THE APPLICATION OF MACHINE LEARNING TECHNIQUES TOWARDS THE OPTIMIZATION OF HIGH ENERGY PHYSICS EVENT SIMULATIONS WITHIN THE ALICE TRD AT CERN



Christiaan Gerhardus Viljoen

Department of Statistics

Faculty of Science

University of Cape Town

This dissertation is submitted in partial fulfilment of the Degree of Master of Science

Dedicated to my mother, Elizabeth Suzanna Bloem Viljoen, who has always inspired me to follow my higher passions, despite the myriad difficulties that life makes us face; and to search fearlessly and incessantly for the deeper truths underlying our everyday world.

“

A man may imagine things that are false, but he can only understand things that are true, for if the things be false, the apprehension of them is not understanding.

”

—Sir Isaac Newton

Declaration

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text.

It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

In accordance with the Department of Statistics guidelines, this thesis is does not exceed 20,000 words.

Signed:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Christiaan Gerhardus Viljoen, B.Sc., B.Sc. (Med.) (Hons.),

Cape Town

Summary / Abstract (Delete as Preferred)

The summary must be written in English and should consist of a piece of connected prose forming an abstract of the dissertation and be about 300 words in length. If at all possible, it should be accommodated on one side of A4 sized paper. It should bear the candidate's name and the exact title of the dissertation at the head of the page.

If candidates submit a soft-bound dissertation in the first instance, they will need when submitting the final hard-bound copy of the dissertation, to provide a further, loose-leaf, copy of this summary, identical to that bound into the final version, for the University Library file.

The summary will be considered by the examiners and, if the dissertation is approved, be deposited in the University Library for consultation and inter-library loan.

Acknowledgements

Firstly, I would like to thank my father, Christiaan Gerhardus Viljoen, for all the support – material, emotional and financial – he has selflessly provided to me throughout my life, and particularly towards my higher education journey. You have no idea how much appreciation I have for all the sacrifices you have made for me, and all the advice you have given me.

Secondly, I want to thank my aunt, Professor Emma Ruttkamp-Bloem, for all the mentoring she has provided to me in navigating the world of academia, and for the inspiration that her own academic career instils in me.

Thirdly, I want to thank Dr Thomas Dietel for providing me with this immense opportunity to be part of the largest scientific experiment in human history, and for the rigorous scientific guidance that he has, and continues to provide to me.

Lastly, I would like to thank my larger family, on both my father’s and mother’s side, for providing the loving and stable environment that makes any place we assemble Home.

Table of Contents

[1 Introduction 1](#_Toc536283904)

[1.1 Background 2](#_Toc536283905)

[1.2 Aims & Goals 3](#_Toc536283906)

[1.3 Summary of Work Done & Major Findings 4](#_Toc536283907)

[1.4 The Structure & Organization of this Dissertation 5](#_Toc536283908)

[2 High Energy Physics & The CERN Experiment 6](#_Toc536283909)

[2.1 A Brief History of Atomic Theory 7](#_Toc536283910)

[2.2 The Standard Model of Particle Physics 8](#_Toc536283911)

[2.2.1 Introduction 8](#_Toc536283912)

[2.2.2 The Fundamental Particles 9](#_Toc536283913)

[2.2.3 The Fundamental Forces 12](#_Toc536283914)

[2.2.4 The Higgs Boson 13](#_Toc536283915)

[2.3 The Quark Gluon Plasma 13](#_Toc536283916)

[2.4 Interactions of Particles with Matter 14](#_Toc536283917)

[2.4.1 The Bethe-Bloch Curve 14](#_Toc536283918)

[2.4.2 Transition Radiation 17](#_Toc536283919)

[2.5 The CERN Experiment 17](#_Toc536283920)

[2.5.1 Hardware 18](#_Toc536283921)

[2.5.2 Software 19](#_Toc536283922)

[2.5.3 Collaborations 20](#_Toc536283923)

[3 The ALICE Collaboration 21](#_Toc536283924)

[3.1 Objectives of the ALICE Experiment 22](#_Toc536283925)

[3.2 Gas Detectors 23](#_Toc536283926)

[3.3 The ALICE Detector 24](#_Toc536283927)

[3.3.1 The Transition Radiation Detector 25](#_Toc536283928)

[4 Deep Learning 27](#_Toc536283929)

[4.1 Deep Learning within the Context of Artificial Intelligence and Machine Learning 28](#_Toc536283930)

[4.2 Mathematical Background for Deep Learning 29](#_Toc536283931)

[4.2.1 Rosenblatt’s Perceptron 29](#_Toc536283932)

[4.2.2 Deep Feedforward Neural Networks 30](#_Toc536283933)

[4.3 Convolutional Neural Networks 36](#_Toc536283934)

[4.4 Variational Autoencoders 37](#_Toc536283935)

[4.5 Generative Adversarial Networks 38](#_Toc536283936)

[5 Data 39](#_Toc536283937)

[6 Methods 40](#_Toc536283938)

