Spring 2021

Interdisciplinary Machine Learning Methods for Particle Physics

Mariel Pettee
Yale University Graduate School of Arts and Sciences, mariel.pettee@gmail.com

Abstract

Interdisciplinary Machine Learning Methods for Particle Physics:

The Search for the Higgs Boson Produced in Association with a Leptonically-Decaying Vector Boson and Decaying to a Tau Pair, Hadronic Tau Identification in the ATLAS High-Level Trigger, and Predictions of Many-Body System Dynamics

Mariel Pettee

2021

Following the discovery of a Higgs boson-like particle in the summer of 2012 at the Large Hadron Collider (LHC) at CERN, the high-energy particle physics community has prioritized its thorough study. As part of a comprehensive plan to investigate the many combinations of production and decay of the Standard Model Higgs boson, this thesis describes a continued search for this particle produced in association with a leptonically-decaying vector boson (i.e. a W or Z boson) and decaying into a pair of tau leptons.

In Run 1 at the LHC, ATLAS researchers were able to set an upper constraint on the signal strength of this process at $\mu = \sigma/\sigma_{SM} < 5.6$ with 95% confidence using 20.3 fb^{-1} of collision data collected at a center-of-mass energy of $\sqrt{s} = 8$ TeV. My thesis work, which builds upon and extends the Run 1 analysis structure, takes advantage of an increased center-of-mass energy in Run 2 of the LHC of $\sqrt{s} = 13$ TeV as well as 139 fb^{-1} of data, approximately seven times the amount used for the Run 1 analysis. While the higher center-of-mass energy in Run 2 yields a higher expected cross-section for this process, the analysis faces the additional challenges of two newly-considered final states, a higher number of simultaneous interactions per event, and a novel neural network-based background estimation technique. I also describe advanced machine learning techniques I have developed to support tau identification in the ATLAS High-Level Trigger as well as predicting and analyzing the dynamics of many-body systems such as 3D motion capture data of choreography.

Interdisciplinary Machine Learning Methods for Particle Physics:

The Search for the Higgs Boson Produced in Association with a Leptonically-Decaying Vector Boson and Decaying to a Tau Pair, Hadronic Tau Identification in the ATLAS High-Level Trigger, and Predictions of Many-Body System Dynamics

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy
in
Physics

by
Mariel Pettee

Dissertation Director: Professor Sarah Demers

June 2021

In honor of my mom,
math major & art history minor,
Carol Nelson Sherwood Pettee
(1958 - 2017)

Acknowledgements

Things that kept my life quasi-together while writing this during a pandemic: Dirty oat milk chai lattes from East Rock Coffee, walks over every permutation of nearby streets, TikTok, eating handfuls of shredded cheese straight from the bag, tofu katsu curry, Outer Wilds, scented baths & candles, *The Crown*, virtual Compline services, Chinese Superhero rolls from Miya's Sushi (RIP), Alecia Omblero/the El Poblano Burrito Express food truck's veggie quesadillas with quinoa, egg & cauliflower kati rolls from the Ali Baba food truck, birdwatching, mood boards, formal themed dinners, Zoom play readings with friends, and my house's very affectionate cat, Bristol.

Thesis-writing music recommendations: HD Crackling Birchwood Fireplace 6 Hours (`https://youtu.be/UraL4IEcu-o`), Roman Roads IV-XI by Land Observations, Steve Reich's Music for 18 Musicians, Philip Glass's String Quartet No. 5, the soundtracks to *Phantom Thread* & *Moonlight*, chill lo-fi hip hop beats for writing a physics PhD dissertation, Grimes, Robyn, and your favorite dance party playlist from five years ago.

The people who made this thesis possible:

I am deeply grateful to my thesis committee (Professors Sarah Demers, Paul Tipton, Walter Goldberger, and Karsten Heeger) for their thoughtful insights and support.

Thanks to Chase Shimmin, Douglas Duhaime, Ilya Vidrin, Raymond Pinto, Justin Berry, Dana Karwas, the Yale Center for Collaborative Arts & Media, the Yale Womens' Faculty Forum, and the Yale Digital Humanities Lab, I was given ample room to grow as an independent researcher during my years here surrounded by brilliant, warm, generous, and deeply inspiring collaborators, thinkers, and friends.

Thanks to Santiago Miret, Somdeb Majumdar, Marcel Nassar, Hanlin Tang, Amir Khosrowshahi, and the rest of the Intel AI Lab, I was offered the exceptionally rare chance to collaborate with expert AI researchers in pursuit of my own strange research goals. In just a few short months, I gained a wealth of research skills, wonderful colleagues, and a sense of confidence in setting my own research agenda.

Thanks to Antonio de Maria, Quentin Buat, Bertrand Martin Dit Latour, Christopher

Deutsch, Pier-Olivier Deviveiros, Duc Bao Ta, and the whole Tau Trigger team & HLeptons analysis group, I consistently found great selflessness and camaraderie even from across an ocean. Thanks to Michela Paganini, Luke de Oliveira, Dan Guest, and more, I was embedded in the ideal support network for thinking about machine learning research and ended up with fantastic friendships as well.

Thanks to Lotte Thomsen, Giani Pezzullo, Savannah Thais, Emma Castiglia, Claire Laffan, and India Bhalla-Ladd, I was so proud to be a part of the Run 2 VH analysis. Your monumental efforts and support permanently shaped me for the better as a scientist and our time together made my PhD experience a total joy.

Thanks to the creative collaborators I've found here, I was able to fully thrive as an artist while pursuing my PhD, something I never imagined would be possible. For that gift, I will be forever grateful. These people include Liam Bellman-Sharpe, the Yale Cabaret, the Yale School of Drama, Elm City Dance Collective, Angharad Davies, Lee Sher/Saar Harari, Amanda Castro & Urban Bush Women, the Paul Taylor 2 Dance Company, Germaul Barnes/Catherine Cabeen/Bill T. Jones, Vicky Shick, and Renee Robinson, the dancers from *A Different Drum*, and Yale Dance Theater/Yale Dance Lab.

Thanks to my fellow Yale Physics cohort (incoming class of 2015), I was given countless late nights filled with silliness, problem-solving, dancing, cheap takeout dinners, cupcakes, and crashing on the LoveSac™. I have the deepest gratitude for them and for my many cherished friendships from the broader Yale Physics community.

Thanks to my lifelong friends from high school, college, University of Cambridge, and more, I have a tireless cheerleading squad composed of those who know me best. Thanks to my New Haven community, roommates, and podmates, I have been surrounded by the most fun, eclectic, innovative, and sincere people, and have never felt alone.

Thanks to Emily Coates and Sarah Demers, I am well aware that I hit the academic mentorship jackpot. Their impact on me has been immense. I have learned from them to carve my own intellectual path with fearlessness, hard work, and empathy.

Finally, I am so grateful to my family and my partner, Danny Seara, for loving me unconditionally and giving my life such bliss, richness, and meaning.

Contents

1 Theoretical Motivation

Leucippus and Democritus had the right idea, at least. Circa 400 BCE, this pair established the notion of atomism, from the Greek *atomos*, meaning "indivisible", in an attempt to define the core elements of physical reality [1]. Atomism asserted that the universe fundamentally consists of indestructible particles swirling and colliding to form temporary structures in an otherwise empty void.

In the millenia since, the particle physics community has learned that the vacuum of space is hardly empty (instead, it's more of a carbonated beverage bubbling with pairs of particles that pop in and out of existence) and that many of these particles are actually quite short-lived. The notion that the universe is a grand assemblage of particles with no constituent parts, however, persists through today in the widely-accepted scientific theory known as **the Standard Model**. Established in the latter decades of the twentieth century, the Standard Model provides a theoretical description of how the universe operates at its most basic level: *what is the all the stuff in the universe made of, and what determines how this stuff behaves?* Particle physicists are therefore concerned with understanding the fundamental pieces of matter (particles) and their many kinds of interactions (forces).

The Standard Model is often considered a triumph within the history of physics and perhaps of human intellectual accomplishment in general. As a theory, it is remarkably powerful – with just a few notable exceptions, experimental results across the board align beautifully with its predictions, and it includes the most precise agreement between theory and experiment in the history of science (over 10 significant digits!). Its success is the result of the interplay of brilliant advances from both theorists and experimentalists over many decades. In this chapter, I'll describe the theoretical and mathematical frameworks underpinning this theory.

1.1 The Standard Model of Particle Physics

The dozens of fundamental particles described by the Standard Model can be neatly categorized based on something called their **spin**, a concept related to the weirdness of angular momentum in quantum mechanics. In the classical (i.e. macroscopic) world, we can measure different forms of angular momentum in, say, the rotation of the Earth. The angular momentum of the Earth is composed of an *orbital* component from its yearly rotation around the Sun as well as a *spin* component from its daily rotation, each of which is calculated differently. In the quantum (i.e. subatomic) world, we can also measure orbital and spin angular momenta, but the rules of quantum mechanics require these calculations to look rather different than their classical analogues. While a classical measurement of angular momentum could result in any number, any measurement of a particle's quantum *orbital* angular momentum will result in a restricted set of quantized values $\hbar\sqrt{l(l+1)}$, where $l = (0, 1, 2, 3, \cdots)$ and $\hbar$, Planck's constant, is $\approx 1.05 \times 10^{-34} \frac{\mathrm{m^2 \cdot kg}}{\mathrm{s}}$. Even more intriguingly, every measurement of a given particle's quantum *spin* angular momentum will always be exactly the same: $\hbar\sqrt{s(s+1)}$, where s could be an integer or half-integer. Since s (the spin quantum number, or just "spin") is fixed for every fundamental particle, it's a very useful fundamental property for organizing our picture of the Standard Model's description of matter and forces.

Each of the fundamental particles of the Standard Model are shown in Figure 1, along with lines describing which particles interact with each other.

1.1.1 Fermions: $s = \pm\frac{1}{2}$

Fermions are the core elements of all known observable matter in our universe. In fact, nearly all matter that we interact with on a daily basis is made up of just three fermions: the electron, the up quark, and the down quark. While fermions are defined as particles with half-integer spin values, in practice, every known fundamental fermion has spin $= \pm\frac{1}{2}$.

The fermions, as illustrated in Table 1, can be further subdivided into **quarks**, which interact with the strong force, and **leptons**, which do not. All of the fermions interact via the weak force, however. Each quark and three of the leptons, the electron (e), muon (μ), and tau (τ), also carry an electric charge. The remaining neutral leptons, neutrinos, interact via the weak force, but not the strong or electromagnetic forces. Each fermion has a corresponding antiparticle with the same mass but opposite essential properties such as charge, lepton number, and chirality. Interestingly, both leptons and quarks have six known

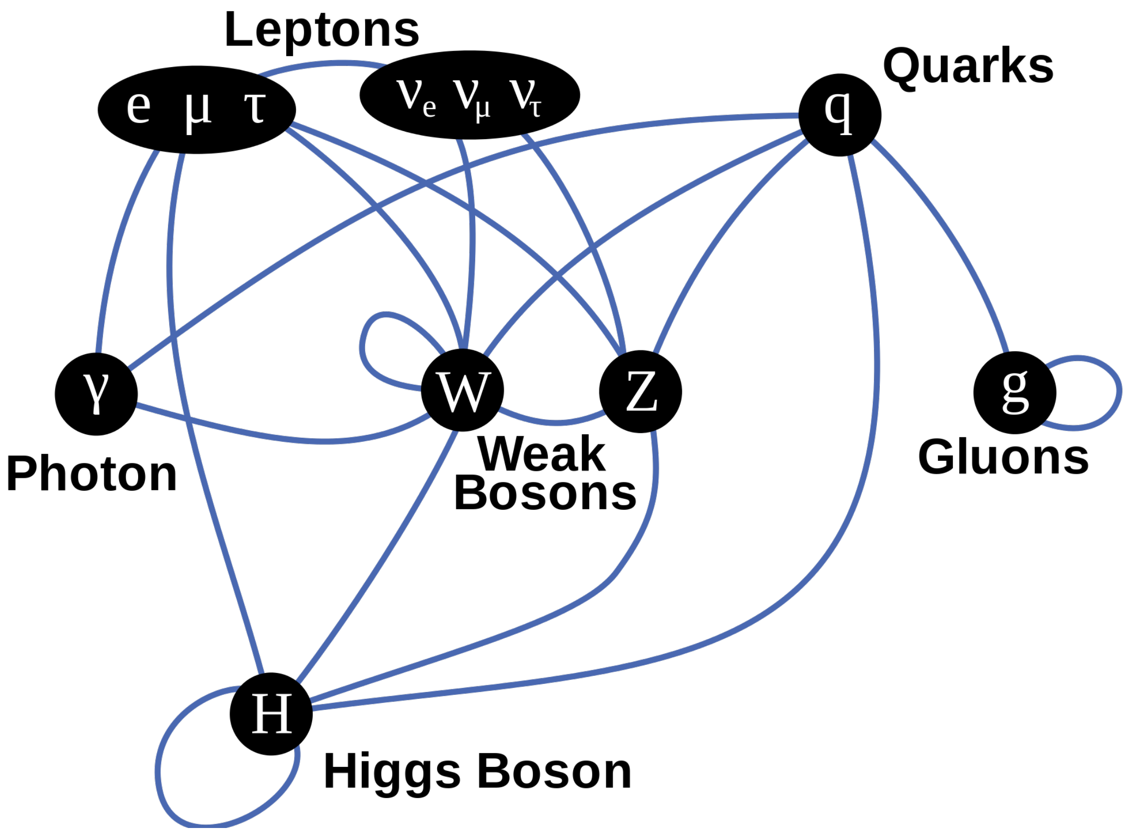


Fig. 1: The fundamental particles of the Standard Model are shown as black circles, while the connections between them indicate how they interact. The photon only interacts with charged particles: the charged leptons and the W boson. The weak bosons, W and Z, interact with all leptons and quarks. Gluons, however, only couple with quarks. Lastly, the Higgs boson couples with every particle that has a mass except neutrinos – the charged leptons, quarks, and weak bosons. (Public domain image from Wikimedia Commons).

fundamental particles that may be conveniently arranged into pairs known as Generations I, II, and III, in order of increasing mass.

Particle	Symbol	Generation	Electric Charge [e]	Mass
Electron	e	I	-1	0.511 MeV
Electron Neutrino	ν_e	I	0	< 2 eV
Muon	μ	II	-1	106 MeV
Muon Neutrino	ν_μ	II	0	< 0.19 MeV
Tau	τ	III	-1	1.78 GeV
Tau Neutrino	ν_τ	III	0	< 18.2 MeV
Up Quark	u	I	$2/3$	2.2 MeV
Down Quark	d	I	$-1/3$	4.7 MeV
Charm Quark	c	II	$2/3$	1.28 GeV
Strange Quark	s	II	$-1/3$	95 MeV
Top Quark	t	III	$2/3$	173 GeV
Bottom Quark	b	III	$-1/3$	4.18 GeV

Table 1: Properties of the Standard Model fermions (antiparticles omitted), all with spin $1/2$. Masses are approximate and up-to-date with 2018 Particle Data Group listings [2].

Particle	Symbol	Spin	Electric Charge [e]	Mass
Photon	γ	1	0	0
W Boson	$W^\pm$	1	$\pm$ 1	80.4 GeV
Z Boson	Z^0	1	0	91.2 GeV
Gluon	g	1	0	0
Higgs Boson	H	0	0	125 GeV

Table 2: Properties of the Standard Model bosons. Masses are approximate and up-to-date with 2018 Particle Data Group listings [2].

1.1.2 Vector Bosons: $s = \pm 1$

Vector bosons are the conduits of the three fundamental forces treated in the Standard Model: electromagnetism, the weak force, and the strong force. Like the fermions, they are fundamental particles, but they exhibit a unique behavior as mediators of particle interactions. This means that each fundamental force has one or more corresponding vector bosons that are exchanged between two fermions during an interaction. The particles belonging to this category (all with spin 1) are the photon (γ), which mediates electromagnetism; the W & Z bosons ($W^\pm$ & Z^0), which mediate the weak force; and the gluon (g), which mediates the strong force.

1.1.3 Scalar Bosons: $s = 0$

Until 2012, the only fundamental bosons observed experimentally had spin $s = \pm 1$, though theoretically in the Standard Model there could exist bosons with spins of any integer. The announcement of the discovery of the Higgs boson at CERN on July 4, 2012 introduced particle physics to its first **scalar boson**, meaning a boson with spin $s = 0$. All bosons, vector and scalar, are summarized in Table 2. While the Higgs boson isn't a force-carrier like the vector bosons, it is a physical excitation of the Higgs field, which couples with several fundamental particles to give them nonzero masses. The theoretical underpinnings of the Higgs boson are central to this thesis, so I will explore them in detail in Section 1.4.

1.1.4 Conservation Rules

Once equipped with a given pair of particles to begin with, one can proceed to construct all possible subsequent interactions by piecing together the various interaction vertices allowed under the conservation rules of particle physics. Particle interactions observed in nature should conserve:

- **Energy and momentum**: This means that massive particles will decay into less massive particles, unless prevented by another conservation law. (A *decay* is a spontaneous conversion into other particles.)

- **Electric charge**: The sum of electric charges of the beginning particles must equal the sum of the electric charges of the final-state particles.

- **Color charge**: All observable particles are color-neutral, where *color* is analogous to electric charge for the strong force.

- **Baryon number**:
$$\frac{(\#\text{ of quarks}) - (\#\text{ of antiquarks})}{3}$$

- **Lepton number**:
$$(\#\text{ of leptons}) - (\#\text{ of antileptons})$$

- **Charged lepton flavor**: The number of leptons of a particular flavor/generation of the charged leptons (e, μ, and τ). Neutrinos, which are electrically neutral, have been observed in recent decades to violate lepton flavor conservation through the process of neutrino oscillations.

1.1.5 Feynman Diagrams

Though particle physics is inherently a study of particles we often cannot see, a theoretical contribution from the famously unconventional physicist Richard Feynman called **Feynman diagrams** provides a visual framework for understanding particle interactions. The apparent simplicity of these diagrams belies their capacity to represent the complex integrals and dynamics that we'll explore in Section 1.2.

A Feynman diagram represents a particle interaction over time. I will use the convention that time moves from left to right across the diagram, meaning that the leftmost lines represent the particles before the interaction, and the rightmost lines represent the particles in the final state of the interaction. Internal lines represent so-called *propagators* or *virtual particles* – these are the mediator particles of the interaction in question, and are not detected in either the initial or final states of the process.

Feynman devised a straightforward visual system of lines and interaction vertices to represent particles and their interactions, as seen in Table 3. Given the guidelines just described in Section 1.1.4, only a limited number of interaction vertices are permitted under the constraints of the Standard Model. These vertices usually have three (or, rarely, four) lines attached, representing particles. From this foundational set of interaction vertices, we can construct full interactions by sticking these basic interactions together like a set of tinker toys. The most important 3-particle allowed interaction vertices are summarized in Figures 2 - 5.

