5-15-2022

Partial Automation of Data Analysis in a Z Prime Boson Search

Ethan Colbert
ecolbert@bellarmine.edu

Partial Automation of Data Analysis in a Z Prime Boson Search

Ethan Colbert

Advisor: Dr. Akhtar Mahmood

Reader: Dr. Robert Kelley

May 15, 2022

Contents

One of the most fundamental questions a physicist can ask is, “what is the world made of?” There are many ways in which this question can be cast. One of those is the more precise question, “if we break a piece of matter down as much as we possibly can, what are we left with?” We have been asking this question for a few millennia (many place the beginning of this investigation at Democritus positing the idea of the atom). Over time, answers to this question have been refined, resulting in our current understanding of the structure of matter. It is an understanding that is beautifully simple at its core yet remains extremely elusive in the details. To start with, we need to describe two things: what particles make up matter and how these particles interact with the forces.

Background: Particles and Forces

Elementary Particles

If a physicist is asked to describe matter, they will almost certainly talk about particles and forces. These are the two essential ingredients that describe the universe. The kinds of particles physicists deal with depends largely on the scale of the objects in question. At an everyday scale, one might consider a baseball launched off a bat to be a particle. In chemistry and biology, particles usually refer to molecules or atoms. In physics, scientists “zoom in” even further, less than the femtometer scale (10^{-15} m), and the word “particle” refers to tiny objects that represent the fundamental building blocks of matter.

When we study physics at scales that small, we refer to them as simply *elementary particles.* These particles, together with the force they interact with, are collectively known as the Standard Model. The particles all fit on a compact chart as shown below.

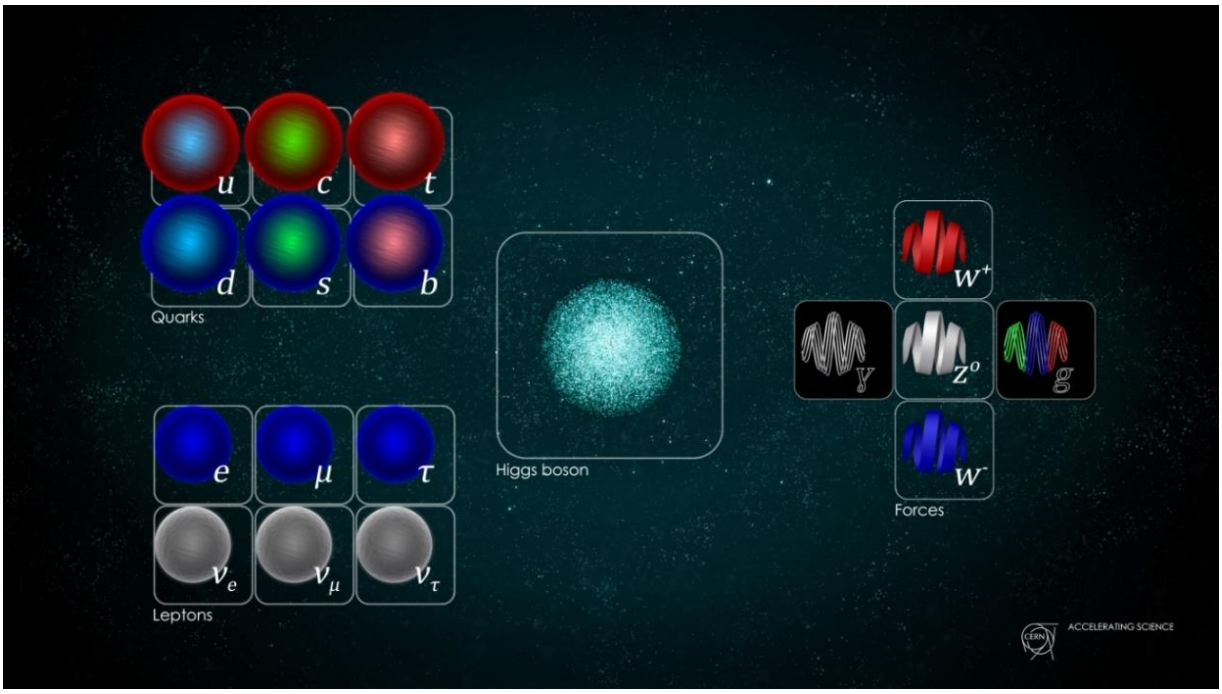


Figure 1: a summary of the elementary particles of the Standard Model. Quarks show up in the upper left, leptons in the lower left, and the various bosons that carriers the forces on the right. The Higgs boson that provides mass, is shown at the center. Image from [1].

Figure 1 has three distinct clusters. The two on the left are occupied by matter particles.

In the lower left, a cluster of six particles are known as *leptons*. Of these, one is familiar: the electron, one of the particles that make up the atom orbiting the nucleus. Electron is used to define the most common unit of charge as -1. This unit is given the abbreviation *e* or e^-, (for 'elementary charge') and the charges of the other particles are expressed in terms of this charge. The electron has a small mass: about 0.511 MeV[1] [2]. In fact, this is the smallest mass of any matter particle, with the exception of neutrinos.

[1] This unit is expressed as MeV/c^2. The electron-volt is a common unit of energy equal to roughly 1.602×10^{-19} Joules (MeV is "mega-electron-volts"). It is convenient in particle physics to express the masses of particles in units of energy, according to Einstein's famous $E_0 = mc^2$. It is also common in notation to drop the division by c^2 when referring to mass.

Next in the lepton cluster is the muon, abbreviated as μ or μ^-. Muons are produced in the upper atmosphere when cosmic rays collide with atomic nuclei [3]. They also have an electric charge of -*e* and carry a mass of roughly 105.66 MeV [2], placing them just over 200 times more massive than the electrons. The final particle in the top row of the lepton cluster is the tau particle, abbreviated as τ or τ^-. It also carries an electric charge of -*e* and has a mass of about 1,776.86 MeV. This makes it about 16.8 times more massive than the muon, extremely heavy for an elementary particle. Its large mass means that it can only be produced in extremely high-energy interactions and will decay very quickly: its average lifetime is only about 0.29×10^{-12} seconds [2].

The bottom row of the lepton cluster is occupied by the neutrinos. The three types of neutrinos correspond to the three leptons. Their abbreviated as ν_e, ν_μ, and ν_τ. All neutrinos are all electrically neutral and have such small masses that they were originally thought to be entirely massless. These neutrinos are exceptionally difficult to detect; the only way that we can account for them in particle accelerator experiments is by finding the "missing" energy – energy we know was present in a collision event but that wasn't captured by the detector instrument. However, neutrinos are by no means rare in nature. The sun produces billions of neutrinos each second [4].

The upper left cluster in the table is occupied by the six types, or "flavors," of quarks. The quarks combine to make up most of the "matter" around us, such as, protons and neutrons. They are different from leptons since quarks are never seen in isolation, but only in combination with other quarks, forming larger particles [5, p. 42] that are commonly referred as *hadrons*. Hadrons are of two distinct types hadrons, one type is called *baryons* (particles made up on three quarks), and the other type is called *mesons* (particles made up of a quark and an antiquark). There are six types of quarks, namely up (u), down (d), strange (s), charm (c), bottom (b), and top (t). The quarks have fractional charges (either +2/3e or -1/3e). Their masses and electric charges are shown in Table 1: The very large mass of the Top quark is still a mystery.

Quark Flavor	Up	Down	Strange	Charm	Bottom	Top
Mass (MeV/c^2)	2.16	4.67	93	1,027	4,018	≈170,000
Electric Charge (*e*)	+2/3	-1/3	-1/3	+2/3	-1/3	+2/3

Table 1: Masses and electric charges of the quarks. Data from [6].

Like the lepton cluster, from left to right across the quark cluster in Figure 1, the mass of the particles increases. Each column in the table represents a *generation* of matter particles. There are three generations of particles; the first generation consisting of the (i) up quark, down quark, electron, and electron neutrino, the second generation consisting of the (ii) strange quark, charm quark, muon, and muon neutrino, and the third generation consisting of (iii) the bottom quark, top quark, tau particle, and tau neutrino. Each successive generation is more massive, and correspondingly rarer, than the previous generation. All the matter particles (quarks and leptons) are categorized as *fermions* (these particles have fractional spins)[2]. The remaining particles in Figure 1 are classified as *bosons* (these particles have integer spins).

Protons and neutrons are each composed of three quarks, and only of up and down quarks. A proton is composed of two up quarks and a down quark bound together (note that the electric charges of all three are added up to positive or +*e*). A neutron is composed of two down quarks and a single up quark (once again, the electric charges sum to the total neutron charge which is zero or neutral). All of the elementary particles that make up an atom are members of the first generation of particles.

[2] By "familiar properties," I mean mass and electric charge. The fermionic or bosonic nature of any given particle does, of course, relate to its spin. Fermions have half-integer spin, while bosons have integer spin. Fermions also must obey the Pauli Exclusion Principle, while bosons are immune to it.

In addition to the particles shown in Figure 1, the Standard Model also incorporates tier corresponding *antiparticles*. The "positron" is the antiparticle of the electron has a positive charged[3]. All matter particles have corresponding antiparticles have precisely the same mass but opposite electric charge to their corresponding matter particles. Most antiparticles are denotated by placing a bar above the abbreviation of their matter particle. The exceptions to this are electrons, muons, and tau particles; for example, e^- for an electron, e^+ for a positron. Antiparticles extends to larger particles of matter, such as protons and neutrons, as well. These are also denotated with an overbar.

In the center of Figure 1 is the Higgs boson, that was discovery in 2012 [7]. Its prominence was largely because it was the final particle included in the Standard Model that had yet to be directly detected in experiments. The Higgs boson plays a unique role in the Standard Model theory in that it is related to our current understanding of how particles obtain their mass [8]. Finally, on the right side of figure 1, there are a cluster of five particles. These are fundamentally different from the others in that they do not make up matter. These five particles are also bosons, and are known as *force carrier particles*.