[6.1 ROOT 41](#_Toc536283939)

[6.2 Data Extraction from WLCG 42](#_Toc536283940)

[7 Results 43](#_Toc536283941)

[8 Discussion 44](#_Toc536283942)

[9 Conclusion 45](#_Toc536283943)

[10 Bibliography 47](#_Toc536283944)

[11 Appendices 48](#_Toc536283945)

[Appendix A: Plotting the Bethe-Bloch Equation 49](#_Toc536283946)

[Appendix B: Plotting Binary Cross-Entropy 56](#_Toc536283947)

List of Tables

[Table 1: The twelve fundamental fermions. 10](#_Toc535711721)

List of Figures

[Figure 1: Bethe-Bloch curve for a pion moving at relativistic speeds through silicon medium 15](#_Toc536283948)

[Figure 2: Bethe-Bloch curve for an electron moving at relativistic speeds through a silicon medium 16](#_Toc536283949)

[Figure 3: Illustration of the descent towards zero, of the Binary Cross Entropy Loss Function as ŷ, or , approaches the true y. 34](#_Toc536283950)

List of Abbreviations and Acronyms

|  |  |
| --- | --- |
|  |  |
| ALICE | A Large Ion Collider Experiment |
| TRD | Transition Radiation Detector |
| CERN | European Organization for Nuclear Research |
| QGP | Quark Gluon Plasma |
| LHC | Large Hadron Collider |
| WLCG | Worldwide LHC Computing Grid |
| QCD | Quantum Chromodynamics |
| QGP | Quark-Gluon Plasma |
| ML | Machine Learning |
| Pb-Pb | Lead-Lead Collisions |
|  | Electron |
|  | Pion |
| QED | Quantum Electrodynamics |
| p | Proton |
| n | Neutron |
|  | Electron Neutrino |
|  |  |
|  |  |

List of Appendices

[Appendix A: Plotting the Bethe-Bloch Equation 49](#_Toc536283951)

[Appendix B: Plotting Binary Cross-Entropy 56](#_Toc536283952)

# Introduction

## Background

This Masters Dissertation seeks to apply cutting edge techniques in Machine Learning (ML) towards the simulation of High Energy Physics (HEP) collision events, which routinely occur at the Large Hadron Collider (LHC) as part of the ongoing fundamental research conducted by the Counsel for European Nuclear Research (CERN).

More specifically, the focus of this thesis centres around the development of Deep Generative Models which are able to produce datasets that are indistinguishable from data produced by the Transition Radiation Detector (TRD) at the ALICE (A Large Ion Collider Experiment) collaboration at CERN, during Lead-Lead (Pb-Pb) heavy ion collisions.

## Aims & Goals

## Summary of Work Done & Major Findings

## The Structure & Organization of this Dissertation

# High Energy Physics & The CERN Experiment

## A Brief History of Atomic Theory

The earliest correct model for the atom can be traced back to 400 BCE, when Democritus proposed that the entire universe consisted of fundamental particles, or “Atoms”, which cannot be divided any further.

In 1803, Dalton refined this model to state that these indivisible atoms can have distinguishing chemical and physical traits and that they combine to form chemical compounds.

Then, in 1987, JJ Thompson discovered the electron and proposed an incorrect theory for subatomic structure in which negatively charged electrons were embedded within positive charges within the atom.

Rutherford, Marsder and Geiger disproved this model in 1911, with their seminal alpha-particle scattering experiment and put forth a more accurate model for the atom, in which most of the atom consists of empty space, with a dense core of positively charged protons.

In 1913, Bohr refined this model further, indicating that electrons orbit the positively charged atomic core at distinct energy levels. While this model did explain the emission spectrum of Hydrogen, it could not explain the emission spectra of any of the other elements.

Between 1924 – 1928, De Broglie, Heisenberg and Schrödinger each separately developed a quantum paradigm, where electrons have wave-like properties and appear in much more complex orbitals. This is still the accepted theory of atomic structure today.

There have been some refinements made to the quantum theory, as new information has come to light: a neutral subatomic particle, the neutron, was discovered in 1932, which solved the puzzle of why atoms were found to be nearly twice as heavy as expected based on proton number; this discovery also disproved Dalton’s second law, which stated that all atoms of a specific element were identical, and resulted in the concept of isotopes (atoms with the same number of protons, but differing numbers of neutrons). In the same year, Cockroft and Walton split the atom for the first time, by bombarding Lithium atoms with electrons, splitting them into two Helium particles.

The 1950s brought about a new era in nuclear physics, in which particle accelerators with collision energies of a few hundreds of MeVs became affordable, along with cosmic ray and inelastic proton-scattering experiments; since this time, a whole host of subatomic elements have been discovered, many of which are unstable. The discovery of these new particles has led, over time, to the development and refinement of the modern Standard Model of Particle Physics.