1.1.6 Physics Beyond the Standard Model

Despite the many successful experimental predictions that have emerged from the Standard Model, it should not be mistaken as a complete theory of the fundamental physics of

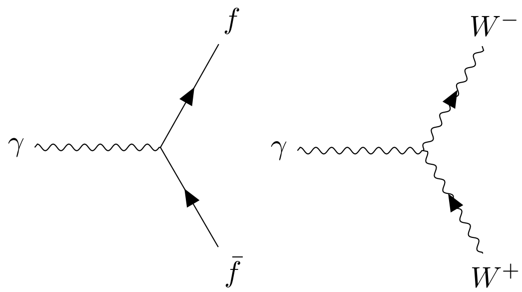


Fig. 2: Primary allowed 3-particle vertices for Quantum Electrodynamics (QED), i.e. the electromagnetic force, where f is a charged fermion of any flavor.

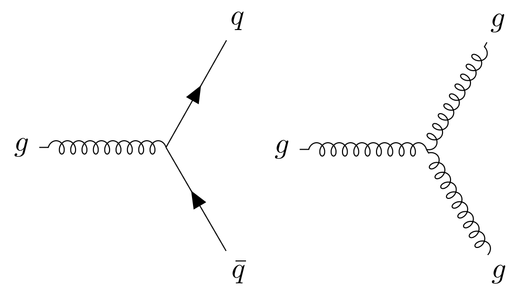


Fig. 3: Primary allowed 3-particle vertices for Quantum Chromodynamics (QCD), i.e. the strong force, where q is any flavor of quark.

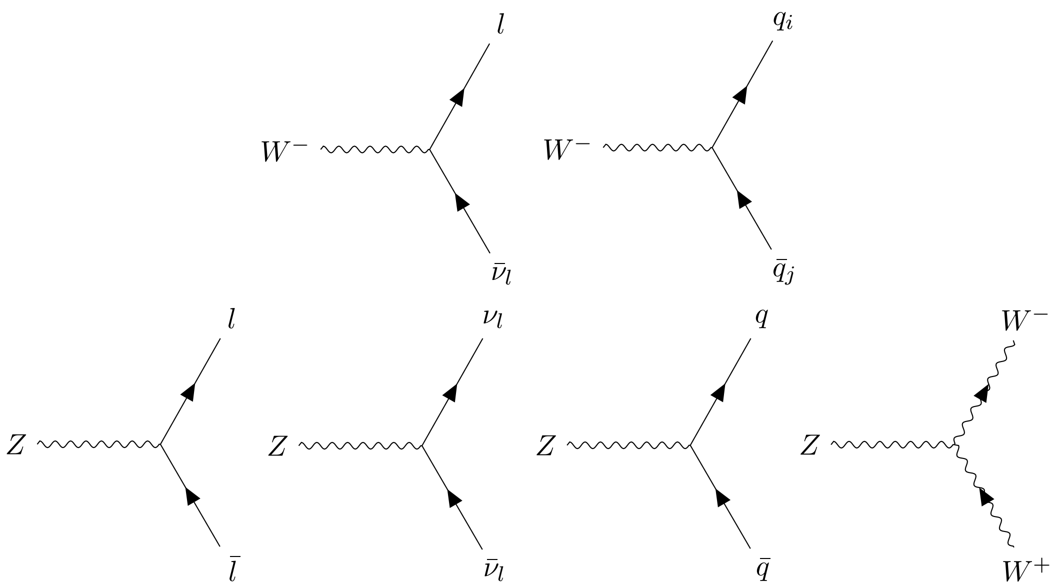


Fig. 4: Primary allowed 3-particle vertices for the electroweak (EW) force. Notably, flavor-changing neutral currents are disallowed, meaning the Z must couple with fermions of the same flavor.

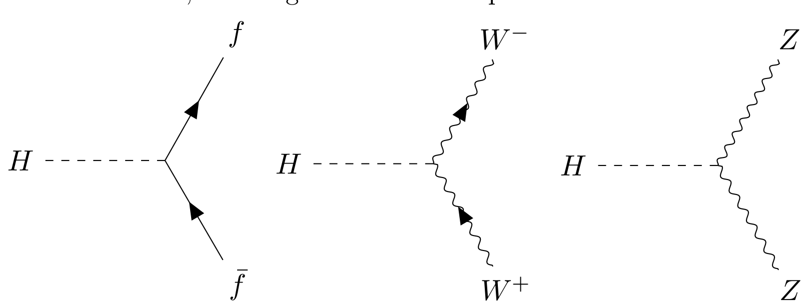


Fig. 5: Primary allowed 3-particle vertices for the Higgs boson.

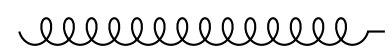
Particle	Symbol
Fermion	
Anti-Fermion	
γ, W, or Z	
Gluon	
Higgs boson	

Table 3: The building blocks of Feynman diagrams are made up of lines to indicate particles. With time reading from left to right, antiparticles may be thought of as particles moving backwards in time (though this shouldn't be taken literally).

the universe. There are several major observed phenomena that lack a theoretical explanation by the Standard Model, including:

- **Gravity:** Compared with the other forces described by the Standard Model (electromagnetism, the weak force, and the strong force), the force of gravity is extremely weak – so much so that it can typically be ignored between subatomic particles because other forces overwhelm it. We don't yet know whether or not it behaves as the other fundamental forces do at the quantum level.

- **Dark Matter:** The consensus of the astrophysics community is that the majority of matter in the universe – about 5 times the amount of known matter – is made up of an unknown substance called *dark matter*. While dark matter has not been directly detected, its presence has been inferred through a variety of indirect calculations relating to angular velocities of galaxies, gravitational lensing, the Cosmic Microwave Background, and the apparent mass distributions of galaxy clusters. Several theories of dark matter suggest that it should take the form of a fundamental particle, but such a particle is not included in the Standard Model.

- **Dark Energy:** Dark energy is believed to make up the majority of the energy content of the universe and to be responsible for the accelerating expansion of the universe. Evidence for this accelerating expansion has been collected from analyses of Type Ia supernovae as well as peaks in the correlation function of baryon acoustic oscillations in the early universe.

- **Matter-Antimatter Asymmetry:** The Standard Model holds that matter and antimatter are created in equal amounts in allowed interactions, with a small amount of asymmetry introduced by a concept called CP violation. However, astrophysical measurements tell us that the nearby universe is overwhelmingly composed of matter, with very little antimatter. The Standard Model's allowance of CP violation alone cannot account for the vast matter-antimatter asymmetry seen in the observable universe.

- **Neutrino Masses:** The Standard Model was extended to include a mathematical description of neutrino oscillations following their discovery. However, this extension necessitates that neutrinos have non-zero masses. Experimental results show that the neutrino masses are many orders of magnitude smaller than the masses of the other Standard Model particles, but the Standard Model does not explain this discrepancy, nor does it outline a mechanism that grants neutrinos their masses.

1.2 Quantum Field Theory (QFT)

The theoretical success of the Standard Model is rooted in Quantum Field Theory (QFT), a framework that extends the physical laws of the subatomic world (*quantum mechanics*) to extremely high energies and fast speeds (*special relativity*). More specifically, QFT becomes relevant when we are interested in measuring distances that are smaller than the Compton wavelength of a relativistic particle $\left(\lambda_C = \frac{\hbar}{mc}\right)$.

To understand why quantum mechanics alone is insuffient to describe particles at distances this small, we must understand that at these length scales, it becomes impossible to identify a specific location of a given particle. Why is this? Consider perhaps the most famous equations representing each of the two fields entwined in QFT:

1. **Special Relativity:**
$$E = mc^2 \tag{1}$$

2. **Quantum Mechanics:** Heisenberg's Uncertainty Principle,
$$\Delta E \Delta t \geq \frac{\hbar}{2} \tag{2}$$

The first of these equations states that energy can be converted into mass, and vice-versa, with an exchange rate of c^2, the speed of light squared, for a particle at rest. The second equation demonstrates the fundamental limits of measuring multiple properties of a particle at once – we can individually precisely measure the energy of a particle or a particular length of time, but cannot measure both at once with perfect accuracy.

In a particle physics experiment for which we are interested in measuring how far a moving particle travels to great precision, we could equivalently try to measure how long a moving particle stays in our detector with great precision. Thus, our uncertainty in our measurement of time, Δt, must be very small. But this has consequences for our uncertainty in ΔE:

$$\Delta E \geq \frac{\hbar}{2\Delta t} \tag{3}$$

If $\Delta t \leq \frac{\lambda_C}{2c}$, then $\Delta E \geq mc^2$. In other words, the uncertainty in our energy measurement is wide enough to include the possibility that a particles of mass m could have spontaneously emerged from that amount of energy. If an identical particle could emerge at our measurement site, how could we be sure which particle we were attempting to measure in the first place? The Compton wavelength therefore marks the threshold at which quantum mechanics and special relativity become incompatible.

To handle this problem, the framework of QFT shifts its perspective from the analysis of single particles of fixed numbers to the analysis of **fields** that permeate all of space and time, from which many particles could frequently be appearing and disappearing. In QFT, these fields are the bedrock of the universe and particles are excited states of these fields.

1.2.1 Lagrangians in QFT

The dynamics of quantum fields are encoded in a mathematical construct called a Lagrangian density, or just **Lagrangian** for short. This Lagrangian is an analogue of the Lagrangians commonly used in classical mechanics, which are defined as

$$L = \text{(Kinetic Energy) - (Potential Energy)}. \tag{4}$$

Classically, once one has determined the Lagrangian describing a desired physical system, one can derive the equations of motion for that system by calculating the Euler-

Lagrange equations for the coordinates q_i and their time derivatives $\dot{q}_i$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0. \tag{5}$$

Extending the Lagrangian L for discrete particles to describe quantum fields that exist throughout spacetime means redefining L as an integral over a Lagrangian density $\mathcal{L}$, a function of one's coordinates ϕ_i and their partial derivatives over the three coordinates of space and the one coordinate of time, all indexed by μ:

$$L = \int \mathcal{L}(\phi_i, \partial_\mu \phi_i)\ \mathrm{d}x^\mu \tag{6}$$

It is this $\mathcal{L}$ that we commonly refer to as the Lagrangian in QFT. Finally, we can follow the same principle[1] that led to the derivation of Equation 5 to derive the Euler-Lagrange equation for a relativistic field ϕ_i:

$$\partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu \phi_i)}\right) - \frac{\partial L}{\partial \phi_i} = 0. \tag{7}$$

1.2.2 Local Gauge Invariance

One additional concept that is crucial to grasping how the Standard Model operates is **symmetry**. In Noether's Theorem, the mathematician and theoretical physicist Emmy Noether revealed a crucial connection between a Lagrangian's symmetries and corresponding conserved quantities: When physicists discuss symmetries, "symmetric" is synonymous with "invariant" or even "redundant" – each of these terms reflects that the Lagrangian can be modified in some way that will have no ultimate effect. Classically, for example, we can use Noether's Theorem to show that the law of conservation of energy is actually a result of classical Lagrangians being invariant in time.

Similarly, the Standard Model is explicitly structured so that the Standard Model Lagrangian obeys a special kind of symmetry called a **local gauge symmetry**. A gauge

[1] The Principle of Least Action can be understood as the notion that objects in the universe follow the most efficient paths that satisfy their constraints from the laws of physics. For example, a beam of light will always follow the shortest path between two points. The *action* (S) is defined as the integral of a Lagrangian over time: $S = \int_{t_1}^{t_2} L\ \mathrm{d}t$. One can derive the path followed by a given object by calculating the path between two states for which the action S is minimized ($\delta S = 0$).

symmetry differs from more familiar forms of symmetry like angular invariance, time invariance, translation invariance, etc. because the symmetry is mathematical, not physical. If there are multiple mathematical descriptions of relativistic quantum fields that result in the same Lagrangian, i.e. the same physical dynamics, then that Lagrangian has an inherent gauge symmetry. Moreover, a *local* gauge symmetry implies that this mathematical symmetry doesn't have to be uniformly applied at every point in spacetime.

The mathematical procedure of promoting global gauge symmetries to local gauge symmetries reveals which particles in the Standard Model interact with each other. This is because converting a global gauge symmetry to a local one necessitates the creation of new dynamical gauge fields that interact with other particles. In the next section, I'll outline this process in a simple case for demonstration purposes, but this same method is applied for each sector of the Standard Model to construct the overall Standard Model Lagrangian and the dynamical gauge fields associated with it. Then, in Section 1.4, I'll show how this central tenet of local gauge invariance necessitates the existence of the Higgs boson.

1.2.3 Promoting a Global Symmetry to a Local Symmetry

Consider the Dirac Lagrangian representing an interaction between fermions:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{8}$$

where ψ represents a fermion and $\overline{\psi} = \psi^\dagger \gamma^0$ is constructed such that $\overline{\psi}\psi$ is Lorentz-invariant. To generate a Lagrangian for this QFT, we will additionally require that $\mathcal{L}$ obey a global (for now) gauge symmetry $\psi \to e^{i\theta}\psi$, for which θ is a constant – essentially a constant phase shift for the fermionic field at every point in spacetime.

Propagating this transformation of ψ through the Lagrangian, we find:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{9}$$

$$= (e^{-i\theta}\overline{\psi})(i\gamma^\mu \partial_\mu - m)(e^{i\theta}\psi) \tag{10}$$

$$= \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{11}$$

$e^{\pm i\theta}$ are each constants that can commute with each of the other terms in the Lagrangian, and they will cancel each other out. The Lagrangian is therefore invariant under this global gauge transformation.

To promote this global symmetry to a local one, we will now require that $\mathcal{L}$ obey $\psi \rightarrow e^{i\theta(x)}\psi$, where $\theta(x)$ is now a function of x^μ, i.e. a point in spacetime. We need not propagate this transformation of ψ through the Lagrangian, though, to know that the Lagrangian will not be invariant under this local gauge symmetry. This is because of the term involving the partial derivative ∂_μ:

$$\partial_\mu(e^{i\theta(x)}\psi) = \partial_\mu(e^{i\theta(x)})\psi + (e^{i\theta(x)})\partial_\mu\psi \tag{12}$$

$$= ie^{i\theta(x)}\partial_\mu\theta(x) + (e^{i\theta(x)})\partial_\mu\psi \tag{13}$$

There is a new term relating to $\partial_\mu\theta(x)$ that prevents this derivative term from achieving local gauge invariance.

The fix to make the field ψ invariant under this local gauge symmetry is actually to introduce a new, massless field A_μ such that under the same local gauge transformation,

$$A_\mu \rightarrow A_\mu - \partial_\mu\theta(x). \tag{14}$$

We have therefore constructed (by hand) a term to cancel out the problematic $\partial_\mu\theta(x)$ term from Equation 13. We can then formulate a quantity that *is* invariant, namely:

$$D_\mu = (\partial_\mu + iA_\mu). \tag{15}$$

We refer to Equation 15 as the **gauge covariant derivative**, i.e. the derivative that transforms just like the fermion field under local gauge invariance:

$$D_\mu\psi \to (D_\mu\psi)' = (\partial_\mu + iA'_\mu)\psi' \tag{16}$$

$$= (\partial_\mu + i(A_\mu - \partial_\mu\theta(x)))(e^{i\theta(x)}\psi) \tag{17}$$

$$= \partial_\mu e^{i\theta(x)}\psi + iA_\mu e^{i\theta(x)}\psi - i\partial_\mu\theta(x)e^{i\theta(x)}\psi \tag{18}$$

$$= \cancel{ie^{i\theta(x)}\partial_\mu\theta(x)} + (e^{i\theta(x)})\partial_\mu\psi + iA_\mu e^{i\theta(x)}\psi - \cancel{ie^{i\theta(x)}\partial_\mu\theta(x)\psi} \tag{19}$$

$$= e^{i\theta(x)}(\partial_\mu + iA_\mu)\psi = e^{i\theta(x)}D_\mu\psi. \tag{20}$$

The theory is now fundamentally altered – it includes a new gauge field, A_μ, that interacts with the fermions already present in the theory. A_μ is necessarily massless because its associated mass term would not be invariant under the same gauge transformation. However, as a whole, the theory is now invariant under the local gauge symmetry, as desired:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi \to \mathcal{L}' = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi \tag{21}$$

1.2.4 Continuous Symmetry Groups in the Standard Model

Each particle in the Standard Model obeys certain physical symmetries (translation-invariance, rotation-invariance, and invariance under Lorentz boosts, or the set of all space-time transformations that preserve a constant speed of light) captured by the Poincaré group $\mathrm{SO}^+(1,3) \rtimes \mathbb{R}^{(1,3)}$ as well as three local gauge symmetries: $\mathrm{SU}(3)_\mathrm{C}$, $\mathrm{SU}(2)_\mathrm{L}$, and $\mathrm{U}(1)_\mathrm{Y}$. Each of these symmetries is continuous (such as a rotation in 3D space, which can be parameterized by continuous parameters like θ and ϕ), as opposed to discrete (such as a reflection across an axis of symmetry, which cannot be described by a continuous parameter). Overall, then, the entire Standard Model Lagrangian is invariant under

$$\mathrm{SO}^+(1,3) \rtimes \mathbb{R}^{(1,3)} \times \mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}. \tag{22}$$

Eugene Wigner laid the foundation for the remarkable correspondence between these continuous symmetry groups, also called **Lie groups**, and the physical particles in the universe in 1939 with a study of the Poincaré group. He found that he could classify the known particles based on irreducible mathematical representations of the Poincaré group.

In particular, many irreducible representations of the Poincaré group can be parameterized by two quantities m and s, where $m \geq 0$ and $s \in \frac{\mathbb{Z}}{2}$. These mathematical quantities directly correspond with the basic physical properties of *mass* and *spin* for fundamental particles! We will soon discover that the irreducible representations of the other groups in the SM Lagrangian also correspond with other essential properties of the fundamental particles.[2]

$\mathrm{SU}(3)_{\mathrm{C}}$ refers to the local gauge invariance in Quantum Chromodynamics (QCD) related to color charge; $\mathrm{SU}(2)_{\mathrm{L}}$ refers to the local *chiral* gauge invariance of **weak isospin**, and $\mathrm{U}(1)_{\mathrm{Y}}$ refers to the local gauge invariance in the weak force of **weak hypercharge**. While $\mathrm{SU}(3)_{\mathrm{C}}$ describes the QCD sector of the Standard Model on its own, $\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_{\mathrm{Y}}$ combine to form the Standard Model's Electroweak (EW) sector. Incorporating these three local gauge symmetries into the Standard Model Lagrangian $\mathcal{L}_{\mathrm{SM}}$ introduces new, massless gauge fields that will correlate directly with the very vector bosons described in Section 1.1.2, though the journey from these three symmetry groups to the SM vector bosons is complicated by an additional step of Electroweak Symmetry Breaking, which I'll discuss in Section 1.4.1. This is because the gauge bosons introduced by applying the steps in Section 1.2.3 are necessarily massless. The Higgs mechanism (described in Section 1.4) will complete the picture by explaining the origins of mass for the W and Z bosons as well as for the charged fermions they couple with.