Interactions Among the Particles: Forces

In addition to describing the particles that make up the ordinary mater in the universe, any successful theory must describe how those particles interact with one another. The theory behind the Standard Model (quantum field theories) are based in the four "fundamental forces" to accomplish this. These four forces are – the Gravitational force (carrier particle is called the Graviton), the Electromagnetic force (carrier particle is called the Photon), the Weak force (carrier particles are called the W^+, W^-, and Z^0 bosons), and the Strong force (carrier particle is called the Gluon). The strengths of these four forces are different and depends on the energies and scale at which their interactions occur. Standard model is based on the three out of these four forces, since the strength of the Gravitational forces is extremely weak compared to the other three forces at the subatomic scale. The three forces that fully

[3] An electron has an electron number of +1, while a positron has electron number of -1.

incorporated into quantum field theories are the electromagnetic force, the strong and weak nuclear forces. The gravitational force has yet been incorporated into quantum field theories since its strength is 10^{40} times weaker than the electromagnetic force. So, the carrier of the Gravitational force, the Graviton is not included in Figure 1 [9]. The development of a fully consistent theory of "quantum gravity" is an unsolved problem in theoretical physics. The force carrier particles serve as "messengers" for the forces, and particles interacting via one of the forces exchange bosons between them.

The electromagnetic force is the force is responsible for nearly every physical interaction we experience in daily life. The "contact forces" that we experience when we touch things are caused, at the microscopic level, by electromagnetic repulsion between electrons in the atoms of the two interacting bodies. The electromagnetic force carrier particle the photon, abbreviated as γ was discovered by Einstein, while explaining the photoelectric effect, and is was one of several landmark events marking the beginning of quantum physics. The photon carries no mass at all. It is also electrically neutral, even though it mediates interactions between electrically charged particles.

The weak and the strong forces only operates inside the nucleus and are often refereed as nuclear forces. The strong nuclear force is responsible for the combining of quarks that forms larger particles (hadrons) such as protons and neutrons. In addition to this, the strong force causes these protons and neutrons to stick together to form the nuclei of atoms, preventing electromagnetic repulsion from pushing the protons away from each other (the strong force is many times stronger than the electromagnetic force, at the nuclear scale). This force is mediated by the gluon, abbreviated as G. The gluon, like the photon, is also massless and electrically neutral [10]. The weak nuclear force is responsible for the nuclear beta decay. It is mediated by the W^+, W^-, and Z bosons as stated earlier. Unlike photons and gluons, the carrier particles for the weak force have mass. The $W^{\pm}$ bosons carry a mass of roughly 80.38 GeV and an electric charge of $\pm e$. The Z boson is electrically neutral and has a mass of roughly 91.2 GeV [10].

The Z' (Z Prime) boson does not appear in the Standard Model. Rather, its existence is predicted by theories that extend the Standard Model that are now part of the Beyond the Standard Model (BSM)

theories. BSM predict that, if interactions occur at extremely high energies, the weak force would sometimes be mediated by a very massive, electrically neutral particle. This new particle that behaves like a Z boson but has a much higher mass, is referred to as a Z' boson. Because the Z' boson particle has never been directly detected in an experiment, we don't know precisely how massive it is. Theorists have been able to make predictions based on the specific model that they use, but such predictions most often take the form of a lower limit of the mass. This lower limit has been refined by experimental searches for the particle. Theoretical predictions of a lower mass limit are already extremely high, upwards of 955 GeV [11]. Refined figures from experimentalists are far higher than this, placing $M_{Z'} \geq 2.57$ TeV [12] (that's 2,570 GeV).

As I mentioned earlier, the four forces are characterized by their respective "strengths." The actual strength of any force on a particle depends on its properties, those of the particle it interacts with, and the distance between the two. Nonetheless, it is possible to account for these factors by considering the interaction of two given particles a given (very short) distance apart, via each of the four forces. The strengths of the four forces at the atomic scale is shown in Table 2:

Force	**Relative Strength at Atomic Scale**
Strong Force	1
Electromagnetic Force	10^{-2}
Weak Force	10^{-13}
Gravitational Force	10^{-42}

Table 2: relative strengths of the four fundamental forces at very small distances. Data from [5, p. 59].

We find that the strengths of the forces vary across many orders of magnitude. As shown on Table 2, surprisingly, the Gravitational force is far weaker than any of the other forces. This is largely the reason why quantum field theories have yet to incorporate the Gravitational force – its effects at subatomic scales are negligible. However, one might ask why we don't experience the nuclear forces in everyday life, while the effects of the Gravitational and Electromagnetic forces seem apparent. The answer is that these strengths are only valid at very small distances, at scale of 10^{-15} meters (femto-meters) or less. At distances greater than femto-meters, the strengths of the strong and weak nuclear forces diminish exponentially and become insignificant. Physicists often refer to the two nuclear forces as "short range forces". The "range" of a force factors into the dependence of its actual strength on distance. Beyond the range, this dependence tends to be of the form:

$$F(r) \propto e^{-\frac{r}{a}} \frac{1}{r^2} \quad (1)$$

Where *a* is the range of the force [5, p. 19]. The strong and weak forces both have a finite (and indeed very small) range. The ranges of the electromagnetic and gravitational forces, however, have infinite range, and therefore we experience their effects at large scales. Taking the limit of the above proportionality as $a \to \infty$ yields the expression familiar from Newton's law of Gravitation and Coulomb's law in classical electromagnetism:

$$F(r) \propto \frac{1}{r^2} \quad (2)$$

Particle Decay

Some of the particles that have been mentioned thus far are stable, meaning they'll continue to exist indefinitely. The most prominent examples of stable particles are electrons and protons (made of up and down quarks). Other particles, however, are less stable, and if they exist, will eventually *decay* into

other particles. Particles decay only via specific processes known as *decay modes* that obey a number of conservation laws depending on the forces that mediate the decay. Additionally, the decay modes, and their probabilities with which they occur, are governed by quantum field theories. One important rule is that all particle decay modes follow the conservation of energy and the relationship between energy and mass, that is, $E_0 = mc^2$. This law places a limit on the total mass of the "daughter" particles produced in a decay and the mass of the "parent" particle. In particular, the total energy of the daughter particles must be equal to the total energy of the parent particle:

$$K_p + m_p c^2 = \sum_d (K_d + m_d c^2) \quad (3)$$

Where K_p and K_d are the kinetic energies of the parent particle and each daughter particle, respectively, and m_p and m_d are their respective masses. So more massive particles tend to decay into less massive ones, because their energy is being divided among multiple daughter particles with each decay. This helps explain why virtually all matter we encounter in daily life is made up of first-generation particles: they are the lightest, and therefore the most stable; second and third generation particles decay into them.

Another rule strictly followed by any decay process is conservation of electric charge. If a parent particle has a charge of $+e$, for instance, then the sum of the charges of all the daughter particles from its decay must also be $+e$. This does not mean that no negatively charged particles can be produced by this particle's decay, but rather that the number of positively charged daughter particles must always exceed the number of negatively charged ones. I will give one example: the decay of the W^+ boson into three "pions." The process can be written:

$$W^+ \rightarrow \pi^+ \pi^+ \pi^- \quad (4)$$

This decay occurs in about 0.000101% of the time a W^+ boson decays (has a branching fraction of 1.01×10^{-6}) [10]. There is a negatively charged daughter particle formed in this process, however, the total electric charge is the same on both sides.

Of particular interest here are the decay modes of the Z boson. The Z boson is heavier than any matter particle except for the top quark, and so we would expect it to have many decay modes. In other words, any Z boson decay would involve enough energy to produce a large variety of potential daughter particles. This is indeed the case [10]. The two most common decay modes are included in this search. They are:

$$Z \rightarrow e^- e^+$$

$$Z \rightarrow \mu^- \mu^+ \qquad (5)$$

Each of these decays involves a Z boson decaying into a lepton and its corresponding antilepton. These decay modes are known as *dilepton channels*.

Because more massive particles are rare in nature, we can only create them in collisions using a high energy particle accelerator and study the decay signatures inside a particle accelerator.

Particle Accelerators

To induce interactions between particles that involve very large amounts of energy, we must first get the particles to move at extremely high speeds, very close to the speed of light. This is accomplished using the electromagnetic force[4]. The electromagnetic force is exerted on any electrically charged particle, which simply pushes a positive charge in the direction *of* the electric field, and a negative charge in the direction *opposite* to the electric field. The accelerator creates an electric field around an evacuated

[4] The classical electric force is, of course, a large-scale approximation of the electromagnetic force described by the quantum field theory. However, it is an accurate enough approximation that its use in conceptualizing the acceleration of particles in an accelerator is appropriate.

(vacuum) *beam pipe* and directed along it. This will cause a group of particles in the beam pipe (known as a *bunch*) to accelerate in the proper direction. Particles are accelerated to extremely high speeds at accelerators like the Large Hadron Collider (LHC) at CERN (Conseil Européen pour la Recherche Nucléaire in French or the European Organization for Nuclear Research in English); CERN is now known as the European Organization for Particle Research. The most recent collisions at the LHC involved protons travelling at roughly 99.99999896% of the speed of light[5] [13]. Under the direction of Dr. Akhtar Mahmood, Bellarmine University has been involved with the LHC experiment and have processed data for the LHC using the 384-core (with 300TB of disk storage space) Tier2 Supercomputer (located in Pasteur Hall) that is linked to the Open Science Grid (OSG) cyberinfrastructure. Over 100 Petabytes of data is produced by the LHC experiments per year.

