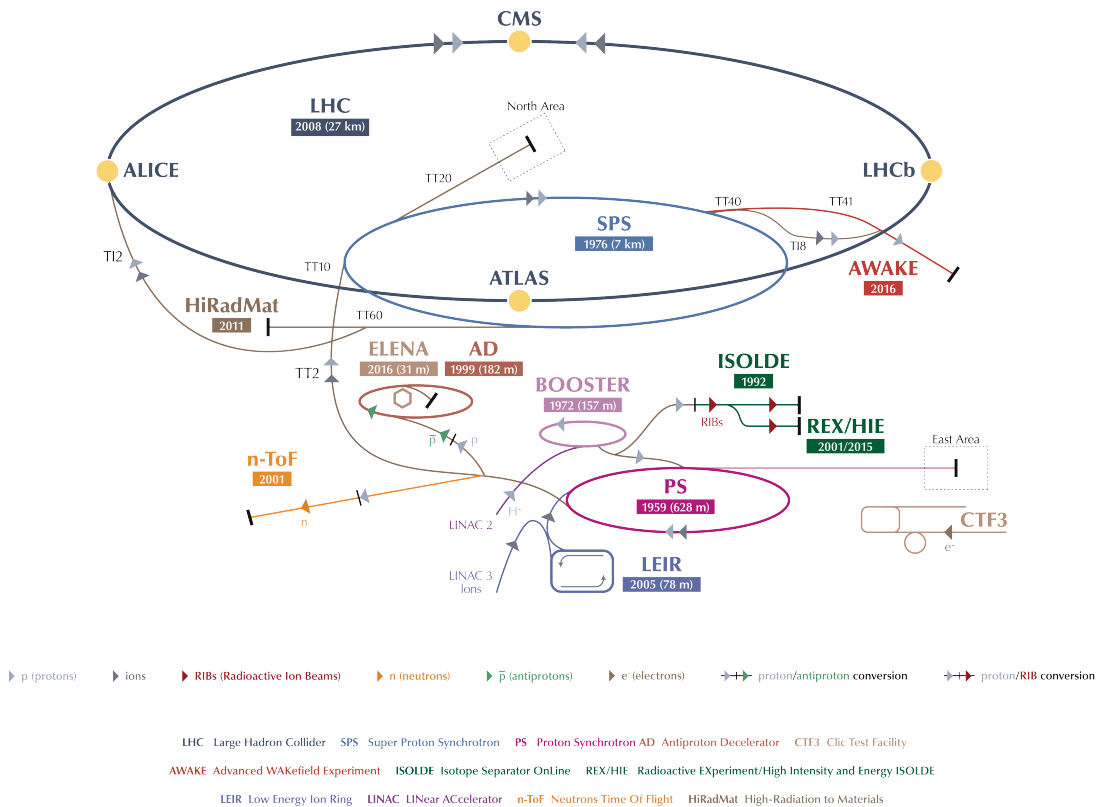
3.2.1 CERN

CERN, ranskaksi *Conseil Européen pour la Recherche Nucléaire* ja vapaasti suomennettuna *Euroopan hiukkasfysiikan tutkimuskeskus*, on hiukkasfysiikan tutkimukseen keskittynyt tutkimuskeskus Meyrinissä, Sveitsissä. Keskus sijaitsee Sveitsin ja Ranskan rajalla levittyen kummankin valtion puolelle. CERNissä tutkimusta tehdään hiukkaskiihdyttimien ja hiukkasilmaisimien avulla. Hiukkaskiihdyttimien avulla hiukkassuihkuja kiihdytetään ja törmäytetään toisiinsa tai kiinteisiin kohteisiin. Törmäyksiä havainnoidaan ja mittaustuloksia tallennetaan hiukkasilmaisimilla. [21]

CERN on perustettu vuonna 1954. Vuonna 2015 CERNissä työskenteli henkilökuntatunnuksella 2531 henkilöä ja muilla tunnuksilla yhteensä 13794 henkilöä [22]. Suomi kuuluu CERNin jäsenmaihin, joita on raportin kirjoitushetkellä 22 kappaletta [21]. Suomen yhteistyöstä CERNin kanssa huolehtii Helsingin yliopiston, Aalto-yliopiston, Jyväskylän yliopiston, Lappeenrannan teknillisen yliopiston ja Tampereen teknillisen yliopiston yhteinen tutkimuslaitos, Fysiikan tutkimuslaitos (Helsinki Institute of Physics, HIP) [23].

Hiukkaskiihdyttimet CERNissä muodostavat kiihdytinkompleksin, jossa hiukkasia saadaan kiihdytettyä yhä suuremmille energioille. Kiihdytettävää hiukkasjoukkoa kutsutaan hiukkassuihkuksi. CERNin kiihdytinkompleksi on esitetty kuviossa 4.

Kiihdytettäessä protoneja protonilähteenä toimii vetyä sisältävä kaasupullo. Sähkökentän avulla elektronit irroitetaan vetyatomien kuorilta, jolloin saadaan tuotettua protoneja. Näiden protonien kiihdyttäminen alkaa Linac 2 -linearikiihdyttimellä. Linac 2:n jälkeen protonisuihku ohjataan Proton Synchrotron Booster -kiihdyttimeen (PSB), josta suihku jatkaa edelleen Proton Synchrotron -kiihdyttimeen (PS). Proton



Kuvio 4. CERNin kiihdytinkompleksi. Ennen suurinta LHC-kiihdytintä protoneja kiihdytetään useilla esikiihdyttimillä. Kuva CERN Document Server -sivustolta [24].

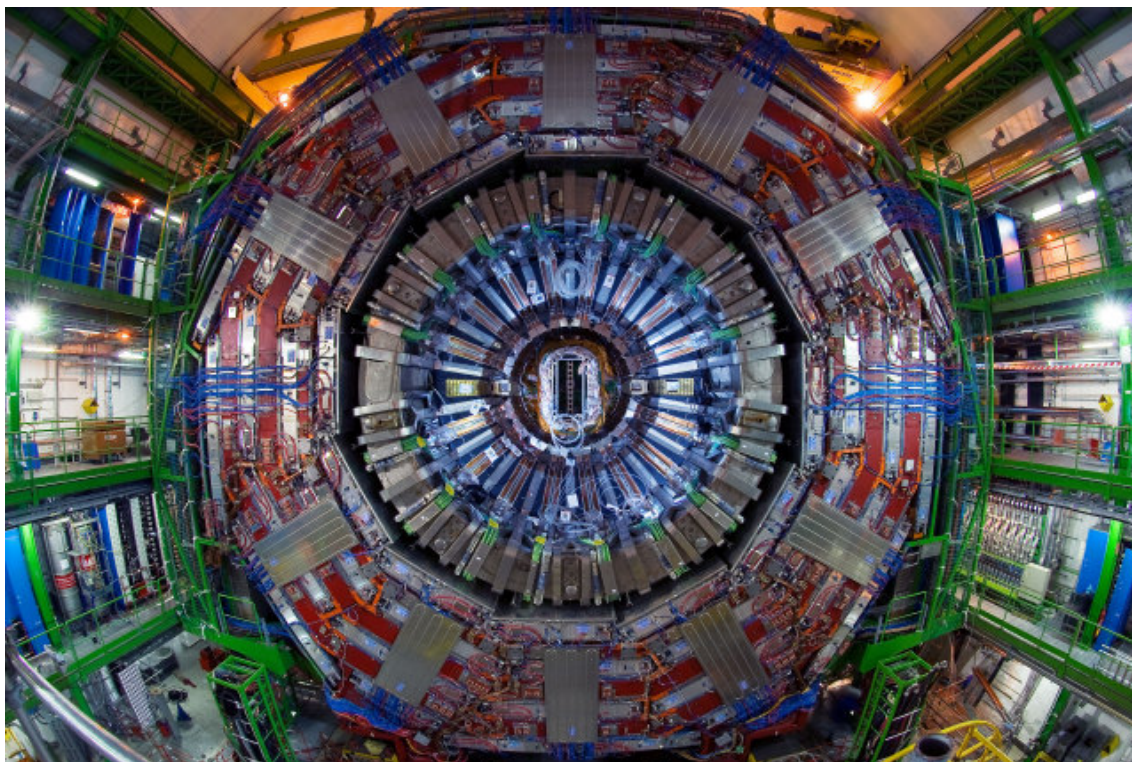
Synchrotonia seuraa vielä Super Proton Synchrotron -kiihdytin (SPS), joka kiihdyttää hiukkaset 450 GeV:n energiaan. [21]

Tämän prosessin jälkeen muodostetaan kaksi protonisuihkua, jotka ohjataan CERNin suurimmalle kiihdyttimelle, Large Hadron Colliderille (LHC). LHC:llä suihkut kulkevat erillisissä tyhjiöputkissa, toinen myötäpäivään ja toinen vastapäivään. LHC:llä protonit saadaan kiihdytettyä n. 6,5 TeV:n maksimienergialle. [21]

Hiukkastörmäyksiä luodaan risteyttämällä LHC:n kaksi hiukkassuihkua. Suihkujen risteytyessä osa suihkuissa olevista hiukkasista törmää. Törmäyksissä kokonaisenergia on 13 TeV:n luokkaa. Tästä energiasta syntyy uusia hiukkasia, joita pyritään havainnoimaan hiukkasilmäyksillä. LHC:llä suihkuja törmäytetään yhteensä neljässä kohdassa kiihdytintä. Näillä kohdilla sijaitsevat LHC:n neljä hiukkasilmäysinta: ALICE, ATLAS, CMS ja LHCb (ks. kuvio 4). [21] Tässä tutkimuksessa keskitytään edellä mainituista CMS-ilmäysinten keräämään hiukkasfysiikan tutkimusdataan.

3.2.2 CMS-ilmaisim

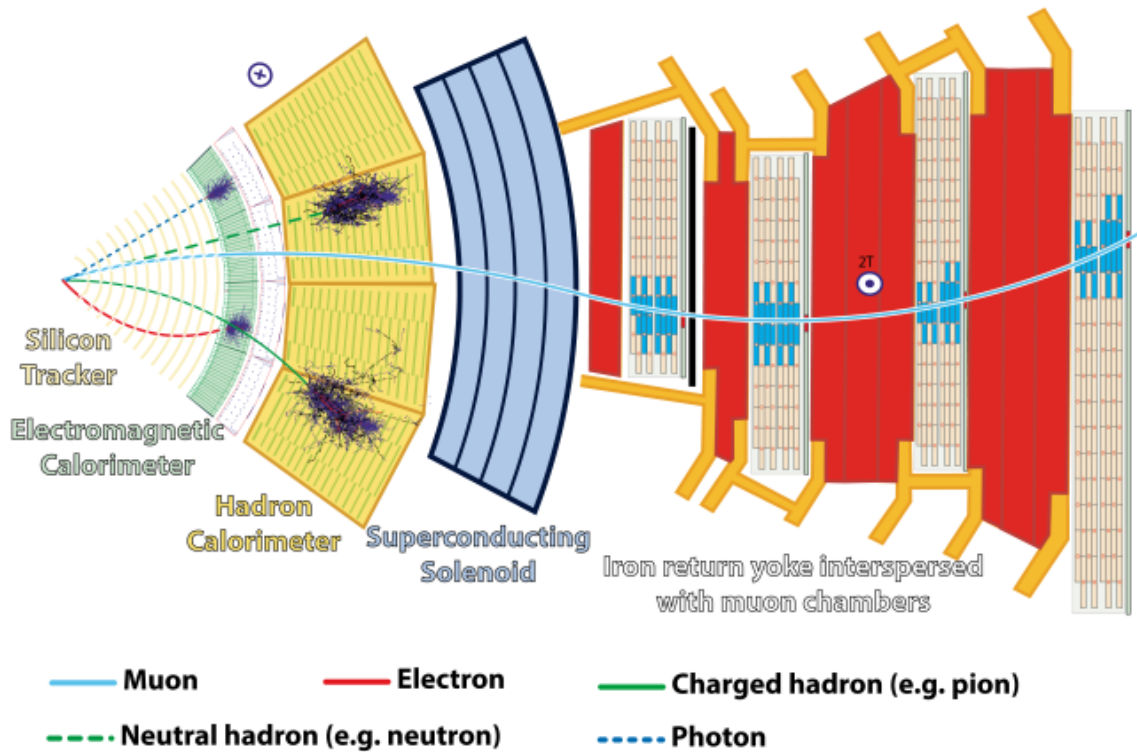
CMS, Compact Muon Solenoid, on niin sanottu yleisilmaisim. Tämä tarkoittaa sitä, että se on suunniteltu havaitsemaan kaikkia uusia fysiikan ilmiöitä, joita LHC voi tuoda ilmi [25]. CMS:n tutkimustehtävät liittyvät mm. standardimalliin, ylimääräisiin ulottuvuuksiin sekä pimeään aineeseen [26]. Valokuva CMS-ilmaisimesta on kuviossa 5



Kuvio 5. CMS-ilmaisim avattuna. Ilmaisim koostuu erilaisista ilmaisimkomponenteista. Kuva CERN Document Server -sivustolta [27].

CMS-ilmaisimesta on esitetty poikkileikkaus kuviossa 6. Ilmaisim koostuu päällekkäisistä kerroksista erilaisia ilmaisimkomponentteja, joiden avulla voidaan havaita törmäyksissä syntyneitä hiukkasia. Lisäksi olennainen osa CMS-ilmaisinta on suuri suprajohtava solenoidi, jonka luoman magneettikentän avulla kaareutetaan sähköisesti varautuneiden hiukkasten lentoratoja. Magneettikenttä kohdistaa liikkeessä oleviin varattuihin hiukkasiin voiman, joka saa hiukkaset ympyräradalle. Ympyräradan säde r saadaan määritettyä lausekkeella $r = \frac{mv}{qB}$, jossa m on hiukkasen massa, v nopeus ja q varaus, sekä B magneettikentän magneettivuon tiheys. Kaareutumisen avulla voidaan määrittää hiukkasten varauksia sekä liikemääriä. [25]

Sisimpänä CMS:ssä on pii-ilmaisim (silicon tracker), jonka avulla voidaan määrittää varattujen hiukkasten lentoratoja. Varatut hiukkaset vuorovaikuttavat pii-ilmaisimen kanssa sähkömagneettisesti ja saavat ilmaisimen luomaan sähköisen pulssin. Pulssit yhdistämällä saadaan muodostettua hiukkasten lentoratoja. [25]



Kuvio 6. CMS-ilmaisimen poikkileikkaus. Sisimpänä ovat pii-ilmaisim, sähkömagneettinen kalorimetri sekä hadronikalorimetri. Näitä ympäröivät suprajohtava solenoidi sekä myonikammiot. Kuva CERN Document Server -sivustolta [28].

Pii-ilmaisimen jälkeen tulevat CMS:n kaksi kalorimetriä, sähkömagneettinen kalorimetri (ECAL) ja hadronikalorimetri (HCAL). Näiden kalorimetrien avulla voidaan määrittää hiukkasten energioita. Sisempänä olevan ECAL-kalorimetri voi määrittää elektronien ja fotonien energioita ja ulompi HCAL taas hadronien (esim. protonit, neutronit ja pionit) energioita. [25]

Uloimpana CMS-ilmaisimessa ovat suuret myonikammiot, joilla voidaan määrittää myoneiden lentoratoja ja liikemääriä. Myonit vuorovaikuttavat heikosti muun materian kanssa, joten ne jatkavat matkaansa myös myonikammioiden läpi ulos ilmaisimesta. [25]

Miten data sitten etenee edellä mainituilta ilmaisinkomponenteilta eteenpäin? LHC-kiihdyttimessä hiukkassuihkut risteävät noin 40 MHz:n taajuudella, eli 40 miljoonaa kertaa sekunnissa. Jokaisessa suihkujen risteämisessä noin 20 protonia törmäävät. Yhteensä CMS-ilmaisimessa tapahtuu siis noin 800 miljoonaa törmäystä sekunnissa. [29, s. 3]

Kaikkien näiden törmäysten tuottaman datan tallentaminen olisi teknisesti mahdotonta, sillä dataa syntyy todella paljon todella lyhyessä ajassa. Tätä varten CMS-ilmaisimessa on ns. liipaisusysteemi (triggering system), jolla karsitaan pois ei-

kiinnostavat törmäystapahtumat. CMS:n liipaisusysteemi koostuu kahdesta eri osasta, *Level-1 Trigger* -liipaisimesta sekä *High-Level Trigger* -liipaisimesta. Näistä Level-1 Trigger prosessoi välittömästi sekunnin aikana tapahtuvista 800 miljoonasta törmäyksestä n. 100 000 mahdollisesti kiinnostavaa törmäystä. High-Level Trigger taas karsii näistä ainoastaan n. 100 törmäystä tallennettavaksi. Eli CMS tallentaa noin 100 törmäystapahtuman tiedot sekunnissa jatkokäsittelyä varten. [29, s. 3]

3.2.3 Hiukkasten epäsuora havaitseminen

Kaikkia hiukkasia ei voida havaita suoraan CMS:n ilmaisinkomponenteilla. Esimerkiksi tässä tutkimuksessa kehitettävässä työssä käsitellään Z -bosonia, jota ei voida havaita suoraan CMS-ilmaisimella. Tämä johtuu yksinkertaisesti siitä, että lyhyen elinikensä ($2,6 \times 10^{-25}$ s, ks. luku 3.1) takia Z hajoaa välittömästi törmäystapahtuman jälkeen ehtimättä edes ensimmäiselle ilmaisinkomponentille.

Epäsuorassa havaitsemisessa Z :n olemassaolo päätellään sen hajoamistuotteiden avulla. Z :lla on useita kymmeniä hajoamistapoja, mutta tässä tutkimuksessa keskitytään ainoastaan Z :n hajoamiseen myoniksi (μ^-) ja antimyoniksi (μ^+) [20] [16, s. 185]. Tästä edespäin käytetään myonista ja antimyonista myös nimitystä myonipari.

Kehitettävässä laboratoriotyössä käytettävään dataan on valittu törmäystapahtumia, joissa on kussakin havaittu myoni ja antimyoni. Nämä myoniparit ovat mahdollisesti peräisin Z :n hajoamisesta, mutta voivat olla myös täysin muita myoneita tai antimyoneita joistakin muista törmäyksen prosesseista. Tarvitaan siis jokin kriteeri, jolla voidaan päätellä onko myonipari peräisin Z :sta.

Päätelyssä käytetään apuna invariantin massan käsitettä. Invariantti massa, M , N :lle hiukkaselle määritellään lausekkeella

$$Mc^2 = \sqrt{\left(\sum_{k=1}^N E_k\right)^2 - \left(c \sum_{k=1}^N \vec{p}_k\right)^2}, \quad (1)$$

jossa c on valon nopeus tyhjiössä, E_k yksittäisen hiukkasen energia ja \vec{p}_k yksittäisen hiukkasen liikemäärävektori. Nimensä mukaisesti invariantin massan arvo on sama kaikissa referenssikoordinaatistoissa. [16, s. 283]

Keskitytään nyt tilanteeseen, jossa yksi M -massainen ja E -energinen hiukkanen hajoaa kahdeksi m_1 ja m_2 -massaiseksi sekä E_1 ja E_2 -energiseksi hiukkaseksi. Hajoisessa energia ja liikemäärä säilyvät, joten $E = E_1 + E_2$ ja $\vec{p} = \vec{p}_1 + \vec{p}_2$.

Energian ja liikemäärän säilymisellä lausekkeesta (1) saadaan siis

$$\begin{aligned} Mc^2 &= \sqrt{(E_1 + E_2)^2 - c^2(\vec{p}_1 + \vec{p}_2)^2} \\ &= \sqrt{E_1^2 + 2E_1E_2 + E_2^2 - c^2\vec{p}_1^2 - 2c^2\vec{p}_1 \cdot \vec{p}_2 - c^2\vec{p}_2^2} \\ &= \sqrt{2E_1E_2 - 2c^2|\vec{p}_1||\vec{p}_2|\cos(\beta) + m_1^2c^4 + m_2^2c^4}. \end{aligned} \quad (2)$$

Relativistinen dispersiorelaatio voidaan saattaa seuraavaan muotoon [16, s. 3]:

$$M^2 c^4 = E^2 - c^2 \vec{p}^2$$

$$E = \sqrt{c^2 \vec{p}^2 + M^2 c^4}.$$

Tästä asettamalla $c = 1$ sekä olettamalla hiukkasten massa liikemäärään verrattuna hyvin pieneksi, saadaan

$$E = \sqrt{\vec{p}^2 + M^2} = |\vec{p}| \sqrt{1 + \frac{M^2}{p^2}} \xrightarrow{M \ll |\vec{p}|} |\vec{p}|.$$

Soveltamalla edellä saatua tulosta $E = |\vec{p}|$ ja asetusta $c = 1$ lausekkeeseen (2), saadaan se sievennettyä muotoon

$$M = \sqrt{2E_1 E_2 (1 - \cos(\beta))},$$

jossa β on hiukkasten liikemäärävektorien välinen kulma. Tällä lausekkeella voidaan laskea ns. invariantti massa hiukkasparille, jos vain tiedetään energiat ja hiukkasten liikemäärien välinen kulma.

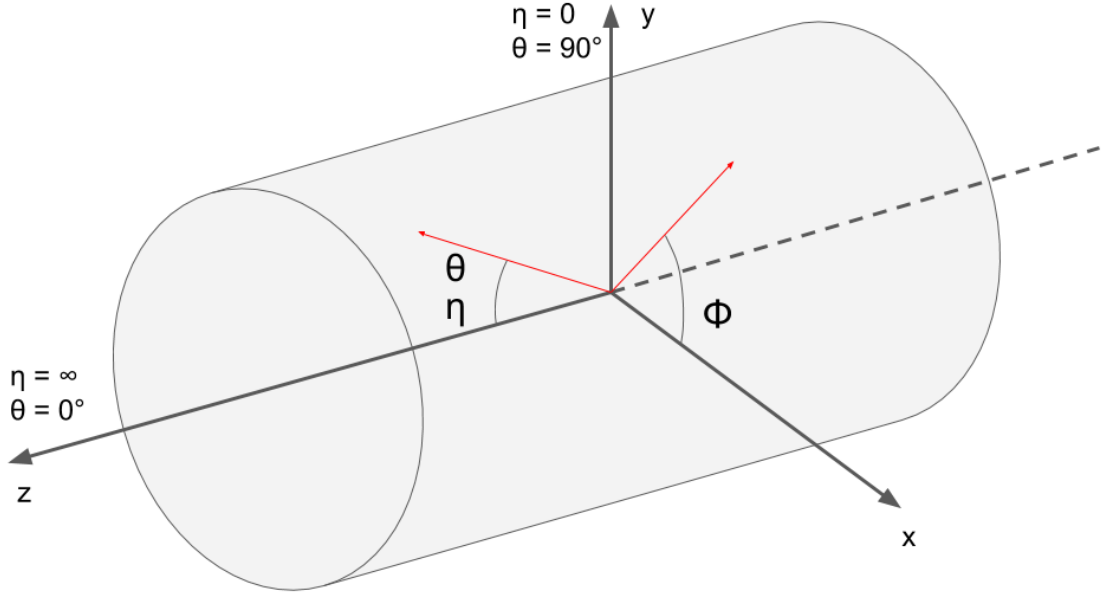
Kokeellisessa hiukkasfysiikassa invariantille massalle käytetään usein seuraavaa, ainoastaan massattomille hiukkasille pätevää lauseketta:

$$M = \sqrt{2p_{T1} p_{T2} (\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))}, \quad (3)$$

jossa p_T on hiukkasen liikemäärän hiukkasuuhkua vastaan kohtisuora komponentti, η pseudorapiditeetti ja ϕ atsimuuttikulma. Pseudorapiditeetti määritellään hiukkasen sirontakulman θ avulla lausekkeella $\eta = -\ln(\tan(\frac{\theta}{2}))$, eli se kuvaa kulmaa. Myös ϕ kuvaa kulmaa. Kuviossa 7 on esitetty θ , η ja ϕ CMS-ilmaisimessa. Hiukkasuuhku kulkee z-akselin suunnassa. Kuvasta nähdään myös, että η :n määritelmästä johtuen sen arvo on 0, kun $\theta = 90^\circ$ ja ∞ , kun $\theta = 0^\circ$.

Mikäli siis myonipari on peräisin Z -bosonin hajoamisesta, vastaa myoniparille laskettu invariantin massan arvo Z :n massaa. Jokaiselle myoniparille voidaan laskea invariantin massan arvo ja näistä arvoista voidaan luoda histogrammi. Histogrammin avulla voidaan todentaa mahdollinen Z -bosonin esiintyminen. Mikäli invarianttien massojen jakaumaan muodostuu piikki, viittaa tilanne siihen, että tapahtumissa on esiintynyt piikin paikan ilmaisemaa massaa vastaava hiukkanen. [16, s. 116–117]

Pelkkä visuaalinen havainto piikistä ei kuitenkaan ole riittävä todiste hiukkasen olemassaolosta. Jotta jakaumasta saataisiin konkreettisemmin tietoa piikkiä vastaavan hiukkasen massasta ja elinajasta, voidaan jakaumaan tehdä funktion sovitus. Piikin



Kuvio 7. CMS-ilmaisimen koordinaatteja. Hiukkassuihku kulkee z-akselin suunnassa. Symbolit θ sekä ϕ kuvaavat erilaisia kulmia, ja η pseudorapiditeettia.

muotoa voidaan approksimoida esimerkiksi Breit-Wigner -jakaumalla, jonka lauseke näyttää seuraavalta: [16, s. 117–118]

$$N(M) = \frac{K}{(M - M_r)^2 + \frac{\Gamma^2}{4}}. \quad (4)$$

Lausekkeessa K on hajoamisten lukumäärästä riippuva vakio, M invariantti massa, M_r piikin maksimikohta ja Γ jakauman puoliarvoveveys (engl. decay width). Jakauman leveys ja resonanssissa havaittavan hiukkasen elinaika τ ovat yhteydessä lausekkeella

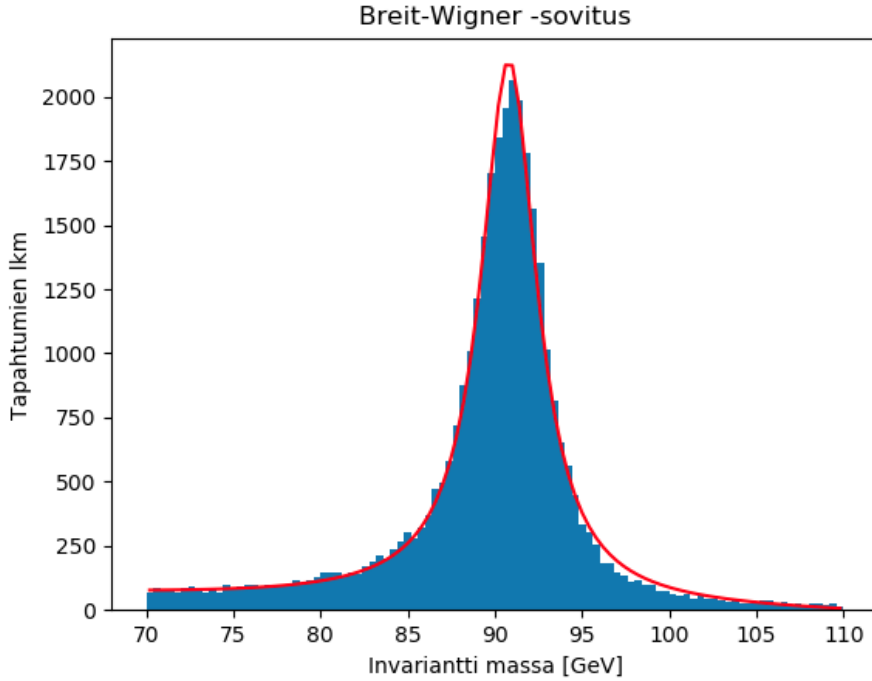
$$\Gamma \equiv \frac{\hbar}{\tau}, \quad (5)$$

jossa \hbar on redusoitu Planckin vakio, $\hbar = \frac{h}{2\pi}$ [16, s. 117].

Invarianttien massojen jakaumassa esiintyvän resonanssin massa ja puoliarvoveveys saadaan siis määritettyä sovittamalla lauseke (4) jakaumaan. Sovituksessa on otettava huomioon myös histogrammissa esiintyvä tausta. Tausta on seurausta siitä, että tapahtumissa on havaittu myös myonipareja, jotka ovat peräisin täysin satunnaisista hajoamisista, eivätkä siis jonkun tietyn hiukkasen hajoamisprosessista.

Breit-Wigner -jakauman lausekkeeseen voidaan lisätä esimerkiksi termi $aM + b + A$, jossa $aM + b$ huomioi taustan lineaarisen osan ja A korkeuden. Termissä M on lausekkeessa (4) esiintyvä invariantti massa, ja a sekä $b + A$ tilannekohtaisia parametreja.

Kuviossa 8 on esitetty esimerkki invarianttien massojen jakaumaa kuvaavaan histogrammiin tehdystä Breit-Wigner -sovituksesta, jossa histogrammin tausta on otettu huomioon edellä mainitulla tavalla.



Kuvio 8. Esimerkki invarianttien massojen jakaumaa kuvaavaan histogrammiin tehdystä Breit-Wigner -sovituksesta. Histogrammin taustan lineaarinen osa on otettu huomioon lisäämällä sovitukseen termi $aM + b$ ja taustan korkeuden vaikutus lisäämällä termi A .

Kehitettävässä laboratoriotyössä perehdytään myös tarkemmin siihen, mistä jakauman leveys muodostuu. Vaikka Z -bosonin massa saataisiin määritettyä edellä kuvatulla tavalla äärettömän pienellä laitteiston mittauserätarkkuudella, syntyisi histogrammiin silti jakauma eikä esimerkiksi yhtä täsmällistä piikkiä. Tämä selittyy aika-energia -epätarkkuusperiaatteella. Aika-energia -epätarkkuusperiaatteen nojalla voidaan todeta ajan epätarkkuuden Δt ja energian epätarkkuuden ΔE noudattavan seuraavaa relaatiota: [30]

$$\Delta t \Delta E \geq \frac{\hbar}{2}. \quad (6)$$

Hiukkasfysiikassa massa ja energia vastaavat toisiaan. Lausekkeesta (6) voidaan siis nähdä, että hiukkasten energian tai massan määrittämiseen liittyy aina epätarkkuutta. Mitä lyhyempi elinaika on, sitä epätarkempi on hiukkasen massa.

Jakauman leveyden lisäksi kehitettävän laboratoriotyön lopussa tarkastellaan pseudorapiditeetin η (ks. kuvio 7) vaikutusta invarianttien massojen histogrammin piikin

leveyteen. Kuvion 7 mukaisesti käytännössä pseudorapiditeetti kertoo, millä kulmalta törmäyksessä syntynyt hiukkanen on kulkenut. Mitä suurempi pseudorapiditeetti hiukkasella on, sitä lähempänä hiukkasuikun linjaa se etenee. Suurilla pseudorapiditeeteilla hiukkaset osuvat siis ilmaisimen päätyyn, ja vastaavasti pienillä pseudorapiditeeteilla ilmaisimen vaipalle.

Pseudorapiditeetti vaikuttaa invarianttien massojen histogrammin piikin leveyteen siten, että mikäli histogrammi tuotetaan suurempien pseudorapiditeettien omaavien myonien invarianteista massoista, saadaan leveämpi piikki kuin jos histogrammissa käytettäisiin pienempien pseudorapiditeettien omaavia myoneja. Tämä selittyy sillä, että ilmaisimen päätyyn suunnanneet myonit vuorovaikuttavat enemmän ilmaisimen materiaalin kanssa kuin vaipalle suunnanneet. Vuorovaikutuksissa myonit menettävät energiaansa. Tämä häiritsee myonien radan sovittamista ja liikemäärän mittaamista. Lisäksi tilanteeseen vaikuttaa se, että CMS-ilmaisimessa liikemäärän mittaus riippuu mm. myonikammioiden suuntauksesta, ilmaisimen materiaalin jakautumisesta sekä magneettikentästä [31]. Voidaan olettaa, että hiukkasen suuntautuminen ilmaisimen päätyyn tai vaipalle vaikuttaa edellä mainittuihin tekijöihin.

3.3 Avoin data

Tutkimuksessa kehitettävässä laboratoriotyössä käytetään CMS-kokeen avointa dataa. Tässä alaluvussa määritellään mitä avoimella datalla tarkoitetaan. Lisäksi kuvaillaan CMS-kokeen avoimen datan julkaisupolitiikkaa sekä datan muotoa.

Avoindata.fi -palvelun määritelmän mukaan avoin data on dataa, joka ”– – on kenen tahansa uudelleen käytettävissä maksutta, luvallisesti ja koneluettavassa muodossa”. Datalla taas tarkoitetaan tietoa, joka on koneellisesti käsiteltävässä muodossa. [32]

Haasio ja Savolainen [33, s. 15] määrittelevät datan hieman eri tavalla. Heidän mukaansa data on informaation raaka-ainetta. Datasta voi tulla informaatiota, mutta kaikki data ei ole informaatiota. Haasio ja Savolainen kuvaavat tätä tiedon arvoketjulla, joka on esitetty kuviossa 9. Arvoketjusta nähdään, että data on ketjun ensimmäinen osa. Datasta voidaan jalostaa informaatiota, informaatiosta tietoa, ja edelleen tiedosta tietämystä sekä viisautta. [33, s. 14]

data \Rightarrow informaatio \Rightarrow tieto \Rightarrow tietämys \Rightarrow viisaus

Kuvio 9. Tiedon arvoketju Haasion ja Savolaisen [33, s. 14] mukaan. Datasta voidaan jalostaa informaatiota, ja informaatiosta edelleen tietoa.

CMS-kokeen dataan liittyen CMS-kollaboraatio on sitoutunut säilyttämään keräämänsä dataa sekä tarjoamaan sitä avoimesti saataville tietyn ajanjakson jälkeen [1]. Data julkistetaan Creative Commons CC0 -lisensoituna². Lisenssi tarkoittaa, että

²<https://creativecommons.org/publicdomain/zero/1.0/deed.fi>

data on julkistettu vapaaseen yleiseen käyttöön luopuen kaikista tekijänoikeuslain-säädännön alaisista oikeuksista [34].

Study Group for Data Preservation and Long Term Analysis in High Energy Physics (DPHEP) on ryhmä, joka vastaa korkeaenergifysiikan datan pitkäaikaissäilytyksestä. DPHEP on määritellyt korkeaenergifysiikan avoimelle datalle neljä eri säilytysmallia, jotka on esitelty taulukossa 1. Myös CMSn avoimen datan säilytys toteutetaan näiden mallien mukaisesti. [35]

Taulukko 1. DPHEP:n säilytysmallit korkeaenergifysiikan avoimelle datalle [35]. Mallit jaetaan neljään eri tasoon datan laajuuden ja käyttötarkoituksen mukaan.

Säilytysmalli	Käyttötarkoitus
Taso 1: ylimääräisen dokumentaation tarjoaminen	julkistamiseen liittyvän tiedon etsiminen
Taso 2: datan säilyttäminen yksinkertaistetussa tallennusmuodossa	”outreach”-toiminta, yksinkertainen analyysi
Taso 3: analyysitason dataformaatin ja tietokoneohjelmiston säilyttäminen	valmiiseen rekonstruktioon perustuva tieteellinen analyysi
Taso 4: perustason datan sekä rekonstruktio- ja simulointiohjelmistojen säilyttäminen	datan koko potentiaali

CERNin Open Data -portaalissa³ julkistettu CMSn avoin data on tasojen 2 ja 3 dataa. Portaalissa on kolmenlaisia datakokoelmia: peruskokoelmia (primary datasets), simuloituja kokoelmia sekä yksinkertaistettuja kokoelmia. Lisäksi portaalista löytyy virtuaalikone, jolla dataa voidaan käsitellä.

Tässä tutkimuksessa laadittavassa laboratoriotyössä käytetään yksinkertaistettua CMSn avointa dataa, joka on CSV-tiedostomuodossa. Kysessä on siis tason 2 dataa (taulukko 1). CSV (comma-separated values) on tiedostomuoto, jossa tiedoston eri arvot erotetaan toisistaan pilkulla [36]. CSV-tiedostot voidaan avata tavallisessa taulukkolaskentaohjelmassa. Työssä käytetyt CSV-tiedostot on johdettu peruskokoelmista portaalista löytyvän ohjelmakoodin⁴ avulla.

Luodut CSV-tiedostot sisältävät ilmaisimen mittaamia ja mittaustuloksista määritettyjä lukuarvoja erilaisille törmäyksissä havaituille hiukkasille. Kuviossa 10 on esitetty osa työssä käytetystä CSV-tiedostosta tekstinkäsittelyohjelmassa avattuna. Tiedostoon on valittu vain sellaisia törmäystapahtumia, joissa on esiintynyt tasan

³<http://opendata.cern.ch/>

⁴<http://opendata.cern.ch/record/552>

kaksi myonia. Törmäyksissä voisi esiintyä toki useampiakin myoneja kerralla, mutta nämä tapahtumat on rajattu pois.

```

Run,Event,Type1,E1,px1,py1,pz1,pt1,eta1,phi1,01,Type2,E2,px2,py2,pz2,eta2,phi2,02,M
165617,74601783,G,9.69873,-9.51843,0.366285,1.86329,9.51748,0.194546,3.10311,-1,6,9.76332,7.32771,-1.15237,6.34728,7.41776,0.775564,-0.155984,1.17.4922
165617,75108943,G,6.20385,-4.26661,0.456545,-4.4793,4.29097,-0.91207,3.03499,-1,6,9.66898,7.27397,-2.82112,-5.7104,7.80188,-0.678625,-0.369977,1.11.5534
165617,75587682,G,19.2892,-4.21208,-0.651623,18.8121,4.26219,2.19046,-2.98811,-1,6,9.82439,4.34394,-0.473505,8.79849,4.36967,1.44967,-0.108575,1.9.16361
165617,75660978,G,7.04268,-6.32678,-0.268545,3.08025,6.33248,0.469033,-3.09917,-1,6,5.58571,4.47476,0.848935,-3.23185,4.55458,-0.660499,0.187488,1.12.4774
165617,75947690,G,7.27507,0.102966,-5.53308,-4.72122,5.53403,-0.773616,-1.55219,-1,6,7.31811,-0.398831,6.94081,2.28247,6.95226,0.322677,1.62819,1.14.3159
165617,74570517,G,14.9422,-9.17441,-10.3203,-5.70786,13.0007,-0.402405,-2.29748,-1,7,2.98793,-0.986225,0.354098,-2.79616,1.04787,-1.70803,2.79688,1.6.82697
165617,74697773,G,41.9538,-1.47942,-12.4648,40.0319,12.5523,1.87664,-1.68893,-1,6,10.1005,1.62377,7.52991,-6.5324,7.783,-0.769739,1.35841,1.39.5356
165617,74704205,G,14.4304,-4.34083,-6.0729,-12.3492,7.46478,-1.27743,-2.19138,1,6,24.7433,9.40612,8.61811,21.2008,12.7572,1.28132,0.741767,-1.37.7372
165617,75167029,G,0.75833,2.39032,-3.78235,8.67145,4.47435,1.41558,-1.0072,-1,6,13.5684,-3.39746,5.19098,12.0665,6.20395,1.41875,2.15032,1.10.5391
165617,75206813,G,10.1623,0.476262,-8.51642,5.52306,8.52973,0.609133,-1.51493,-1,6,19.7754,2.7508,-13.9871,13.7059,14.255,0.853872,-1.37661,1.3.11319
165617,75201957,G,9.63652,1.65156,-5.43133,7.78616,5.67688,1.12134,-1.2756,1,6,7.02632,-1.64493,4.44183,5.18867,4.73663,0.947273,1.92546,-1.10.4079
165617,75330864,G,12.5078,-2.86936,5.65922,-10.7794,6.34508,-1.30016,2.80405,1,6,13.5187,3.08626,-7.81458,-10.5901,8.48194,-1.0541,-1.19466,-1.14.6994
165617,75386136,G,14.5274,6.51727,-6.20374,11.405,8.99785,1.05849,-0.768757,1,6,17.162,7.98902,-6.62707,13.6668,10.3799,1.08857,-0.692485,-1.0.751482
165617,75393376,G,6.01118,3.36137,2.88831,4.0598,4.43184,0.820752,0.709849,-1,6,7.75571,-5.04158,-3.01265,5.06422,5.87312,0.780559,-2.60297,1.10.1786
165617,74431562,G,26.416,-12.835,-5.88741,-22.3247,14.1209,-1.23885,-1.71153,1,6,38.5893,13.6568,10.292,-34.5932,17.1007,-1.45384,0.645814,-1.31.0799
165617,74452546,G,12.0723,-3.73219,2.51036,-11.2026,4.4979,-1.64374,2.54948,-1,6,10.4401,3.35463,-3.42022,-9.27537,4.79076,-1.41468,-0.795879,1.9.29997
165617,74496886,G,6.22365,-0.589158,4.20978,-4.55424,4.24046,-0.932742,1.69116,-1,6,5.75664,-0.530833,-5.4727,1.70163,5.49838,0.304739,-1.66749,1.11.5201
165617,74635178,G,18.2409,7.54322,1.87995,16.5011,7.53925,1.49691,0.244751,1,6,4.14056,-3.9333,-1.27875,0.16435,4.13594,0.0397265,-2.82726,-1.14.4845
165617,74642386,G,31.2704,-0.127271,-0.85753,30.214,8.05853,2.83205,-1.555,1,6,15.8815,-1.15408,-7.70752,13.8374,7.79344,1.33849,-1.71943,-1.5.76065
165617,74751994,G,44.2247,8.09262,7.92033,42.7504,11.3235,2.08374,0.774639,-1,6,17.1959,-2.85587,-3.23714,16.6449,4.31683,2.05913,-2.2937,1.13.9764
165617,74783970,G,21.4339,5.10969,3.8105,20.4639,6.37407,1.88299,0.640769,-1,6,5.13736,-3.41692,-3.69615,1.02199,5.03357,0.201664,-2.31696,1.15.5406
165617,74969122,G,59.9226,-46.1516,29.4115,-24.407,54.7266,-0.432382,2.57421,1,6,52.4465,30.1097,-16.3989,-39.6876,34.2859,-0.988511,-0.498717,-1.89.9557
165617,75075850,G,14.9805,-3.94723,7.78366,12.0767,8.72731,1.12852,2.04012,-1,6,5.47065,-4.55124,0.221568,-3.02555,4.55663,-0.622915,3.09295,1.14.0268
165617,75215378,G,9.18082,-4.49262,-6.6885,4.24562,8.05727,0.585169,-2.16227,-1,6,11.1368,-5.26164,-0.29218,5.25099,9.82064,0.512021,-2.13622,1.0.320191
165617,75244714,G,6.34141,4.70274,-2.52771,-3.42013,5.33901,-0.60332,-0.493193,1,6,7.27079,-5.46354,1.06529,-4.67634,5.55079,-0.76355,2.94905,-1.10.8177
165617,75392933,G,5.17516,-2.72562,-4.07641,-1.65075,4.90369,-0.33058,-2.16016,-1,6,6.87391,4.21752,2.06896,4.60663,5.10082,0.81118,0.597347,1.11.5221
165617,75447381,G,17.619,5.76253,0.763444,16.6321,5.81288,1.77363,0.131717,1,6,8.97265,-4.5331,-0.11933,7.74171,4.53467,1.30448,-3.11527,-1.10.5406
165617,75548869,G,15.0038,2.53353,-6.30229,-13.3778,6.79247,-1.42991,-1.18857,1,6,19.2025,-4.03516,17.8477,-5.82243,18.2982,-0.313059,1.79315,-1.25.8044
165617,75783677,G,19.1211,4.12483,-4.60754,18.0931,6.18415,1.79469,-0.84062,1,6,4.64652,-2.16233,3.95677,-1.1168,4.50907,-0.245215,2.07094,-1.16.5054

```

Kuvio 10. Ruutukaappaus CSV-tiedoston rakenteesta. Ylin rivi sisältää eri sarakkeiden otsikot. Arvot erotetaan toisistaan pilkuilla.

Kunkin myoniparin tiedot esitetään tiedostossa omalla rivillään pilkuilla erotettuina arvoina. Ylimmällä otsikkorivillä määritellään, mihin suureeseen tai tietoon mikäkin arvo viittaa. Esimerkiksi $E1$ on ensimmäisen myonin energia ja $E2$ toisen myonin energia. Otsikkorivin eri nimet viittaavat arvoihin seuraavasti:

- Run = ajon numero, jonka aikana data on kerätty,
- Event = törmäystapahtuman numero,
- Type = myonin tyyppi; globaali myoni (G) on mitattu sekä pii-ilmaisimessa että myonikammiossa, tracker-myoni (T) on mitattu vain pii-ilmaisimessa (luokittelut ovat hypoteeseja, tyyppiä ei tiedetä täysin varmasti),
- E = myonin energia
- px, py, pz = myonin liikemäärän erisuuntaiset komponentit
- pt = poikittaisliikemäärä, eli myonin liikemäärän hiukkassuihkua vastaan kohtisuora komponentti,
- eta = η , eli pseudorapiditeetti, kulmaa kuvaava koordinaatti (ks. kuvio 7),
- phi = ϕ , eli atsimuuttikulma, myös kulmaa kuvaava koordinaatti (ks. kuvio 7),
- Q = myonin sähköinen varaus
- M = myoniparille tiedostoon valmiiksi laskettu invariantti massa (tieto poistetaan laboratoriotyössä)

3.4 TIM-alusta

TIM (The Interactive Material, vuorovaikutteinen materiaali) on ”Jyväskylän yliopiston tietotekniikan laitoksella kehitetty avoimen lähdekoodin (MIT-lisenssi) alusta helpottamaan vuorovaikutteisen materiaalin tuottamista verkkoon” [37]. MIT on Massachusetts Institute of Technologyssa kehitetty ohjelmistolisenssi [38, s. 85][39].

TIM-alustalla voidaan luoda TIM-dokumentteja. Dokumenteissa pystytään yhdistämään tavallisen tekstin ohkeen esimerkiksi kuvia, videoita ja vuorovaikutteisia komponentteja (plugins). Komponentit voivat sisältää esimerkiksi avoimia tekstivastauksia, monivalintatehtäviä tai eri ohjelmointikielillä ajettavia ja muokattavia ohjelmia. Dokumentin käyttäjä voi kirjoittaa, muokata ja ajaa koodia suoraan TIM-alustassa. [37]

TIM toimii verkkoselaimessa osoitteessa <https://tim.jyu.fi/>. Selainpohjaisuus mahdollistaa käytön millä tahansa laitteella, jolla käyttäjä pääsee verkkoon (puhelin, tabletti, tietokone). TIMin käyttö vaatii joko Jyväskylän yliopiston Korppi-järjestelmän tunnukset tai TIM-tunnukset, jotka käyttäjä voi luoda itselleen. TIM toimii ns. pilvipalveluna, eli kaikki TIMillä tehty ja luotu tallentuu verkkoon, ei käyttäjän omalle laitteelle. [37]

Alustan käyttö on rajoitettu vain Suomen sisälle mahdollisten ulkomailta tulevien verkkohyökkäysten vähentämiseksi. Suomen ulkopuolelta käyttö onnistuu kuitenkin VPN-yhteyden avulla. VPN on julkisia yhteyksiä käyttäen rakennettu tietoverkko, jonka avulla voidaan yhdistää yksityiseen sisäiseen tietoverkkoon, esimerkiksi tässä tapauksessa Jyväskylän yliopiston verkkoon [40].

3.5 Hiukkasfysiikka Jyväskylän yliopistossa

Jyväskylän yliopiston fysiikan laitoksella hiukkasfysiikan tutkimus tapahtuu kahdella pääalalla sekä kokeellisesti että teoreettisesti. Pääalat ovat *neutriinofysiikka ja kosmologia* sekä *ultrarelativististen raskasioneitörmäysten fysiikka*. CERNin LHC-kiihdyttimen kokeista Jyväskylä on mukana ALICE-kokeessa. [41]

Hiukkasfysiikkaa opetetaan fysiikan laitoksella muutamilla eritasoisilla kursseilla. Ensimmäisen kerran hiukkasfysiikkaa opetetaan kurssilla *FYSA2002 Moderni fysiikka, osa B*, jossa käsitellään mm. standardimallin alkeishiukkasia, erityisen suhteellisuusteorian perusteita sekä hiukkasreaktioiden kinematiikkaa [42]. Tutkimuksessa kehitettävä laboratoriotyö otettiin käyttöön juuri tällä kurssilla yhdeksi vaihtoehdoksi ns. lapputyöksi. Lapputyöllä tarkoitetaan fysiikan laitoksella laboratoriotyötä, josta ei palauteta erillistä työselostusta. Kurssilla annetaan kaksi mahdollisuutta laboratoriotöiden tekemiseen: opiskelija voi tehdä joko kolme lapputyötä tai kaksi lapputyötä ja yhden arvosteltavan työn, josta palkitaan ylimääräisillä kurssipisteillä [43].

Tarkemmin hiukkasfysiikkaa käsitellään mm. kursseilla *FYSS4300 Hiukkasfysiikka*

ja FYSS4456 Hiukkasfysiikan kokeelliset menetelmät [44]. Tutkimuksen tekohetkelä fysiikan laitoksen laboratoriotöissä ei vielä ollut hiukkasfysiikkaa käsitteleviä töitä [45].

3.6 CERNin avoimen datan opetuskäyttö

Tässä alaluvussa perehdytään jo toteutettuihin sovelluksiin ja oppimateriaaleihin, joissa on käytetty CERNissä kerättyä hiukkasfysiikan tutkimusdataa opetustarkoituksessa. Useat eri tahot käyttävät CERNissä kerättyä hiukkasfysiikan tutkimusdataa opetuksessa. Dataa hyödynnetään sekä lukio- että yliopistotasolla. Sovellukset keskittyvät mm. data-analyysin harjoitteluun, hiukkasten tunnistamiseen törmäystapahtumista sekä opiskelijoiden tutustuttamiseen fyysikon työhön.

Barnett ym. [46] esittelevät artikkelissaan ”Learning with the ATLAS experiment at CERN” erilaisia opetusprojekteja ja -tilanteita, joissa on hyödynnetty ATLAS-ilmaisimen keräämää tutkimusdataa. Opetusprojektit on toteutettu *Learning with ATLAS@CERN* -projektin alla. Yksi esimerkki tällaisesta on Tukholmassa kymmenelle Tukholman yliopiston ja AlbaNovan yliopistokeskittymän House of Science -keskuksen yliopisto-opiskelijalle järjestetty tilaisuus. Tilaisuuden alussa opiskelijoille esiteltiin hiukkasfysiikkaa, ATLAS-koetta sekä ”Learning with ATLAS” -projektia. Tämän jälkeen opiskelijat tutkivat ATLASin MINERVA-työkalun⁵ avulla törmäystapahtumia ja määrittivät K^0 -hiukkasen massan ja elinajan. Tilaisuudesta kerätyn palautteen perusteella opiskelijat pitivät tilanteesta ja harjoituksesta sekä suosittelisivat sitä muille opiskelijoille. [46, s. 30–31]

Iso-Britanniassa Birminghamin yliopistossa kehitettiin tietokonepohjainen laboratoriotyö, jossa käsitellään lyhytikäisten Z - ja W -bosonien havaitsemista ATLAS-kokeessa. Ensimmäisen kerran työ esiteltiin Birminghamin yliopistossa kevätlukukaudella 2010. Työ tehdään parityönä kahdessa neljän tunnin osassa ilman erillisiä ohjaajia. Tavoitteena työssä on tutustua hiukkasten havaitsemiseen ja mittaamiseen liittyviin tekniikoihin MINERVA-työkalun avulla. MINERVAn avulla pystytään määrittämään leptonipareille invariantti massa, joka taas edelleen mahdollistaa Z - tai W -bosonin massan määrittämisen. Artikkelin julkaisuvuoteen 2012 mennessä työn oli tehnyt yli sata yliopisto-opiskelijaa. Palaute opiskelijoilta oli ollut hyvin positiivista. [46, s. 31]

Myös Kreikassa on tehty hyvin vastaavaa työtä kuin Birminghamissa. Ateenan yliopistossa kehitettiin neljännen vuoden opiskelijoille suunnattu laboratoriotyö Z - ja W -bosonien havaitsemisesta ja mittaamisesta. Myös tässä työssä tavoitteena on tutustuttaa opiskelijat ATLAS-kokeessa käytettäviin analyysitekniikoihin. Erona Birminghamin työhön on kuitenkin se, että Ateenassa työ tehdään kolmen hengen ryhmässä ja ennen työtä opiskelijoille pidetään n. 1,5 tunnin luento työn teoreettisista ja käytännöllisistä asioista. [46, s. 32]

⁵<http://atlas-minerva.web.cern.ch/atlas-minerva/>

Itse työ koostuu kolmesta osasta. Ensimmäisessä opiskelijat tunnistavat Z -bosonin hajoamisesta syntyneitä myoneita ja elektroneita. Toisessa osassa lasketaan HYPATIA-ohjelmiston⁶ avulla invariantteja massoja sekä muodostetaan niistä histogrammi Z :n massan selvittämiseksi. Kolmannessa osassa opiskelijat tunnistavat kymmenen törmäystapahtuman joukosta yhden Higgsin bosonin sisältävän tapahtuman. Kotitehtävänä opiskelijat raportoivat tuloksistaan sekä vastaavat niihin liittyviin kysymyksiin. [46, s. 32–33]

Artikkelin [46] mukaan yhteensä 60 opiskelijaa oli suorittanut Ateenan yliopistossa kehitetyn työn. Valtaosa opiskelijoista piti HYPATIAn käyttöä kiinnostavana ja motivoivana. Opiskelijat arvostivat, että työssä käytettiin oikeaa hiukkasfysiikan dataa sekä samankaltaisia työkaluja kuin fyysikotkin käyttävät. Opiskelijat kokivat haastavana ja motivoivana sen, että analysointia tehtiin lähes reaaliajassa (online) ja että he antoivat oman panoksensa hiukkasten löytämiseen. [46, s. 33]

Edellä mainittujen lisäksi ATLAS-kokeen avointa dataa on hyödynnetty myös esimerkiksi Kreetassa opettajien kesäkoulussa [46, s. 33] sekä artikkelin julkaisuhetkellä lähes valmiina olleessa *ATLAS Multimedia Educational Lab for Interactive Analysis* (CAMELIA) -ohjelmassa [46, s. 35]. CAMELIA mahdollisti törmäystapahtumien interaktiivisen visualisoinnin. Ohjelma on kuitenkin todennäköisesti lopetettu, sillä sen verkkosivuja ei enää löytynyt aktiivisina.

Artikkelissa mainittujen sovelluskohteiden lisäksi ATLAS on julkaissut verkkosivullaan uutisen, jossa kerrotaan Dortmundin teknillisen yliopiston laboratorioskursista, jossa käytetään ATLASin keräämää tutkimusdataa [47]. Dortmundin kerrotaan olevan ensimmäinen yliopisto, jossa ATLASin avointa dataa hyödyntävä työ sisällytetään osaksi opetussuunnitelmaa. Toisaalta tämä on ristiriidassa Barnett ym:n artikkelin kanssa, sillä Birminghamin harjoituksen mainittiin tapahtuvan yliopistokurssilla. [46, s. 31] [47]

Yksi laajalti käytetty CERNin eri kokeiden tutkimusdatan opetuksellisen hyödyntämisen muoto on Masterclass-harjoitus [48]. Masterclass-harjoituksessa yhden päivän aikana lukio-opiskelijat osallistuvat luennolle hiukkasfysiikasta, analysoivat hiukkasfysiikan dataa sekä käyvät läpi tuloksia videokonferenssissa esimerkiksi CERNin tutkijoiden tai jonkun muun instituutin edustajien kanssa. Masterlasseille osallistuu vuosittain yli 13000 lukio-opiskelijaa 52:sta maasta. Suomessa Masterlasseja on järjestetty Helsingin ja Jyväskylän yliopistoissa. [48]

Esimerkiksi CMS Masterclassissa käytetään datan tarkasteluun työkaluna *iSpy-webgl*⁷-nimistä visualisointisovellusta [49]. Sovelluksen avulla opiskelijat voivat tarkastella visuaalisesti törmäystapahtumia ja niissä havaittuja hiukkasia. Tulokset kirjataan yhteiseen taulukkoon, jolloin myös videokonferenssin kautta osallistuvat ammattilaiset pääsevät käsiksi tuloksiin.

Helsingin yliopiston fysiikan laitokselle tehdyssä pro gradu -tutkielmassa *Avoin*

⁶<http://hypatia.phys.uoa.gr/>

⁷<https://www.i2u2.org/elab/cms/ispy-webgl/>

hiukkasfysiikan tutkimusdata opetuskäytössä Suoniemi [50] on toteuttanut osana tutkielmaa tapaustutkimuksen Helsingissä vuonna 2013 järjestetystä Masterclass-tapahtumasta. Tutkimuksessa selvitettiin osallistuvan havainnoinnin ja kyselylomakkeen avulla 21 lukiolaisen mielipiteitä tapahtuman vahvuuksista ja heikkouksista. Tapahtuman tärkeimmiksi anneiksi mainittiin mielenkiintoiset luennot, hiukkasfysiikkaan tutustuminen ja tutkimustyön tekeminen. Tapahtumaa pidettiin onnistuneena ja data-analyysia mielenkiintoisena. Tapahtuman yhteydessä olleet luennot koettiin raskaiksi tai vaikeiksi. [50, s. 32–36]

Edellä mainittujen sovelluskohteiden lisäksi CMS-ilmaisimen avoimen datan opetuskäyttöä kartoitettiin lyhyen, englanninkielisen sähköpostikyselyn avulla [51]. Kysely lähetettiin sähköpostilistalle, joka kattoi kaikki CMS-kollaboraatioissa työskentelevät henkilöt. Viestissä kysyttiin seuraavat kysymykset (suomennokset: Rikkilä) [51]:

- Oletko käyttänyt tai tiedätkö ketään kuka olisi käyttänyt CMS:n avointa dataa opetuksessa yliopisto- tai lukiotasolla?
- Jos vastasit kyllä, niin mikä oli käyttötapaus (tavallinen fysiikan kurssi, hiukkasfysiikan kurssi, joku muu...) ja onko käytöstä saatavilla dokumentaatiota?
- Miten käsittelit(te) dataa? (CMSSW VM:llä, ROOT, Jupyter Notebook tai Python Notebook, visualisointiohjelma, taulukkolaskenta tai muu...)
- Onko käyttötapaus jo toteutettu vai onko se suunnitteilla?