1.3 The Standard Model Lagrangian

The Standard Model Lagrangian may be concisely written as:

[2] This connection between mathematics and the physical world was so striking to some eminent physicists (such as Werner Heisenberg, Steven Weinberg, and Abdus Salam [3]) that a kind of folklore emerged within a subsection of the particle physics community arguing that fundamental particles literally *are* irreducible representations of Lie groups – or, rather, that these continuous symmetries are ontologically more fundamental to the universe than the particles themselves. This controversial concept obviously ventures far outside the scope of this thesis and into the territory of philosophy of physics, so I won't dwell on it, but will instead leave the reader with a quote by Eugene Wigner later in life, from his essay *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*:

> *"The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps also to our bafflement, to wide branches of learning."* [4]

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge fields}} + \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{Higgs}}. \tag{23}$$

As the reader will recall from Equation 22, the SM Lagrangian must obey three local gauge symmetries, and each of these symmetries corresponds with a dynamical gauge field introduced due to the formalism demonstrated in Section 1.2.3. We'll name these three fields (by historical convention) B_μ for $U(1)_Y$, W_μ for $SU(2)_L$, and A_μ for $SU(3)_C$. The gauge covariant derivative applied throughout the SM Lagrangian is

$$\mathrm{D}_\mu = \partial_\mu - ig_1 \frac{Y}{2} B_\mu - ig_2 W_\mu^j t^j - ig_3 A_\mu^k T^k. \tag{24}$$

1.3.1 $\mathcal{L}_{\textit{gauge fields}}$

$\mathcal{L}_{\text{gauge fields}}$ refers to the terms associated with the three gauge fields introduced to promote each portion of this overall symmetry from a global to a local symmetry. This term,

$$\mathcal{L}_{\text{gauge fields}} = -\frac{1}{4} \sum_{X\mu} F_{\mu\nu}^j F^{j\mu\nu}, \tag{25}$$

is an expression involving the Standard Model field strength tensor:

$$F_{\mu\nu}^a = \partial_\mu X_\nu^a - \partial_\nu X_\mu^a - g_x f^{abc} X_\mu^b X_\nu^c \tag{26}$$

where X_μ is one of the three gauge fields (B_μ, W_μ, or A_μ) and g_x is the coupling constant (g_1, g_2, or g_3) associated with each gauge field X_μ. The f^{abc} are real, nonzero, and totally antisymmetric under permutations of any two indices for the non-Abelian groups $SU(2)_L$ and $SU(3)_C$, since non-Abelian groups are noncommutative, and are zero for the Abelian group $U(1)_Y$.

1.3.2 $\mathcal{L}_{\textit{fermions}}$

In the covariant derivative (Equation 39), t^j ($j = \{1, 2, 3\}$) and T^k ($k = \{1, \cdots, 8\}$) refer to two bases of Hermitian matrices that obey the Lie algebra

$$[T^a, T^b] = if^{abc} T^c. \tag{27}$$

This means that they are the generators of the gauge groups $SU(2)$ and $SU(3)$. In general, a group $SU(N)$ has dimension N^2-1, as it is generated by traceless Hermitian matrices. $SU(2)$ is therefore three-dimensional and $SU(3)$ is eight-dimensional. Each group has the same number of generators as it has dimensions.

The group $SU(2)$ has three generators: $t^j = -i\frac{\sigma_j}{2}$, where the σ_j are the three Pauli matrices. The group $SU(3)$ has eight generators: $T^k = \frac{\lambda_k}{2}$, where the λ_k are the eight Gell-Mann matrices – the 3×3 $SU(3)$ analogues of the 2×2 Pauli matrices of $SU(2)$.

With these definitions in mind, we can, in principle, expand out all the terms of $\mathcal{L}_{\text{fermions}}$:

$$\mathcal{L}_{\text{fermions}} = -\sum_f \bar{f}\gamma^\mu D_\mu f + \mathcal{L}_{\text{Yukawa}}{}^3 \tag{28}$$

Some of these terms will contain couplings between the fermions and A_μ, the field introduced by the $SU(3)_C$ symmetry from QCD. C here stands for *color*, a metaphorical term given to the label indexing each of the three fields in the triplet representation of each quark in $SU(3)$. QCD is mathematically quite similar to QED, with the exception that QCD is a non-Abelian theory while QED is Abelian. This means that the structure constants f^{abc} are nonzero for QCD, leading to self-interaction terms with A_μ (i.e. the gluons), while they are zero for QED, meaning the photon is not self-interacting. In fact, these theories are so similar that the matrix element of quark-quark scattering needs only a constant correction factor of $\frac{2}{9}$ accounting for the various color options of the quarks involved to modify the QED matrix element for quark-quark scattering.

The remaining terms not involving A_μ can be re-organized according to four different operators – one corresponding to $U(1)_Y$ (i.e. B_μ) and three corresponding to $SU(2)_L$: W^1_μ, W^2_μ, and W^3_μ. We can further group these terms by defining

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu), \tag{29}$$

which, when acting on doublets of particles such as $l = \begin{pmatrix} \nu_e \\ e \end{pmatrix}$ or $q = \begin{pmatrix} u \\ d \end{pmatrix}$, aligns with the electric charge of the gauge boson. However, this arrangement has a clear flaw: it violates

[3] See Section 1.4.3.

parity, i.e. mirror flips across all three spatial dimensions: $\vec{x} \to -\vec{x}$. This is a consequence of the weak force being **chiral** – it couples differently with fermions that are left-handed versus right-handed. "Handedness" here refers to the orientation of the projection of a particle's spin onto its momentum, otherwise known as *helicity*, which in the limit of a massless particle moving at the speed of light is a binary notion called *chirality*. Chirality is analagous to the handedness established by the right-hand rule in physics: just as only the right-hand rule can tell us the direction of a magnetic force on a charged current in a magnetic field, the weak force cannot couple to massless particles with right-handed chirality nor massless antiparticles with left-handed chirality. This is empirically known as well: all observed neutrinos have left-handed chirality, and all observed antineutrinos have right-handed chirality. Mathematically, introducing handedness into the $SU(2)$ gauge symmetry means including the helicity projection operators $P_{L,R} = \frac{1}{2}(1 \mp \gamma^5)$, where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is a combination of the four Dirac matrices. Unfortunately, introducing γ^5 into the Lagrangian yields terms such as $\overline{\psi}\gamma^5\gamma^\mu\partial_\mu\psi$ that violate parity invariance.

The Standard Model explains this chiral behavior of the weak force by asserting that only left-handed fermions transform under $SU(2)_L$, while right-handed fermions are singlets of $SU(2)_L$, meaning they don't transform under this group. However, right-handed fermions should still couple with the photon and the Z boson. To address this, the remaining two operators not used to construct $W\mu^\pm$ – B_μ and W^3_μ – are proposed to *mix together*, parameterized by the Weinberg angle or "weak mixing angle" θ_W, to create the physical eigenstates corresponding to the photon and Z boson, i.e. A_μ and Z_μ:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \tag{30}$$

This methodology of uniting the electromagnetic and weak forces into the *electroweak force* turns out to explain the behavior of the fermions with great precision, down to the quantization of their electric charges and the relationships between them. However, by using different representations of $SU(2) \times U(1)$ for left-handed and right-handed fermions, we eliminate the possibility of including mass terms for fermions. This is on top of the local, chiral gauge symmetry already forbidding masses for both the fermions and the gauge bosons associated with the symmetry itself. The seamless reintroduction of these masses comes due to electroweak symmetry breaking, addressed in Section 1.4.1.

1.3.3 $\mathcal{L}_{Higgs}$

The SM Higgs model contains a doublet of complex scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{31}$$

for which

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}(\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \text{V}(\phi), \tag{32}$$

where $V(\phi)$ is the Higgs potential:

$$\text{V}(\phi) = \mu^2(\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2. \tag{33}$$

ϕ is a doublet because it was designed as a member of the electroweak sector of the SM and therefore should transform under the $SU(2)_L \times U(1)_Y$ local gauge symmetry. Thus, we represent it as a doublet of weak isospin with the top element (the charged field ϕ^+) of weak isospin $\frac{1}{2}$ and the bottom element (the neutral field ϕ^0) of weak isospin $-\frac{1}{2}$. Both fields are necessary because the Higgs mechanism has to explain the mass of the neutral Z^0 boson as well as the charged $W^\pm$ bosons, and as we'll see in Section 1.4.1, the four degrees of freedom introduced by using two complex fields (ϕ_1, ϕ_2, ϕ_3, and ϕ_4) are necessary to the process of electroweak symmetry breaking that will eventually yield four massive bosons (W^+, W^-, Z^0, and H^0).

It's worth noting that Equation 33 is written in terms of μ^2 and λ in order to suggest something like a mass term (where the coefficient of $\phi^\dagger\phi$ is taken to be $-\frac{m^2}{2}$) and a self-interaction term (where the coefficient of $(\phi^\dagger\phi)$ is related to the self-coupling strength). However, we can only properly interpret $\mu^2(\phi^\dagger\phi)$ as a mass term if it has the correct sign – that is, if $\mu^2 < 0$. Keep this in mind as we explore the Higgs potential in more depth in the next section.

1.4 The Higgs Boson

As the most recently-discovered fundamental particle, the Higgs boson is the subject of many current particle physics studies probing every aspect of its behavior. Its discovery was announced on July 4th, 2012, at CERN (my third day of work at the experiment!) to great fanfare.

According to the Standard Model, we expect to be able to create a Higgs boson in a few different ways, each with different likelihoods. Furthermore, the Higgs boson is a short-lived particle, and as soon as it forms, it is expected to quickly decay in one of several different ways, each with different likelihoods. The Standard Model gives clear predictions for the likelihoods of each of these Higgs boson production and decay mechanisms. While the distribution of Higgs boson decays is independent of how each Higgs boson is produced, the frequency of each production mode within a particle accelerator may be calculated based on the energies and types of particles collided.

Before explaining how Higgs bosons are made at the LHC, however, it's important to tackle one further concept: the theoretical prediction that led to the Higgs boson discovery and explains how the massive gauge bosons and fermions in the SM acquired a mass without breaking their respective symmetries: **electroweak symmetry breaking**.

1.4.1 Electroweak Symmetry Breaking

Returning to Equation 33 describing the Higgs potential energy as a function of μ^2 and λ, we will investigate the properties of this potential by looking for a stable ground state (i.e. a minimum point of the potential) and looking at small perturbations around that minimum. However, it's evident that the potential itself will look different depending on the values of μ^2 and λ in its definition. There are two relevant configurations (illustrated in Figure 6) to consider based on the values of these two parameters:

1. $\mu^2 > 0$, $\lambda > 0$: A stable equilibrium at $\phi = 0$.
2. $\mu^2 < 0$, $\lambda > 0$: A multiply-degenerate stable equilibrium at $|\phi| = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}}$.

These situations mark two different configurations of the Higgs field over time: (1) in the first $< 10^{-12}$ seconds of the early universe, when the temperature of space exceeded the

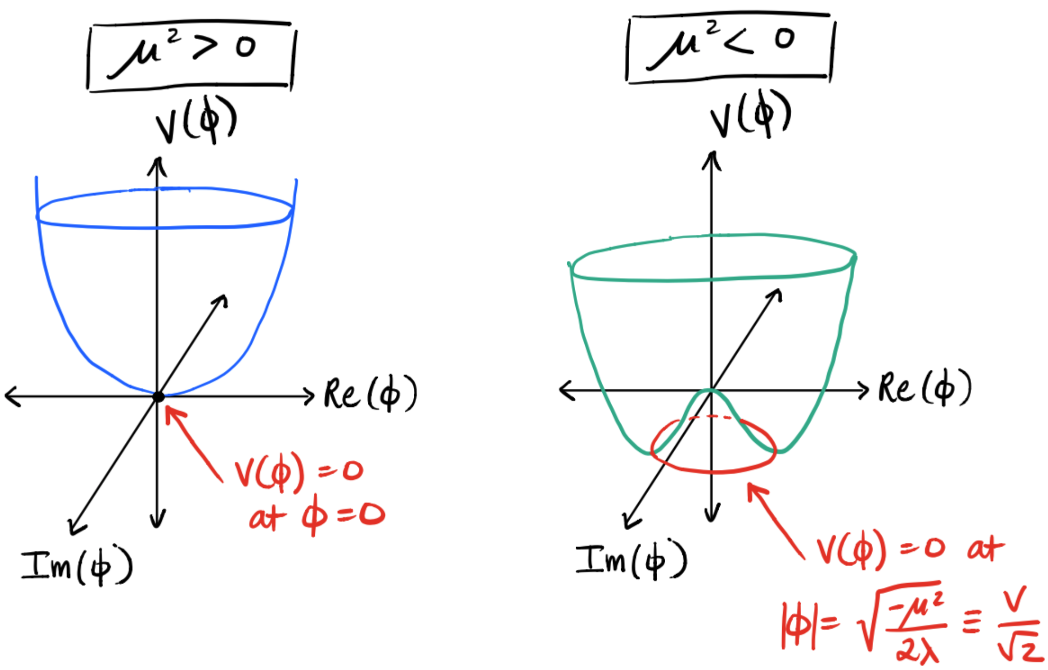


Fig. 6: On the left, the Higgs potential (Equation 33) for a complex scalar field ϕ when $\mu^2 > 0$ and the $SU(2)$ symmetry is preserved. There is one stable equilibrium location at $|\phi| = 0$. This is the configuration of the Higgs field for a tiny fraction of a second following the Big Bang. On the right, the Higgs potential when $\mu^2 < 0$, i.e. when the $SU(2)$ symmetry is spontaneously broken in the ground state. There are infinitely many stable equilibrium locations along the circle $|\phi| = \frac{v}{\sqrt{2}}$ in the complex ϕ plane. This is the configuration of the Higgs field for the remainder of our universe's history.

critical temperature $T_C \approx 160$ GeV[4], and (2) the remainder of the history of our universe, when the universe's temperature cooled below this point. Similarly to how water vapor smoothly condenses into a liquid below a certain temperature, the Higgs field also went through a transition around this point that affected the configuration of its potential. It is at this point in time ($T \leq T_C$) that the process of electroweak symmetry breaking occurred and resulted in the massive bosons and charged fermions in the SM.

Recall from Equation 31 that the Higgs field ϕ transforms as a doublet of $SU(2)$, i.e.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_0 + i\phi_1 \\ \phi_2 + i\phi_3 \end{pmatrix}, \tag{34}$$

where ϕ^+ is electrically charged and ϕ^0 is electrically neutral. Without loss of generality, we can choose to focus on one of the infinitely-many options for the ground state of this potential along the circle $|\phi| = \frac{v}{\sqrt{2}}$ to be $\phi_0 = \phi_1 = \phi_3 = 0$ and $\phi_2 = \frac{v}{\sqrt{2}}$ (a convention called the *unitary gauge*):

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \tag{35}$$

This is a good choice of ground state because it sets ϕ to be nonzero only in its ϕ^0, i.e. electrically neutral, component, which is consistent with experimental measurements that the vacuum of space is electrically neutral.

The act of choosing one of these degenerate ground states for the potential is, in fact, the breaking of the symmetry of the Higgs potential. Just as a pencil stood vertically on its tip will temporarily exhibit a rotational symmetry before tipping over, breaking that symmetry, the necessity of choosing a single ground state out of the infinite degenerate ground states breaks the symmetry of the Higgs potential.

[4]Otherwise known as around 10^{15} degrees Fahrenheit, or about a trillion times the temperature of the surface of our sun.

1.4.2 The Higgs Mechanism

The Higgs mechanism[5] refers to the process sparked by spontaneously breaking the symmetry of the Higgs potential that results in massive $W^{\pm}$ and Z^0 bosons.

Now that we have chosen a minimum point and broken a symmetry, we can introduce small fluctuations about the equilibrium by adding a scalar field $h(x)$ that corresponds to excitations of the Higgs field, i.e the Higgs boson itself. The Higgs field can now be written as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + h(x)) \end{pmatrix}. \tag{36}$$

Plugging this expressing for ϕ back into Equation 33, we will find that a mass term has emerged for the new field $h(x)$:

$$V(\phi) = \frac{1}{2}\mu^2 h(x)^2 + \cdots , \tag{37}$$

suggesting that the mass of this field (i.e. the mass of the Higgs boson particle, the excitation of the Higgs field) is

$$m_h = \sqrt{-\mu^2} = \sqrt{2\lambda v^2}. \tag{38}$$

We can also use our new expression for ϕ with the gauge covariant derivative corresponding to the electroweak sector:

$$\mathrm{D}_\mu = \partial_\mu - ig_1 \frac{Y}{2} B_\mu - ig_2 W_\mu^j t^j, \tag{39}$$

and look just at the mass terms emerging from the quantity $|D_\mu \phi|^2$, noting that since the Higgs doublet in Equation 31 has "down-type" weak isospin $-\frac{1}{2}$ and no electric charge, $Y = 2(Q - I_W) = 1$. We will find that these gauge boson mass terms can be concisely written as

[5] You can thank me for not referring to this concept the way Peter Higgs himself sometimes does to acknowledge the other physicists who contributed to its formulation: "the ABEGHHK'tH mechanism," for Anderson, Brout, Englert, Guralnik, Hagen, Higgs, Kibble, and 't Hooft.

$$\frac{v^2}{8}\begin{pmatrix} W^1 & W^2 & W^3 & B^0 \end{pmatrix}^* \begin{pmatrix} g_2^2 & & & \\ & g_2^2 & & \\ & & g_2^2 & -g_1g_2 \\ & & -g_1g_2 & g_1^2 \end{pmatrix} \begin{pmatrix} W^1 \\ W^2 \\ W^3 \\ B^0 \end{pmatrix}. \tag{40}$$

Since the W^1 and W^2 fields can be combined to form the $W^{\pm}$ bosons (Equation 29), we can conclude that

$$m_W = \frac{1}{2}g_2 v. \tag{41}$$

However, the mass terms for W^3 and B^0 are mixed via the matrix

$$\frac{v^2}{8}\begin{pmatrix} g_2^2 & -g_1g_2 \\ -g_1g_2 & g_1^2 \end{pmatrix}. \tag{42}$$

To see how these masses will translate to the physical Z_μ and A_μ fields corresponding to the Z^0 boson and photon (γ^0), we can rewrite these fields in terms of the Weinberg mixing angle from Equation 30:

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{43}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W, \tag{44}$$

and choose a new basis that diagonalizes the matrix from Equation 42 such that we can identify terms that look like

$$\frac{1}{2}\begin{pmatrix} A_\mu & Z_\mu \end{pmatrix}\begin{pmatrix} m_A^2 & 0 \\ 0 & m_Z^2 \end{pmatrix}\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix}. \tag{45}$$

This matrix in diagonalized form is:

$$\frac{v^2}{8}\begin{pmatrix} 0 & 0 \\ 0 & g_1^2 + g_2^2 \end{pmatrix}, \tag{46}$$

from which we can directly read off the masses of the A_μ and Z_μ:

$$m_A = 0 \text{ and } m_Z = \frac{v}{2}\sqrt{g_1^2 + g_2^2}. \tag{47}$$

Amazingly, the Higgs mechanism has broken the electroweak symmetry and resulted in three massive bosons (W^+, W^-, and Z^0) and one massless boson (A_μ, i.e. the photon). How can we understand this result in the context of the process shown in Section 1.2.3 that outlined clear instructions for achieving local gauge invariance only with the introduction of a *massless* gauge field? The Lagrangian itself remains unchanged by the Higgs mechanism, i.e. we are still talking about the same physical system. However, we have rewritten the field ϕ in terms of fluctuations around a minimum ground state and chosen an explicit form for it in a convenient gauge. We should remember, too, that gauge symmetries are not physical symmetries of spacetime, but rather *redundancies* in the mathematical description of a system. Fixing a gauge by choosing a particular ground state doesn't have physical repercussions, but it does allow us to learn about the mass and interaction terms for the Higgs field. Physicists sometimes explain the appearance of masses for the three new gauge field as well as the scalar Higgs field as these fields "eating" the massless Goldstone bosons that would have emerged in our calculations had we not broken the symmetry by choosing a particular ground state.