The bunch of particles must also be steered. This is accomplished via the electromagnetic force. The electromagnetic force is essentially the combination of both the magnetic force and the electric force. The magnetic force differs from the electric force in that the force experienced by a particle is perpendicular to both the magnetic field and the direction in which the particle moves. This causes a particle moving in a direction perpendicular to a magnetic field to be pushed around in a circle. Accelerators use magnetic fields directed vertically to cause particles to move in a horizontal circle. But, there is a problem, however. The magnitude of the magnetic force on a particle is not only dependent on its charge, but also on its speed; the two are in fact proportional. The (classical) equation for the magnetic force is $\overrightarrow{F_B} = q\vec{v} \times \vec{B}$, where $\overrightarrow{F_B}$ is the magnetic force (the arrows merely denote vector quantities), q is the particle's charge, $\vec{v}$ is the particle's velocity, $\vec{B}$ is the magnetic field, and '×' denotes the vector cross product. If velocity and magnetic field are perpendicular, the magnitude of the force reduces to $F_B = qvB$. The force required to keep an object moving in a circle (known as centripetal force) is proportional

[5] This figure was calculated using $E = (\gamma - 1)m_p c^2$ (where γ has its usual relativistic meaning) with E = 6.5 TeV, the reported proton energy.

not to speed, but rather to speed *squared*, the equation being $F_c = m\frac{v^2}{r}$ where F_c is the centripetal force and m is the particle's mass. Putting the two equations together, we can derive an expression for the radius of the circle in which a charged particle moving perpendicular to a magnetic field travels:

$$r = \frac{mv}{qB} \tag{6}$$

So, if the magnetic field was kept constant as the particle sped up, it would travel in ever larger circles. Early particle accelerators, known as cyclotrons, started the particle near the center of the accelerator and allowing it to accelerate until it reached the outer edge [14]. However, as physicists desired accelerators to reach ever higher energies and ever higher speeds, the size of particle accelerators needed to increase. As a result, the synchrotron was developed, which keeps the radius of rotation constant. It achieves this by increasing the magnitude of the magnetic field as the particle bunches speed up [15]. The size of synchrotron accelerators can be very large: the circumference of the LHC is roughly 26.7 km [13] (about 16.6 miles).

There are accelerators that are called fixed target accelerators that only accelerate one set of bunches (a collection of bunches forms a *beam*) and eventually collide them with a stationary target, for which a constant electric field works well. However, at the LHC, two beams of proton are accelerated at very close to the speed of light, in opposite directions. This is possible by applying an alternating voltage to the devices providing the electric fields, effectively "switching" them at the frequency of the voltage. The two beams travel in separate beam pipes until they are collided inside a particle detector.

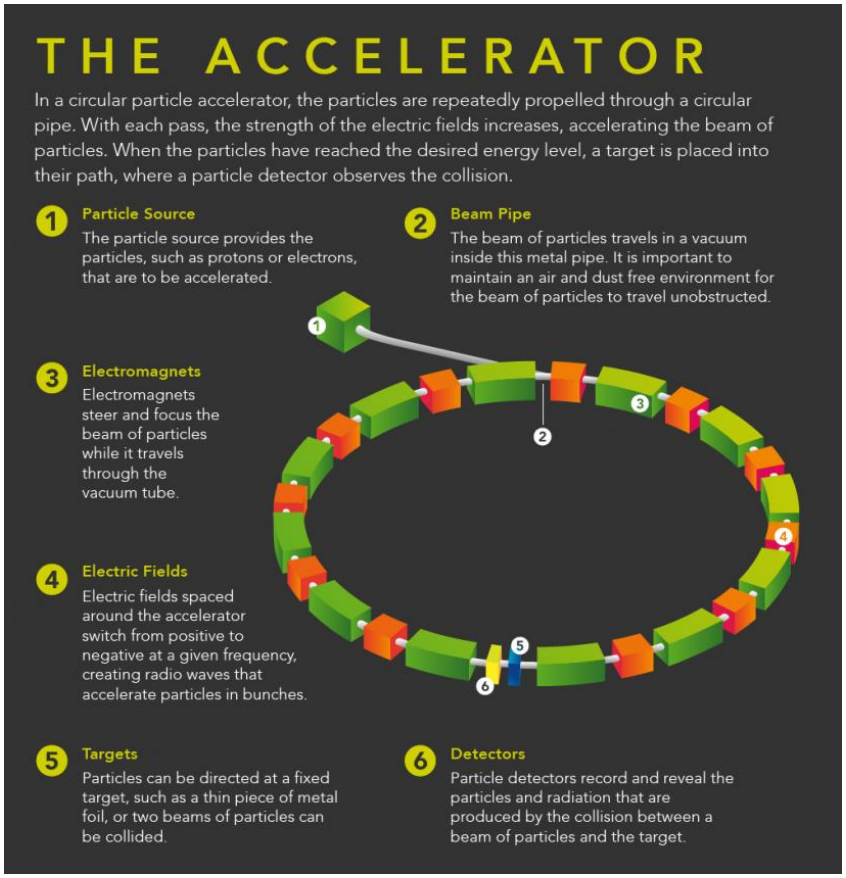


Figure 2: a diagram of a particle accelerator, such as a synchrotron. Note that, for the LHC, the targets (5) do not exist. Rather, two beams of particles travelling in opposite directions collide inside a large detector (6). Image from the U.S. Department of Energy [16].

The LHC is the largest particle accelerator ever built to date. A section of the accelerator tunnel is located along the border between Switzerland and France, and the remaining portion of the LHC tunnel is in France, as shown in Figure 3. LHC tunnel is roughly 100 meters below the ground [13]. LHC has four particle detectors, called ATLAS, CMS, LHCb, and ALICE. The center-of-mass energy of the LHC is currently 14 TeV (7 TeV per beam). At this high energy, the LHC is able to recreate the exact conditions of the very early universe inside the particle detectors when the universe was only a trillionth of second old after the Big Bang. On average, at the LHC, about a billion proton-proton collisions take place per second. Each collision is like a mini Big Bang.

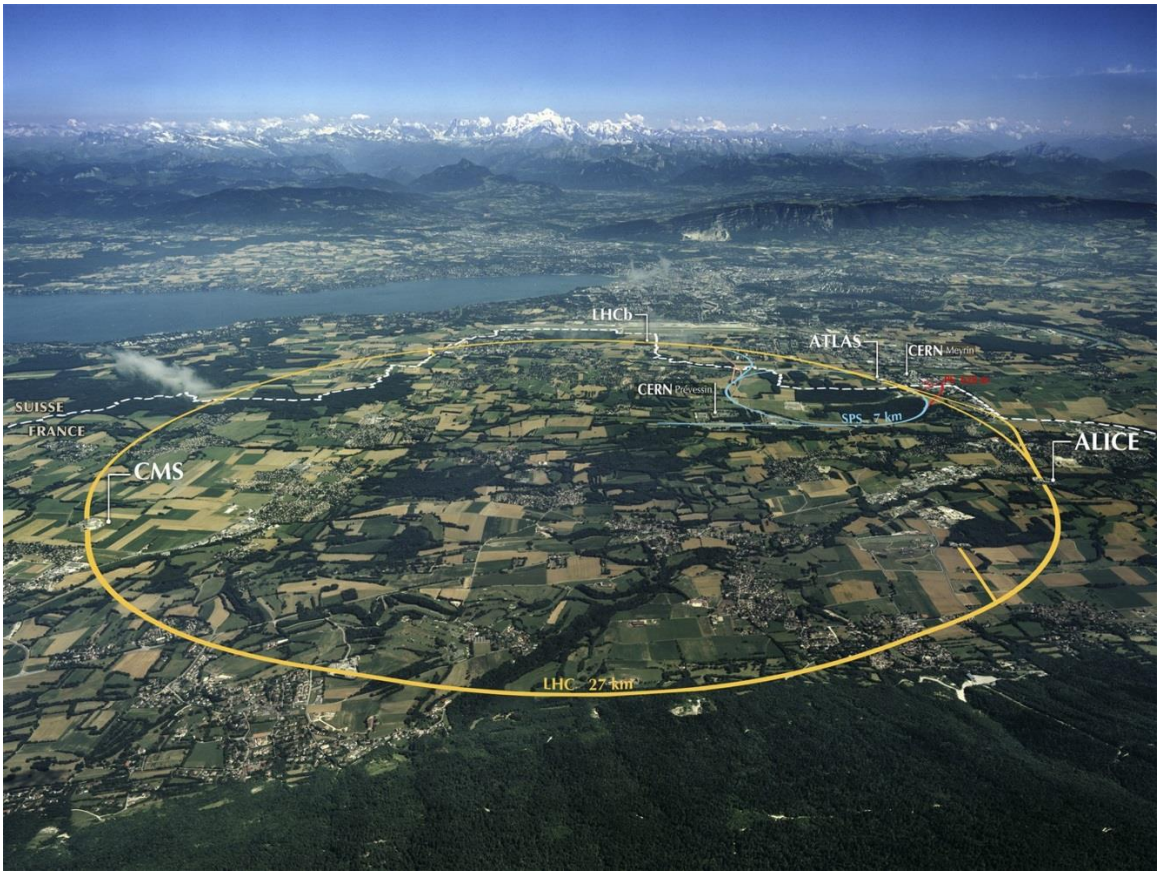


Figure 3: an aerial view of the CERN complex, with the 27 km LHC accelerator highlighted in yellow. The locations of the detectors are labeled. Image from [17].

At the LHC, once bunches of protons are accelerated to speeds very close to the speed of light cross with each other inside the particle detectors. Protons in the bunches collide, interacting and producing huge quantities of new particles inside the detector, which then decay and interact with each other. It is the physicists' task to decipher and precisely identify all the particles that are produced in these collisions by studying and analyzing their decay modes.