Sähköpostiin vastasi yhteensä 19 henkilöä ja vastauksista 13 sisälsivät kartoituksen kannalta oleellista tietoa. Vastajista 7 oli käyttänyt dataa yliopistotasolla ja 4 lukiotasolla. Pääosin dataa oli käsitelty virtuaalikoneella (CMSSW) ja ROOT-ohjelmalla. Kaksi vastaajaa oli hyödyntänyt Jupyter Notebookia datan käsittelyssä. Käyttökohteiksi vastaajat mainitsivat mm. hiukkasfysiikan kurssin, toisen asteen koulutuksen päättötestin (matura thesis), Masterclass-harjoituksen sekä lopputyön.

Edellä mainittujen lisäksi verkosta löytyy myös muita hiukkasfysiikan opetuksellisia sovelluskohteita, joista osa käyttää myös avointa tutkimusdataa. Sovelluksia on koottu esimerkiksi QuarkNet⁸- ja Go-Lab⁹-verkkosivuille.

⁸<https://quarknet.i2u2.org/data-portfolio>

⁹<http://www.golabz.eu/>

3.7 Fysiikan käytännön töiden tavoitteet

Tutkimuksessa luotavan laboratoriotyön arvioimisessa ja kehittämisessä hyödynnetään Nivalaisen, Asikaisen ja Hirvosen [6] vuonna 2013 julkaistussa artikkelissa *Pre-service teachers' objectives and their experience of practical work* esittämiä fysiikan kokeellisten töiden tavoitteita. Artikkelissaan Nivalainen ym. tutkivat fysiikan opettajaopiskelijoiden näkemyksiä fysiikan käytännön töiden tavoitteista koulussa ja yliopistossa. Käytännön työksi tutkijat määrittelevät kaikki kokeelliset aktiviteetit, mukaanlukien demonstraatiot, simulaatiot, tietokonemallinnukset sekä kursseille osallistuvien tekemät työt. [6, s. 1]

Nivalainen ym. havaitsivat aikaisempiin fysiikan kokeellisten töiden tavoitteiden määritelmiin tutustuessaan, että jotkut tavoitteet esiintyivät artikkeleissa toisia useammin. Kuitenkin myös eroja eri tutkijoiden välillä esiintyi. Tästä syystä Nivalainen ym. muodostivat viisi uutta tavoitekategoriaa perustuen seitsemään aiempaan tutkimukseen. [6, s. 2–3]

Näitä tavoitekategorioita käytetään tutkimuksessa luotavan laboratoriotyön arvioimisessa ja kehittämisessä. Nivalaisen ym. kategorioita voidaan pitää hyvin käytännön työn tavoitteita kuvaavina, sillä kategoriat on koostettu useampaan aiempaan tutkimukseen perustuen. Tavoitekategoriat ovat seuraavat [6, s. 3] (käännökset: Rikkilä):

1. Käytännöllisten tai kokeellisten taitojen kehittäminen
2. Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen
3. Motivaation lisääminen
4. Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittäminen
5. Sosiaalisten ja oppimisen taitojen parantaminen

Käytännöllisten tai kokeellisten taitojen kehittämisellä Nivalainen ym. tarkoittavat opiskelijoiden taitojen ja kykyjen kehittämistä esimerkiksi hypoteesien muodostamisessa, kokeiden suunnittelussa, havainnoinnissa, datan tulkitsemisessä, virheiden käsittelyssä ja raportoinnissa. Tämä kategoria esiintyi muodossa tai toisessa jokaisessa seitsemässä aiemmassa tutkimuksessa. [6, s. 2]

Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen nähtiin Nivalaisen ym. mukaan monissa aiemmissa tutkimuksissa yhtenä tärkeimmistä tavoitteista. Käytännön töiden vaikutusta opiskelijoiden tieteellisen ymmärryksen kehittämiseen oli kuitenkin myös kritisoitu aiemmissa tutkimuksissa. Siitä huolimatta Nivalaisen ym. mielestä käytännön työt tarjoavat opiskelijoille mahdollisuuden tutkia tieteellistä sisältöä eri näkökulmasta kuin puhtaasti teoreettinen opetus. Tämän takia he pitävät käytännön töitä tärkeänä lisänä muuhun opetukseen. [6, s. 2–3]

Nivalaisen ym. tarkastelemissa aiemmissa tutkimuksissa motivaation lisääminen mainittiin yhdeksi käytännön töiden tavoitteeksi. Kategoriolla tarkoitetaan tavoitetta lisätä opiskelijoiden motivaatiota tieteen oppimista kohtaan. Sekä aiemmissa kokeellisissa tutkimuksissa että kannanotoissa esitettiin Nivalaisen ym. mukaan, että motivaation lisäämiseen liittyen käytännön töiden tavoitteina on luoda kiinnostuksen, hovin ja ilon tunteita. [6, s. 3]

Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittämistä ehdotettiin Nivalaisen ym. mukaan aiemmissa artikkeleissa yhdeksi käytännön töiden tavoitteeksi. Tieteen luonteella tarkoitetaan kuvausta siitä, mitä tiede on, miten se toimii, miten tutkijat toimivat sosiaalisena ryhmänä ja miten yhteiskunta reagoi tieteeseen ja ohjaa sitä [52, s. 4]. Myös tieteellisen prosessin ymmärtämistä ehdotettiin käytännön töiden tavoitteeksi aiemmissa tutkimuksissa. Nivalainen ym. yhdistivät nämä tavoitteet yhdeksi tavoitekategoriaksi. [6, s. 3–4]

Viimeiseksi viidestä tavoitekategoriasta Nivalainen ym. valitsivat sosiaalisten ja oppimisen taitojen parantamisen. Tavoitteeseen liittyvät esimerkiksi yhteistyö- ja kommunikaatiotaidot sekä mahdollisuudet opiskelijan ja opettajan väliseen vuorovaikutukseen. Nivalainen ym. mainitsevat, että sosiaalisten ja oppimisen taitojen parantamisen tavoitetta ei esiintynyt ennen 1990-lukua ilmestyneessä kirjallisuudessa. He arvelevat, että mahdollisesti oppimisteorioiden kentän kehittyminen on johtanut tämän tavoitteen ilmaantumiseen 1990-luvulla. [6, s. 4]

3.8 Teoreettisen ongelma-analyysin yhteenveto

Teoreettisessa ongelma-analyysissä alaluvuissa 3.1–3.7 perehdyttiin hiukkasfysiikkaan, sen kokeelliseen tutkimukseen CERNissä, avoimen datan käsitteeseen, TIM-alustaan sekä hiukkasfysiikan opetukseen Jyväskylän yliopiston fysiikan laitoksella. Lisäksi tutustuttiin aikaisempiin sovelluksiin ja materiaaleihin, joissa on käytetty CERNin avointa dataa opetustarkoituksessa, sekä Nivalaisen, Asikaisen ja Hirvosen [6] laatimiin fysiikan käytännön töiden tavoitekategorioihin.

Tässä alaluvussa kootaan yhteen edellä mainituista tehtyjä havaintoja. Lisäksi näiden perusteella määritellään tutkimuksen kehittämisen tarpeita, mahdollisuuksia, rajoitteita ja haasteita.

Hiukkasfysiikka ja sen kokeellinen tutkimus ovat monimutkaisia kokonaisuuksia, kuten alaluvuista 3.1 ja 3.2 huomataan. Jyväskylän yliopiston fysiikan laitoksella ei perehdytä tarkasti hiukkasfysiikkaan ennen modernin fysiikan kurssia, jolle työ kehitetään. Kehittämisen haasteena on laatia työ siten, että opiskelijat varmasti ymmärtävät, mitä työssä tehdään.

Hiukkasten havaitsemisen ymmärtämiseen liittyy myös haasteita. Opiskelijoille on hyvin todennäköisesti uutta tietoa se, mitä eri ominaisuuksia hiukkasista voidaan mitata hiukkasilmäimellä. Lisäksi opiskelijoiden tulee ymmärtää, että kaikkia hiuk-

kasia ei voida havaita suoraan hiukkasilmaisimella. Epäsuoraan havaitsemiseen liittyvä invariantin massan käsite voi myös olla haastava sisäistä.

Toisaalta hiukkasfysiikan käsittelyn voi mieltää myös mahdolliseksi motivoivaksi tekijäksi. Aikaisemmissa vastaavissa hiukkasfysiikan opetusinterventioissa opiskelijoiden palaute on ollut positiivista [46, s. 30–31, 39]. Lisäksi aidon tutkimusdatan ja -menetelmien hyödyntämistä arvostettiin [46, s. 33].

Kehitettävä laboratoriotyö luodaan Jyväskylän yliopiston tietotekniikan laitoksella kehitetylle TIM-alustalle. Fysiikan laitoksella TIM-alustaa käytetään erilaisilla kursseilla, joten tutun alustan hyödyntäminen voi vähentää opiskelijoiden työssä kohtaamia haasteita. Tässä rajoitteena on kuitenkin se, että TIM-alusta ei ole käytettävissä Suomen ulkopuolelta. Rajoitteeseen vastataan luomalla työstä englanninkielinen versio Jupyter Notebook -alustalle. Jupyter Notebook on ominaisuuksiltaan TIM-alustan kaltainen avoimen lähdekoodin sovellus [53]. Englanninkielinen, avoimesti julkaistu versio tarjoaa mahdollisuuden työn hyödyntämisen opetuksessa eri puolilla maailmaa.

Työssä käytetään avoimesti julkistettua CMS-ilmaisimen tutkimusdataa. CMS-kokeen julkistama avoin data on Creative Commons CC0 -lisensioitua. CC0-lisenssi tarkoittaa, että data on julkistettu vapaaseen yleiseen käyttöön. Tämä mahdollistaa ylipäänsä koko laboratoriotyön toteuttamisen.

CMS:n avointa dataa on saatavilla erilaisissa tiedostomuodoissa. Työssä data on CSV-muotoista. Tämä mahdollistaa sen, että dataa voidaan käsitellä TIM-alustalla Python-ohjelmointikielen¹⁰ avulla. Monimutkaisemmat tiedostomuodot vaatisivat erilliset ohjelmistot, joiden käyttäminen vaatisi lisää osaamista opiskelijoilta.

Python-ohjelmointikielen käyttäminen luo myös haasteita, sillä Pythonilla ohjelmointi ei välttämättä ole opiskelijoille aikaisemmin täysin tuttua. Fysiikan laitoksen opetukseen sisältyy jatkossa Python-ohjelmointia osana kursseja, mutta tarkkaa opiskelijoiden Python-osaamisen tasoa on hankala ennustaa. Tähän haasteeseen vastataan siten, että työssä osa ohjelmointikoodista on valmiiksi kirjoitettuna. Tällä tavalla työn voi tehdä ilman varsinaista aiempaa kokemusta Python-ohjelmoinnista.

Se, että kaikki CMS:n avoin data ei ole valmiiksi CSV-muotoista, ei rajoita kehittämistä tässä tapauksessa. Laboratoriotyö kehitetään sellaiseksi, että se toimii datalla, joka löytyy jo valmiiksi CSV-muodossa. Kuten alaluvussa 3.3 mainittiin, CSV-tiedostoja voisi myös johtaa itse peruskokoelmista ohjelmointikoodilla. Tähän ei kuitenkaan ollut tarvetta.

Alaluvussa 3.5 käsiteltiin lyhyesti hiukkasfysiikan opetusta Jyväskylän yliopiston fysiikan laitoksella. Huomionarvoista on se, että fysiikan laitoksella ei vielä ole yhtään laboratoriotyötä, joka käsitelisi kokeellista hiukkasfysiikkaa. Tutkimuksen kehittämistuotoksella vastataan tähän tarpeeseen.

Aikaisemmin toteutettuihin sovelluksiin ja oppimateriaaleihin, joissa on käytetty

¹⁰<https://www.python.org/>

CERNissä kerättyä hiukkasfysiikan tutkimusdataa opetustarkoituksessa, perehdyttiin alaluvussa 3.6. Barnettin ym. [46] artikkelin perusteella erilaiset ATLAS-ilmaisimen tutkimusdatalla toteutetut opetusprojektit ja -tilanteet olivat onnistuneet hyvin. Opiskelijoiden palautteet olivat olleet positiivisia ja he olivat mm. arvostaneet, että työssä käytettiin oikeaa hiukkasfysiikan dataa sekä samankaltaisia työkaluja kuin fyysikotkin käyttävät [46, s. 30–31, 39].

Alaluvussa 3.7 tutustuttiin Nivalaisen ym. [6] määrittelemiin fysiikan kokeellisten töiden tavoitekategorioihin. Nämä kategoriat liittyvät kehittämisen tavoitteen määrittämiseen. Kehittämisen tavoitteena on luoda em. kategoriat mahdollisimman hyvin täyttävä laboratoriotyö. Kehittämisellä tavoitellaan myös tietoa kehittämisprosessin aikana ilmenneistä haasteista ja mahdollisuuksista. Lisäksi tutkimuksessa pyritään määrittelemään suosituksia onnistuneen, hiukkasfysiikan avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämiseen.

4 Ensimmäinen kehittämisvaihe

Tässä luvussa kuvaillaan kehittämisprosessin ensimmäisen syklin jälkimmäinen osio: ensimmäinen kehittämisvaihe. Ensimmäinen kehittämisvaihe pitää sisällään työn konkreettisen luomisen alusta alkaen.

Yksi kehittämistutkimuksen laadukkuuden kriteereistä (ks. luku 2.3) on se, että tutkimuksen kaikki syklit dokumentoidaan tarkasti. Lisäksi kehittämistutkimuksesta raportoidessa kaikki kehittämisen vaiheet tulee kuvailla [9, s. 38–39]. Seuraavissa alaluvuissa kerrotaan tarkasti ensimmäisen kehittämissyklin eri vaiheet kronologisessa järjestyksessä.

Ensimmäisessä kehittämisvaiheessa loin työn omiin kokemuksiini, asiantuntijoiden apuun sekä teoreettiseen ongelma-analyysiin perustuen. Kuvailen tekstissä perusteluni erilaisille kehittämisprosessissa tekemilleni valinnoille. Tässä kehittämisvaiheessa en vielä tietoisesti hyödyntänyt viitekehyksenä Nivalaisen, Asikaisen ja Hirvosen [6] fysiikan käytännön töiden tavoitteita. Käytin näitä tavoitteita varsinaisesti vasta työn onnistumisen arviointiin toisessa ongelma-analyysissa, joka on esitetty luvussa 5.

4.1 TIM-alustaan tutustuminen

Olimme sopineet Sami Räsäsen, Sakari Juutisen ja Jussi Maunukselan kanssa, että laadin kehittämäni laboratoriotyön TIM-alustalla (ks. luku 3.4) toimivaksi. TIM on Jyväskylän yliopiston tietotekniikan laitoksella kehitetty alusta, jonka avulla voidaan tuottaa vuorovaikutteista materiaalia verkkoon. TIMiä käytetään esimerkiksi joidenkin Jyväskylän yliopiston fysiikan laitoksen kurssien materiaali- tai tehtäväalustana. Koska TIM ei ollut minulle kovinkaan tuttu, aloitin kehittämisen tutustumalla TIM-alustan toimintaan.

Ennen tutkimuksen aloittamista olin muutamia kertoja kokeillut TIMiä ja saanut pienen perehdytyksen sen käyttöön fysiikan laitoksen yliopistontutkija Taneli Kalvakselta. En kuitenkaan ollut tarkemmin perehtynyt TIM-alustan ominaisuuksiin ja TIM-dokumenttien luomiseen.

Olin kuullut, että TIM-alusta tallentaa dokumentin haltijan nähtäväksi muutoksia, joita dokumentin käyttäjät ovat tehneet. Esimerkiksi jos käyttäjät tekevät TIMin kautta luotuun ohjelmakoodiin muutoksia, pitäisi näiden muutosten näkyä dokumentin haltijalle. Tämä tieto oli hyödyllistä varmistaa aineistonkeruun ja työn käytettävyyden kannalta, joten otin sähköpostilla yhteyttä TIMin kehittäjätiimiin asias-

ta. Kehittäjätiimistä minulle vastasivat Vesa Lappalainen ja Mika Lehtinen. He vahvistivat ominaisuuden olemassaolon.

Aloitin TIMiin tutustumisen luomalla uuden dokumentin. Tutustuin TIMin käyttöliittymään ja kokeilin sen eri ominaisuuksia. Loin uuden kappaleen (paragraph), jossa kokeilin *Edit paragraph* -valikon perusominaisuuksia, kuten otsikoiden, listojen ja kappalejakojen luomista. Tästä siirryin testaamaan kuvan sisällyttämistä dokumenttiin ja \LaTeX -ladontaa. Perusominaisuuksien hallinnan jälkeen otin käsittelyyn laboratoriotyön kannalta olennaiset ominaisuudet: tiedostojen lataamisen TIM-alustaan sekä interaktiivisten komponenttien (plugins) luomisen.

Tiedostojen laataminen TIMiin onnistuu alustan oman *Upload file/image* -toiminnon kautta. Haluttu tiedosto tallentuu TIMin sisäiseen hakemistoon, johon viittaamalla tiedostoa voidaan käyttää.

Kehittämisen ensimmäinen tekninen haaste tuli vastaan työssä käytettävien datatiedostojen lataamisessa TIMiin edellä mainitulla tavalla. Ilmeni, että suurin 72 Mt:n datatiedosto ylitti alustalle asetetun 50 Mt:n rajan ladattavan tiedoston koon suhteen. Olin asiasta yhteydessä TIMin kehittäjätiimiin. He ratkaisivat ongelman nostamalla väliaikaisesti rajan 80 Mt:hen. Väliaikaisen noston aikana sain ladattua myös suurimman datatiedoston TIMiin.

Tiedostojen lataamisen ohella tutustuin myös TIMin interaktiivisiin komponentteihin. Komponentit ovat dokumentin interaktiivisia osia, joita voi luoda joko valmiista komponenttipohjista tai alusta asti itse ohjelmoiden. Valmiista pohjista löytyy komponentteja esimerkiksi monivalintakysymysten luomiseen, graafien tekemiseen ja komentorivin upottamiseen dokumenttiin. Kehittämisen kannalta tärkein on komponentti, joka luo dokumenttiin upotuksen, johon käyttäjä voi kirjoittaa ajettavaa ohjelmointikoodia. Koodikomponentit mahdollistavat CSV-tiedostojen käsittelyn kehitettävässä työssä. Kuviossa 11 on esitetty ruutukaappauskuva yhdestä työn koodikomponentista.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')

print("Törmäystapahtumien lukumäärä = %d" % len(ds))
```

Aja

Highlight

Kuvio 11. Esimerkki TIMin koodikomponentista. Komponentti sisältää koodia, jota käyttäjä voi muokata. Koodi ajetaan alareunassa olevaa *Aja*-painiketta klikkaamalla. Painikkeella *Highlight* on mahdollista korostaa koodin eri osia.

Koodikomponentteja luodessa täytyi valita, millä ohjelmointikielellä käyttäjät voivat kirjoittaa komponenttiin. Valitsin työssä käytettäväksi kieleksi Python-ohjelmointikielen

(ks. luku 3.8). Olin jo aiemmin käsitellyt CMS-kokeen avointa dataa Pythonilla ja todennut sen toimivaksi. Pythonin valintaa tuki myös se, että Jyväskylän yliopiston fysiikan laitos oli suunnitellut tutkimuksen tekohetkellä lisäävänsä Pythonin käyttöä fysiikan peruskursseilla [54].

Loin ensimmäiset Python-komponentit ja totesin niiden toimivuuden. Komponentit toimivat halutulla tavalla, eli pystyin Pythonilla ohjelmoiden esimerkiksi hakemaan TIMiin ladatun CSV-muotoisen datatiedoston ja nimeämään sen. Lisäksi komponenttien yhteydessä oleva *Alusta*-painike osoittautui toimivaksi. Painikkeen avulla koodikomponentin pystyi palauttamaan siihen tilaan, jossa se oli työn alussa. Eli jos käyttäjä lisäisi annettuun koodiin jotakin toimimatonta, pääsisi *Alusta*-painikkeen avulla takaisin tilaan, jossa on vain komponentissa valmiina ollut koodi.

Yhden koodikomponentin jälkeen siirryin kokeilemaan monen komponentin yhdistämistä. Työn kannalta tärkeä ominaisuus on se, että yhdessä komponentissa tehdyt asiat ovat käytettävissä myös myöhemmissä komponenteissa. Esimerkiksi jos opiskelija hakisi työn ensimmäisessä koodikomponentissa datatiedoston käyttöönsä ja nimeäisi sen, täytyisi tiedoston toimia uudella nimellä myös seuraavissa komponenteissa. Tässä ilmeni teknisiä haasteita. Komponentit eivät vuorovaikutaneet keskenään automaattisesti, eli yhdessä koodikomponentissa tehdyt asiat eivät välittyneet seuraavaan. Myös tähän haasteeseen kysyin apua TIMin kehittäjätiimiltä.

Ratkaisuna oli luoda jokaisen koodikomponentin alkuun käyttäjältä piilossa oleva koodi, joka ajaa aina edellisen komponentin ennen kyseessä olevan komponentin koodiin siirtymistä. Tällöin syntyy ketju, jossa esimerkiksi kolmas komponentti ajaakin ensin toisen komponentin, joka ajaa taas ensimmäisen komponentin ennen itseään. Valitsin tämän ketjutustavan komponenttien vuorovaikutuksen toteutustavaksi. Haittapuolena ketjutuksessa on se, että ajaminen käy sitä raskaammaksi, mitä enemmän komponentteja ketjutetaan, sillä jokainen komponentti ajaa aina kaikki edelliset.

Ennen varsinaisen työn luomisen aloittamista kokeilin vielä kuvaajien tuottamista TIMillä. Hiukkasfysiikan analyysissa kuvaajat ovat tärkeässä osassa, joten tahdoin varmistua, että niiden luominen onnistuu sujuvasti. Valitsin kuvaajien luomiseen Pythonin *matplotlib.pyplot* -moduulin¹¹, sillä olin jo aiemmin todennut sen toimivan hyvin CSV-muotoisen datan kanssa. Totesin, että moduuli toimii myös TIMissä.

¹¹<https://matplotlib.org/index.html>

4.2 Työn rakenteen luominen

Olimme suunnitelleet jo ennen työn luomista Sami Räsäsen kanssa työn perusrungon. Sovimme, että työ sisältää taustatietoja ja teoriaa, jolloin työ opettaa myös hiukkasfysiikan teoreettista puolta ja käsitteitä. Konkreettiseksi tehtäväksi sovimme datassa esiintyvien myoniparien invariantin massan laskemisen sekä sen perusteella tehtävän pienimuotoisen data-analyysin. Lisäksi päätimme, että työssä käsitellään mittauksiin liittyvää epätarkkuutta.

Seuraavat alaluvut sisältävät kehittämisprosessin kulun kuvailun, sekä eri prosessin vaiheissa tekemäni valinnat perusteluineen. Prosessi on jaettu työn eri osien luomista kuvaileviin alalukuihin. Kuvailussa viitataan liitteenä olevaan PDF-vedokseen (LIITE A) työn siitä versiosta, jota fysiikan laitoksen opiskelijoista koostuva testiryhmä testasi. Huomionarvoista on, että PDF-vedos ei näytä visuaalisesti samalta kuin oikea, TIM-alustalla tarkasteltava työ. Esimerkiksi osa TIMin painikkeista jää näkymättömiin ja piilossa olevat avattavat vihjeet taas ovat vedoksessa näkyvissä. Vedoksen sisältö vastaa kuitenkin täysin työn sisältöä.

Lisäksi tärkeä huomio on se, että liitteenä oleva vedos on jo kerran testattu ja hieman kehitetty versio. Kuten luvussa 2.2 mainittiin, toteutin työn testaamisen kahdessa eri vaiheessa. Ensimmäisessä testauksessa työn teki pareittain yhteensä kuusi CERNissä kesäharjoittelussa ollutta matemaattis-luonnontieteellisten alojen yliopisto-opiskelijaa. Kukin pari kommentoi keskenään ääneen työn tekemistä. Lisäksi havainnoin parien työskentelyä työn aikana. Työn tekemisen jälkeen haastattelin kaikki kolme paria. Myös ohjaajani Kati Lassila-Perini kokeili työtä ja antoi kehittämis ehdotuksia.

Testaamisen toisessa vaiheessa taas yhdeksän Jyväskylän yliopiston fysiikan laitoksen opiskelijaa testasi itsenäisesti työtä. Havainnoin testitulannetta sekä tallensin ruutukaappausvideot testaajien näytöiltä. Työn tekemisen jälkeen jokainen testaja osallistui myös yksilöhaastatteluun.

Näiden kahden eri vaiheen välissä muokkasin työtä CERNin testiparien havainnoinnin perusteella. Koska CERNin ja Jyväskylän testausten välissä oli ainoastaan neljä päivää aikaa, päädyin käsittelemään CERNissä tehtyjä testauksia vain pienimuotoisesti. Jätin huomioimatta parien haastattelut ja käytin ainoastaan parien havainnoinnin aikana tekemiäni huomioita sekä Lassila-Perinin kommentteja. Näiden perusteella tein työhön noin 40 erilaista muutosta. Suurin osa muutoksista oli luonteeltaan pieniä, kuten esimerkiksi kirjoitusvirheiden korjaamista, faktojen ja tekstin tarkentamista sekä matemaattisten lausekkeiden tarkempaa avaamista. Suurimpana muutoksena lisäsin mukaan kaksi ylimääräistä sanallista kysymystä. Tämä testivaihe kuvaillaan tarkemmin myöhemmin alaluvussa 5.2

Koska muutokset olivat pieniä, liitteenä on ainoastaan edellä mainittu, hieman kehitetty versio työstä (LIITE A). Tämä on siis Jyväskylässä testattavana ollut versio. Seuraavissa alaluvuissa on ilmaistu kohdat, joissa eri versioiden välillä on ollut merkittäviä eroja.

4.2.1 Teoriaosuus

Aloitin kehitettävän työn luomisen päättämällä alustavan rakenteen. Päätin, että työn alkuun tulee teoriaosa, joka sisältää tarvittavat teoreettiset lähtökohdat työn tekemiseen. Teoriaosan lisäämiseen päädyin siksi, koska pohdimme yliopistontutkija Räsäsen kanssa teoriaosion lisäävän työstä oppimista. Lisäksi uskoin, että modernin fysiikan kurssin luennoilla ei välttämättä käydä kaikkea työssä tarvittavaa teoriaa läpi. Myös useimmissa muissa fysiikan laitoksen laboratoriotöiden työohjeissa on mukana työhön liittyvää teoriaa käsittelevä osio.

Teoriaosiota laatiessa tajusin, että ennen työn julkaisua tulee selvittää joitakin käytännön asioita fysiikan laitoksen kanssa. Kuka esimerkiksi vastaa työn tarkastamisesta ja miten opiskelijat palauttavat tekemänsä työn vastuuhenkilölle? Päätin pohtia näitä tarkemmin vasta työn valmistuttua. Lisäksi aloin tässä vaiheessa miettiä työlle nimeä. Päädyin myöhemmin antamaan työlle nimen *FYSA2002: Hiukkasfysiikan data-analyysi aidolla CERNin tutkimusdatalla*.

Havaitsin teoriaosaan liittyen myös sen, että osa teoriasta täytyy valita jätettäväksi työn ulkopuolelle. Kaikkea työhön liittyvää teoriaa ei pystyisi sisällyttämään mukaan, joten kehittäessä tuli pohtia mitkä mahdollisesti olivat jo tuttuja asioita ja mitkä eivät. Päädyin sisällyttämään teoriaosaan seuraavat asiat:

- hiukkasten kiihdyttäminen ja törmäyttäminen,
- törmäyksissä syntyneiden hiukkasten havaitseminen,
- datan tallentaminen,
- hiukkasten epäsuora havaitseminen,
- invariantti massa -käsite ja
- Z -bosonin havaitseminen.

Työn teoriaosa (LIITE A-3–A-8) muistuttaa siis rakenteeltaan tämän raportin lukuja 3.2, jossa kuvaillaan kokeellista hiukkasfysiikan tutkimusta CERNissä.

Teoriaosaa kirjoittaessa huomasin, että osiosta tulee suhteellisen pitkä. Päätin rytmittää osiota esittämällä noin puolivälissä sekä lopussa lukijalle sanalliset tehtävät teoriasta. Ensimmäinen tehtävä sijoittuu työssä hiukkasten havaitsemisen teorian jälkeen (LIITE A-6). Lukijaa pyydetään kuvailemaan myonien käyttäytymistä ja havaitsemista CMS-ilmaisimessa. Toisessa, teoriaosion loppuun sijoittuvassa tehtävässä lukijaa pyydetään kuvailemaan omin sanoin Z -bosonin esiintymisen todistamista invariantin massan käsitteen avulla (LIITE A-8).

4.2.2 Datan hakeminen

Teoriaosan valmistuttua siirryin työstämään osiota, jossa aloitetaan varsinainen data-analyysi (LIITE A-9). Osion alkuun kirjoitin osion tavoitteen, kerroin käytettävän ohjelmointikielen olevan Python sekä avasin käytettävän datan taustoja. Kerroin myös, miten data on saatu CSV-tiedostomuotoon liittämällä mukaan linkin koodiin, jolla dataa on käsitely. Tätä linkkiä ja prosessin avaamista ei ollut työn aivan ensimmäisessä versiossa, vaan lisäsin sen mukaan CERNissä tehtyjen testausten perusteella.

Päätin lisätä osion alkuun myös kolme teknistä huomiota työhön liittyen (LIITE A-9). Ensimmäisessä huomiossa varoitin käyttäjiä mahdollisesta TIMin laskenta-ajan ylittymisestä syntyvästä virheilmoituksesta. On mahdollista, että käytössä olevan suuren datatiedoston käsittely aiheuttaa satunnaisesti liian suuren laskenta-ajan. Tämä voi kuitenkin korjautua ajamalla koodi uudestaan. Toisena teknisenä huomiona korostin TIMistä löytyvän *Alusta*-painikkeen käyttöä (ks. alaluku 4.1). Painikkeen avulla käyttäjä pääsee aina alustamaan koodikomponentin siihen tilaan, jossa se oli työtä aloitettaessa. Kolmantena huomiona päätin kertoa, että koodikomponenttien toiminta perustuu piilossa olevaan koodiin, joka ajaa aiempia komponentteja. Halusin tällä tiedolla vähentää mahdollista epätietoisuutta siitä, mitä koodikomponenteissa tapahtuu. Työn tekijä saattaisi ihmetellä, miten aiemmissä komponenteissa tehdyt asiat toimivat myös seuraavissa, vaikka komponenteissa ei näy tähän viittaavia komentoja. Tämän kolmannen huomioon lisäsin työhön vasta CERNissä tehtyjen testausten jälkeen Lassila-Perinin ehdotuksesta.

Ensimmäisessä työvaiheessa opiskelija hakee käsittelyyn tutkimusdataa sisältävän tiedoston (LIITE A-9). Tiedoston hakemista varten loin koodikomponentin, johon kirjoitin käyttäjälle valmiiksi tarvittavan Python-koodin. Käyttäjän tulee siis työtä tehdessä ainoastaan ajaa komponentin koodi. Komponentin sisältö on esitetty esimerkissä 1.

Esimerkki 1. Työn ensimmäinen koodikomponentti ilman piilotettua koodia. Näkyvissä olevilla komennoilla haetaan tarvittavat moduulit ja käytettävä data-tiedosto, sekä tulostetaan tiedostossa olevien törmäystapahtumien lukumäärä.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')

print('Törmäystapahtumien lukumäärä = % d' % len(ds))
```

Kuten alaluvussa 4.1 mainitsin, sisällytin komponentteihin myös käyttäjältä piilossa olevaa koodia. Piilossa olevaan koodiin kirjoitin komentoja, joita opiskelijan ei tarvitse huomioida. Esimerkin 1 koodikomponentin sisältö kokonaisuudessaan piilo-

tetun koodin kanssa on esitetty esimerkissä 2.

Esimerkki 2. Työn ensimmäinen koodikomponentti piilotetun koodin kanssa.

```
''' {plugin=''csPlugin'' #p11}
type: py
filename: eka
path: user
showCodeOn:
fullprogram: !!!
#
// BYCODEBEGIN
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')

print('Törmäystapahtumien lukumäärä = % d' % len(ds))
// BYCODEEND
ds = ds.drop('M', 1)
#
!!
'''
```

Tässä ennen `// BYCODEBEGIN` -komentoa piilossa on ainoastaan koodikomponentin luomiseen TIMissä tarvittavia komentoja. Komennoilla mm. määritetään komponentin kieleksi Python sekä annetaan sille nimi. Nimeämisen avulla tämä ensimmäinen komponentti voidaan ajaa käyttäjältä piilossa seuraavia komponentteja ajettaessa. Varsinainen ohjelma alkaa aina komennosta `fullprogram: !!!`.

Komentoa `// BYCODEEND` seuraava piilossa oleva rivi poistaa haetusta CSV-tiedostosta sarakkeen *M*, joka sisältää jo valmiiksi datatiedostoon laskettuna myoniparien invariantit massat. Sarakkeen poistamisella halusin estää tilanteen, jossa käyttäjä saisi massat suoraan tiedostosta, sillä työssä tavoitteena on laskea massat itse.

Korostin ensimmäisen koodikomponentin yhteydessä, että muuttujan `ds` nimeä ei saa vaihtaa (LIITE A-9). Mikäli näin tekee, niin mm. esimerkin 2 lopussa oleva sarakkeen *M* poistaminen ei toimi, sillä se olettaa muuttujan nimen olevan `ds`.

Käyttäjälle näkyvissä olevassa osassa `import` -alkuiset rivit hakevat käyttöön työssä tarvittavat kolme Python-moduulia. Moduulit ovat kooditiedostoja, jotka sisältävät erilaisia valmiiksi laadittuja funktioita. Moduulien avulla käyttäjän ei siis tarvitse itse luoda kaikkia tarvitsemiaan funktioita. Työssä käytetyistä moduuleista *pandas* sisältää data-analyysissä tarvittavia funktioita, *numpy* numeerisen laskennan funktioita ja *matplotlib.pyplot* kuvaajien luomisessa tarvittavia funktioita.

Rivillä `ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')` haetaan työssä

käytettävä CSV-tiedosto *pandas*-moduulin avulla. Tämän jälkeen vielä tulostetaan komenolla `print()` tiedostossa olevien törmäystapahtumien lukumäärä. Tulostuksella halusin korostaa opiskelijalle sitä, että tiedostossa todella on runsas määrä dataa (yhteensä 475 465 riviä).

Työn ensimmäisessä versiossa olin valinnut muuttujan `ds` nimeksi `datasetti`. CERNin testauksissa tekemieni havaintojen perusteella päädyin lyhentämään nimen. Ainakin yksi työtä CERNissä testanneista pareista koki pitkän nimen kirjoittamisen turhana toistona.

Esimerkissä 1 kuvatun koodikomponentin suunnittelun yhteydessä törmäsin tekniiseen haasteeseen. Yritin suunnitteluvaiheessa samassa komponentissa ladata CSV-tiedoston ja tarkastella sen sisältöä, mutta sain TIMissä virheilmoituksen ”Runtime exceeded, maybe loop forever.” Virheilmoitus tarkoitti, että koodin ajaminen oli kestänyt kauemmin kuin TIMille asetettu suurin laskenta-aika sallii. Kokeilin myös jakaa koodia eri komponentteihin, mutta sain myös tällä tapaa saman virheilmoituksen. Tiedostokooltaan pienempien CSV-tiedostojen käyttäminen toimi ilman virheilmoitusta, joten ajattelin ongelman johtuvan liian suuresta tiedostokoosta. Kysyin ongelmaan apua TIMin kehittäjätiimiltä. He ratkaisivat tilanteen lataamalla työssä käytössä olevan CSV-tiedoston apuhakemistoon, josta TIM lukee tiedoston nopeammin. Tämän jälkeen tiedoston ajaminen toimi hyvin.

Datan hakemista seuraavaksi osioksi loin tyhjän koodikomponentin, johon opiskelijan tulee itse kirjoittaa koodi, joka tulostaa haetun tiedoston viisi ensimmäistä riviä (LIITE A-10). Ohjeistin tarvittavat komennot osion yhteydessä. Tulostuksen avulla opiskelijalle tulee näkyviin, mitä eri arvoja datatiedoston rivit sisältävät. Näitä arvoja tullaan käyttämään työn seuraavassa vaiheessa suoritettavassa laskussa. Tällä osiolla sain myös itse varmistettua, että koodikomponenttien ketjutus toimii. Mikäli ketjutus toimii, eli myöhemmät komponentit ajavat aina kaikki edelliset, tulostuu myös tapahtumien lukumäärä aina seuraavissa komponenteissa, sillä ensimmäinen komponentti tekee tulostamisen. Tämä oli suunnitteluvaiheessa hyvä testi työn toimivuudesta.

Työn rakennetta luodessa tajusin, että kehittämisessä olisi hyvä varmistaa, ettei käyttäjä voi jäädä täysin jumiin mahdollisten ohjelmointiongelmiensa kanssa. Keksinkin, että työhön voisi lisätä vihjeitä, jotka auttavat tarvittaessa eteenpäin. Löysin TIMistä ominaisuuden, jolla dokumenttiin voi luoda klikattaessa aukeavia tekstiosioita. Tätä ominaisuutta hyödyntäen loin työhön ensimmäisen vinkin (LIITE A-10). Vinkkiä klikatessa esiin tulee teksti ”Vinkki: `print(tiedostonimi.head())`”, joka on tiedostonimeä lukuun ottamatta valmis koodi edellä mainitun osion komponenttiin. Korostin vihjeen yhteydessä, että käyttäjän tulisi avata vihje vasta itse yrittämisen jälkeen. Liitteenä olevassa työn PDF-vedoksessa vinkki näkyy jo valmiiksi avattuna ja ilman graafista painiketta.

Tein vastaavanlaiset vihjeet muihinkin työn koodikomponentteihin, joissa opiskelijan tulee itse kirjoittaa koodia. Komponentista riippuen loin 1–4 asteittain enemmän ratkaisua paljastavaa vihjettä. Tällä halusin varmistaa työn sujuvuuden. Työtä

tehdessä opiskelija ei voi saada välittömästi apua, sillä työ tehdään ilman ohjaajan avustusta. Myös muut vinkit näkyvät aiemmin mainitulla tavalla liitteenä olevassa PDF-vedoksessa eri tavalla kuin oikeassa työssä.

4.2.3 Invariantin massan laskeminen

Työn seuraavaksi osioksi tein koodikomponentin, jossa opiskelijan tulee laskea haetun datatiedoston arvojen avulla jokaisen törmäystapahtuman myoniparille invariantin massan arvo (LIITE A-10). Kirjoitin komponenttiin valmiiksi muuttujan nimen sekä komennot, joilla koodi tulostaa viisi ensimmäistä laskettua arvoa. Tulostuksen avulla pyrin tuomaan opiskelijalle tarkemmin ilmi sen, mitä tilanteessa ollaan laskemassa. Tämä komponentti on esitetty kokonaisuudessaan, piilotettu koodi mukaan lukien, esimerkissä 3. Esimerkissä käyrillä nuolilla (\leftrightarrow) viitataan siihen, että koodissa rivi jatkuisi katkeamatta.

Esimerkki 3. Invariantin massan laskemisen koodikomponentti piilotetun koodin kanssa.

```
''' {plugin="csPlugin" #p13}
type: py
filename: kolmas
path: user
showCodeOn:
fullprogram: !!!
#
file = 'eka.py'
exec(compile(open(file).read(), file, 'exec'))
// BYCODEBEGIN
invariantti_massa =
print('Viisi ensimmäistä laskemaasi arvoa:')
print(invariantti_massa[0:5])
// BYCODEEND
varmistus = np.sqrt(2*ds.pt1*ds.pt2*(np.cosh(ds.eta1-ds.eta2) - np.
    ↪ cos(ds.phi1-ds.phi2)))
if invariantti_massa.values[8] == varmistus.values[8]:
    print('Invariantit massat laskettu oikein!')
else:
    print('Tarkasta vielä lasku, yllä olevat arvot eivät ole
    ↪ oikein. Jos jäät jumiin, katso vinkki.')
#
!!
'''
```

Käyttäjälle näkyvä osuus on taas komentojen `// BYCODEBEGIN` ja `// BYCODEEND` välissä. Piilossa ylhäällä komennon `fullprogram: !!!` jälkeen on koodia, joka ajaa aiemmat koodikomponentit. Käyttäjälle näkyvässä osiossa on kirjoitettuna valmiiksi muuttujan nimi sekä kaksi tulostuskomentoa. Opiskelijan tehtäväksi jää määritel-

lä muuttuja `invariantti_massa` siten, että siihen tallentuu jokaiselle datatiedoston myöniparille laskettu invariantti massa. Invariantin massan laskeminen onnistuu yleisesti monella erilaisella lausekkeella. Työssä käytettäväksi valitsin muodon, jolla lasku onnistuu suoraan datatiedostossa olevien arvojen avulla ilman välituloksia.

Alhaalle käyttäjältä piiloon päätin kirjoittaa koodia, joka kertoo ovatko invariantit massat laskettu oikein. Toteutin vertailun luomalla muuttujan `varmistus`, johon tallensin oikealla tapaa lasketut invariantin massan arvot. Tämän jälkeen koodi vertaa if-else -rakenteella käyttäjän laskemien massojen kahdeksatta arvoa muuttujan `varmistus` kahdeksanteen arvoon. Mikäli arvot täsmäävät, saa käyttäjä viestin laskun onnistumisesta. Mikäli taas arvot eivät ole samat, käyttäjää kehoitetaan tarkastamaan lasku ja katsomaan tarvittaessa vinkkejä.

Toteutustapa arvoja vertailemalla ei ole täysin riskitön. Periaatteessa voisi olla mahdollista, että opiskelija onnistuisi laskemaan arvot oikein, mutta laskutavasta tai muusta tekijästä johtuen arvojen jokin merkityksetön desimaali voisi erota muuttujan `varmistus` arvosta. Tällöin if-else -vertailu ilmoittaisi arvojen eroavan ja kehoittaisi opiskelijaa tarkastamaan laskun. Tätä ongelmaa ei kuitenkaan ilmennyt missään kohti työn luomista ja testaamista.

Esimerkin 3 laatimisen jälkeen aloitin työn siihen asti valmistuneen osuuden testaamisen. Testaamista varten loin TIMiin sähköpostilla alaluvussa 3.4 mainitun TIM-tunnuksen. Tällä yliopiston tunnuksestani erillisellä tunnuksella pystyin simuloimaan työtä tekevää opiskelijaa, jolla on työhön vain käyttöoikeudet.

Testitunnuksilla huomasin, että koodikomponentit toimivat kuten pitikin. Käyttäjälle näkyi ainoastaan komponenttien valmis koodi eikä esimerkiksi tekemiäni testiajoja. Lisäksi ylläpitäjän nähtäväksi tallentuivat testitunnuksilla koodikomponenteissa tehdyt ajot. Tämän ominaisuuden avulla ylläpitäjä voi siis nähdä, minkälaisia koodin ajoja opiskelija on yrittänyt ennen varsinaista onnistunutta koodia.

Testin aikana tuli ilmi, että ylläpitäjä ei näe, onko opiskelija avannut vinkkejä vai ei. Jotta vinkeistä olisi jäänyt jälki, olisi ne pitänyt toteuttaa klikkaamalla avautuvan tekstin sijaan ohjelmointikoodilla, sillä koodin ajaminen jättäisi ylläpitäjälle jäljen. Testatessa huomasin myös, että koodikomponenttien alla oli *Näytä koko koodi* -painike, jota klikkaamalla TIM nosti esiin myös piilotetuiksi tarkoittamani osiot. Tämä olisi työn kannalta huono ominaisuus, sillä tällöin käyttäjä voisi esimerkiksi kopioida invariantin massan laskemiseen tarvittavan lausekkeen suoraan esimerkin 3 yhteydessä mainitsemaani varmistusosiosta. Päätin siis poistaa painikkeen käyttäjiltä. Tämä onnistui antamalla painikkeen nimeksi tyhjän tekstikentän komennolla `showCodeOn:`, jonka seurauksena painike katosi.

4.2.4 Histogrammin laatiminen

Välissä tapahtuneen testaamisen jälkeen siirryin työstämään osiota, jossa luodaan histogrammi lasketuista invariantin massan arvoista (LIITE A-11). Aloitin tämän työvaiheen pohtimalla parasta toteutustapaa histogrammin tuottamiseen Python-kielellä. Kokeilin aluksi monirivistä koodia, mutta totesin sen liian monimutkaiseksi kirjoitettavaksi opiskelijoille. Aloin siis etsiä, olisiko histogrammi mahdollista luoda yksinkertaisemmalla komennolla. Päädyin käyttämään yhdellä rivillä toimivaa *matplotlib.pyplot*-moduulin `plt.hist()` -komentoa. Totesin komennon niin yksinkertaiseksi, että opiskelijan on mahdollista tuottaa se ohjeiden avulla itse.

Histogrammin laatiminen -osion alkuun kirjoitin tekstiä siitä, mitä histogrammilla tarkoitetaan, ja miten histogrammi luodaan `plt.hist()` -komennolla (LIITE A-11). Sisällytin tekstiin myös linkin *matplotlib.pyplot*-moduulin dokumentaatioon¹², joka sisältää lisää tietoa histogrammikomennosta. Linkkiä laatiessani huomasin, että TIM ei avaa linkkejä automaattisesti uuteen välilehteen, vaan siirtyy TIM-dokumentista linkin osoitteeseen. Tämä ominaisuus ei olisi työn kannalta hyvä, sillä dokumentista poistuttaessa tallentamatonta tietoa tai ajamatonta koodia voisi kadota. Päätin liittää linkin mukaan pelkkänä tekstinä sekä huomauttaa, että linkki kannattaa avata uuteen välilehteen.

Lisäsin mukaan myös huomautuksen siitä, että virheellisen koodin ajaminen voi tuottaa varsinaisen virheilmoituksen lisäksi myös ylimääräisen ”*No File: ...test.png*” -ilmoituksen. Ilmoitus johtuu siitä, että virheellisen koodin ajaminen pysähtyy ennen loppua, jolloin lopussa oleva komento, jolla png-muotoinen kuvaaja tallennetaan, jää ajamatta. Tämän virheen saaminen voisi johtaa siihen, että opiskelija ryhtyisi-kin varsinaisen virheen sijaan pohtimaan ylimääräisen virheilmoituksen korjaamista. Tätä huomautusta ei ollut työn ensimmäisessä versiossa, vaan lisäsin sen mukaan CERNissä tehtyjen testien havainnoinnin perusteella.

Osion alustamisen jälkeen loin esimerkissä 4 kuvatun koodikomponentin. Käyttäjältä piilotetussa koodissa ennen komentoa `# BYCODEBEGIN` ja komennon `# BYCODEEND` jälkeen alustetaan kuvaajalle tiedosto komennolla `imgsource: test2.png` sekä ajetaan kaikki edeltävät koodikomponentit. Käyttäjälle näkyvässä osassa taas kehoitetaan kommentissa kirjoittamaan alle tarvittava komento histogrammin tuottamista varten. Lisäksi annetaan valmiiksi komennot histogrammin akseleiden ja otsikon nimeämistä varten.

Esimerkki 4. Histogrammin laatimisen koodikomponentti.

```
''' {plugin="csPlugin" #simpplottti}
path: user
filename: plotti2
imgsource: test2.png
type: py
```

¹²https://matplotlib.org/devdocs/api/pyplot_summary.html

```

showCodeOn:
fullprogram: |!!
file = 'kolmas.py'
exec(compile(open(file).read(), file, 'exec'))
# BYCODEBEGIN
# Kirjoita tähän alle komento, jolla saat luotua histogrammin.

# Nimetään akselit ja otsikko. Älä muuta näitä.
plt.xlabel('Invariantti_massa [GeV]')
plt.ylabel('Tapahtumien_lukumäärä')
plt.title('Kahden_myonin_invariantin_massan_histogrammi\n')
# BYCODEEND
plt.savefig('test2.png')
!!
'''

```

Jotta käyttäjä saa luotua histogrammin, tulee komponenttiin lisätä muotoa `plt.hist(x, bins=0, range=(0,0))` oleva komento. Komennossa x korvataan sen muuttujan nimellä, jossa histogrammiin kuvattava data sijaitsee. Kohta *bins* kuvaa histogrammin jakovälien lukumäärää ja *range* tarkasteluvälin ylä- ja alarajoja. Osion alustavassa tekstissä on tarjottu tarkasteluväliksi esimerkiksi väliä (60,120) sekä kehoitettu käyttäjää kokeilemaan erilaisia jakovälien lukumääriä.

Halusin histogrammin laatimisen yhteyteen myös osion, jossa opiskelija voisi tarkastaa, onko luonut histogramminsa oikein. Toteutin tarkastamisen klikkaamalla avattavalla mallihistogrammilla (LIITE A-12). Malliin vertailemalla opiskelija voi todeta, onko hänen histogramminsa kunnossa, vai vaatiiko se vielä muutoksia. Jätin mallihistogrammista tarkoituksellisesti akselien lukuarvot pois. Tällä halusin varmistaa, ettei työtä voisi tehdä loppuun mallista otetuilla arvoilla, vaan histogrammi tulisi laatia itse. Liitteenä olevassa PDF-vedoksessa mallihistogrammi on jo valmiiksi avattuna, kun oikeassa työssä käyttäjän tulee klikata histogrammi auki.

Osion loppuun lisäsin vielä sanallisen tehtävän (LIITE A-12). Tehtävässä pyydetään kuvailemaan histogrammia ja kertomaan, mitä tietoa histogrammista voi saada. Tällä halusin herättää työn tekijää pohtimaan histogrammin merkitystä ja välttää tilanteen, jossa opiskelija vain toteuttaisi pyydettyjä komentoja pohtimatta niiden avulla saatuja tuloksia.

4.2.5 Funktion sovittaminen histogrammiin

Histogrammiosion jälkeen siirryin suunnittelemaan sitä, miten luotua histogrammia analysoitaisiin. Olimme puhuneet Kati Lassila-Perinin kanssa, että histogrammin lisäksi työssä olisi hyvä saada konkreettisia perusteita Z -bosonin esiintymisen todentamiseen. Tällöin analyysi ei jäisi ainoastaan visuaalisten havaintojen tasolle. Päätimme, että histogrammiin voisi sovittaa funktion, jonka kautta saataisiin ulos konkreettisia lukuarvoja. Ensimmäisessä kehittämisvaiheessa valitsin sovitettavaksi funktioksi normaalijakauman, sillä invarianttien massojen arvot noudattavat työssä melko hyvin normaalijakaumaa.

Loin siis työn seuraavaksi vaiheeksi osion, jossa käyttäjä sovittaa luomaansa histogrammiin normaalijakauman, jonka avulla saadaan selville tilanteen keskihajonta ja odotusarvo (LIITE A-13). Odotusarvo kuvaa jakauman todennäköisyysmassan painopistettä. Keskihajonta taas kuvaa keskimääräistä poikkeamaa odotusarvosta.

Aloitin osion työstämisen kokeilemalla erilaisia tapoja sovittaa funktio histogrammiin. Tässä havaitsin, että osa sovitustavoista ei ota huomioon histogrammia luotaessa *range*-komennolla tehtyä rajausta Z -bosonin piikin lähetyville. Tämä johti siihen, että myös jakauman alku- ja loppupään invarianttien massojen arvot vaikuttivat sovitukseen, ja sovitus ei osunut hyvin histogrammin piikin kohdalle.

Edellä mainittu ongelma ratkesi *scipy.optimize* -moduulin¹³ *curve_fit* -optimointikomennon avulla. Komento optimoi sille annetuilla alkuarvoilla haluttuun dataan halutunlaisen funktion. Lisäksi päätin, että funktiota ei soviteta suoraan edellisessä osiossa laadittuun histogrammiin. Tähän valintaan päädyin siksi, koska sovitusta varten histogrammin tulisi olla normitettu ja histogrammiin tulisi valita vain piikin lähellä olevia invariantin massan arvoja. Näiden ehtojen toteuttaminen alaluvussa 4.2.4 kuvailussa osiossa olisi ollut mielestäni liian monimutkainen opiskelijan laadittavaksi.

Sovitusosion alkuun lisäsin tietoa normaalijakaumasta, osion tavoitteesta sekä matemaattisten kommentojen kirjoittamisesta Pythonilla (LIITE A-13). Tämän jälkeen loin esimerkissä 5 esitetyn koodikomponentin.

Esimerkki 5. Normaalijakauman sovittamisen koodikomponentti.

```

''' {plugin="csPlugin" #fittaus}
path: user
filename: fittaus
imgsource: test4.png
type: py
showCodeOn:
fullprogram: !!!

file = 'kolmas.py'
exec(compile(open(file).read(), file, 'exec'))

```

¹³<https://docs.scipy.org/doc/scipy/reference/optimize.html>

```

# BYCODEBEGIN
# Rajaudutaan vain histogrammin piikin lähelle, jossa
# histogrammi on normaalijakautunut.
alaraja = 88
ylaraja = 94
binit = 100

# Valitaan rajojen sisällä olevien massojen arvot.
rajatutmassat = invariantti_massa[(invariantti_massa > alaraja) & (
    ↪ invariantti_massa < ylaraja)]

# Luodaan valituista normitettu histogrammi normaalijakaumaa varten
    ↪ .
histogrammi = plt.hist(rajatutmassat, bins=binit, range=(alaraja,
    ↪ ylaraja), normed=True)

# Y-akselille tapahtumien määrä kullakin jakovälillä (saadaan
    ↪ muuttujasta histogrammi),
# x-akselille jakovälit luomalla lista, jossa binien määrä
    ↪ numeroita tasavälein
# histogrammin rajauksen matkalta.
y = histogrammi[0]
x = np.linspace(alaraja,ylaraja,binit)

# Määritellään, että halutaan sovittaa normaalijakauman funktio.
def gaussinen(x, sigma, mu):
    return #POISTA TÄMÄ KOMMENTTI JA KIRJOITA TÄHÄN
        ↪ NORMAALIJAKAUMAN LAUSEKE

# Alkuarvot optimointia varten. Ensimmäinen on keskihajonta (sigma)
# ja toinen odotusarvo (mu)
alkuarvot = [#TÄHÄN ALKUARVO SIGMALLE , #TÄHÄN ALKUARVO MY:LLE ]

# Haetaan optimointiin käytettävä moduuli ja tulostetaan tehty
    ↪ sovitus
# sekä siitä saadut keskihajonnan ja odotusarvon arvot.
from scipy.optimize import curve_fit
parhaat, kovarianssi = curve_fit(gaussinen, x, y, p0=alkuarvot)
print("Keskihajonnan_ (sigma)_ arvo_ =_ %f" % parhaat[0])
print("Odotusarvon_ (mu)_ arvo_ =_ %f" % parhaat[1])
plt.plot(x, gaussinen(x, *parhaat), 'r-', label='sigma_ =_ {},_ mu_ =_
    ↪ {}'.format(parhaat[0], parhaat[1]))
plt.xlabel('Invariantti_massa_[GeV]')
plt.ylabel('Normitettu_tapahtumien_lkm')
plt.title('Normaalijakauman_ sovitus')
plt.legend()

# BYCODEEND
plt.savefig('test4.png')
!!
'''

```

Tässä komponentissa piilotin käyttäjältä ainostaan edelliset komponentit ajavan sekä kuvaajan alustuksen ja tallentamisen tekevät komennot. Koko muu `#BYCODEBEGIN` ja `#BYCODEEND` -komentojen välissä oleva koodi on valmiina käyttäjän nähtävillä.

Näkyvissä olevan koodin alussa valitaan muuttujilla `alaraja` ja `ylaraja` vain histogrammin piikin lähellä olevat invariantin massan arvot. Muuttajalla `binit` taas asetetaan jakovälien määrä tilanteeseen sopivaksi. Opiskelijan ei siis tarvitse pohtia näille sopivia arvoja, vaan ne ovat jo annettuna.

Rajojen ja jakovälien valinnan jälkeen lisäsin kaksi komentoa, joista toinen valitsee mukaan vain rajojen määräämät arvot ja toinen luo näistä arvoista normitetun histogrammin. Lisäksi muotoilin luodusta histogrammista saatavat arvot sovituskomennolle sopivaan muotoon luomalla muuttujat `x` ja `y` sekä määritin valmiiksi rungon normaalijakauman funktion luomista varten. Annoin funktiolle nimen `gaussinen`.