## The Standard Model of Particle Physics

### Introduction

The Standard Model of Particle Physics is a framework which allows us to understand the fundamental structure and dynamics of our universe in terms of elementary particles, where all interactions between elementary particles are similarly facilitated by an exchange of particles. In summary, based on our current understanding, our entire universe consists of a very sparse array of fundamental particles once we delve into the subatomic realm (1).

At an energy scale of electron Volts (an electron Volt is a unit of energy, equivalent to the amount of work required to accelerate a single electron through a potential difference of 1 Volt), the low energy manifestation of Quantum Electrodynamics (QED) allows atoms to exist in bound states with negatively charged electrons () orbiting a positively charged nucleus consisting of positively charged protons () and electrically neutral neutrons (), based on the electrostatic attraction of these opposing electrical charges (1).

Quantum mechanics explains the emergence of unique physical properties in different elements, which arise from their exact electronic structures. Quantum Chromodynamics (QCD) is the fundamental theory of the strong interaction, which binds protons and neutrons together within the nucleus of the atom. Similarly, at this energy scale, the weak force causes nuclear β-decays of radioactive isotopes and is involved in the nuclear fusion processes that occur within stars; the nearly massless electron neutrino ) is produced during both of the abovementioned processes (1).

Therefore, almost all physical phenomena that occur under normal circumstances can be explained by the Electromagnetic-, Strong- and Weak Forces, Gravity (which is very weak, but explain the large-scale structure of the universe), and just four fundamental particles: the electron, proton, neutron and electron neutrino (1).

### The Fundamental Particles

At higher energy scales, of the order of electron Volt (or giga-electron Volt, 1 GeV), protons and neutrons are understood to be bound states of truly fundamental particles called quarks, in the following manner: protons consist of two up-quarks and a down-quark p(uud), whereas neutrons consist of two down-quarks and an up-quark n(ddu) (1).

At the lowest energy level of the standard model, the first generation of particles are then the electron, electron neutrino, the up-quark and the down-quark; these are currently considered to be truly elementary, in that they cannot be subdivided (1).

Higher energy scales, such as those achieved at modern particle accelerators, result in the second and third generation of the four elementary particles; these are heavier versions of the first generation: for example, the muon () is essentially a version of an electron which is 200 × heavier than a low energy electron, i.e. . The tau-lepton () is the third generation of the electron, and is much heavier, i.e. . These mass differences do have physical consequences, but the fundamental properties and interactions of the various generations remain identical (1).

Current experimental evidence indicates that there are no further generations than these three, and so all matter in the universe seems to be circumscribed by the following twelve fundamental fermions, reproduced from (1):

Table 1: The twelve fundamental fermions.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Leptons | | | | Quarks | | |
|  | Particle | Q | Mass/GeV | Particle | Q | Mass/GeV |
| First Generation | Electron () | -1 | 0.005 | Down (d) | -1/3 | 0.003 |
| Neutrino () | 0 | < | Up (u) | +2/3 | 0.005 |
| Second Generation | Muon () | -1 | 0.106 | Strange (s) | -1/3 | 0.1 |
| Neutrino () | 0 | < | Charm (c) | +2/3 | 1.3 |
| Third Generation | Tau () | -1 | 1.78 | Bottom (b) | -1/3 | 4.5 |
| Neutrino () | 0 | < | Top (t) | +2/3 | 174 |

While it is accepted that neutrinos are not massless, their masses are so small that they have not been precisely determined, however, the upper bounds for the estimated masses for neutrinos are around 9 orders of magnitude smaller than the other fermions (1).

The Dirac equation describes the state of each of the twelve fundamental fermions and indicates that for each fermion there is an antiparticle which has the same mass but opposite charge, which is indicated by a horizontal bar over the particle’s symbol, or a charge symbol of the opposite sign, e.g. the anti-down quark is indicated by d̅, whereas the antimuon is indicated by (1).

Interactions between particles are facilitated by the four fundamental forces, but the effect of gravity at this scale is sufficiently negligible that it can be ignored without loss of accuracy. All particles take part in weak interactions and are therefore subject to the weak force. The neutrinos are all electrically neutral and therefore are not involved in electromagnetic interactions and are, so to speak, invisible to this force. Quarks carry what is termed as “colour charge” by QCD and are therefore the only particles that feel the strong force (1).

The strong force confines quarks to confined states within hadrons and are therefore not freely observed under normal circumstances (1).

### The Fundamental Forces

Classical electromagnetism explained the electrostatic interaction between particles using a scalar potential, Newton himself that matter could interact with any other matter without the mediation of direct contact (1).

Quantum Field Theory circumvents this non-material explanation and encompasses the description of each of the fundamental forces. Electromagnetism is explained by Quantum Electrodynamics (QED), the Strong Force by Quantum Chromodynamics (QCD), the weak force by the Electroweak Theory (EWT), Gravity has not been explained by the Standard Model yet; therefore, Einstein’s General Theory of Relativity is still the best explanation of this force, but it falls within the bounds of Classical Physics. As such, the search to incorporate gravity into the Standard Model is an ongoing area of research and has resulted in exciting new theoretical research avenues such as string theory and loop quantum gravity arising (1).