Particle Detectors

Particle detectors are themselves very large apparatuses, comprised of many complex components as shown in figure 4:

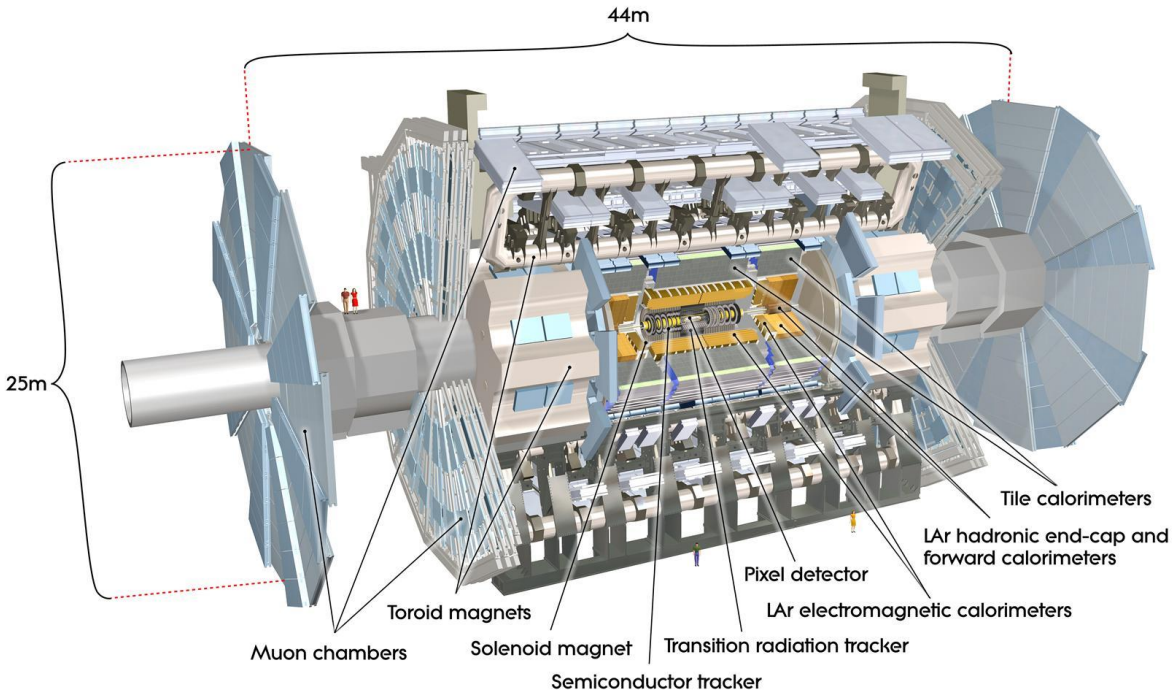


Figure 4: a schematic of the ATLAS detector showing the various detector components. It is about five-stories high. Image from [18].

When particles collide, they do so inside a detector. Detectors are structured such that they can capture the particles produced in a collision and record the results, precisely. A particle detector at the LHC is structured in layers (like an onion), centered around the beam pipe:

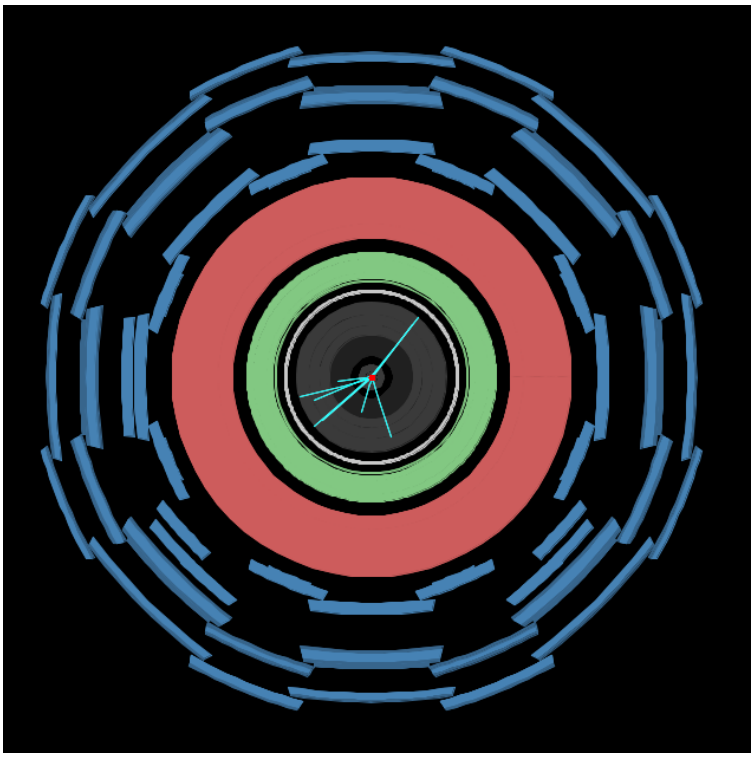


Figure 5: a visualization of the ATLAS detector at the LHC, showing a frontal view along the beam pipe. The tracker is shown in grey near the center, the electromagnetic and hadronic calorimeters in green and red, and the muon detectors in blue. Image made with Hypatia software [19].

The innermost layer is the tracker. It contains an array of wires spaced throughout its volume, held at a specific voltage. The volume is also filled with Xenon gas [20]. As an electrically charged particle moves through the space at high speed, it causes gas molecules to become ionized along the way. This ionization causes a measurable, short-lived current in the wires nearest it, allowing a "pulse" to be measured. Cohesive tracks such as the light blue-green ones seen in figure 4 are reconstructed from these pulses [21]. This means that tracks can be recorded for any electrically charged particles produced in a collision. Neutral particles, however, cannot be tracked and are detected solely by the calorimeters.

Just outside the tracker is the electromagnetic calorimeter, or "e-cal." Unlike trackers, calorimeters stop the particles that they capture, totally absorbing them. They do this in order to provide a measurement of the particle's total energy. The e-cal captures and measures the energy of two rather

specific types of particles: electrons (as well as positrons) and photons [22]. This allows physicists to immediately identify these two types of particles: if a particle leaves a track and is caught in the e-cal, then it must be an electron or positron. If it didn't leave a track but is caught in the e-cal, then it must be a photon. Distinguishing between electrons and positrons is more challenging since it involves the curvature in a particle's track. There is a uniform, vertical magnetic field throughout the detector which will cause any electrically charged particle to follow a curved path. The direction of this curve, according to the equation for magnetic force, depends on the particle's electric charge. The charge, and whether it is an electron or a positron, can be deduced based on whether the particle's track curves clockwise or counterclockwise[6].

Outside the e-cal is the hadronic calorimeter, or "h-cal." This calorimeter captures and measures the energy of any particles comprised of quarks (these particles are classified as "hadrons"). Most notably, protons and neutrons, and host of other baryons and mesons are among these particles, and are distinguishable in the same way as electrons/positrons and photons. While other hadrons are produced in plentiful supply within the ATLAS detector, they tend to have very short lifetimes and decay before reaching the h-cal. If a negatively charged hadron were to be captured by this detector layer, we would be able to identify it from the curvature in its track, just as we would with electrons and positrons. The outermost layer of the ATLAS detector detects muons and consists of many individual concentric layers. Since muons are electrically charged, they also leave tracks.

The only particles that aren't captured and measured are neutrinos[7]. Neutrinos are extremely difficult particles to detect because they are neutral and there is not a specific detector layer that can detect these neutrinos. The only indication of the presence of neutrinos in the ATLAS detector is so-

[6] The particles are, of course, moving at a high enough speed that the curvature in their tracks is often not obvious at a glance.

[7] Gluons are only found inside hadrons, binding quarks, and the W and Z bosons are short lived and decay before reaching the calorimeters.

called *missing energy*. Missing energy represents energy that we know was present in a collision event, but that wasn't captured in a calorimeter. While we can deduce that this missing energy is carried by escaping neutrinos, it still represents a limitation in the information that can be extrapolated from ATLAS data. We don't get information about the individual neutrinos, so studies of any decay process involving neutrinos can be challenging.

With the information from all layers of the detector combined, it is possible to identify the final products of a collision event. Combining the known properties of the different species of particles with the specific types of information gathered by different layers of the detector, we can determine a "signature" for each particle species. Since electrons, and positrons are electrically charged, they can be tracked in the e-cal. Thus, any particle that leaves a track and is caught by the e-cal must be an electron or positron (we can tell which based on the direction of their track's curvature). Muon and antimuons are also electrically charged, they can be tracked in the muon detector layers. Similar identifying characteristics can be determined for the other types of particles as shown in figure 6.

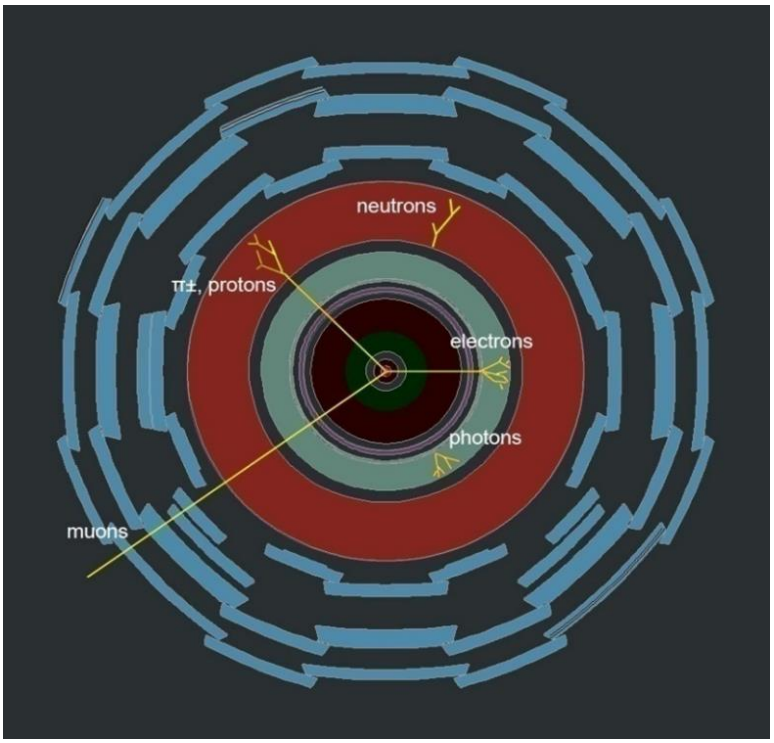


Figure 6: signatures of the various particles in the detector. Caught in the e-cal are electrons and photons, distinguishable by whether they are tracked. Caught in the h-cal are neutrons, which leave no track, protons, and pions ($\pi^{\pm}$), which are another type of charged particle composed of quarks. Muons are tracked and pass through both calorimeters.