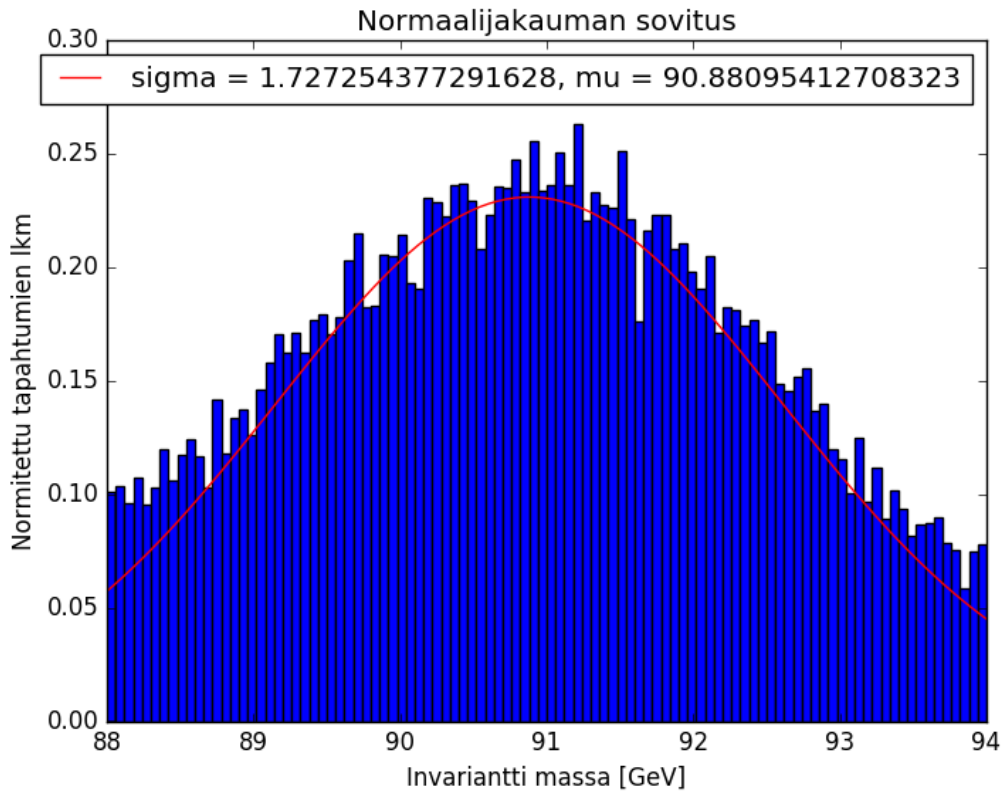
Opiskelijan tehtäväksi jää siis kirjata funktion `gaussinen` määrittelyyn normaalijakauman lauseke. Kirjoitin tarvittavan lausekkeen sekä vinkkejä osion alustukseen. Opiskelijan tulee myös valita tilanteeseen sopivat alkuarvot keskihajonnalle σ ja odotusarvolle μ . Näitä varten loin valmiiksi muuttujan `alkuarvot`, johon käyttäjä voi täydentää valitsemansa arvot.

Alkuarvojen päättely onnistuu työn edellisessä osiossa (LIITE A-12) laaditun histogrammin avulla. Arvojen valinnalla ei kuitenkaan ole suurta merkitystä, sillä käytettävä optimointikommento ei ole kovin tarkka annetuista alkuarvoista. Riittää, että arvot ovat suurin piirtein samaa suuruusluokkaa kuin sovitetussa funktiossa ne tulevat olemaan.

Loppuosaan koodikomponentin näkyvää osiota kirjoitin komentoja, joilla haetaan sovitusta varten tarvittava moduuli, tehdään sovitus sekä luodaan histogrammi. Lisäksi komennolla nimetään histogrammin akselit sekä tulostetaan sovituksesta saadut lukuarvot keskihajonnalle σ ja odotusarvolle μ . Nämä kaikki annetaan käyttäjälle valmiina.

Esimerkin 5 koodikomponentti tuottaa siis edellä kuvattujen komentojen avulla normitetun histogrammin invarianttien massojen arvoista. Lisäksi histogrammiin tulostuu siihen sovitetun normaalijakauman kuvaaja tunnuslukuineen. Tuotettava histogrammi on esitetty kuviossa 12.

Sisällytin koodikomponentin yhteyteen neljä vinkkiä, jotka antavat asteittain enemmän apua sovituksen laatimiseen (LIITE A-14). Ensimmäinen vinkki kehottaa kiinnittämään huomiota sulkujen oikeaoppiseen käyttöön. Päätin korostaa tätä, sillä syötettävä normaalijakauman lauseke sisältää monia sulkuja lausekkeessa esiintyvien potenssien ja murtolukujen takia. Toinen ja kolmas vinkki antavat apua keskihajonnan ja odotusarvon alkuarvojen määrittelyyn. Neljäntenä vinkkinä päätin antaa valmiin, toimivan normaalijakauman lausekkeen. Tällä pyrin välttämään tilannetta, jossa työn tekijä ei pääsis työssä eteenpäin esimerkiksi lausekkeessa olevan pienen näppäilyvirheen takia, sillä syötettävä lauseke on suhteellisen pitkä.



Kuvio 12. Histogrammiin luotu normaalijakauman sovitus. Arvot σ ja μ ovat jakauman keskihajonta ja odotusarvo.

Lisäksi kirjoitin koodikomponentin jälkeen huomion siitä, että käyttäjän tulisi varmistua, että sovitus näyttää normaalijakaumalta (LIITE A-14). Tällä pyrin vähentämään mahdollisuutta tilanteeseen, jossa opiskelija jatkaisi työssä eteenpäin virheellisellä sovituksella, ja siten myös virheellisillä keskihajonnan ja odotusarvon arvoilla. Koska työ tehdään ilman aktiivista ohjausta, virheelliset arvot selviäisivät pahimmillaan vasta työn palautuksen jälkeen. Tämän huomion lisäsin työhön vasta CERNissä tehtyjen testausten jälkeen.

Edellä kuvaillun osion tekemisen aikana tapahtui yhden kerran tilanne, jossa työ lakasi toimimasta. TIMissä suoritettua huoltokatkoa jälkeen huomasin, että työssä käytettävää CSV-tiedostoa ei enää löytynyt koodikomponenteissa. Sain komponentteja ajettaessa virheilmoituksen `"OSError: File b'/cs/data/DoubleMuRun2011A.csv' does not exist."` Olin ongelmasta yhteydessä TIMin kehittäjätiimiin, jolta sain heti vastauksen. Huoltokatkoa yhteydessä CSV-tiedoston sisältänyt apuhakemisto oli unohtunut kopioida paikalleen. Kopioinnin jälkeen tiedoston hakeminen toimi taas normaalisti.

4.2.6 Histogrammin analysointi

Histogrammiin tehtyjen sovitusten jälkeen päätin luoda työhön osion, jossa saatuja tuloksia analysoidaan (LIITE A-15). Tulosten analysointia varten laadin kuuden sanallisen kysymyksen sarjan. Tein kysymykset TIMin valmiilla kysymyspohjalla. Tällä tapaa opiskelijoiden sanalliset vastaukset tallentuvat työn tarkastajan nähtäväksi.

Ensimmäiseksi kysymykseksi kirjoitin seuraavan tekstin: *”Mitä voit sanoa Z-bosonin esiintymisestä histogrammin ja siihen sovitetun funktion pohjalta? Perustelee vastauksesi teoriaosaa ja muita lähteitä apunasi käyttäen.”* Tällä pyrin saamaan vastauksen työn tärkeimpään tavoitteeseen, Z-bosonin esiintymisen vahvistamiseen. Opiskelija voi todeta histogrammin perusteella, että myoniparien invarianttien massojen arvot näyttäisivät piikittyvän vahvasti noin 91 GeV:n kohdalle. Tarkemman lukuarvon opiskelija saa normaalijakauman sovituksesta saadulla odotusarvolla. Tämän ja teoriaosan perusteella opiskelija voi todeta, että tapahtumissa on havaittavissa hiukkanen, jonka massa on noin 91 GeV. Tämä vastaa hyvin teoriaosassa annettua Z-bosonin massan teoreettista ennustetta.

Toisen kysymyksen kirjoitin muotoon *”Mikä on tulostesi perusteella Z-bosonin massa? Miten selvität sen?”* Tällä halusin, että opiskelija yhdistää analyysistä saadun odotusarvon Z-bosonin massaan. Lisäksi tavoittelin perustelua massan selvittämiseen.

Seuraavaan kysymykseen sisällytin saadun tuloksen vertailua työn ulkopuolisen lähteen teoreettiseen Z-bosonin massan arvoon. Muotoilin kysymyksen seuraavasti: *”Avaa uuteen välilehteen Particle Data Groupin verkkosivut (<http://pdg.lbl.gov/>). Etsi sieltä arvo Z-bosonin massalle ja vertaa sitä saamaasi arvoon. Mitä havaitset?”* Päätin ottaa mukaan tämän kysymyksen siksi, koska halusin tuoda opiskelijalle vertailukohtaa työssä saadulle tulokselle. Lisäksi kysymys esittelee opiskelijalle Particle Data Group -sivuston, joka sisältää runsaasti erilaisia tuloksia hiukkasfyysiikan alalta.

Kysymyssarjan neljänneksi kysymykseksi kirjasin seuraavan: *”Milloin Z-bosoni havaittiin ensimmäisen kerran ja mikä Z:n massan fysikaalinen merkitys on?”* Tämän kysymyksen tarkoituksena on saada opiskelija perehtymään analyysin kohteena olevan Z-bosonin taustoihin. Pyrin luomaan kysymyksellä merkitystä tehdyille analyysille.

Työn aivan ensimmäiseen versioon sisällytin vain nämä neljä ensimmäistä kysymystä. CERNissä tehtyjen testausten ja ohjaaja Kati Lassila-Perinin kommenttien jälkeen päätin lisätä mukaan vielä kaksi kysymystä, joissa käsitellään virhetarkastelua ja epätarkkuuksia.

Sarjan toiseksi viimeisessä kysymyksessä tutustutaan analyysissa syntyneen histogrammin piikin leveyteen. Kirjoitin kysymyksen muotoon *”Miksi saamasi histogrammin leveys on suurempi kuin Particle Data Groupin Z-bosonille ilmoittama leveys*

$2,4952 \pm 0,0023 \text{ GeV?}$ ” Tarkoitukseni oli tällä kysymyksellä ottaa työhön mukaan myös virhetarkastelua ja epätarkkuuksien arviointia, sillä analyysin teossa ei erikseen huomioida mittausepä-tarkkuuksia. Ideana on, että opiskelija hahmottaa eron teoreettisen piikin leveyden ja kokeellisella analyysillä saadun leveyden välillä. Kokeellisessa analyysissä piikki on leveämpi siksi, koska energioilla ja liikemäärillä on tietty mittatarkkuus eli resoluutio. Ja koska invariantti massa lasketaan mitattujen energioiden ja liikemäärien avulla, vaikuttaa resoluutio saatuun tulokseen.

Osion viimeiseksi kysymykseksi valitsin seuraavan: ”*Jos energia ja liikemäärä saataisiin mitattua äärettömän tarkasti, saataisiinko Z :n massan kohdalle histogrammiin yksi täsmällinen piikki vai jakauma? Perustelee valintasi.*” Idea tämän aiheen käsittelemiseen työssä tuli Kati Lassila-Periniltä. Kysymyksen tarkoitus on tuoda ilmi, että ideaalisessakin mittaustilanteessa piikillä on leveys, sillä tämä on täysin fyysikaalinen ominaisuus. Tämä selittyy alaluvussa 3.2.3 käsitellyn aika-energia-epätarkkuusperiaatteen avulla. Sen mukaan mitä lyhyempi hiukkasen elinaika on, sitä epätarkempi on hiukkasen massa.

4.2.7 Koko datatiedoston histogrammi

Työn loppuun päätin vielä sisällyttää osion, jossa luodaan histogrammi kaikista käytetyn datatiedoston invarianttien massojen arvoista rajautumatta Z -bosonin piikin kohdalle (LIITE A-17). Tämän osion ideana on havainnollistaa opiskelijalle sitä, että käytetystä datasta löytyy analyysin kohteena olleen Z :n lisäksi myös muita hiukkasia. Tämä voisi jäädä muuten ymmärtämättä, sillä itse analyysissä rajauduttiin ainoastaan Z :n piikin lähetyville.

Datatiedoston kaikkien invarianttien massojen histogrammin luomista varten laadin esimerkissä 6 esitetyn koodikomponentin. Käyttäjältä piilossa olevassa osiossa ennen `#BYCODEBEGIN` -komentoa koodi ajaa esimerkin 3 koodikomponentin sekä ketjutuksen takia myös sitä edeltävät komponentit. Tällä saadaan käyttöön histogrammiin tulevat invarianttien massojen arvot. Lopussa, `#BYCODEEND` -komennon jälkeen käyttäjältä piilossa tapahtuu komponentissa luodun histogrammin tallentaminen.

Esimerkki 6. Koko datatiedoston histogrammin luomisen koodikomponentti.

```
''' {plugin="csPlugin" #kokoplotti}
path: user
filename: plotti3
imgsource: test3.png
type: py
showCodeOn:
fullprogram: !!!
file = 'kolmas.py'
exec(compile(open(file).read(), file, 'exec'))
# BYCODEBEGIN
binien_lkm = 500
# Lasketaan massoista logaritmit ja painotus.
```

```

inv_massa_log = np.log10(invariantti_massa)
painot = []
for a in invariantti_massa:
    painot.append(binien_lkm/np.log(10)/a)

# Plotataan painotettu histogrammi.
plt.hist(inv_massa_log, binien_lkm, range=(0,2.5), weights=painot,
        ↪ lw=0, color="darkgrey")
plt.yscale('log')

# Akselien ja otsikon nimeäminen.
plt.xlabel('log10(invariantti_massa) [log10(GeV)]')
plt.ylabel('Tapahtumien_lukumäärä')
plt.title('Kahden myonin invariantin massan histogrammi\n')
# BYCODEEND
plt.savefig('test3.png')
!!
'''

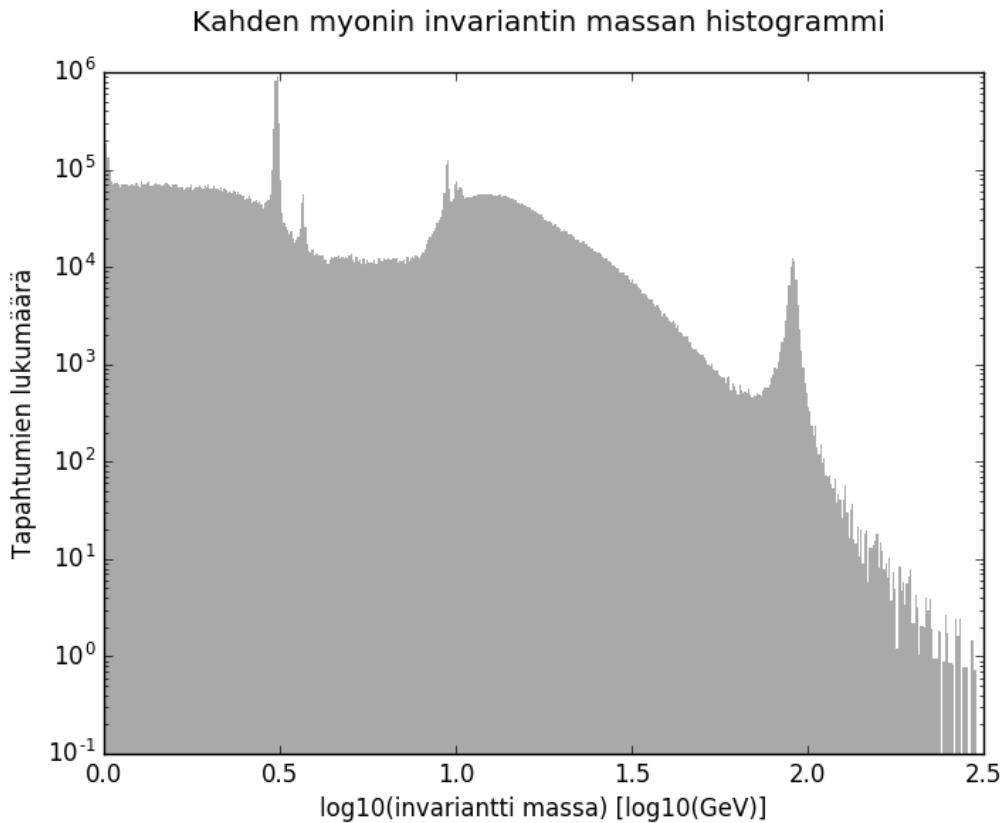
```

Käyttäjälle näkyvään koodiin loin aluksi muuttujan `binien_lkm`, johon tallensin riittävän suuren jakovälien määrän. Valitsin suuremman jakovälien määrän kuin aiemmissa histogrammeissa, sillä nyt mukana oli runsaasti laajempi määrä invarianttien massojen arvoja. Tällä tavalla histogrammi ja siihen syntyvät piikit piirtyvät siistimmin.

Jakovälien määrittämisen jälkeen kirjoitin komentoja, joilla luodaan massojen arvoista kymmenkantaiset logaritmit sekä niiden avulla painotukset histogrammia varten. Histogrammin luominen toteutetaan jälleen `plt.hist()` -komennolla, mutta tällä kertaa painotuksilla ja logartimisilla akseleilla. Komponentin loppuun kirjoitin komennot, joilla nimetään histogrammin akselit ja otsikko. Lisäksi käyttäjältä piiloon lisäsin koodirivin, joka tallentaa luodun histogrammin kuvana.

Yllä kuvattujen valintojen avulla histogrammista tulee samankaltainen, kuin osioon liittämäni CMS-kokeen julkaisema histogrammi (LIITE A-17). Tämä mahdollistaa histogrammien vertailun keskenään. Työssä luotava histogrammi on esitetty kuviossa 13.

Vertailua varten lisäsin osioon työn viimeisen sanallisen kysymyksen: *”Vertaa luomaasi histogrammia alla olevassa kuvassa 10 näkyvään CMS-kokeen julkaisemaan histogrammiin. Mitä muita hiukkasia Z-bosonin lisäksi havaitset?”* Kysymyksellä pyrin saamaan opiskelijan hahmottamaan sitä, että käytetyssä datatiedostossa on havaittavissa myös muita massojen piikittymisiä, eli käytännössä hiukkasia. Ilman koko dataa kattavaa histogrammia opiskelijalle voisi jäädä väärä kuva datan sisällöstä. Lisäksi tavoittelin työn merkittävyyden havainnollistamista, sillä vertailemalla käytetystä datasta luotua histogrammia CMS-kokeen histogrammiin, opiskelija voi huomata, että ne muistuttavat hyvin paljon toisiaan. Tällä pyrin korostamaan sitä, että analyysissa käytetty data todella on aitoa CMS-kokeen dataa, ja että sillä voi tuottaa vastaavia tuloksia kuin CMS-kollaboraatiossa on saatu tehtyä.



Kuvio 13. Koko datatiedostoa kuvaava myoniparien invarianttien massojen histogrammi. Z -bosonia vastaavan piikin lisäksi havaittavissa on muita hiukkasista vastaavia piikkejä.

Koko datatiedoston kattavan histogrammin jälkeen loin työn loppuun vielä lyhyen lopetusosion (LIITE A-18). Siinä kerroin, että työn vastauksia ei tarvitse erikseen tallentaa tai palauttaa, vaan vastaukset tallentuvat suoraan TIMiin tarkastajan nähtäväksi. Lisäksi liitin mukaan linkin CERNin Open Data -portaaliin¹⁴, josta löytyy lisää tietoa CERNin avoimesta datasta. Aivan loppuun kirjasin vielä työssä käytetyt lähteet.

Laboratoriotyön valmistuttua testasin vielä työtä itsenäisesti. Käytin testaamiseen jälleen testitunnuksia, joilla pystyin simuloimaan työtä tekevää opiskelijaa. Havaittiin mm. sen, että sanallisiin kysymyksiin vastatessa käyttäjä näkee *saved*-tekstin tallentaessaan vastauksensa. Lisäksi käyttäjä voi tarkastella aiempia tekstikenttään tallentamia vastauksia. Vastaukset eivät myöskään tallennu toistensa päälle, vaan jokainen tallennus tuottaa uuden vastauksen huolimatta siitä, oliko käyttäjä tarkastelemassa aiempia vastauksiaan.

Testasin myös mitä tapahtuu, jos TIMistä poistuu kesken työn tekemisen. TIM antoi varoituksen mikäli yritin sulkea selaimen tilanteessa, jossa koodikomponentissa

¹⁴<http://opendata.cern.ch/>

oli ajamatonta koodia. Tämä on työn kannalta hyvä ominaisuus, sillä tällöin opiskelija ei voi vahingossa poistua keskeneräisestä työstä esim. työn linkkejä tarkastellessa. Kokeilin lisäksi poistua työstä varoituksesta huolimatta. Tämäkään ei aiheuttanut suuria ongelmia, sillä palatessani takaisin työhön selaimen edellinen sivu -painikkeella, löytyivät työssä tehdyt tallennukset ja aiemmat ajot. Myös ylläpitäjälle näkyivät nämä tallennetut ja ajatut asiat työstä poistumisesta huolimatta.

4.3 Ensimmäisen kehittämisvaiheen yhteenveto

Alaluvuissa 4.1 ja 4.2 kuvailin kehittämisprosessin ensimmäisen kehittämisvaiheen kulun. Tässä vaiheessa loin TIM-alustalla laboratoriotyön ensimmäisen version. Laatimisen ohella myös testasin itsenäisesti työn toimivuutta.

Prosessin aikana nousi esiin muutamia teknisiä haasteita. Analysoitava datatiedosto ylitti kooltaan TIMiin asetetun suurimman tiedostokoon 50 Mt:n rajan. Tiedoston lataaminen TIMiin onnistui kuitenkin siten, että TIMin kehittäjät nostivat rajaa hetkellisesti suuremmaksi.

Suuri tiedostokoko aiheutti myös toisen haasteen. Työn rakentamisen alkupuolella datatiedoston käsittely ylitti TIMiin asetetun suurimman mahdollisen laskentaajan. Tämä johti virheilmoitukseen ja ajamisen keskeytymiseen. Ongelma poistui sen jälkeen, kun TIMin kehittäjät siirsivät käytettävän datatiedoston apuhakemistoon, josta tiedoston lukeminen onnistui nopeammin.

Lisäksi teknistä haastetta ilmeni koodikomponenttien vuorovaikutuksessa. Työn eri osissa olevien koodikomponenttien tuli olla toisiinsa yhteydessä sillä tapaa, että aiemmissa komponenteissa ohjelmoidut asiat toimisivat myös seuraavissa. Oletuksena TIMistä ei löytynyt tällaista ominaisuutta. TIMin kehittäjätiimin avustuksella toteutin komponenttien vuorovaikutuksen ketjuttamalla komponentit siten, että yksittäinen komponentti ajaa aina itsensä lisäksi myös edellisen komponentin. Edellisen komponentin ajamiseen tarvittavan koodin kirjoitin piiloon käyttäjältä. Ketjutuksen avulla sain vuorovaikutuksen toimimaan. Pienenä uhkatekijänä ketjutus kasvattaa tarvittavaa laskenta-aikaa, sillä esimerkiksi viimeisen komponentin ajon yhteydessä ajetaan jo useampi edellinen. Laskenta-aika ei kuitenkaan ylittynyt ketjutusten takia.

Haasteiden lisäksi kohtasin kehittämisprosessin aikana tilanteita, joissa jouduin tekemään erilaisia valintoja työn suhteen. Ensimmäinen suurempi valinta koski työssä käytettävän ohjelmointikielen valitsemista. Päädyin Python-kieleen, sillä minulla oli jo aiempaa kokemusta avoimen datan käsittelystä Pythonilla, sekä olin mieltänyt kielen suhteellisen yksinkertaiseksi ymmärtää ja käyttää. Lisäksi päätöstä tuki tieto siitä, että Jyväskylän yliopiston fysiikan laitos oli tutkimuksen tekohetkellä suunnitellut lisäävänsä Pythonin käyttöä fysiikan kursseilla.

Kohtasin valintoja mm. ohjelmointiin liittyvissä tilanteissa. Päädyin esimerkiksi käyttämään histogrammien luomiseen lyhyttä komentoa pidemmän, monirivisen

komennon sijaan. Mielestäni pidempi komento olisi ollut liian haastava opiskelijoiden itse tuotettavaksi. Myös normaalijakauman sovittamisen histogrammiin tekevän koodin päätin antaa opiskelijalle lähes valmiiksi kirjoitettuna samoin perustein.

Lisäksi tein ohjelmointiin liittyviä valintoja erilaisten asioiden toteutustapoihin liittyen. Esimerkiksi invarianttien massojen tarkistamiseen valitsemani vertailun olisi voinut toteuttaa monilla eri tavoilla. Päädyin kirjoittamaan vertailukoodin siten, että vertailussa tarkastetaan, vastaavatko opiskelijan laskemien massojen ja itse laskemieni massojen kahdeksannet arvot täsmälleen toisiaan. Vertailu tällä tapaa sisältää pienen riskin, sillä koodi voi tulkita vain merkityksettömän vähän toisistaan eroavat arvot eri arvoiksi, johtaen kehotukseen laskea lasku uudestaan. En kuitenkaan törmännyt tällaiseen tilanteeseen työtä laatiessani, joten päätin luoda vertailun yllä kuvatun kaltaisena.

Valintoja liittyi myös työn sisältöön. Prosessin aikana tuli esimerkiksi päättää, kuinka paljon teoriaa sisällytän työn teoriaosioon. Päädyin suhteellisen kattavaan teoriaosuuteen, sillä uskoin, että kaikkea työssä tarvittavaa teoriaa ei tulisi käsittelemään työhön liittyvällä kurssilla. Toinen suurempi sisällöllinen päätös oli sanallisten kysymysten sarjan lisääminen työn loppupuolelle. Päätin, että data-analyysia tulkitaan työssä näiden sanallisten kysymysten kautta.

Haasteiden ja valintojen lisäksi kehittämisprosessin aikana tein erilaisia oivalluksia työhön liittyen. En ollut esimerkiksi prosessia aloittaessani keksinyt, että työhön voisi lisätä avattavia vihjeitä. Tämä tuli mieleeni prosessin aikana. Myös teoriaosaa rytmittävät kaksi sanallista kysymystä päätin lisätä mukaan teoriaosan luomisen yhteydessä.

Edellä mainittujen suurimpien huomioiden lisäksi prosessissa tapahtui monia erilaisia pieniä tilanteita. Näistä joitakin olen maininnut alalukujen 4.1 ja 4.2 kuvailussa. Todennäköisesti mainitsematta on kuitenkin jäänyt erilaisia tiedostamattomia pieniä valintoja, joita olen kehittäjän roolissa huomaamattani tehnyt.

5 Toinen ongelma-analyysi

Ensimmäisen kehittämisvaiheen valmistuttua siirryin tutkimuksen toiseen kehittämissykliin (ks. kuvio 2). Tämän syklin aloitin toisella ongelma-analyysillä, jonka kulku on kuvailtuna tässä luvussa. Toisessa ongelma-analyysissä testasin kehittämäni laboratoriotyötä kahdessa eri vaiheessa luonnontieteiden yliopisto-opiskelijoilla. Keräsin testauksista tietoa erilaisilla aineistonkeruumenetelmillä. Analysoin keräämääni aineistoa sisällönanalyysin ja sisällön erittelyn keinoin.

Alaluvussa 5.1 kerron tarkemmin aineistonkeruumenetelmäni. Keräsin aineistoa haastatteluiden, havainnoinnin ja kyselylomakkeen avulla. Alaluvussa 5.2 kuvailen työn testaamisen ensimmäisen vaiheen kulun. Tässä vaiheessa työtä testasi pareittain yhteensä kuusi matemaattis-luonnontieteellisten alojen opiskelijaa, jotka olivat kesäharjoittelussa CERNissä.

Työn testaamisen toisen vaiheen kulun kuvailen alaluvussa 5.3. Toisessa vaiheessa työtä testasi yhteensä yhdeksän Jyväskylän yliopiston fysiikan laitoksen opiskelijaa. Testaajat tekivät työn itsenäisesti. Välittömästi työn valmistuttua testaajat vastasivat kyselylomakkeeseen. Lisäksi kukin testaja osallistui myöhemmin yksilöhaastatteluun.

Testeistä sain käyttööni runsaan määrän erilaista aineistoa. Käytin aineiston analyysimenetelmänä sisällönanalyysia ja sisällön erittelyä, joista kerron tarkemmin alaluvussa 5.4. Analyysin kulun ja tulokset kuvailen alaluvussa 5.5. Luvun lopussa kerään toisessa ongelma-analyysissä tehdyt havainnot ja saadut tulokset yhteen. Tämä yhteenveto on alaluvussa 5.6.

5.1 Aineistonkeruumenetelmät

5.1.1 Haastattelu

Haastattelu on henkilökohtainen tilaisuus, jossa haastattelijä esittää suullisesti kysymyksiä haastateltavalle, eli tiedonantajalle [14, s. 85]. Tutkimuksessa keräsin osan aineistosta haastatteleamalla kehitettyä laboratoriotyötä testanneita opiskelijoita. Haastatteluilla pyrin selvittämään miten työ toimi opiskelijoiden mielestä – oliko työ sujuva, kaipasiko se kehitystä, auttoiko työ oppimaan? Lisäksi haastatteluilla tavoittelin tietoa siitä, miten työ täytti Nivalaisen, Asikaisen ja Hirvosen [6] esittämät tavoitteet fysiikan kokeellisille töille.

Toteutin tutkimuksen haastattelut teemahaastatteluina, eli puolistrukturoituina haastatteluina. Teemahaastatteluksi määritellään haastattelu, joka etenee etukäteen valittujen teemojen sekä niihin liittyvien tarkentavien kysymysten varassa. Teemahaastattelu sijoittuu haastattelumuotona lomakehaastattelun ja avoimen haastattelun väliin. [14, s. 87–90]

Esimerkiksi kyselylomakkeeseen verrattuna haastattelun merkittävin etu on joustavuus. Haastattelijä voi tarvittaessa oikaista väärinkäsityksiä, esittää tarkentavia kysymyksiä sekä keskustella haastateltavan kanssa. Paperi- tai verkkokyselyissä nämä eivät olisi mahdollisia. Haastattelussa vältytään siis helpommin mahdollisilta väärinymmärretyiltä kysymyksiltä. [14, s. 85]

Tuomen ja Sarajärven [14, s. 85] mukaan haastattelussa suositellaan, että tiedonantajille kerrotaan haastattelun aihe etukäteen tutustuttavaksi. Tämä on myös eettisesti perusteltua, sillä tiedonantajan tulisi tietää mitä haastattelussa tullaan käsittelemään. Tässä tutkimuksessa kerroin tiedonantajille informointilomakkeen (LIITTEET B JA C) yhteydessä tutkimuksen tavoitteet, mutta tarkkaa haastattelun aihetta en kirjallisesti ilmoittanut. Voidaan kuitenkin olettaa, että tiedonantajat ymmärsivät haastattelun liittyvän testattavaan laboratoriotyöhön. Lomakkeessa kerroin tiedonantajille myös, miten haastatteluista kerättyä tietoa käsitellään tutkimuksessa.

Teemahaastattelussa kysymysten järjestys ja sanamuodot voivat joko olla yhtenäiset kaikkien tiedonantajien välillä tai vaihdella eri haastatteluissa [14, s. 88]. Pidän tutkimuksen haastatteluissa kysymysten järjestyksen lähtökohtaisesti samana. Myös kysymysten sanamuodon pidin vakiona, sillä luin kysymykset aina tietokoneelle kirjoitetusta pohjasta. Joidenkin haastateltavien kohdalla kysyin myös haastattelupohjasta poikkeavia kysymyksiä.

Laadullisen tutkimuksen aineiston koosta Tuomi ja Sarajärvi [14, s. 98] kertovat seuraavasti: ”Voidaan todeta, että laadullisissa tutkimuksissa, ovat ne sitten ammatikorkeakouluissa tehtyjä opinnäytetöitä tai yliopistojen pro gradu -töitä, aineiston koot eivät vaihtele merkittävästi, vaan pääsääntöisesti aineiston koko on pieni tai vähäinen verrattuna määrälliseen tutkimukseen.” Heidän mukaansa opinnäytetyössä aineiston kokoa ei tule pitää merkittävimpänä kriteerinä, mutta siihen on syytä kiinnittää huomioita.

Laadullisessa tutkimuksessa on tärkeää, että haastatteluiden tiedonantajilla on tietoa ja kokemusta tutkittavasta ilmiöstä [14, s. 98]. Tiedonantajien valinta tulee siis olla harkittua. Tähän tutkimukseen valitsin tiedonantajia kahdesti, sillä suoritin työn testaukset kahdessa eri vaiheessa.

Ensimmäiset tiedonantajat, jotka tekivät työn pareittain ja joita haastattelin pareittain, olivat CERNissä kesäharjoittelussa olevia matemaattis-luonnontieteellisten alojen yliopisto-opiskelijoita. Tähän vaiheeseen osallistui yhteensä kuusi henkilöä, eli kolme paria. Kolme testaaajaa luki pääaineinaan fysiikkaa, yksi matematiikkaa sekä yksi bioinformaatioteknologiaa. Yksi testaaajista oli jättänyt esitietolomakkeen

(LIITE B) *Pääaine*: -kentän tyhjäksi. Mukaan lähteneet sopivat hyvin tiedonantajiksi, sillä kehitetty työ tulee käyttöön fysiikan kurssille, jota käyvät sekä fysiikan pääaineopiskelijat että sivuaineopiskelijat.

Toisen vaiheen tiedonantajat olivat Jyväskylän yliopiston fysiikan laitoksen opiskelijoita, jotka ilmoittautuivat vapaaehtoisina mukaan testaamiseen ja haastatteluun. Tiedonantajat palkittiin yhden opintopisteen suuruisella suorituksella, jonka pystyi lukemaan osaksi fysiikan syventäviä laboratoriotöitä. Palkinnolla pyrin varmistamaan riittävän osallistujien määrän. Tämä tavoite toteutui, sillä fysiikan laitokselta tiedonantajia tuli yhteensä yhdeksän kappaletta.

Fysiikan laitoksen opiskelijat soveltuivat hyvin tutkimuksen tiedonantajiksi, sillä kehitetty työ tulee käyttöön juuri Jyväskylän yliopiston fysiikan laitokselle. Tiedonantajat edustavat siis hyvin työn kohderyhmää. Täten voidaan olettaa, että heiltä saatava tieto on relevanttia tutkimuksen kannalta.

Haastattelussa on mahdollista, että haastateltavat pyrkivät kertomaan sosiaalisesti suotavia vastauksia [55, s. 206]. Tämä voi heikentää haastattelun luotettavuutta. Esimerkkinä voidaan tarkastella haastatteluissa käyttämäni kysymystä ”*Kuinka vaikea työ mielestäsi oli?*” On mahdollista, että haastateltava on kokenut työn hyvin vaikeaksi, mutta pitää tämän myöntämistä esimerkiksi nolona tai heikkouden osoittamisena. Tässä tilanteessa haastateltava voi kertoa työn olleen helpompi kuin hän on itse kokenut. Lisäksi on mahdollista, että haastateltavat puhuvat haastattelutilanteessa toisin kuin normaalisti [55, s. 206]. Nämä tekijät vaikuttavat siihen, että haastattelun avulla saatuja tuloksia ei aina voi yleistää varauksetta.

Olen kuvailut tarkemmin haastattelujen käytännön toteutusta erikseen testauksen molemmissa vaiheissa. Alaluvussa 5.2 kerron ensimmäisen testauksen yhteydessä tekemistäni haastatteluista. Alaluvussa 5.3 taas kuvailen toisen testausvaiheen haastattelujen käytännön toteutusta.

5.1.2 Havainnointi

Hyödynsin tutkimuksen aineistonkeruussa myös havainnointia. Havainnointi voidaan jakaa neljään eri muotoon: piilohavainnointiin, havainnointiin ilman osallistumista, osallistuvaan havainnointiin ja osallistavaan havainnointiin. Tutkimuksessa käytin näistä pääsääntöisesti havainnointia ilman osallistumista sekä kerran osallistuvaa havainnointia. Havainnoinnissa ilman osallistumista tiedonantajat tietävät osallistuvansa tutkimukseen, eli havainnointia ei tapahdu heiltä salaa. Tutkija tarkkailee tiedonantajia ulkopuolelta ja vuorovaikuttamatta heidän kanssaan. Lisäksi havainnoinnissa ilman osallistumista tiedonantajilta tulee kysyä lupa havainnointiin. Kerroin havainnoinnista kummankin testivaiheen osallistujille informointilomakkeiden yhteydessä (LIITTEET B JA C). Osallistuvassa havainnoinnissa taas tutkija on aktiivinen toimija yhdessä tiedonantajan kanssa. Tutkija voi siis vaikuttaa havainnoinnin tapahtumien kulkuun. [14, s. 93–95]

Käytin havainnointia eri osissa tutkimusprosessia. Työn testaamisen ensimmäisessä vaiheessa CERNissä havainnoin työtä testaavia pareja. Kirjoitin ylös havaintoja testaajien keskusteluista ja siitä, mitä he tekivät. Olin sijoittunut siten, että näin osittain testaajien näytön, jolla he tekivät työtä. Parien lisäksi havainnoin osallistuvasti tilannetta, jossa testasimme työtä yhdessä ohjaajani Kati Lassila-Perinin kanssa.

Testaamisen toisessa vaiheessa havainnoin työtä tekeviä fysiikan laitoksen opiskelijoita testaamisen aikana. Omien havaintojeni lisäksi tallensin jokaisen testaajan tietokoneelta ruutukaappausvideon, jossa näkyy mitä testaaja on työn aikana tehnyt. Tässä havainnoinnissa en nähnyt testaajien näyttöjä omin silmin. Kirjasin ylös havaintoja siitä, mitä testaamisen aikana sanottiin tai mitä testitilassa tapahtui. Näitä havaintoja hyödynsin enemmän taustatietoina kuin varsinaisena tutkimusaineistona. Myös ruutukaappausvideoita käytin vain tiettyjen työn kohtien tarkkailuun, eli en käynyt järjestelmällisesti läpi videoiden sisältöä.

Kuten yllä olevasta kuvailusta selviää, havainnointi toimi tutkimuksessa kerättävän tiedon monipuolistajana. Pääasiallisina aineistonkeruumenetelminä käytin teema-haastattelua ja kyselylomakkeita.

5.1.3 Kyselylomake

Kysely on aineistonkeruumenetelmä, jossa aineistoa kerätään standardoidusti, eli kaikilta vastaajilta täsmälleen samalla tavalla [55, s. 193]. Kysely voidaan toteuttaa esim. posti- tai verkkokyselynä. Tässä tutkimuksessa käytin kahta erilaista kyselylomaketta työn toisen vaiheen testauksessa. Ennen työn tekemistä keräsin testaajien taustoista tietoa paperisella esitietolomakkeella (LIITE C). Toinen käyttämäni lomake oli verkossa täytettävä kysely, jolla keräsin työn toisen vaiheen testaajilta tietoa välittömästi testin jälkeen. Kuvailen lomakkeiden käyttöä ja tiedonkeruuta tarkemmin alaluvussa 5.3.

Kyselylomakkeen tulee olla helposti täytettävä ja ulkoasultaan moitteeton [55, s. 204]. Toteutin verkkokyselyn *Google Forms* -palvelun¹⁵ avulla. Olin jo aiemmin kandidaatintutkielmaa tehdessäni todennut palvelulla tuotetut kyselyt selkeiksi vastata ja käsitellä. Kuvio 14 kuvaa lomakkeen ensimmäisen sivun visuaalisen ulkoasun.

Lomakkeen alussa kysyin vastaajan nimen, jotta voisin yhdistää vastaukset kunkin testaajan haastatteluihin ja muihin aineistoihin. Lisäksi tiedustelin, oliko vastaaja käyttänyt TIM-alustaa aiemmin ennen testauksia. Näiden tietojen jälkeen lisäsin lomakkeeseen kahdeksan sanallista kysymystä tai kehotusta, jotka jaoin aina uudelle lomakkeen sivulle edellisen jälkeen. Tällä pyrin siihen, että vastaaja ei säikähtäisi kysymysten määrää ja keskittyisi vain yhteen kohtaan kerrallaan.

Paperisen esitietolomakkeen, joka jaettiin testaajille ennen työn tekemistä, rungon laadin itse. Lähetin lomakkeen kommentoitavaksi ohjaajalleni Jan Sarénille sekä Jy-

¹⁵<https://www.google.com/forms/about/>

väskylän yliopiston opettajankouluslaitoksen professori Jouni Viirille. Kommenttien perusteella muokkasin lomakkeen muutamia kohtia. Käytin hieman erilaisia lomakkeita testaamisen ensimmäisessä vaiheessa CERNissä (LIITE B) ja toisessa vaiheessa Jyväskylässä (LIITE C).

Kysymyksiä testattavasta työstä

Vastaa vielä rauhassa ja ajatuksella seuraaviin kysymyksiin, kun olet saanut suoritettua työn testaamisen loppuun.
Vastauslaatikot laajenevat automaattisesti rivinvaihdolla.

Sähköpostiosoitteesi (paturikk@student.jyu.fi) tallennetaan, kun lähetät tämän lomakkeen. Jos et ole tämä henkilö, [vaihda tiliä](#).

Nimesi:

Oma vastauksesi

Olitko koskaan käyttänyt TIM-alustaa ennen tätä työtä?

Kyllä

En

Kerro omin sanoin, mitä teit työssä.

Oma vastauksesi

SEURAAVA Sivu 1 / 8

Älä koskaan lähetä salasanaa Google Formsin kautta.

Kuvio 14. Kyselylomakkeen ensimmäinen sivu. Alareunassa oleva palkki kertoo lomakkeen etenemisestä.

5.2 Kuvailu ensimmäisestä testauksesta

Toinen ongelma-analyysi alkoi tilanteesta, jossa olin saanut valmiiksi kehitettävän laboratoriotyön ensimmäisen version. Kuten aiemmin mainitsin, testasin työtä kahdessa eri vaiheessa. Tässä alaluvussa kuvailen, miten toteutin käytännössä testaamisen ensimmäisen vaiheen.

Tutkimusta aloittaessa aion testata kehitettävää työtä ainoastaan Jyväskylän yliopiston fysiikan laitoksen opiskelijoilla. Vasta myöhemmässä vaiheessa tutkimusprosessia heräsi idea siitä, että työtä testattaisiin jo ennen sovittua Jyväskylän testiä. Aloitin tutkimuksen tekemisen CERNissä, Genevessä, jossa olin myös ensimmäisen version valmistuttua. Tällöin päätin ohjaajieni kanssa, että työtä voisi testata

jo etukäteen CERNissä olleilla suomalaisilla kesäharjoittelijoilla ennen varsinaista Jyväskylän testaamista.

CERNissä tiedustelin kesäharjoittelijoille järjestetyn tapaamisen yhteydessä vapaaehtoisia osallistujia testaamiseen. Yhteensä kuusi harjoittelijaa ilmaisi kiinnostuksensa. Osallistujat muodostivat kolme paria, joista kullekin sovin oman ajankohdan työn tekemistä varten.

Testitilanteita varten varasin häiriöttömän tilan, jossa pari pystyi tehdä työn rauhassa. Jokainen pari toi testausta varten paikalle oman kannettavan tietokoneen. Kaksi pareista teki työn Windows-käyttöjärjestelmällä, kun taas yhdellä pareista oli käytössä Linux-järjestelmä. Olin itse testannut työtä Mac OS X -järjestelmällä, joten tällä tapaa sain hyvän alustavan kuvan siitä, toimiiko työ yleisimmillä käyttöjärjestelmillä.

Ennen työn aloittamista ohjeistin parit työskentelemään yhdessä omaan tahtiin. Kehotin pareja pohtimaan ja kommentoimaan ääneen työn aikana kohdattavia tilanteita. Äänitin koko testitilanteen, jotta parien kommentit jäisivät talteen. Lisäksi asetuin parien taakse siten, että näin osittain heidän näyttönsä.

Testauksen aikana havainnoin tilannetta ja kirjasin ylös tekemiäni havaintoja. En kommentoinut tai avustanut parien työskentelyä muutamaa huomautusta lukuun ottamatta. Yhdelle parille huomautin, että työtä tehdessä saa käyttää haluamiaan ulkopuolisia lähteitä. Toista paria taas ohjeistin jättämään huomiotta TIMissä olevat klikattavat palkit, joilla TIMin käyttäjä voi merkitä dokumentista lukemansa kohdat. Muuten annoin parien tehdä työn omin avuin keskenään.

Kaikki kolme paria saivat työn valmiiksi. Testaajien työskentelyyn käyttämät ajat vaihtelivat reilun tunnin ja kahden tunnin välillä. Näiden havaintojen perusteella pystyin jo toteamaan, että työn ensimmäisessä versiossa ei ollut vakavia, työn tekemisen estäviä puutteita tai ongelmia. Lisäksi työ vaikutti pituudeltaan sopivalta. Fysiikan laitoksen laboratoriotöiden oletetaan vievän maksimissaan neljä tuntia aikaa työtä kohden.

Heti työn tekemisen jälkeen haastattelin jokaisen pareista. Toteutin haastattelut teemahaastatteluina. Käytin jokaisessa haastattelussa seuraavaa runkoa:

- Kertokaa omin sanoin, mitä teitte työssä.
- Koetteko, että työ kehitti fysiikan kokeellisia taitojanne? Millä tavalla?
- Millä tavalla tietonne hiukkasfysiikasta muuttuivat, vai muuttuivatko?
- Millä tavalla ohjelmointitaitonne muuttuivat, vai muuttuivatko?
- Miten motivoivaksi koitte työn?
- Kertokaa työn sujuvuudesta.
- Mitä haasteita kohtasitte työtä tehdessä?

- Mitä opitte työstä?
- Miten olisitte voineet oppia lisää työstä?
- Mitä muuttaisitte työstä?
- Kuinka vaikea työ mielestänne oli?
- Mitä jäi sanomatta tai tuliko muita huomioita?

Yllä olevan rungon lisäksi reagoin haastatteluissa esiin tulleisiin tilanteisiin. Esi-
tin esimerkiksi tarkentavia kysymyksiä sekä keskustelin haastateltavien kanssa työn
sisällöstä. Nauhoitin jokaisen haastattelun ja myöhemmin myös litteroin, eli sana-
sanaisesti puhtaaksi kirjoitin [55, s. 222], tallenteet.

Lisäksi CERNissä Fysiikan tutkimuslaitoksen puolen ohjaajani Kati Lassila-Perini
testasi työtä ja antoi siihen kehittämisehdotuksia. Tässä testauksessa teimme työn
yhdessä ja keskustelimme samalla testauksessa ilmenneitä asioita. Käytin siis osal-
listuvaa havainnointia, jossa tutkija toimii aktiivisesti tiedonantajan kanssa [14, s.
94–95]. En nauhoittanut käymäämme keskustelua, vaan kirjoitin esille tulleet hu-
miot itselleni talteen.

Testauksen ensimmäisestä vaiheesta keräsin siis tietoa havainnoimalla, äänittämällä
testaajien väliset keskustelut työn tekemisen aikana sekä äänittämällä haastattelut.
Lisäksi sain ohjaajaltani kehittämisehdotuksia. Päätin käyttää työn kehittämiseen
kuitenkin ainoastaan havainnoinnin aikana tekemiäni muistiinpanoja sekä Lassila-
Perinin huomioita. Tähän päätökseen päädyin osittain aikataulusyistä. Viimeisen
parin haastattelu ajoittui tiistaille 13.6.2017, kun taas Jyväskylässä tapahtuvan tes-
taamisen toisen vaiheen olin aikatauluttanut perjantaille 16.6.2017.

Päätin tehdä CERNin testien perusteella ainoastaan nopean, pienimuotoisen kehittä-
misen työhön. Havainnointien ja Lassila-Perinin ehdotusten perusteella tein työ-
hön noin 40 pientä muutosta. Luonteeltaan muutokset olivat sellaisia, että niiden
toteuttaminen ei vaatinut tarkempaa analysointia. Muutoksissa esimerkiksi korja-
sin kirjoitusvirheitä, tarkensin faktoja sekä avasin tarkemmin matemaattisia lausek-
keita. Suurin muutos oli se, että lisäsin työssä histogrammin analysointivaiheeseen
kaksi sanallista kysymystä lisää. Olen kuvannut tarkemmin tekemiäni suuremmat
muutokset alaluvussa 4.2.

Tämän pienimuotoisen kehittämisen avulla sain sujuvoitettua työn kulkua sekä pois-
tettua karkeimmat virheet. Pyrin tällä siihen, että Jyväskylässä suoritettavassa tes-
tauksen toisessa osassa työstä voisi tulla ilmi jo syvällisempiä kehityskohteita pienten
epäkohtien sijaan. Lisäksi sain CERNin testauksista tietoa ja kokemusta Jyväskylän
testien toteuttamista varten. Pystyin toteamaan, että työ toimii eri käyttöjärjes-
telmillä sekä eri henkilöiden tekemänä. Tämä vähensi huolta siitä, että Jyväskylän
testauksessa ilmenisikin jokin este, mikä pilaisi testitilaisuuden. Haastatteluissa taas
pääsin harjoittelemaan haastattelijan roolia sekä totesin äänityksen toimivuuden.

Testauksen ensimmäisen vaiheen perusteella tekemiäni muutosten jälkeen työ oli

valmis Jyväskylässä tapahtuvaan testauksen toiseen vaiheeseen.

5.3 Kuvailu toisesta testauksesta

Tässä alaluvussa kuvailen Jyväskylässä tapahtuneen testausprosessin kulun. Alaluku on jaettu eri testausprosessin vaiheita kuvaaviin osioihin. Alaluvussa 5.3.1 kerron valmisteluista, joita tein ennen testauksia. Alaluvussa 5.3.2 kuvailen pohjaluentoa, jolle testajat osallistuivat ennen työn tekemistä. Itse testitilanteen kulusta ja havainnoinnista kerron alaluvussa 5.3.3.

Heti työn tekemisen jälkeen testajat vastasivat kyselylomakkeeseen, josta kerron alaluvussa 5.3.4. Lisäksi myöhemmin jokainen testaja osallistui yksilöhaastatteluuun. Haastattelujen toteutuksen kuvailen alaluvussa 5.3.5.

Tästä luvusta eteenpäin käytän testaaajista lyhenteitä T1–T9, jotta testausprosessin eri vaiheissa tapahtuneet asiat voidaan liittää eri testajiin.

5.3.1 Testauksen valmistelu

Aloitin toisen testausvaiheen suunnittelun jo tutkimuksen alussa. Päätin, että tulen testaamaan kehittämäni työtä Jyväskylän yliopiston fysiikan laitoksen opiskelijoilla, sillä valmis työ tulee käyttöön juuri tälle kohderyhmälle. Lisäksi päätin haastatella opiskelijoita testaamisen jälkeen.

Fysiikan laitoksen yliopistontutkija Sami Räsänen sekä yliopistonlehtori Jussi Mauuksela avustivat minua testitilanteen suunnittelussa. Ideoimme Räsänen ja Mauukselan kanssa, että testiryhmän voisi yrittää koota fysiikan laitoksen kesäopiskelijoista. Suunnittelimme testitilaisuuden ajoittuvan kesäkuulle, sillä olin 15.6.2017 asti tekemässä tätä tutkimusta CERNissä.

Osallistujien määrän varmistamiseksi päätimme antaa opiskelijoille testaamisesta palkkioksi yhden opintopisteen suorituksen. Fysiikan laboratoriotöistä vastaavan vanhemman lehtorin Sakari Juutisen päätöksellä opiskelijat pystyivät lukemaan tämän suorituksen osaksi fysiikan syventäviä laboratoriotöitä. Lisäksi suoritus katsottiin osaksi kesäopintoja.

Päätimme Räsänen kanssa myös sisällyttää osaksi testaamista pohjaluennon (LIITE D), jossa käsitellään kokeellista hiukkasfysiikkaa ja hiukkasten havaitsemista CMS-ilmaisimella. Räsänen lupautui pitämään luennon. Pohjaluennolla tavoittelimme mahdollisimman hyvin aitoa tilannetta simuloivaa testiä. Laboratoriotyö tulee osaksi modernia fysiikkaa käsittelevää kurssia, joten aidossa tilanteessa kurssilla oleva opiskelija saa ennen työn tekemistä tietoa kurssin luennoilta.

Sovimme luennon ajankohdaksi torstain 15.6.2017 ja testipäiväksi sitä seuraavan perjantain 16.6.2017. Tiiviillä aikataululla tavoittelimme sitä, että kukaan testaukseen osallistuvista ei jättäisi prosessia kesken. Aikataulutuksen jälkeen aloitin tes-

taajien hakuprosessin välittämällä fysiikan laitoksen opiskelijoiden sähköpostilistalle 4.5.2017 seuraavan viestin:

”Fysiikan aineenopettajaopiskelija Paavo Rikkilä tutkii ja kehittää pro gradu -tutkielmassaan laboratoriotyötä uudistuvaa modernin fysiikan kurssia varten. Työ toteutetaan ns. lapputyönä, ja sen aiheena on hiukkasfysiikan kokeellinen tutkimus CERNin CMS-kokeen aidon datan avulla.

Työn ensimmäisen version testaaajiksi kesälle 2017 haetaan fysiikan opiskelijoita. Testaaminen ei vie kauaa aikaa, ja **siitä palkitaan yhdellä opintopisteellä (1 op)**, joka luetaan fysiikan syventäväksi laboratoriotyöksi (<https://www.jyu.fi/science/opiskelu-ohjeet/fysiikka/tyoosasto/syventavien-opintojen-laboratoriotyot-ohjelmatyot>). Opintopiste voidaan lukea kesän aikana suoritettuihin opintoihin, jotka mahdollistavat opintotukien nostamisen kesän ajalta.

Testaaminen koostuu seuraavista osista:

- to 15.6. klo 12:15–14:00 pohjaluento hiukkasfysiikasta, Sami Räsänen (2 h)
- pe 16.6. klo 12:30–16:30 työn tekeminen tietokoneella (4 h)
- valitsemasi haastattelu-aika viikolta 25 (30–45 min), ajat nähtävissä osoitteessa <https://docs.google.com/spreadsheets/d/1d1Smujw5SbHbUQPhmAiM2WERbR25e50nQ6t4Lhg87rk/edit#gid=0>

Osallistumalla annat tärkeän panoksen fysiikan opetuksen kehittämiseen ja tutkielman laatimiseen. Osallistuaksesi voit olla opinnoissasi missä vaiheessa tahansa. Osallistuminen arvioidaan asteikolla hyväksytty/hylätty ja työstä ei kirjoiteta erillistä raporttia tai selostusta.

Ilmoittaudu testiryhmään Korpissa osoitteessa <https://korppi.jyu.fi/kotka/course/student/courseInfo.jsp?course=157502>. Kurssille ilmoittautumisen jälkeen klikkaa kohtaa ”Laboratoriotyöt” ja ilmoittaudu ryhmään ”Uuden laboratoriotyön testaaminen”. Muista myös valita itsellesi sopiva haastattelu-aika, ohjeet löytyvät lomakkeesta.

Mahdollisia kysymyksiä aiheesta voi lähettää osoitteeseen paavo.t.rikkila@student.jyu.fi.

Ystävällisin terveisin,
Paavo Rikkilä”

Varsinaisen viestin lisäksi välitin sähköpostilistalle myös vajaa kaksi viikkoa myöhemmin 16.5.2017 muistutusviestin. Ennen muistutusta mukaan oli ilmoittautunut kuusi testaaajaa, ja muistutuksen jälkeen ilmoittautuneita tuli vielä kolme lisää. Yhteensä siis testaamiseen ilmoittautui yhdeksän fysiikan laitoksen opiskelijaa, joista kaikki osallistuivat koko testausprosessiin.

Ilmoittautumisvaiheessa kaksi osallistujaa (T3, T6) kertoivat, että perjantaina 16.6. tapahtuva testausajankohta ei sopisi heille. Päätin järjestää toisen mahdollisen ajan

työn tekemiselle keskiviikolle 28.6., joka sopi kummallekin osallistujalle. Lisäksi kaksi opiskelijaa (T2, T3) kertoivat, että eivät pääsisi osallistumaan 15.6. pidettävälle pohjaluennolle. Tämän ongelman ratkaisemiseksi päätimme Sami Räsänen kanssa videoida luennon. Hän järjesti luennolle videokameran sekä latsi videotallenteen osallistujien nähtäväksi. Kaksi testajaa siis katsoi luennon itsenäisesti videolta ennen työn tekemistä, ja loput seitsemän osallistuivat luennolle.

5.3.2 Pohjaluento

Testiprosessi alkoi torstaina 15.6.2017 pohjaluennolla. Alussa Räsänen avasi testajille tutkimuksen tavoitetta ja taustoja. Varsinainen luento käsitteli mm. hiukkasfysiikan standardimallia, hiukkasten hajoamista, kokeellista hiukkasfysiikkaa CERNissä sekä invariantin massan laskemista. Pohjaluennon luentodiat ovat raportin liitteenä (LIITE D).

Luennon jälkeen osallistujat täyttivät informointi- ja esitietolomakkeen (LIITE C). Kaksi luennon itsenäisesti videolta katsontua osallistujaa täyttivät lomakkeen testitulanteessa ennen työn tekemistä. Informointilomakkeessa osallistujille kerrottiin tutkimuksen tavoite, avattiin millä eri tavoin testauksesta kerätään aineistoa sekä kerrottiin kerätyn aineiston käsittelystä. Lisäksi lomakkeessa mainittiin, että palkinnoksi tarjotun opintopisteen saaminen edellyttää kaikkiin tutkimuksen osa-alueisiin osallistumista. Osallistujia pyydettiin informointilomakkeen lopussa vahvistamaan suostumuksensa tutkimukseen.

Esitietolomakkeella pyrin keräämään tietoa testaukseen osallistuvien taustoista sekä mahdollisista tulosten tulkinnan kannalta hyödyllisistä asioista. Lomakkeen alussa kysyin osallistujien pääaineen sekä sivuaineet. Tämän jälkeen tiedustelin monivalintakysymyksillä, olivatko testajat käyneet fysiikan laitoksella kolmea työhön liittyvää kurssia. Sisällytin *kyllä* ja *ei* -vaihtoehtojen lisäksi vastausvaihtoehdon *en ole varma*.

Lisäksi esitin lomakkeessa kolme väittämää, joihin vastattiin 5-portaisilla Likertin asteikoilla. Ensimmäinen väittämä oli muotoa *”Olen mielestäni hyvä ohjelmoimaan.”* Toisessa väittämässä tiedustelin osallistujien hiukkasfysiikan tuntemusta muodossa *”Hiukkasfysiikka on minulle tuttua.”* Kolmannella väittämällä pyrin saamaan tietoa osallistujien aiemmasta kokemuksesta Python-ohjelmointikielen parissa. Tämän väittämän kirjoitin muotoon *”Olen ohjelmoinut Python-ohjelmointikielillä.”* ja vastausvaihtoehdoiksi valitsin 5-portaisen asteikon väliltä *erittäin usein* ja *en ollenkaan*.