Looking at electromagnetism, the interaction between charged particles occurs via the exchange of massless virtual photons, which explains momentum transfer via a particle exchange and circumventing the issue of a non-physical potential as the medium of interaction (1).

Similarly, there are virtual particles (gauge bosons) for both the Strong Force (i.e. the massless gluon) and Weak Force (i.e. and bosons, which are around 80 times heavier than the proton and the Z boson, which facilitates a weak neutral-current interaction). The gauge bosons all have spin 1, compared to the fermions whom all have spin ½ (1).

### The Higgs Boson

The Higgs Boson, whose existence was confirmed by the CMS and ATLAS collaborations at CERN in 2012, but proposed in 1964 by three separate theoretical papers, breaks rank with the other particles outlined by the standard model in that it is many orders of magnitude heavier and is a scalar particle which endows other standard model particles with mass, a property without which all particles would constantly move at the speed of light, (1).

The Higgs boson manifests as a disturbance of the Higgs field, which is non-zero in a vacuum, in contrast to the other fundamental particles which all have a vacuum expectation value of zero, i.e. (1).

In QFT, an expectation value is a real number calculated as the average over the expected values of an observable, weighted according to their respective likelihood.

On their own, all particles are massless, but interacting with the Higgs Field, which is always non-zero, the Higgs mechanism gives them their distinguishing masses (1).

## The Quark Gluon Plasma

## Interactions of Particles with Matter

In order to study subatomic particle, they need to be detected. Most particles produced during High Energy Physics Experiments are unstable and therefore decay within a specific characteristic mean lifetime . Those particles with will traverse several meters before decaying and are therefore directly detectable by particle detectors installed at the Large Hadron Collider (LHC) at CERN. Particles with shorter lifespans are usually detected indirectly, by the interaction of their decay products with detector material (1).

### The Bethe-Bloch Curve

The Bethe-Bloch equation describes the energy lost by a charged particle moving at relativistic speed through a medium, as a result of electromagnetic interactions with atomic electrons. A single charged particle with velocity , passing through a medium with atomic number and density , will lose energy as a result of ionisation of the medium, as a function the distance travelled in the medium, according to the Bethe-Bloch formula (1):

Please see Appendix A for the code used to plot Figure 1 and Figure 2, illustrating the characteristic energy loss curves for the two subatomic particles studied in this project, the pion and the electron .



Figure 1: Bethe-Bloch curve for a pion moving at relativistic speeds through silicon medium



Figure 2: Bethe-Bloch curve for an electron moving at relativistic speeds through a silicon medium

### Transition Radiation

## The CERN Experiment

At the end of 1951, a resolution was agreed upon to establish a European Council for Nuclear Research (CERN: *Conseil Européen pour la Recherche Nucléaire)* at an intergovernmental UNESCO meeting in Paris. The final draft of the CERN commission was signed by twelve nations in 1953 (2).

Today, CERN is a truly international organization, with 22 member states, who contribute to operating costs and are involved in major decision making, many countries with observer status, and even more non-member countries with co-operation agreements, including South Africa (3).

CERN’s research mandate revolves around finding answers to fundamental questions about the structure and evolution of our universe, as well as its origins; it aims to achieve these goals by providing access to its particle accelerator facilities and compute resources to international researchers, who perform research that advances the forefront of human knowledge, for the benefit of humanity as a whole. As such, CERN is politically neutral and advocates for evidence-based reasoning, knowledge transfer from fundamental research to industry and grass-roots development of future generations of scientists and engineers (4).

### Hardware

In order to fulfil its ambitious goals, CERN’s facilities, located under the Franco-Swiss border (see Figure 3 for geographical context), boasts an intricate system of particle accelerators and -detectors and a data centre with over 174,000 processor cores, 150,000 Terabytes (TB) of Disk space and over 1,000 TB of random access memory (RAM) (5); this main datacentre is connected both to its extension in Budapest, Hungary and the multi-tier Worldwide LHC Computing Grid (WLCG), all of which operates at a data transfer rate of around 10 Gigabytes/second (GiB/s).