Key Parameters

While a multitude of parameters are recorded, and can be measured, for any collision event, three are of special importance when searching for Z (or Z') boson. First among these is the transverse momentum (p_t). This refers to momentum of the particle in the plane perpendicular to the direction of the original proton beams in the ATLAS detector. Measuring transverse momentum serves two purposes: it is used as a filter criterion from the background particles. To calculate the transverse momentum of an electron or muon, it is necessary to measure the radius of curvature in their paths through the detector. This is accomplished using information from their tracks. Once the radius is determined, momentum can be calculated simply using the formula, adapted from equation 7:

$$mv = p_t = Bqr \tag{7}$$

Where q is the electric charge of the particle and B is the strength of the magnetic field present inside the detector.

Calculation of the mass of the Z (or Z') boson from its two daughter particles (electron-positron pairs or muon-antimuon pairs) requires measurements of their transverse momentum, scattering angle (the angle at which a particle's path deviates from the direction of the original proton beams), and total energy. First, transverse momentum and scattering angle can be used to determine each daughter particle's total momentum, via the formula:

$$p = \frac{p_t}{sin\,\theta}$$

(8)

Where p is total momentum, p_t is transverse momentum, and θ is scattering angle. The mass calculation follows from Einstein's equation for energy, which states that $E^2 = (mc^2)^2 + (pc)^2$, and from the principles of conservation of energy and momentum. For a Z (or Z') boson decaying into two daughter particles, we can determine the mass of the Z (or Z') boson as follows:

$$(m_Z c^2)^2 = {E_Z}^2 - (\overrightarrow{p_Z} c)^2$$

$${m_Z}^2 = \left(\frac{E_1}{c^2} + \frac{E_2}{c^2}\right)^2 - \left(\frac{\overrightarrow{p_1}}{c} + \frac{\overrightarrow{p_2}}{c}\right)^2$$

$${m_Z}^2 = \frac{{E_1}^2 + {E_2}^2 + 2E_1E_2}{c^4} - \frac{{p_1}^2 + {p_2}^2 + 2\overrightarrow{p_1} \cdot \overrightarrow{p_2}}{c^2}$$

$${m_Z}^2 = \left(\left(\frac{E_1}{c^2}\right)^2 - \left(\frac{p_1}{c}\right)^2\right) + \left(\left(\frac{E_2}{c^2}\right)^2 - \left(\frac{p_2}{c}\right)^2\right) + 2\left(\frac{E_1E_2}{c^4} - \frac{\overrightarrow{p_1} \cdot \overrightarrow{p_2}}{c^2}\right)$$

$${m_Z}^2 = {m_1}^2 + {m_2}^2 + 2\left(\frac{E_1E_2}{c^4} - \frac{\overrightarrow{p_1} \cdot \overrightarrow{p_2}}{c^2}\right)$$

$$m_Z = \sqrt{{m_1}^2 + {m_2}^2 + 2\left(\frac{E_1E_2}{c^4} - \frac{\overrightarrow{p_1} \cdot \overrightarrow{p_2}}{c^2}\right)}$$

(9)

Where a subscript '1' refers to the first daughter particle, a subscript '2' refers to the second; all masses refer to the rest masses of the respective particles, and '·' refers to the vector dot product, present in the cross-term involving momentum.

Another important parameter is called *pseudorapidity*. It is related to the angle at which the final products of a collision event deviate (scatter) from the direction of the original proton beams. It is denoted by 'η' and calculated using the formula:

$$\eta = -ln\left(tan\frac{\theta}{2}\right)$$

(10)

Where θ is the angle at which a particle deviates from the direction of the beam pipe. Pseudorapidity is used as a filter criterion to distinguish bona fide particle tracks from the low momentum background tracks. A particle's pseudorapidity is lower if the particle scatters at a larger angle, reaching a value of zero if the particle scatters perpendicular to the beam direction (θ = 90°) and becoming infinitely large if a particle does not scatter at all and remains parallel to the beam direction. Figure 7 shows a graph of pseudorapidity as a function of scattering angle, along with a visualization of the relationship:

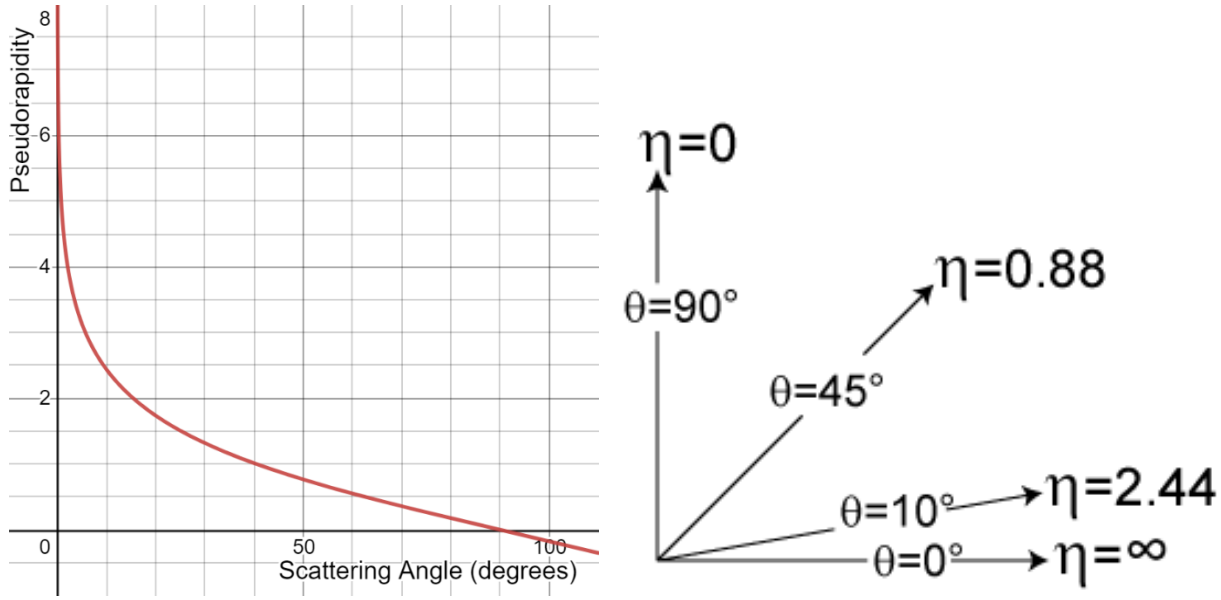


Figure 7: Left: a graph of pseudorapidity as a function of scattering angle. Notice the vertical asymptote at θ = 0 and the intersection with the θ-axis at θ = 90°. Right: a graphic showing various scattering angles from a horizontal beam pipe, and their corresponding pseudorapidity values.

Typically, an analysis would filter out events whose products have pseudorapidities above a specific value (most often excluding events with decay products having $\eta > 2$), thus selecting for decay products that scatter sharply away from the original proton beams which will most likely not be tracks from background events.

The final parameter of importance is the missing energy. Missing energy is also used as a filter criterion. We typically exclude events with missing energy greater than 25 GeV. A missing energy of 25GeV or more is considered to be a good signature of a bonafide neutrino.

ATLAS Data Analysis

At the LHC, the number of protons in a single bunch is on the order of 10^{11}, and each beam consists of nearly 3,000 bunches [13]. This means that each collision event that occurs at the LHC includes an extremely large number of interactions and decay products. This is advantageous from the standpoint of data generation – a greater number of interactions implies more possibility finding a new particle that can lead to a new particle discovery, especially when we are searching for very rare processes that are less common (such as those involving the Z' boson). However, it also provides a challenge in data analysis: the high number of interactions in each collision event means there are a lot of processes that are not relevant to a particular analysis, and therefore that is a greater chance that relevant processes will be lost[8].

The ATLAS detector itself has fast electronic systems in place that decide what processes and particle information to record and what to discard. These systems are called *trigger systems* that use a variety of methods to make these determinations [23]. Once the raw data is recorded, it is first analyzed by CERN's Tier0 Grid supercomputer computer that is equipped with several thousand cores. This initial analysis consists primarily of identification of particles and reconstruction of tracks from the raw data reported by the ATLAS detector [24]. The newly reconstructed data is then distributed to over 150 sites through a tiered computing infrastructure using the Open Science Grid (OSG) cyberinfrastructure. Bellarmine University (BU) is one of those OSG grid computing sites. The ATLAS data to which we

[8] This problem is referred to as "pileup," and efforts are constantly being made to find novel ways of mitigating it.

have access to at BU is stored in the form of many XML files, with one file representing one collision event.

The problem of pileup presents a challenge when analyzing the ATLAS data. In addition to the presence of a plethora of measurements in any collision event that may be irrelevant to an analysis, all of the interactions during a collision occur in a very small volume of space. This can make it very difficult to reconstruct particle trajectories from raw tracker data. Even after this is accomplished, we must ensure that specific decay particles originated from one parent particle from the same interaction point in the detector. This is particularly relevant for analyses of Z (or Z') boson decay: it is possible for a pair of muons or electrons to be among the final products of two separate processes that occurred in the same collision. These particles can constitute as a "fake signal," mimicking a Z (or Z') boson decay. It is necessary to eliminate the number of "fake signals" in the data analysis. This can be done by closely examining the particles' tracks and ensuring that two daughter particles included in an analysis originated from the same interaction point inside the detector.