Note that the massive gauge boson masses are related: $m_W/m_Z = \cos\theta_W$. The verification of this relationship between the $W^\pm$ and Z masses and the cosine of the Weinberg angle was one of the important steps of validating the Higgs mechanism that led particle physicists to eagerly seek out the detection of the Higgs boson. Based on measurements of m_W and g_2, we can also approximate the value of v as 246 GeV/c^2. Physicists refer to v as the **electroweak scale**. Once the values for v and the Higgs boson mass $m_h \approx 125$ GeV/c^2 were determined, the value of the Higgs self-coupling constant λ was also fixed to be ≈ 0.13.

In addition to the gauge boson mass terms (v^2VV, where $V = \{W^\pm, Z^0\}$), the new electroweak Lagrangian will also contain terms proportional to hVV and $hhVV$, corresponding to the triple and quadruple vertexes between two V bosons and one or two Higgs bosons h. The coefficients of these terms can be used to calculate the expected coupling strength between the Higgs bosons and the V bosons. In each case, the coupling strength is proportional to the mass of the V boson.

1.4.3 Fermion masses

Just as mass terms for the gauge bosons are explicitly forbidden under the electroweak $SU(2)_L \times U(1)_Y$ symmetry, fermion mass terms are also forbidden under the same symmetry. Introducing the doublet of complex scalars $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ from Equation 31 also allows us to rewrite the Lagrangian with new terms corresponding to fermion mass terms ($\mathcal{L}_{\text{Yukawa}}$) that do obey the $SU(2)_L \times U(1)_Y$ symmetry. For example, the mass term for the tau lepton would take the form:

$$\mathcal{L}_\tau = -g_\tau \left[\begin{pmatrix} \overline{\nu}_\tau & \bar{\tau} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tau_R + \bar{\tau}_R \begin{pmatrix} \phi^- & \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \right] \tag{48}$$

By choosing the same gauge as before for ϕ that only keeps ϕ^0 nonzero, the mass terms for the neutrinos disappear, leaving only mass terms for the charged leptons. When acting on doublets of quarks, however, this choice would also remove mass terms for the up-type quarks (u, c, t), which is problematic. We can fix this by conjuring another Higgs doublet

$$\phi_c = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix}, \tag{49}$$

called its *conjugate*, that transforms just as ϕ does. Writing ϕ in this way results in additional mass terms for the up-type quarks.

The Yukawa coupling parameters g_f in the fermion mass terms are free parameters in the Standard Model, meaning the theory has no preference for their values – they must be experimentally measured. However, they all follow the pattern

$$g_f = \sqrt{2}\frac{m_f}{v}, \tag{50}$$

meaning they scale with the masses of the fermions in question. These Yukawa couplings vary from $g_e \approx 3 \times 10^{-6}$ (smallest) to $g_t \approx 0.995$ (largest).

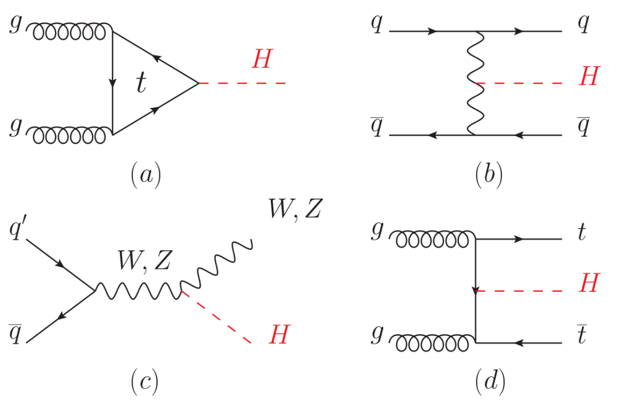


Fig. 7: Higgs boson production mechanisms. [5]

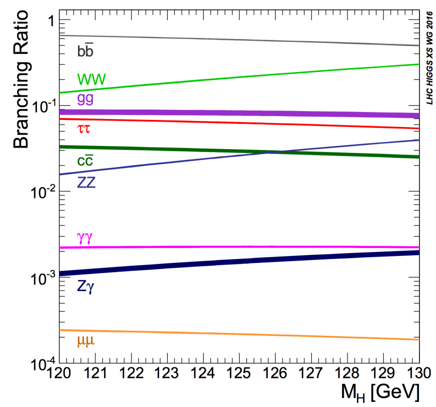


Fig. 8: Higgs boson branching ratios. [5]

1.4.4 Production Mechanisms

Higgs boson events at the LHC, which primarily collides two protons at center-of-mass energies of around $\sqrt{s} = 13$ TeV, will mostly feature four main Higgs boson production processes, each described in Table 4. 88% of Higgs bosons produced at the LHC at $\sqrt{s} = 13$ TeV come from gluon-gluon fusion via a quark loop (Figure 7-a). Given that the Higgs boson couples most strongly to the heaviest quarks, this loop typically involves top or bottom quarks. The remaining 12% of Higgs bosons are primarily generated from vector boson fusion (Figure 7-b), associated production with a vector boson, i.e. a W or Z boson (Figure 7-c), and associated production with a pair of top quarks (Figure 7-d).

Production Mode	Symbol	% of Total Higgs Boson Production
Gluon-Gluon Fusion	ggF	88.2%
Vector Boson Fusion	VBF	6.86%
Associated Production with a Vector Boson	VH	4.08%
Associated Production with 2 Top Quarks	ttH	0.91%

Table 4: Predicted frequencies of Higgs boson production mechanisms at the Large Hadron Collider in proton-proton collisions at $\sqrt{s} = 13$ TeV. Percentages reflect the ratio of expected cross-section to total Higgs boson cross-section according to the 2018 PDG review [5].

1.4.5 Decay Mechanisms

The Standard Model also predicts the relative frequency of Higgs boson decays into various collections of lighter fundamental particles as a function of the Higgs boson mass. These decays occur nearly instantaneously with Higgs boson productions, as the lifetime of the Higgs boson is only about 10^{-23} seconds, meaning that a Higgs boson will traverse a distance orders of magnitude smaller than the radius of a proton before it decays. The ATLAS and CMS detectors at the LHC can only begin tracking particles a couple of centimeters away from the central beam line, meaning they will only ever detect the decay products of each Higgs boson. The majority of 125 GeV Higgs bosons are expected to decay into a bottom-antibottom quark pair, followed by (in order of descending likelihood) two W bosons, two gluons, and a tau-antitau pair. The branching ratios of the main decay channels for a 125 GeV Higgs boson are shown in Figure 8 and detailed in Table 5.

Decay Mode	Symbol	% of Total Higgs Boson Decays
Bottom Quark Pair	$b\bar{b}$	57.1%
W Boson Pair	$W\bar{W}$	22.0%
Gluon Pair	gg	8.53%
Tau Lepton Pair	$\tau\bar{\tau}$	6.26%
Charm Quark Pair	$c\bar{c}$	2.88%
Z Boson Pair	ZZ	2.73%
Photon Pair	$\gamma\gamma$	0.23%
Other	$-$	0.27%

Table 5: Decay channels for a 125 GeV Higgs boson [6]

2 The ATLAS Experiment at CERN

Physicists love a good acronym, and sometimes go to great lengths to compress their experiment names into impressive abbreviations. This thesis represents an analysis of data collected at one of the best examples of this phenomemon: the ATLAS Experiment at CERN, where ATLAS stands for "**A** **T**oroidal **L**HC **A**pparatu**S**."

Two more acronyms are relevant to undertanding the context for the ATLAS Experiment: CERN, or *Conseil Européen pour la Recherche Nucléaire* (more commonly referred to in English as the European Organization for Nuclear Research), and the LHC, or Large Hadron Collider.

2.1 CERN

CERN refers to the organization and physical site of the largest particle physics laboratory in the world, located just outside of Geneva, Switzerland (see Figure 9). CERN is a massive international research organization comprising nearly 18,000 employees representing dozens of countries. Founded in 1954, CERN was established in the wake of World War II as an explicit endeavor to promote peace through international collaboration and scientific discovery for the public good. In fact, its charter proclaims:

> "The Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character, and in research essentially related thereto. The Organization shall have no concern with work for military requirements and the results of its experimental and theoretical work shall be published or otherwise made generally available." [7]

Over its nearly 70-year history, CERN has been the site of several particle accelerators of increasing size and power, beginning in 1957 with the 600 MeV Synchrocylotron [8]. Most activities at CERN currently involve the Large Hadron Collider (LHC), described in the next section. CERN is also notable as the birthplace of the World Wide Web, which was invented by Tim Berners-Lee in 1989 in order to facilitate sharing information with other physicists around the world collaborating on experiments at CERN [9].

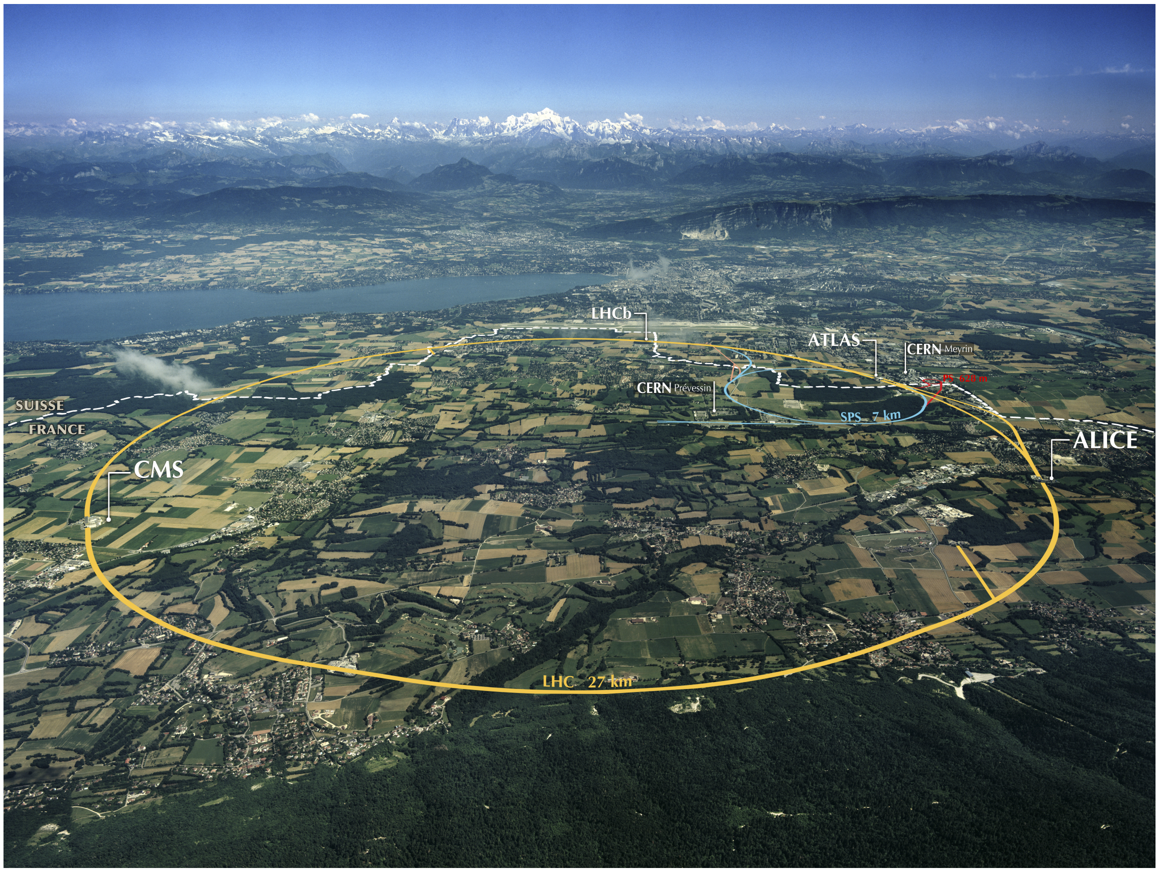


Fig. 9: The campus of CERN, the largest particle physics laboratory in the world, is located outside of Geneva, Switzerland. Here, we can see the city of Geneva in the distance, bordering Lake Geneva, under a canopy of the Alps, including Mont Blanc. Outlined in yellow is the circumference of the Large Hadron Collider (LHC), a circular tunnel that lies underground. Labels along the circumference of the LHC indicate the locations of the major experiments that study the LHC's collisions. [10]

2.2 The Large Hadron Collider

Historically, experimental particle physics has made many of its largest scientific strides with the aid of *particle colliders*. Of course, fundamental particles are all around us, but the particles we interact with on an everyday basis are only a fraction of the particles predicted by the Standard Model. Experimental particle physicists are often interested in studying particles that we rarely see on Earth. Even if we were able to produce these rare particles easily, their quantum mechanical natures would require us to make frequent measurements to collect sufficient data on their many possible interactions. However, producing these rarer particles with any regularity is not easy, particularly given that the LHC uses protons, which are not fundamental particles at all (see Section 2.2.2). Additionally, since the Standard Model makes statistical predictions, statistically-significant numbers of observations are needed for compelling experimental results. These goals have led the field to construct elaborate machinery in order to repeatedly produce and measure high-energy particle collisions in order to observe rare particle processes in a controlled manner.

High-energy collisions are a worthwhile target because, in general, the higher the energy of the particle collision, the more massive the particles that may be produced in the aftermath of that collision. As the reader might recall from the discussion of the Standard Model particles in Section 1.1, some of our most familiar particles (electrons, up quarks, and down quarks, say) are also among the least massive particles. More massive particles are often rarer and therefore more difficult to produce and study. Higher energies also correlate with extremely short length scales. Recalling our discussion about the necessity of QFT to study quantum mechanical behaviours at high energies (Section 1.2), we found that measuring small distances (such as lengths smaller than the Compton wavelength of a particle, $\lambda_C = \frac{\hbar}{mc}$) requires larger and larger amounts of energy. One can therefore think of a particle collider as a kind of superpowered microscope: by producing large amounts of energy, it allows us to measure properties of particles and forces at extremely small length scales.

The **Large Hadron Collider** (LHC) is currently the largest particle collider on Earth. By some estimates, it is among the largest and most sophisticated machines ever constructed by humans. Located around 300 feet underground below the site of the CERN campus outside of Geneva, it takes the form of a massive circular tunnel that stretches 17 miles in circumference. It also creates the highest-energy particle collisions ever produced by humans. Most frequently, the particle beams in the LHC are made up of protons, though the "hadron" in its name is a more general term referring to a particle composed of three quarks. Its circular shape is motivated by the goal of facilitating high-energy particle collisions: just as two trains colliding head-on at full speed would produce a larger crash than a single train hitting a wall, two particles colliding head-on from opposite directions will make a larger burst of energy than a single particle beam hitting a static target. However, we must remember that we're not colliding trains – we're colliding some of the smallest objects in the universe. The procedure to set up a particle collision therefore requires a staggering amount of precision. The circular tunnel configuration of the LHC allows us to recycle particles that did not manage to collide on previous passes around the ring. When this occurs, the particles are guided around the circular tunnel to the next interaction point and we re-try the collision. This maximizes the amount of experimental collision data we can observe.

Of course, the process of producing a high-energy particle collision is anything but simple. The design of the LHC is the result of decades of invention and reinvention of the concept of a particle collider since the end of World War II. Particle physicists at CERN are hardly the only beneficiaries of this technological innovation – in fact, particle accelerators have led to breakthrough cancer treatments such as proton therapy, advanced cargo screening for homeland security, and even more efficient ways to produce safe, durable packaging for food products. In the next sections, however, I'll focus on detailing the operation of the LHC for its intended purpose: colliding particles at very high speeds. Then, in Section 2.3, I'll discuss the ATLAS detector that measures the outputs of these particle collisions.

2.2.1 Definitions & Units

Measuring the properties of some of the smallest objects in the universe has led to a unique set of measurement units. In particular, the energies of fundamental particles are historically reported in units of **electron volts (eV)**, i.e. the amount of energy an electron gains when it is accelerated through an electrical potential difference of 1 Volt. One electron volt corresponds to 1.6×10^{-19} Joules, where Joules are a unit often used to describe an energy level comparable to that of a household light bulb. For particle physics at the LHC, we will frequently encounter energy levels in terms of millions, billions, or trillions of electron volts (MeV, GeV, and TeV, respectively). While these prefixes might suggest extremely high energies, it's important to keep in mind that these energies are still far smaller than 1 Joule, so they are still small by most human standards.

Circular colliders like the LHC are categorized by their center-of-mass energy E_{cm}, also known as $\sqrt{s}$. This Lorentz-invariant quantity summarizes the combined energy & momentum of the two particles colliding head-on. If two particles collide with equal and opposite momenta, their relativistic four-vectors will look like $p_1 = (E_1, \vec{p})$ and $p_2 = (E_2, -\vec{p})$. One Lorentz-invariant quantity combining these two four-vectors is s:

$$
\begin{aligned}
(p_1 + p_2)^2 &= (E_1 + E_2)^2 - (\vec{p} + (-\vec{p}))^2 && (51) \\
&= (E_1 + E_2)^2 && (52) \\
&= (E_{\text{center of mass}})^2 \equiv s && (53)
\end{aligned}
$$

In its most recent years of operation, the LHC accelerated proton beams with energies $E_{cm} = \sqrt{s} = 6.5$ TeV each in order to produce collisions with a combined center-of-mass energy of up to 13 TeV.