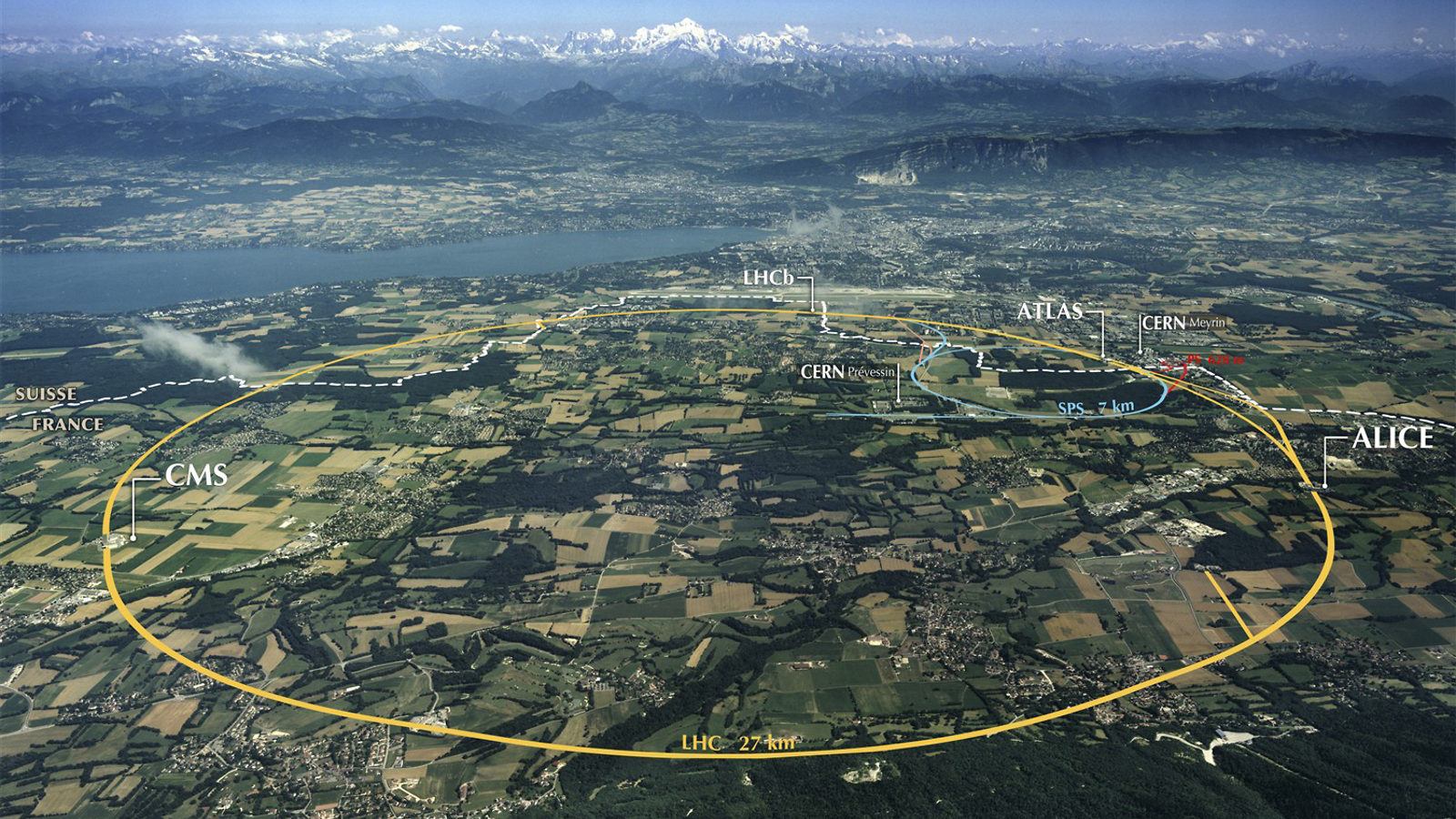


Figure : CERN facilities in geographical context

#### Accelerators

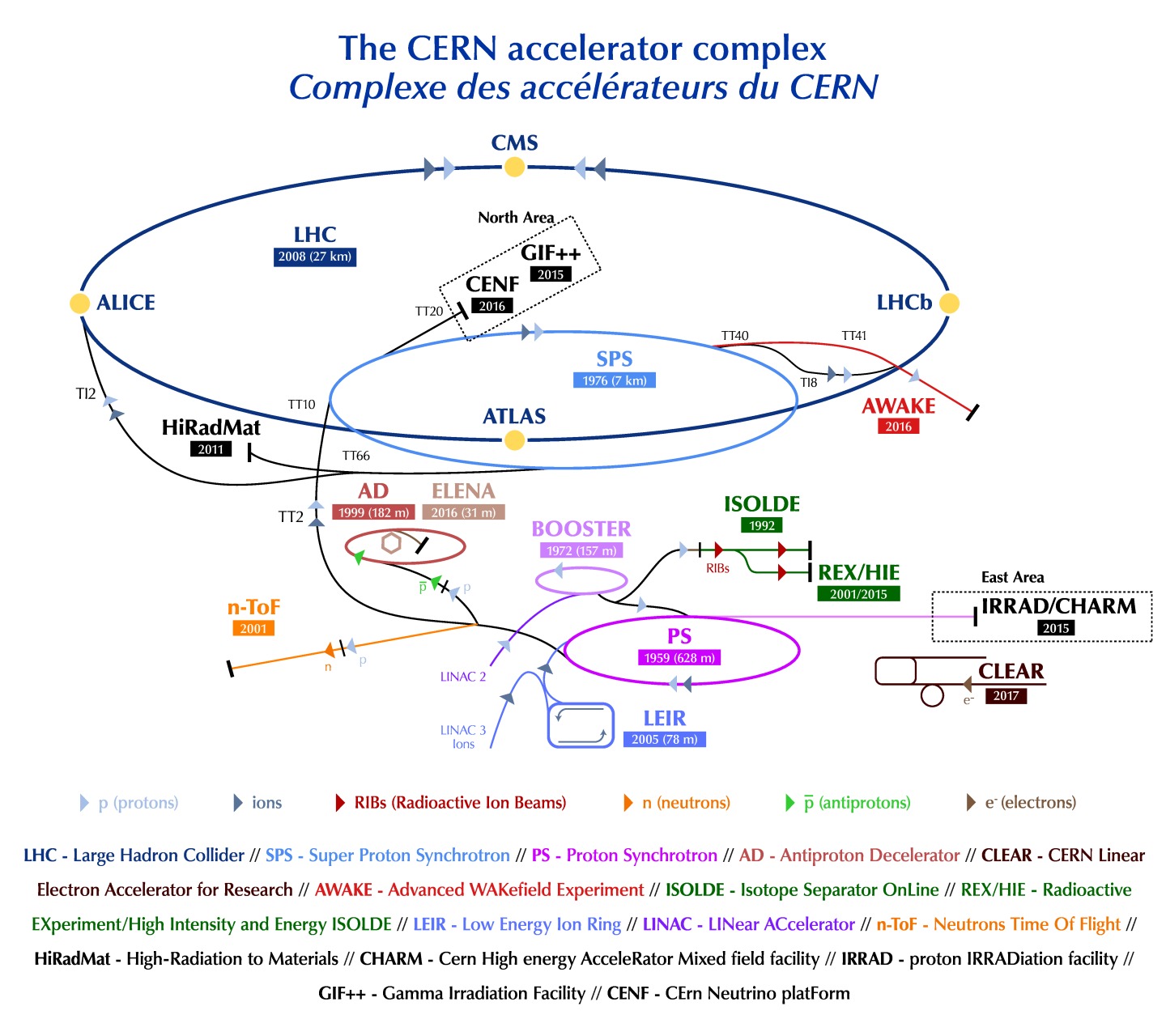


Figure : The CERN accelerator complex (5).

#### Detectors

#### The Worldwide Large Hadron Collider Computing Grid (WLCG)

### Software

### Collaborations

# The ALICE Collaboration

## Objectives of the ALICE Experiment

## Gas Detectors

## The ALICE Detector

### The Transition Radiation Detector

# Deep Learning

## Deep Learning within the Context of Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) is a branch of Computer Science concerned with getting computers to perform tasks that are characteristic of those performed by the human mind. The field of AI encompasses both hard-coded rule-based programs (known as the knowledge base approach to AI, which has largely remained ineffective), as well as Machine Learning, which is an approach to AI which aims to get computers to perform these tasks without explicitly coding the solutions for them (3).

The success of Machine Learning algorithms is largely determined by the representation of the data fed through them. Often, a large amount of an AI practitioner’s time is dedicated to engineering the right feature-set to hand to a simple machine learning algorithm (3).

In the case of machine learning for image classification, which loosely ties back to some of the aims in this project, it is not always immediately obvious as to which features will be informative to an ML algorithm. For example, feeding raw pixel values into a linear regression model should not be very effective, since images vary in terms of positional information, lighting, sharpness, rotation, etc. (3)

Representation learning is a solution to feature generation in which ML is applied, not only to map from a feature set to an output, but also towards automatically learning the most useful representation of the data; usually this representation will encompass identifying the major factors of variation which effectively explain the observed data and discarding those which are not useful to the algorithm (3).

Deep Learning is an approach to representation learning which constructs useful representations based on a combination of simpler representations. In fact, the basic unit of a neural network is the perceptron, which in itself is a very simple function, but once compiled into a Multi-layer Perceptron, the rich texture of the input data distribution can be very accurately captured, since useful features discovered in the first layers of such a neural network can be combined in various ways to create additional useful features (3). Continuing with the image classification example, an early layer of a convolutional neural network may detect edges in an image, the next layer may detect corners and shadows, and layers further down will ideally detect actual visual elements (faces, car lights, arms, etc.) (3).

## Mathematical Background for Deep Learning

### Rosenblatt’s Perceptron

The original Rosenblatt paper (4) outlining the concept of the “perceptron” aimed to develop a theory to explain: 1. How sensory information is detected by biological organisms, 2. how that information is subsequently processed and stored and 3. how mental comprehension or organismal behaviour (which he termed “*preference for a particular response*”) was driven by the first two processes.