Previous Data Analysis at Bellarmine University

Our focus at BU has been primarily on the search for evidence of the Z' boson. We have conducted analysis using the Hypatia software application, which is based on the ATLANTIS application [19]. This software allows its users to visualize a collision event and perform many types of analyses. Figure 8 shows the Hypatia interface:

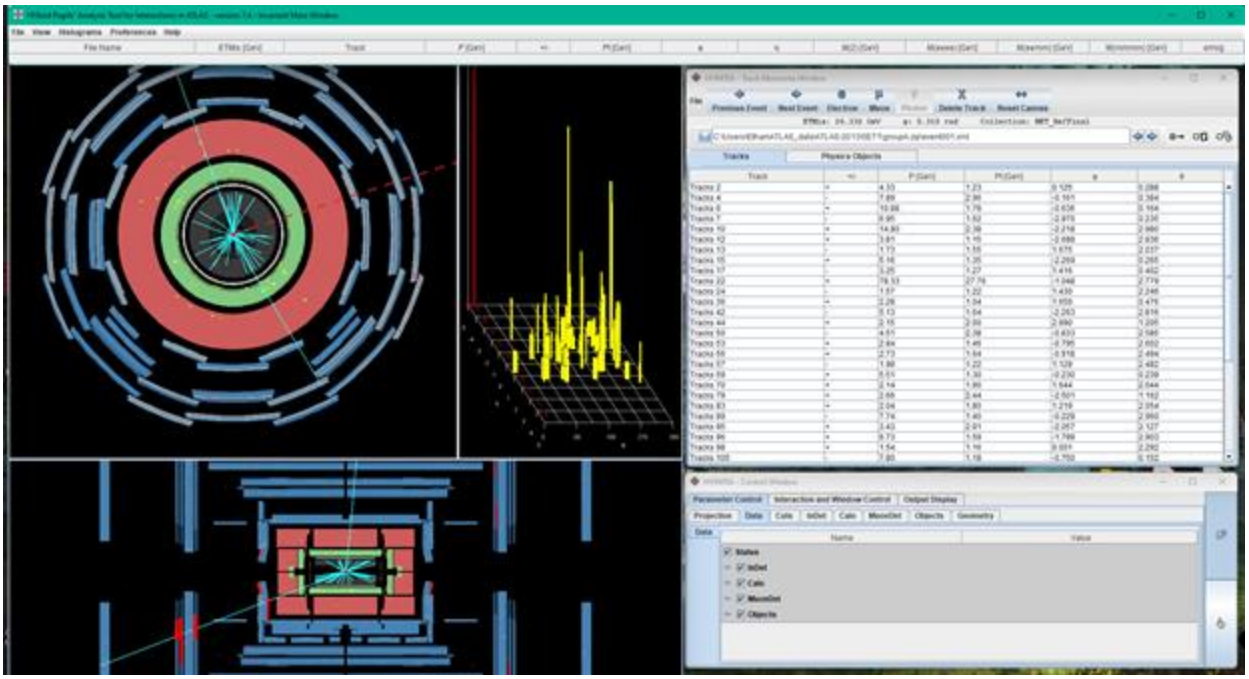


Figure 8: the Hypatia user interface. It includes a visualization of the ATLAS detector along with numerous analysis tools.

Hypatia has specific filter parameters that we can choose when analyzing a collision event to search for muon-antimuon or electron-positron pairs that are likely the products of a Z (or Z') boson decay. If such a pair of particles is found after the filter parameters are applied, Hypatia can directly provide measurements of all the necessary parameters: such as transverse momentum and pseudorapidity for each particle, missing energy for the collision event, and the mass of the Z (or Z') boson[9]. The effect of filter criteria can be seen visually, as shown in Figure 9:

[9] This is technically the calculated center-of-mass energy of the vertex from which the two product particles originated. This vertex, in our analysis, is taken to be the location of a Z or Z' boson decay.

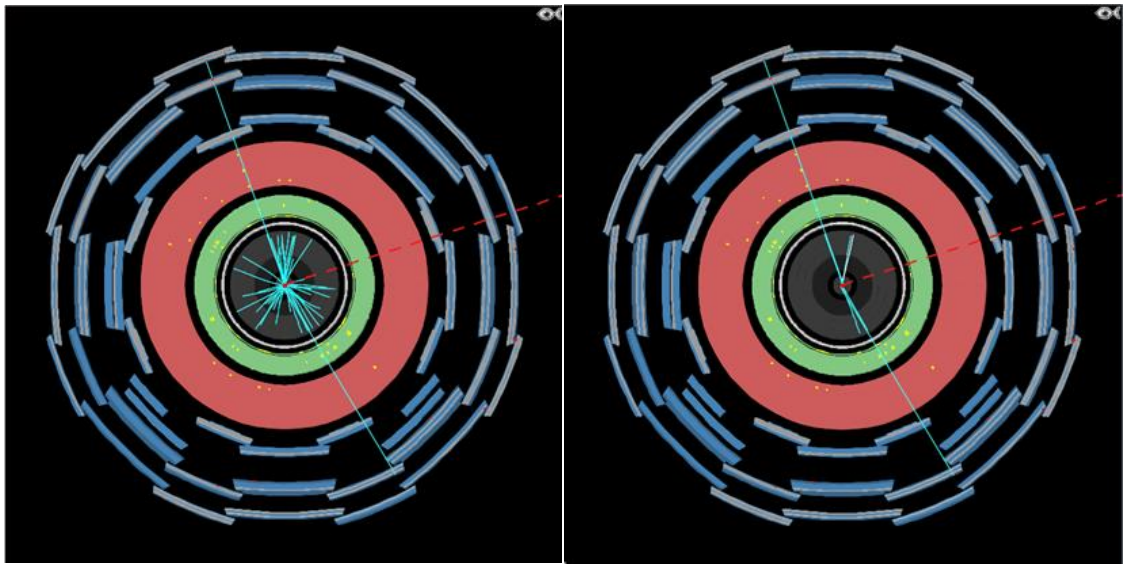


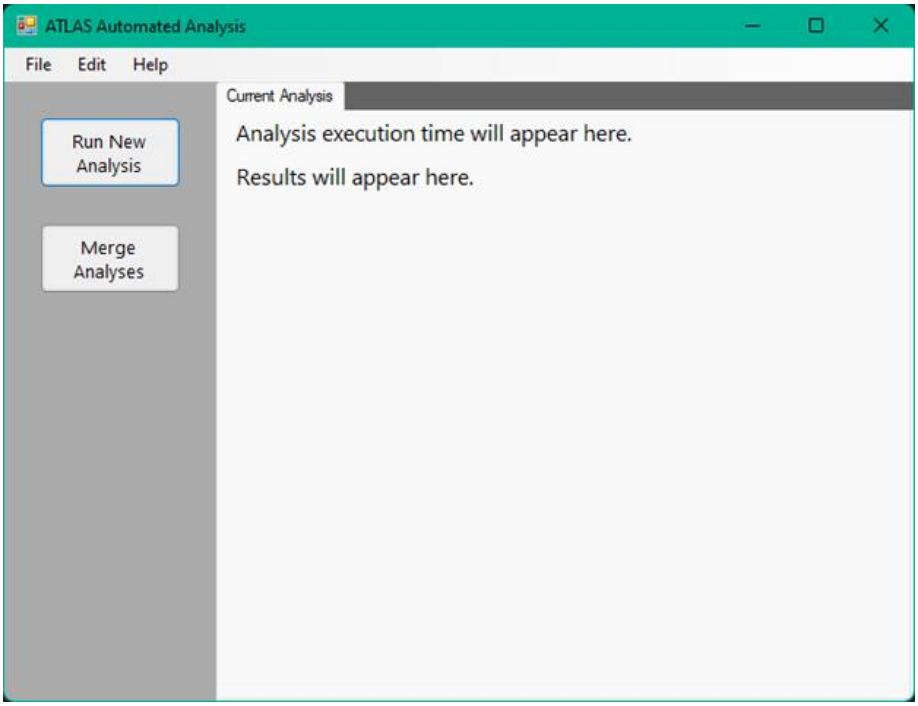
Figure 9: A collision event visualized in Hypatia before any filter criteria are applied (left) and after filtering for transverse momentum greater than 10 GeV/c (right). This event has a muon-antimuon pair that meets the transverse momentum criteria.

If no muon-antimuon or electron-positron pairs meet the filter criteria, the event is discarded. The Hypatia platform thus allows for robust analysis of ATLAS collision events. However, analyzing ATLAS data in Hypatia is also very time consuming because each collision event must be examined individually and manually.