Another key metric for understanding the power of particle colliders is **instantaneous luminosity** ($\mathcal{L}$). While the center-of-mass energy $E_{cm} = \sqrt{s}$ contains information about which particles that collider might be able to produce, instantaneous luminosity contains information about how frequently the collider will be able to produce the particles in question. At the LHC, physicists refer to the collision of two particles as an *event*. The instantaneous luminosity of the LHC tells us how frequently we can observe events within the collider. Relatedly, a quantity called **integrated luminosity** (L) tells us how many events we potentially observed over a given period of time.

The instantaneous luminosity $\mathcal{L}$ of the LHC scales with the frequency of collisions within the experiment (f), but it requires a little more information than the collision frequency alone. Rather than collide a single particle with a single other particle, which is prohibitively difficult to consistently accomplish, at the LHC we collide tightly-packed bunches of particles with other bunches of particles. This increases the likelihood of collisions. The instantaneous luminosity therefore also depends on the number of particles in each bunch (n_1 and n_2) as well as on the distribution of particles within each bunch. If we assume that the particles in each bunch resemble a 2D Gaussian distribution along their axes perpendicular to the beamline, then we can write the instantaneous luminosity as a function of f, n_1, n_2, and the beam standard deviations in the $\hat{x}$ and $\hat{y}$ directions σ_x and σ_y:

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{54}$$

This quantity has units of $\frac{1}{\text{time}\cdot\text{area}}$. If we are interested in counting the number of events at the LHC over a given period of time, we can use the integrated luminosity L, which has units of inverse area, and **cross section** (σ), which has units of area:

$$\text{Number of events} = \sigma \int \mathcal{L}(t)\ \mathrm{d}t. \tag{55}$$

The cross section is of crucial importance for particle physicists, as it is our metric for understanding the likelihood of certain particle processes occurring. A very rare process will have a small cross section, while a common process will have a large cross section. Clearly, then, in order to observe a very rare process at the LHC multiple times, it is essential for the beam design to result in a large enough $\mathcal{L}$ (or at least a long enough time frame t) that can counteract the smallness of the σ in question.

The convention of using a cross section to describe the likelihood of events is a historical one, and in the context of particle beams colliding, it is somewhat unintuitive at first. It originated from earlier particle collider designs that used a single beam of particles colliding with a fixed target material. In a fixed-target experiment, one can think of particle processes as different-sized targets: it's easy for a particle to hit a large target, but difficult to hit a tiny target. Cross-sections, with units of surface area, are therefore a useful way of thinking about relative likelihoods of particle processes. This picture is more complicated in collider experiments like the LHC, where each bunch of protons is simultaneously the particle beam and the (moving) target material, but we still use the same convention.

An additional historical quirk of particle collider notation is the unit of area used to report cross sections and luminosities: the **barn**[6]. One barn is 10^{-28} m^2, comparable to the size of a uranium nucleus. In the context of luminosities, which are measured in units of inverse area, the relevant unit becomes inverse barns, usually with various size prefixes

[6]As with other weird terminology in particle physics, this unit was coined humorously. In this case, it originated from physicists working on the Manhattan Project in 1942 who needed to invent a secret word to conceal the nature of their calculations. Apparently, one of the physicists had a rural upbringing, and he suggested that for an atomic nucleus, a target of this size would be as big as a barn. [11]

in scientific notation. The most common luminosity units seen at the LHC are "inverse picobarns" and "inverse femtobarns" to refer to 10^{-40} m^2 and 10^{-43} m^2, respectively.

2.2.2 Proton PDFs

The protons circulating in the LHC are, crucially, *not* fundamental particles. This means that a proton has constituent parts, i.e. quarks and gluons, that are collectively called *partons*. Each proton's quantum numbers are defined by its three *valence quarks*, two up quarks and one down quark, with gluons mediating their interactions. These gluon exchanges have significant effects on the dynamic interactions within the proton, and must be accounted for to effectively understand the resulting particle collisions following the intersection of two proton beams. Physicists therefore use **parton distribution functions**, or PDFs, to aggregate information about the likely proton dynamics as a function of $x \in [0, 1]$, where x represents the fraction of the proton's longitudinal momentum carried by each parton, and Lorentz-invariant energy scale Q^2. These PDFs cannot be calculated *a priori* with the usual methods of perturbation theory because the coupling constant defining parton interactions, $\alpha_S \approx 1$, is far too large. Particle physicists have therefore constructed proton PDFs from experimental measurements using external particles as probes.

Example proton PDFs at different values of Q^2 are shown in Figure 10. At low values of Q^2, when the proton has lower energy, more of the fraction of its overall momentum tends to be carried by its valence quarks. The limit of $x = 1$ represents the fully-elastic scenario in which a given quark carries all of the momentum of the proton, and therefore the proton will behave as a single point-like particle. As Q^2 increases, however, it becomes more and more likely that the proton will contain a number of quark-antiquark pairs, called *sea quarks*, generated by gluons. The overall fraction x of proton momentum shouldered by a given valence quark therefore decreases, as the momentum is distributed across the additional sea quarks and gluons.

2.2.3 Synchrotron Mechanics

The LHC's goal is to collide protons (or, more precisely, the partons contained within protons), but in order to get to the point of collision, it must first accelerate those protons

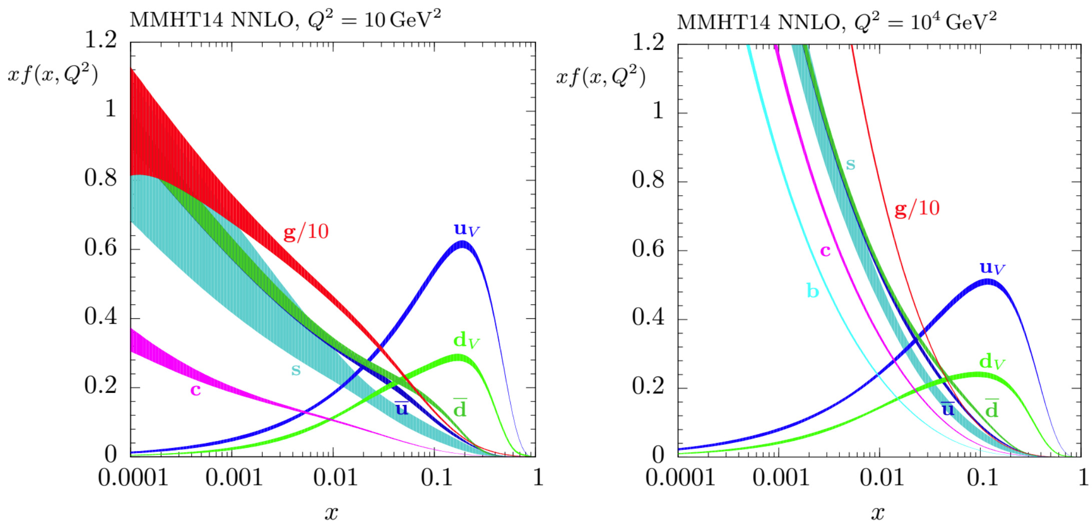


Fig. 10: A global analysis of proton PDFs shows different inner dynamics of the proton at lower energy scales ($Q^2 = 10$ GeV^2, on the left) versus higher energy scales ($Q^2 = 10,000$ GeV^2, on the right). The valence quarks u_V and d_V are shown in dark blue and lime green, while gluons are shown in red, and the remaining colors correspond to sea quarks of various flavors. The thickness of the colored bands corresponds to 68% confidence levels. The probability of finding primarily valence quarks within the proton increases at first with x and peaks around $x = \frac{1}{3}$, indicating that each of the three quarks carries an equal fraction of the proton's longitudinal momentum, though the probability of any given valence quark carrying all of a proton's momentum is small. As Q^2 increases, the proton is more likely to have its longitudinal momentum carried significantly by gluons and sea quarks. [12]

to high speeds. In particular, the LHC is a type of particle accelerator called a *synchrotron*. This name derives from the concept that the magnetic fields of the magnets within the accelerator change simultaneously as the particles contained within on a closed-loop path increase in speed.

The protons themselves are obtained from stores of hydrogen gas. Applying an electromagnetic field to the gas strips away the electrons in the gas, leaving the bare protons. These protons are then fed into a series of linear and circular accelerators from the past decades of CERN's operation that gradually increase their speed: the Linac 2, the PS Booster, the Proton Synchrotron, the Super Proton Synchrotron, and finally, the largest ring: the LHC itself (see Figure 11).

Synchrotrons accelerate charged protons using powerful radio-frequency (RF) systems in resonant metallic cavities that serve as acceleration points for the particles circling the

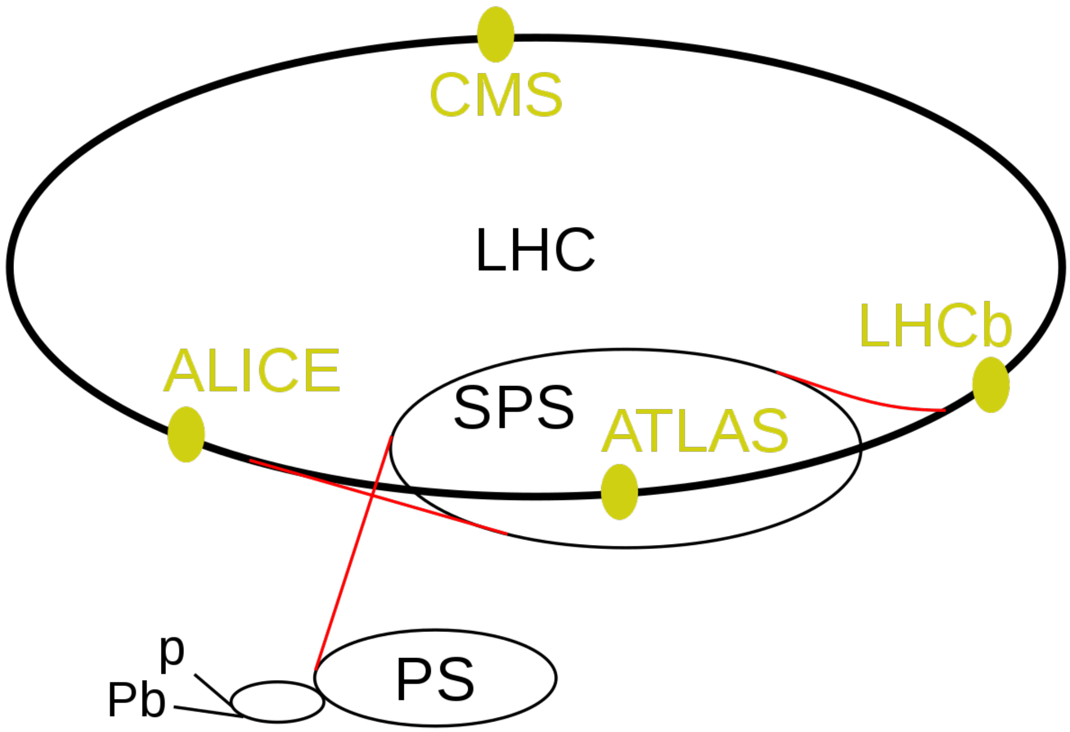


Fig. 11: This schematic of the LHC shows the path of protons (labeled "p" in the bottom left) as they move through a series of linear and circular accelerators before finally reaching the LHC's main ring, where they receive their final bursts of energy from the RF cavities in order to reach their maximum speed. The four main LHC experiments (in yellow) show the locations of beam crossings and therefore particle collisions.

synchrotron ring. The LHC has 16 RF cavities containing time-varying electromagnetic fields that oscillate at 400 MHz, yielding a maximum voltage of 2 MV each. The timing of these oscillations allows for protons passing through the time-dependent field to accelerate due to an incremental transfer of energy. The RF cavities also result in the bunched structure of the proton beam, as any protons with slightly higher or lower energies than the target energy will arrive in the cavity at different times and will accelerate or decelerate accordingly.

While the RF cavities accelerate the protons forward, the particles would move linearly if not for very strong (8 Tesla, or more than 100,000 times more powerful than the Earth's own magnetic field) dipole magnets placed along the circumference of the ring designed to bend the particles in a circle. The strength of these dipole magnets is essential to turning the proton bunches around the LHC's circumference, as weaker magnets would necessitate a much larger circular tunnel structure. Each dipole magnet produces two immensely strong opposing magnetic fields nestled closely together in order to steer each of the two paths of proton bunches in opposite directions along the LHC's perimeter.

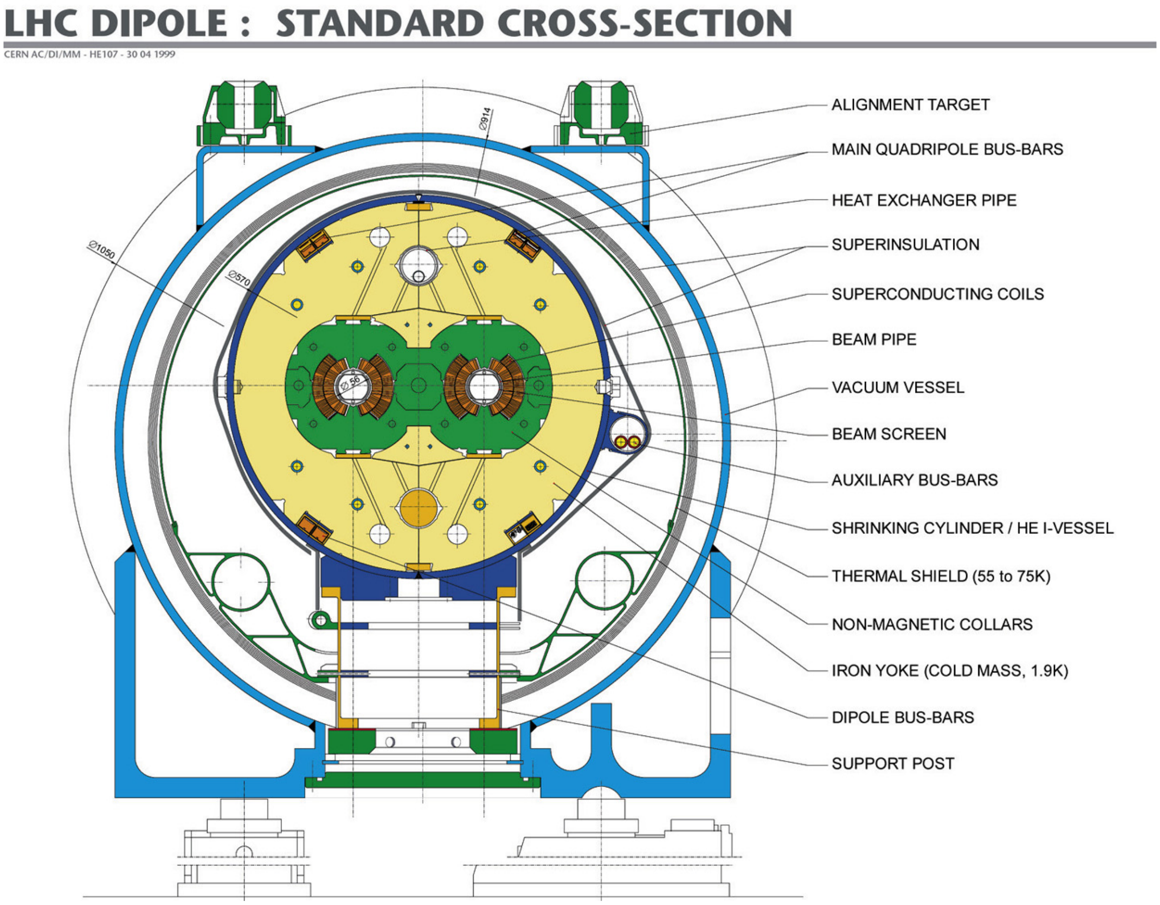


Fig. 12: A cross-sectional view of a dipole magnet at the LHC showing the two beam pipes surrounded by superconducting coils and an iron yoke cooled to a temperature of around 2 Kelvin – colder than the average temperature of empty space.

These magnets operate on the principle of the Lorentz force $\vec{F}$ on a charged particle q moving at velocity $\vec{v}$ through a magnetic field $\vec{B}$ and electric field $\vec{E}$:

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \tag{56}$$

When a nonzero magnetic field is present, the cross-product in this formula causes a charged particle's trajectory to bend. This same principle is also important for the beam-focusing quadrupole magnets placed at the four main LHC particle detectors that are responsible for squeezing the opposite-moving beams even more closely together in order to cross and maximize the likelihood of the proton beams intersecting.

The ability of synchrotrons to increase the energies of the charged particles within is limited by the effect of *synchrotron radiation*, a kind of electromagnetic radiation emitted by charged particles when they experience an acceleration perpendicular to their velocity (i.e. when they curve). While synchrotron radiation is sometimes a desired effect in a laboratory setting, at the LHC, it is an undesirable loss of energy for the protons we are so actively trying to accelerate. However, the amount of synchrotron radiation emitted by protons is far less than what would be emitted in the same context by electrons because protons have significantly larger masses and therefore don't accelerate as much as electrons would with the same energy transfer by the RF cavities.

2.2.4 Operation

At the LHC, the frequency of collisions is $f = 40$ MHz, i.e. 40 million times per second, with new bunches of protons colliding every 25 nanoseconds. This means that approximately 3,000 bunches of protons can, in theory, fit along the circumference of the LHC ring during any given run. In practice, several of these bunches are left empty (the so-called "abort gap") to allow for safe beam dumps during a run. Each bunch travels at a speed very close to the speed of light, looping over 11,000 times around the entire circumference *every second*. At the start of each datataking run, there are approximately 120 billion protons tightly compacted together within each bunch. While the energy of each proton within each bunch is small by human standards – 6.5 TeV, or around one millionth of a Joule – together, all the protons within all the bunches within an LHC beam combine to an energy of around 350 million Joules, which is more comparable to the kinetic energy of an aircraft carrier moving at 150 mph. It's all the more remarkable, then, that that massive amount of energy is squeezed into such a tiny area at the collision point: less than 100 μm^2, or around the width of a human hair.

This thesis examines data from the entirety of Run 2 datataking at the LHC. As shown in the long-term LHC schedule (Figure 13), Run 2 extended from the years 2015 through 2018. The first datataking run at the LHC, Run 1, took place from 2011 - early 2013.

Following a long shutdown period from 2019 - 2020, Run 3 of datataking was scheduled to commence in 2021, but this timeline was extended due to the COVID-19 pandemic, and Run 3 is now slated to commence in March 2022 [13].