5.3.3 Työn tekeminen

Pohjaluentoa seuraavana päivänä 16.6. järjestin ensimmäisen testitilaisuuden, johon osallistui seitsemän yhdeksästä testaaajasta. Loput kaksi (T3, T6) testasivat työtä 28.6. järjestämässäni toisessa tilaisuudessa. Varasin testauksia varten fysiikan laitoksen tietokoneluokan. Olin pyytännyt yliopiston IT-palveluita asentamaan etukäteen jokaiselle luokan koneelle *CamStudio*-ohjelman, jonka avulla pystyin tallentamaan ruutukaappausvideot testaaajien näytöistä. Olin myös testannut *CamStudio*n käyttöä itsenäisesti etukäteen.

Aloitin ensimmäisen testitilaisuuden toivottamalla osallistujat tervetulleiksi ja avaamalla heille testauksen taustoja. Kerroin, että kehitettävänä oleva työ tulee valittavaksi uudistuvalla modernin fysiikan kurssille itsenäisesti tehtävänä lapputyönä. Lapputyöllä tarkoitetaan fysiikan laitoksella laboratoriotyötä, josta ei palauteta erillistä työselostusta. Kerroin myös, että testitilaisuudella ja pohjaluennolla simuloidaan kurssin tilannetta, joten työ tehdään itsenäisesti. Mainitsin, että osallistujat voivat unohtaa roolini työn tekijänä sekä havainnoijana. Kehotin myös osallistujia olemaan ja toimimaan kuten normaalisti.

Ennen aloittamista pyysin kaikkia testaaajia asettamaan koneillensa oikeat *CamStudio*n asetukset esimerkiksi näyttäen. Asetusten säätämisen jälkeen kerroin, että osallistujat voivat tarvittaessa pyytää apua testaamisen aikana. Päätin, että en kuitenkaan välttämättä antaisi apua heti pyytämisen jälkeen tilanteesta riippuen. Mainitsin myös, että tulisin nauhoittamaan tai videoimaan keskustelut, joita voisi syntyä minun ja testaaajien välille. En kuitenkaan päätenyt tekemään näin, vaan sen sijaan kirjasin havaintoihini ylös tärkeimmät huomiot käydyistä keskusteluista.

Tajusin jälkikäteen muutamia asioita, joita olisin voinut korostaa tarkemmin ennen testaamista. Olisin voinut avata vielä tarkemmin sitä, mitä kaikkea tutkimukseni käsittää. Lisäksi osallistujille olisi voinut olla hyvä korostaa, että työn voi tehdä rauhassa välittämättä siitä, missä ajassa muut saavat työn valmiiksi. Esimerkiksi viimeisille testaaajille muiden poistuminen voisi aiheuttaa sosiaalista painetta siitä, että he ovat tehneet työtä pidempään.

Edellä mainittujen valmistelujen jälkeen annoin noin kello 12:50 osallistujille luvan aloittaa työn tekeminen. Testauksen alettua aloin havainnoida tilannetta ja kirjata havaintojani muistiin. Käytin menetelmänä havainnointia ilman osallistumista, jossa tutkijan rooli on tarkkailla tiedonantajia ulkopuolelta vuorovaikuttamatta heidän kanssaan. Vuorovaikutusta syntyi työn tekemisen aikana ainoastaan tilanteissa, joissa osallistuja pyysi apua johonkin ongelmaan.

Havainnoinnin alkupuolella pohdin, että minun täytyisi varoa auttamasta testaaajia liian äkkiä mahdollisissa ongelmatilanteissa. Kurssilla työn tekemiseen ei tule olemaan saatavilla välitöntä apua, joten avustaminen vääristäisi simulointia.

Noin tunnin kuluttua työn aloittamisesta varmistin omalla koneellani, että ruutukaappausvideon tallentaminen onnistuu. Video vaikutti tallentuvan normaalisti.

Tämän jälkeen testaaaja T2 pyysi apua ongelmaan. Testaaaja oli saanut työssä virheilmoituksen, mutta ei tiennyt mistä ilmoitus tulee. Kerroin hänelle, että en vielä tarjoa apua, mutta käyn myöhemmin uudestaan tarkistamassa tilanteen. Huomasin havainnoijana olemisen roolista tässä vaiheessa sen, että olisin halunnut auttaa testaaajaa, mutta minun tuli välttää vielä avun antamista.

Noin 17 minuuttia myöhemmin sama testaaaja sai työssään virheilmoituksen *"Runtime exceeded, maybe loop forever."* Päätin tässä vaiheessa avustaa häntä, sillä virheilmoitus viittasi TIMin laskenta-ajan ylittymiseen, joka ei välttämättä olisi työn tekijästä riippuva virhe. Testaaaja oli jumissa koodikomponentissa, jossa luodaan histogrammi invarianttien massojen arvoista (LIITE A-12). Koetin muuttaa testaaajan kirjoittamaa koodia sekä päivittää verkkosivun, jossa TIM oli auki. Tämän jälkeen alustin koodikomponentin ja kirjoitin itse komponenttiin tarvittavan koodin. Tämän jälkeen koodi toimi ja ajo tuotti oikeanlaisen histogrammin. Annoin siis testaaajalle T2 valmiiksi oikean muotoisen histogrammin tuottamiseen vaadittavan komennon.

Tarkistin myöhemmin tilanteen ruutukaappausvideolta. Tarkastuksessa löysin syyn ajoajan ylittymisestä varoittavalle virheilmoitukselle. Ilmoitus johtui siitä, että tekijä oli kirjoittanut koodikomponentin loppuun ylimääräisen `plt.show()` -komennon. Normaalisti Pythonilla ohjelmoitaessa komento ei aiheuttaisi ongelmia, mutta jostakin syystä TIMin koodikomponenteissa tämä komento saa aikaa ajoajan ylittymisen.

Ensimmäiseksi työn sai valmiiksi noin tunnin ja kahdenkymmenen minuutin kuluttua aloittamisesta testaaaja T4. Jokaisen testaaajan kohdalla kävin työn valmistuttua sammuttamassa ruutukaappausvideon tallennuksen sekä ohjasin testaaajan täyttämään verkkokyselylomakkeen. Kerron kyselylomakkeen sisällöstä takemmin alaluvussa 5.3.4. Näytin lomakkeeseen vievän osoitteen testitilan valkokankaalla, mutta en esimerkiksi lähettänyt osoitetta osallistujien sähköpostiin. Tällä varmistin sen, että kaksi myöhemmin työtä testaavaa osallistujaa eivät pääsisi tutustumaan lomakkeeseen etukäteen.

Kello 14:24, eli noin puolitoista tuntia työn aloittamisesta, testaaaja T1 kysyi minulta apua. Hän kysyi, miksi hänen saamansa *invalid syntax* -muotoinen virheilmoitus viittasi normaalifunktion sovittamista käsittelevässä osiossa (ks. alaluku 4.2.5) esiintyvään muuttujaan `alkuarvot`. Kerroin, että en vielä auttaisi testaaajaa, mutta tulisin myöhemmin katsomaan tilannetta uudestaan.

Kahdentoista minuutin jälkeen palasin edellä mainitun testaaajan luokse. Testaaaja oli yhä jumissa samassa kohtaa. Kysyin, miten hän voisi yrittää koodin ajamista uudestaan. Tämän jälkeen testaaaja kopioi käyttämänsä normaalijakauman lausekkeen tallenteen ja painoi koodikomponentin *Alusta*-painiketta. Ohjeistin testaaajaa yrittämään tällä tapaa uudestaan koodin ajoa. Kerroin myös, että mikäli vika onkin TIMistä riippuva, niin autan sen selvittämisessä.

Kävin tarkistamaan tilanteen vielä muutaman kerran tämän jälkeen. Kerroin myös vinkkinä sen, että `alkuarvot` tulisi kirjata muodossa `alkuarvot = [,]` ja että arvojen päättely onnistuu normaalijakauman lausekkeen avulla. Testaaaja kertoi ko-

keilleensa jo tätä muotoa. Lopulta ongelman syyksi osoittautui se, että normaali-jakauman lausekkeen lopusta puuttui sulkumerkintä. Virheilmoitus tästä viittasi kuitenkin hämäävästi muuttujaan `alkuarvot`, sillä normaalijakauman lausekkeen ja muuttujan välissä oli ainoastaan koodiin vaikuttamattomia kommenttirivejä (ks. esimerkki 5).

Seuraava havainnoinnista ylös kirjaamani tapahtuma oli testaaajan T8 kello 15:18 esittämä kysymys. Testaaja kysyi minulta CMS-kokeen tuottamassa histogrammissa (LIITE A-17) esiintyneiden hiukkasten nimiä. Vastasin, että en kerro niitä testauksen aikana, mutta työn valmistuttua voin paljastaa ne. Pohdin, että kysymys saattoi mahdollisesti liittyä siihen, että testaaja halusi vastata avoimeen sanalliseen kysymykseen mahdollisimman tarkasti. Tämä sai minut ajattelemaan, että työn sanallisissa kysymyksissä voisi vielä tarkastaa, saako niistä kuvan siitä, kuinka laajasti niihin oletetaan vastattavan.

Kello 15:37 työtä oli tekemässä enää kaksi osallistujaa (T1, T8), joista T8 toinen sai juuri työn valmiiksi. Pohdin havainnoinnin aikana, että tämä voisi mahdollisesti vaikuttaa viimeisen testaaajan T1 työskentelyyn. Viimeisenä olemisen voisi luoda painetta työn valmiiksi saamiseen. Viimeinen testaaja ilmoitti olevansa valmis kello 15:42. Samalla hän kysyi uudestaan, olisinko voinut kertoa CMS-kokeen histogrammissa esiintyneiden hiukkasten nimet. Kerroin vastaavani kysymykseen vasta myöhemmin haastattelun jälkeen, sillä tieto olisi voinut vaikuttaa testaaajan vastauksiin työn jälkeen tehtävissä kyselylomakkeessa ja haastattelussa.

Viimeinen testaaja palautti työn jälkeen verkossa täytettävän kyselylomakkeen noin kello 15:55, ja poistui tämän jälkeen. Testauksen päätyttyä kävin siirtämässä kaikkien testaaajien koneille tallennetut ruutukaappausvideot itselleni talteen.

Testaaajien työhön käyttämä aika vaihteli noin 1 tunnin 20 minuutin ja 2 tunnin 50 minuutin välillä. Ensimmäisten seitsemän testaaajan aikakeskiarvo oli noin 2 tuntia 7 minuuttia. Ajat sisälsivät ainoastaan työn tekemiseen kuluneen ajan, eivätkä esimerkiksi kyselylomakkeen täyttämistä.

Toinen testitilaisuus keskiviikkona 28.6. oli myös fysiikan laitoksen tietokoneluokassa. Tähän tilaisuuteen osallistuivat testaaajat T3 ja T6, joille ensimmäisen testin aikataulu ei ollut sopiva. Alustin testin samalla tapaa kuin ensimmäisessä tilaisuudessa, mutta hieman epävirallisemmalla tyyllillä. Ainoana lisänä mainitsin alustuksessa, että työssä saa mennä niin kauan aikaa kuin tekeminen vaatii.

Toinen testaus alkoi kello 13:26. Välittömästi aloituksen jälkeen tietokoneluokan verkkoyhteys katosi hetkeksi. Tämän seurauksena testaaja T6 joutui sammuttamaan koneensa ja aloittamaan työn uudestaan. Testaaajan T3 koneeseen yhteyskatko ei vaikuttanut, joten hän jatkoi työskentelyä. Kello 13:36 myös T6 pääsi aloittamaan työn uudestaan.

Hetken kuluttua tästä, kello 13:45, huomasin, että olin asettanut ruutukaappausvideot tallentumaan väärään sijaintiin. Keskeytin hetkeksi testin kulun ja korjasin asian. Testi jatkui noin kello 13:48. Hieman tämän jälkeen testaaja T6 kysyi, tulee-

ko hänen katsoa teoriaosaan liitetty video (LIITE A-5) nyt, kun verkkoyhteys taas toimi. Kehotin häntä katsomaan videon.

Havainnoinnissa ei hetkeen tapahtunut mitään merkittävää. Kello 14:30 havaitsin, että testaaaja T3 oli ollut ilmeisesti jonkin aikaa jumissa osiossa, jossa lasketaan invarianttien massojen arvot (ks. alaluku 4.2.3). Kävin kysymässä noin kymmenen minuutin kuluttua, oliko hänellä kaikki kunnossa työn kanssa. Testaaaja vastasi olleensa jo kauan aikaa jumissa kyseisessä kohdassa. Huomasin, että hän oli käyttänyt laskemiseen oikean näköistä lauseketta, mutta siitä puuttuivat ainakin arvot suureille η ja ϕ .

Testaaaja kysyi, tarvitaanko laskussa pseudorapiditeetin lauseketta $\eta = -\ln(\tan(\frac{\theta}{2}))$. Vastasin, että tarvetta tämän lausekkeen käyttöön ei ole ja kerroin vinkkinä, että arvot suureille η ja ϕ voi päätellä edellisessä osiossa tulostettujen datatiedoston rivien avulla. Tämän jälkeen testaaaja sai invarianttien massojen laskun toimimaan.

Kello 15:05 huomasin testaaajan T3 jumiutuneen hetkeksi osioon, jossa luodaan histogrammi lasketuista invariantin massan arvoista (ks. alaluku 4.2.4). En reagoinut tähän, vaan palasin muutaman kerran uudestaan tarkastelemaan tilannetta. Hetken kuluttua testaaaja oli edennyt kohdasta itsenäisesti eteenpäin.

Puolen tunnin päästä tästä, noin kello 15:38, testaaaja T6 sai työn valmiiksi. Ohjasin hänet täyttämään verkkokyselylomaketta. Tämän jälkeen kello 15:46 tilan internet-yhteys kaatui jälleen. Tämän seurauksena testaaajan T6 lomakkeeseen jo täytetyt vastaukset katosivat, sillä hän ei ollut päässyt vielä lomakkeen loppuun asti tallentamaan vastauksia. Kello 15:49 verkkoyhteys palasi takaisin. Testaaaja T6 kertoi pitävänsä hieman turhauttavana sitä, että hän joutuisi taas kirjoittamaan lomakkeen vastaukset uudestaan. Vastasin hänelle, että olisi mainiota, jos hän tästä huolimatta jaksaisi kirjoittaa samat asiat, jotka hän oli lomakkeelle kirjoittanut ennen yhteyden kaatumista.

Tässä vaiheessa huomasin, että verkkoyhteyden häiriöt saattoivat johtua yliopiston IT-palveluiden etukäteen tiedottamasta runkoverkon huoltotyöstä. En ollut huomannut ennen testauksia tätä. Kello 15:52 testaaaja T2 sai työn valmiiksi ja siirtyi täyttämään lomaketta. Hän sai lomakkeen valmiiksi kello 16:04. Testaaaja T6 taas palautti vastauksensa kello 16:13, jonka jälkeen testitilaisuus päättyi. Lopuksi siirsin ruutukaappausvideot itselleni talteen ja poistin ne testaaajien koneilta.

Toisen testitilaisuuden testaaajat käyttivät työhön aikaa noin 1 tunnin ja 55 minuuttia sekä noin 2 tuntia ja 23 minuuttia. Nämä ajat sisälsivät vain työn tekemiseen kuluneen ajan. Ajoista on vähennetty edellä mainitun videoiden väärän tallennuspaikan aiheuttama tauko.

Seuraavassa alaluvussa kerron tarkemmin testitilaisuuksissa käytetystä verkkokyselylomakkeesta. Jokainen testaaaja täytti lomakkeen välittömästi työn valmistuttua ennen poistumistaan testitilasta.

5.3.4 Verkkokyselylomake

Heti työn tekemisen jälkeen täytettävällä kyselylomakkeella tavoittelin välitöntä palautetta ja tunteuksia työstä. Lomaketta täyttäessä testajat hyvin todennäköisesti muistivat tarkasti työn aikana tapahtuneet asiat. Lisäksi kirjallisesti annettavat vastaukset voivat olla rehellisempiä kuin haastattelussa kasvotusten kerrotut. Haastattelussa testajat voivat mahdollisesti kaunistella palautetta, sillä he tietävät minun toimineen myös työn tekijänä.

Tein kyselylomakkeen Google Forms -palvelun avulla. Lomakkeen ensimmäisellä sivulla (ks. kuvio 14) kysyin vastaajan nimen, jotta voisin yhdistää lomakkeen muuhun vastaajasta kerättyyn aineistoon. Kysyin myös, oliko vastaaja käyttänyt TIM-alustaa aikaisemmin ennen testitulannetta.

Edellä mainittujen esitietojen jälkeen esitin lomakkeessa yhteensä kahdeksan sanallista kehoitusta tai kysymystä. Jaoin jokaisen näistä omalle lomakkeen sivulle, jotta vastaaja voisi keskittyä rauhassa yhteen osaan kerrallaan. Esitin lomakkeella seuraavat kohdat:

- Kerro omin sanoin, mitä teit työssä.
- Kerro omin sanoin työn sujuvuudesta.
- Miten motivoivaksi koit työn?
- Mitä opit työstä?
- Miten olisit voinut oppia lisää työstä?
- Mitä haasteita kohtasit työtä tehdessä?
- Mitä muuttaisit työstä?
- Jäikö jotakin, mitä haluat sanoa työstä tai testaamisesta ylipäänsä?

Lomakkeen täytön ajan vastaajat näkivät alareunassa etenemistä kuvaavan palkin sekä lomakkeen sivujen määrän. Lopuksi lomakkeen lähetyksen jälkeen vastaukset tallentuivat tarkasteltavakseni Google Formsiin. Kerron lomakkeen vastauksista ja niiden analysoinnista tarkemmin alaluvussa 5.5, jossa kuvailen aineiston analysointia.

5.3.5 Haastattelut

Testausprosessin viimeisessä vaiheessa haastattelin jokaisen työn tehneen testaaajan. Toteutin haastattelut teemahaastatteluina (ks. alaluku 5.1.1). Sovin jokaisen testaaajan kanssa erikseen sähköpostitse haastatteluajan. Kuusi testaaajaa (T1, T2, T4, T7, T8, T9) osallistui haastatteluihin kolmen päivän päästä työn tekemisestä. Loput kolme osallistujaa (T3, T5, T6) varasivat haastatteluaikinsa myöhemmälle ajankohdalle, osallistujat T3 ja T6 päivän päähän toisesta testitilaisuudesta ja osallistuja T5 vajaan kahden viikon päähän ensimmäisestä tilaisuudesta.

Varasin jokaista haastattelua varten fysiikan laitokselta huoneen, jossa tilaisuuden pystyi järjestämään rauhassa. Haastattelujen nauhoittamiseen käytin kahta erilaista nauhoitusvälinettä, puhelinta ja ääninauhuria. Kahden laitteen käyttö osoittautui hyödylliseksi, sillä yhden testaaajan kohdalla unohdin laittaa varsinaisen nauhurin päälle. Sain kuitenkin haastattelun talteen puhelimen tallennuksesta.

Jokaisen haastattelun alussa toivotin testaaajan tervetulleeksi ja ohjasin hänet istumaan vastapäätä minua. Kerroin, että tallennan haastatteluista ääniraidan ja laitoin nauhurit päälle. Tämän jälkeen avasin testaaajalle haastattelun ideaa kertomalla, että tarkoituksena on kerätä aineistoa tehtyyn työhön liittyen kysymysten ja keskustelun kautta. Puhuin lisäksi tutkimukseni tavoitteista.

Kerroin myös muutamia asioita, joilla pyrin parantamaan haastatteluiden luotettavuutta. Avasin tulevan haastattelun kulkua kertomalla, että kyseessä on puoli-strukturoitu haastattelu. Mainitsin, että haastattelussa minulla on valmiina sarja kysymyksiä, mutta niiden lisäksi kumpikin voi kysyä muita mieleen tulevia asioita. Lisäksi muistutin, että testaaaja voi unohtaa roolini työn tekijänä ja kehittäjänä, ja sanoa haastattelussa sekä hyviä että huonoja puolia työstä. Kehotin myös haastateltavaa olemaan mahdollisimman rennosti ja keskustelemaan normaalisti, vaikka tilanne saattaisikin tuntua viralliselta esimerkiksi nauhureiden takia. Nämä seikat mainitsemalla pyrin saamaan mahdollisimman aitoja vastauksia.

Varsinaisissa haastatteluissa käytin seuraavaa runkoa:

- Kerro omin sanoin mitä teit työssä.
- Koetko, että työ kehitti kokeellisen työskentelyn taitojasi? Millä tavalla?
- Muuttuivatko tietosi hiukkasfysiikasta? Millä tavalla?
- Muuttuivatko ohjelmointitaitosi? Millä tavalla?
- Muuttuivatko tietosi tieteen luonteesta ja tieteellisestä prosessista? Millä tavalla?
- Miten motivoivaksi koit työn?
- Kerro työn sujuvuudesta.
- Mitä haasteita kohtasit työtä tehdessä?

- Mitä opit työstä?
- Miten olisit voinut oppia lisää työstä?
- Mitä muuttaisit työstä?
- Kuinka vaikea työ mielestäsi oli?
- Jäikö jotakin sanomatta tai tuliko muita huomioita?

Osassa haastatteluja esitin myös rungon ulkopuolisia tarkentavia kysymyksiä sekä lisäkysymyksiä, jotka muodostin haastattelujen aikana. Lisäksi seitsemän haastattelurungon kohdalla lisäsin jo valmiiksi runkoon yhden tai useamman tarkentavan kysymyksen tai huomion erilaisten havainnoinnissa ja kyselylomakkeessa esiin nousseiden asioiden perusteella. Esimerkiksi testaaajan T3 kohdalla huomasin havainnoinnissa, että hän oli jäänyt jumiin työn yhteen kohtaan noin puolen tunnin ajaksi. Tästä syystä lisäsin hänen haastattelurunkoonsa kysymyksen ”*Miltä tuntui, kun olit työssä jumissa 30 minuuttia?*”

Haastattelujen kesto vaihteli väliltä 7 – 40 minuuttia. Osa testaaajista vastasi kysymyksiin lyhyesti eikä esittänyt lisäkysymyksiä, kun taas osassa haastatteluja syntyi pidempää keskustelua. Huomasin haastattelujen jälkeen, että ensimmäisessä tilaisuudessa (testaaja T2) jännitin hieman rooliani haastattelijana. Tuntui, että tätä seuraavat haastattelut sujuivat luontevammin saadessani kokemusta haastattelijana toimimiseen.

Haastatteluiden jälkeen testauksen toisen vaiheen testausprosessi oli ohi. Jokainen testaaaja suoritti koko prosessin loppuun asti. Kerron testausprosessissa keräämäni aineiston käsittelystä ja analysoinnista tarkemmin alaluvussa 5.5. Seuraavassa alaluvussa kuvailen käyttämäni analyysimenetelmien, *sisällönanalyysin* ja *sisällön erittelyn*, teoriaa ja analyysin kulkua.

5.4 Sisällönanalyysi ja sisällön erittely

Tuomen ja Sarajärven [14] mukaan sisällönanalyysi on terminä monimerkityksinen. Sisällönanalyysi voi viitata yksittäiseen analyysimetodiin, jolla ”– pyritään saamaan tutkittavasta ilmiöstä kuvaus tiivistetyssä ja yleisessä muodossa.” [14, s. 117]. Lisäksi sisällönanalyysilla voidaan tarkoittaa ”– kirjoitettujen, kuultujen tai nähtyjen sisältöjen analyysia väljänä teoreettisena kehyksenä.” [14, s. 103].

Tässä tutkimuksessa viitataan sisällönanalyysilla yksittäiseen analyysimetodiin, jota käytin kerätyn aineiston analysoinnissa. Sisällönanalyysi mahdollistaa erilaisten dokumenttien systemaattisen ja objektiivisen analysoinnin. Dokumentteina voidaan pitää lähes mitä tahansa kirjalliseen muotoon saatettua materiaalia. Sisällönanalyysin avulla tutkittava ilmiö pyritään saamaan kuvattuna tiivistetyssä ja yleisessä muodossa johtopäätöksien tekemistä varten. Analyysilla pyritään lisäämään aineiston informaatioarvoa luomalla selkeyttä siihen. [14, s. 117, 122]

Sisällönanalyysin lisäksi sovelsin aineistoon sisällön erittelyä. Tuomi ja Sarajärvi [14, s. 199] tarkoittavat sisällön erittelyllä ”– – dokumenttien analyysia, jossa kuvataan kvantitatiivisesti esimerkiksi tekstin sisältöä.” Sisällön erittely eroaa sisällönanalyysistä siis siten, että ensimmäisenä mainitussa kuvaus on määrällistä ja jälkimmäisessä sanallista [14, s. 199].

Sisällönanalyysi voidaan jakaa kolmeen eri tyyppiin: aineistolähtöiseen, teoriaohjajaan ja teorialähtöiseen sisällönanalyysiin [14, s. 121]. Toteutin aineiston analyysin tässä tutkimuksessa teorialähtöisesti. Teorialähtöinen sisällönanalyysi haastatteluaineistoon muokattuna etenee seuraavan rungon mukaisesti [14, kuvio s. 123, muokattu]:

1. Analyysirungon muodostaminen
2. Haastattelujen kuunteleminen ja aukikirjoitus sana sanalta
3. Aineiston lukeminen ja sisältöön perehtyminen
4. Pelkistettyjen ilmausten etsiminen ja alleviivaaminen
5. Pelkistettyjen ilmausten listaaminen
6. Samankaltaisuuksien ja erilaisuuksien etsiminen pelkistetyistä ilmauksista
7. Pelkistettyjen ilmausten ryhmittely tai yhdistäminen ja alaluokkien muodostaminen
8. Alaluokkien yhdistäminen analyysirungossa määriteltyihin yläluokkiin

Ensimmäisenä tapahtuvalla analyysirungon muodostamisella tarkoitetaan sitä, että analyysia varten luodaan runko, joka sisältää erilaisia luokituksia tai kategorioita [14, s. 128]. Runko perustuu johonkin aiempaan käsitejärjestelmään [14, s. 127]. Tässä tutkimuksessa perustin analyysirunkoni Nivalaisen, Asikaisen ja Hirvosen [6] esittämiin fysiikan kokeellisten töiden tavoitekategorioidiin (ks. alaluku 3.7). Analyysirunko on kuvattu taulukossa 2. Kirjasin runkoon myös kolme luokitusta, joilla tavoittelin lisätietoa työn toimivuudesta. Nämä luokitukset ovat vaakaviivalla erotettuna taulukon lopussa.

Analyysirungossa kirjattuna ovat siis ainoastaan yläluokat, jotka perustuvat Nivalaisen ym. kategorisointiin sekä yleisesti työn toimivuuteen. Sarakkeet *Alkuperäinen ilmaus*, *Pelkistetty ilmaus* ja *Alaluokka* ovat mukana havainnollistamassa analyysin etenemistä. Analyysissa sarakkeisiin luodaan sisältöä lähtien alkuperäisistä ilmauksista ja päätyen alaluokkien muodostamiseen. Lopuksi eri alaluokat ryhmitellään sopiviin yläluokkiin. Yhteen yläluokkaan voi siis kuulua useita eri alaluokkia.

Ennen kuin aineistosta voidaan lähteä etsimään edellä mainittuja alkuperäisiä ilmauksia, pelkistettyjä ilmauksia ja alaluokkia, tulee aineisto kirjoittaa auki ja perehtyä sen sisältöön. Nämä vaiheet on merkitty edellä listattuun analyysin etenemiseen numeroilla 2) ja 3). Lisäksi analyysia varten on määritettävä analyysiyksikkö, joka voi olla esimerkiksi yksittäinen sana, kirjain, lause tai useita lauseita käsittä-

Taulukko 2. Tutkimuksen analyysirunko. Ensimmäiset viisi yläluokkaa muodostuvat Nivalaisen, Asikaisen ja Hirvosen [6] määrittelemistä fysiikan käytännön töiden tavoitekatgorioista. Loput kolme alaluokkaa muodostettiin tutkimustehtäviin perustuen.

Yläluokka	Alkuperäinen ilmaus	Pelkistetty ilmaus	Alaluokka
Käytännöllisten tai kokeellisten taitojen kehittäminen			
Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen			
Motivaation lisääminen			
Tieteen luonteen ja tieteellisen prosessin ymmärtäminen			
Sosiaalisten ja oppimisen taitojen parantaminen			
Työssä kohdatut haasteet			
Työssä kehitettävää			
Työssä onnistunutta			

vä ajatuskokonaisuus. Käytin analyysiyksikkönä ajatuskokonaisuutta, johon sisältyi yksi tai useampi lause. [14, s. 122–123]

Keräsin työn toisessa testauksessa aineistoa havainnoimalla, ruutukaappausvideoilla, TIMiin tallentuneilla testaaajien vastauksilla, kyselylomakkeella sekä haastatteluilla. Analysoin näistä tarkemmin sisällönanalyysin ja sisällön erittelyn keinoin kyselylomakkeen vastauksia ja testaaajien haastatteluja. Muuta osaa aineistosta käytin tiedon täydentämiseen ja varmistamiseen. Esimerkiksi ennen haastatteluita tutustuin testaaajien kyselylomakkeen vastauksiin sekä TIMissä tehtyihin ajoihin. Tarvittaessa käytin ruutukaappausvideota varmistaakseni, mitä testaaaja oli työssä tehnyt. Näiden avulla lisäsin osiin haastattelurunkoja valmiiksi tarkentavia kysymyksiä.

Analysoitavien kyselylomakkeen vastausten ja haastattelujen käsittelyn aloitin lomakkeen vastauksiin tutustumalla. Tämän jälkeen litteroin haastattelunauhut. Tällä tapaa sain haastattelujen sisällön helpommin analysoitavaan muotoon. Litteroin haastattelut suhteellisen tarkasti kirjaamalla ylös kaikki puheessa esiintyneet sanat. En kirjannut litterointeihin kaikkia äännähdyksiä tai merkinnyt sanojen painotuksia. Merkitsin litteroinneissa omia puheitani kirjaimella M ja haastateltavien puhetta kirjaimella H. Esimerkiksi yksi haastattelun osio näyttää litteroinnissa tältä:

”M: Ja tota mulla on tässä tosiaan sarja kysymyksiä, et tää on semmonen puolistrukturoitu haastattelu et sit mä saatan kysyä myös jotain jos tulee

mieleen ja sä voit kans kysyä multa että niinku.

H: Juu.”

Kyselylomakkeen vastaukset kopioin Google Forms -palvelusta käsiteltäväksi tekstinkäsittelyohjelmaan. Nämä testaaajien avoimiin kysymyksiin kirjoittamat vastaukset olivat jo valmiiksi kirjallisessa muodossa.

Litterointien valmistuttua aloin etsiä aineistosta analyysirunkoon liittyviä ilmaisuja. Aloitin prosessin haastatteluista etsimällä erikseen jokaiseen Nivalaisen ym. [6] esittämiin fysiikan kokeellisten töiden tavoitekategorioidiin liittyviä ilmaisuja. Käytännössä toteutin tämän ottamalla PDF-muotoisesta litterointitekstistä viisi kopiota ja yliviivaamalla näistä yhteen aina kuhunkin kategoriaan liittyvät ilmaisut. Lisäksi etsin erikseen ilmaisuja, joissa mainittiin työssä kohdattuja haasteita, kehittämistä kaipaavia asioita sekä onnistumisia koskevia seikkoja. Kyselylomakkeen vastauksia pidin suoraan alkuperäisinä ilmauksina, sillä testaaajat olivat vastanneet kysymyksiin jo valmiiksi tiiviisti.

Alkuperäisten ilmaisujen etsimisen jälkeen muodostin niistä pelkistettyjä ilmaisuja, eli siirryin analyysiprosessin neljänteen vaiheeseen. Pelkistettyjen ilmaisujen avulla alkuperäiset ilmaisut saadaan tiiviimpään ja helpommin tulkittavaan muotoon. Esimerkiksi yhdessä litteroinnissa yliviivasin seuraavan alkuperäisen ilmauksen liit-tyen tavoitekategoriaan ”Käytännöllisten tai kokeellisten taitojen kehittäminen”. Lainauksessa ”– –” -merkinnällä viitataan epäolennaisen tekstin pois jättämiseen.

”H: No joo, en mä nyt ollut ihan niin tarkkaan tiennyt miten ne CERNissä vähän niinku vaikka tämmösen bosonin olemassaolon todistaminen, niin että minkälaila se tulee – –. Että kyl se tota vähän niinku avas sitä maailmaa siitä, että minkälaila sitä hiukkasfysiikkaa voi tehdä.

M: Mmm.

H: Ja minkä, minkälaisella perusteella sit voidaan justiin tota jotain uusia hiukkasia sitten olemassaoloa todistaa.”

Tämän ilmauksen tiivistin pelkistetyksi ilmaukseksi ”*Työ avasi hiukkasten löytymisen periaatetta.*”. Pelkistin samalla tapaa myös kyselylomakkeen vastaukset. Listasin pelkistetyt ilmaukset uuteen tekstitiedostoon kategorioittain tai aihealueittain. Tämä oli siis analyysiprosessin viides vaihe.

Kuudennessa ja seitsemännessä vaiheessa etsin pelkistetyistä ilmauksista samankaltaisuuksia ja eroavaisuuksia sekä muodostin ilmauksista erilaisia alaluokkia. Yksittäinen pelkistetty ilmaus saattoi sopia useampaan eri alaluokkaan. Esimerkkinä alaluokkien muodostamisesta kyselylomakkeen kysymyksen ”Mitä opit työstä?” vastauksien pelkistetyt ilmaukset ja niistä muodostetut alaluokat on kirjattu tauluk- koon 3.

Analyysin lopuksi yhdistin aineistosta nousseet alaluokat analyysirungossa määri- tettyihin yläluokkiin (ks. taulukko 2). Osaan yläluokista kirjasin myös alaluokkia

Taulukko 3. Esimerkki alaluokkien muodostamisesta lomakkeen kysymyksen ”Mitä opit työstä?” pelkistetyistä ilmauksista.

Pelkistetty ilmaus	Alaluokka
Invariantin massan käyttö ja sovellukset. Hieman lisää Python-kieltä.	hiukkasfysiikan teoria, ohjelmointi
Kertausta aiemmin käsitellystä asiasta.	ei uutta
Kiihdyttimien toimintaperiaatteesta lisää. Vähän data-analyysia.	hiukkasfysiikan teoria, data-analyysi
Hieman ohjelmointia. Z:n olemassaolon todistaminen.	ohjelmointi, hiukkasten havaitseminen
Hiukkasten etsiminen massajakaumasta.	hiukkasten havaitseminen
Hieman kokeellista hiukkasfysiikkaa.	kokeellinen hiukkasfysiikka
Itsehillintää. Hiukkasten ominaisuuksien määrittämisen teoriaa ja laskennan käytäntöä.	itsehillintä, hiukkasfysiikan teoria
Hiukkasfysiikka ja Python olivat uusia asioita.	hiukkasfysiikka, ohjelmointi
Kuinka CERNissä tehdään tutkimusta ja miten hiukkaa voi havaita. Ohjelmoinnissa ei mitään uutta.	kuinka tutkitaan, hiukkasten havaitseminen

yhdistäviä tekijöitä, jotka helpottivat alaluokkien jäsentämistä. Seuraavassa alaluovussa kerron tarkemmin yläluokittain analyysin tuloksista.

5.5 Analyysin tulokset

Tässä alaluvussa kuvailen aineistoon tekemäni sisällönanalyysin ja sisällön erittelyn tulokset. Luku alkaa esitietolomakkeen vastausten kuvailusta. Tämän jälkeen esitän analyysin tulokset analyysirungon yläluokkien mukaisessa järjestyksessä.

5.5.1 Esitietolomake

Kaikki yhdeksän testatajaa täyttivät ennen työn tekemistä esitietolomakkeen (LIITE C). Lomakkeen vastauksilla pyrin kartoittamaan tiedonantajien lähtötietoja ja taustoja. Näiden tietojen avulla pystyin arvioimaan, kuinka hyvin tiedonantajien joukko sopi tutkimukseen.

Lomakkeen alussa kysyttiin osallistujan pääainetta, sivuaineita sekä pääaineen opiskeluaikaa lukuvuosissa. Kaikki testatajat opiskelivat pääaineenaan fysiikkaa. Opiskeluaikat vaihtelivat kahden ja kahdeksan vuoden välillä, keskiarvon ollessa noin 3,7 vuotta. Sivuaineina testatajat opiskelivat matematiikkaa (7), kemiaa (7), tietotekniikkaa (3), kasvatustiedettä (2) sekä solu- ja molekyylibiologiaa (1).

Lomake jatkui kolmella monivalintakysymyksellä, jotka käsittelivät vastaajien osallistumista kolmelle eri fysiikan kurssille. Kysymykset olivat muotoa ”Olen käynyt kurssin – –”, ja vastausvaihtoehdot olivat *kyllä*, *en* sekä *en ole varma*. Kurssin FYSP160: Moderni fysiikka oli käynyt jokainen vastaajista. Lähes kaikki (8) olivat käyneet myös kurssin FYSP120: Numeeriset menetelmät. Ainoastaan yksi (1) vastaaja valitsi vaihtoehdon *en*. Kurssin FYSH300: Hiukkasfysiikka taas oli käynyt vain yksi (1) vastaajista. Loput (8) vastasivat vaihtoehdolla *en*.

Aikaisemmin käytyjen kurssien perusteella tiedonantajien joukko osoittautui hyvin tutkimukseen sopivaksi. Testattava laboratoriotyö tulee käyttöön kurssilla FY-SA2002 Moderni fysiikka, osa B, joka ajoittuu fysiikan aineopintoihin. Perusopintovaiheessa ei vielä tutustuta tarkemmin hiukkasfysiikkaan. Tästä syystä on hyvä, että testaaajista ainoastaan yksi oli osallistunut aiemmin hiukkasfysiikan kurssille. Myös se, että suurin osa testaaajista oli käynyt numeerisia menetelmiä käsittelevän kurssin FYSP120, oli testaamisen kannalta hyvä asia. Tutkimuksen tekoheikellä fysiikan laitos otti käyttöön uuden opetussuunnitelman, jossa lisättiin aineopintovaiheessa Python-ohjelmointikielen käyttöä. On siis mahdollista, että laboratoriotyötä tulevaisuudessa tekevillä opiskelijoilla on jo hieman taustaa ohjelmointiin tai ohjelmalliseen ajatteluun. Täten on hyvä, että myös tiedonantajilla suurimmasta osasta oli kokemusta numeerisesta laskennasta, jossa hyödynnetään ohjelmallista ajattelua.

Seuraavana vastaajilta kysyttiin 5-portaisella Likertin asteikolla heidän tuntemuksiinsa ohjelmointitaidoistaan. Kysymys oli muodossa ”Olen mielestäni hyvä ohjelmoimaan.” Vastausvaihtoehdot ja kunkin vaihtoehdon valinneiden testaaajien lukumäärät olivat seuraavat:

- samaa mieltä (0)
- osittain samaa mieltä (4)
- ei samaa eikä eri mieltä (1)
- osittain eri mieltä (2)
- eri mieltä (2)

Tiedonantajien kokemukset ohjelmointitaidoistaan olivat vaihtelevat. Tiedonantajien joukkoa voidaan pitää tältä osin satunnaisena otantana, joka sopii tutkimukseen hyvin. Valmista työtä tulee tekemään erilaisia opiskelijoita, joilla on todennäköisesti monenlaisia ohjelmointitaitoja. Tiedonantajien kokemukset ohjelmointitaidoistaan kuvaavat hyvin tätä.

Toiseksi viimeisenä kohtana lomakkeessa oli Likert-asteikollinen väittämä ”Hiukkasfysiikka on minulle tuttua.” Väittämän vastausvaihtoehdot olivat samat kuin edellisessä väittämässä. Tähän testaaajat vastasivat seuraavasti:

- samaa mieltä (1)
- osittain samaa mieltä (3)

- ei samaa eikä eri mieltä (4)
- osittain eri mieltä (1)
- eri mieltä (0)

Vastaukset painottuivat siihen suuntaan, että hiukkasfysiikka oli ollut testaaajille jo aiemmin tuttua. Tällä ei ole suurta merkitystä tutkimuksen kannalta. Hiukkasfysiikka terminä kattaa sekä teoreettisen että kokeellisen hiukkasfysiikan, joten väittämäs- tä ei voi suoraan päätellä testaaajien kokemusta työssä käsiteltävästä kokeellisesta hiukkasfysiikasta. Lisäksi kysymyksessä en varsinaisesti korostanut sitä, kysyttiinkö testaaajien hiukkasfysiikan tuntemusta ennen vai jälkeen pohjaluennon. Kyselyloma- ke täytettiin pohjaluennon jälkeen, jolloin luento saattoi vaikuttaa testaaajien koke- muksiin siitä, kuinka tuttua hiukkasfysiikka heille oli. Olin siis muotoillut tämän kysymyksen hieman huonosti.

Viimeisenä lomakkeessa esitettiin Likertin asteikolla varustettu väittämä ”Olen oh- jelmoinut Python-ohjelmointikielellä.” Tähän testattavien vastaukset jakaantuivat seuraavalla tavalla:

- erittäin usein (0)
- usein (0)
- jonkin verran (0)
- muutamia kertoja (3)
- en ollenkaan (6)

Vastausten perusteella tiedonantajat olivat käyttäneet Python-kieltä vain muutamia kertoja tai eivät ollenkaan. Tutkimukseen tämä vaikuttaa kahdella eri tavalla. Tie- donantajien joukko ei Python-taitojen osalta todennäköisesti simuloi täysin tulevia työn tekeviä opiskelijoita, sillä fysiikan laitoksen kursseilla tullaan käsittelemään Python-ohjelmointia. Toisaalta testaamisen kannalta osallistujien kokemattomuus Pythonin kanssa on hyvä asia. Testauksesta voidaan tämän perusteella saada tietoa siitä, miten kokematon Pythonin käyttäjä suoriutuu työn tekemisestä.

Esitietolomakkeen vastausten perusteella voidaan todeta tutkimuksen tiedonanta- jien soveltuvan tutkimuksen tarkoituksiin. On mahdollista, että tiedonantajat ovat normaalia aktiivisemmin opetukseen osallistuvia opiskelijoita, sillä he ilmoittau- tuivat vapaaehtoisesti mukaan testaamiseen. Toisaalta testauksesta tarjottu yhden opintopisteen palkinto pienentää hieman tätä mahdollisuutta. Tiedonantajat eivät tarjoa tietoa siitä, miten kehitettävä työ soveltuu fysiikan sivuaineopiskelijoille, sillä kaikkien testaaajien pääaineena oli fysiikka. Tiedonantajat edustavat kuitenkin hy- vin fysiikan aineopinto-opiskelijoita, joille laboratoriotyö laaditaan. Myös tiedonan- tajiin taustatiedot hiukkasfysiikasta ja ohjelmoinnista simuloivat hyvin työn kohde- ryhmää.

Seuraavissa alaluvuissa esittelen kyselylomakkeen vastausten ja haastattelujen sisäl-

lönanalyysistä sekä sisällön erittelystä saamani tulokset. Käyn tulokset läpi analyysirungon (ks. taulukko 2) yläluokkien mukaisessa järjestyksessä aloittaen viidestä fysiikan kokeellisten töiden tavoitteesta.

5.5.2 Fysiikan kokeellisten töiden tavoitteet

Ensimmäinen Nivalaisen ym. [6] nimeämä fysiikan kokeellisten töiden tavoitekategoria on ”Käytännöllisten tai kokeellisten taitojen kehittäminen”. Tällä tarkoitetaan taitojen ja kykyjen kehittymistä mm. hypoteesien muodostamisessa, kokeiden suunnittelussa, datan tulkitsemisessa ja virheiden käsittelyssä [6, s. 2]. Tähän analyysin yläluokkaan liittyvät aineistosta nousseet alaluokat on esitetty taulukossa 4. Kunkin alaluokan perässä oleva numero ilmaisee, kuinka monta kertaa haastatteluissa tai lomakevastauksissa mainittiin kuhunkin alaluokkaan liittyvä ilmaisu.

Taulukko 4. Yläluokkaan ”Käytännöllisten tai kokeellisten taitojen kehittäminen” liittyvät alaluokat haastatteluissa ja kyselylomakevastauksissa. Suluisissa oleva numero kertoo, kuinka monta kertaa kuhunkin alaluokkaan liittyvä ilmaisu mainittiin.

Haastattelut	Kyselylomake
tunnistamisprosessi (10)	tunnistamisprosessi (3)
data-analyysi (6)	ohjelmointi (3)
ohjelmoinnin syntaksinen oppi (4)	data-analyysi (1)
fysiikon työ (2)	
numeriikka (1)	
ohjelmointitaidot (1)	

Haastatteluissa taulukon 4 alaluokat liittyivät usein esittämäni kysymyspariin ”Koetko, että työ kehitti kokeellisen työskentelyn taitojasi? Millä tavalla?”. Käytännöllisistä tai kokeellisista taidoista mainittiin eniten kehittyneen tunnistamisprosessiin liittyviä asioita. Myös kyselylomakkeen vastauksista nousi kolme ilmaisua tunnistamisprosessiin liittyvästä oppimisesta vastauksena lomakkeen kysymykseen ”Mitä opit työstä?” Alaluokan *tunnistamisprosessi* nimellä viitataan hiukkasten havaitsemiseen tai niiden olemassaolon osoittamiseen liittyviin asioihin. Tähän sisältyvät esimerkiksi työssä Z -bosonin olemassaolon todistamiseksi tehdyt eri vaiheet (invariantin massan laskeminen, histogrammin laatiminen, histogrammin analysointi). Taulukossa 5 on esimerkkinä kaksi tähän alaluokkaan liittyvää ilmausta.

Työ kehitti myös testaaajien taitoja *data-analyysiin* liittyen. Näitä ilmaisuja mainittiin haastatteluissa yhteensä kuusi kertaa ja kyselylomakkeessa kerran. Data-analyysiin liittyivät esimerkiksi taulukossa 6 esitetyt ilmaisut.

Tunnistamisprosessiin ja data-analyysiin liittyvien asioiden kehittyminen on työn onnistumisen kannalta hyvä asia. Laboratoriotyön alussa (LIITE A-2) mainitaan, että

Taulukko 5. Kaksi esimerkkiä alaluokkaan ”tunnistamisprosessi” liittyvistä alkuperäisistä ja pelkistetyistä ilmauksista.

Alkuperäinen ilmaus	Pelkistetty ilmaus
H: – – näki sitä, että mitenkä mitenkä niinku oikeesti kokeellisesti tää asia on havaittu niin se havainnollisti aika hyvin tai tälleen että.	Työ havainnollisti sitä, miten Z-bosoni on kokeellisesti havaittu.
M: – – Aaa mitä sä opit tästä työstä? H: – – nimenomaan se, et millai sä niinku ku sä etit jotain hiukkasta niin millai sä sen niinku tunnistat sieltä ni ihan tämmöst ihan konkreettista.	Oppi hiukkasten tunnistamiseen liittyviä käytännön asioita.

tavoitteena on ”– – tehdä yksinkertaista kokeellista hiukkasfysiikan data-analysis – –” sekä tehtävänä ”– – todeta Z-bosonin esiintyminen, selvittää Z:n massa analyysin perusteella ja verrata saatua massaa Particle Data Groupin (<http://pdg.lbl.gov/>) arvoon.” Taulukon 4 perusteella testajat kokivat näihin liittyvien alaluokkien kehittyneen.

Muita kehittyneitä käytännöllisiä tai kokeellisia taitoja olivat mm. numeeriset taidot, fyysikon työn hahmottaminen sekä ohjelmointitaidot. Jaoin haastatteluissa mainitut ohjelmointitaidon kehittymiset vielä kolmeen eri luokkaan testajien kertoman perusteella. Vastaukset jakautuivat seuraavasti:

- ei muutosta (5 vastaajaa)
- vain syntaksista oppia (3 vastaajaa)
- kehittyivät (1 vastaaja)

Vain syntaksisella opilla viitataan tilanteisiin, joissa testaja mainitsi joko suorasti tai epäsuorasti ainoastaan ohjelmointikielen syntaksin ymmärtämisen tai koodin lukemisen kehittyneen. Jaottelun perusteella työ olisi voinut kehittää paremmin testajien ohjelmointitaitoja. Pääosin testajat kokivat ohjelmointitaitojensa pysyneen ennallaan tai kehittyneen vain syntaksisella opilla.

Sisällönanalyysillä muodostettujen alaluokkien lisäksi sain tietoa testajien kokemuksista käytännöllisten tai kokeellisten taitojen kehittymisestä haastattelurungossa olleen kysymyksen avulla. Esitin jokaiselle haastateltavalle kysymyksen ”Koetko, että työ kehitti kokeellisen työskentelyn taitojasi?” Tähän luokittelin vastaukset seuraavasti:

- ei muutosta (3 vastaajaa)
- kehittyivät vähän (3 vastaajaa)
- kehittyivät (2 vastaajaa)
- ei osaa sanoa (1 vastaaja)

Taulukko 6. Kaksi esimerkkiä alaluokkaan ”data-analyysi” liittyvistä alkuperäisistä ja pelkistetyistä ilmauksista.

Alkuperäinen ilmaus	Pelkistetty ilmaus
<p>M: Koetsä että tää työ kehitti sun kokeellisen työskentelyn taitoja? H: – – ehkä jonkun verran semmosta niinku data-analyysitekemistä totta kai vähän harjoittelua siihen ni. Joo, kyllä. M: Haluutsä viel tarkentaa että niinku millä tavalla se niinku? H: Niin siis se, että – – mitenkä käsitellään dataa ja miten sitä voi näin käsitellä ja sitten vähän että niinku tota plottaamista ja muuta vastaavaa.</p>	<p>Kehitti hieman datan analysointiin liittyviä taitoja.</p>
<p>M: – – et mitä sä opit tost työstä? H: – – sitten koodaamispuolelle että niinku tietokoneella pystyy tekemään aika helposti kuitenkin sitten tommosta tilastollista analyysia. Ja sinne saadaan sitte tota suuremman, tosi suuren joukon dataa ni hyvin nopeesti – –</p>	<p>Työ opetti, että tietokoneella voidaan tehdä helposti tilastollista analyysia suurelle datamäärälle.</p>

Vastausten perusteella pienen enemmistön käytännölliset tai kokeelliset taidot kehittyivät jollakin tavalla. Toisaalta työ ei saavuttanut tätä tavoitetta kolmen testaaajan kohdalla, jotka kertoivat etteivät taidot muuttuneet.

Toisena analyysin yläluokkana oli Nivalaisen ym. kategorian tavoite ”Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen.” Nimensä mukaisesti tällä tarkoitetaan käsitteisiin ja ns. ”teoreettiseen sisältöön” liittyviä asioita, joita usein käsitellään puhtaasti teoriavetoisessa opetuksessa [6, s. 2–3]. Työn aiheeseen liittyen tällaisia ovat esimerkiksi invariantin massan käsite, standardimalli sekä hiukkasfysiikan teoria yleisesti.

Analyysin toiseen yläluokkaan liittyvät haastatteluista esiin nousseet alaluokat on kuvattu taulukossa 7. Alaluokat on ryhmitelty niitä kuvaavien yhdistävien tekijöiden alle. Yhdistävät tekijät jakavat alaluokat *tietoa tarjonneisiin, opittuihin, kerrattuihin* ja *mahdollisesti selventäviin* luokkiin.

Taulukon 7 ensimmäisestä yhdistävästä tekijästä nähdään, että työ tarjosi testaa-
jien mielestä tietoa esimerkiksi hiukkasista, Z -bosonin määrittämisestä ja siitä, miten hiukkasfysiikan tutkimusta tehdään. Toisen yhdistävän tekijän alaluokkien mukaan työ myös opetti erilaisia tieteelliseen sisällön ymmärtämiseen ja käsitteelliseen ymmärtämiseen liittyviä asioita. Näistä eniten mainintoja tuli metodista, jolla invariantti massa tai hiukkasen olemassaolo voidaan selvittää.

Yksittäisten mainintojen perusteella työ myös kertasi analyysiprosessin kulkua sekä Z -bosonin kinematiikkaa. Lisäksi yksi haastateltava mainitsi, että työ voi selventää hiukkasten olemassaolon osoittamisen prosessia.

Taulukko 7. Yläluokkaan ”Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen” liittyvät haastatteluista nousseet alaluokat ja niitä yhdistävät tekijät.

Yhdistävä tekijä	Alaluokka
tietoa	hiukkasista (4)
	Z :n määrittämisestä (3)
	miten hiukkasfysiikkaa tehdään (2)
	pseudorapiditeetin hyödyntämisestä invariantin massan etsimisessä (1)
	CERNin laitteistosta (1)
opittuja asioita	massan ja energian ekvivalenssista (1)
	metodi invariantin massan tai hiukkasten löytämiseen (3)
	hiukkasfysiikan teoriaa yleisesti (2)
	invariantti massa (2)
	miten dataa voidaan käyttää (1)
	aaltofunktion symboli (1)
	piikit massajakaumassa tarkoittavat hiukkasia (1)
	pseudorapiditeetin hyödyntäminen (1)
standardimallin sisältö (1)	
kertausta	analysointiprosessista (1)
	Z :n kinematiikasta (1)
voi selventää	hiukkasen olemassaolon osoittamisen prosessia (1)

Tiivistetysti taulukosta 7 voi todeta, että haastattelujen perusteella työ kehitti monipuolisesti tieteelliseen sisältöön ja käsitteisiin liittyvää ymmärtämistä. Testaajat mainitsivat sekä yksittäisiä opittuja käsitteitä (esim. invariantti massa, aaltofunktion symbolin oppiminen) että yleisemmin hiukkasfysiikan teoriaan liittyviä asioita (esim. Z -bosonin määrittäminen, hiukkasfysiikan kokeellinen tutkimus).

Haastatteluissa esitin edellä käsitellyyn tavoitekategoriaan ”Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen” liittyvän kysymysparin ”Muuttuivatko tietosi hiukkasfysiikasta? Millä tavalla?” Jaoin testaajien vastaukset kolmeen eri luokkaan, jotka olivat *eivät*, *muuttuivat vähän* sekä *muuttuivat*. Vastaukset jakautuivat seuraavasti:

- eivät (4 vastaajaa)
- muuttuivat vähän (3 vastaajaa)
- muuttuivat (2 vastaajaa)

Pieni enemmistö, viisi testaajaa, ilmoitti tietojensa hiukkasfysiikasta muuttuneen jollakin tapaa. Neljä testaajaa koki tietojen pysyneen ennallaan. Vastausten perus-

teella työ voisi siis lisätä vielä enemmän testaaajien tietoa hiukkasfysiikasta.

Kyselylomakkeesta tähän kategoriaan nousi kolme alaluokkaa, jotka luokittelin yhdistävän tekijän *opittuja asioita* alle. Alaluokat on kirjattu taulukkoon 8. Taulukosta nähdään, että lomakevastausten perusteella opitut asiat liittyivät kaikki jollakin tapaa hiukkasfysiikkaan. Kolmen ilmaisun mukaan työssä oppi hiukkasfysiikan teoriaa, kun kaksi yksittäistä ilmaisua liittyivät kokeelliseen hiukkasfysiikan ja hiukkasfysiikan yleisesti oppimiseen.

Taulukko 8. Yläluokkaan ”Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen” liittyvät kyselylomakkeesta nousseet alaluokat ja niitä yhdistävät tekijät.

Yhdistävä tekijä	Alaluokka
	hiukkasfysiikan teoriaa (3)
opittuja asioita	kokeellista hiukkasfysiikkaa (1)
	hiukkasfysiikkaa (1)

Kolmas analyysin yläluokka oli nimeltään ”Motivaation lisääminen”. Tällä tarkoitetaan Nivalaisen ym. [6, s. 3] mukaan sitä, että käytännön töiden tulisi lisätä opiskelijan motivaatiota tieteen oppimiseen. Tähän liittyen käytännön töillä tavoitellaan kiinnostuksen, ilon ja huvun tunteiden luomista. Kategoriaan ei siis liity se, kuinka motivoituneita opiskelijat ovat tekemään yksittäistä fysiikan kokeellista työtä.

Esitin jokaisessa haastattelussa tähän liittyen kysymyksen ”Miten motivoivaksi koit työn?” Myöhemmin tajusin, että kysymys liittyy juuri edellä mainitsemaani työn tekemisen motivaatioon, joka ei kuulu kolmanteen tavoitekategoriaan. Jätin siis tuloksista pois suoraan työn tekemisen motivaatioon liittyneet vastaajien ilmaisut.

Haastatteluissa tuli ilmi myös itse kategoriaan liittyviä ilmaisuja. Taulukossa 9 on esitetty näistä pelkistettyjä ilmaisuja sekä niiden pohjalta muodostettuja alaluokkia. Tässä alaluokat kuvaavat sitä, mikä on testaaajan mukaan aiheuttanut motivaation kasvun tai kiinnostuksen tunteen.

Tulosten perusteella motivaation lisäämistä synnyttivät analyysissä käytetty data, tunnistamisprosessiin liittyvät asiat sekä itse työn tekeminen. Yksi testaaaja kertoi, että työ ei lisännyt motivaatiota muita asioita kohtaan. Yhdellä testaaajalla työ sai aikaan motivaation vähenemisen Python-ohjelmointia kohtaan.

Kyselylomakkeen vastauksista nousi kolme motivaation lisäämiseen liittynyttä pelkistettyä ilmaisua, jotka muodostivat kaikki oman alaluokkansa. Nämä ilmaisut ja alaluokat on kirjattu taulukkoon 10.

Haastattelujen ja kyselylomakkeen perusteella työ sai aikaan kokemuksia, jotka lisäsivät motivaatiota tieteen oppimista kohtaan, tai motivaatioon liittyviä tunteita. Toisaalta työ aiheutti myös motivaation laskua, joka on ei-toivottava ominaisuus. Verrattuna aiempiin CERNin avoimen datan opetuskokeiluihin (ks. alaluku 3.6),

Taulukko 9. Yläluokkaan ”Motivaation lisääminen” haastatteluista nousseet alaluokat ja pelkistetyt ilmaisut.

Alaluokka	Pelkistetty ilmaisu
data (2)	Muualla luodun datan käyttäminen oli mielenkiintoista. Oikean datan käyttäminen oli hyvä asia.
tunnistamisprosessi (2)	Hiukkasten löytämisen prosessi oli mielenkiintoinen. Z -bosonin olemassaolon todistaminen oli ”jännä” lopputulos.
työ (2)	Työ voi olla silmiä avaava toisen vuoden opiskelijoille. Työ voi herättää kiinnostuksen hiukkasfysiikkaa kohtaan.
muu (2)	Työ ei lisännyt motivaatiota muihin asioihin. Työ laski motivaatiota Python-ohjelmointiin.