Partial Automation as a Desktop Application

Because the individual analysis of collision events in Hypatia is time consuming, it would be advantageous to reduce the number of events that need to be analyzed. Automating the process of applying filter criteria to large sets of data would eliminate the need to examine events that fail to meet those criteria individually. Because a significant fraction of events are ultimately discarded, my application would appreciably reduce the amount of time needed to analyze large sets of the data. I have created a desktop application that accomplishes this task. My application is able to receive a large set of event files that can programmatically apply filter criteria to each of event file, and return a list of events

that are likely to contain Z (or Z') boson decays. The user interface for the application is shown in Figure 10:



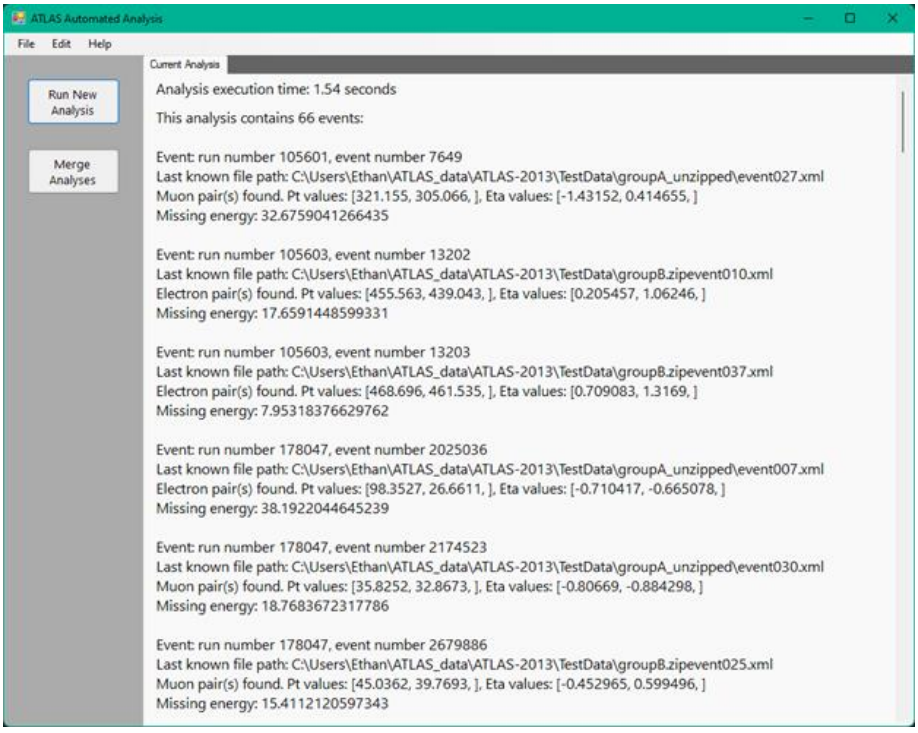


Figure 10: the user interface of my application on startup (upper image) and after running an analysis on a test set containing 100 event files (lower image). In under 2 seconds, 34 events were discarded based on filter criteria, a process that would take close to an hour using Hypatia.

Currently, the only criterion used to exclude events is transverse momentum, and the application simply records pseudorapidity and missing energy values. This is because the specific pseudorapidity and missing energy values used to exclude events may be different for different analyses (the values mentioned previously are simply those used most often). In the future, I hope to implement a tool that allows the user to set the specific filter criteria and values that can be used to exclude those events in an analysis.

It is important to note that this application cannot conduct a full analysis of the ATLAS datasets. Most importantly, it cannot examine particle tracks – any event that contains at least a muon-antimuon or an electron-positron pair that meet the transverse momentum criteria are included in this automated analysis. Because of this, the events identified by the application must still be examined using Hypatia to verify the presence of a Z boson process. Even with this limitation, the application is still useful for its ability to reduce the quantity of events that must be individually examined. Events that do not contain particle pairs that meet the filter criteria can be quickly discarded and individual analysis can focus on events more likely to contain Z (or Z') boson processes.

A key necessity of this application is the ability to analyze data files that are packaged into zipped archives, as this is a common way that the ATLAS data is distributed. Unzipping this data to a PC's hard drive would greatly increase both time and space requirements, making it highly undesirable. This application avoids this by accessing data files within zipped archives directly, rather than unzipping them. However, this technique precludes the ability to examine the contents of subdirectories or additional zipped archives located within zipped archives. These items are ignored by the application. The data being analyzed at BU consists primarily of zipped archives which only contain event files, making this limitation acceptable.

Users will be able to save their analysis (list of events that meet filter criteria) as an XML file. The analysis can be reopened in the application or used in any way the user wishes to conduct the analysis. The application's source code is available at https://github.com/EthanColbert8/ThesisApp1. The application is open-source, and anyone may use or adapt the source code freely.

Implementation Framework

I have implemented this application as a Windows Forms app using the .NET Framework (the C# programming language). This framework has two prominent advantages that make it particularly suited for this application. First, it is a very straightforward framework in which to design graphical user interfaces for desktop applications. As a result, I was able to design a logical user interface with minimal hindrance. Second, the .NET API has a well-developed set of tools for handling XML files. Since this is the format in which the event files are stored, this toolset is extremely useful in reading the XML files.

The framework has one significant drawback: the application runs on Windows-only and will not run on computers using Linux or the Mac operating systems [25]. I hope to use the Mono Project framework to create a version of the application that runs on Linux machines [26]. I do not plan to develop a Mac version of the application.

Analysis of Algorithm

The algorithm my application uses to analyze large sets of files can be summarized with the following pseudocode:

```
Input: a directory name, "TopDirectory"

define a Stack "Directories" of directory names

push TopDirectory to Directories

while (number of items on Directories > 0)
{
      pop "CurrentDirectory" from Directories

      for (each “Item” in CurrentDirectory)
      {
            if (Item is a directory)
                  push Item to Directories

            if (Item is an XML file)
                  **analyze Item as a collision event

            if (Item is a zipped archive)
            {
                  for (each “Entry” in Item)
                  {
                        if (Entry is an XML file)
                              **analyze Entry as a collision event
                  }
            }
      }
}

Output: a list of analyzed collision events that met filter criteria
```

A purely theoretical analysis of this algorithm is difficult because few of the operations used are, even approximately, basic computational operations. In particular, analyzing a single event file involves tasks such as setting up an XML reader, and scanning through the file, and interpreting specific items within it, which are highly implementation dependent. However, the simplifying assumption that analyzing a single event file runs in approximately constant time allows some inferences to be made. The following analysis ignores several factors, such as the overhead of iterating over empty directories, and is intended merely to give an approximate idea of the expected behavior.

First, consider the case in which there are no subdirectories or zipped archives within the top directory. In this case, the algorithm will simply iterate over the directory’s contents and execute an *O(1)* analysis for each XML file, ignoring any other contents. Therefore, in this simple case, the algorithm’s time complexity reduces to approximately *O(n)*, where *n* is the number of event files analyzed.

Next, consider the case in which there are no subdirectories within the top directory, but there may be zipped archives. In this case, the algorithm processes each event file in the top directory, then additionally iterates through each of the zipped archives, processing each event file in each of them. As a result, each zipped archive can be analyzed in the same way as the previous case. Therefore, letting the letter *z* index the zipped archives, the number of operations taken by the algorithm is in:

$$O(n_T) + \sum_z O(n_z)$$

(11)

Where n_T is the number of event files in the top directory and n_z is the number of event files in the z^{th} zipped archive. This is equivalent to:

$$O(n_T) + O\left(\sum_z n_z\right) = O(n_T) + O(n_Z) = O(n)$$

(12)

Where n_Z is the total number of event files in all the zipped archives and $n = n_T + n_Z$ is the "grand total" number of event files analyzed.

Finally, consider the case in which there may be event files, zipped archives, and subdirectories (which may themselves contain event files, zipped archives, and further subdirectories) in the top directory. Notice that each subdirectory is pushed to a stack and subsequently treated in the same way as the top directory. This means that, for each subdirectory, the previous case occurs. Letting the letter *m* index the subdirectories, this gives us a time complexity of:

$$O\left(\sum_m n_m\right) = O(n)$$

(13)

where n_m is the total number of event files in the m^{th} subdirectory and n is once again the grand total number of event files to be analyzed.

While this does not constitute a full theoretical analysis, it does give a rough idea of the algorithm's expected behavior: we expect the algorithm's runtime to increase approximately linearly with the number of event files analyzed. Additionally, it bears noting that a typical use case would involve a relatively shallow directory tree in which the number of event files is much larger than the number of subdirectories or zipped archives, meaning most of the runtime would indeed be spent analyzing event files.

A preliminary testing of the program confirms this analysis. The program has been run on test sets containing 50, 100, 150, 200, and 250 event files. Five of these test sets contained between one and five subdirectories, each with 50 event files, while another five contained between one and five zipped archives (and no subdirectories), each containing 50 event files, as shown in Figure 11:

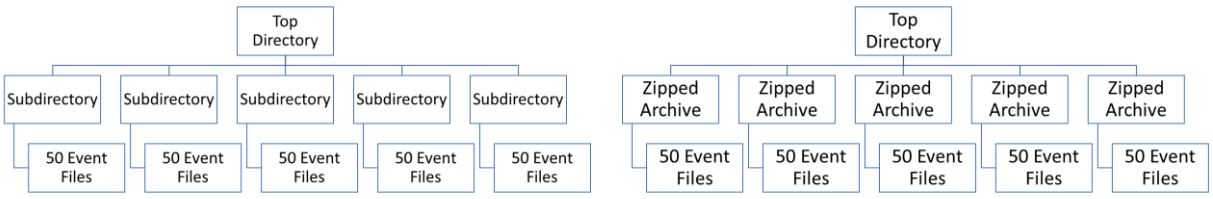


Figure 11: The structure of the test sets used to gather execution time data for my application. Sets with fewer than 250 event files had proportionally fewer subdirectories or zipped archives, but no event files were stored directly in the top directory.

The program has been run 5 times on each of the ten test sets, and the mean execution times were as follows[10]:

Mean Execution Times (n=5) (sec)		
Number of Event Files Analyzed	Unzipped	Zipped
50	0.668	0.748
100	1.235	1.446
150	1.852	2.183
200	2.421	2.868
250	3.000	3.542

Table 3: Mean execution times of my application on test sets of varying size. Unzipped data seems to be processed slightly faster than data in zipped archives, but both samples show a linear trend as the number of event files increases.

Figure 12 shows graphs of mean execution time as a function of number of event files analyzed:

[10] Actual execution times are, of course, highly machine dependent. The reported execution times are from a PC with an AMD Ryzen 5 5600X processor, 32 GB of 3200 MHz RAM, and a 2 TB SATA SSD hard drive.