He outlined a mathematical framework for these mechanisms, at the hand of the following constructs:

1. **S-points:** sensory units which can possess any of a number of response curves based on the signal strength of incoming information

2. **A-units:** association cells located in an “association area” , which in some of his models was preceded by a “projection area”

3. S-points are connected in specific ways to A-units and forward their stimulus response to them, in the form of an inhibitory or an excitatory impulse

4. : A threshold value assigned to each A-unit dictates whether it will fire, based on the algebraic sum of excitatory and inhibitory signals received, from either S-points or preceding A-units

5. The connections between S-points and A-units, and between A-units themselves is random, and not all elements of such a network are connected to each other

6. Response units, , receive a large number of inputs from the set, called its source-set, and have feedback mechanisms to A-units in its source set. (4)

He put forth various models for response curve summation and how these networks would learn (4), but while the mathematical constructs he proposed were oversimplifications of the complexity of biological brains, they were found to be extremely useful in training computers to emulate their capabilities.

### Deep Feedforward Neural Networks

At its most basic level, an artificial neural network (ANN) is an approximation of a mapping function , which maps from a set of input features to a response, . Feedforward neural networks have one-way information flow from input features to output, whereas recurrent neural networks have feedback connections (3).

Also called multilayer perceptrons (MLPs), deep feedforward networks are composed of an arbitrary number of nested approximating mapping functions, of the form:

The superscript of these functions, , indicates the layer index of the function in an ANN, with indicating the depth of such a neural network. It is this concept of chained functions of arbitrary depth from which the term Deep Learning is derived (4).

The process of training such a network, , to give the closest approximation to the desired output, , is an iterative process, involving passing many observations, each having the same feature set through the MLP, assessing the output, , according to an error metric, , and individually adjusting each of the mapping functions according to their contribution to the differential of the magnitude of error at the conclusion of each training step . In other words, a parameter set , pertaining to each is iteratively adjusted according to . (3).

The set of nested approximation functions outlined above are commonly referred to as hidden layers, the dimensionality of the outputs of each layer is known as its width, or as the number of neurons in that particular hidden layer (3).

In order to produce subtle derived features from the input feature set, nonlinear transformations are applied to the output of each layer in the network, which in itself is a simple linear function of the form , where is a vector of weights of the same length as the set of input features, which are essentially a set of coefficients for each in the chain of functions, and is a real-valued bias term, which is essentially an intercept term for each (3).

It is easy to see that chaining such a set of linear models without applying nonlinear transformations (denoted as ) to what are essentially an arbitrary number of linear regression functions (), one would simply arrive at another linear model (3).

A commonly used nonlinear transformation , or activation function, in modern deep learning algorithms is the rectified linear unit (the ReLU function), which is simply an affine transformation, reminiscent of the response curves envisioned in Rosenblatt’s paper, of the form (3).

Combining the concepts explained above, then gives us a representation for a single hidden layer in an ANN as follows:

And, by extension, for a neural network with three hidden layers:

We now have a vector of weights multiplied by a vector of input features, which can be the original features fed to , or the weighted outputs of previous hidden units in . Since we essentially have a vector of hidden units, we also have a vector of bias terms, and all of these hyperparameters, collectively referred to as , need to be optimized to arrive at a reasonable approximation of a theoretically optimal mapping function (3).

To achieve the optimization of , most deep learning models utilize the concept of maximum likelihood, to minimize a loss function , for example, binary cross entropy:

,

where is the model’s estimate for the probability of an observation of being of a particular class (3).

Please see Appendix B for the code used to generate Figure 3, which shows how, as approaches the true (in this binary classification example, ), the binary cross entropy loss function approaches 0.



Figure 5: Illustration of the descent towards zero, of the Binary Cross Entropy Loss Function as ŷ, or , approaches the true y.

The chain rule of calculus is employed by backpropagation to enable the derivative of the loss function to be redistributed through the network, based on the partial derivative of each hyperparameter with respect to the derivative of the loss function (3):

For k = hidden layers, , we compute the element-wise gradient on the layer’s output (before the non-linear activation function is applied):

And the gradients on the weights and the bias term:

Here, represents the weight decay penalty, where the size of the weights are constrained, in a manner inversely proportional to . A regularizer is added to the loss, where contains all the weight and bias parameters.

This gradient is then propagated to the activations of the preceding layer:

## Convolutional Neural Networks

## Variational Autoencoders

## Generative Adversarial Networks

# Data

# Methods

## ROOT

## Data Extraction from WLCG

# Results

# Discussion

# Conclusion

# Bibliography

1. *Modern Particle Physics.* **Thomson, Mark.** Cambridge, UK : Cambridge University Press, 2013. ISBN 978-1-107-03426-6.

2. **CERN.** About CERN: Who we are: Our History. *CERN.* [Online] CERN. [Cited: 26 January 2019.] https://home.cern/about/who-we-are/our-history.

3. **Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron.** *Deep Learning.* Cambridge, Masachusetts : The MIT Press, 2016. ISBN 9780262035613.

4. *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.* **Rosenblatt, F.** 6, 1958, Psychological Review, Vol. 65.