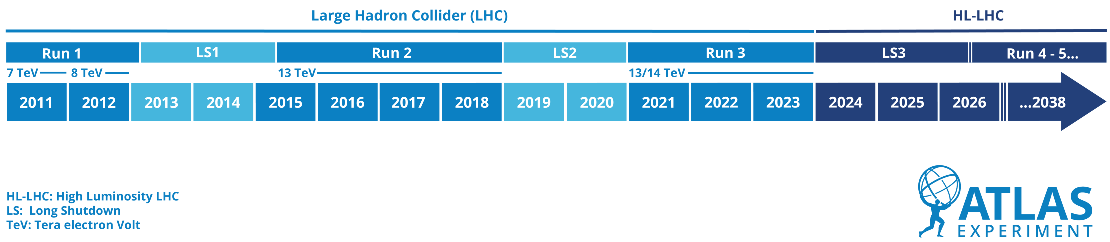


Fig. 13: The long-term schedule of LHC datataking shows Run 1 (2011-2013), Run 2 (2015-2018), and upcoming Run 3 (2022-2024, pushed back approximately 1 year due to COVID-19 pandemic). The High-Luminosity LHC, or HL-LHC, is shown on the right in dark blue. It is scheduled to begin in 2024 and run for approximately a decade.

Run 1 at the LHC operated at 7 and 8 TeV center-of-mass energy and produced about 30 fb^{-1} of integrated luminosity. Run 2 operated at a higher center-of-mass energy of 13 TeV and produced about 150 fb^{-1} of integrated luminosity (see Figure 14). The High-Luminosity LHC, or HL-LHC, is planned to operate at the same energy as Run 3 (designed for 14 TeV), but a much higher instantaneous luminosity in order to produce potentially more than 3,000 fb^{-1} of integrated luminosity across its lifetime of approximately a decade, starting in 2024.

2.3 The ATLAS Detector

At 44 meters long and 25 meters tall [15], the gargantuan ATLAS detector has been recording collisions within the Large Hadron Collider (LHC) since late 2009 [16]. It was designed as a general purpose detector, meant to seek out any and all new physics phenomena accessible at the unprecedented energy scales of the LHC. Roughly cylindrical in shape, the detector is aligned with the beamline of the LHC.

2.3.1 Coordinate System

The center of the ATLAS detector, i.e. the interaction point of the colliding proton beams, is defined as the origin for the experiment's common coordinate system. The $\hat{z}$ axis

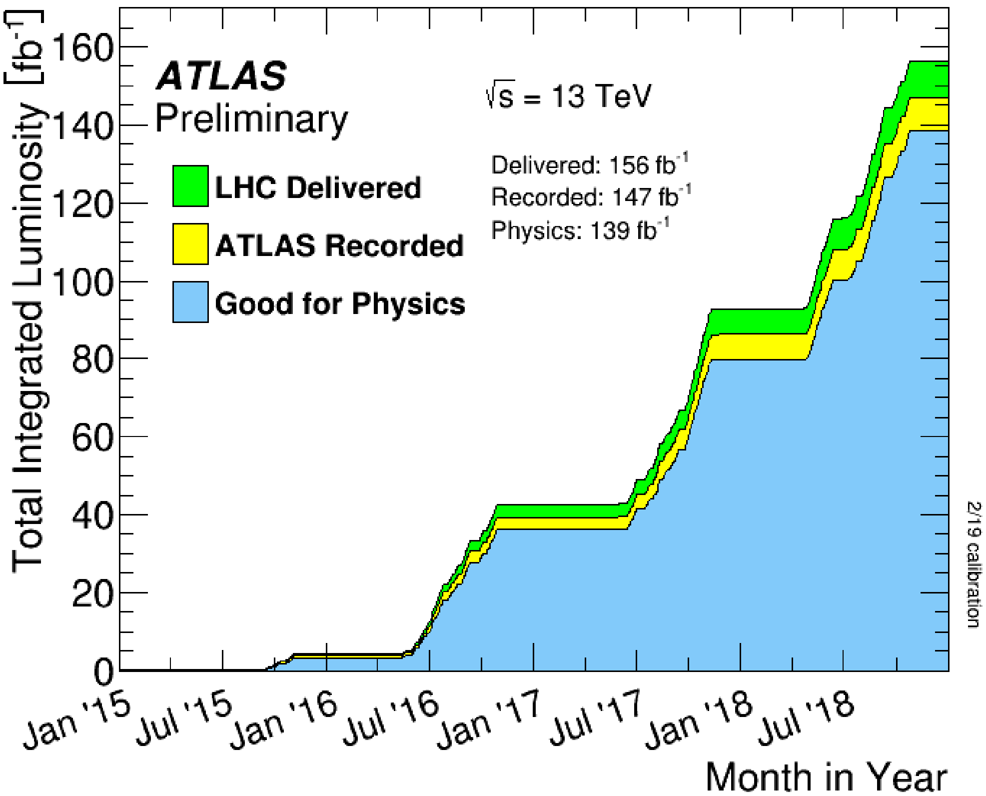


Fig. 14: The cumulative integrated luminosity collected at the LHC over time during Run 2 shows a total amount of data delivered by the LHC to the ATLAS Experiment of 156 fb^{-1} (in green). The vast majority of this data is recorded by ATLAS (in yellow), but some data is lost at this stage due to inefficiencies in the data acquisition process and the need to ramp up the voltages of the tracking detectors and pixel system preamplifiers before recording data. Finally, "good for physics" (in blue) denotes data where all physics objects have been reconstructed with good data quality. [14]

runs along the beamline of the LHC and through the center of the detector. The $+\hat{x}$ axis points from the interaction point to the center of the LHC, while the $+\hat{y}$ axis points up from the interaction point to the surface of the Earth. The $\hat{x}-\hat{y}$ plane is therefore perpendicular to the beamline.

Rather than using $\hat{x}$ and $\hat{y}$ to describe positions in the transverse plane, the ATLAS experiment uses the coordinates ϕ and η. ϕ represents the azimuthal angle around the beamline. η, a Lorentz-invariant quantity also known as *pseudorapidity*, is defined in terms of the polar angle θ:

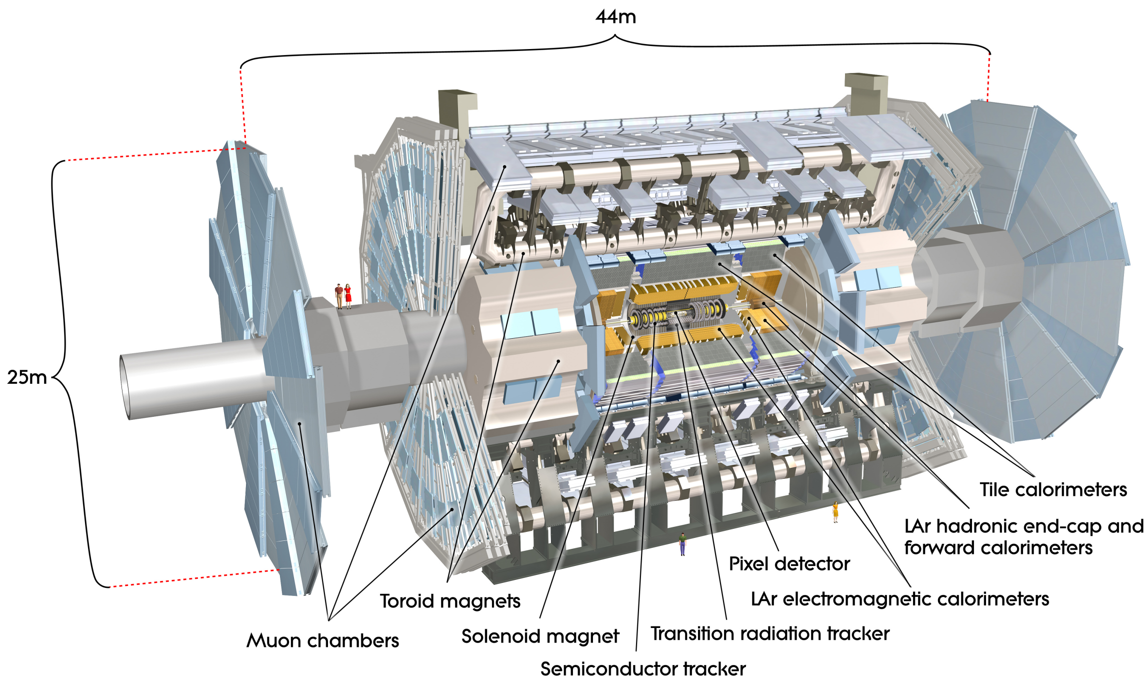


Fig. 15: The ATLAS detector, with two human figures shown standing near the leftmost muon chambers to convey a sense of scale.

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \tag{57}$$

Radial distance from the interaction point is commonly reported in terms of ΔR:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \tag{58}$$

2.3.2 Inner Detector

The innermost layers of the detector surrounding the beamline constitute the mechanisms for tracking charged particles emanating from the collision site. This Inner Detector region is also placed within a magnet system that supplies a magnetic field parallel to the beamline of approximately 2 Tesla. Closest to the beamline lies the Pixel Detector, densely packed with roughly 80,000,000 silicon pixels to ensure high-granularity tracking close to the event [18]. The Pixel Detector was augmented in 2014 with the Insertable B-Layer (IBL), which further improves reconstruction of the impact parameter while providing

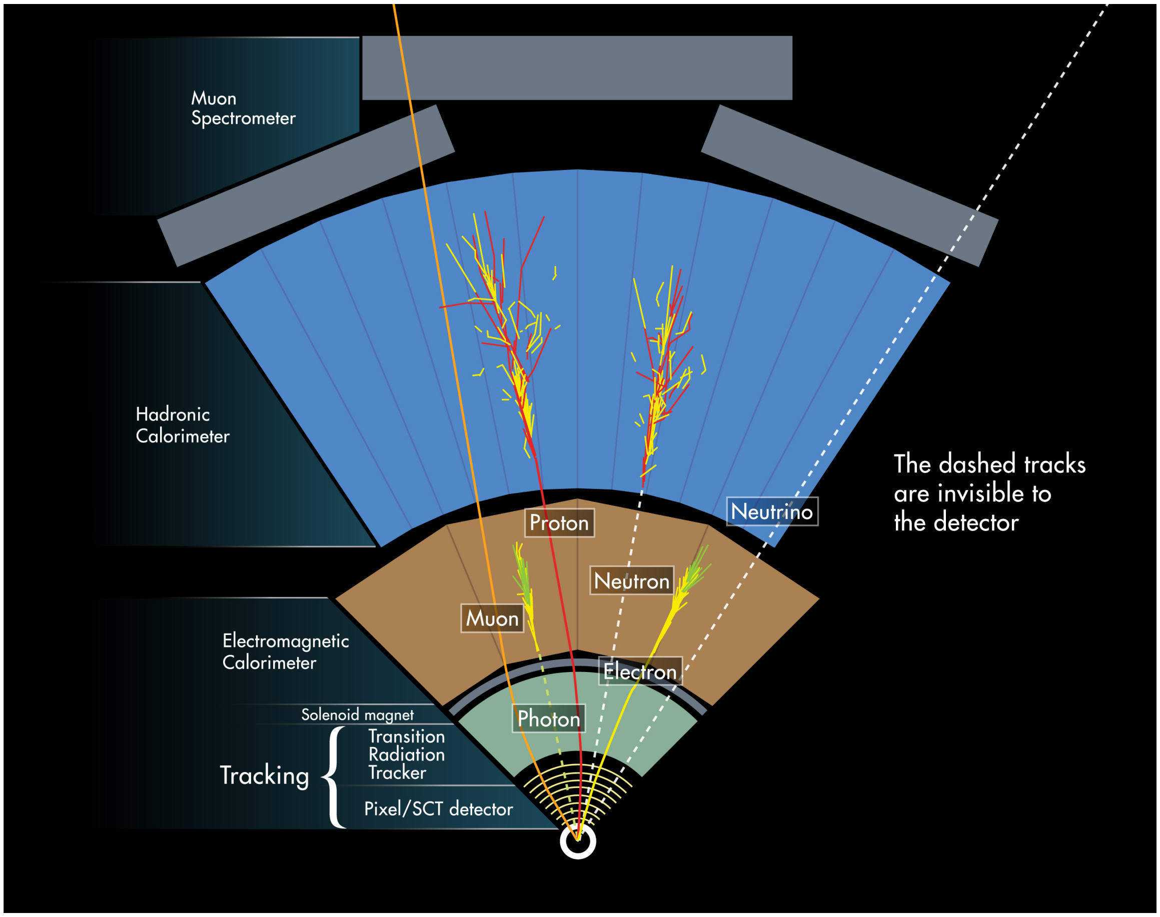


Fig. 16: A cross-sectional view of one portion of the ATLAS detector, with its Inner Detector, EM and Hadronic Calorimeters, and Muon Spectrometer. [17]

needed support to the innermost pixel tracker layers that experienced significant radiation damage during Run 2 operations [19]. Surrounding the Pixel Detector are the Semiconductor Tracker and the Transition Radiation Tracker, which collect tracking information for charged particles with silicon microstrips and 4mm straw tubes, respectively [20].

2.3.3 Calorimeters

Outside of the Inner Detector and its magnetic field are the calorimeters, meant to extract the energy deposited by particles they absorb. The innermost Electromagnetic (EM) Calorimeter uses lead to absorb particles (and thereby initiate particle showers as a particle interacts with the lead) and liquid argon as its sampling material. A particle's energy is measured using the information from how its shower forms through the layers of absorbing

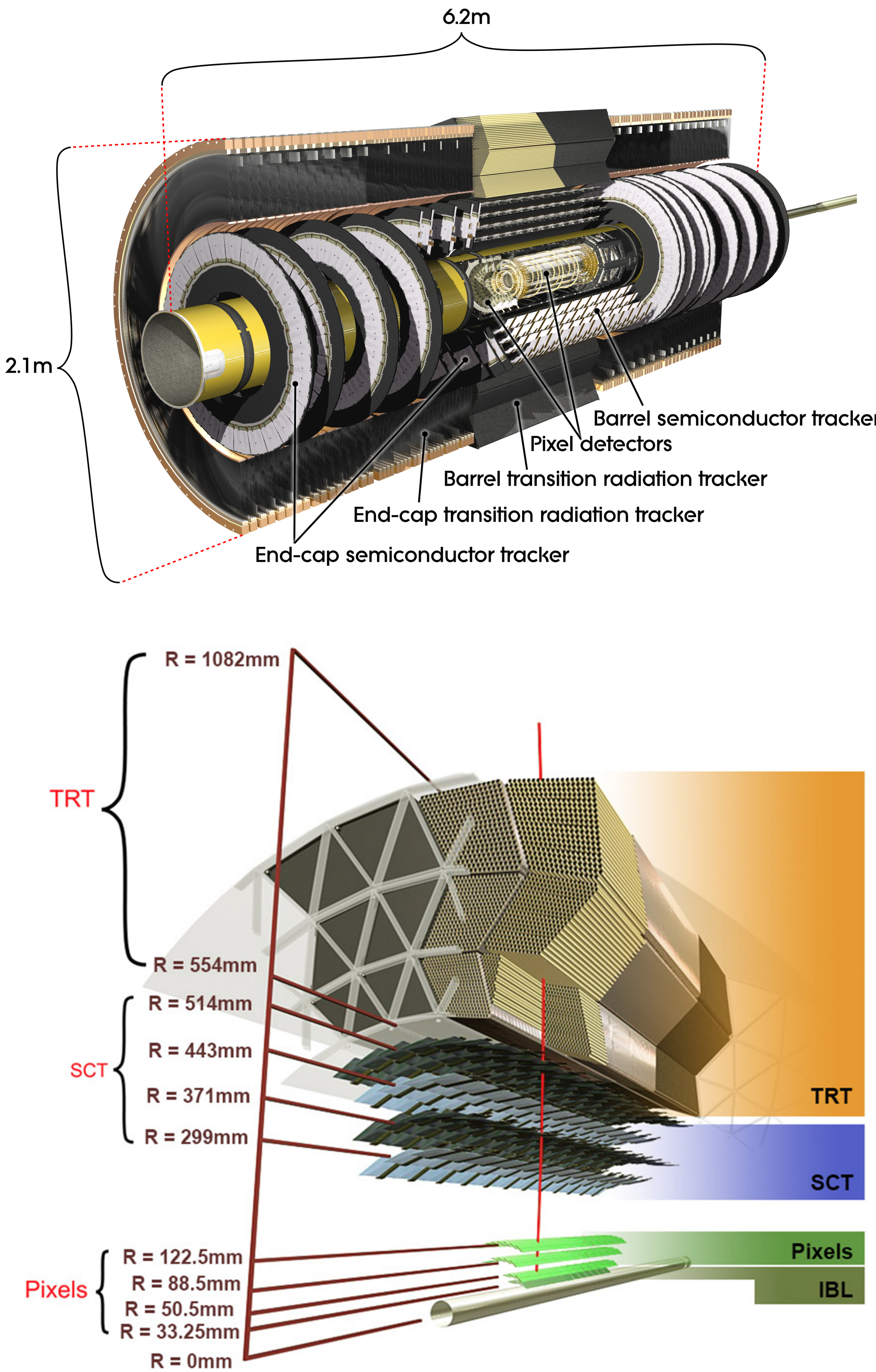


Fig. 17: Above: The ATLAS Inner Detector, including the Transition Radiation Tracker (TRT), Semiconductor Tracker (SCT), and Pixel detector. **Below:** A cross-sectional view of the ATLAS Inner Detector subsystem, including the IBL closest to the beamline.

and sampling material. The EM Calorimeter is surrounded by the Hadronic Calorimeter, which instead uses steel plates as its absorbing material with scintillating tiles interspersed between them to sample the shower [21]. As seen in Figure 16, the EM Calorimeter is optimized to absorb energy from particles that can interact via the electromagnetic force such as electrons and photons. The Hadronic Calorimeter, however, is designed to absorb energy from hadrons, i.e. particles containing quarks, such as protons and neutrons. As the interaction length of hadrons tends to be larger than the radiation length of electrons and photons in dense materials[7], the Hadronic Calorimeter lies outside of the EM Calorimeter. Together, both of these calorimeters are typically able to stop most particles emanating from the collision site other than muons and neutrinos.