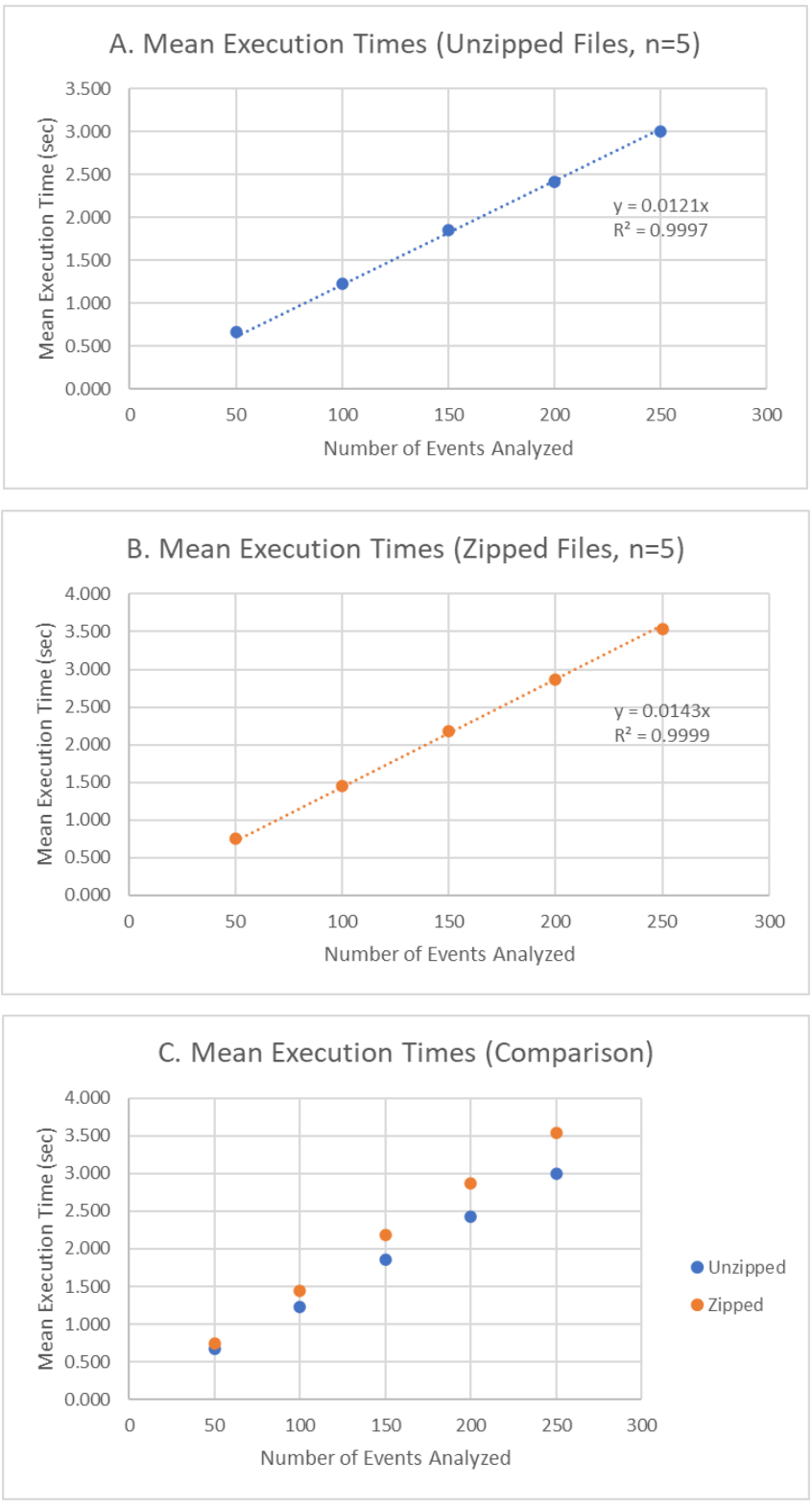


Figure 12: graphs of mean execution times as a function of number of event files analyzed, for A) unzipped data, and B) zipped data. C shows both on the same graph for comparison.

Notice that, in both cases, a linear regression model, with execution time directly proportional to number of events analyzed, fits extremely well (both coefficients of determination are above 0.999). Therefore, the actual behavior of the application aligns with the expected behavior as the size of the data set grows.

While linear time complexity is a good starting point, it can and should be improved upon. The main avenue for performance improvement is parallelization. In the future, I hope to incorporate multithreading into the application's central algorithm, allowing the use of a larger share of the machine's resources to complete analyses more quickly. This could potentially be applied on a larger scale to utilize a cluster computing system.

Conclusions

This new desktop application successfully applies filter criteria to large sets of ATLAS data events, identifying events likely to contain Z (or Z') boson processes and allowing users to save the information recorded in the analysis. The amount of time it takes to execute grows linearly with the number of event files analyzed, as was initially expected. In this sense, the development of this application has been successful. However, there are several significant improvements that can be made, including:

1. Development of a working Linux version of the application (hopefully using the Mono Project).
2. Additional extensions of basic functionalities and "quality of life" improvements, such as giving users the ability to set filter criteria.
3. Performance improvement via parallelization.

I plan to continue developing the application in the future, implementing these improvements and further developing the application's capabilities.

References

[1] CERN, "The Standard Model," 2021. [Online]. Available: https://home.cern/science/physics/standard-model. [Accessed 21 November 2021].

[2] "Lepton Summary Table," in *Review of Particle Physics*, Progress of Theoretical and Experimental Physics, 2020, pp. 34-36.

[3] S. Cecchini and M. Spurio, "Atmospheric Muons: Experimental Aspects," *Geoscientific Instrumentation, Methods and Data Systems,* vol. 1, no. 2, pp. 185-196, 2012.

[4] J. N. Bahcall, "Solar Neutrinos. I. Theoretical," *Physical Review Letters,* vol. 12, no. 11, pp. 300-302, 1964.

[5] D. Griffiths, Introduction to Elementary Particles, 2nd Revised ed., Weinheim: WILEY-VCH Verlag GmbH & Co., 2010.

[6] "Quark Summary Table," in *Review of Particle Physics*, Progress of Theoretical and Experimental Physics, 2020, p. 37.

[7] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Physics Letters B,* vol. 716, no. 1, pp. 30-61, 2012.

[8] CERN, "The Higgs Boson," 2022. [Online]. Available: https://home.cern/science/physics/higgs-boson. [Accessed 23 March 2022].

[9] CERN, "Extra Dimensions, Gravitons, and Tiny Black Holes," 2022. [Online]. Available: https://home.cern/science/physics/extra-dimensions-gravitons-and-tiny-black-holes. [Accessed 23 March 2022].

[10] "Guage and Higgs Boson Summary Table," in *Review of Particle Physics*, Progress of Theoretical and Experimental Physics, 2020, pp. 31-33.

[11] E. Accomando, A. Belyaev, L. Fedeli, S. F. King and C. Shepherd-Themistocleous, "Z' Physics with Early LHC Data," *Physical Review D,* vol. 83, no. 7, 2011.

[12] CMS Collaboration, "Search for Physics Beyond the Standard Model in Dilepton Mass Spectra in Proton-Proton Collisions at $\sqrt{s}$=8 TeV," *Journal of High Energy Physics,* 2015.

[13] CERN, "Facts and Figures About the LHC," 2022. [Online]. Available: https://home.cern/resources/faqs/facts-and-figures-about-lhc. [Accessed 28 March 2022].

[14] E. O. Lawrence, "Method and Apparatus for the Acceleration of Ions". United States of America Patent 1,948,384, 20 February 1934.

[15] E. D. Courant and H. S. Snyder, "Theory of the Alternating-Gradient Synchrotron," *Annals of Physics,* vol. 3, no. 1, pp. 1-48, 1958.

[16] B. Dotson, "How Particle Accelerators Work," 18 June 2014. [Online]. Available: https://www.energy.gov/articles/how-particle-accelerators-work. [Accessed 23 March 2022].

[17] M. Brice, "Aerial View of the CERN taken in 2008," July 2008. [Online]. Available: https://cds.cern.ch/record/1295244. [Accessed 22 April 2022].

[18] CERN, "ATLAS Detector at the LHC," 2020. [Online]. Available: http://opendata.atlas.cern/release/2020/documentation/atlas/experiment.html. [Accessed 28 March 2022].

[19] "Hypatia," University of Athens, 2014. [Online]. Available: http://hypatia.phys.uoa.gr/. [Accessed 28 March 2022].

[20] A. Vogel, "ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates," *Nuclear Instruments and Methods in Physics Research Section A,* vol. 732, pp. 277-280, 2013.

[21] G. Charpak and F. Sauli, "Multiwire Proportional Chambers and Drift Chambers," *Nuclear Instruments and Methods,* no. 162, pp. 405-428, 1979.

[22] D. E. Groom, "Calorimeters," in *Review of Particle Physics,* Progress of Theoretical and Experimental Physics, 2020, pp. 573-579.

[23] ATLAS Collaboration, "Operation of the ATLAS Trigger System in Run 2," *Journal of Instrumentation,* vol. 15, no. 10, 2020.

[24] J. Elmsheuser and A. Di Girolamo, "Overview of the ATLAS Distributed Computing System," *EPJ Web Conf.,* vol. 214, 2019.

[25] Microsoft Corporation, "Desktop Guide (Windows Forms .NET)," 16 November 2021. [Online]. Available: https://docs.microsoft.com/en-us/dotnet/desktop/winforms/overview/?view=netdesktop-5.0. [Accessed 21 April 2022].

[26] Mono Project, "Supported Platforms," 2022. [Online]. Available: https://www.mono-project.com/docs/about-mono/supported-platforms/. [Accessed 21 April 2022].

[27] B. A. Dobrescu and S. Willocq, "Z'-Boson Searches," in *Review of Particle Physics,* Progress of Theoretical and Experimental Physics, 2020, pp. 900-904.