# Appendices

[Appendix A: Plotting the Bethe-Bloch Equation 49](#_Toc536283953)

[Appendix B: Plotting Binary Cross-Entropy 56](#_Toc536283954)

Appendix A: Plotting the Bethe-Bloch Equation

Create a Bethe-Bloch function:

#Planck's constant:  
h <- 6.62607004e-34  
  
#Speed of light m/s  
c <- 299792458  
  
#Fine structure constant  
alpha <- 1/137  
  
#Mass of an electron Mass/GeV  
  
m.e <- 0.005  
  
#Density n, atomic number Z, the fraction of the speed of light the particle is moving at, beta, and the particle's velocity v are specified as parameters to the equation  
  
  
dE.dx <- function(n,Z,v,beta){  
 -4 \* pi \* h^2 \* c^2 \* alpha^2 \* ((n \* Z)/(m.e \* v^2)) \* log(((2 \* beta^2 \* gamma^2 \* c^2 \* m.e)/(I.e)) - beta^2,base=exp(1))  
}  
  
#For an electron traversing a silicon detector:  
  
v <- seq(0.1\*c,c,100000)  
  
beta <- v/c  
  
#Lorentz factor  
  
gamma <- 1/(sqrt(1-(v^2/c^2)))  
  
n <- 1  
  
  
  
Z <- 14  
  
#Effective ionization potential of the material  
  
I.e <- 10 \* Z  
  
electron.y = dE.dx(n=n,Z=Z,v=v,beta=beta)  
  
require(latex2exp)

## Loading required package: latex2exp

m.e <- 273.13\*m.e  
  
pion.y = dE.dx(n=n,Z=Z,v=v,beta=beta)  
  
  
  
plot(x=beta\*gamma, y=-pion.y,type="l",main="Bethe-Bloch Curve of a Pion moving through Silicon", xlab = TeX("$\\beta\\cdot\\gamma$"),ylab=TeX("$-dE/dx$"),col="blue",cex.main=0.8)



plot(x=beta\*gamma, y=-electron.y,type="l",main="Bethe-Bloch Curve of an Electron moving through Silicon", xlab = TeX("$\\beta\\cdot\\gamma$"),ylab=TeX("$-dE/dx$"),col="red",cex.main=0.8)



v <- seq(0.8\*c,c,100000)  
  
beta <- v/c  
  
#Lorentz factor  
  
gamma <- 1/(sqrt(1-(v^2/c^2)))  
  
n <- 1  
  
m.e <- 0.005  
  
electron.y = dE.dx(n=n,Z=Z,v=v,beta=beta)  
  
m.e <- 273.13\*m.e  
  
pion.y = dE.dx(n=n,Z=Z,v=v,beta=beta)  
  
plot(x=beta\*gamma, y=-pion.y,type="l",main="Bethe-Bloch Curve of a Pion moving through Silicon \nat Speeds Upwards of 80% of the Speed of Light", xlab = TeX("$\\beta\\cdot\\gamma$"),ylab=TeX("$-dE/dx$"),col="blue",cex.main=0.8)



plot(x=beta\*gamma, y=-electron.y,type="l",main="Bethe-Bloch Curve of an Electron moving through Silicon \nat Speeds Upwards of 80% of the Speed of Light", xlab = TeX("$\\beta\\cdot\\gamma$"),ylab=TeX("$-dE/dx$"),col="red",cex.main=0.8)



v <- seq(0.9\*c,c,100000)  
  
beta <- v/c  
  
#Lorentz factor  
  
gamma <- 1/(sqrt(1-(v^2/c^2)))  
  
n <- 1  
  
m.e <- 0.005  
  
electron.y = dE.dx(n=n,Z=Z,v=v,beta=beta)  
  
m.e <- 273.13\*m.e  
  
pion.y = dE.dx(n=n,Z=Z,v=v,beta=beta)  
  
plot(x=beta\*gamma, y=-pion.y,type="l",main="Bethe-Bloch Curve of a Pion moving through Silicon \nat Speeds Upwards of 90% of the Speed of Light", xlab = TeX("$\\beta\\cdot\\gamma$"),ylab=TeX("$-dE/dx$"),col="blue",cex.main=0.8)



plot(x=beta\*gamma, y=-electron.y,type="l",main="Bethe-Bloch Curve of an Electron moving through Silicon \nat Speeds Upwards of 90% of the Speed of Light", xlab = TeX("$\\beta\\cdot\\gamma$"),ylab=TeX("$-dE/dx$"),col="red",cex.main=0.8)



Appendix B: Plotting Binary Cross-Entropy

Define a function to plot the binary cross-entropy loss function:

cross.entropy <- function(y,p){  
 -(y \* log(p,base = 10) + ((1-y)\*(1 - log(p,base=10))))  
}  
  
#if the predicted class is 1:  
  
y <- 1  
  
p <- seq(0,1,0.01)  
  
loss <- cross.entropy(y,p)  
  
require(latex2exp)

## Loading required package: latex2exp

plot(x=p,y=loss, type="b", col=rainbow(250),cex=0.5, main = TeX("J($\\theta$) = -(y log(p)-(1-log(p)))"),ylab = "Cross Entropy", xlab = TeX("$\\hat{y}$"))