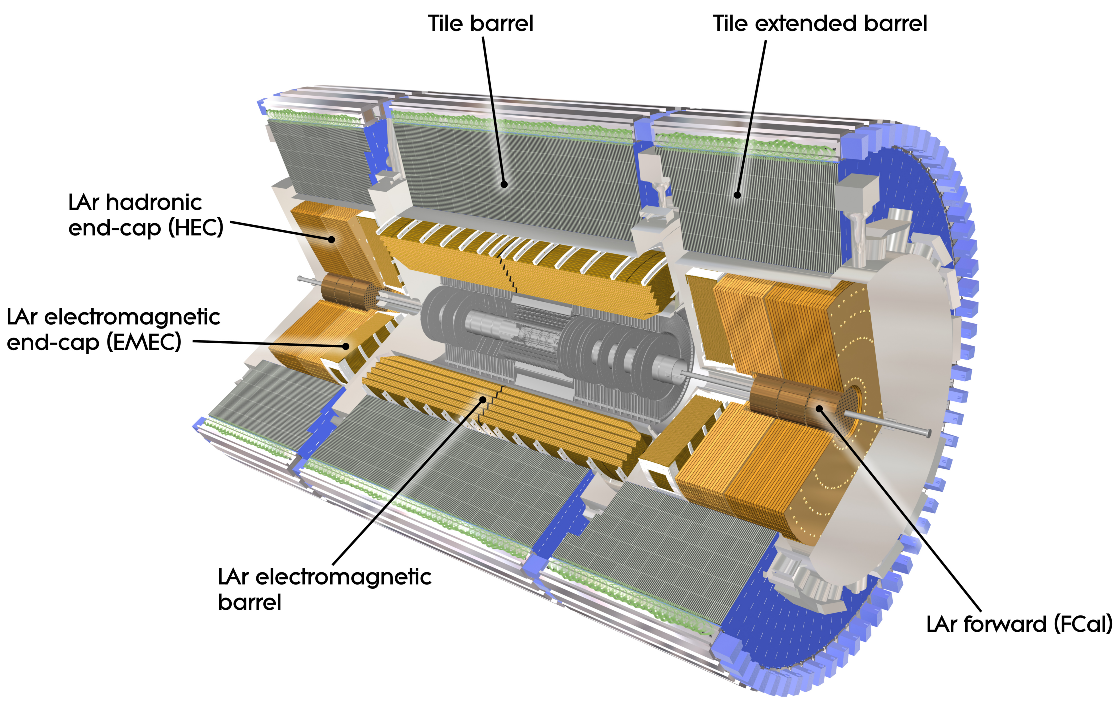


Fig. 18: The ATLAS calorimeter system, including the EM (LAr) & Hadronic (Tile) calorimeters.

[7]Both the nuclear interaction length and the electromagnetic radiation length give approximate length scales for characterizing electromagnetic vs. hadronic particle showers. In particular, the nuclear interaction length refers to the mean distance a hadronic particle travels between nuclear interactions, while the electromagnetic radiation length corresponds to 7/9 of the mean free path for photon pair production or the mean length over which an electron loses all but $1/e$ of its original energy.

2.3.4 Muon Spectrometer

At the outer edges of the ATLAS detector lies the Muon Spectrometer. The muon spectrometer system was designed to identify and measure the momenta of muons as they exit the ATLAS calorimeters [22]. The spectrometer is embedded in a 0.5 Tesla magnetic field, and muons are tracked within three levels of monitored drift tubes. The monitored drift tubes, with the aid of cathode strip chambers on either end of the detector, also measure muon p_T in the bending plane. Timing information for muons passing through the spectrometer is received from resistive plate chambers, which are used for triggering purposes as well as measurements of muon p_T in the non-bending plane. [23]

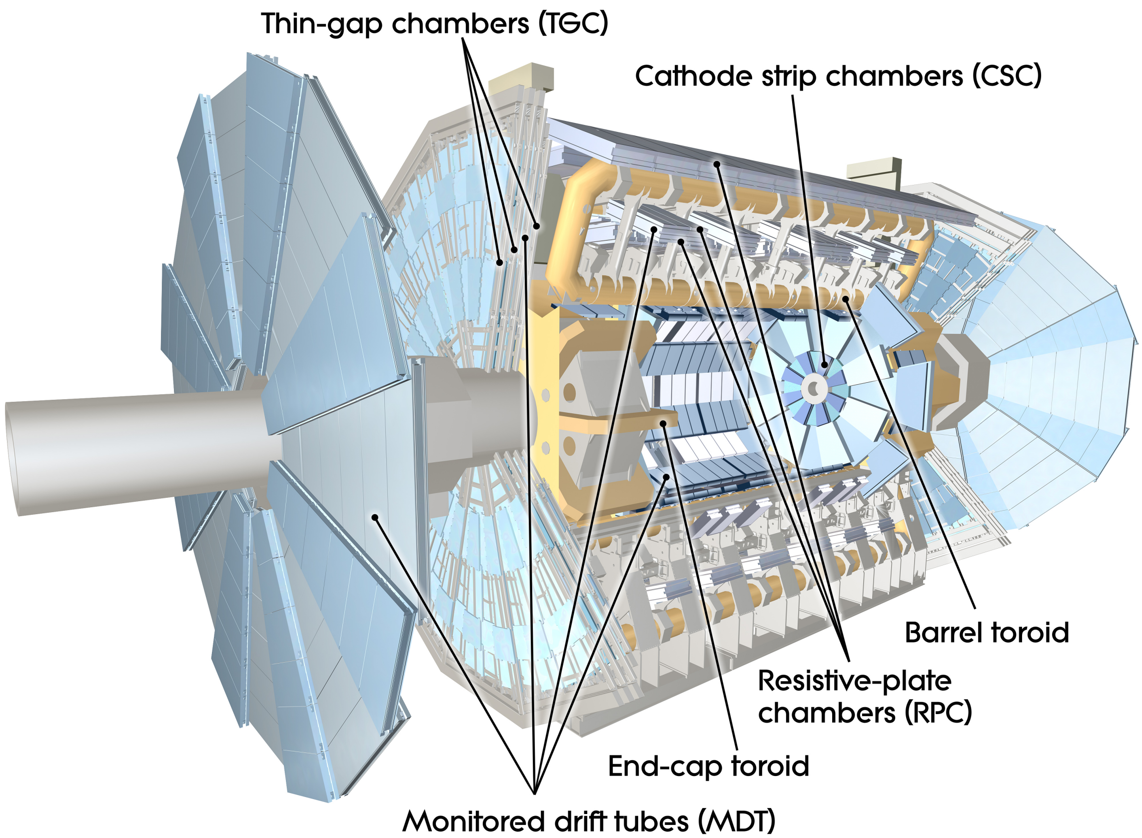


Fig. 19: The ATLAS muon spectrometer subsystem.

2.4 Trigger & Data Acquisition at ATLAS

As we learned in Section 2.2.4, LHC particle collisions occur every 25 nanoseconds, meaning a frequency of 40 million per second (40 MHz). From each of these crossing of proton beams comes a number of particle collision events – an average of 33.7 interactions per bunch crossing during Run 2 (see Figure 20).

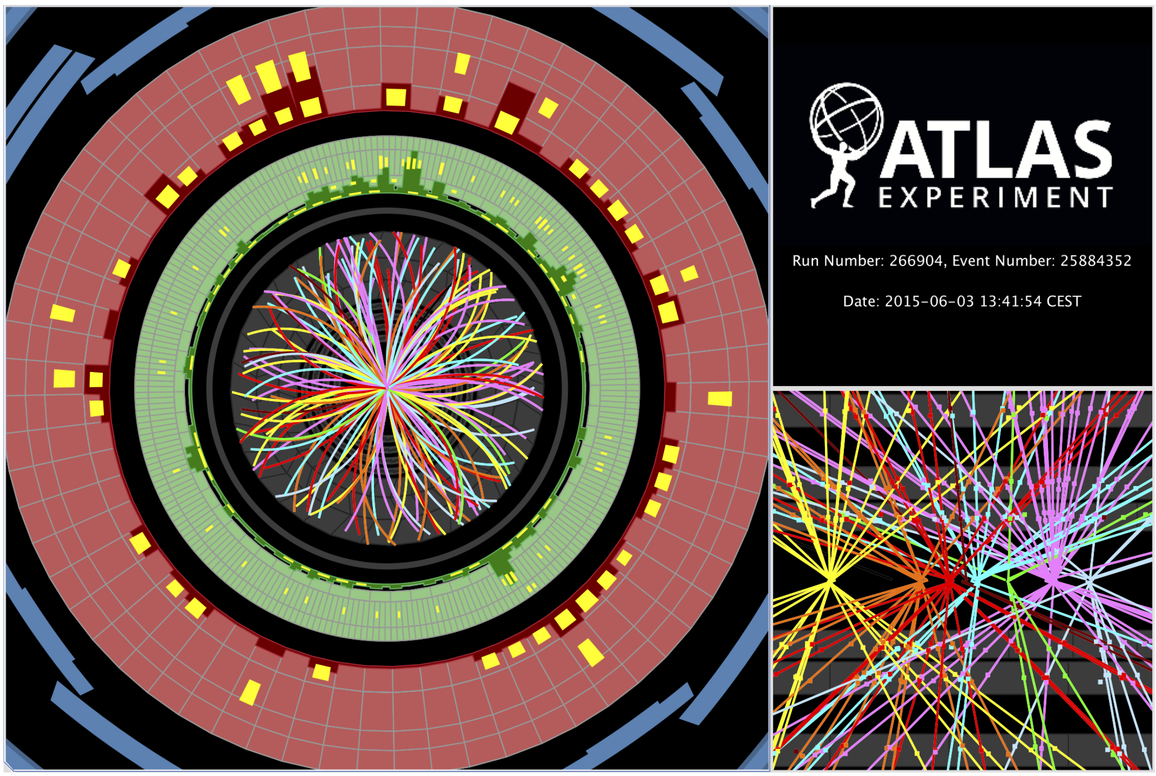


Fig. 20: An event display of a real ATLAS event from June 2015, during the first stable LHC beams at 13 TeV. **Left:** A perspective along the LHC beamline, with curved lines indicating charged particle trajectories in the tracking detectors and the green and yellow rectangles indicating the magnitude of energy deposits in the calorimeters. **Right:** The same event from a perspective perpendicular to the beamline. This angle makes it clear that there were several particle interactions, here visualized as different-colored lines emanating from distinct vertices, during this single beam crossing. It is likely that several of these vertices correspond to lower-energy collisions, also called "pileup events", that serve as undesirable background noise for many particle physics analyses.

The amount of data produced during an LHC run is therefore staggeringly large, and it is neither possible nor, even from a physics standpoint, desirable to store every piece of data from every collision. Many of the interactions per bunch crossing will yield low-energy, "soft" collision events that won't contain the interesting particles researchers care about, like taus or Higgs bosons. The ATLAS Trigger & Data Acquisition (TDAQ) system was therefore designed to record LHC data at a more manageable rate of 1 kHz during Run 2. The TDAQ system is tasked to decide in real time whether or not a given physics event should be one of the fewer than 1% of total collision events recorded to disk for later analysis.

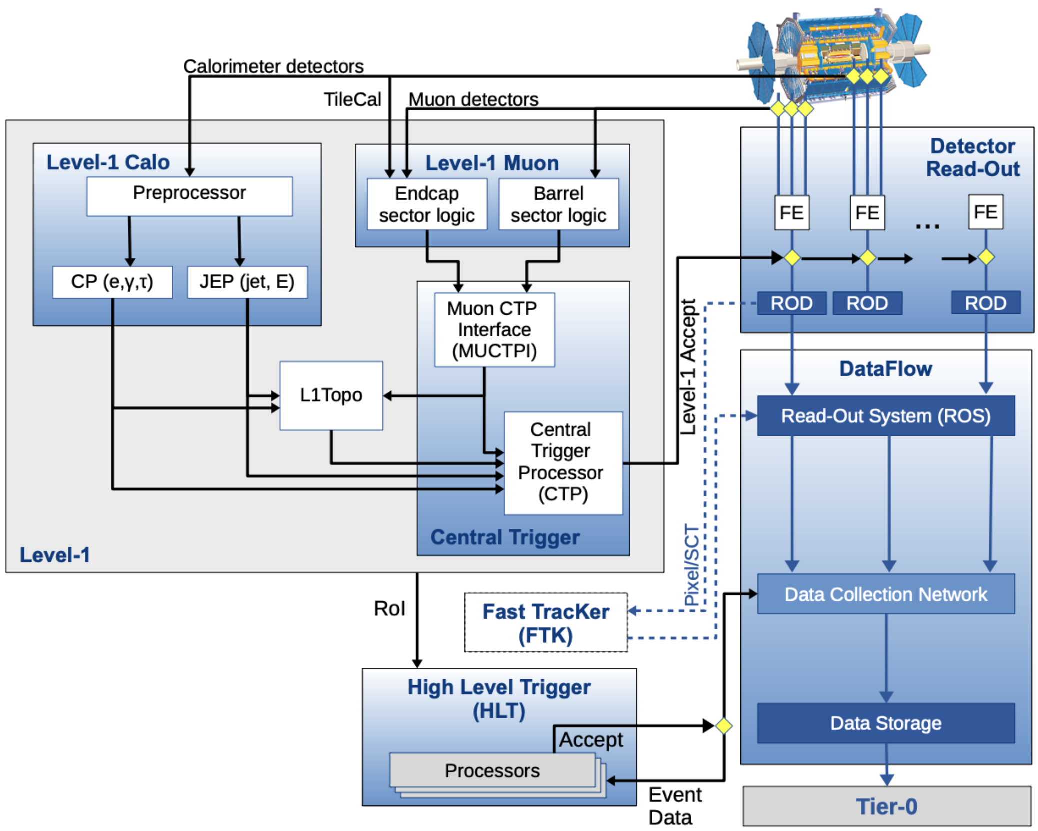


Fig. 21: A schematic of the ATLAS Trigger & Data Acquisition system, including the Level-1 (L1) trigger, the High Level Trigger (HLT), and the flow of data from the detector to permanent storage in the offline Tier-0 computing facility. The Fast TracKer (FTK), shown in this diagram, was planned for rapid track reconstruction at the L1 accept rate in Run 2, but the project was cancelled in 2019 and it was not used by the HLT in Run 2. [24]

The ATLAS TDAQ system, detailed in Figure 21, consists of two stages: the Level-1 (L1) trigger, a hardware-based trigger, and the High-Level Trigger (HLT), a software-based trigger.

- The hardware-based **L1 trigger** acts on a variety of information including event-level quantities, object multiplicities, reduced-granularity information from the calorimeters (L1Calo) and muon detectors (L1Muon), and topological (meaning kinematic or geometric) requirements (L1Topo). This information is synthesized in the Central Trigger Processor (CTP) to form the final L1 trigger decision for a given event. The L1 trigger has a maximum readout rate of 100 kHz and operates with a latency of 2.5 μs. Once the L1 trigger has accepted an event, detector information is streamed from the Front-End (FE) detector electronics to the Read-Out Drivers (RODs) for processing and then to the Read-Out System (ROS) for buffering the data. The data is then passed by request from the ROS to the HLT to inform the second stage of the triggering process.

- The **High-Level Trigger (HLT)** is software-based, consisting of many dedicated algorithms to identify the approximately 1,500 specific event signatures of interest in the trigger menu. These algorithms are deployed on a computing farm of around 40,000 processors. When requested, they can also make use of information from the full detector. The algorithms are generally designed to operate on a region of interest (RoI) identified by the L1 trigger. After extracting certain features from the RoI, the HLT then uses one or more custom methods to determine its overall trigger decision. The ATLAS HLT had an average readout of approximately 1.2 kHz and an average throughput rate of physics data to permanent offline storage of 1.2 GB/s in Run 2.

2.5 Event Simulation

ATLAS analyses depend crucially on simulations of particle physics events to inform our expectations for which events we should see during LHC datataking runs and how those collisions will interface with the detector. Claiming evidence of a new particle or physics process requires a thorough understanding of the data we would expect to have collected in the absence of this new behavior. We can then calculate how likely it is that the data we observed could have emerged from a *null hypothesis*, i.e. a scenario aligned with our current understanding of the Standard Model, versus a process unexplained by the Standard Model. Because these simulation methods are grounded in Monte Carlo methods, meaning repeated random sampling techniques, they are often referred to as **Monte Carlo (MC)**

simulations.

The MC simulation process begins with the generation of the underlying physics interaction and its immediate particle decays. For events involving QCD-induced hadronization, this process (described in Section 2.5.2) is considerably more involved. There are then steps to simulate the interaction of that physics process with the ATLAS detector itself, followed by a digitization step that outputs raw electronic voltages and currents in the same format that we would see from the real readout of a physics event.

MC generators at ATLAS include generators of the matrix element only (e.g. MadGraph, Powheg) as well as general-purpose generators that include parton showering capabilities in addition to matrix element calculations. The MC simulations used for this thesis work are primarily from the generators Powheg+Pythia8 and Sherpa [25].

2.5.1 Matrix Element Calculations

Calculating the expected rate of a particle physics interaction, at its core, springs from Fermi's Golden Rule:

$$\Gamma_{fi} = 2\pi |T_{fi}|^2 \rho(E_f) \tag{59}$$

where Γ_{fi} is the number of transitions per unit time from an initial state i to a final state f, $|T_{fi}|$ is the transition **matrix element** of the superposition of an unperturbed Hamiltonian and a perturbing interaction potential, and $\rho(E_f)$ is the density of states that accounts for the kinematic likelihood of the transition. Imposing requirements for Lorentz invariance, energy conservation, and momentum conservation modifies this core structure into a more useful form:

$$\Gamma_{fi} = \frac{(2\pi)^4}{2E_i} \int |\mathcal{M}_{fi}|^2 \delta(E_i - \sum_f E_f) \delta^3(\vec{p}_i - \sum_f \vec{p}_f) \prod_f \frac{\mathrm{d}^3\vec{p}_f}{(2\pi)^3 2E_f} \tag{60}$$

This form allows us to integrate over all possible final states allowed by energy and momentum conservation given a specific initial state, but the integral itself is now Lorentz invariant. The delta functions impose the conservation rules explicitly, the matrix element T_{fi} is replaced with a Lorentz-invariant analogue $\mathcal{M}_{fi}$, and the density of states $\rho(E_f)$ is now a product of normalized Lorentz-invariant phase spaces for each final-state particle.

The fundamental particle physics laws defining the likelihood of a given interaction are all encoded in the matrix element $\mathcal{M}_{fi}$, so the approximation of these matrix elements are a primary component of our MC simulation techniques.

2.5.2 Hadronization

For many QCD processes at the LHC, a formula like Equation 60 is insufficient for describing the full dynamics of the particle event, as it only considers the products of the hard-scattering (high-momentum transfer, or high Q^2) interaction, not the softer (lower-momentum, or low Q^2) aftereffects of QCD-initated gluon and quark radiation from the core underlying event. Following the hard-scattering interaction, final-state partons then undergo a period of *parton showering* during which they radiate cascades of partons that typically align with the directions of the original partons. Finally, the showers begin the process of *hadronization*, during which they develop into multi-parton bound states called hadrons. It is at this point that nonperturbative QCD effects become nonnegligible. Figure 22 shows an illustration of each of these steps over time.