Taulukko 10. Yläluokkaan ”Motivaation lisääminen” kyselylomakkeen vastauksista nousseet alaluokat ja pelkistetyt ilmaisut.

Alaluokka	Pelkistetty ilmaisu
kokeellinen hiukkasfysiikka (1)	Kokeellisen hiukkasfysiikan kertaus oli mukavaa.
data-analyysi (1)	Data-analyysi oli mielenkiintoista.
työ (1)	Työ lisäsi hieman kiinnostusta kokeelliseen hiukkasfysiikkaan.

motivaation lisäämiseen liittyvistä alaluokista löytyi yhteneväisyyksiä. Ateenan yliopistossa kehitettyä, Z - ja W -bosonien havaitsemista käsittelevää laboratoriotyötä tehneet opiskelijat arvostivat myös aidon datan käyttöä [46, s. 33]. Lisäksi Suonien [50] pro gradu -tutkielman mukaan lukiolaiset kokivat Helsingissä järjestetyssä Masterclass-tapahtumassa data-analyysin mielenkiintoiseksi.

Neljäntenä analyysin yläluokkana oli tavoitekategoria ”Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittäminen”. Tavoitteessa tieteen luonteella tarkoitetaan kuvausta siitä, mitä tiede on, miten se toimii, miten tutkijat toimivat sosiaalisena ryhmänä ja miten yhteiskunta reagoi tieteeseen ja ohjaa sitä [6, s. 3–4].

Tähän yläluokkaan kuuluvat haastatteluista ja lomakevastauksista nousseet alaluokat on esitetty taulukossa 11. Alaluokat kuvaavat sitä, minkä aiheen testajat mainitsivat kehittyneen.

Eniten kehittymisen kokemuksia mainittiin liittyvän hiukkasten löytämisen prosessiin, jolla viitataan sekä hiukkasten havaitsemiseen että niiden olemassaolon todistamiseen. Testaajat kuvailivat työn selventäneen, mahdollisesti selventäneen, konkreettisoineen sekä havainnollistaneen hiukkasten löytämisprosessia.

Lisäksi testaaajien mielestä työ konkretisoi ja avasi kokeellista hiukkasfysiikkaa sekä

Taulukko 11. Yläluokkaan ”Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittäminen” liittyvät alaluokat haastatteluissa ja kyselylomakevastauksissa.

Haastattelut	Kyselylomake
löytämisprosessi (5)	hiukkasfysiikan tutkiminen (1)
kokeellinen hiukkasfysiikka (3)	
invariantin massan hankkiminen (1)	
fyysikon työ (1)	

sen tutkimista. Esimerkiksi yksi testaaaja kertoi haastattelun lopuksi seuraavasti:

H: – – se kumminkin voi herättää kiinnostuksen siihen hiukkasfysiikkaan, voisin kuvitella että. Antaa vähän sitä konkretiaa siitä, että miten se oikeesti toimii että.

M: Mm

H: Siitä kokeellisesta työskentelystä.

Työ myös konkretisoi yhden testaaajan mielestä sitä, mitä fyysikko tekee työkseen. Yhdelle testaaajalle työ opetti, miten invariantti massa voidaan ”hankkia”.

Tulosten perusteella voidaan todeta, että tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittymistä koskevat ilmaisut liittyivät työssä tehtyyn prosessiin sekä yleisesti kokeelliseen hiukkasfysiikkaan. Työ ei testaaajien mainintojen perusteella kehittänyt ymmärrystä fysiikasta yleisesti tieteenä eikä avannut yhteiskunnan suhdetta tieteseen. Toisessa kehittämisvaiheessa olisi siis hyvä ottaa huomioon myös nämä tavoitteet.

Viimeisenä Nivalaisen ym. [6] fysiikan kokeellisten töiden tavoitekategoriiana on ”Sosiaalisten ja oppimisen taitojen parantaminen”. Tällä viitataan mm. yhteistyö- ja kommunikaatiotaitoihin sekä opiskelijan ja opettajan väliseen vuoropuheluun [6, s. 4].

Osittain työn toteutustavasta johtuen tämä tavoite jäi täyttymättä. Päätin alunperin, että työ tullaan tekemään itsenäisesti ja vapaasti valittavana ajankohtana. Olisin todennäköisesti voinut myös sopia, että työ tehtäisiin työosaston vuorolla itsenäisen työskentelyn sijaan. Tällöin paikalla olisi laboratorioassistentti opastamassa ja auttamassa tarvittaessa. Päädyin kuitenkin itsenäisesti tehtävään työhön siksi, koska tämä vaihtoehto oli ylipäänsä mahdollinen. Kehittämäni työ ei vaadi erillistä laitteistoa, vaan ainut tarvittava asia on tietokone ja verkkoyhteys.

Itsenäisesti tehtävä työ voi olla vuorolla tehtävää sujuvampi laboratoriotöiden yleisen organisoinnin kannalta, sillä tällöin työn tekeminen ei vaadi osastovuoron järjestämistä. Jyväskylän yliopiston fysiikan laitoksella on muutama vastaava laboratoriotyö, joihin ei tarvita erillistä työosaston vuoron varausta. Toisaalta itsenäisesti

tehtävä työ vaatii jälkikäteen tapahtuvan tarkastamisen, mikä muutoin tehtäisiin osastovuorolla.

Tavoitteeseen sosiaalisten ja oppimisen taitojen parantamisesta voisi vastata myös laatimalla työn pareittain tehtäväksi. Tällöin työ voisi kehittää myös yhteistyö- ja kommunikaatiotaitoja. Lisäksi parityöskentely voisi herättää itse tekemistä enemmän pohdintaa työn kysymyksiin liittyen. Toisessa kehittämisvaiheessa olisi siis hyvä pohtia, voisiko työ olla myös pareittain tehtävissä.

Nivalaisen ym. [6] tavoitekategorioiden lisäksi kirjasin analyysirunkoon (ks. taulukko 2) kolme yläluokkaa, joilla pyrin selvittämään yleisesti, kuinka toimiva työ oli. Nämä yläluokat käsittelevät työssä kohdattuja haasteita, kehittämistä vaativia asioita sekä onnistuneita asioita. Pyrin saamaan yläluokkien avulla lisää tietoa työn toimivuudesta fysiikan kokeellisten töiden tavoitekategorioiden ohella. Seuraavissa alaluvuissa kuvailen kolmen viimeisen yläluokan analyysin tulokset.

5.5.3 Työssä kohdatut haasteet

Yläluokan ”Työssä kohdatut haasteet” avulla pyrin selvittämään, millaisia haastavia asioita työ sisälsi testaaajien mielestä. Tätä varten sisällytin sekä haastatteluihin että kyselylomakkeeseen kysymyksen ”Mitä haasteita kohtasit työtä tehdessä?” Haasteita nousi esiin tämän kysymyksen lisäksi myös muista haastatteluiden ja lomakkeen kohdista.

Analyysia tehdessä tajusin, että sanavalintana ”haaste” on monitulkintainen. Sana sisältää sekä negatiivisen että positiivisen merkityksen. Negatiivisessa merkityksessä haasteen voi mieltää asiana, joka haittaa tai häiritsee. Laboratoriotyössä haastava kohta voisi tarkoittaa esimerkiksi työn osiota, jossa opiskelijan on vaikea edetä ilman apua. Positiivisessa merkityksessä taas haaste voi viitata asiaan, joka motivoi ja tarjoaa ajateltavaa. Tässä mielessä laboratoriotyö ilman haasteita voisi tarkoittaa liian helppoa työtä, joka ei opettaisi mitään uutta.

Haastattelurunkoa ja kyselylomaketta laatiessa tavoittelin haaste-sanalla sen negatiivista merkitystä. Tämän avulla pyrin selvittämään mahdollisia liian hankalia tai toimimattomia kohtia työssä. Analyysin vastauksia tulkitessa on kuitenkin otettava huomioon edellä mainittu monitulkintaisuus. On mahdollista, että testaaaja on kokenut haaste-sanana positiivisella merkityksellä.

Analyysin tuloksena haastatteluista ja kyselylomakkeesta yläluokkaan ”Työssä kohdatut haasteet” nousseet alaluokat on esitetty taulukossa 12. Alaluokkien perässä oleva luku kertoo, kuinka monta kertaa kyseiseen alaluokkaan liittyvä ilmaus mainittiin. Esimerkiksi yksi testaaajista mainitsi kolmesti eri kohdissa haastattelua ohjelmoinnin tuottaneen pieniä haasteita, jolloin kirjasin jokaisen näistä maininnoista erikseen alaluokkaan ”ohjelmointi”.

Taulukosta 12 nähdään, että haastatteluiden ja lomakkeen neljä eniten mainintoja saanutta alaluokkaa ovat lähes samat. Eniten haasteita mainittiin esiintyneen ohjel-

Taulukko 12. Yläluokkaan ”Työssä kohdatut haasteet” liittyvät alaluokat haastatteluissa ja kyselylomakevastauksissa.

Haastattelut	Kyselylomake
ohjelmointi (16)	ohjelmointi (10)
sanalliset tehtävät (4)	sanalliset tehtävät (3)
epäselvä tehtävänanto (2)	ei haasteita (2)
ei haasteita (1)	virhelaskut (1)
virhelaskut (1)	
informaation sisäistäminen (1)	
aaltofunktion merkintä (1)	
massojen yksiköiden puuttuminen (1)	
invariantin massan laskeminen (1)	
logaritmiset asteikot (1)	

mointiin liittyen. En eriteltyt *ohjelmointi*-alaluokassa sitä, kertoiko testaaaja ohjelmointihaasteen olleen pieni vai suuri. Esimerkkinä tästä esitän kahden eri testaaajan maininnat ohjelmointiin liittyvistä haasteista. Seuraava alkuperäinen ilmaisu liittyi testaaajan kokemaan pieneen haasteeseen ohjelmoinnissa:

M: Joo. Öö kerro sit tota omin sanoin sen työn sujuvuudesta.

H: No siis. Se se suju niinku muuten iha jees paitsi et siin oli just ne pari pari just ne sulku ohjelmointipätkää mitkä ny tietty teki sillai sen ny taas. En mä tiedä oliko se siin ite omist ohjelmointitaidoist lähinnä vai siitä omast niinku huolimattomuudest kiinni ne mitkä niinku tökki – –.

Toinen testaaaja taas koki ohjelmoinnin huomattavasti suurempana haasteena. Tämä nähdään seuraavasta haastattelukatkelmasta:

H: Se oli helvetin vaikeeta kun koodista puuttu se yks sulku. Voi että se oli vaikeeta ja tuskasta ja ärsyttävää ja kaikkee mahdollista, hajottavaa että. Yhtä samanlaista tuskaa, en tiijä milloin sitä kokee mutta se kun sä et pääse millään tavalla eteenpäin – –.

Ohjelmoinnissa haastavaksi mainittiin edellä olevien sulkujen puuttumisen lisäksi mm. virheilmoitusten tulkitseminen, kokemattomuus Python-kielestä sekä yleisesti pienet haasteet. Haastavaksi koettiin myös esimerkiksi se, että työssä väitettiin sekä komennon `ds.head()` että komennon `print()` tulostavan käytetyn datatiedoston ensimmäiset rivit.

Toiseksi eniten haasteita mainittiin liittyvän sanallisiin tehtäviin. Sanallisissa tehtävissä haastavaksi mainittiin mm. sen hahmottaminen, minkälaisia vastauksia tehtävillä haettiin. Esimerkiksi yksi testaaaja kertoi haastattelussa seuraavasti:

H: Paitti nämä niin sanalliset kysymykset on aina mulle kauheen hankalia. Niihin on vaikee keksiä mitä tässä niinkun halutaan vastaukseks.

Sama testaja vastasi myös kyselylomakkeen kysymykseen ”Mitä haasteita kohtasit työtä tehdessä?” seuraavalla vastauksella:

– – Sanallisten tehtävien merkityksen avaaminen tuntuu aina haastavalta.

Muut haastavaksi mainitut asiat olivat sekä suurempia että yksittäisempiä osuuksia työstä. Esimerkiksi informaation sisäistämisen haaste liittyy työhön kokonaisuutena, kun taas logaritmisien asteikon tuoma haaste on yksittäinen haastava tekijä. Lisäksi osa testajista koki, että he eivät kohdanneet työssä haasteita. Esimerkiksi kaksi testajaa vastasi kyselylomakkeen kysymykseen ”Mitä haasteita kohtasit työtä tehdessä?” seuraavilla vastauksilla:

En juuri mitään. Kaikki toimi niinkuin pitkin.

En kohdannut, sillä ohjeet olivat selkeät.

Haaste-sanana monitulkintaisuuden takia haasteettomuus voisi olla työn toimivuuden kannalta huono asia. Vastausten perusteella kuitenkin vaikuttaa, että vastaajat ovat käsittäneet haasteettomuuden onnistumisena. Vastauksista ei näy, että haasteettomuus olisi liittynyt liian helppoon työhön, vaan ennemminkin työssä etenemistä haittaavien asioiden puuttumiseen.

5.5.4 Työssä kehitettävät asiat

Toiseksi viimeinen analyysin yläluokka oli ”Työssä kehitettävää.” Tällä yläluokalla tavoittelin testajien mielipiteitä siitä, mitä työssä tulisi vielä kehittää testauksen perusteella. Pyrin saamaan tietoa sekä konkreettisista kehityskohteista että abstrakteista kehitystä kaipaavista työn osa-alueista. Yläluokkaa varten sisällytin sekä haastattelurunkoon että kyselylomakkeeseen kysymykset ”Mitä muuttaisit työstä?” ja ”Miten olisit voinut oppia lisää työstä?”

Taulukossa 13 on esitetty haastatteluista ja lomakevastauksista nousseet kehitettäviin asioihin liittyvät alaluokat sekä niihin kuuluvien ilmaisujen lukumäärät. Lisäksi taulukkoon on kirjattu kaksi alaluokkaa yhdistävää tekijää, *suuremmat muutokset* ja *pienemmät muutokset*. Näillä alaluokat on jaettu suurempia ja pienempiä muutoksia vaativiin kehityskohteisiin.

Suurempia muutoksia vaativia kehitettäviä asioita tarkastelemalla nähdään, että sekä haastatteluiden että lomakevastausten perusteella työ voisi tarjota enemmän tekemistä. Tähän alaluokkaan sisältyvät ilmaukset vaihtelivat pienistä asioista laajempiin kokonaisuuksiin. Tämän hahmottamiseksi taulukkoon 14 on kirjattu alaluokkaan *lisää tekemistä* sisältyvät pelkistetyt ilmaukset.

Taulukko 13. Työssä kehitettäviin asioihin haastatteluista ja kyselylomakkeesta nousseet alaluokat sekä niitä yhdistävät tekijät.

Yhdistävä tekijä	Alaluokka (haastattelu)	Alaluokka (kyselylomake)
suuremmat muutokset	lisää tekemistä (15)	lisää tekemistä (4)
	lisää tietoa (8)	lisää tietoa (3)
	käytettävyys* (7)	ohjelmointi liian helppoa (1)
	ohjelmointi ongelmallista (6)	mahdollisuus käyttää Matlabia (1)
	ohjelmointi liian helppoa (1)	vähemmän ohjausta (1)
pienemmät muutokset	voisi olla konkreettisempi (1)	
	tarkempi ohjeistus (10)	paremmat sanalliset kysymykset (3)
	käytettävyys* (7)	
	paremmat sanalliset kysymykset (6)	
	logartimiset asteikot haastavia (1)	

* alaluokka ”käytettävyys” sisälsi sekä suuria että pieniä muutoksia

Taulukosta 14 voidaan myös havaita, että alaluokkaan *lisää tekemistä* liittyvät kehitysmaininnat jakautuivat konkreettisiin ja abstrakteihin kohteisiin. Konkreettisia kehityskohteita olivat mm. hiukkasten lentoratojen tutkimisen lisääminen, laskujen välivaiheiden laskenta sekä yksittäisen kysymyksen lisääminen. Abstraktit kehitysehdotukset koskettivat taas ohjelmoinnin ja sisällön lisäämistä.

Suuremmista muutoksista toiseksi eniten mainintoja liittyi taulukon 13 mukaan alaluokkaan *lisää tietoa*. Tähän liittyen mainittiin, että työstä olisi voinut oppia enemmän mm. teoriaa, hiukkasilmäimien toimintaa sekä Python-kieltä. Yksittäisten testaaajien mielestä myös pohjaluentoa olisi voinut syventää ja käytetyn datan alkuperää olisi voinut avata tarkemmin. Lisäksi mainittiin, että työssä kaiken ohjelmointikoodin olisi voinut laittaa näkyviin käyttäjälle.

Kuten taulukkoon 13 on merkitty, kirjasin alaluokan *käytettävyys* kummankin yhdistävän tekijän alle. Käytettyyteen liittyen suurempina muutoksina kehitettäväksi mainittiin mahdollisuus ohjelmoida erikseen Matlab-ympäristössä, TIMin monimutkainen käytettävyys sekä työn kaatuminen. Pienempinä muutoksina mainittiin käytettävien matemaattisten lausekkeiden parempi sijoittaminen, mahdollisuus nähdä työ myös jälkikäteen sekä virheilmoitusten epälooginen sijainti ja toimimattomuus.

Suurempiin muutoksiin liittyen haastatteluissa ohjelmointi mainittiin kuudessa ilmaisussa ongelmalliseksi ja yhdessä liian helpoksi. Kyselylomakkeessa yksi testaaaja mainitsi ohjelmoinnin olleen liian helppoa. Näiden lisäksi testaaajat mainitsivat yksittäisiä suurempia muutoksia vaativia kehityskohteita, jotka on esitetty taulukossa 13.

Pienempiin muutoksiin liittyivät jo mainitun alaluokan *käytettävyys* lisäksi yksittäinen maininta logartimisten asteikoiden haastavuudesta sekä eniten mainittuina alaluokat *tarkempi ohjeistus* ja *paremmat sanalliset kysymykset*. Tarkempaan ohjeis-

Taulukko 14. Yläluokan ”Työssä kehitettävää” alaluokkaan ”lisää tekemistä” kuuluvat haastattelujen ja lomakevastausten pelkistetyt ilmaisut.

Pelkistetty ilmaisu (haastattelu)	Pelkistetty ilmaisu (kyselylomake)
Datan voisi hakea itse.	Myonien liikeratoja voisi tutkia.
Teoriaosan voisi kirjoittaa tai etsiä itse.	Teoriaosan voisi kirjoittaa itse.
Analyysin voisi toistaa kahdesti.	Itse tehtävää ohjelmointia voisi olla enemmän.
Hiukkasten lentoratoja voisi tutkia.	Työ voisi olla pidempi.
Kysymyksiä tai tehtäviä voisi lisätä.	
Ohjelmoitavaa voisi olla enemmän.	
Työn voisi laajentaa pakolliseksi työselostukseksi.	
Verrokkihistogrammin sijaan hiukkasia voisi identifioida Particle Data Groupin avulla.	
Laskujen välivaiheita voisi laskea paperille ennakkoon.	
Osan laskuista olisi voinut laskea käsin sekä verrata näitä koneen laskemiin.	
Voisi lisätä kysymyksen logaritmisesta ja ei-logaritmisestä eroista.	
Voisi lisätä kysymyksen siitä, miten eri hiukkaset liittyvät toisiinsa.	
Ohjelmoitavaa voisi olla enemmän.	
Sisältöä voisi olla enemmän.	
Ohjelmoitavaa voisi olla enemmän.	

tukseen sisällyttiin ilmaisut, joissa mm. mainittiin kehityskohteiksi työn infotekstien ja termien tarkentaminen, energian kokoluokan avaaminen sekä tieto siitä, tuleeko lasketulle invariantille massalle selvittää virherajat. Lisäksi mainittiin, että sanallisten kysymysten yhteydessä voisi tarkentaa, minkälaisia vastauksia kysymyksiin haetaan. Alaluokkaan sisältyi myös muita yksittäisiä kehityskohteita.

Alaluokka *paremmat sanalliset kysymykset* viittaa nimensä mukaisesti ilmaisuihin, joissa kehitettävä asia liittyy sanallisten kysymysten kehittämiseen. Sanallisten kysymysten mainittiin esimerkiksi sisältävän toistoa, olevan pinnallisia sekä epäselviä. Edellä mainittu kysymysten tarkastaminen olisi voinut sisältyä hyvin myös tähän alaluokkaan. Myös kyselylomakkeista ilmeni yksi maininta tähän alaluokkaan liit-

tyen.

5.5.5 Työssä onnistuneet asiat

Haasteiden ja kehityskohteiden lisäksi pyrin selvittämään, mitkä asiat testajat kokivat työssä onnistuneiksi. Tätä varten loin analyysirungon viimeiseksi yläluokaksi luokan ”Työssä onnistunutta”. En lisännyt haastattelurunkoon tai kyselylomakkeeseen erillisiä kysymyksiä tähän yläluokkaan liittyen. Yläluokkaan kuuluvat alaluokat tulivat ilmi eri kysymysten yhteydessä. Sekä haastatteluiden että lomakevastausten perusteella nousseet alaluokat on esitetty taulukossa 15.

Taulukko 15. Yläluokkaan ”Työssä onnistunutta” liittyvät alaluokat haastatteluissa ja kyselylomakevastauksissa.

Haastattelut	Kyselylomake
käytettävyys (9)	käytettävyys (4)
ohjelmointi (6)	ohjeistus (4)
havainnollistavuus (5)	opettavuus (3)
ohjeistus (3)	vaikeustaso (3)
teoriaosa (3)	teoriaosa (1)
muu (13)	muu (4)

Tulosten perusteella työn onnistuneimpia osa-alueita olivat mm. käytettävyys, ohjeistus, ohjelmointi, havainnollistavuus sekä opettavuus. Myös teoriaosa mainittiin toimivaksi sekä työn vaikeustaso sopivaksi. Alaluokkaan *muu* sisältyi pelkistettyjä ilmaisuja, joissa mm. kuvailtiin työtä (mukava, mielenkiintoinen, motivoiva), kerrottiin työn olevan sopivan mittainen sekä koettiin oikean datan käyttö onnistuneeksi. Alaluokan *muu* havainnollistamiseksi kaikki siihen liittyneet pelkistetyt ilmaisut on kirjattu taulukkoon 16.

5.6 Toisen ongelma-analyysin yhteenveto

Tämä alaluku sisältää yhteenvedon toisessa ongelma-analyysissä sisällönanalyysillä ja sisällön erittelyllä saaduista tuloksista. Tulosten perusteella pohdin, kuinka hyvin laboratoriotyön testattu versio toteutti Nivalaisen ym. [6] fysiikan kokeellisten töiden tavoitekategoriat. Lisäksi selvitän toista kehittämisvaihetta varten, millä tapaa työtä tulee kehittää tulosten perusteella.

Alaluvussa 3.7 tarkemmin esitellyt, Nivalaisen ym. määrittelemät fysiikan kokeellisten töiden tavoitekategoriat ovat kertauksena seuraavat [6, s. 3] (käännökset: Rikkilä):

1. Käytännöllisten tai kokeellisten taitojen kehittäminen

Taulukko 16. Yläluokan ”Työssä onnistunutta” alaluokkaan ”muu” kuuluvat haastattelujen ja lomakevastausten pelkistetyt ilmaisut.

Pelkistetty ilmaisu (haastattelu)	Pelkistetty ilmaisu (kyselylomake)
Muulla luodun datan käyttäminen oli mielenkiintoista	Työ ei vaadi muutoksia.
Muulla luodun datan käyttäminen oli tulevien analyysien kannalta hyödyllistä	Työ ei ehkä vaadi muutoksia.
Työ oli mielekäs	Työn tekeminen oli mukavaa.
Työ oli motivoiva	Työ oli mielenkiintoinen.
Asioiden ”mutkan kautta” löytämisen pohittaminen oli mielenkiintoista.	
Työ oli kokonaisuudessaan hyvin hiottu.	
Työn pituus oli lapputyöksi hyvä.	
Oikean datan käyttö oli hyvä asia.	
Työ oli motivoiva.	
Työ oli hauska.	
Työ oli palkitseva, kun koodin toimisessa työssä pääsi eteenpäin.	
Työhön meni sopivasti aikaa.	
Työ oli aika kattava, mitään ei jäänyt käsittelemättä.	

2. Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittämisen
3. Motivaation lisääminen
4. Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittäminen
5. Sosiaalisten ja oppimisen taitojen parantaminen

Näistä ensimmäisen työ saavutti tulosten perusteella melko hyvin. Kysyttäessä kokeellisten taitojen kehittymisestä, testaaajista viisi vastasi taitojen kehittyneen tai kehittyneen vähän. Kolmen testaaajan mielestä kokeelliset taidot eivät kehittyneet, ja yksi testaaaja ei osannut kertoa muuttuivatko taidot. Eniten mainittiin kehittyneiksi tunnistamisprosessiin, data-analyysiin ja ohjelmointiin liittyviä taitoja. Maininnat viittaavat siihen, että laboratoriotyön alkuun kirjattu tavoite täyttyi osalla testaaajia. Työn tavoitteeksi kerrottiin data-analyysin tekeminen sekä Z -bosonin esiintymisen toteaminen.

Työ olisi voinut kehittää enemmän tekijöiden ohjelmointitaitoja. Viisi testajaa koki, että työn tekeminen ei muuttanut heidän ohjelmointitaitojaan. Kolme testajaa kertoi ainoastaan syntaksin ymmärtämisen tai koodin lukemisen kehittyneen. Ainoastaan yksi testaja ilmaisi varsinaisten ohjelmointitaitojensa kehittyneen. Tämän perusteella toisessa kehittämisvaiheessa voisi tavoitella muutoksia, joiden avulla työ opettaisi laajemmin ohjelmointia.

Toisaalta huomiot ohjelmoinnista eivät olleet yksiselitteisiä. Tulosten perusteella työssä kohdatuista haasteista nousi selvästi eniten mainintoja ohjelmointiin liittyen. Maininnat sisälsivät sekä pieniksi koettuja että suurempia haasteita. Myös testajien kertomia työn kehitysehdotuksia liittyi ohjelmointiin. Neljässä ilmaisussa ohjelmointia ehdotettiin lisättävän, kuuden ilmaisun perusteella ohjelmointi oli ongelmallista ja kahden mukaan liian helppoa. Esitietolomakkeen mukaan testajien ohjelmointitaidot ennen työn tekemistä vaihtelivat. Testaajilla oli ennen työn tekemistä kokemusta Python-kielestä enintään muutamien ohjelmointikertojen verran tai ei ollenkaan.

Toisessa kehittämisvaiheessa ohjelmoinnin oppimista tulee siis yrittää lisätä, mutta samalla täytyy varmistaa, ettei ohjelmoinnin taso nouse liian vaikeaksi. Testaajista suurin osa sai tehtyä työn itsenäisesti ilman apua, mutta yhdelle annoin valmiiksi histogrammin tuottamiseen vaadittavan komennon sekä kahta autoin ongelmatilanteessa. Ohjelmointiin liittyvää oppia voisi siis mahdollisesti lisätä esimerkiksi avaamalla tarkemmin koodin toimintaa tai lisäämällä opiskelijan tehtäväksi yksinkertaista ohjelmointia. Monimutkaisemman ohjelmoinnin vaatiminen opiskelijalta voi olla havainnointien perusteella riski, sillä työ tulee olla itsenäisesti tehtävissä. Lisäksi ohjelmointiin liittyen piilossa olevaa koodia voisi muuttaa enemmän opiskelijan nähtäväksi. Yksi testaja mainitsi, että tämä selventäisi koodikomponenttien toimintaa.

Toinen fysiikan töiden tavoitekatgoria liittyi tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittämiseen. Myös tämä tavoite täyttyi tulosten perusteella melko hyvin. Työ tarjosi monipuolisesti hiukkasfysiikkaan liittyvää tietoa sekä opetti erilaisia asioita. Testaajat mainitsivat saaneensa tietoa esimerkiksi hiukkasista, Z -bosonin määrittämisestä sekä kokeellisen hiukkasfysiikan tekemisestä. Opituiksi asioiksi mainittiin mm. invariantin massan käsite, metodi invariantin massan tai hiukkasten löytämiseen sekä pseudorapiditeetin hyödyntäminen analyysiprosessissa.

Työ olisi kuitenkin voinut tarjota vielä paremmin tietoa hiukkasfysiikan teoriasta. Testajien hiukkasfysiikan tietojen muuttumisesta kysyttäessä kaksi testajaa kertoi tietojen muuttuneen ja kolme tietojen muuttuneen vähän. Neljän testaajan mielestä työ ei muuttanut heidän tietojaan hiukkasfysiikasta. Esitietolomakkeen perusteella ainoastaan yksi tiedonantajista oli käynyt hiukkasfysiikkaa käsittelevän kurssin FYSH300: Hiukkasfysiikka. Toisaalta lomakkeesta ilmeni, että neljä tiedonantajaa koki hiukkasfysiikan jollakin tapaa tutuksi, neljän jättäessä mielipiteensä antamatta ja yhden ollessa osittain toista mieltä.

Esitietojen perusteella työssä käsiteltävä hiukkasfysiikan teoria olisi siis ainakin osittain uutta asiaa tiedonantajille. Se, että testaa- jien tiedot hiukkasfysiikasta enimmäkseen pysyivät ennallaan tai muuttuivat vähän, viestii tarpeesta pohtia työn kehittämistä hiukkasfysiikan teorian osalta. Teoriaosaa voisi esimerkiksi syventää, tai työhön voisi lisätä tehtävän, jossa opiskelijan tulisi soveltaa hiukkasfysiikan teoriaa tai käsitteitä.

Kolmanteen tavoitekategoriaan, motivaation lisäämiseen, työ vastasi tyydyttävästi. Analyysin perusteella työ sai aikaan kokemuksia, jotka lisäsivät motivaatiota tieteen oppimista kohtaan. Ilmaisuis- sa mainittiin esimerkiksi työn olevan mahdollisesti ”silmiä avaava” sekä kiinnostuksen hiukkasfysiikkaa kohtaan herättävä. Testaajat kokivat työssä myös motivaation lisääntymiseen liittyviä tunteita. Työ herätti esimerkiksi mielenkiintoa ja mukavuuden tunnetta.

Tavoitteessa onnistumista laskee se, että työn tekeminen aiheutti yhden testaa- jan mielestä motivaation laskua Python-ohjelmointia kohtaan. Tämä ei ole ohjelmointia hyödyntävälle laboratoriotyölle hyvä ominaisuus. Lisäksi yksi testaja kertoi, että työ ei lisännyt motivaatiota muita asioita kohtaan. Toisessa kehittämisvaiheessa tulee pyrkiä siihen, että työ lisäisi vielä enemmän motivaatiota yleisesti tieteen oppimista kohtaan, eikä missään tapauksessa vähentäisi sitä.

Neljäs analyysissä käytetyistä tavoitekategorioista oli nimeltään ”Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittäminen”. Myös tähän tavoitteeseen työ vastasi tyydyttävästi. Tulosten perusteella työ kehitti testaa- jien ymmärrystä mm. hiukkasten löytämisprosessista, kokeellisesta hiukkasfysiikasta sekä fyysikon työstä. Kaikki tämän tavoitekategorian ilmaisut liittyivät jollakin tapaa työssä tehtyyn analyysiprosessiin tai yleisesti kokeelliseen hiukkasfysiikkaan. Testaajat eivät maininneet työn kehittäneen heidän ymmärrystään fysiikasta yleisesti tieteenä. Mainintoja ei myöskään noussut siitä, että työ olisi avannut yhteiskunnan suhdetta tieteeseen. Näiltä osin työtä voisi siis vielä parantaa toisessa kehittämisvaiheessa.

Viidettä tavoitekategoriaa, sosiaalisten ja oppimisen taitojen parantamista, työ ei täyttänyt. Osittain tämä johtuu valitusta toteutustavasta. Itsenäisesti tietokoneella tehtävän lapputyön ideana on, että työn voi tehdä ilman fysiikan työosaston vuoron varausta. Tällöin saatavilla ei ole tekemisen aikana ohjaavaa assistenttia tai henkilökunnan jäsentä. Tästä syystä työhön ei liittynyt yhteistyö- ja kommunikaatiotaitojen kehittymistä tai opiskelijan ja opettajan välistä vuoropuhelua.

Tavoitteen täyttymistä voisi edistää esimerkiksi muuttamalla työ pareittain tehtäväksi. Tällöin opiskelijoiden välille syntyisi keskustelua sekä todennäköisesti argumentointia esim. ohjelmointiin ja sanallisiin kysymyksiin vastaamiseen liittyen. Toisessa kehittämisvaiheessa tulisi pohtia parityön mahdollisuutta.

Työn kehittämistä varten analysoin edellä mainittujen tavoitekategorioiden lisäksi työssä kohdattuja haasteita sekä kehitettäväksi ja onnistuneiksi mainittuja asioita. Haasteita liittyi tulosten perusteella eniten ohjelmointiin. Ohjelmoinnissa haastavaksi koettiin mm. sulkujen puuttumisesta johtuneet virheet, virheilmoitusten tulkitse-

minen sekä yleisesti pienet haasteet. Toiseksi eniten haasteita mainittiin liittyneen sanallisiin tehtäviin. Tehtävissä koettiin haastavaksi esimerkiksi se, että tehtävänannoista ei saanut kuvaa siitä, kuinka laajasti tehtäviin tuli vastata. Edellä olevissa tuloksissa on huomioitava se, että sana ”haaste” on monitulkintainen sisältäen sekä positiivisen että negatiivisen merkityksen.

Toisessa kehittämisvaiheessa tulee siis kiinnittää huomiota ohjelmoinnin haasteiden poistamiseksi. Tätä varten esimerkiksi virheilmoituksista voisi tarjota enemmän tietoa koodikomponenttien yhteydessä. Lisäksi ohjelmointia voisi pyrkiä sujuvoittamaan pohtimalla, mitkä kohdat voivat olla erityisen haastavia, sekä selkeyttämällä niitä. Sanallisissa tehtävissä esiintyneiden haasteiden poistamiseksi tehtävänantoja voisi selkeyttää. Tehtävien yhteydessä voisi myös ilmaista vastausten toivotun laajuuden, joko suorasti tai epäsuorasti. Ohjelmoinnin ja sanallisten tehtävien lisäksi myös muita testaaajien mainitsemissa yksittäisiä haasteita tulee pohtia toisessa kehittämisvaiheessa.

Työssä kehitettäväksi ehdotettujen asioiden analyysissä ilmeni, että testaaajat mainitsivat sekä suurempia että pienempiä muutoksia vaativia kehityskohteita. Suurempia muutoksia vaativista ehdotuksista eniten mainintoja liittyi siihen, että työ voisi tarjota lisää tekemistä. Tähän liittyen mainittiin mm. ohjelmoinnin ja sisällön lisäämistä, sekä yksittäisiä opiskelijan tehtäväksi lisättäviä asioita. Toiseksi eniten suuremmista muutoksista mainittiin tiedon lisäämiseen liittyviä kehitysehdotuksia. Työstä olisi mm. voinut oppia lisää teoriaa, hiukkasilmäimien toimintaa sekä Python-kieltä. Muita suurempia muutoksia vaativia mainintoja olivat esimerkiksi käytettävyyteen ja ohjelmointiin liittyneet ilmaisut.

Pienempiä muutoksia vaativiksi kehityskohteiksi mainittiin esimerkiksi ohjeistuksen tarkentaminen, käytettävyys sekä sanallisten kysymysten muokkaus toimivimmaksi. Testaaajien mukaan esimerkiksi infotekstejä, termejä ja tehtävänantoja voisi tarkentaa. Käytettävyyteen liittyen lausekkeiden sijoittelua voisi parantaa sekä virheilmoitusten logiikkaa avata. Sanallisista kysymyksistä voisi poistaa toistoa sekä muuttaa niitä aiempaa syvällisemmiksi.

Toisessa kehittämisvaiheessa on pohdittava sekä edellä mainittuja suurempia että pienempiä kehityskohteita. Varsinkin pienempiä muutoksia vaativat ehdotukset ovat helppoja kehittää. Suuremmista muutoksista eniten mainitut tekemisen lisääminen ja tiedon lisääminen sopivat havainnointien perusteella työhön hyvin. Testitilaisuuksissa ilmeni, että pisimmilläänkin työn tekemiseen kului aikaa alle kolme tuntia (ks. alaluku 5.3.3), kun maksimissaan työ saisi kestää neljä tuntia. Tämä viittaa siihen, että työssä on vielä tilaa laajemmalle sisällölle.

Viimeisenä analyysissä käsiteltiin työssä onnistuneita asioita. Tulosten perusteella työssä koettiin onnistuneeksi mm. käytettävyyteen, ohjeistukseen, ohjelmointiin sekä havainnollistavuuteen liittyviä asioita. Työn kerrottin myös olevan opettavainen ja vaikeustasoltaan sopiva. Lisäksi yksittäisinä mainintoina onnistuneina pidettiin mm. aidon datan käyttöä sekä työn mielenkiintoisuutta ja palkitsevuutta.

Edellä mainitut onnistumiset auttavat arvioimaan toisessa kehittämisvaiheessa tehtäviä valintoja. Esimerkiksi ohjelmointiin liittyi tulosten perusteella sekä kehitettäviä ongelmakohtia että onnistumisia. Kehittämisessä on siis pohdittava, mitä osia ohjelmoinnista tulee muuttaa ja mitä pitää ennallaan.

6 Toinen kehittämisvaihe

Toisessa kehittämisvaiheessa tein työhön muutoksia, joilla pyrin vastaamaan toisesta ongelma-analyysistä esiin nousseisiin kehityskohteisiin. Toisen ongelma-analyysin testausten ja saadun aineiston analyysin lisäksi keräsin työn toimivuudesta tietoa Jyväskylän yliopiston fysiikan laitoksen henkilökunnalta. Esittelin työtä kahdessa eri tilaisuudessa Jan Sarénille, Sami Räsäselle, Sakari Juutiselle sekä Sami Nurmel-le. Yliopistonlehtori Sami Nurmi luennoi syksyllä 2018 kurssin FYSA2002 Moderni fysiikka, osa B, jossa työ otettiin ensimmäisen kerran käyttöön.

Aloitin toisen kehittämisvaiheen kokeilemalla työn koodilohkojen toimintaa ajamalla lohkot läpi. Näissä ajoissa huomasin, että kuvaajia tuottavissa lohkoissa esiintyi häiriöitä. Osalla ajokerroista lohko tuotti vain osan kuvaajan yläreunasta. Lisäksi viimeinen, koko datatiedoston histogrammin tuottava lohko ei toiminut ollenkaan. Lohkon ajaminen antoi vain *"No file:png"* -muotoisen virheilmoituksen. Tämä viittasi siihen, että koodi pysähtyi jostakin syystä ennen viimeistä `plt.savefig()` -komentoa, jolla tuotettu kuvaaja tallennetaan png-tiedostomuodossa.

Perehdyin ongelmaan selvittämällä, kuinka kauan kuvaajia tuottavien koodilohkojen ajaminen vei laskenta-aikaa. Erilaisissa tilanteissa laskenta kesti maksimissaan n. 3,7 sekunnin verran. Tästä päättelin, että ongelma ei todennäköisesti johtunut laskenta-ajan ylittymisestä. Tämän lisäksi yritin selvittää, oliko *matplotlib.pyplot* -moduuli päivittynyt työn testivaiheiden jälkeen. Lähetin myös TIMin kehittäjille sähköpostiviestin aiheesta.

Ongelma avautui TIMin kehittäjien tekemillä testeillä. Todennäköisesti syynä toimimattomuuteen oli se, että kuvaajia tuottavat koodilohkot ylittivät jostakin syystä TIMiin asetetun rajan ns. pinon (stack) koolle. Pino on eräs tietojenkäsittelyssä käytettävä tietorakenne [56]. Ongelma ratkesi TIMin kehittäjien antamalla komennolla, jolla pinon kokoa pystyi manuaalisesti kasvattamaan. Tämän jälkeen osiot toimivat moitteettomasti.

Pidin työn teoriaosan alkupuolen lähes muuttumattomana. Lisäsin teoriaosan alkuun tietoa Higgsin bosonin löytymisestä CERNissä. Tällä pienellä lisäyksellä pyrin avaamaan yhteiskunnan suhdetta tieteeseen. Kirjoitin seuraavan tekstin:

”Vuonna 2012 CERNissä CMS ja ATLAS -koeasemat ilmoittivat havainneensa uuden hiukkasen, jonka massa vastasi Higgsin bosonin ennustettua massaa. Higgsin bosoni ja siihen liittyvä Higgsin kenttä avaavat hiukkasten massan syntymistä. Löytöä seuraavana vuonna Higgsin bosonin teoreettisesti ennustaneille Peter Higgsille ja François Englertille myönnettiin fysiikan Nobelpalkinto.”

Teoriaosan luvussa ”Hiukkastörmäysten tutkiminen” muokkasin ilmaisimien osien kuvauksia. Lisäsin kunkin ilmaisinkomponentin yhteyteen enemmän tietoa siitä, mihin komponenttien toiminta perustuu. Näillä tarkennuksilla pyrin lisäämään työn tarjoamaa tietoa hiukkasfysiikasta ja ilmaisimien toiminnasta. Esimerkiksi sähkömagneettisen kalorimetrin kuvailun yhteyteen lisäsin seuraavan selostuksen:

”ECAL-kalorimetri koostuu lyijyvolframaatista (lead tungstate) tehdyistä kristalleista, jotka tuottavat valoa elektronien ja fotonien läpäistessä ne. Tuotetun valon määrä on verrannollinen hiukkasen energiaan. Valoilmaisimien avulla voidaan siis määrittää kalorimetrin pysäyttämän hiukkasen energia. Myös HCAL-kalorimetrin toiminta perustuu valon havaitsemiseen.”

Toisen ongelma-analyysin perusteella teoriaosan lopussa oleva toinen sanallinen tehtävä koettiin liian pinnalliseksi, sillä vastaus tähän kerrottiin juuri ennen kysymystä. Korvasin tämän osion toisella tehtävällä, jossa ideana on laskea annettujen yksittäisen tapahtuman mittaustulosten avulla yksittäisen myoniparin invariantti massa tietokoneella tai laskimella. Valitsin arvot siten, että lasku tuottaa tulokseksi n. 91 GeV. Laskun lisäksi tehtävänä on etsiä Particle Data Group -sivustolta Z -bosonin massan teoreettinen ennuste, verrata tulosta siihen sekä pohtia, voiko havainnosta tehdä varmoja johtopäätöksiä. Lisäksi poistin teoriaosassa aiemmin olleen valmiiksi annetun Z -bosonin massan tehtävää varten.

Tehtävällä pyrin lisäämään työn tekijöiden hiukkasfysiikan tietojen kehittymistä. Invariantin massan lausekkeen käyttäminen laskimella laskettavaan laskuun voi auttaa lausekkeen hahmottamisessa. Lisäksi tavoittelin tehtävän tarjoavan ymmärrystä siitä, että n. 91 GeV omaava myonipari voi myös olla peräisin jostakin muusta tapahtumasta, kuin Z :n hajoamisesta. Todennäköistä toki on, että kyseessä ovat Z :sta peräisin olevat myonit.

Seuraavat työhön tekemäni muutokset liittyivät osioon, jossa aloitetaan data-analyysin tekeminen. Osiossa opiskelijaa pyydetään käymään Open Data -portaalissa tarkastelemassa alkuperäisen, AOD-muotoisen datatiedoston tietoja. Lisäsin tähän kehotuksen selvittää sivustolta, kuinka suuri datatiedosto on kyseessä. Tarkensin myös selostukseen, että työssä käytettävä CSV-muotoinen tiedosto on tallennettu TIMin palvelimelle. Näillä lisäyksillä vastasin toisessa ongelma-analyysissä mainittuun kehitysehdotukseen datan alkuperän tarkemmasta avaamisesta.

Analyysin alussa esitettäviin teknisiin huomioihin lisäsin yhden uuden huomion. Kerroin, että ohjelmoinnissa saatavat virheilmoitukset eivät aina viittaa suoraan itse virheeseen, vaan mahdollisesti virheen läheisyydessä olevaan kohtaan koodia. Kehotin opiskelijaa virheilmoituksen saadessaan tarkastamaan koko koodin toiminnan. Tällä lisäyksellä pyrin selkeyttämään virheilmoitusten tulkintaa ja virheiden paikannusta.

Muutin ensimmäistä koodikomponenttia, jossa haetaan työssä käytettävä data ja tulostetaan törmäystapahtumien lukumäärä. Jätin komponenttiin valmiiksi datan

hakevan komennon, mutta siirsin tapahtumien lukumäärän tulostamisen opiskelijan tehtäväksi. Tätä varten lisäsin osioon tietoa Pythonin komennoista `print()` ja `len()`, sekä kirjoitin avuksi kaksi avattavaa vihjettä. Avasin myös komponentin infotekstissä tarkemmin datan hakemisessa käytettävää `read_csv()` -komentoa. Muutokset liittyivät ohjelmoinnin lisäämiseen sekä sen oppimisen kehittämiseen.

Seuraavassa koodikomponentissa tulostetaan haetun datatiedoston viisi ensimmäistä riviä. Toisen ongelma-analyysin perusteella tässä aiheutti hämmennystä se, että infotekstissä väitettiin sekä komennon `head()` että komennon `print()` tulostavan. Todellisuudessa `head()` palauttaa sille annettavan muuttujan viisi ensimmäistä riviä, mutta ei vielä tulosta niitä. Korjasin tämän virheen infotekstiin. Lisäksi tämän osion yhteydessä selvensin hieman kuvailua siitä, että saadussa tulosteessa `\`-merkki kertoo tiedoston rivin jatkuvan tulosteen seuraavalla rivillä.

Yleisesti koodikomponentteihin liittyen toisessa ongelma-analyysissä ilmeni, että työssä voisi näyttää käyttäjälle komponenttien koko koodin. Tämä voisi selkeyttää ohjelmoinnin toiminnan ymmärtämistä. Lisäsin siis jokaiseen koodikomponenttiin painikkeen ”*Näytä piilossa oleva koodi*”, jota klikkaamalla käyttäjä saa näkyviin koko koodin, mutta ei pääse muokkaamaan alussa piilossa olevaa osuutta. Painikkeen avulla käyttäjä ei näe komponenttien komentoa `fullprogram: |!` edeltäviä komentoja, joiden avulla määritellään itse koodikomponentti TIMiin.

Muutoksen seurauksena muokkasin koodikomponenttia, jossa lasketaan invariantin massan arvot datatiedoston myonipareille. Työn testatussa versiossa olin toteuttanut laskun tarkistuksen siten, että piilossa olevassa tarkistuskoodissa suoritettiin lasku samalla lausekkeella, ja verrattiin yksittäistä arvoa käyttäjän laskemaan (ks. alaluku 4.2.3, esimerkki 3). Lisättyäni painikkeen koko koodin näkemiseen, opiskelijan olisi ollut mahdollista kopioida tarvittava invariantin massan lauseke suoraan tarkastuskoodista. Muokkasin tarkastusta siten, että koodi tarkastaa, osuuko opiskelijan laskemien massojen viides arvo lukujen 14,31 ja 14,32 väliin. Oikein laskettu arvo on noin 14,314303 GeV:n suuruinen. Tällä tapaa vertailu ei anna apua laskun suorittamiseen.

Lisäksi tässä osiossa kirjoitin komponentin infotekstiin huomautuksen siitä, että muuttujan `invariantti_massa` nimeä ei saisi muuttaa. Toinen nimi johtaisi siihen, että vertailukoodi ei toimisi. Lisäsin tämän huomautuksen myös virheilmoitukseen, jonka käyttäjä saa, jos lasku on laskettu väärin. Muutin myös onnistuneen laskun tulostetta siten, että tulosteessa kerrotaan invarianttien massojen yksikkö, gigaelektronivoltti (GeV).

Seuraavaan koodikomponenttiin, jossa laaditaan invarianttien massojen arvoista histogrammi, tein kaksi muutosta. Lisäsin histogrammin laatimisen yhteyteen huomautuksen siitä, että koodikomponentin loppuun ei tule kirjoittaa komentoa `plt.show()`. Normaalisti Pythonin `matplotlib.pyplot` -moduulilla kuvaajien näyttämiseen käytetään edellä olevaa komentoa, mutta TIM-alustalla tämä johti toisen ongelma-analyysin perusteella koodin ajoajan ylittymiseen. Huomautuksella pyrin vähentämään mahdollisia ongelmatilanteita ohjelmoinnin kanssa. Toisena muutoksena muu-

tin osion kolmatta vinkkiä siten, että vinkissä ei anneta suoraan sopivaa jakovälien määrää, vaan sen sijaan kehoitetaan kokeilemaan arvoja lukujen 50 ja 200 väliltä. Tällä muutoksella pyrin antamaan enemmän vastuuta opiskelijalle valmiin vastauksen sijaan.

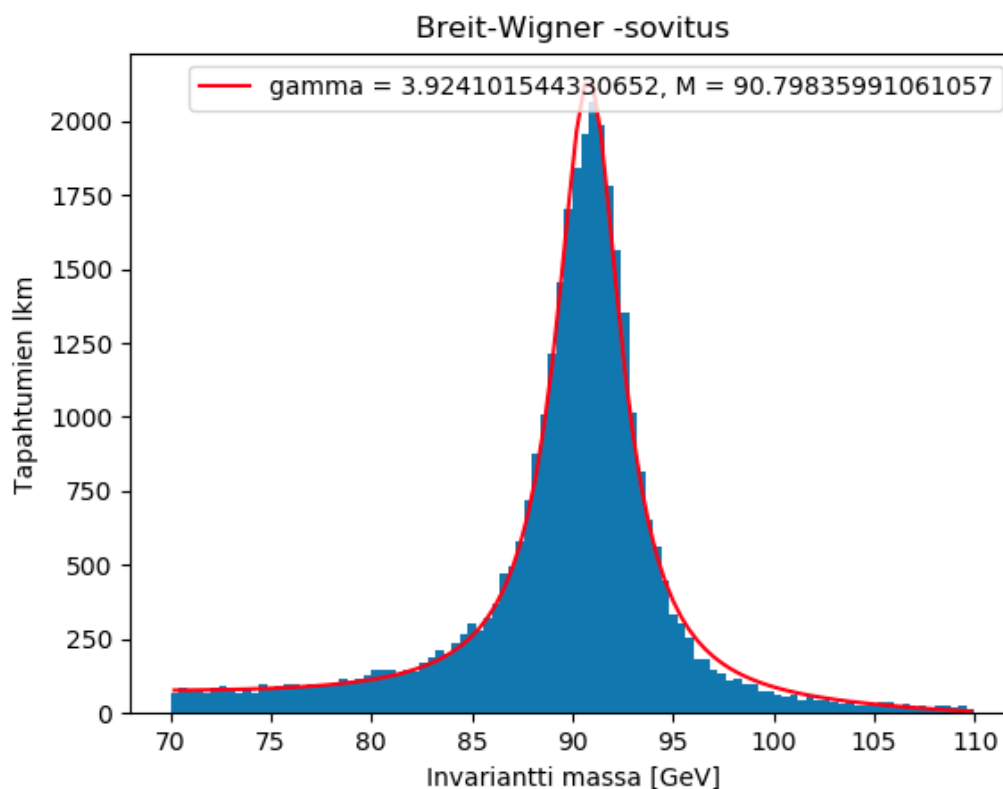
Kahteen seuraavaan osioon, ”Funktion sovittaminen histogrammiin” sekä ”Histogrammin analysointi” tein suurempia muutoksia. Muutoksien tarkastelun helpottamiseksi viitataan liitteenä olevaan edellä mainituista kahdesta osiosta luotuun PDF-vedokseen (LIITE E), joka kuvaa osioiden sisältöä toisen kehittämisvaiheen jälkeen. Osioon, jossa histogrammiin tehdään funktion sovitus, lisäsin fysiikan laitoksen henkilökunnan kommenttien perusteella alkuun tiedon siitä, että sovituksella tavoitellaan havaitun resonanssin odotusarvoa sekä resonanssia vastaavan hiukkasen elin-aikaa (LIITE E-2). Aiemmin osiossa ei mainittu tarkemmin syytä sovituksen tekemiselle. Lisäksi suurempana muutoksena vaihdoin fysiikan laitoksen henkilökunnan kommenttien perusteella sovituksen normaalijakaumasta Breit-Wigner -jakaumaan (ks. alaluku 3.2.3, lauseke (4)), sillä Breit-Wigner -jakauma kuvaa tilannetta normaalijakaumaa paremmin. Jakauman vaihtamisen myötä päivitin osion alkuun normaalijakauman sijaan tietoa Breit-Wigner -jakaumasta (LIITE E-2). Lisäsin myös alaluvun 3.2.3 lausekkeen (5) mukaisen yhteyden jakauman leveyden Γ ja resonanssissa havaittavan hiukkasen elinajan τ välille.

Sovituksen muutoksen jälkeen myös opiskelijan tehtävä muuttui hieman. Päätin antaa jakauman valmiiksi kirjoitettuna koodiin, jolloin opiskelijan pohdittavaksi jää ainoastaan jakaumaan liittyvien alkuarvojen päättely. Muutin myös pääteltävät alkuarvot normaalijakauman tunnuslukujen sijaan Breit-Wigner -jakauman maksimiksi M_r sekä puoliarvoveveydeksi Γ .

Korjasin myös osioon kirjoitettua koodia fysiikan laitoksen henkilökunnan kommenttien perusteella. Testatussa työn versiossa koodissa olevaan muuttujaan x oli jäänyt virhe. Muuttujassa olisi pitänyt olla taulukko histogrammin jakovälien keskikohtista, mutta testattu versio tuotti vääränlaisen taulukon. Päivitin koodin vastaamaan jakovälien keskikohtia. Lisäsin myös optimointikomennon painotukset osiossa tehtävälle sovitukselle sekä löyhensin sovituksessa tehtävää rajautumista Z -bosonin piikin lähelle.

Sopivilla alkuarvoilla osion koodi (LIITE E-3) tuottaa ajamisen jälkeen kuviossa 15 esitetyn kuvaajan. Kuvaajan lisäksi koodin ajaminen tulostaa kuviossa 16 esitetyn tekstin. Tulosteesta selviää `curve_fit`-optimointikomennon avulla saadut arvot histogrammissa esiintyneen resonanssin leveydelle, jakauman maksimille sekä taustan huomioimisessa käytetyille parametreille a , b ja A . Lisäksi jokaiselle arvolle tulostuu virherajat, jotka lasketaan optimointikomennon palauttaman kovarianssimatriisin avulla.

Lisäsin sovitusosion loppuun myös kaksi uutta huomiota liittyen taustan huomioon ottamiseen sekä siihen, että oikeaoppisesti sovitus tulisi tehdä iteroimalla useampia askelia (LIITE E-3). Huomioilla pyrin lisäämään työstä oppimista sekä välttämään mahdollisten virhekäsitysten syntymistä. Säilytin ennallaan huomion siitä, että mi-



Kuvio 15. Laboratoriotyössä tuotettava Breit-Wigner -sovitus invarianttien massojen jakaumassa esiintyvään resonanssiin. Kuvaajasta on luettavissa myös arvot jakauman maksimikohdalle sekä puoliarvovevyydelle.

käli tehty sovitus ei näytä osuvan histogrammiin, tulee opiskelijan palata tarkastamaan syötetyt alkuarvot.

Toisen ongelma-analyysin perusteella osiossa "Histogrammin analysointi" oleviin sanallisiin kysymyksiin liittyi kehitettävää. Kysymyksiin mainittiin esimerkiksi sisältäneen toistoa sekä olleen pinnallisia ja tehtävänannoiltaan hieman epäselviä. Toiston poistamiseksi yhdistin kysymyssarjan kaksi ensimmäistä kohtaa, jotka olivat testatussa versiossa seuraavat:

"Kysymys 4:

Mitä voit sanoa Z-bosonin esiintymisestä histogrammin ja siihen sovitetun funktion pohjalta? Perustele vastauksesi teoriaosaa ja muita lähteitä apunasi käyttäen."

"Kysymys 5:

Mikä on tulostesi perusteella Z-bosonin massa? Miten selvität sen?"

Nämä yhdistin seuraavaan muotoon (LIITE E-4):

```

Resonanssin leveyden (gamma) arvo = 3.924101544330652 +- 0.07790529087047438
Jakauman maksimin (M) arvo = 90.79835991061057 +- 0.030519006859533925
a = -1.6400566977915547 +- 0.12667558716422678
b = 167.95314789958195 +- 12.787290711707136
A = 13097.55841142805 +- 180.36500067550008

```

Kuvio 16. Kuvakaappaus laboratoriotyössä tuotetusta tulosteesta. Tulosteesta selviää arvot virherajoiheen histogrammissa esiintyneen resonanssin leveydelle, jakauman maksimille sekä taustan huomioimisessa käytetyille parametreille a , b ja A .

”Kysymys 4:

Mitä voit sanoa Z -bosonin esiintymisestä histogrammin ja siihen sovitetun funktion pohjalta?

Pystytkö määrittämään Z :n massan virheineen? Miten?

Perustele vastauksesi teoriaosaa ja muita lähteitä apunasi käyttäen.”

Ennen muutosta oli mahdollista, että opiskelija pohti Z -bosonin massaa jo ensimmäisessä kysymyksistä, ja tämän jälkeen huomasi massaa kysyttävän vasta toisessa. Lisäksi uudella kysymyksellä tarkensin sitä, että kysytylle massalle tulee ilmoittaa virherajat, jotka saadaan sovitustuloksesta. Virhearvojen laskemisen pakollisuus nousi toisessa ongelma-analyysissä yhdeksi tarkennettavaksi asiaksi.

Kysymyssarjan toisen kysymyksen muotoilin seuraavasti (LIITE E-4):

”Kysymys 5:

Laske sovituksen avulla Z -bosonin elinaika τ virheineen.

Vertaa saamaasi arvoa Z :n tunnettuun elinaikaan. Mitä havaitset? Mikä voisi mahdollisesti selittää havaintosi?”

Tässä opiskelijan tehtävänä on siis laskea sovituksesta saadun jakauman leveyden Γ avulla elinaika τ luvussa 3.2.3 esitetyn lausekkeen (5) relaation avulla. Vertailussa Z :n tunnettuun elinaikaan huomataan, että työssä saadaan tunnettua arvoa pienempi elinaika. Tämä selittyy sillä, että työssä Z -bosonia vastaavan resonanssin leveydeksi saadaan esimerkiksi Particle Data Groupin ilmoittamaa leveyttä $2,4952 \pm 0,0023$ GeV suurempi arvo [20]. Suurempi leveyden Γ arvo taas vastaa pienempää elinajan τ arvoa lausekkeen (5) mukaisesti.

Pidin ennallaan kysymyksen ”Milloin Z -bosoni havaittiin ensimmäisen kerran ja mikä Z :n fysikaalinen merkitys on?” (LIITE E-4). Tähän ei liittynyt suoria kehitysehdotuksia. Arvioin kysymyksen palvelevan opiskelijoiden tieteellisen sisällön ja prosessin ymmärtämisen kehittymistä.

Sarjan viimeinen kysymys liittyi jakauman leveyteen. Testatussa versiossa kysymys oli seuraavan muotoinen:

”Jos energia ja liikemäärä saataisiin mitattua äärettömän tarkasti, saataisiinko Z :n massan kohdalle histogrammiin yksi täsmällinen piikki vai jakauma? Perustele vastauksesi.”

Tällä kysymyksellä tavoittelin opiskelijoille ymmärrystä alaluvussa 3.2.3 käsitellystä resonanssin jakauman leveydestä. Testaajien vastauksia tarkastelemalla havaitsin kuitenkin, että osa perusteli jakauman syntymistä hieman eri tavalla, kuin kysymyksen tavoitteena oli. Esimerkiksi kolme vastausta olivat seuraavanlaisia:

- ”Jakauma, sillä mitattuja hajoamisia, joissa syntyy myoneja, on muitakin kuin pelkkä Z -bosonin hajoaminen.”
- ”Jakauma, sillä mittauksessa esiintyy kuitenkin muitakin reaktioita, jotka tuottavat myös myoni- antimyoni pareja.”
- ”Jakauma, sillä myonit voivat tulla muistakin lähteistä kuin Z -bosonista.”

Jakauman syntymistä perusteltiin siis sillä, että myös satunnaisille, ei Z -bosonista peräisin olleille, myonipareille laskettu invariantin massan arvo voi olla n. 91 GeV. Tämä perustelu on totta, mutta pikemminkin piikin leveyden sijaan nämä vaikuttavat piikin alla olevaan ns. ”taustaan”. Pysin ohjaamaan opiskelijoiden päättelyä enemmän epätarkkuusperiaatteella perustelun suuntaan muuttamalla kysymyksen seuraavaan muotoon:

”Jos energia ja liikemäärä saataisiin mitattua äärettömän tarkasti, saataisiinko Z :n massan kohdalle histogrammiin muusta jakaumasta erottuva yksi täsmällinen piikki vai jakauma? Perustele vastauksesi.”

Kysymyssarjan jälkeen työssä laadittiin koko datatiedoston kattava histogrammi, jota verrattiin CMS-kokeen julkaisemaan histogrammiin. Tein tämän osion koodikomponenttiin kaksi muutosta. Muutin alussa olevaa koodia siten, että komponentti ei hae dataa ja invariantin massan arvoja ajamalla edelliset koodikomponentit. Tämän sijaan lisäsin suoraan komponenttiin komennot, joilla haetaan tarvittavat moduulit ja datatiedosto, sekä poimitaan datatiedostosta suoraan invarianttien massojen arvot sarakkeelta M . Tällä muutoksella pyrin keventämään koodin ajamista, sillä tämä komponentti ei toiminut oikein toisen kehittämisvaiheen alussa. Lisäksi vaihdoin komennossa `plt.hist()` määrätyksi histogrammin alarajaksi arvon 0 sijaan arvon $-0,5$. Testatussa versiossa en ollut huomannut, että x-akselin määrittävän seurauksena histogrammi jatkuu negatiiviselle puolelle, joten siinä osa histogrammista rajautui vahingossa pois.

Tein muutoksia myös edellä mainittuun x-akselin määrittämiseen liittyen. Histogrammissa x-akseli kuvaa arvoa $\log_{10}(M)$, jossa M on invariantti massa. Tämä koettiin toisen ongelma-analyysin perusteella vaikeaksi hahmottaa. Tilanteen selvittämiseksi lisäsin osion infotekstiin esimerkin akselin lukemisesta. Lisätty osio alkaa kohdasta ”Eli esimerkiksi x-akselin...”:

”Aja alla oleva koodi luodaksesi histogrammin. Huomaa, että y-akseli on logaritminen ja x-akselilla on invarianttien massojen arvojen 10-kantaiset logaritmit, eli $\log_{10}(\text{massan arvo})$. Eli esimerkiksi x-akselin kohta 0,5 vastaavan invariantin massan arvon selvittäminen yksiköissä GeV onnistuu seuraavasti:

$$\begin{aligned}\log_{10}(\text{massa}) &= 0,5 \\ 10^{\log_{10}(\text{massa})} &= 10^{0,5} \\ \text{massa} &= 10^{0,5} \approx 3,1622 \text{ GeV}''\end{aligned}$$

Tämän jälkeen työssä esitettiin testatun version viimeinen kysymys, jossa pyydettiin vertailemaan saatua histogrammia CMS-kokeen vastaavaan. Tähän kysymykseen lisäksi huomautuksen, jossa kehotin opiskelijaa käyttämään tarvittaessa apuna Particle Data Group -sivustoa. Päädyin huomautuksen lisäämiseen toisessa ongelma-analyysissä ilmenneen huomion perusteella siitä, että hiukkasia voisi tunnistaa verrokkihistogrammin sijaan myös PDG-sivustolta.

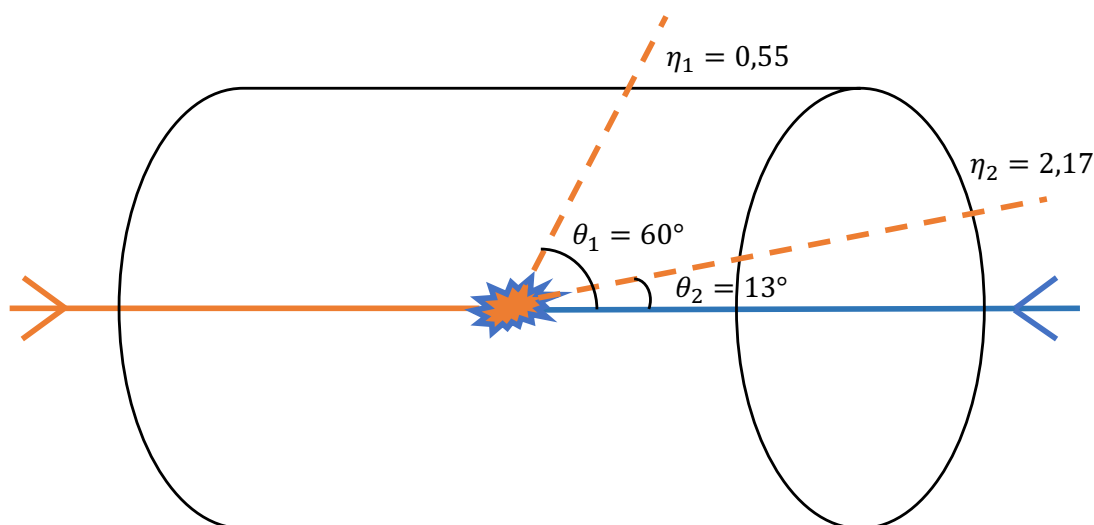
Aivan työn lopussa olevaan infotekstiin lisäsin huomautuksen siitä, että käyttäjän ei tule muuttaa vastauksiaan tai koodikomponenttien ajoja, ennen kuin hän on saanut palautteen työstä. Huomiolla pyrin välttämään tilannetta, jossa opiskelija palaisikin työn tekemisen jälkeen esimerkiksi kokeilemaan erilaisia vaihtoehtoja koodikomponenttiin. Tällöin työn tarkastaja voisi luulla kokeiluja opiskelijan vastauksiksi, sillä ne jäisivät näkymään viimeisimpinä ajoina aiempien, varsinaisten ajojen lisäksi. Valmiin työn tarkastus toimii siten, että opiskelija ilmoittaa sähköpostilla tarkastajalle tehneensä työn. Tämän jälkeen tarkastaja käy opiskelijan vastaukset läpi, antaa mahdollisesti sanallista palautetta sähköpostilla ja lopuksi hyväksyy tai hylkää työn. Jyväskylän yliopiston fysiikan laitoksella ns. lapputöistä ei anneta erillistä arvosanaa.

Edellä mainittujen testattuun runkoon tehtyjen muutosten lisäksi loin laboratorio-työhön vielä täysin uuden osion (LIITE F), jonka sijoitin työn loppuun. Uuden osion luomisella pyrin vastaamaan useampaan toisessa ongelma-analyysissä ilmenneeseen kehittämistarpeeseen. Analyysin perusteella työssä olisi voinut olla enemmän tekemistä. Tähän liittyivät mm. kehitysehdotukset ohjelmoinnin ja sisällön lisäämisestä. Myös testausten havainnoinnin perusteella työtä voisi pidentää, sillä jokainen testaaaja sai työn tehtyä alle kolmessa tunnissa, kun aikaa työn tekemiseen on maksimissaan neljä tuntia. Lisäksi toinen ongelma-analyysi antoi ilmi, että työ voisi tarjota enemmän tietoa esimerkiksi hiukkasfysiikan teoriasta, ilmaisimien toiminnasta tai Python-kielestä.

Lisätyssä osiossa perehdytään alaluvussa 3.2.3 käsiteltyyn asiaan siitä, miten pseudorapiditeetit vaikuttavat invarianttien massojen jakaumaan. Osiossa lajitellaan työssä käytetystä datatiedostosta erilleen törmäystapahtumat, joissa kummankin myonin pseudorapiditeetin itseisarvo on ollut suuri, ja kummankin pieni. Näistä tuotetaan kaksi histogrammia Z-bosonin piikin kohdalta vertailtavaksi. Histogrammeista on tavoitteena huomata, että pienempien pseudorapiditeettien tapahtumilla piikki

on hieman kapeampi kuin suuremmilla. Histogrammien luomisen jälkeen opiskelija pyydetään kertomaan ja selittämään havaintonsa. Lopuksi opiskelijalle annetaan selitys siitä, miksi histogrammit eroavat.

Osion alussa kerrataan laboratoriotyön teoriaosassa esitettyä tietoa pseudorapiditeetista (LIITE F-2). Uutena esitetään kuvion 17 mukainen kuva kahden eri pseudorapiditeetin omaavan hiukkasen suuntautumisesta CMS-ilmaisimessa. Kuvan tarkoitus on havainnollistaa opiskelijalle sitä, että eri pseudorapiditeetin omaavat hiukkaset suuntautuvat ilmaisimen eri osiin.



Kuvio 17. Kaksi eri pseudorapiditeetin omaavaa hiukkasta. Suuremman pseudorapiditeetin hiukkanen osuu ilmaisimen päättyyn. Pienemmällä pseudorapiditeetillä taas hiukkanen suuntautuu ilmaisimen keskiosaa kohti.

Teoriakertauksen jälkeen siirrytään valitsemaan erikseen suurten ja pienten pseudorapiditeettien tapahtumat (LIITE F-3). Tätä varten loin koodikomponentin, jossa suurin osa koodista on valmiiksi kirjoitettuna, mutta käyttäjän tehtäväksi jää täydentää koodissa tehtävä valintakriteeri toimivaksi. Komponentin sisältö on esitetty kokonaisuudessaan esimerkissä 7.

Esimerkki 7. Törmäystapahtumien luokittelun pseudorapiditeettien perusteella tekevä koodikomponentti piilotetun koodin kanssa.

```
''' {#pseudolohko1 plugin="csPlugin"}
type: py
filename: pseudorapiditeetti
path: user
ulimit: -f 1000 -t 10 -s 20000
showRuntime: True
showCodeOn: Näytä piilossa oleva koodi
fullprogram: !!!
```

```

//
// BYCODEBEGIN
# Haetaan tarvittavat moduulit.
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

# Tällä rivillä haetaan data ja tallennetaan se muuttujaan "ds".
ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')

suuret_etat = #TÄHÄN TOINEN JAKO
pienet_etat = #TÄHÄN TOINEN JAKO

# Tulostetaan tietoa valinnasta.
print('Kaikkien_tapahtumien_lukumäärä=_d' % len(ds))
print('Niiden_tapahtumien_lkm,_joissa_kummankin_myonin_
      ↳ pseudorapiditeetti_ollut_suuri=_d' % len(suuret_etat))
print('Niiden_tapahtumien_lkm,_joissa_kummankin_myonin_
      ↳ pseudorapiditeetti_ollut_pieni=_d' % len(pienet_etat))
// BYCODEEND
!!
'''

```

Käyttäjälle näkyvä osuus on komentojen `// BYCODEBEGIN` ja `// BYCODEEND` välissä. Alussa `import` -komentoilla haetaan tarvittavat Python-moduulit ja komennolla `pd.read_csv()` käytettävä CSV-muotoinen data. Opiskelijan tehtävänä on täydentää koodia siten, että muuttujaan `suuret_etat` tallentuvat tapahtumat, joissa kummankin myonin pseudorapiditeetit ovat olleet suuria, ja vastaavasti muuttujaan `pienet_etat` ne, joissa pseudorapiditeetit ovat olleet pieniä. Lopuksi koodi tulostaa `print()` -komentoilla kaikkien tapahtumien lukumäärän, sekä erikseen suurten ja pienten pseudorapiditeettien tapahtumien lukumäärät.

Jaottelua varten määrittelin osion infotekstissä ”suuriksi” pseudorapiditeeteiksi yli 1,52 olevat arvot ja ”pieniksi” alle 0,38 olevat. Testasin, että näillä arvoilla kumpaankin ryhmään tulee lähes sama määrä törmäystapahtumia. Opiskelija saa siis koodin täydennettyä toimivaksi kirjoittamalla esimerkissä 8 esitetyt kaksi riviä koodikomponenttiin.

Esimerkki 8. Tarvittavat kaksi riviä, joilla tapahtumat saadaan luokiteltua pseudorapiditeettien perusteella pienten ja suurten ryhmiin.

```

suuret_etat = ds[(np.absolute(ds.eta1) > 1.52) & (np.absolute(ds.
      ↳ eta2) > 1.52)]
pienet_etat = ds[(np.absolute(ds.eta1) < 0.38) & (np.absolute(ds.
      ↳ eta2) < 0.38)]

```

Kirjoitin komponentin infotekstiin listan komennoista, joita luokittelun laatimiseen tarvitaan (LIITE F-3). Avasin kommentojen toimintaa ja kerroin esimerkin yhden komennon käytöstä. Lisäksi kerroin, että Pythonissa desimaalierottimena toimii pilkun sijaan piste. Lisäsin tämän huomautuksen siksi, koska aiemmin työssä käyttäjän ei tarvitse syöttää Pythonilla desimaalilukuja. Pisteiden käyttäminen erottimena voi siis tulla opiskelijalle uutena asiana.

Komentojen lisäksi loin komponentin yhteyteen kolme klikkaamalla avautuvaa vinkkiä (LIITE F-4). Kahdessa ensimmäisessä vinkissä kehotin varmistamaan erilaisten komentoissa vaadittavien yksityiskohtien olevan oikein. Kolmannessa vinkissä annoin suoraan tarvittavat kaksi kommentoa. Tällä pyrin välttämään sitä, että opiskelija jäisi jumiin työn tähän vaiheeseen. Komennot sisältävät runsaasti erilaisia sulkuja ja viittauksia, joten on mahdollista, että opiskelijan on haastava löytää mahdollista virhettä.

Tapahtumien jaottelun jälkeen osiossa ajetaan koodikomponentti, joka tuottaa omat histogrammit pienten ja suurten pseudorapiditeettien omaavien tapahtumien invarianteista massoista (LIITE F-4). Histogrammit luodaan Z -bosonin piikin kohdalta. Tämän koodikomponentin sisältö on esitetty esimerkissä 9.

Esimerkki 9. Pienten ja suurten pseudorapiditeettien tapahtumista erilliset histogrammit luovan koodikomponentin sisältö.

```
''' {#pseudolohko2 plugin="csPlugin"}
type: py
filename: pseudorapiditeetti2
path: user
imgsource: pseudotupla.png
ulimit: -f 1000 -t 10 -s 20000
showRuntime: True
showCodeOn: Näytä piilossa oleva koodi
fullprogram: !!!
#
file = 'pseudorapiditeetti.py'
exec(compile(open(file).read(), file, 'exec'))
// BYCODEBEGIN
# Erotellaan suurten ja pienten etan tapahtumien invariantit massat
# histogrammien luomista varten.
inv_massa_suuret = suuret_etat['M']
inv_massa_pienet = pienet_etat['M']

# Luodaan matplotlib.pyplot -moduulilla muokatun kokoinen kuva,
  ↪ johon
# plotataan kaksi histogrammia.
f = plt.figure(1)
f.set_figheight(15)
f.set_figwidth(15)
plt.subplot(211)
plt.hist(inv_massa_suuret, bins=120, range=(60,120))
```

```

plt.ylabel('suuret_etas_tapahtumien_lkm', fontsize=20)
plt.subplot(212)
plt.hist(inv_massa_pienet, bins=120, range=(60,120))
plt.ylabel('pienet_etas_tapahtumien_lkm', fontsize=20)
plt.xlabel('invariantti_massa_GeV', fontsize=20)

// BYCODEEND
plt.savefig('pseudotupla.png')
!!
'''

```

Alussa koodi ajaa edeltävän komponentin, jotta siinä haettu data ja datan jaottelu saadaan käyttöön. Tämän jälkeen koodissa poimitaan suurten ja pienten pseudorapiditeettien tapahtumien invariantit massat suoraan muuttujien sarakkeilta M . Lopuksi komponentissa luodaan yksi kuva, johon sisällytetään erikseen kummankin ryhmän histogrammin kuvaaja. Komponentin ajaminen tuottaa kuvion 18 mukaisen kahden histogrammin kuvan. Histogrammien päällekkäinen sijainti helpottaa niiden vertaamista keskenään.

Histogrammien laatimisen jälkeen osiossa vertaillaan histogrammeja keskenään, sekä pohditaan selitystä vertailussa tehdyille havainnoille (LIITE F-4). Tätä varten kirjoitin seuraavan tehtävänannon:

Vertaile yllä olevia histogrammeja keskenään. Millä tapaa myonien pseudorapiditeetit vaikuttavat massajakaumaan?

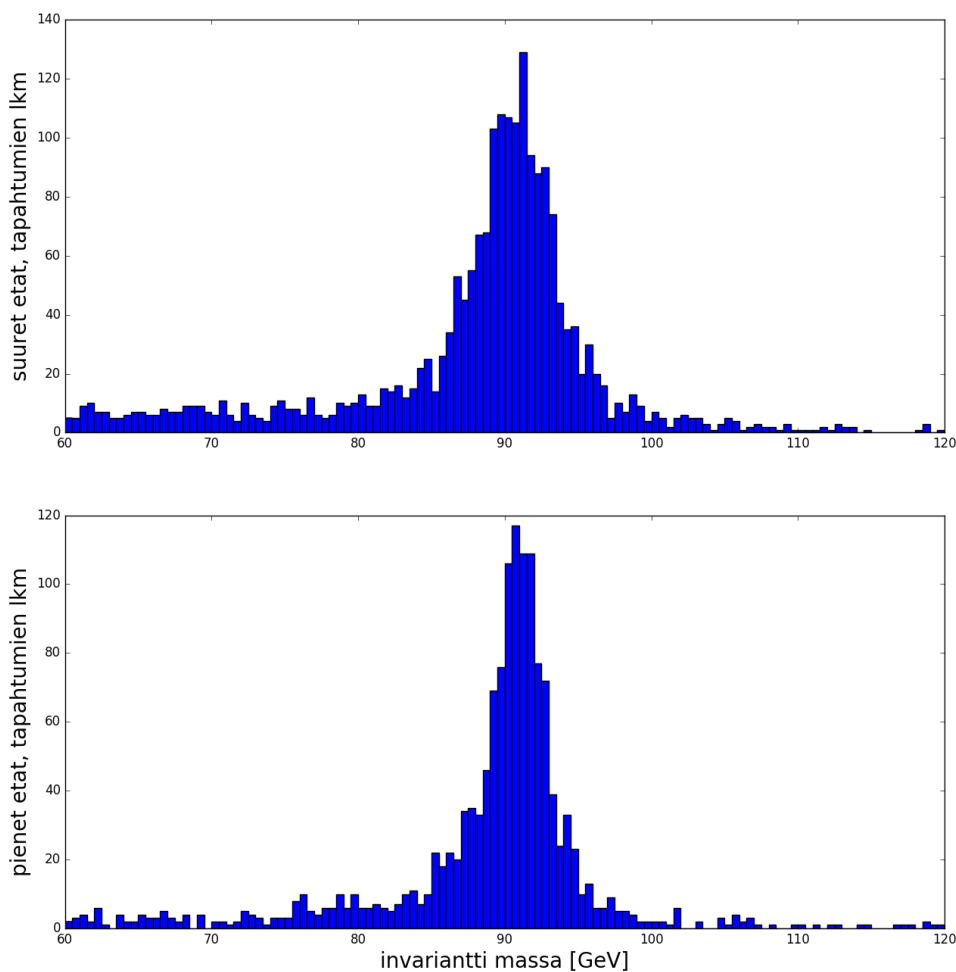
Mikä voisi mahdollisesti selittää havaintosi?

Pohdi tilannetta ensin itse, ja tallenna vastauksesi alla olevaan kenttään. Lopuksi avaa kentän jälkeen esitetty selitys tilanteelle. Ei haittaa, vaikka oma pohdintasi eroaa tästä selityksestä, vastausta ei tarvitse korjata.

Tehtävänannolla pyrin herättämään opiskelijan ajattelemaan mahdollisia selityksiä sille, miksi suurempien pseudorapiditeettien tapahtumien massajakauman resoluutio on huonompi kuin pienempien. Tarjosin kuitenkin myös klikkaamalla avautuvan selityksen tilanteelle (LIITE F-5). Tähän päädyin siksi, koska selitys on suhteellisen monimutkainen ja haastava päätellä vain työn teoriaosan perusteella. Osion tavoitteena on ennemminkin tarjota siinä havaittavan ilmiön selitys, kuin testata osaamista siihen liittyen.

Tämän jälkeen työ päättyy edellisen version tapaan lopetusosioon, jossa kerrotaan tietoa tarkastamisesta sekä tarjotaan linkki Open Data -portaaliin, jossa on saatavilla lisää tietoa CERNin avoimesta datasta.

Edellä kuvattujen muutosten lisäksi päätin toisen kehittämisvaiheen lopuksi muuttaa työn myös pareittain tehtäväksi. Toisen ongelma-analyysin perusteella työ ei parantanut tekijöiden sosiaalisia ja oppimisen taitoja. Työn aikana ei syntynyt vuo-



Kuvio 18. Työn viimeisen koodikomponentin tuottama kuva, joka sisältää erikseen invarianttien massojen histogrammit pseudorapiditeettien perusteella jaetuille törmäystapahtumille. Ylempään histogrammiin on valittu tapahtumia, joissa kummankin myonin pseudorapiditeetti on ollut suuri, ja alempaan tapahtumia, jossa pseudorapiditeetit ovat olleet pieniä.

rovaikutusta muiden opiskelijoiden tai ohjaajien kanssa. Tavoitteen parempaan täyttymiseen pyrin päättämällä, että työ on mahdollista tehdä joko itsenäisesti tai parin kanssa. Tämän päätöksen myötä työ oli myös paremmin linjassa kurssin FYSA2002 muita laboratoriotöitä koskevan ohjeistuksen kanssa. Fysiikan laitoksen työosaston ohjeissa kerrotaan, että kurssin muut lapputyöt on mahdollista tehdä itsenäisesti tai pareittain [43]. Tutkimuksen tekohetkellä fysiikan laboratoriotöistä vastaava yliopistonlehtori Sakari Juutinen hyväksyi päätöksen parityöskentelyn mahdollisuudesta.

Tätä muutosta varten lisäsin aivan työn alkuun tiedon siitä, että työ on mahdollista tehdä itsenäisesti tai pareittain. Loin alkuun myös komponentin, johon opiskelijan tai opiskelijoiden tulee kirjoittaa nimensä ja sähköpostiosoitteensa. Lisäksi työn lopussa olevaan infoon lisäsin, että mikäli työ tehtiin pareittain, tulisi myös vastuuhenkilölle lähetettävään sähköpostiin kirjoittaa molempien vastaajien nimet ja sähköpostiosoitteet. Näillä lisäyksillä pyrin siihen, että työn vastuuhenkilön on mahdollisimman aukoton nähdä, oliko palautetun työn tehnyt yksittäinen opiskelija vai kaksi tekijää.

7 Kehittämistuotos ja luotu kehittämisen viitekehys

Toinen kehittämisvaihe johti valmiiseen kehittämistuotokseen, TIM-alustalla tehtävään laboratoriotyöhön. Työ otettiin ensimmäisen kerran käyttöön Jyväskylän yliopiston fysiikan laitoksella syksyn 2018 toisessa periodissa. Kehittämistuotos tuli osaksi kurssia *FYSA2002 Moderni fysiikka, osa B* yhdeksi vaihtoehdoksi ns. lapputyöksi.

Kehittämistuotoksen sisältö ja sen muodostuminen on kuvailtu tutkimuksen kehittämisvaiheista kertovissa luvuissa 4 ja 6. TIMissä olevan työn tarkastelua on rajattu siten, että työ on nähtävissä vain Jyväskylän yliopiston Korppi-järjestelmän tunnuksilla. Tällä rajauksella vältetään se, että yliopiston ulkopuolinen tuntematon käyttäjä pääsisi ajamaan työn koodikomponentteja, ja sitä kautta mahdollisesti kuormittamaan TIMiä.

Työtä on kuitenkin mahdollista tarkastella sen englanninkielisenä versiona. TIM-alustalla oleva työ käännettiin myös englanniksi Jupyter Notebook -alustalle. Tämä mahdollistaa työn opetuskäytön myös Jyväskylän yliopiston ja Suomen ulkopuolella. Jupyter Notebook -muotoinen työ on saatavilla osoitteessa <https://github.com/cms-opendata-education/zboson-exercise>.

Kehittämistuotoksen lisäksi tutkimuksessa muodostettiin kehittämisen viitekehys tulevia avoimen datan opetusinterventioita varten. Viitekehukseen koottiin tutkimuksen perusteella viisi suositusta, jotka liittyvät onnistuneen, hiukkasfysiikan avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämiseen. Suositukset eivät ole suoraan tutkimuksessa kehitetyn työn testatun version piirteitä, sillä tulosten perusteella versio ei ollut vielä täysin onnistunut. Suositukset muodostettiin testatusta työstä saatujen tulosten ja kehittämisprosessissa havaittujen asioiden perusteella.

Seuraavassa on kuvattuna kehittämisen viitekehysten viisi suositusta, sekä niiden tarkemmat selitykset. Jokainen suosituksesta on esitetty kursivoituna ennen selitystä.

Ohjelmoinnin määrän suhteen on hyvä varautua kompromisseihin.

Hiukkasfysiikan tutkimusdataa voidaan käsitellä ohjelmointia hyödyntäen. Tutkimuksen perusteella ohjelmoinnin suhteen voi kuitenkin joutua tekemään kompromisseja. Kompromissit liittyvät siihen, kuinka paljon laboratoriotyössä ohjelmointia annetaan opiskelijan tehtäväksi ja kuinka paljon tarjotaan valmiina.

Tutkimus osoitti, että kehitetty laboratoriotyö olisi voinut lisätä paremmin opiskelijoiden ohjelmointitaitoja. Toisaalta ohjelmointi koettiin haastavaksi. Työssä suurin osa ohjelmointikoodista annettiin opiskelijalle valmiina, ja itse kirjoitettaviin osuuksiin tarjottiin sekä tietoa että vihjeitä. Suurin osa testaaajista sai työn tehtyä itsenäisesti. Kolmea testaajaa avustettiin työn tekemisen aikana.

Ohjelmoinnin suhteen on siis mahdollisesti tehtävä kompromisseja siitä, pyritäänkö työllä opettamaan mahdollisimman paljon ohjelmointia, vai käytetäänkö ohjelmointia ainoastaan datan käsittelyn teknisenä työkaluna.

Python-kieli on teknisiltä ominaisuuksiltaan riittävä datan käsittelyyn.

Tutkimuksen perusteella Python-ohjelmointikieli riittää hyvin hiukkasfysiikan avoimen tutkimusdatan käsittelyyn yliopistotasolla. Pythonin käyttö ei rajoittanut laboratoriotyön kehittämistä tai toimintaa. Kieli ja sen moduulit tarjosivat riittävät komennot suuren CSV-tiedoston lukemiseen, käsittelyyn ja kuvaamiseen graafisesti. Myös funktion sovittaminen histogrammiin onnistui Python-kielillä.

Kehittämisen aikana ilmeni teknisiä haasteita, mutta nämä liittyivät Pythonin sijaan TIM-alustan toimintaan ja kehittäjän omiin ohjelmointiratkaisuihin. Suuri datatiedosto ja siitä luotavat kuvaajat ylittivät kehittämisvaiheessa TIMiin asetettuja resurssirajoituksia. Haasteet selvitettiin kuitenkin helposti yhteistyöllä TIMin kehittäjien kanssa.

Työhön on hyvä sisällyttää monipuolisesti teoriaa kokeellisesta hiukkasfysiikasta.

Kokeellisen hiukkasfysiikan teorialla tarkoitetaan tässä mm. hiukkasfysiikan käsitteisiin, hiukkasilmäisimien toimintaan ja hiukkasten havaitsemiseen liittyvää tietoa. Tutkimuksen perusteella on hyvä, että laboratoriotyössä käsitellään kokeellisen tehtävän lisäksi kokeellisen hiukkasfysiikan teoriaa. Testattu työ antoi monipuolisesti tietoa sekä oppia hiukkasfysiikan tieteelliseen sisältöön ja käsitteelliseen ymmärtämiseen liittyen.

Toisaalta testaaajat kokivat, että työstä olisi voinut oppia vielä enemmän teoriaa ja ilmaisimien toimintaa. Työ olisi myös voinut paremmin lisätä testaaajien hiukkasfysiikan tietoja. Toisessa kehittämisvaiheessa työhön lisättiin tarkempaa tietoa ilmaisimista sekä osio, jossa perehdyttiin tarkemmin pseudorapiditeettiin liittyvään teoriaan.

Oikean datan ja analyysikeinojen käyttäminen tukee työn onnistumista.

Tutkimuksessa testatussa työssä käytettiin aitoa CMS-kokeen tutkimusdataa. Datan avulla tehtiin yksinkertaistettu analyysi, jolla havainnollistettiin oikean hiukkasfysiikan data-analyysin kulkua. Tutkimuksen perusteella oikean datan ja analyysiprosessin käyttäminen sai aikaan työn onnistumista tukevia asioita. Työ kehitti testaaajien taitoja hiukkasten tunnistamiseen ja data-analyysiin liittyen. Testaaajat mainitsivat työn olleen havainnollistava. Lisäksi yksittäisinä mainintoina aidon datan käyttöä pidettiin mielenkiintoisena ja hyvänä asiana.

Työssä olisi myös ollut mahdollista oikaista analyysin tekemisessä esimerkiksi antamalla opiskelijoille valmiina tarvittavat datan perusteella tehdyt histogrammit ja sovitukset. Tämä voisi johtaa edellä mainittujen onnistumisten puuttumiseen.

Työ on hyvä toteuttaa tehtäväksi joko pareittain tai ohjaajan avustamana.

Kehitetty laboratoriotyö ei toteuttanut Nivalaisen ym. [6] fysiikan kokeellisten töiden tavoitekategoriaa ”Sosiaalisten ja oppimisen taitojen parantaminen”. Testaajat tekivät työn itsenäisesti ilman vuorovaikutusta vertaisten tai ohjaajien kanssa. Itsenäisen työskentelyn sijaan laboratoriotyö on hyvä kehittää joko pareittain tehtäväksi tai sellaiseksi, että työn tekemisen aikana opiskelijalla on mahdollisuus vuorovaikuttaa ohjaajan kanssa.

Parityöskentely luo vuorovaikutusta ja voi herättää keskustelua työssä kohdatuista pohdinnoista ja havainnoista. Ohjaajan saatavuus taas poistaisi turhia ohjelmointiin liittyviä ongelmia. Tutkimuksessa havaittiin, että ohjelmointikoodissa olevan pienen virheen löytäminen voi olla haastavaa. Ohjaajan avustuksella ohjelmoinnissa tapahtuvat virheet eivät aiheuttaisi tilannetta, jossa opiskelija ei pääse etenemään työssä. Tällöin ohjelmointia voitaisiin myös antaa enemmän opiskelijan itse tehtäväksi.

8 Johtopäätökset ja pohdinta

8.1 Tutkimuskysymyksiin vastaaminen

Tämän tutkimuksen tavoitteena oli kehittää CERNin CMS-kokeen avointa tutkimusdataa hyödyntävä fysiikan laboratoriotyö, sekä muodostaa kehittämisen viitekehys tulevia opetusinterventioita varten. Tutkimuksen tutkimuskysymykset olivat seuraavat:

1. Mitä haasteita ja valintoja kohdataan CERNin avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämisessä?
2. Miten kehitetyn työn testattu versio täytti Nivalaisen ym. [6] fysiikan kokeellisten töiden tavoitekategoriat?
3. Minkälaisia asioita on hyvä ottaa huomioon, kun halutaan kehittää onnistunut hiukkasfysiikan avointa tutkimusdataa hyödyntävä laboratoriotyö?

Tutkimuskysymyksistä ensimmäiseen vastattiin ensimmäisen kehittämisvaiheen yhteenvedossa alaluvussa 4.3. Kehittämisen aikana kohdatut haasteet liittyivät laboratoriotyön tekniseen toteutukseen. Datatiedoston koko ylitti TIM-alustaan asetetun tiedostokoon rajoituksen. Ongelman ratkaisemiseksi rajoitusta nostettiin väliaikaisesti tiedoston lataamisen ajaksi. Lataamisen lisäksi tiedoston käsittelyn suuri ajoaika aiheutti TIM-alustalla haastetta. Kehittämisen alussa datatiedostoa käsittelevien koodikomponenttien ajaminen ylitti TIMiin asetetun suurimman mahdollisen ohjelmointikoodin ajoajan. Ongelma ratkesi, kun TIMin kehittäjät siirsivät datatiedoston parempaan sijaintiin, josta sen lukeminen oli nopeampaa.

Haastetta liittyi myös työssä olevien koodikomponenttien ketjuttamiseen. Oletuksena TIMiin luotavat koodikomponentit eivät vuorovaikuta keskenään siten, että edellisessä komponentissa ohjelmointikoodilla tehdyt asiat välittyisivät seuraavaan. Työssä kuitenkin tarvittiin tätä ominaisuutta. Haasteeseen vastattiin TIMin kehittäjien avulla luomalla koodikomponentteihin komentoja, jotka ajavat valitut edelliset komponentit ennen varsinaisen sisällön ajamista. Ratkaisu osoittautui toimivaksi.

Haasteita tarkastelemalla voidaan todeta, että pääosin työn kehittäminen sujui helposti. Mikään haasteista ei rajoittanut suunnitellun työn toteuttamista. Lisäksi haasteet ratkesivat hyvin nopeasti TIMin kehittäjien avustuksella.

Kehittämisen aikana kohdattiin valintoja ohjelmointiin ja työn sisältöön liittyvissä tilanteissa. Työn koodikomponenttien yhteydessä oli valittava, kuinka paljon ohjelmointia tuli opiskelijan tehtäväksi, ja kuinka paljon annettiin valmiina. Myös ohjel-

mointiosioden tekniseen toteutukseen liittyi valintoja. Ohjelmonnissa yhden asian voi usein toteuttaa hyvin monella eri tavalla. Työhön valitut toteutustavat eivät välttämättä olleet kaikkein optimaalisimpia, mutta niihin ei myöskään liittynyt suuria ongelmia.

Työn sisällölliset valinnat liittyivät työhön kirjoitettavan teorian määrään sekä siihen, miten työssä tehtyä data-analyysia tulkittaisiin. Työhön sisällytettiin melko runsaasti kokeellisen hiukkasfysiikan teoriaa. Tällä pyrittiin varmistamaan, että työn aihe olisi opiskelijan ymmärrettävissä. Data-analyysin tulkinnessa päädyttiin siihen, että tulkinta tehdään sanallisten kysymysten ja tehtävien avulla.

Toiseen tutkimuskysymykseen vastattiin toisen ongelma-analyysin yhteenvedossa alaluvussa 5.6. Analyysin perusteella työn testattu versio täytti melko hyvin Niivalaisen ym. [6] fysiikan käytännön töiden tavoitekategoriat ”Käytännöllisten tai kokeellisten taitojen kehittäminen” sekä ”Tieteellisen sisällön ymmärtämisen ja käsitteellisen ymmärtämisen kehittäminen”. Tyydyttävästi testattu versio vastasi tavoitekategorioihin ”Motivaation lisääminen” ja ”Tieteen luonteen ja tieteellisen prosessin ymmärtämisen kehittäminen”. Tavoitekategoriaa ”Sosiaalisten ja oppimisen taitojen parantaminen” työn testattu versio ei täyttänyt.

Tavoitekategorioihin pohjautuvan arvioinnin perusteella siis ilmeni, että työn ensimmäisessä versiossa oli vielä kehitettävää. Kehittämistarpeeseen vastattiin luvussa 6 kuvaillussa toisessa kehittämisvaiheessa. Suurimpina muutoksina työn pituutta lisättiin uudella osiolla, sekä päätettiin, että työ on mahdollista tehdä myös pareittain työskennellen. Näiden lisäksi tehtiin useampia pienempiä muutoksia.

Toisen kehittämisvaiheen seurauksena saatu laboratoriotyö tulee käyttöön Jyväskylän yliopiston fysiikan laitokselle. Työtä pyrittiin kehittämään siten, että se vastaisi mahdollisimman hyvin fysiikan käytännön töiden tavoitekategorioita. Tutkimuksessa tätä lopullista työtä ei enää testattu tai muulla tapaa arvioitu. On mahdollista, että työtä testaamalla löytyisi edelleen osa-alueita, joita olisi mahdollista vielä kehittää. Ja edelleen testausta ja kehittämistä sisältäviä syklejä voitaisiin toteuttaa uudestaan. Tutkimuksen rajallisten resurssien takia kahden syklin kehittämistä voidaan kuitenkin pitää riittävänä työn julkaisemista varten.

Kolmanteen tutkimuskysymykseen vastattiin laatimalla luvussa 7 esitetty kehittämisen viitekehys. Viitekehukseen koottiin viisi suositusta, jotka liittyvät onnistuneen, hiukkasfysiikan tutkimusdataa hyödyntävän laboratoriotyön kehittämiseen. Suositukset ovat listattuna seuraavat:

- Ohjelmoinnin määrän suhteen on hyvä varautua kompromisseihin.
- Python-kieli on teknisiltä ominaisuuksiltaan riittävä datan käsittelyyn.
- Työhön on hyvä sisällyttää monipuolisesti teoriaa kokeellisesta hiukkasfysikasta.
- Oikean datan ja analyysikeinojen käyttäminen tukee työn onnistumista.

- Työ on hyvä toteuttaa tehtäväksi joko pareittain tai ohjaajan avustamana.

Viitekehyksen suositukset muodostettiin koko kehittämisprosessin aikana saatujen tulosten, tietojen ja kokemusten perusteella. Suosituksilla pyritään helpottamaan tulevia hiukkasfysiikan avoimen tutkimusdatan opetuskokeiluja. Suositukset koskevat avointa tutkimusdataa hyödyntävän laboratoriotyön kehittämistä, mutta varauksella niitä on mahdollista soveltaa myös muiden avoimen datan opetuskokeilujen kehittämiseen.

Suositukset eivät ole kehittämiseen liittyviä ehdottomia ohjeita tai käskyjä. Ne pohjautuvat yhden kehittämisprosessin monipuoliseen tutkimiseen mutta on mahdollista, että toisenlaisessa prosessissa nousee eriäviä huomioita. Viitekehyksellä pyritään kuitenkin tarjoamaan tuleville kehityksille suuntaviivoja, sekä hyödyllistä tietoa tutkitusta yksittäisestä kehittämisprosessista.

8.2 Tutkimuksen eettisyys ja luotettavuus

Tutkimusta voidaan pitää eettisesti toteutettuna. Tutkimus laadittiin Tutkimuseettisen neuvottelukunnan [57] hyvää tieteellistä käytäntöä noudattaen. Tiedonantajia informoitiin kirjallisesti tutkimukseen osallistumisesta (LIITTEET B JA C). Jokainen tiedonantaja myös antoi kirjallisen suostumuksensa tutkimukseen osallistumiseen.

Tiedonantajilta kerättyä aineistoa käsiteltiin luottamuksellisesti. Tutkimusraportista ei ole mahdollista tunnistaa yksittäisiä tiedonantajia. Aiheen puolesta tutkimukseen ei liittynyt erityisiä eettisiä päätöksiä vaativia asioita.

Kehittämistutkimuksen luotettavuutta voidaan arvioida erityisen kriteeristön avulla (ks. alaluku 2.3). Kertauksena kriteerit ovat seuraavat:

- Kehittämisen tulee olla kokonaisvaltaista, jolloin kehittämistuloksena saadaan sekä ohjaavia malleja ja teorioita että kuvailevia teorioita.
- Kehittämisen tulee edetä sykleittäin ja sisältää jatkuvaa kehittämistä ja arviointia.
- Kehittämisessä tulee pyrkiä teorioihin, jotka ovat siirrettävissä kentälle opettajien tai muiden opetusalan ammattilaisten käyttöön.
- Kehittämisprosessiin tulee sisältyä testaamista autenttisissa olosuhteissa.
- Kehittämistutkimuksen kaikki syklit tulee dokumentoida tarkasti. [3, s. 20]

Tässä tutkimuksessa kehittämisprosessi tuotti sekä ohjaavan mallin että kuvailevan teorian. Ohjaavana mallina tutkimuksesta saatiin kehittämisen viitekehys tuleville hiukkasfysiikan avoimen datan opetusinterventioille. Kuvailevaa teoriaa taas tuotettiin kehittämisprosessin kuvailusta. Prosessissa kuvattiin haasteita ja valintoja, joita

ilmeni avointa hiukkasfysiikan tutkimusdataa hyödyntävän laboratoriotyön kehittämisessä. Toisaalta kuvailu ei ole täysin yleistettävissä, sillä vastaava laboratoriotyö voidaan toteuttaa myös muuten, kuin tutkimuksessa käytetyllä TIM-alustalla.

Tutkimus toteutettiin kahdessa kehittämissyklissä, jotka on esitetty alaluvussa 2.1 kuviossa 2. Kummassakin syklissä ongelma-analyysi ja kehittämisvaihe kulkivat liittämättä. Kehittämisvaiheen aikana tarvittaessa palattiin ongelma-analyysiin, ja sieltä takaisin kehittämiseen.

Kehittämistutkimuksessa pyritään luomaan teorioita, joita voidaan siirtää käyttöön opettajille tai opetusalan ammattilaisille. Tämä tutkimus johti kehittämisen viitekehukseen, jolla pyritään edesauttamaan tulevia hiukkasfysiikan avoimen datan opetusinterventioita. Lisäksi konkreettisenä tuotoksena tutkimuksessa kehitettiin Jyväskylän yliopiston fysiikan laitokselle käyttöön tuleva laboratoriotyö. Työstä käännettiin myös yliopiston ulkopuolella toimiva englanninkielinen versio.

Kehittämisprosessin aikana kehitettävää työtä testattiin mahdollisimman hyvin aitoa tilannetta simuloiden. Työtä testasi joukko Jyväskylän yliopiston fysiikan laitoksen opiskelijoita, jotka esitietojen perusteella kuvasivat hyvin työn kohderyhmää, toisen vuoden fysiikan yliopisto-opiskelijoita. Ennen työn tekemistä testaaajille järjestettiin pohjaluento hiukkasfysiikasta. Tällä pyrittiin simuloimaan aitoa kurssitilannetta, jossa opiskelijoille tarjotaan luennoilla teoriaa. Myös testitilanteessa pyrittiin autenttisuuteen siten, että testaaajat tekivät työn itsenäisesti mahdollisimman vähin avuin.

Tutkimuksen dokumentoinnissa pyrittiin tarkkuuteen ja kattavuuteen. Kummassakin kehittämisvaiheessa tehdyt valinnat ja muutokset kuvattiin yksityiskohtaisesti. Aineistonkeruun ja aineiston analyysin käsittelyssä tavoiteltiin kattavaa selostusta. Lisäksi tutkimukseen liittyviä erilaisia dokumentteja sisällytettiin mukaan raportin liitteiksi.

Edellä mainittujen kriteerien tarkastelun lisäksi tutkimuksen luotettavuutta hieman lisäävänä tekijänä voidaan pitää metodisen triangulaation käyttöä. Tutkimuksen aineistoa kerättiin useilla eri aineistonkeruumenetelmillä. Triangulaation kannalta oleellista oli se, että aineiston analyysissä hyödynnettiin rinnakkain sekä haastattelulla että kyselylomakkeella kerättyä aineistoa. Eri aineistoista saatuja tuloksia ei vertailtu systemaattisesti, mutta yleisesti tuloksia tarkastellessa ei havaittu merkittäviä eroavaisuuksia näiden välillä.

8.3 Suositukset jatkotutkimuksille

Kehittämistutkimuksen luonteesta johtuen tutkimus tarjoaa runsaasti mahdollisuuksia jatkotutkimuksille. Yksinkertaisimmillaan tutkimuksen jatkaminen onnistuisi uusia kehittämissyklejä luomalla. Sykleissä tutkimuksessa kehitettyä laboratoriotyötä testattaisiin uudelleen, ja testausten perusteella kehitettäisiin edelleen. Sykliä lisääminen johtaisi yhä jalostetumpaan työhön.

Toisaalta saman tutkimusasetelman toistamisen sijaan tutkimus tarjoaa myös muita mielenkiintoisia vaihtoehtoja jatkotutkimuksille. Kehitetyn laboratoriotyön käyttöönoton jälkeen työn käytännön vaikutuksien tutkiminen olisi mahdollista. Tutkimuskohteena voisi olla esimerkiksi se, millä tapaa työ vaikuttaa yliopisto-opiskelijoiden hiukkasfysiikan oppimiseen. Tähän olisi mahdollista perehtyä esimerkiksi määrällisenä tutkimuksena, jossa opiskelijoille jaettaisiin ennen työn tekemistä ja työn tekemisen jälkeen testejä, joilla pyrittäisiin mittaamaan oppimista. Toinen lähestymistapa voisi olla esimerkiksi laadullinen tutkimus siitä, miten opiskelijat ovat kokeneet työn edistävän oppimista.

Lisäksi kiinnostavia tutkimuskohteita voitaisiin saada siirtämällä työ pois yliopisto-opetuksen kontekstista. Yksi tällainen vaihtoehto voisi olla tutkimus siitä, minkälainen hiukkasfysiikan avointa tutkimusdataa hyödyntävä kokeellinen työ toimisi lukio-opetuksessa.

Lähteet

- [1] Various Authors, “Cms data preservation, re-use and open access policy.” PDF-dokumentti, 2012. Saatavilla: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=6032>. Viitattu 27.10.2017.
- [2] A. C. Tazon, D. Colling, M. A. J. Huffman, K. Lassila-Perini, T. McCauley, A. M. Rao, A. Y. R. Marrero, and E. Sexton-Kennedy, “Open access to high-level data and analysis tools in the CMS experiment at the LHC,” *Journal of Physics: Conference Series*, vol. 664, no. 3, p. 032027, 2015. Saatavilla: <http://stacks.iop.org/1742-6596/664/i=3/a=032027>. Viitattu 2.1.2019.
- [3] J. Perna, “Kehittämistutkimus tutkimusmenetelmänä,” in *Kehittämistutkimus opetuslalla* (J. Perna, ed.), pp. 9–26, Jyväskylä: PS-kustannus, 1 ed., 2013.
- [4] D. C. Edelson, “Design research: What we learn when we engage in design,” *The Journal of the Learning Sciences*, vol. 11, no. 1, pp. 105–121, 2002. http://dx.doi.org/10.1207/S15327809JLS1101_4.
- [5] The Design-Based Research Collective, “Design-based research: An emerging paradigm for educational inquiry,” *Educational Researcher*, vol. 32, no. 1, pp. 5–8, 2003. <http://www.jstor.org/stable/3699927>.
- [6] V. Nivalainen, M. A. Asikainen, and P. E. Hirvonen, “Preservice teachers’ objectives and their experience of practical work,” *Phys. Rev. ST Phys. Educ. Res.*, vol. 9, p. 010102, tammikuu 2013. DOI: 10.1103/PhysRevSTPER.9.010102. Saatavilla: <https://link.aps.org/doi/10.1103/PhysRevSTPER.9.010102>. Viitattu 25.11.2017.
- [7] J. Perna (toim.), *Kehittämistutkimus opetuslalla*. Jyväskylä: PS-kustannus, 1. ed., 2013. ISBN: 978-952-451-580-1.
- [8] A. L. Brown, “Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings,” *The Journal of the Learning Sciences*, vol. 2, no. 2, pp. 141–178, 2012. <http://www.jstor.org/stable/1466837>.
- [9] A. Collins, “Towards a design science of education,” in *New Directions in Educational Technology* (E. Scanlon and T. O’Shea, eds.), pp. 14–21, Berlin: Springer-Verlag, 1 ed., 1992. ISBN: 978-3-642-77750-9 (eBook).
- [10] T. Plomp, “Educational design research: An introduction,” in *Educational Design Research, Part A: An Introduction* (T. Plomp and N. Nieveen, eds.), pp. 10–

- 51, Enschede: Netherlands Institute for Curriculum Development (SLO), 1 ed., 2013.
- [11] M. Aksela and J. Perna, “Kehittämistutkimus pro gradu -tutkielman tutkimusmenetelmänä,” in *Kehittämistutkimus opetuslalla* (J. Perna, ed.), pp. 181–200, Jyväskylä: PS-kustannus, 1 ed., 2013.
- [12] A. Collins, D. Joseph, and K. Bielaczyc, “Design research: Theoretical and methodological issues,” *The Journal of the Learning Sciences*, vol. 13, no. 1, pp. 15–42, 2004. http://dx.doi.org/10.1207/s15327809jls1301_2.
- [13] P. Rikkilä, “Open data in CMS.” Diaesitys, 2016. Saatavilla: <http://users.jyu.fi/~tulappi/hitupva16puheet/CMS-opendata-Rikkila.pdf>. Viitattu 22.8.2017.
- [14] J. Tuomi and A. Sarajärvi, *Laadullinen tutkimus ja sisällönanalyysi*. Helsinki: Tammi, 1. ed., 2018. Uudistettu laitos. ISBN: 978-951-3199-53-1.
- [15] Y. S. Lincoln and E. G. Guba, *Naturalistic inquiry*. Newbury Park: SAGE Publications, Inc., 1985. ISBN 0-8039-2431-3.
- [16] B. R. Martin and G. Shaw, *Particle Physics*. Chichester: John Wiley & Sons Ltd, 3. ed., 2008. ISBN: 978-0-470-03294-7 (pbk.).
- [17] D. Griffiths, *Introduction to Elementary Particles*. Weinheim: WILEY-VHC Verlag GmbH & Co. KGaA, 2. ed., 2008. ISBN: 978-3-527-40601-2.
- [18] Wikimedia-käyttäjä MissMJ, “Standard model of elementary particles.” Kuva, 2006. Saatavilla: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. CC Attribution 3.0 Unported. Viitattu 24.11.2017.
- [19] CERN, “The Standard Model.” Verkkosivu. Saatavilla: <https://home.cern/topics/higgs-boson/origins-brout-englert-higgs-mechanism>. Viitattu 17.8.2017.
- [20] C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.*, vol. C40, no. 10, p. 100001, 2016.
- [21] CERN, “About CERN.” Verkkosivu. Saatavilla: <https://home.cern/about>. Viitattu 15.5.2017.
- [22] CERN, “Annual Report 2015.” Vuosiraportti. ISBN: 978-92-9083-8 (sähköinen versio). Saatavilla: <https://cds.cern.ch/record/2235983/files/2015AnnualReport.pdf>. Viitattu 15.5.2017.
- [23] Fysiikan tutkimuslaitos, “Tutkimuslaitos.” Verkkosivu. Saatavilla: https://www.hip.fi/?page_id=765. Viitattu 15.5.2017.
- [24] C. De Melis, “The CERN accelerator complex. Complexe des accélérateurs du CERN,” heinäkuu 2016. General Photo, OPEN-PHO-ACCEL-2016-

009. © CERN. Saatavilla: <https://cds.cern.ch/record/2197559>. Viitattu 15.5.2017.
- [25] The CMS Experiment at CERN, “Detector.” Verkkosivu. Saatavilla: <https://cms.cern/detector>. Viitattu 18.8.2017.
- [26] CERN, “CMS.” Verkkosivu. Saatavilla: <https://home.cern/about/experiments/cms>. Viitattu 18.8.2017.
- [27] S. Boreham, M. Brice, P. Ginter, C. Marcelloni, M. Hoch, and C. Collaboration, “Photos from the CMS Photo Book.” CMS Collection, elokuu 2008. CMS-PHO-GEN-2008-026. © CERN. Saatavilla: <http://cds.cern.ch/record/1344500>. Viitattu 24.11.2017.
- [28] D. Barney, “CMS Detector Slice.” CMS Collection, tammikuu 2016. CMS-PHO-GEN-2016-001. CC-BY-4.0. © CERN. Saatavilla: <http://cds.cern.ch/record/2120661>. Viitattu 24.11.2017.
- [29] S. Cittolin, A. Rácz, and P. Sphicas, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS, Geneva: CERN, 2002.
- [30] J. S. Briggs, “A derivation of the time-energy uncertainty relation,” *Journal of Physics: Conference Series*, vol. 99, no. 1, p. 012002, 2008. Saatavilla: <http://stacks.iop.org/1742-6596/99/i=1/a=012002>. Viitattu 4.1.2019.
- [31] The CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV,” *Journal of Instrumentation*, vol. 7, no. 10, p. P10002, 2012. Saatavilla: <http://stacks.iop.org/1748-0221/7/i=10/a=P10002>. Viitattu 4.1.2019.
- [32] Avoindata.fi-palvelu, “Sanasto.” Verkkosivu. Saatavilla: <https://www.avoindata.fi/fi/content/sanasto>. Viitattu 27.10.2017.
- [33] A. Haasio and R. Savolainen, *Tiedonhankintatutkimuksen perusteet*. Helsinki: BTJ Kirjastopalvelu Oy, 1. ed., 2004. ISBN: 951-692-577-4.
- [34] Creative Commons, “CC0 1.0 Yleismaailmallinen (CC0 1.0) Public Domain -lausuma.” Verkkosivu, 2017. Saatavilla: <https://creativecommons.org/publicdomain/zero/1.0/deed.fi>. Viitattu 27.10.2017.
- [35] DPHEP Study Group, “Data preservation in high energy physics.” Raportti, 2009. Saatavilla: <http://arxiv.org/abs/0912.0255>. Viitattu 27.10.2017.
- [36] A. Poikola, P. Kola, and K. A. Hintikka, *Julkinen data – johdatus tietovarantojen avaamiseen*. Helsinki: Edita Prima Oy, 2010. ISBN 978-952-243-146-2 (verkkójulkaisu). Saatavilla: <https://www.scribd.com/doc/28845102/Julkinen-data>. Viitattu 27.10.2017.

- [37] TIM, “TIM (The Interactive Material, vuorovaikutteinen materiaali).” Verkkosivu. Saatavilla: <https://tim.jyu.fi/view/tim/historia>. Viitattu 12.5.2017.
- [38] L. Rosen, *Open source licensing: software freedom and intellectual property law*. New Jersey: Prentice Hall Professional Technical Reference, 1. ed., 2005. ISBN: 0-13-148787-6.
- [39] Open Source Initiative, “The MIT License.” Verkkosivu, 2017. Saatavilla: <https://opensource.org/licenses/MIT>. Viitattu 12.5.2017.
- [40] webopedia, “VPN – virtual private network.” Verkkosivu, 2017. Saatavilla: <http://www.webopedia.com/TERM/V/VPN.html>. Viitattu 28.12.2018.
- [41] Jyväskylän yliopiston fysiikan laitos, “Hiukkasfysiikka.” Verkkosivu. Saatavilla: <https://www.jyu.fi/fysiikka/tutkimus/suurenergia>. Viitattu 23.5.2017.
- [42] Jyväskylän yliopiston matemaattis-luonnontieteellinen tiedekunta, “FYSA2002 Moderni fysiikka, osa B (6 op).” Verkkosivu. Saatavilla: <https://www.jyu.fi/ops/fi/science/fysiikan-kandidaattiohjelma/unit/16181>. Viitattu 17.5.2018.
- [43] Jyväskylän yliopiston fysiikan laitos, “FYSA2002 Moderni fysiikka, B-osa (6 op).” Verkkosivu. Saatavilla: <https://www.jyu.fi/science/fi/fysiikka/opiskelu/tyoosasto/tyot/fysa2002>. Viitattu 12.8.2018.
- [44] Jyväskylän yliopiston matemaattis-luonnontieteellinen tiedekunta, “Fysiikan maisteriohjelma, fysiikan opintosuunta.” Verkkosivu. Saatavilla: <https://www.jyu.fi/ops/fi/science/fysiikan-maisteriohjelma-fysiikka>. Viitattu 15.8.2018.
- [45] Jyväskylän yliopiston fysiikan laitos, “Kurssien työt ja työohjeet.” Verkkosivu. Saatavilla: <https://www.jyu.fi/science/fi/fysiikka/opiskelu/tyoosasto/tyot>. Viitattu 15.8.2018.
- [46] R. M. Barnett *et al.*, “Learning with the ATLAS experiment at CERN,” *Physics Education*, vol. 47, no. 1, p. 28, 2012. <http://stacks.iop.org/0031-9120/47/i=1/a=28>.
- [47] S. Sparaciari, “Dortmund students work with ATLAS data.” Verkkosivu, 2016. Saatavilla: <https://atlas.cern/updates/atlas-news/dortmund-students-work-atlas-data>. Viitattu 23.5.2017.
- [48] International Particle Physics Outreach Group (IPPOG), “International Masterclasses – hands on particle physics.” Verkkosivu. Saatavilla: <http://physicsmasterclasses.org/>. Viitattu 23.5.2017.
- [49] K. Cecire, T. McCauley and QuarkNet LHC fellows, “CMS Masterclass 2017 Documentation.” Verkkosivu. Saatavilla: <https://quarknet.i2u2.org/page/cms-masterclass-2017-documentation>. Viitattu 23.5.2017.

- [50] S. Suoniemi, “Avoin hiukkasfysiikan tutkimusdata opetuskäytössä.” Pro gradu -tutkielma, 2014. Helsingin yliopisto, fysiikan laitos. Saatavilla: <http://urn.fi/URN:NBN:fi-fe2017112252158>. Viitattu 2.3.2018.
- [51] P. Rikkilä, “Request for help for master’s thesis related to CMS Open Data.” Sähköpostiviesti, 2017. Viesti lähetetty ”cms-members” -sähköpostilistalle, joka kattaa kaikki CMS-kollaboraatioissa työskentelevät henkilöt. Viitattu 2.1.2019.
- [52] W. McComas, M. Clough, and H. Almazroa, “The role and character of the nature of science in science education,” in *The nature of science in science education: Rationales and strategies*, vol. 7, pp. 3–39, Kluwer Academic Publishers, tammikuu 1998. ISBN: 0-306-47215-5 (eBook).
- [53] Project Jupyter, “Project Jupyter.” Verkkosivu. Saatavilla: <http://jupyter.org/>. Viitattu 25.11.2017.
- [54] J. Sarén and T. Kalvas. Suullinen tiedonanto, kevät 2017. Viitattu 2.1.2019.
- [55] S. Hirsjärvi, P. Remes, and P. Sajavaara, *Tutki ja kirjoita*. Helsinki: Tammi, 15. ed., 2009. ISBN: 978-951-31-4836-2.
- [56] Tieteen termipankki, “Tietojenkäsittelytiede:pino.” Verkkosivu. Saatavilla: <http://tieteentermipankki.fi/wiki/Tietojenkäsittelytiede:pino>. Viitattu 9.8.2019.
- [57] Tutkimuseettinen neuvottelukunta, “Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen suomessa.” Tutkimuseettisen neuvottelukunnan ohje, 2012. Saatavilla: http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf. Viitattu 8.8.2018.

A Laboratoriotyön Jyväskylässä testattu versio

Liite A sisältää PDF-vedoksen työn siitä versiosta, jota Jyväskylän yliopiston fysiikan laitoksen opiskelijoista koostuva testiryhmä testasi. Vedos ei näytä visuaalisesti samalta kuin oikea, TIM-alustalla tarkasteltava työ. Esimerkiksi osa TIMin painikkeista on näkymättömissä sekä piilossa olevat vihjeet näkyvät jo valmiiksi aukaisuina. Vedoksen sisältö vastaa kuitenkin täysin työn sisältöä.

FYSA2002: Hiukkasfysiikan data-analyysi aidolla CERNin tutkimusdatalla

Tervetuloa tekemään modernin fysiikan kurssin B-osan lopputyötä *Hiukkasfysiikan data-analyysi aidolla CERNin tutkimusdatalla*. Varaa aikaa työn tekemiseen normaalin lopputyön verran, eli 4 tuntia. Työ kuitataan tehdyksi ilmoittamalla vastuuhenkilölle työn tekemisestä sähköpostilla (ohjeet lopussa). Vastauksesi tallentuvat suoraan TIMiin vastuuhenkilön nähtäväksi.

Ongelmatilanteissa voit olla yhteydessä XXXXX.

Työn rakenne ja tavoitteet

Tämän työn tavoitteena on tehdä yksinkertaista kokeellisen hiukkasfysiikan data-analyysia aidolla, CERNin CMS-ilmaisimella kerätyllä tutkimusdatalla. Tehtävänä on todeta Z-bosonin esiintyminen, selvittää Z:n massa analyysin perusteella ja verrata saatua massaa Particle Data Groupin (<http://pdg.lbl.gov/>) arvoon.

Työssä lasketaan CMS-ilmaisimella havaittujen myoniparien invariantteja massoja, luodaan näistä massoista histogrammi sekä analysoidaan histogrammia. Näiden vaiheiden kautta päästään käsiksi Z-bosonin esiintymiseen ja sen massaan.

Työn rakenne on seuraava:

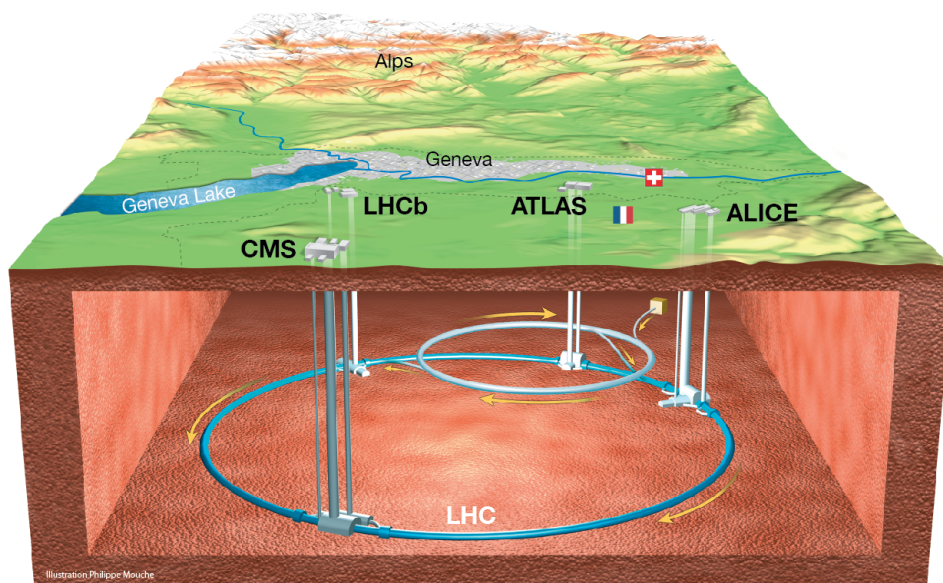
- Teoriatausta
- Invariantin massan laskeminen
- Histogrammin laatiminen
- Funktion sovittaminen histogrammiin
- Histogrammin analysointi
- Koko datasetin histogrammi

Ota nyt rento asento ja lue alussa huolellisesti teoriaosuus, sillä sen ymmärtäminen on oleellista työn tavoitteen saavuttamiseksi ja työstä oppimiseksi. Työssä on varattu reilusti aikaa teorian läpikäyntiin.

Teoriatausta

Hiukkasfysiikka on fysiikan osa-alue, jossa tutkitaan aineen ja säteilyn rakenteita sekä vuorovaikutuksia niiden välillä. Kokeellisessa hiukkasfysiikassa tutkimusta tehdään kiihdyttämällä ja törmäyttämällä hiukkasia toisiinsa tai kiinteisiin kohteisiin *hiukkaskiihdytimillä* sekä tutkimalla näitä törmäyksiä *hiukkasilmäisimillä*.

Euroopan hiukkasfysiikan tutkimuskeskus CERNissä sijaitsee maailman suurin hiukkaskiihdytin, Large Hadron Collider (LHC). LHC on 27 kilometriä pitkä ympyränmuotoinen synkrotronikiihdytin. Kiihdytin sijaitsee n. 100 metrin syvyydessä kulkevassa tunnelissa Ranskan ja Sveitsin rajalla (kuva 1).



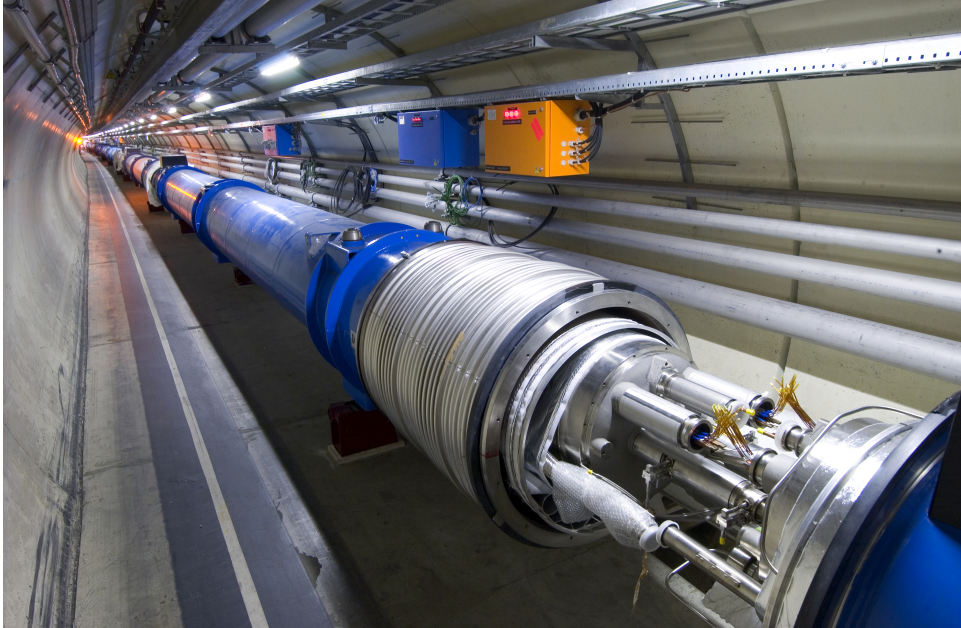
Kuva 1: LHC-kiihdytin ja sen neljä koeasemaa. © CERN [1]

Hiukkasten kiihdyttäminen

LHC:lla kiihdytetään pääasiassa protoneja. Protonilähteenä toimii vetyä sisältävä kaasupullo. Sähkökentän avulla vetyatomien kuorilta irrotetaan elektronit, jolloin saadaan tuotettua protoneja.

Kiihdyttämisprosessi alkaa jo ennen LHC-kiihdytintä. Protoneja kiihdytetään sähkökenttien avulla ja ohjallaan magneettikenttien avulla ennen LHC:lle saapumista Linac 2, Proton Synchrotron Booster, Proton Synchrotron ja Super Proton Synchrotron -kiihdytimillä, jolloin protonit saavuttavat 450 GeV:n suuruisen energian. Lisäksi näissä kiihdytimissä protonit ohjataan tasaisin välimatkoin sijaitseviksi kimpuiksi (bunches) kahteen eri protonisuihkuun (proton beams). Kummassakin suihkussa on tasaisesti n. 7,5 metrin välein 2808 protonikimppua. Kussakin kimpuissa taas on $1,2 \cdot 10^{11}$ protonia.

Tämän jälkeen nämä kaksi protonisuihkuja ohjataan LHC-kiihdytimelle. Suihkut kiertävät LHC:ssa eri suuntiin kahdessa eri tyhjiöputkessa. Kuvassa 2 on avattuna osa LHC-kiihdytintä. Rakenteen sisällä näkyvät tyhjiöputket. Kumpikin protonisuihku saavuttaa maksimissaan LHC:ssa n. 7 TeV:n (7000 GeV) suuruisen energian.



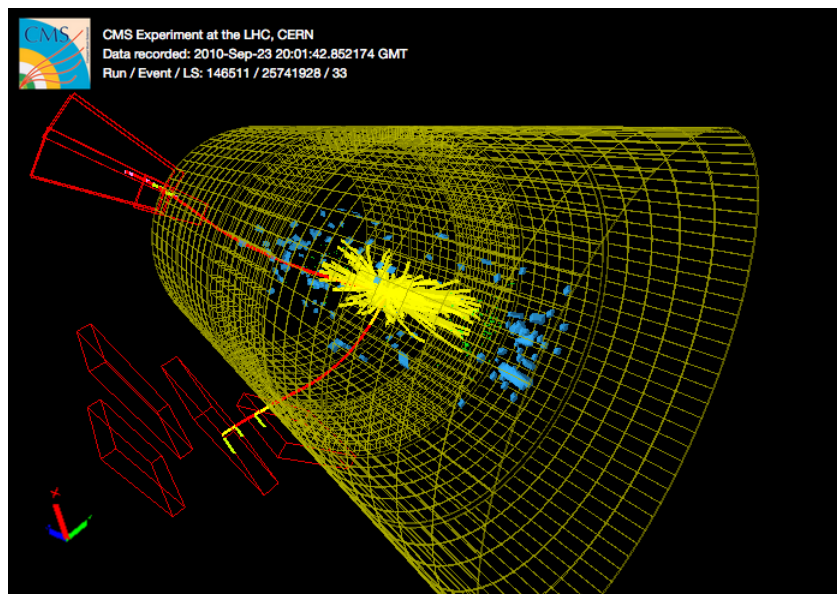
Kuva 2: Osa LHC-kiihdytintä avattuna, sisällä näkyvät mm. tyhjiöputket © CERN [2].

Hiukkastörmäyksiä luodaan risteyttämällä nämä eri suuntiin kulkevat suihkut. Kahden protonikimpun risteytyessä kuitenkin suurin osa protoneista ei törmää toisiinsa. Vain n. 40 protonia törmää keskenään muodostaen siis n. 20 törmäystä. Kimput kulkevat kuitenkin niin nopeasti, että ne risteävät LHC:ssä sekunnin aikana 40 miljoonaa kertaa. Yhteensä törmäyksiä on siis 800 miljoonaa kappaletta sekunnissa.

Maksimissaan törmäyksissä on energiaa yhteensä 14 TeV:n verran. Kuitenkin lähes aina törmäysenergia on tätä pienempi, sillä protonien törmätessä itse asiassa kvarkit ja gluonit törmäävät, jolloin protonien kaikki energia ei suinkaan välity törmäykseen.

Protonien törmätessä törmäysenergia voi muuttua massaksi ($E = mc^2$), eli törmäyksissä voi syntyä uusia hiukkasia. Törmäyksissä syntyneitä hiukkasia tutkimalla ja mittaamalla pyritään ymmärtämään paremmin esimerkiksi pimeää ainetta, antimateriaa ja sitä, mistä kaikki aine koostuu.

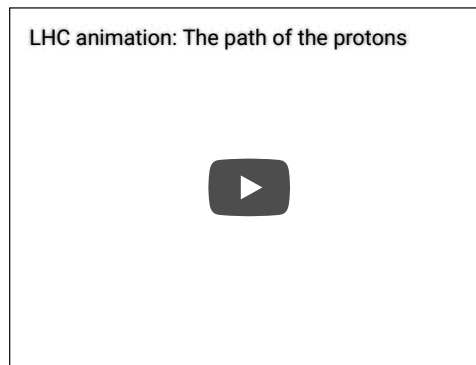
Kuvassa 3 on esitettyä visualisointi osasta yhdessä törmäystapahtumassa syntyneistä hiukkasista, jotka on havaittu CMS-ilmaisimella.



Kuva 3: Visualisoitu törmäystapahtuma.

Video

Kiihdytys- ja törmäytysprosessi on tiivistetty hyvin oheisella lyhyellä videolla. Katso video alusta kohtaan 1:15 asti saadaksesi kuvan siitä, miten prosessi etenee käytännössä.



Zoom: - r + hide video

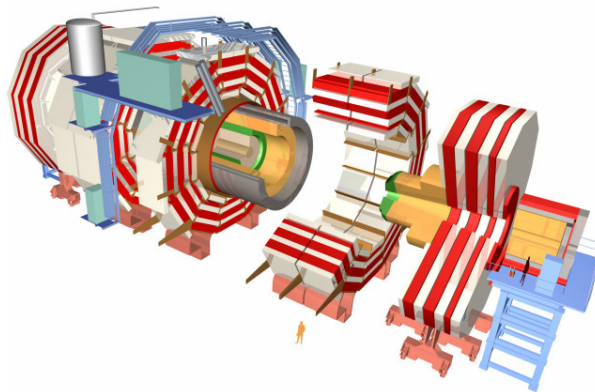
Video 1: CERNin julkaisema video LHC:n kiihdytysprosessista.

Hiukkastörmäysten tutkiminen

Hiukkastörmäyksiä tutkitaan *hiukkasilmäsimillä*. LHC:lla hiukkassuihkuja risteytetään neljässä eri kohtaa. Näillä kohdilla sijaitsevat LHC:n eri hiukkasilmäsimet: ATLAS, LHCb, ALICE ja CMS (ks. kuva 1). Tässä laboratoriotyössä keskitytään CMS-ilmaisimeen ja sen keräämän aidon tutkimusdatan analysointiin.

CMS, Compact Muon Solenoid, on yleiskäyttöinen hiukkasilmäsimin. CMS-kokeen tavoitteisiin kuuluvat mm. standardimallin tutkiminen, ylimääräisten ulottuvuuksien etsiminen sekä pimeää ainetta muodostavien hiukkasten löytäminen.

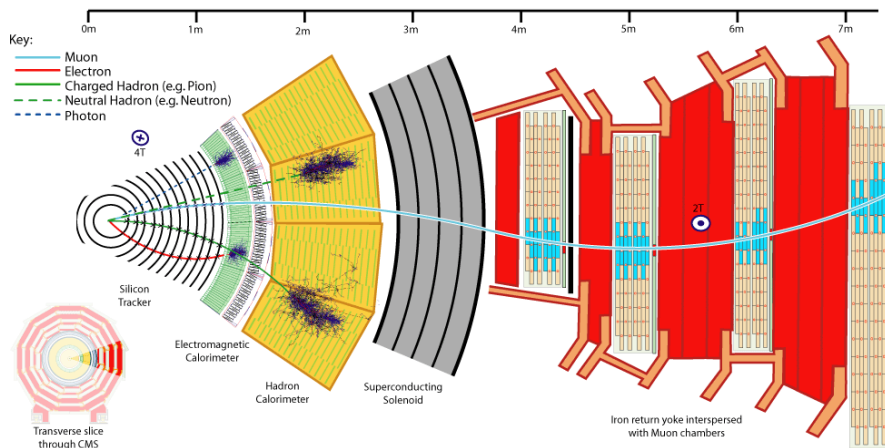
Tiivistettynä CMS-ilmaisimen tehtävä on havaita törmäyksissä esiintyneitä hiukkasia ja mitata niiden ominaisuuksia. Ilmaisim koostuu erilaisista detektoreista, joilla hiukkasia voidaan havaita ja mitata. Kuvassa 4 on esitettyä CMS-ilmaisimen rakenne.



Kuva 4: Havainnekuva CMS-ilmaisimesta. © CERN [3]

Detektorit muodostavat poikkileikkaukseltaan sipulimaisen rakenteen CMS:lle. Tällä rakenteella varmistetaan, että mahdollisimman moni törmäyksessä syntynyt hiukkanen havaitaan.

Eri hiukkasot käyttäytyvät eri tavoin CMS:n detektoreissa. Kuvassa 5 on esitettyä CMS:n poikkileikkaus siten, että hiukkassuihku kulkevat kuvaan ja kuvasta ulos. Kuva 5 havainnollistaa myös erilaisten hiukkasten käyttäytymistä CMS:ssa.



Kuva 5: Poikkileikkaus CMS-ilmaisimesta ja eri hiukkasten käyttäytyminen siinä. © CERN [4]

Sisimpänä olevan pii-ilmaisimen (silicon tracker) avulla voidaan rekonstruoida varattujen hiukkasten liikeradat. Voimakas magneettikenttä kaareuttaa varattujen hiukkasten liikeratoja. Kaarevuuden avulla voidaan laskea varattujen hiukkasten liikemäärät.

Hiukkasten energioita voidaan mitata kalorimetrien avulla. Elektronit ja fotonit pysähtyvät sähkömagneettiseen kalorimetriin (ECAL). Hadronit, kuten esimerkiksi protonit tai neutronit, taas läpäisevät ECAL:in, mutta pysähtyvät hadronikalorimetriin (HCAL).

Tästä pidemmälle CMS:ssa pääsevät vain myonit ja heikosti vuorovaikuttavat hiukkaset, kuten esimerkiksi neutriinot. Myoneiden liikemäärät ja energiat voidaan mitata myonikammioiden (muon chambers) avulla. Neutriinoja ei voida havaita suoraan CMS:lla, mutta niiden olemassaoloa voidaan päätellä puuttuvan energian (missing energy) avulla.

Kysymys 1

Työssä keskitytään CMS-ilmaisimella havaittuihin myoneihin. Miten kuvalet myonien käyttäytymistä ja havaitsemista CMS:ssä?

Tallenna vastauksesi tallennuspainikkeella. Oikealla alhaalla olevasta *wrap*-arvosta ei tarvitse välittää.

Kirjoita tähän vastauksesi

Datan tallentaminen

Kuten yllä mainittiin, LHC-kiihdyttimessä tapahtuu noin miljardi hiukkastörmäystä sekunnissa. CMS-ilmaisimella pystyy havaitsemaan nämä kaikki, mutta kaiken törmäyksistä saatavan datan tallentaminen olisi mahdotonta. Tätä varten välittömästi törmäyksen jälkeen erilaiset liipaisinsysteemit (trigger systems) päättävät, onko törmäys ollut potentiaalisesti mielenkiintoinen vai ei. Ei-mielenkiintoisia törmäyksiä ei tallenneta. Tämä monivaiheinen triggerointiprosessi vähentää tallennettavien törmäysten lukumäärän miljardista noin tuhanteen törmäykseen sekunnissa.

Törmäyksistä kerätty data tallennetaan ROOT-ohjelmistolla (<https://root.cern.ch/>) luettaviin AOD-tiedostoihin (Analysis Object Data). Tiedostojen rakenteet ovat monimutkaisia, eikä niitä voi käsitellä esimerkiksi yksinkertaisina dataa sisältävinä taulukkoina.

Tässä työssä käytetään AOD-tiedostojen sijaan CSV-tiedostomuotoa (comma separated values). CSV-tiedosto on tavallinen tekstitiedosto, joka sisältää pilkulla erotettuja arvoja (ks. kuva 6). Tällaisia tiedostoja voidaan helposti lukea ja käsitellä esimerkiksi Python-ohjelmointikiellä, kuten tässä työssä tehdään.

```
Run,Event,E1,px1,py1,pz1,pt1,eta1,phi1,Q1,E2,px2,py2,pz2,pt2,eta2,phi2,Q2,M
147115,366639895,58.7141,-7.31132,10.531,-57.2974,12.8202,-2.20267,2.17766,1,;
147115,366704169,6.61188,-4.15213,-0.579855,-5.11278,4.19242,-1.02842,-3.0020
147115,367112316,25.5419,-11.4809,2.04168,22.7246,11.661,1.42048,2.9656,1,15,;
147115,366952149,65.3959,7.51214,11.8871,63.8662,14.0619,2.21838,1.00721,1,25,;
147115,366523212,61.4504,2.95284,-14.6227,-59.6121,14.9179,-2.09375,-1.37154,
```

Kuva 6: Esimerkki CSV-tiedoston rakenteesta.

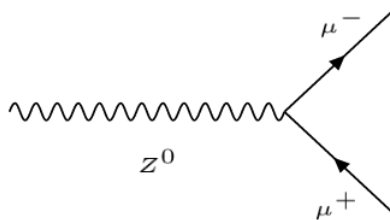
Hiukkasten epäsuora havaitseminen

Läheskään kaikkia hiukkasia ei voida havaita yllä mainitulla tavalla suoraan CMS:lla tai muilla hiukkasilmaisimilla. Etsittävät ja mielenkiintoiset prosessit ovat usein lyhytikäisiä. Niitä etsitään pitkäikäisten prosessien kautta, jolloin havaitseminen on siis epäsuoraa.

Esimerkiksi Z-bosonia, heikon vuorovaikutuksen välittäjähiukkasta, ei voida havaita suoraan CMS:lla, sillä Z:n elinikä on vain 0,000000000000000000000003 sekuntia. Z-bosoni ehtii siis jo hajota ennen kuin se edes saavuttaa pii-ilmaisinta.

Miten Z-bosoni sitten voidaan havaita? Ratkaisu kysymykseen löytyy Z:n hajoamisprosessista. Mikäli Z:n hajoamisesta peräisin olevat hiukkaset voidaan havaita CMS:lla, voidaan niiden avulla päätellä myös Z:n olemassaolo. Havaitseminen on siis epäsuoraa.

Z-bosoni voi hajota 24:llä eri tavalla, mutta tässä työssä tarkastelemme näistä vain yhtä, Z:n hajoamista myoniksi μ^- ja antimyoniksi μ^+ . Tämä prosessi on esitettyä Feynmanin diagrammina kuvassa 7.



Kuva 7: Z-bosonin hajoaminen myoniksi ja antimyoniksi.

Prosessissa syntyneet myonit voidaan havaita CMS:lla. Pelkkien myonien havaitseminen ei kuitenkaan ole riittävä todiste Z:n olemassaolosta, sillä havaitut kaksi myonia voisivat olla peräisin myös mistä tahansa muusta hiukastörmäyksessä tapahtuvasta prosessista (törmäyksissä on valtava määrä erilaisia prosesseja). Tämän takia myös Z:n massa täytyy rekonstruoida.

Invariantti massa

Z:n massaun päästään käsiksi *invariantti massa* -nimisen käsitteen avulla. Johdetaan seuraavaksi (löyhästi) lauseke invariantille massalle.

Tarkastellaan tilannetta, jossa M -massainen ja E -energinen hiukkanen hajoaa kahdeksi m_1 ja m_2 -massaiseksi sekä E_1 ja E_2 -energiseksi hiukaseksi. Hajoamisessa energia ja liikemäärä säilyvät, eli $E = E_1 + E_2$ ja $\vec{p} = \vec{p}_1 + \vec{p}_2$.

Hiukkaset noudattavat relativistista dispersiorelaatiota:

$$Mc^2 = \sqrt{E^2 - c^2\vec{p}^2}.$$

Energian ja liikemäärän säilymisellä tästä saadaan siis

$$\begin{aligned} Mc^2 &= \sqrt{(E_1 + E_2)^2 - c^2(\vec{p}_1 + \vec{p}_2)^2} \\ &= \sqrt{E_1^2 + 2E_1E_2 + E_2^2 - c^2\vec{p}_1^2 - 2c^2\vec{p}_1 \cdot \vec{p}_2 - c^2\vec{p}_2^2} \\ &= \sqrt{2E_1E_2 - 2c^2|\vec{p}_1||\vec{p}_2|\cos(\theta) + m_1^2c^4 + m_2^2c^4}. \end{aligned} \quad (1)$$

Relativistinen dispersiorelaatio voidaan saattaa muotoon

$$M^2c^4 = E^2 - c^2\vec{p}^2$$

$$E = \sqrt{c^2\vec{p}^2 + M^2c^4},$$

josta asettamalla $c = 1$ (hyvin yleinen tapa hiukkasfysiikassa) sekä olettamalla hiukkasten massa liikemäärään verrattuna hyvin pieneksi, saadaan

$$E = \sqrt{\vec{p}^2 + M^2} = |\vec{p}|\sqrt{1 + \frac{M^2}{\vec{p}^2}} \xrightarrow{M \ll |\vec{p}|} |\vec{p}|.$$

Soveltamalla edellä saatua tulosta $E = |\vec{p}|$ ja asetusta $c = 1$ lausekkeeseen (1), saadaan se sievennettyä muotoon

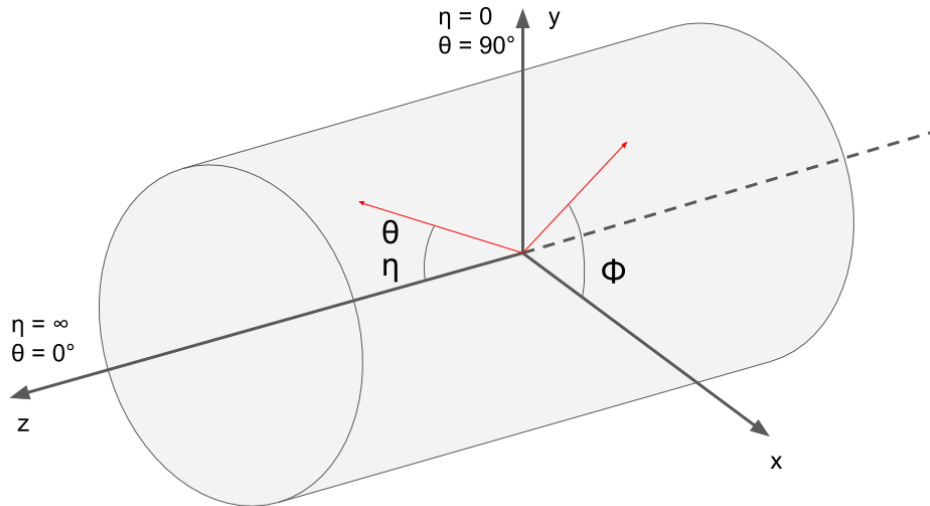
$$M = \sqrt{2E_1E_2(1 - \cos(\theta))},$$

jossa θ on hiukkasten liikemäärävektorien välinen kulma. Tällä lausekkeella voidaan laskea ns. **invariantti massa** hiukkasparille, jos vain tiedetään energiat ja hiukkasten liikemäärien välinen kulma.

Kokeellisessa hiukkasfysiikassa invariantille massalle käytetään usein seuraavaa lauseketta:

$$M = \sqrt{2p_{T1}p_{T2}(\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))}, \quad (2)$$

jossa p_T on hiukkasen liikemäärän hiukkasuuhkua vastaan kohtisuora komponentti, η pseudorapiditeetti ja ϕ atsimuuttikulma. Pseudorapiditeetti määritellään θ :n avulla lausekkeella $\eta = -\ln(\tan(\frac{\theta}{2}))$, eli se kuvaa kulmaa. Myös ϕ kuvaa kulmaa. Kuvassa 8 on esitettyä θ , η ja ϕ CMS-ilmaisimessa. Hiukkasuuhku kulkee z-akselin suunnassa. Kuvasta nähdään myös, että η :n määritelmästä johtuen sen arvo on 0, kun $\theta = 90^\circ$ ja ∞ , kun $\theta = 0^\circ$.



Kuva 8: CMS-ilmaisimen koordinaatteja.

Lisätietoa invariantin massan käsitteestä löydät osoitteesta <https://profmattstrassler.com/articles-and-posts/particle-physics-basics/mass-energy-matter-etc/mass-and-energy/>

Z:n massan rekonstruointi

Invariantin massan avulla päästään käsiksi Z:n olemassaoloon. Työssä keskitytään kuvan 7 mukaiseen Z:n hajoamiseen kahdeksi myoniksi.

Työssä käytettävään dataan on valittu sellaisia törmäystapahtumia, joissa on esiintynyt kaksi myonia (monien muiden hiukkasten lisäksi). Jokaisessa tapahtumassa esiintyvälle myoniparille voidaan laskea invariantti massa lausekkeen (2) avulla. Tätä toistetaan suurelle määrälle tapahtumia.

Mikäli myoniparille laskettu invariantti massa on arvoltaan yhtä suuri kuin Z-bosonin massa, voidaan sanoa, että kyseinen myonipari on peräisin Z:n hajoamisesta. Mikäli taas myoniparin invariantti massa on jotakin muuta, ovat myonit peräisin joistakin muista törmäyksessä esiintyneistä tapahtumista. **Eli invarianttia massaa voidaan käyttää todisteena Z:n olemassaolosta.**

Z:n tunnistaminen

Käytännössä Z:n tunnistamiseksi lasketaan kahden myonin invariantti massa suurelle määrälle törmäystapahtumia. Lasketuista massoista muodostetaan histogrammi, joka kertoo kuinka monta invariantin massan arvoa osuu millekin histogrammin jakovälille.

Mikäli histogrammiin syntyy piikki, eli paljon arvoja tietylle alueelle verrattuna muihin, voidaan saada todistetta että tapahtumissa on havaittu piikkiä vastaavan massan omaava hiukkanen. Histogrammiin voidaan sovittaa funktio, jonka parametreista saadaan oletetun Z-bosonin massa ja leveys. Z-bosonin massan teoreettinen ennuste on n. 91 GeV.

Kysymys 2

Miten voit käyttää invarianttia massaa apuna Z-bosonin esiintymisen todistamiseksi? Kuvaile omin sanoin tätä prosessia.

Kirjoita tähän vastauksesi

Teoriaosa päättyy tähän. Seuraavaksi voit siirtyä analysoimaan dataa.

Invariantin massan laskeminen

Tässä osiossa aloitetaan data-analyysi laskemalla törmäystapahtumissa esiintyneiden myoniparien invariantteja massoja edellä mainitulla tavalla. Laskemisessa ja histogrammin laatimisessa käytetään Python-ohjelmointikieltä.

Analyysissa käytetään CMS-ilmaisimen vuonna 2011 keräämää dataa. Alkuperäisestä datatiedostosta, joka löytyy sivulta <http://opendata.cern.ch/record/17> (avaa linkki uuteen välilehteen jotta et poistu TIMistä), on valittu CSV-tiedostoon vain ne törmäystapahtumat, joissa on esiintynyt täsmälleen kaksi myonia. Valinta on tehty linkin <http://opendata.cern.ch/record/552> kaltaisella koodilla. Käytännössä koodi valitsee alkuperäisestä datatiedostosta halutut arvot ja kirjoittaa ne CSV-tiedostoon.

Haetaan seuraavaksi CSV-datatiedosto ja aletaan töihin!

Teknisiä huomioita

- Jos saat työn aikana koodeja ajettaessa virheilmoituksen `Runtime exceeded, maybe loop forever`, kokeile aja samaa koodia vielä muutaman kerran uudestaan. Kyseinen ilmoitus kertoo TIMin laskenta-ajan ylittymisestä. Koska työssä käsitellään isoa datatiedostoa, voi tämä aika joskus ylittyä. Tällöin ajamista kannattaa kokeilla vielä muutaman kerran uudestaan. Toki syy voi olla myös kirjoittamassasi koodissa.
- Jos vahingossa muokkaat tai poistat koodia siten, että se ei enää toimi, pääset aina takaisin alkutilaan koodilohkon alle ilmestyvällä *Alusta*-painikkeella.
- Työ on rakennettu siten, että koodilohkot ajavat edellisiä aiempia lohkoja piilossa olevan koodin avulla.

Alustus ja datan hakeminen

Alla olevassa koodilohkossa haetaan työssä tarvittavat Python-moduulit *pandas*, *numpy* ja *matplotlib.pyplot* sekä nimetään ne lyhenteillä *pd*, *np* ja *plt*. Moduulit ovat tiedostoja, jotka sisältävät Pythonin määritelmiä ja komentoja. Moduulit haetaan siksi, että Pythonin oletuskomennoilla ei voida tehdä kaikkia työssä tarvittavia asioita.

Lisäksi haetaan TIMin serverille tallennettu datatiedosto käyttöön tallentamalla se muuttujaan, jolle annetaan nimeksi `ds`. **Älä vaihda muuttujan nimeä.** Tulostetaan myös datatiedoston rivien määrä (eli törmäystapahtumien määrä).

Aja alla oleva koodi painamalla *Aja*-nappia. Jos saat tulostettua törmäystapahtumien määrän, voit siirtyä eteenpäin. Koodi voi antaa kaksi virheilmoitusta liittyen `matplotlib.pyplot`-moduuliin, mutta näistä ei tarvitse välittää.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')
print("Törmäystapahtumien lukumäärä = %d" % len(ds))
```

Miltä tiedosto näyttää?

Tiedosto tallentui *pandas*-moduulin *DataFrame*-rakenteena (käytännössä taulukko) muuttujaan `ds`. Tulosta seuraavaksi datatiedoston 5 ensimmäistä riviä, jotta pääset tarkastelemaan mitä tiedosto sisältää. Komennolla `print()` voit tulostaa sulkujen sisälle asetettavan muuttujan. Komennolla `muuttujannimi.head()` taas saat tulostettua valitsemasi datatiedoston 5 ensimmäistä riviä vaihtamalla *muuttajannimen* tilalle muuttujasi nimen.

Kirjoita alle koodi, joka tulostaa datatiedoston 5 ensimmäistä riviä ja aja koodi "Aja"-napilla. Koeta ensin itse, mutta jos jäät jumiin, klikkaa alta vinkkiä.

Klikkaa avataksesi vinkin jos jäät jumiinVinkki: `print(tiedostonimi.head())`

Kirjoita koodi tähän:

Tulosteeseen ilmestyvät \-merkit kertovat, että rivi jatkuu tulosteen seuraaville riveille. Ylin rivi näyttää mitä kaikkea tietoa myonipareista tiedosto sisältää. Esimerkiksi E1 on ensimmäisen myonin energia ja E2 toisen myonin energia jne. Tässä listattuna eri arvot:

- Run = ajon numero, jonka aikana data on kerätty
- Event = törmäystapahtuman numero
- Type = myonin tyyppi, globaali myoni (G) on mitattu sekä pii-ilmaisimessa että myonikammiossa, tracker-myoni (T) on mitattu vain pii-ilmaisimessa (nämä luokittelut ovat hypoteeseja, tyyppiä ei tiedetä täysin varmasti)
- E = myonin energia
- px, pz, py = myonin liikemäärän erisuuntaiset komponentit
- pt = myonin liikemäärän hiukkassuuhkua vastaan kohtisuora komponentti
- eta = η eli pseudorapiditeetti, kulmaa kuvaava koordinaatti (ks. kuva 8)
- phi = ϕ eli atsimuuttikulma, myös kulmaa kuvaava koordinaatti (ks. kuva 8)
- Q = myonin sähköinen varaus

Invariantin massan laskeminen

Seuraavaksi laske datatiedostossa olevien arvojen avulla jokaisen tapahtuman myoniparille invariantti massa. Sinun tarvitsee kirjoittaa haluamasi laskun lauseke vain kerran, sillä ajettaessa koodi laskee kirjoitetun lausekkeen automaattisesti kaikille datatiedoston riveille.

Esimerkiksi jos haluaisit summata jokaiselle tapahtumalle myonien varaukset yhteen ja tallentaa ne muuttujaan *varaukset*, onnistuisi se seuraavasti:

```
varaukset = ds.Q1 + ds.Q2
```

Koodissa tulee siis kertoa, että Q1:llä ja Q2:lla viitataan muuttujassa *ds* oleviin arvoihin. Tämä kerrotaan lisäämällä muuttujan nimi ja piste halutun arvon eteen, kuten yllä on tehty.

Invariantin massan lausekkeessa esiintyvät neliöjuuren, kosinin ja hyperbolisen kosinin saat lataamamme *numpy*-moduulin avulla, jonka nimesimme *np*:ksi. Neliöjuuren saa komennolla `np.sqrt()`, kosinin komennolla `np.cos()` ja hyperbolisen kosinin `np.cosh()`. Sulkeiden sisään luonnollisesti tulevat ne asiat, jotka ovat laskussakin neliöjuuren ym. sisällä.

Kirjoita alle koodi, joka laskee invariantin massan jokaiselle datasetin törmäystapahtumalle. Tallenna laskemasi massat muuttujaan *invariantti_massa*, joka on jo kirjoitettu koodilohkoon valmiiksi.

Ajamisen jälkeen koodi tulostaa viisi ensimmäistä laskemaasi arvoa sekä kertoo tulosteen viimeisellä rivillä, laskitko massat oikein. Tarkistus tehdään piilotetulla koodilla.

Jos lasku ei onnistunut, korjaa käyttämäsi lauseketta kunnes saat ilmoituksen "Invariantit massat laskettu oikein!" Voit aina palauttaa koodin lähtötilanteeseen *Alusta*-painikkeella. Painikkeella *Highlight* voit vaihtaa editorin korostamaan ohjelmointikielen komentoja.

Apua invarianttien massojen laskemiseen löydät teoriaosasta. Lisäksi voit avata alla olevia vinkkejä, jos jäät täysin jumiin. Kokeile kuitenkin ensin itse erilaisia vaihtoehtoja ja yritä selvittää ilman vinkkejä!

Ensimmäinen vinkki

Vinkki: käytä laskemiseen lauseketta $\sqrt{2p_{T1}p_{T2}(\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))}$.

Toinen vinkki

Vinkki: kun kirjaat lausekkeeseen suureita, muista viitata mistä muuttujasta haluat suureet. Esim. jos tarvitset suureen *pt1*, kirjoita koodiin `ds.pt1`

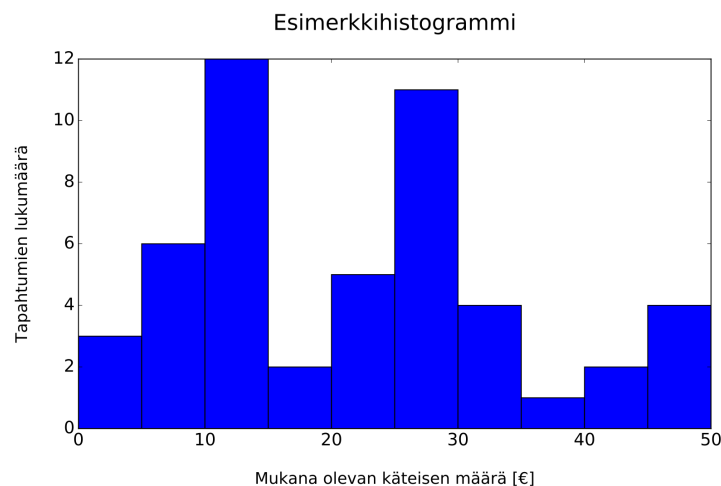
Kolmas vinkki

Vinkki: käytännössä kirjoita koodiin tekstin *invariantti_massa* = jälkeen vinkissä 1 mainittu lauseke yhdelle riville. Muista, että saat kosinin, hyperbolisen kosinin ja neliöjuuren *numpy*:n avulla yllä olevalla tavalla. Muista myös kertoa, mistä muuttujasta haluat suureet vinkin 2 tavalla.

```
invariantti_massa =
print('Viisi ensimmäistä laskemaasi arvoa:')
print(invariantti_massa[0:5])
```


Histogrammin laatiminen

Seuraavaksi luodaan histogrammi lasketuista invarianttien massojen arvoista. Histogrammi kuvaa arvojen jakaumaa, eli kuinka monta tapahtumaa on osunut millekin arvovälille. Kuvassa 9 on histogrammi, joka kuvaa mukana olevan käteisen määrän jakautumista jollekin ihmisjoukolle. Histogrammista nähdään esimerkiksi se, että yleisin käteisen määrä on ollut 10–15 euroa (12 henkilöllä).



Kuva 9: Esimerkkihistogrammi käteisen määrän jakautumisesta.

Histogrammin luominen

Pythonilla histogrammin luominen onnistuu alussa ladatun ja *plt*:ksi nimetyn *matplotlib.pyplot*-moduulin avulla. Komennolla `plt.hist()` voidaan luoda histogrammi antamalla sulkeiden sisälle erilaisia parametreja, joita voit tarkastella osoitteesta https://matplotlib.org/devdocs/api/_as_gen/matplotlib.pyplot.hist.html (avaa uuteen välilehteen, jotta et poistu TIMistä).

Nyt tarvitset ainoastaan kolmea ensimmäistä parametria: muuttujaa, jonka arvoista histogrammi luodaan (*x*), jakovälien määrää (*bins*) sekä tarkasteluväliä (*range*).

Kirjoita alle koodi, joka luo histogrammin laskemiasi invarianttien massojen arvoista. Koska pyritään tarkastelemaan Z-bosonin massan lähellä olevia invarianttien massojen arvoja, aseta tarkasteluväliksi esimerkiksi (60,120). Kokeile, mikä on sopiva jakovälien määrä selkeän histogrammin tuottamiseksi. Voit kokeilla, miten eri arvot vaikuttavat histogrammiin.

Koodissa on jo valmiina komennot histogrammin akseleiden ja otsikon nimeämiseksi. Lisäksi #-merkillä alkavat rivit ovat kommentteja, jotka eivät vaikuta koodin toimintaan.

Mikäli jäät jumiin, klikkaa alla olevia vinkkejä. Koeta kuitenkin luoda histogrammi ilman vinkkejä!

Huom: Jos saat ajaessa virheilmoituksen, mukana voi olla myös ilmoitus `No File: ...test.png`. Älä välitä tästä, vaan tätä edeltävästä virheilmoituksesta, joka kertoo mistä virhe johtuu.

Ensimmäinen vinkki

Vinkki: laskemasi invariantit massat ovat muuttujassa `invariantti_massa`.

Toinen vinkki

Vinkki: komento on muotoa `plt.hist(x, bins=0, range=(0,0))`, jossa `x` korvataan sen muuttujan nimellä, jossa histogrammiin plotattava data (tässä invariantit massat) sijaitsee, sekä nollat korvataan jakovälien lukumäärällä ja tarkasteluvälin ala- ja ylärajalla.

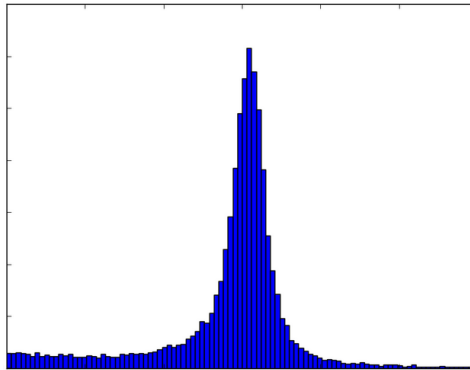
Kolmas vinkki

Vinkki: sopiva jakovälien määrä on n. 100 kappaletta.

```
# Kirjoita tähän alle komento, jolla saat luotua histogrammin.  
  
# Nimetään akselit ja otsikko. Älä muuta näitä.  
plt.xlabel('Invariantti massa [GeV]')  
plt.ylabel('Tapahtumien lukumäärä')  
plt.title('Kahden myonin invariantin massan histogrammi \n')
```

Tarkastus

Kun olet mielestäsi saanut luotua oikeanlaisen histogrammin, voit verrata sitä alta löytyvään mallihistogrammiin, josta on tarkoituksella poistettu lukuarvot. Mikäli oma histogrammisi eroaa mallista suuresti, palaa takaisin edelliseen kohtaan ja yritä uudestaan.

Klikkaa tästä avataksesi mallihistogrammin**Kysymys 3**

Kuvaile histogrammia. Mitä tietoa saat siitä?

```
Kirjoita tähän vastauksesi
```

Funktion sovittaminen histogrammiin

Jotta invarianttien massojen jakaumasta saataisiin ulos lukuarvoja, tulee histogrammin arvoihin sovittaa funktio. Arvot näyttävät normaalijakautuneilta, joten sovitetaan histogrammiin normaalijakauman funktio:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

jossa σ on keskihajonta ja μ odotusarvo.

Alla olevan koodin avulla voidaan optimoida normaalijakauman funktio histogrammin arvoihin. Sinulla on kaksi tehtävää:

1. Kirjoita koodiin `return` -sananjälkeen normaalijakauman lauseke.
2. Päättele aiemmin piirtämästäsi histogrammista, mitä σ :n ja μ :n arvot voisivat suurinpiirtein olla. Kirjaa nämä optimoinnin alkuarvaukset koodiin kohtaan `alkuarvot = [,]`

Koodiin nämä kohdat on merkitty isoilla kirjaimilla kirjoitetuilla `#KOMMENTEILLA`. Korvaa siis ko. kommentit yllä olevilla asioilla. Muut kommentit voit antaa olla ennallaan, saat niistä tietoa mitä koodissa tapahtuu.

Normaalijakauman lauseketta kirjoittaessa tarvitset seuraavia komentoja:

- π : `np.pi`
- neliöjuuri: `np.sqrt()`
- e^x : `np.exp(x)`
- potenssi Pythonissa merkitään `**`, eli esim 5^2 merkitään `5**2`
- keskihajonta: `sigma` (koodissa määritellään tämän niminen muuttuja keskihajonnaksi)
- odotusarvo: `mu` (koodissa määritellään tämän niminen muuttuja odotusarvoksi)

Muista, että muutosten jälkeen voit tarvittaessa palauttaa koodin lähtötilaan `Alusta` -painikkeella.

Tässä vielä normaalijakauman funktio toistamiseen, jotta näet sen koodia kirjoittaessasi:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Ensimmäinen vinkki

Vinkki: ole tarkkana normaalijakauman lausekkeen sulkeiden kanssa. Eli esim. jakolasku $\frac{1}{x-5}$ merkitään $1/(x-5)$

Toinen vinkki

Vinkki: mitä σ ja μ kuvaavat normaalijakaumassa? Katso luomastasi histogrammista mitä nämä arvot suurinpiirtein ovat.

Kolmas vinkki

Vinkki: jos päättelit esimerkiksi alkuarvoiksi (huom. täysin keksityt esimerkkiluvut) $\sigma = 12$ ja $\mu = 1300$, merkkää ne koodiin `alkuarvot = [12, 1300]`

Neljäs vinkki

Vinkki: Toimivan normaalijakauman lausekkeen saat alla olevalla koodilla: $(1/(\sigma * \text{np.sqrt}(2 * \text{np.pi}))) * \text{np.exp}(-(x - \mu) ** 2 / (2 * \sigma ** 2))$

```
# Rajaudutaan vain histogrammin piikin lähelle, jossa
# histogrammi on normaalijakautunut.
alaraja = 88
ylaraja = 94
binit = 100

# Valitaan rajojen sisällä olevien massojen arvot.
rajatutmassat = invariantti_massa[(invariantti_massa > alaraja) & (invariantti_massa
< ylaraja)]

# Luodaan valituista normitettu histogrammi normaalijakaumaa varten.
histogrammi = plt.hist(rajatutmassat, bins=binit, range=(alaraja,ylaraja), normed=True)

# Y-akselille tapahtumien määrä kullakin jakovälillä (saadaan muuttujasta histogrammi)
# x-akselille jakovälit luomalla lista, jossa binien määrä numeroita tasavälein
# histogrammin rajauksen matkalta.
y = histogrammi[0]
x = np.linspace(alaraja,ylaraja,binit)

# Määritellään, että halutaan sovittaa normaalijakauman funktio.
def gaussinen(x, sigma, mu):
    return #POISTA TÄMÄ KOMMENTTI JA KIRJOITA TÄHÄN NORMAALIJAKAUMAN LAUSEKE

# Alkuarvot optimointia varten. Ensimmäinen on keskihajonta (sigma)
# ja toinen odotusarvo (mu)
alkuarvot = [#TÄHÄN ALKUARVO SIGMALLE , #TÄHÄN ALKUARVO MY:LLE ]

# Haetaan optimointiin käytettävä moduuli ja tulostetaan tehty sovitus
# sekä siitä saadut keskihajonnan ja odotusarvon arvot.
from scipy.optimize import curve_fit
parhaat, kovarianssi = curve_fit(gaussinen, x, y, p0=alkuarvot)
print("Keskihajonnan (sigma) arvo = %f" % parhaat[0])
print("Odotusarvon (mu) arvo = %f" % parhaat[1])
plt.plot(x, gaussinen(x, *parhaat), 'r-', label='sigma = {}, mu = {}'.format(parhaat
plt.xlabel('Invariantti massa [GeV]')
plt.ylabel('Normitettu tapahtumien lkm')
plt.title('Normaalijakauman sovitus')
plt.legend()
```

Huom: mikäli sovittamasi funktio ei näytä normaalijakaumalta, palaa vielä takaisin tarkastamaan lauseke. Katso tarvittaessa apua vinkeistä yksi kerrallaan.

Histogrammin analysointi

Kysymys 4

Mitä voit sanoa Z-bosonin esiintymisestä histogrammin ja siihen sovitetun funktion pohjalta? Perustele vastauksesi teoriaosaa ja muita lähteitä apunasi käyttäen.

Kirjoita tähän vastauksesi

Kysymys 5

Mikä on tulostesi perusteella Z-bosonin massa? Miten selvität sen?

Kirjoita tähän vastauksesi

Kysymys 6

Avaa uuteen välilehteen Particle Data Groupin verkkosivut (<http://pdg.lbl.gov/>). Etsi sieltä arvo Z-bosonin massalle ja vertaa sitä saamaasi arvoon. Mitä havaitset?

Kirjoita tähän vastauksesi

Kysymys 7

Milloin Z-bosoni havaittiin ensimmäisen kerran ja mikä Z:n fysikaalinen merkitys on?

Kirjoita tähän vastauksesi

Kysymys 8

Miksi saamasi histogrammin leveys on suurempi kuin Particle Data Groupin Z-bosonille ilmoittama leveys $2,4952 \pm 0,0023$ GeV?

Kirjoita tähän vastauksesi

Kysymys 9

Jos energia ja liikemäärä saataisiin mitattua äärettömän tarkasti, saataisiinko Z:n massan kohdalle histogrammiin yksi täsmällinen piikki vai jakauma? Perustele vastauksesi.

A-16

Kirjoita tähän vastauksesi

Koko datasetin histogrammi

Lopuksi luodaan vielä esimerkkinä histogrammi koko datasetin invarianteista massoista, ilman rajautumista Z-bosonin piikin kohdalle.

Aja alla oleva koodi luodaksesi histogrammin. Huomaa, että y-akseli on logaritminen ja x-akselilla on invarianttien massojen arvojen 10-kantaiset logaritmit, eli $\log_{10}(\text{massan arvo})$.

```
binien_lkm = 500
# Lasketaan massoista logaritmit ja painotus.
inv_massa_log = np.log10(invariantti_massa)
painot = []
for a in invariantti_massa:
    painot.append(binien_lkm/np.log(10)/a)

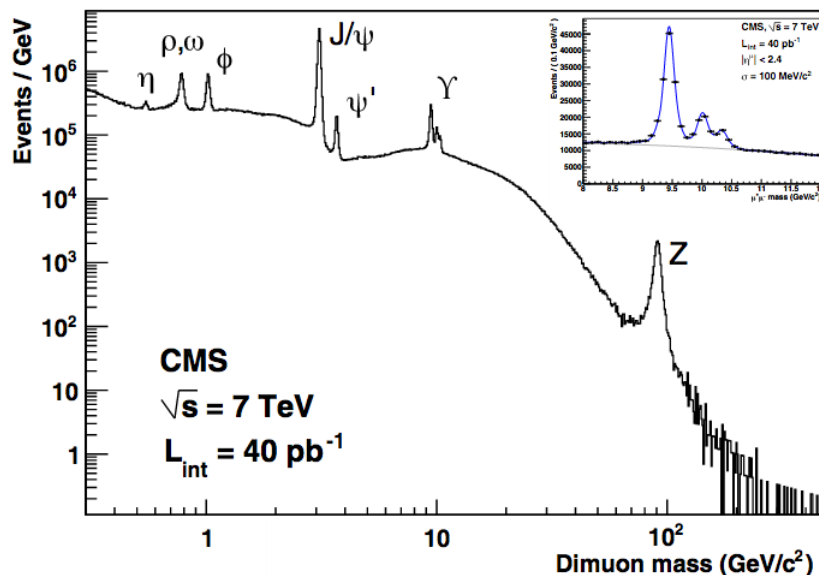
# Plotataan painotettu histogrammi.
plt.hist(inv_massa_log, binien_lkm, range=(0,2.5), weights=painot, lw=0, color="dark")
plt.yscale('log')

# Akselien ja otsikon nimeäminen.
plt.xlabel('log10(invariantti massa) [log10(GeV)]')
plt.ylabel('Tapahtumien lukumäärä')
plt.title('Kahden myonin invariantin massan histogrammi \n')
```

Kysymys 9

Vertaa luomaasi histogrammia alla olevassa kuvassa 10 näkyvään CMS-kokeen julkaisemaan histogrammiin. Mitä muita hiukkasia Z-bosonin lisäksi havaitset?

Kirjoita tähän vastauksesi



Kuva 10: Invarianteista massoista muodostettu histogrammi. © CMS Collaboration [5].

Lopuksi

Kaikki työssä tekemäsi koodin ajot sekä tallentamasi vastaukset ovat automaattisesti TIMissä työn vastuuhenkilön nähtävillä. Työtä ei siis tarvitse erikseen tallentaa.

Ilmoita kuitenkin nyt vastuuhenkilölle osoitteeseen XXXXXXXX, että olet tehnyt työn. Tarkastuksen jälkeen työstä kirjataan suoritusmerkintä Korppiin.

Lisätietoa CERNin avoimesta datasta löydät verkkosivulta <http://opendata.cern.ch/>.

Lähteet

[1] P. Mouche, *Overall view of the LHC. Vue d'ensemble du LHC*, 2014. Url: <https://cds.cern.ch/record/1708847>.

[2] M. Brice, *View of an open LHC interconnection. Vue d'une interconnection ouverte*, 2005. Url: <https://cds.cern.ch/record/905940>

[3] CMS Collaboration, *Detector Drawings*, 2012. Url: <https://cds.cern.ch/record/1433717>.

[4] M. Lapka, D. Barney, E. Quigg et al., *Interactive slice of CMS detector*, 2010. Url: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=4172>.

[5] CMS Collaboration, *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*, 2012. Url: [arXiv:1206.4071](https://arxiv.org/abs/1206.4071).

B Ensimmäisten testaajien informointi- ja esitietolomake

Liite B sisältää informointi- ja esitietolomakkeen, joka jaettiin työtä CERNissä testanneille opiskelijoille. Lomakkeen ensimmäisellä sivulla kerrotaan tietoa tutkimuksen kulusta ja aineistonkeruusta. Lisäksi testaaajalta pyydetään suostumus tutkimukseen osallistumiseen. Toisella sivulla olevalla esitietolomakkeella kerättiin tietoa mm. testaajien opinnoista, ohjelmointikokemuksista sekä hiukkasfysiikan tiedoista. Opiskelijat täyttivät tämän kaksipuolisen informointi- ja esitietolomakkeen ennen työn testaamista.

Kirjallinen informointi

Pro gradu -tutkielman kehittämistutkimus

Paavo Rikkilä

Hei,

olet osallistumassa Paavo Rikkilän pro gradu -tutkielmassa sovellettavaan kehittämistutkimukseen. Tutkimuksen tavoitteena on luoda ja kehittää Jyväskylän yliopiston fysiikan laitoksen modernin fysiikan kurssille laboratoriotyö sekä raportoida kehittämisprosessia. Lisäksi tutkimuksessa pyritään luomaan yleistettävää teoriaa laboratoriotyön kehittämisestä.

Laboratoriotyö käsittelee hiukkasfysiikan data-analyysia oikealla tutkimusdatalla. Työ tehdään Jyväskylän yliopiston tietotekniikan laitoksen TIM-alustalla tietokoneella.

Tutkimusta varten työtä testataan testiryhmällä, johon olet osallistumassa. Testaajia myös haastatellaan työhön liittyen. Testaamisesta kerätään seuraavaa aineistoa:

- Esitietolomake.
- TIM-alustaan tallentuvat tiedot työn kysymysten vastauksista ja koodilohkojen ajamisista.
- Havainnoijan (Paavo) kirjalliset muistiinpanot testaamisen ajalta.
- Äänitallenne haastatteluista. Haastattelun jälkeen äänitallenne litteroidaan ja anonymisoidaan sekä äänite tuhotaan.

Kerättyä aineistoa ei käytetä mihinkään muuhun kuin tämän tutkimuksen tarkoituksiin. Tutkimusaineistoa käsittelee vain yksi henkilö, Paavo Rikkilä. Julkaistusta tutkimuksesta ei voi tunnistaa tutkimukseen osallistuneita testaajia.

Lisätietoa tutkimuksesta voi pyytää Paavo Rikkilältä osoitteesta paavo.t.rikkila@student.jyu.fi

() Suostun osallistumaan yllä kuvattuun tutkimukseen.

Allekirjoitus

Paikka ja päivämäärä

Esitietolomake

Nimi: _____

1) Mikä on pääaineesi ja mitä sivuaineita olet lukenut vähintään 20 opintopistettä?

Pääaine:

Sivuaineet (väh. 20 op luettuna):

 kemia matematiikka tietotekniikka jokin muu, mikä/mitkä (kirjoita alle)

2) Kuinka monta lukuvuotta olet opiskellut tämänhetkistä tutkintoasi varten?

 0–1 2 3 4 5 muu (kirjoita luku)

3) Olen käynyt modernia fysiikkaa käsittelevän kurssin/kursseja

 kyllä en en ole varma

4) Olen käynyt ohjelmointiin liittyvän kurssin/kursseja

 kyllä en en ole varma

5) Olen käynyt hiukkasfysiikkaan liittyvän kurssin/kursseja

 kyllä en en ole varma

6) Olen mielestäni hyvä ohjelmoimaan

 samaa mieltä osittain samaa mieltä ei samaa eikä eri mieltä osittain eri mieltä eri mieltä

6) Hiukkasfysiikka on minulle tuttua

 samaa mieltä osittain samaa mieltä ei samaa eikä eri mieltä osittain eri mieltä eri mieltä

7) Olen joskus ohjelmoinut Python-ohjelmointikielellä

 kyllä en

C Toisten testaajien informointi- ja esitietolomake

Liite C sisältää informointi- ja esitietolomakkeen, joka jaettiin työtä Jyväskylässä testanneille opiskelijoille. Lomakkeen ensimmäisellä sivulla kerrotaan tietoa tutkimuksen kulusta ja aineistonkeruusta. Lisäksi testaaajalta pyydetään suostumus tutkimukseen osallistumiseen. Toisella sivulla olevalla esitietolomakkeella kerättiin tietoa mm. testaajien opinnoista, ohjelmointikokemuksista sekä hiukkasfysiikan tiedoista. Opiskelijat täyttivät tämän kaksipuolisen informointi- ja esitietolomakkeen ennen työn testaamista.

Kirjallinen informointi

Pro gradu -tutkielman kehittämistutkimus

Paavo Rikkilä

Hei,

olet osallistumassa Paavo Rikkilän pro gradu -tutkielmassa sovellettavaan kehittämistutkimukseen. Tutkimuksen tavoitteena on luoda ja kehittää Jyväskylän yliopiston fysiikan laitoksen modernin fysiikan kurssille laboratoriotyö sekä raportoida kehittämisprosessia. Lisäksi tutkimuksessa pyritään luomaan yleistettävää teoriaa laboratoriotyön kehittämisestä.

Laboriortyö käsittelee hiukkasfysiikan data-analyysia oikealla tutkimusdatalla. Työ tehdään Jyväskylän yliopiston tietotekniikan laitoksen TIM-alustalla tietokoneella.

Tutkimusta varten työtä testataan testiryhmällä, johon olet osallistumassa. Testaajia myös haastatellaan työhön liittyen. Testaamisesta kerätään seuraavaa aineistoa:

- Esitietolomake.
- Videotallenne pohjaluennoista.
- Testaajan tietokoneelta ruutukaappausvideo testauksen ajan.
- TIM-alustaan tallentuvat tiedot työn kysymysten vastauksista ja koodilohkojen ajamisista.
- Video- tai äänitallenteet mahdollisista keskusteluista testaajan ja havainnoijan (Paavo) välillä testaamisen aikana. Koko testaustapahtumaa ei videoida.
- Havainnoijan (Paavo) kirjalliset muistiinpanot testaamisen ajalta.
- Äänitallenne haastatteluista. Haastattelun jälkeen äänitallenne litteroidaan ja anonymisoidaan sekä äänite tuhoaan.

Kerättyä aineistoa ei käytetä mihinkään muuhun kuin tämän tutkimuksen tarkoituksiin. Tutkimusaineistoa käsittelee vain yksi henkilö, Paavo Rikkilä. Julkaistusta tutkimuksesta ei voi tunnistaa tutkimukseen osallistuneita testaajia.

Tutkimukseen osallistumisesta palkitaan yhdellä opintopisteellä. Opintopisteen saamisen edellytyksenä on kaikkiin tutkimuksen kolmeen osa-alueeseen osallistuminen (pohjaluento, testaaminen, haastattelu).

Lisätietoa tutkimuksesta voi pyytää Paavo Rikkilältä osoitteesta paavo.t.rikkila@student.jyu.fi

() Suostun osallistumaan yllä kuvattuun tutkimukseen.

Allekirjoitus

Paikka ja päivämäärä

Esitietolomake

Nimi: _____ Pääaine: _____

1) Kuinka monta lukuvuotta olet opiskellut pääainettasi?

 0–1 2 3 4 5 muu (kirjoita luku)

2) Mitä sivuaineita olet lukenut vähintään 20 opintopistettä?

 kemia matematiikka tietotekniikka kasvatustiede muu, mikä/mitkä _____

3) Olen käynyt kurssin FYSP106: Moderni fysiikka

 kyllä en en ole varma

4) Olen käynyt kurssin FYSP120: Numeeriset menetelmät

 kyllä en en ole varma

5) Olen käynyt kurssin FYSH300: Hiukkasfysiikka

 kyllä en en ole varma

6) Olen mielestäni hyvä ohjelmoimaan

 samaa mieltä osittain samaa mieltä ei samaa eikä eri mieltä osittain eri mieltä eri mieltä

6) Hiukkasfysiikka on minulle tuttua

 samaa mieltä osittain samaa mieltä ei samaa eikä eri mieltä osittain eri mieltä eri mieltä

7) Olen ohjelmoinut Python-ohjelmointikielellä

 erittäin usein usein jonkin verran muutamia kertoja en ollenkaan

D Pohjaluento

Liite D sisältää Jyväskylässä työtä testanneille opiskelijoille esitetyn pohjaluennon luentodiat. Luennon piti ja diat laati Jyväskylän yliopiston fysiikan laitoksen yliopistontutkija Sami Räsänen. Luentodiat käsittelevät mm. hiukkasfysiikan standardimallia, hiukkasten hajoamista, kokeellista hiukkasfysiikkaa CERNissä sekä invariantin massan laskemista.

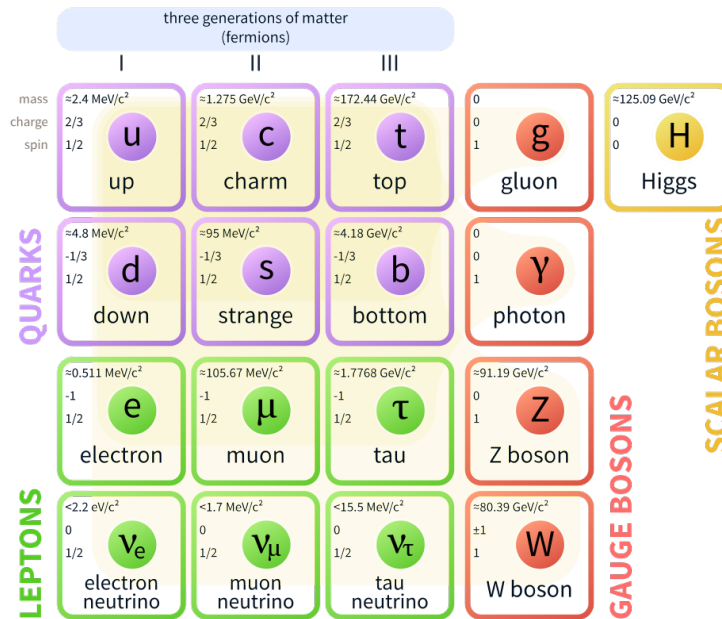
Uusi laboratoriotyö hiukkasfysiikasta – pohjaluento

**15.6.2017
Sami Räsänen**

Content of the lecture:

1. Z –boson in the Standard Model of particle physics
2. Invariant mass method

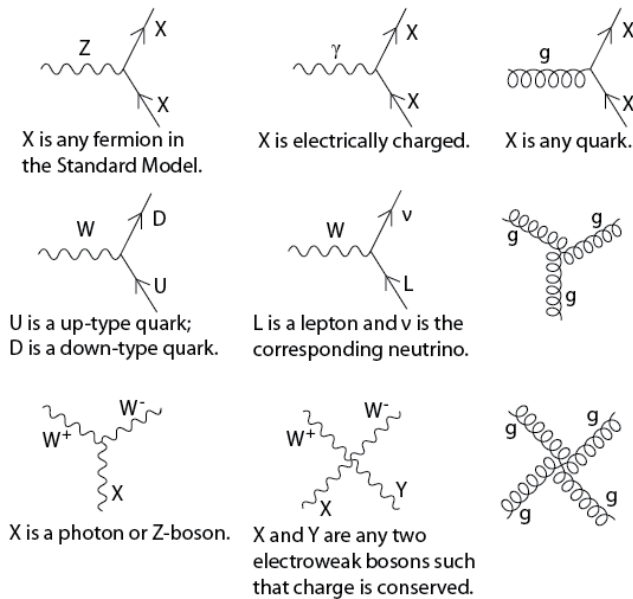
Standard Model of Elementary Particles



Our focus today is on Z-boson and muon μ

Interactions in the Standard Model – Feynman diagrams

Standard Model Interactions (Forces Mediated by Gauge Bosons)



Blackboard examples:

$$e^- + e^- \rightarrow e^- + e^-$$

$$q + g \rightarrow q + \gamma$$

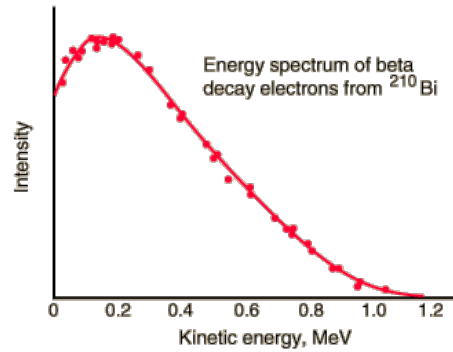
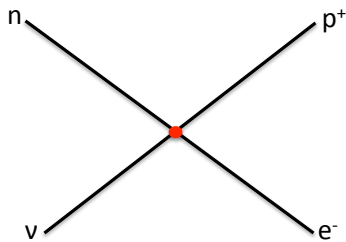
$$g + g \rightarrow H \rightarrow \gamma + \gamma$$

Some history: nuclear beta-decays

Beta-decays conserve mass number (A) but change atomic number (Z) by one unit.

$$\beta^- : (Z, A) \rightarrow (Z+1, A) + e^- + \bar{\nu}$$

$$\beta^+ : (Z, A) \rightarrow (Z-1, A) + e^+ + \nu$$



Fermi's model for beta-decays was based on contact interactions.

This model has a problem: violates conservation of probability at high energies

Electroweak force

Weinberg-Salam-Glashow (1968):

- Electromagnetic and weak interactions are different aspects of the same "electroweak" force
- All quarks and leptons have "weak charge"

Predictions:

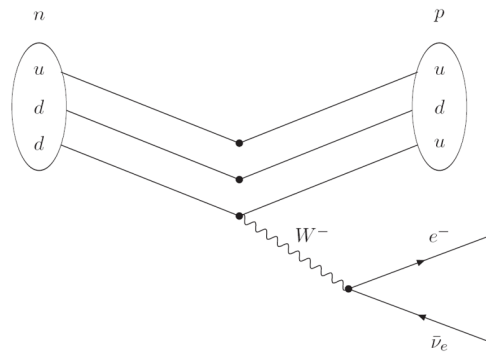
1. There are W-bosons that cause e.g. nuclear beta-decays
2. There is a "heavy photon", called Z-boson
3. Masses of W and Z should be of the order of 50-100 GeV $\gg m_p \sim 1$ GeV
4. Neutrinos are always "left-handed"

Nuclear beta-decays: W-boson emission

Beta-decays conserve mass number (A) but change atomic number (Z) by one unit.

$$\beta^- : (Z, A) \rightarrow (Z+1, A) + e^- + \bar{\nu}$$

$$\beta^+ : (Z, A) \rightarrow (Z-1, A) + e^+ + \nu$$



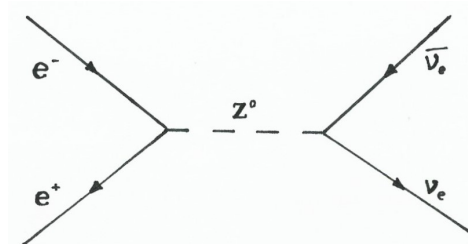
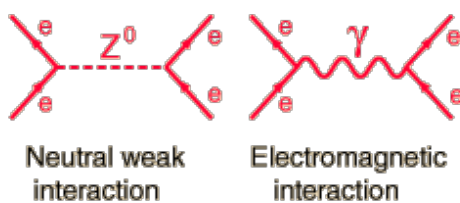
View of the particle physicist:

d quark emits W^- boson and changes to u \Leftrightarrow neutron changes to proton ($Z+1$)
 u quark emits W^+ boson and changes to d \Leftrightarrow proton changes to neutron ($Z-1$)


Neutral Z-boson

- Z-boson is similar with photon and photon interactions can always be mediated by Z as well.
- However, Z is so heavy that EM-reactions via photon are *usually* much more probable. Except around Z-peak that we shall discuss later.
- One key difference also is that Z-boson couples to neutrinos while photons do not, since neutrinos do not have electric charge.

Examples: $e^+ + e^- \rightarrow \gamma, Z \rightarrow \mu^+ + \mu^-$
 $e^+ + e^- \rightarrow Z \rightarrow \nu + \bar{\nu}$



Decay channels of the Z-boson:

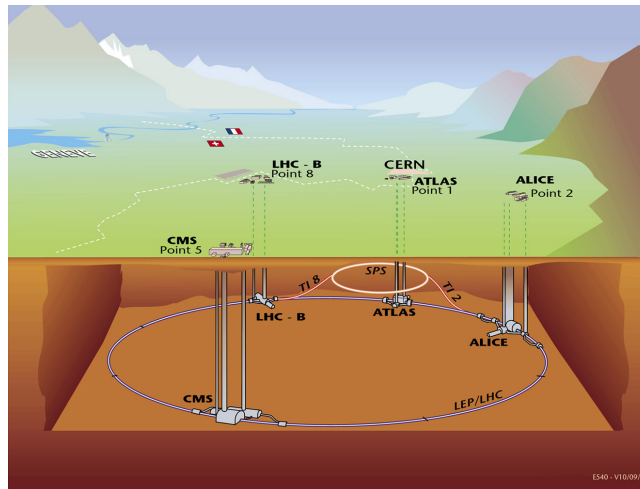
Z DECAY MODES		
Mode	Fraction (Γ_i/Γ)	Scale factor/ Confidence level
Γ_1 $e^+ e^-$	(3.363 \pm 0.004) %	
Γ_2 $\mu^+ \mu^-$	(3.366 \pm 0.007) %	
Γ_3 $\tau^+ \tau^-$	(3.370 \pm 0.008) %	
Γ_4 $\ell^+ \ell^-$	[a] (3.3658 \pm 0.0023) %	
Γ_5 invisible	(20.00 \pm 0.06) %	
Γ_6 hadrons	(69.91 \pm 0.06) %	
Γ_7 $(u\bar{u} + c\bar{c})/2$	(11.6 \pm 0.6) %	
Γ_8 $(d\bar{d} + s\bar{s} + b\bar{b})/3$	(15.6 \pm 0.4) %	
Γ_9 $c\bar{c}$	(12.03 \pm 0.21) %	
Γ_{10} $b\bar{b}$	(15.12 \pm 0.05) %	

Roughly 3.4 % of the Z-bosons decay to di-muon pairs

Research Questions in the lab work:

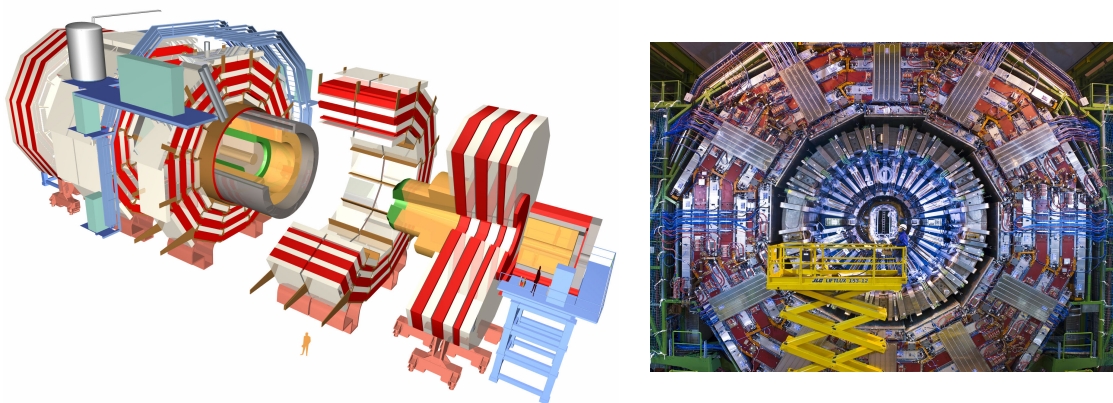
- how do we observe the Z-boson in the experiment?
- How do we measure the mass of the Z-boson?

LHC Collider at CERN:



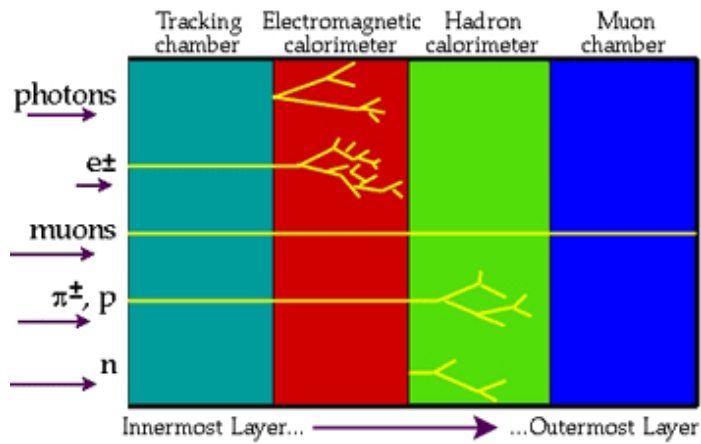
- 27 km circumference (old LEP tunnel), collision energy 13 TeV in pp
- 4 major experiments: ALICE, ATLAS, CMS and LHCb
- ATLAS and CMS: general experiments, new particles and interactions
- ALICE: heavy ion physics
- LHCb: heavy quarks and exotic hadronic states
- Finnish participation: Jyväskylä in ALICE and Helsinki in CMS

CMS-experiment at the CERN LHC:



- CMS = Compact Muon Solenoid
- Size: (length x diameter) = 21.6 m x 15 m, weights 12'500 tons
- Among other characteristics, very impressive detection of muons (which is important in this lab work)

Typical particle physics detectors:

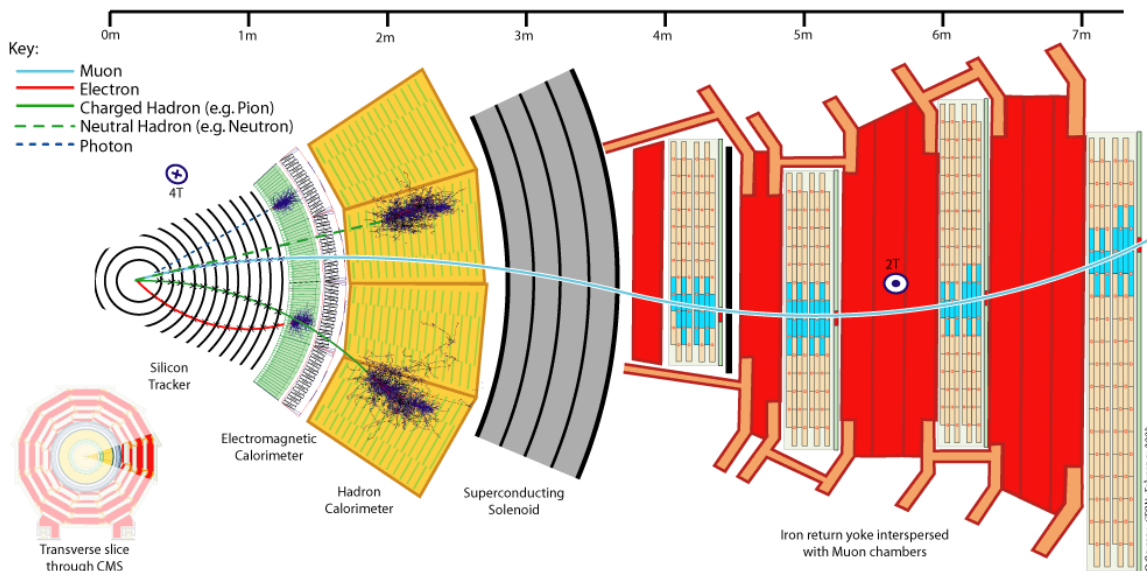


Particle must interact with the detector to be observed.

Two basic types:

1. Tracking detectors – do not absorb (destroy) the particle
2. Calorimeters – absorb the particle and measure its energy

CMS detector and particle measurements:



Muons interact fairly weakly with the matter => traverse through the whole detector

Partial summary:

- Z-boson acts in Standard model similarly as photon, except that it is very massive and couples also to neutrinos.
(Photon does not couple, because neutrinos do not have charge.)
- Mass of the Z-boson is $m_z = 91.1876 \pm 0.0021$ GeV. (Recall: $m_p \sim 1$ GeV.)
- How do we measure the Z-mass using open CMS-data?

We know that the Z-boson decays into di-muon pairs ($Z \rightarrow \mu^+ \mu^-$) in 3.4 % of the cases. CMS-experiment can measure muons. But how in practice we construct the mass of the particle from its decay products?

Lorentz transformations:

Let's take these for granted: If observer K' moves with constant velocity v with respect to the another observer K , then energies (E' and E) and momenta (p' and p) are related as:

$$E' = \gamma(E - vpc)$$

$$p' = \gamma\left(p - \frac{vE}{c^2}\right),$$

$$\text{where } \gamma \equiv 1 / \sqrt{1 - \left(\frac{v}{c}\right)^2}$$

Lorentz transformations:

Take another piece of information for granted:

The rest energy of the particle is given by its mass, $E = mc^2$.

Then we can find out what are the energy and momentum of the particle moving with velocity v . Since at the rest frame $p = 0$, we can boost ourselves to coordinate system with velocity $-v$ and find out:

$$E' = \gamma(E - (-v)pc) = \gamma(mc^2 + v \cdot 0) = \gamma mc^2$$

$$p' = \gamma\left(p - \frac{(-v)E}{c^2}\right) = \gamma\left(0 + \frac{vmc^2}{c^2}\right) = \gamma mv.$$

These relations are familiar e.g. from the high-school books and tables (MAOL).

Relativistic dispersion relation:

Relativistic (total) energy and relativistic momentum of a particle are:

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \equiv \gamma mc^2 \quad \text{and} \quad \vec{p} = \frac{m\vec{v}}{\sqrt{1 - \frac{v^2}{c^2}}} \equiv \gamma m\vec{v}$$

These obey (exercise) relativistic dispersion relation

$$E = \sqrt{c^2 \vec{p}^2 + m^2 c^4}$$

Classical limit, speed small as compared to speed of light ($v/c \ll 1$)

$$E \approx mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}m\frac{v^4}{c^2} + \dots, \quad \vec{p} \approx m\vec{v} + \frac{1}{2}m\frac{v^2}{c^2}\vec{v} + \dots, \quad v \equiv |\vec{v}|.$$

Relativistic total energy includes also the rest energy of the particle, $E = mc^2$.

Classical kinetic energy $E_{\text{kin}} = E - mc^2$ and momentum when $v/c \ll 1$.

If mass is zero or small as compared to momentum (i.e., when $v \rightarrow c$)

$$E = \sqrt{c^2 \vec{p}^2 + m^2 c^4} = c|\vec{p}| \sqrt{1 + \frac{m^2 c^2}{\vec{p}^2}} \xrightarrow{mc \ll |\vec{p}|} c|\vec{p}|$$

All observers agree on the mass of the particle:

Let's prove this statement using relativistic dispersion relation: Start from mass seen by the new observer (K') and calculate that it is the same as the original (K)

$$\begin{aligned}
 (m'c^2)^2 &= (E')^2 - (p'c)^2 = [\gamma(E - vp)]^2 - \left[\gamma \left(pc - \frac{vE}{c} \right) \right]^2 \\
 &= \gamma^2 \left(E^2 - 2vpE + \left(\frac{v}{c} \right)^2 (pc)^2 - (pc)^2 + 2vpE - \left(\frac{v}{c} \right)^2 E^2 \right) \\
 &= \gamma^2 \left(1 - \frac{v^2}{c^2} \right) (E^2 - (pc)^2) = \gamma^2 \frac{1}{\gamma^2} m^2 c^4 = (mc^2)^2 \\
 \Rightarrow m' &= m.
 \end{aligned}$$

We have proven that mass of the particle does not depend on the movement of the observer and is always constant. That information can be used in following.

Energy-momentum conservation in decay:

Particle with energy E and mass M decays into two new particles with energies E_1 and E_2 and masses m_1 and m_2 .

We know that energy and momentum is conserved, hence

$$E = E_1 + E_2$$

$$\vec{p} = \vec{p}_1 + \vec{p}_2$$

Relativistic dispersion gives

$$\begin{aligned}
 Mc^2 &= \sqrt{E^2 - c^2 \vec{p}^2} = \sqrt{(E_1 + E_2)^2 - c^2 (\vec{p}_1 + \vec{p}_2)^2} \\
 &= \sqrt{2E_1 E_2 - 2c^2 |\vec{p}_1| |\vec{p}_2| \cos \theta + m_1^2 c^4 + m_2^2 c^4} \\
 &\approx \sqrt{2E_1 E_2 (1 - \cos \theta)}, \text{ where } \theta \angle (\vec{p}_1, \vec{p}_2)
 \end{aligned}$$

The last approximation holds when $m_{1,2}c^2/E_{1,2} \ll 1$.

!! INVARIANT MASS METHOD – THIS IS WHAT YOU WILL DO !!

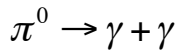
- Particle M decays as $M \rightarrow m_1 + m_2$
- Decay products m_1 and m_2 can be detected with the experiment
- Event-by-event, calculate invariant mass

$$m_{12}c^2 \equiv \sqrt{(E_1 + E_2)^2 - c^2(\vec{p}_1 + \vec{p}_2)^2} \approx \sqrt{2E_1E_2(1 - \cos\theta)}$$

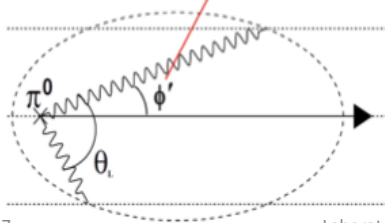
of the **all pairs** m_1 and m_2 measured in a single event

- Repeat procedure in millions or billions of events
- Decays of M are such that $m_{12} = M \Rightarrow$ peak! Rest combinatorial background.

Example:

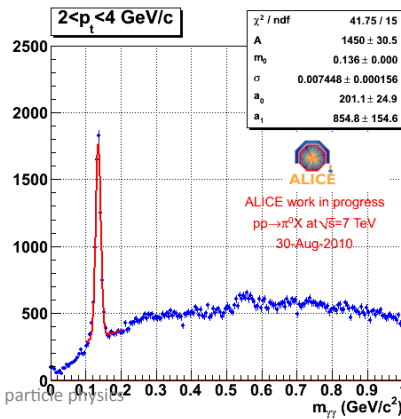


measurable quantity !



15.6.17

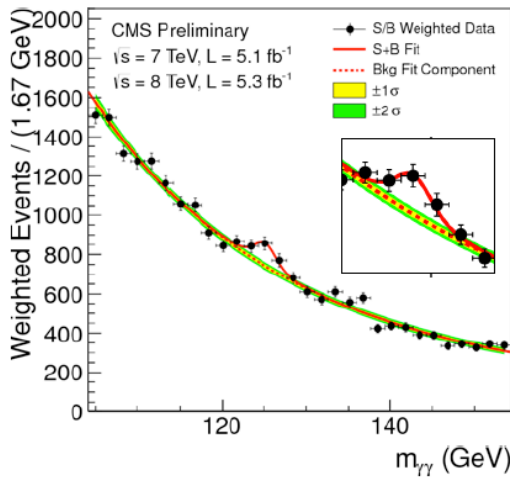
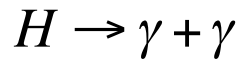
Laboratory work on particle physics



21

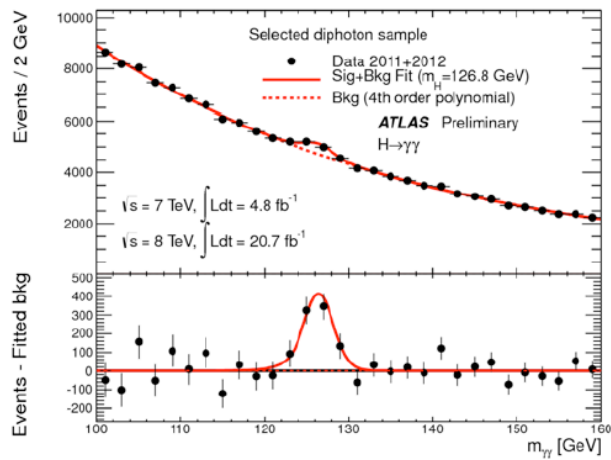
Examples on the use of the invariant mass method (1/3):

Detection of Higgs by ATLAS and CMS



15.6.17

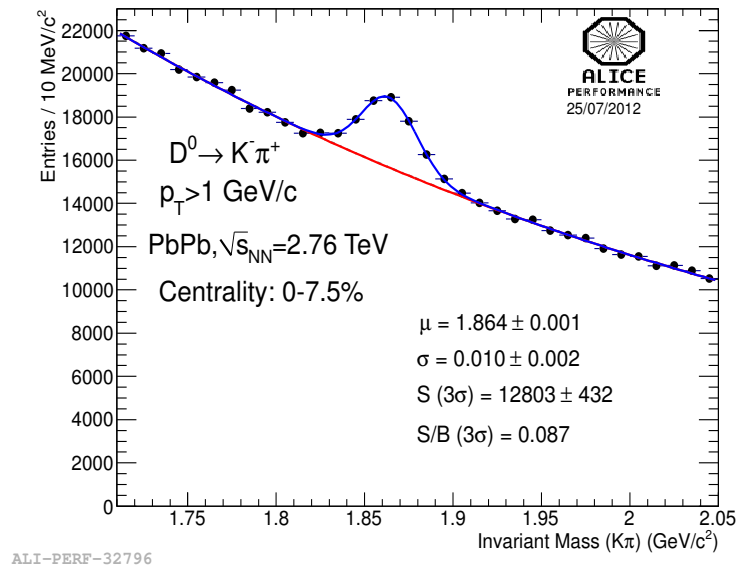
Laboratory work on particle physics



22

Examples on the use of the invariant mass method (2/3):

Detection of D^0 -meson by ALICE



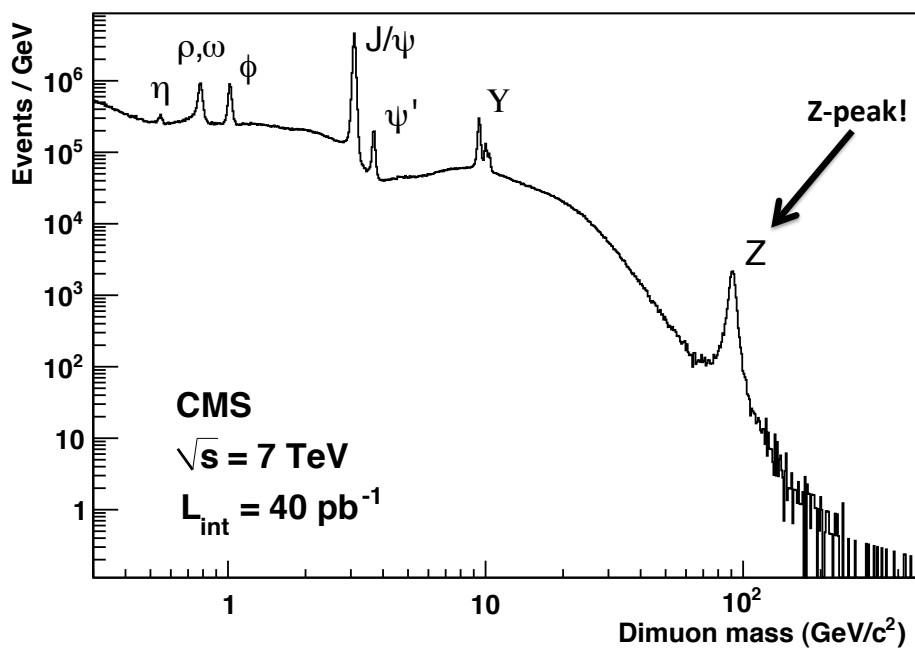
15.6.17

Laboratory work on particle physics

23

Examples on the use of the invariant mass method (3/3):

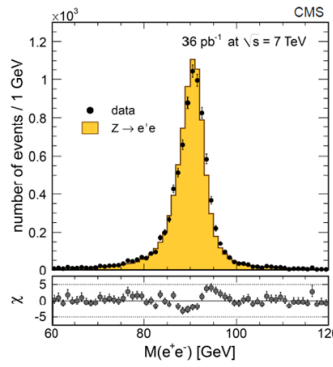
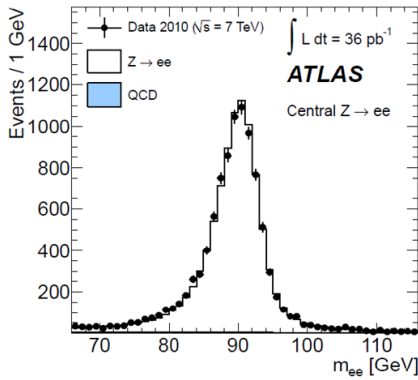
Beautiful di-muon mass distribution by CMS



15.6.17

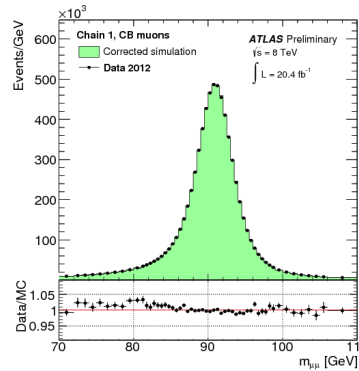
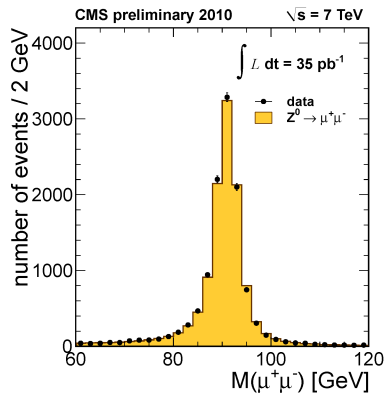
Laboratory work on particle physics

24



Different channels seen, examples to lepton channels

$$Z \rightarrow e^+ + e^-$$



$$Z \rightarrow \mu^+ + \mu^-$$

YOUR TASKS TOMORROW:

- Construct Z-peak from CMS data for di-muon pairs
- measure the mass of Z

E Työn osiot ”Funktion sovittaminen histogrammiin” ja ”Histogrammin analysointi” toisen kehittämisvaiheen jälkeen

Liite E sisältää PDF-vedoksen kehitetyn laboratoriotyön osioista ”Funktion sovittaminen histogrammiin” ja ”Histogrammin analysointi” toisen kehittämisvaiheen jälkeen. Näiden osioiden sisällöt muuttuivat merkittävästi toisessa kehittämisvaiheessa. Vedos ei näytä visuaalisesti samalta kuin oikea, TIM-alustalla tarkasteltava työ. Esimerkiksi osa TIMin painikkeista on näkymättömissä sekä piilossa olevat vihjeet näkyvät jo valmiiksi aukaistuina. Vedoksen sisältö vastaa kuitenkin täysin työn sisältöä.

Funktion sovittaminen histogrammiin

Jotta invarianttien massojen jakaumasta saataisiin tietoa havaitun resonanssin massasta ja elinajasta, tulee histogrammin arvoihin sovittaa jakaumaa kuvaava funktio. Tässä tilanteessa arvot noudattavat Breit-Wigner -jakaumaa:

$$N(E) = \frac{K}{(E - M)^2 + \frac{\Gamma^2}{4}},$$

jossa E on energia, M jakauman maksimi (vastaa resonanssissa havaittavan hiukkasen massaa), Γ jakauman puoliarvoveveys (decay width) sekä K vakio.

Breit-Wigner -jakauma voidaan ilmaista myös muodossa

$$\frac{2\sqrt{2}M\Gamma\sqrt{M^2(M^2+\Gamma^2)}}{\pi\sqrt{M^2+\sqrt{M^2(M^2+\Gamma^2)}}(E^2 - M^2)^2 + M^2\Gamma^2},$$

jossa vakio K on avattu.

Jakauman leveys Γ ja resonanssissa havaittavan hiukkasen elin aika (lifetime) τ ovat yhteydessä seuraavalla tavalla:

$$\Gamma \equiv \frac{\hbar}{\tau},$$

jossa \hbar on redusoitu Planckin vakio.

Alla olevan koodin avulla voidaan optimoida Breit-Wigner jakaumaa kuvaava funktio histogrammin arvoihin. Koodiin on jo kirjoitettu valmiiksi jakauman funktio. Sinun tehtäväksesi jää päätellä aiemmin tuotetun histogrammin avulla, mitä jakauman maksimin M ja jakauman leveyden Γ arvot voisivat suurinpiirtein olla. Kirjaa nämä optimoinnin alkuarvaukset koodiin kohtaan `alkuarvot = [#TÄHÄN ALKUARVO GAMMALLE , #TÄHÄN ALKUARVO M:LLE , -2, 200, 13000]`

Korvaa siis ko. kommentit päätelemilläsi arvoilla. Huomaa, että parametrien a , b ja A alkuarvot on jo annettu valmiina. Muut kommentit voit antaa olla ennallaan, saat niistä tietoa mitä koodissa tapahtuu.

Ajamisen jälkeen ohjelma tulostaa optimoinnin tuloksena saadut parametrien arvot epätarkkuuksineen, sekä tulostaa sovituksen kuvaajan. Parametrien epätarkkuudet saadaan laskettua sovituskomennon `curve_fit` palauttaman kovarianssimatriisin avulla.

— Ensimmäinen vinkki

Vinkki: Pohdi, miten M ja Γ voidaan määrittää kuvaajan avulla. Katso luomastasi histogrammista mitä nämä arvot suurinpiirtein ovat.

— Toinen vinkki

Vinkki: jos päättelit esimerkiksi alkuarvoiksi (huom. täysin keksityt esimerkkiluvut) $\Gamma = 12$ ja $M = 1300$, merkkää ne koodiin alkuarvot = [12, 1300, -2, 200, 13000]

```
# Rajaudutaan vain histogrammin piikin lähelle.
alaraja = 70
ylaraja = 110
binit = 100

# Valitaan rajojen sisällä olevien massojen arvot.
rajatutmassat = invariantti_massa[(invariantti_massa > alaraja) & (invariantti_massa
# Luodaan valituista massojen arvoista histogrammi.
histogrammi = plt.hist(rajatutmassat, bins=binit, range=(alaraja,ylaraja) )

# Y-akselille tapahtumien määrä kullakin jakovälillä (saadaan muuttujasta histogramm
# x-akselille jakovälkien keskikohtat
y = histogrammi[0]
x = 0.5*( histogrammi[1][0:-1] + histogrammi[1][1:] )

# Määritetään Breit-Wignerin jakaamaa kuvaava funktio sovitusta varten.
# E on energia, gamma on resonanssin leveys (decay width),
# M jakauman maksimi sekä a,b ja A parametreja,
# joilla huomioidaan sovituksessa taustatapahtumien vaikutus.
def breitwigner(E, gamma, M, a, b, A):
    return a*E+b+A*( (2*np.sqrt(2)*M*gamma*np.sqrt(M**2*(M**2+gamma**2)))/(np.pi*np.

# Alkuarvot optimointia varten seuraavassa järjestyksessä:
# gamma (resonanssin leveys)
# M (jakauman maksimi)
# a (lineaarisen taustan huomioimisessa käytetty kulmakerroin)
# b (lineaarisen taustan huomioimisessa käytetty y-akselin leikkauskohta)
# A (Breit-Wignerin jakauman "korkeus")
alkuarvot = [#TÄHÄN ALKUARVO GAMMALLE , #TÄHÄN ALKUARVO M:LLE , -2, 200, 13000]

# Haetaan optimointiin käytettävä moduuli, suoritetaan optimointi sekä
# lasketaan optimoitujen parametrien epätarkkuudet.
from scipy.optimize import curve_fit
parhaat, kovarianssi = curve_fit(breitwigner, x, y, p0=alkuarvot, sigma=np.sqrt(y))
virhe = np.sqrt(np.diag(kovarianssi))

# Tulostetaan sovituksesta saadut arvot virheineen
print("Sovituksesta saadut arvot virherajoineen:")
print("")
eka = "Resonanssin leveyden (gamma) arvo = {} +- {}".format(parhaat[0], virhe[0])
toka = "Jakauman maksimin (M) arvo = {} +- {}".format(parhaat[1], virhe[1])
kolmas = "a = {} +- {}".format(parhaat[2], virhe[2])
neljas = "b = {} +- {}".format(parhaat[3], virhe[3])
viides = "A = {} +- {}".format(parhaat[4], virhe[4])
print(eka)
print(toka)
print(kolmas)
print(neljas)
print(viides)

plt.plot(x, breitwigner(x, *parhaat), 'r-', label='gamma = {}, M = {}'.format(parhaa
plt.xlabel('Invariantti massa [GeV]')
plt.ylabel('Tapahtumien lkm')
plt.title('Breit-Wigner -sovitus')
plt.legend()
```

Huomio 1: mikäli sovittamasi funktio ei näytä osuvan hyvin histogrammiin, palaa vielä takaisin tarkastamaan alkuarvot. Katso tarvittaessa apua vinkeistä yksi kerrallaan.

Huomio 2: Sovituksessa otettiin huomioon se, että massajakauma sisältää myös niinsanottua taustaa, eli myonipareja, jotka ovat peräisin muista hajoamisista. Tausta huomioidaan koodissa Breit-Wignerin määrittelyssä komentoa `def breitwigner` seuraavalla rivillä. Sovitus sovitetaan osumaan taustaan termillä $aE + b$, jossa $aE + b$ huomioi taustan lineaarisen osan ja A korkeuden.

Huomio 3: Vielä oikeaoppisempi tapa sovitukseen ja sen parametrien virhearvojen määrittämiseen olisi iteroida sovitus useamman kerran. Iteroinnissa seuraava askel ottaisi alkuarvot aina edellisestä sovituksesta.

Histogrammin analysointi

Kysymys 4

Mitä voit sanoa Z-bosonin esiintymisestä histogrammin ja siihen sovitetun funktion pohjalta?

Pystytkö määrittämään Z:n massan virheeseen? Miten?

Perustele vastauksesi teoriaosaa ja muita lähteitä apunasi käyttäen.

Kirjoita tähän vastauksesi

Kysymys 5

Laske sovituksen avulla Z-bosonin elinaika τ virheeseen.

Vertaa saamaasi arvoa Z:n tunnettuun elinaikaan. Mitä havaitset? Mikä voisi mahdollisesti selittää havaintosi?

Kirjoita tähän vastauksesi

Kysymys 6

Milloin Z-bosoni havaittiin ensimmäisen kerran ja mikä Z:n fysikaalinen merkitys on?

Kirjoita tähän vastauksesi

Kysymys 7

Jos energia ja liikemäärä saataisiin mitattua äärettömän tarkasti, saataisiinko Z:n massan kohdalle histogrammiin muusta jakaumasta erottuva yksi täsmällinen piikki vai jakauma? Perustele vastauksesi.

Kirjoita tähän vastauksesi

F Työhön toisessa kehittämisvaiheessa lisätty osio

Liite F sisältää työhön toisessa kehittämisvaiheessa lisätyn osion PDF-vedoksen. Vedos ei näytä visuaalisesti samalta kuin oikea, TIM-alustalla tarkasteltava työ. Esimerkiksi osa TIMin painikkeista on näkymättömissä sekä piilossa olevat vihjeet näkyvät jo valmiiksi aukaistuina. Vedoksen sisältö vastaa kuitenkin täysin työn sisältöä.

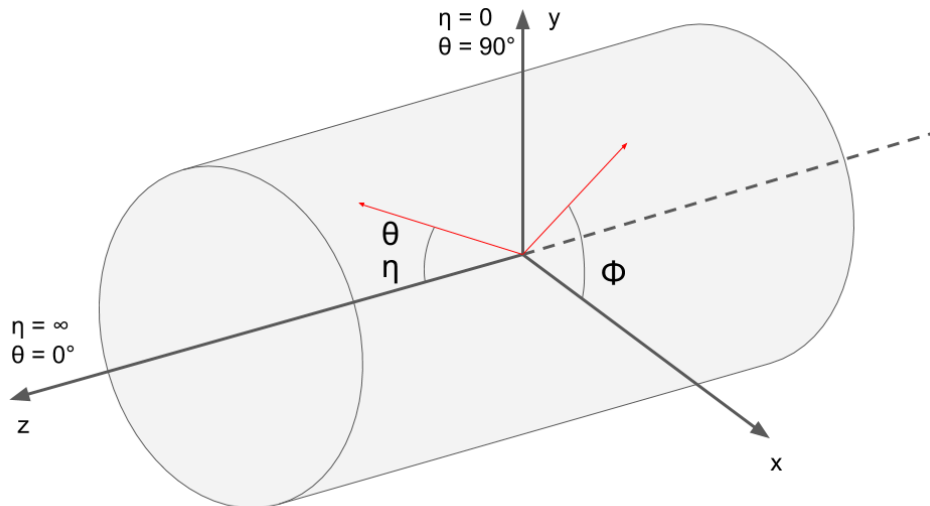
Pseudorapiditeetin η vaikutus massajakaumaan

Tässä viimeisessä osiossa perehdytään lyhyesti siihen, miten CMS-ilmaisimessa havaittujen myonien pseudorapiditeetit vaikuttavat massajakaumaan.

Kuten teoriaosassa kerrottiin, pseudorapiditeetti η kuvaa kulmaa, jolla havaittu hiukkanen on törmäyksen jälkeen poikennut hiukkassuihkusta (z-akselilta). Pseudorapiditeetti määritellään em. kulman θ avulla lausekkeella

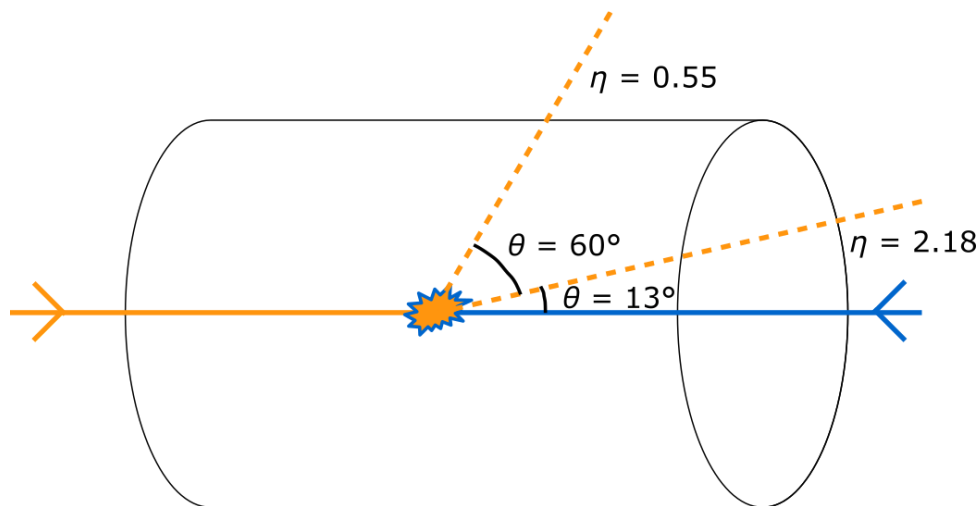
$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$$

Kertauksen vuoksi teoriaosan kuva 8 on esitetty alla. Kuvasta nähdään, että mitä pienempi pseudorapiditeetti hiukkasella on ollut, sitä enemmän se on poikennut törmäyksen jälkeen z-akselin suunnasta. Ja vastaavasti toisinpäin: suuren pseudorapiditeetin omaavat hiukkasat ovat jatkaneet törmäyksen jälkeen matkaansa lähes z-akselia (hiukkassuihkua) pitkin.



Kuva 8: CMS-ilmaisimen koordinaatteja.

Alla oleva kuva 11 havainnollistaa tilannetta, jossa vasemmalta ja oikealta tulevat hiukkassuihkut törmäävät. Kuvassa on esitettyä kaksi eri pseudorapiditeettia omaavaa myonia. Pienemmän pseudorapiditeetin myoni osuu ilmaisimen keskiosaan, kun taas suuremman pseudorapiditeetin myoni kohtaa ilmaisimen päädyn. Myös ilmaisimen päädyssä on myonikammio, joka rekisteröi osuman.



Kuva 11: Kaksi eri pseudorapiditeetin omaavaa hiukkasta.

Tässä viimeisessä osiossa tutkitaan, millä tapaa myonien pseudorapiditeetit vaikuttavat syntyvään massajakumaan. Tutkimusta varten tuotetaan kaksi erilaista histogrammia: toiseen valitaan vain suuren pseudorapiditeetin omaavia myonipareja, ja toiseen vain pienen. Histogrammit luodaan työn aiemmasta osiosta tutulla metodilla.

Tapahtumien valitseminen

Luodaan seuraavaksi tapahtumien jaottelua varten kaksi muuttujaa, `suuret_eta` ja `pienet_eta`. Ensimmäiseen tallennetaan vain ne törmäystapahtumat, joissa kummankin havaitun myonin pseudorapiditeetin itseisarvo on ollut suuri (esim. yli 1,52). Ja vastaavasti jälkimmäiseen vain ne, joissa kummankin myonin pseudorapiditeetin itseisarvo on ollut pieni (esim. alle 0,38). Itseisarvoja käytetään siksi, koska η voi saada myös negatiivisia arvoja.

Täydennä alla olevaa koodilohkoa määrittelemällä muuttujat `suuret_eta` ja `pienet_eta` siten, että edellä kuvattu jako tapahtuu. Tarvitset apuna seuraavia komentoja:

- `ds[vaatimus]` -komennolla saat valittua muuttujasta `ds` vain ne tapahtumat, jotka täyttävät sulkeiden sisään kirjoitettavan vaatimuksen. Vaatimuksia voi olla myös useampia, jolloin komento on muotoa `ds[(vaatimus1) & (vaatimus2)]`
- esimerkkinä tästä komento, jolla voisi valita muuttujasta `keksitty` vain ne rivit, joissa sarakkeiden `a` ja `b` arvot ovat kumpikin suurempia kuin 8: `keksitty[(keksitty.a > 8) & (keksitty.b > 8)]`
- itseisarvon saat numpy-moduulin komennolla `np.absolute()`
- ensimmäisen myonin pseudorapiditeetti on `ds.eta1` ja toisen `ds.eta2`
- "suurempi kuin" ja "pienempi kuin" -vertailut onnistuvat Pythonissa suoraan merkeillä `>` ja `<`
- desimaalierottimena on pilkun sijaan piste (esim. 0.38)

– Vinkki 1

Vinkki: muista valita pieniin mukaan ne tapahtumat, joissa sekä `eta1` että `eta2` ovat olleet pienempiä kuin 0,38. Ja vastaavasti suuriin.

– Vinkki 2

Vinkki: pseudorapiditeetteja hakiessasi muista kertoa, mistä muuttujasta näitä haetaan (kirjoita `ds.eta1` tai `ds.eta2`). Itseisarvoissa muista `np.` -alku ja lukuarvoissa se, että desimaalierottimena on piste.

– Vinkki 3

Vinkki: ensimmäinen muuttuja ehtoineen on `suuret_etat = ds[(np.absolute(ds.eta1) > 1.52) & (np.absolute(ds.eta2) > 1.52)]` ja toinen `pienet_etat = ds[(np.absolute(ds.eta1) < 0.38) & (np.absolute(ds.eta2) < 0.38)]`

```
# Haetaan tarvittavat moduulit.
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

# Tällä rivillä haetaan data ja tallennetaan se muuttujaan "ds".
ds = pd.read_csv('/cs/data/DoubleMuRun2011A.csv')

suuret_etat = #TÄHÄN TOINEN JAKO
pienet_etat = #TÄHÄN TOINEN JAKO

# Tulostetaan tietoa valinnasta.
print('Kaikkien tapahtumien lukumäärä = %d' % len(ds))
print('Niiden tapahtumien lkm, joissa kummankin myonin pseudorapiditeetti on ollut s')
print('Niiden tapahtumien lkm, joissa kummankin myonin pseudorapiditeetti on ollut p
```

Histogrammien luominen

Aja alla oleva koodilohko, jolla tuotetaan suuren ja pienen pseudorapiditeetin myoniparien omaavista tapahtumista erilliset histogrammit. Lohko hakee kummallekin jaottelulle invariantit massat, ja tuottaa näistä Z-bosonia vastaavaa piikkiä kuvaavat histogrammit.

```
# Erotellaan suurten ja pienten etan tapahtumien invariantit massat
# histogrammien luomista varten.
inv_massa_suuret = suuret_etat['M']
inv_massa_pienet = pienet_etat['M']

# Luodaan matplotlib.pyplot -moduulilla muokatun kokoinen kuva, johon
# plotataan kaksi histogrammia.
f = plt.figure(1)
f.set_figheight(15)
f.set_figwidth(15)
plt.subplot(211)
plt.hist(inv_massa_suuret, bins=120, range=(60,120))
plt.ylabel('suuret etat, tapahtumien lkm', fontsize=20)
plt.subplot(212)
plt.hist(inv_massa_pienet, bins=120, range=(60,120))
plt.ylabel('pienet etat, tapahtumien lkm', fontsize=20)
plt.xlabel('invariantti massa [GeV]', fontsize=20)
```

Histogrammien vertailu

Vertaile yllä olevia histogrammeja keskenään. Millä tapaa myonien pseudorapiditeetit vaikuttavat massajakaumaan?

Mikä voisi mahdollisesti selittää havaintosi?

Pohdi tilannetta ensin itse, ja tallenna vastauksesi alla olevaan kenttään. Lopuksi avaa kentän jälkeen esitetty selitys tilanteelle. Ei haittaa, vaikka oma pohdintasi eroaa tästä selityksestä, vastausta ei tarvitse korjata.

Kirjoita tähän vastauksesi

- Kun olet ensin itse pohtinut kysymystä, klikkaa tästä avataksesi selityksen tilanteelle

Histogrammeista voidaan havaita, että ne tapahtumat, joissa kummankin myonin pseudorapiditeetti on ollut pieni, tuottavat kapeamman piikin kuin suurten pseudorapiditeettien tapahtumat. Eli suurilla pseudorapiditeeteilla invarianttien massojen resoluutio on huonompi.

Invarianttien massojen huonompi resoluutio seuraa siitä, että suurilla pseudorapiditeeteilla myoneille mitatun poikittaisen liikemäärän p_t (liikemäärän hiukkassuihkua vastaan kohtisuora komponentti) resoluutio on huonompi. Tämän voi nähdä esimerkiksi CMS-kokeen julkaisun "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV" (<https://arxiv.org/pdf/1206.4071.pdf>) kuvassa 21 sivulla 32. Avaa linkki uuteen välilehteen, jotta et poistu työstä.

Selitys pseudorapiditeetin vaikutukseen resoluutioon on se, että ilmaisimen päätyyn suuntautuvat suuremman pseudorapiditeetin omaavat myonit vuorovaikuttavat todennäköisemmin ilmaisimen materiaalin kanssa kuin pienemmän η :n myonit (ks. kuva 11). Vuorovaikutuksissa myonit menettävät energiaansa. Tämä sotkee hieman myonien radan sovittamista ja p_t :n mittausta. Lisäksi tilanteeseen vaikuttaa se, että p_t :n mittausta riippuu herkästi mm. myonikammioiden suuntauksesta, ilmaisimen materiaalin jakautumisesta sekä magneettikentästä. Voidaan olettaa, että edellä mainitut tekijät ovat huonommin tunnettuja suuren pseudorapiditeetin omaavilla hiukkasilla.