Just as we dealt with non-perturbative QCD effects for the proton with parton distribution functions (PDFs) in Section 2.2.2, MC generators also make use of PDFs to calculate cross-sections of QCD processes at the LHC. To n^{th} order in perturbation theory, the inclusive cross-section for $pp \to X$ is:

$$\sigma^{(n)} = \text{PDF}(x_1, \mu_F) \otimes \text{PDF}(x_2, \mu_F) \otimes \hat{\sigma}^{(n)}(x_1, x_2, \mu_F, \mu_R), \tag{61}$$

where the two PDFs correspond to each incoming proton at factorization energy scale μ_F and $\hat{\sigma}^{(n)}$ is the hard-scattering, parton-level cross-section that depends on μ_F as well as the renormalization scale μ_R. These energy scales are set not by an *a priori* calculation, and so their values are chosen by hand. The inclusion of energy scales μ_F and μ_R is necessitated by divergences in the matrix element calculation: introducing cutoff scales prevents the integrals from diverging to infinity. μ_F accounts for the infrared (IR) divergences, while μ_R accounts for the ultraviolet (UV) divergences.

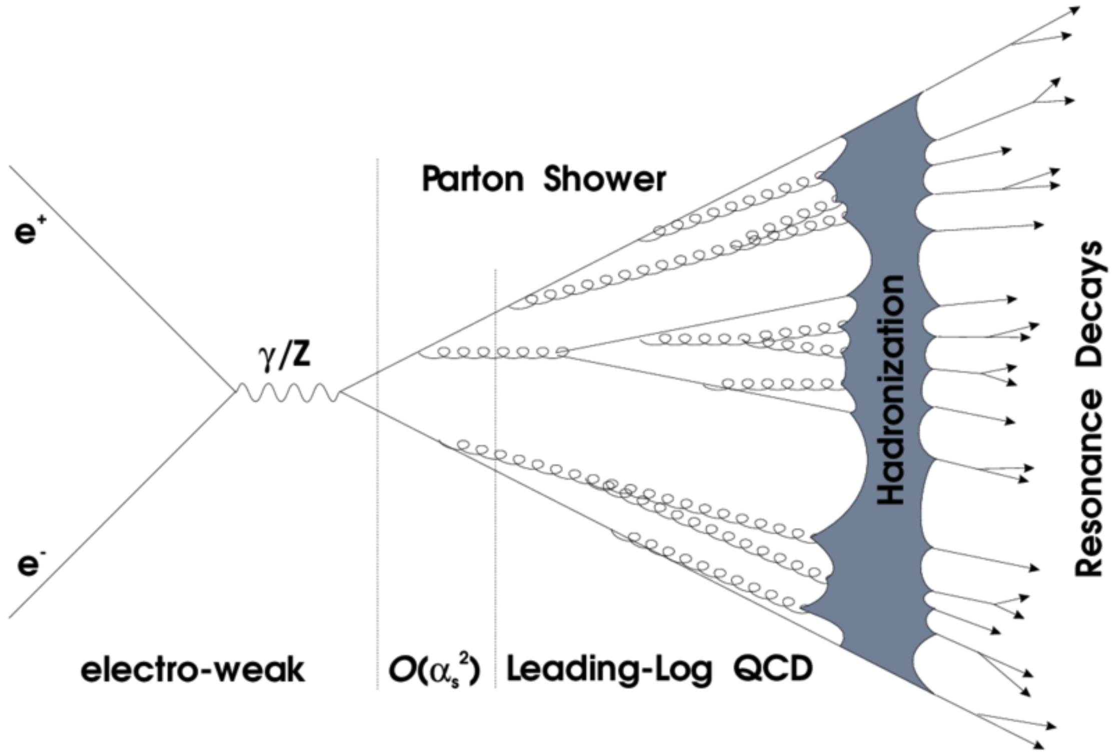


Fig. 22: An example schematic of a QCD hadronization process emerging from electron-positron annihiliation into two quarks. From the two final state quarks, we see cascades of partons radiating in parton showers, and then a hadronization stage, followed by the decays of those hadrons. [26]

2.5.3 *Detector Simulation and Digitization*

Once generated, physics events are then run through a software called Geant4 that propagates a record of all the stable "truth" particles in an event through a complex simulation of the ATLAS detector. This simulation accounts for not only detector geometry and its possible misalignments, but also qualities relating to how particles could interact with the materials of the detector. Following the propagation of the event through the simulated ATLAS detector, additional custom software elements convert the simulated event into raw electronic outputs. This stage also includes simulations of various sources of electronic noise from the detector. The digitized outputs, called Raw Data Objects (RDOs), can be easily converted into bytestream form in order to match the actual ATLAS detector outputs. This allows ATLAS physicists to run the same trigger-level algorithms on both simulated and real data.

2.6 Reconstruction and Identification of Analysis Objects

Up until this point, we have covered in detail the steps taken to produce two beams of protons, accelerate them in opposite directions around the tunnel of the LHC, and intersect their beams in order to rapidly produce high-energy particle collisions. The aftermath of these collisions is then converted into electrical signals by the ATLAS detector, and these signals are filtered by the TDAQ system in order to store the events for offline analysis.

However, this story omits a critical step: the reconstruction and identification of objects that will then be fed into physics analyses. In other words, how do we convert each dazzling spray of electronic signals into a structured list of electrons, muons, taus, and more? The ATLAS Experiment has devised bespoke methods for the classification of detector signals into physics objects such as electrons, photons, muons, QCD jets, and missing E_T that I will summarize in this section.

Note: As tau objects are particularly important for this analysis, I will discuss tau reconstruction and identification at length in Chapter 4.

2.6.1 Electrons

As charged particles, electrons leave charged tracks in the ATLAS inner detector. They also leave localized energy deposits, particularly within the EM calorimeter, corresponding with their energies. Electron reconstruction therefore involves a careful matching of suitable charged tracks with corresponding clusters of energy deposits in the EM calorimeter. Figure 23 shows a detailed schematic of the typical path of an electron from the beam axis, through the inner detector, and finally to the EM calorimeter.

Electron track reconstruction begins with the formation of a track seed of three hits from silicon tracker layers. A pattern recognition algorithm is then run in hopes of extending the track seed to a full track of at least seven silicon hits for track candidates with $p_T > 400$ MeV. Any distinct track candidates sharing silicon tracker hits are then passed to algorithms designed for ambiguity resolution to unequivocally assign silicon track hits to individual track candidates. Then, to account for the energy loss resulting from charged particles interacting with detector materials, a track-fitting method based on a Kalman filter called the Gaussian-sum filter (GSF) is run.

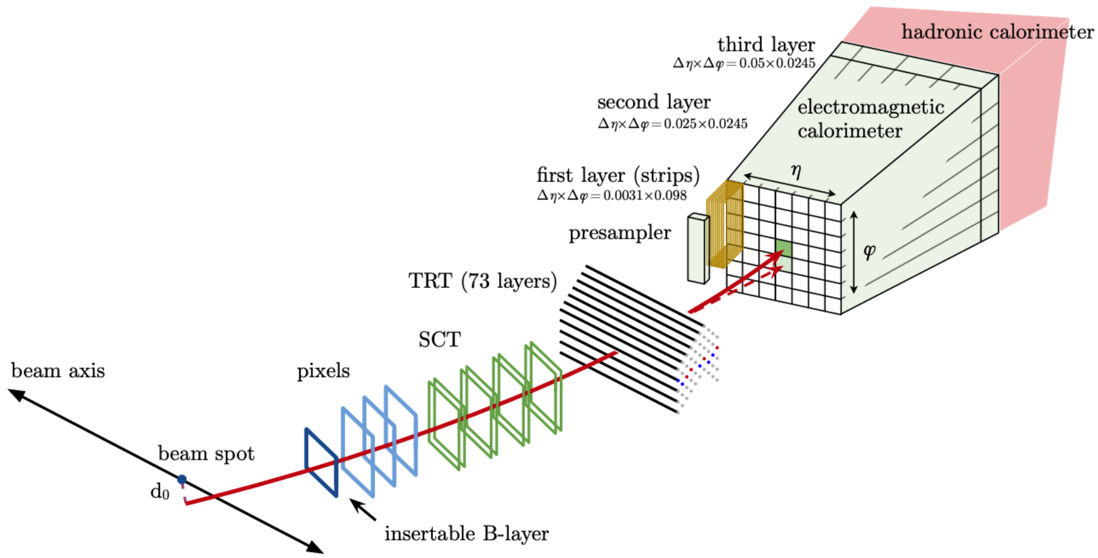


Fig. 23: The red arrow shows an example path of an electron moving through the trackers of the ATLAS inner detector in a curved trajectory before interacting with the EM calorimeter. The red dashed arrow indicates the path of a photon produced as a result of the electron interacting with the tracking material. [27]

In Run 2, a preexisting method for reconstructing calorimeter seed clusters based on a sliding-window method over fixed-sized clusters of calorimeter cells was replaced with a dynamic clustering method to create variable-size clusters called *superclusters*. Topo-clusters, or topologically-connected calorimeter cell clusters, form the seeds of these superclusters. The topo-clusters first emerge by finding calorimeter cells initiating a cluster that contain energies greater than four times the expected cell noise from electronic and pileup noise. Neighbor cells then join these initial proto-clusters if their energies exceed twice the expected cell noise. In general, only the energy deposits from the EM calorimeter are summed for the electron reconstruction. To transform topo-clusters into superclusters, the EM topo-clusters are considered in descending order in E_T and tested to see if they pass a minimum $E_T > 1$ GeV and are matched to a track with at least four silicon tracker hits.

Given both a GSF-filtered track candidate and a candidate calorimeter supercluster, we apply a matching procedure to require that the track and calorimeter supercluster are close together in η and ϕ: $|\eta_{\text{cluster}} - \eta_{\text{track}}| < 0.05$ and $-0.1 < -q \times (\phi_{\text{cluster}} - \phi_{\text{track}}) < 0.05$. The overall cluster energy is calibrated to match the original energy of the incoming electron using multivariate techniques, and the final track parameters of the electron candidate are taken from the best-matched track to the supercluster. As shown in Figure 24, electron reconstruction efficiency (defined as the percent of true electrons that are reconstructed as an electron candidate with good track quality, i.e. one pixel hit and seven silicon tracker hits) is better than 97% above $E_T > 15$ GeV.

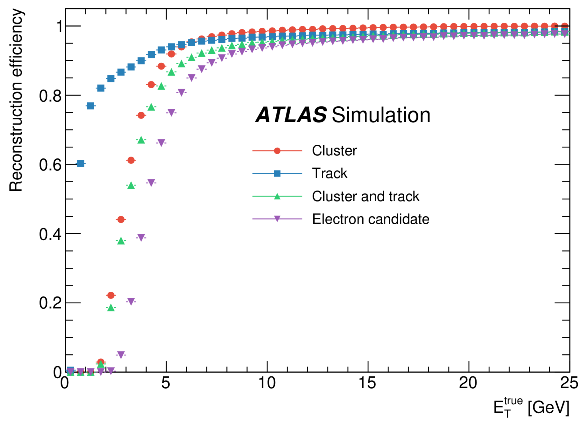


Fig. 24: The reconstruction efficiency for simulated electrons from a single-electron sample is shown as a function of true E_T, or transverse energy, during each step of the electron reconstruction process. Above $E_T > 15$ GeV, the reconstruction efficiency is higher than 97%. [27]

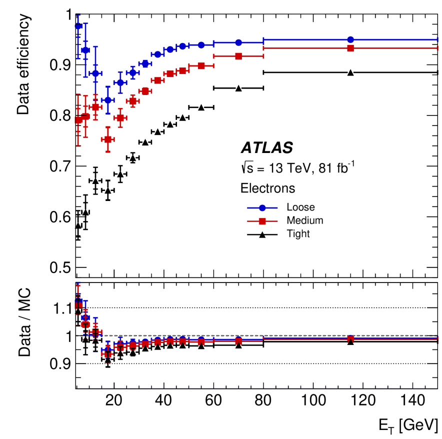


Fig. 25: Electron identification efficiencies measured in $Z \rightarrow ee$ data events as a function of E_T are shown for three working points: loose (blue), medium (red), and tight (black). [27]

Additional electron quality criteria are applied to reconstructed electron candidates in the form of electron identification. This is valuable for separating prompt, isolated electrons from background processes such as photon conversions or hadronic decays from other processes. A suite of variables including basic track and cluster parameters are fed into a data-driven likelihood discriminant model (an adaptive kernel density estimator, or KDE) in bins of E_T and η. The performance of this electron identification scheme in Run 2 is shown as a function of E_T in Figure 25.

2.6.2 Photons

Photon reconstruction follows essentially the same calorimeter-clustering techniques as are used for electrons. As photons are electrically-neutral, they will not leave charged tracks in the inner detector. However, between 20-65% of photons will convert to an electronpositon pair in the inner detector depending on their position in $|\eta|$. Therefore an unconverted photon corresponds to a calorimeter supercluster with no associated ID track, while a converted photon corresponds to a calorimeter supercluster with an associated conversion vertex in the ID. For converted photons only, the superclustering step will incorporate new topo-clusters that match with the same conversion vertex as the seed cluster.

Given the similarities in supercluster development for both electrons and photons, an ambiguity resolution scheme is sometimes needed to decide whether to assign a given supercluster to an electron or photon object based on the presence of a quality track candidate attached to the supercluster.

Photon identification uses a cut-based selection based on calorimeter shower shape variables. The identification efficiencies for unconverted and converted photons at the tight working point are shown in Figure 26.

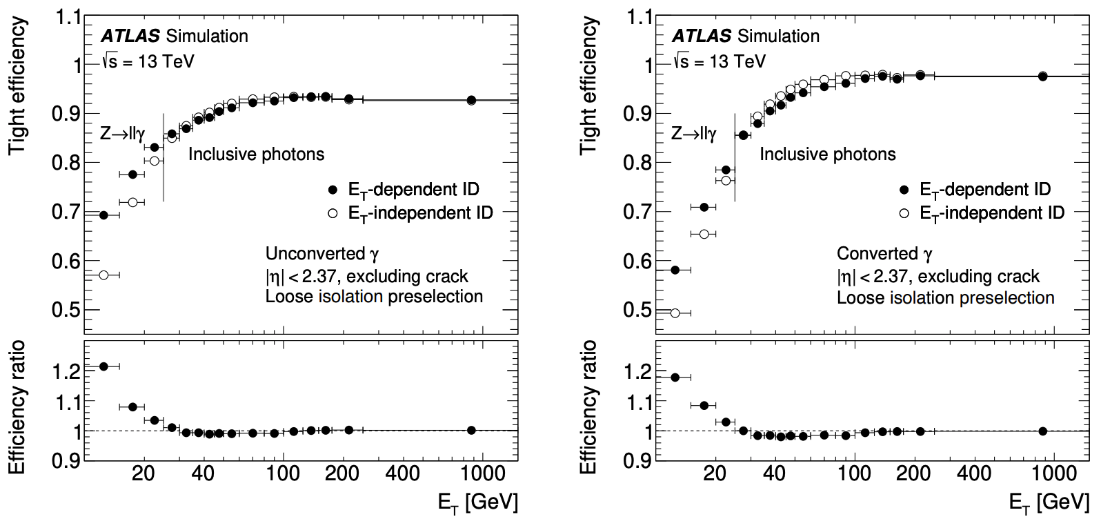


Fig. 26: Identification efficiencies versus E_T for the tight working point for unconverted (left) and converted (right) photons in $|\eta| < 2.37$. For $E_T < 25$ GeV, events come from $Z \to ll\gamma$, and for $E_T > 25$ GeV, events come from inclusive photon production. [27]

2.6.3 Muons

Like electrons, muons will also leave charged tracks in the tracking detectors, but unlike electrons, muons at the LHC are often produced at energies corresponding to a minimum in their stopping power, i.e. how much energy is lost as they interact with materials in the detector. A muon with this trait is referred to as a minimally-ionizing particle. Muons' calorimeter energy deposits are therefore very unlike those of electrons, photons, or hadrons. The Muon Spectrometer (MS) subsystem was designed to augment muon reconstruction and identification with additional tracking signatures of muon candidates beyond the inner detector trackers.

Track reconstruction in the MS involves a global χ^2 fit of the muon's trajectory through hits in the MS, taking into account the muon's expected interactions with the detector materials. The track candidate is then extended to include additional hits and is re-fit. As with electron track reconstruction, ambiguities with tracks sharing multiple track hits are resolved by eliminating lower-quality tracks that share many hits with another, higher-quality track. When possible, MS tracks are matched to inner detector (ID) tracks to formed a "combined" muon candidate. Otherwise, muon candidates can also be constructed by extrapolating ID track candidates out to possible candidate hits in the MS. MS track candidates with no matching ID track candidates can be extrapolated inwards to the beamline. Additionally, muon candidates can emerge from calorimeter tagging, in which ID tracks are extrapolated through the calorimeters and find the energy signatures characteristic of a minimally-ionizing particle.

Muon identification methods are then applied to reconstructed muon candidates to apply further quality criteria when desired. This is particularly useful when distinguishing prompt muons from non-prompt muons emerging from processes such as hadron decays. For muons with an ID track, all muon identification working points require at least one pixel detector hit and five silicon detector hits. Muon reconstruction and identification efficiencies across Loose, Medium, and Tight working points measured in both data and MC are shown as a function of p_T and η in Figure 27.

2.6.4 Jets

The broad term "jets" refers to a wide class of sprays of particles initiated by the parton showering and hadronization processes affiliated with QCD interactions. While the jet itself is not a well-defined physics object, the identity of the original parton seeding the jet will define many kinematic properties of the jet, and therefore jets can be associated (or

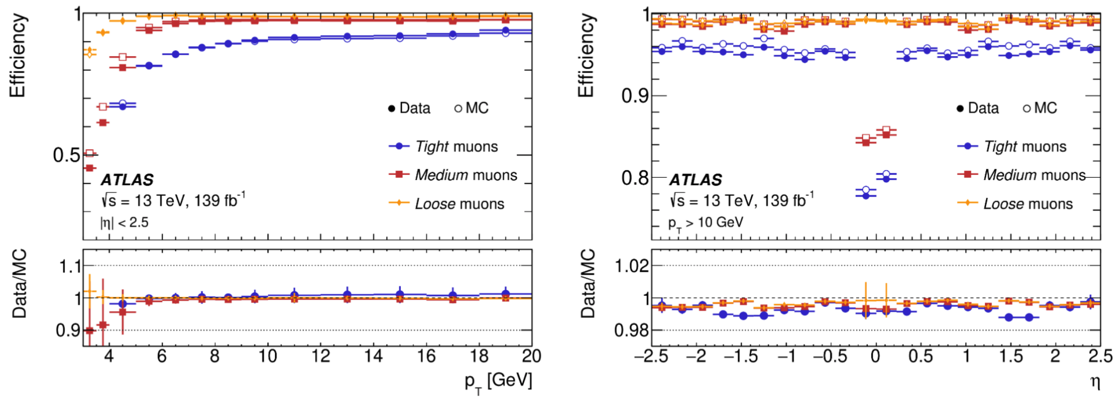


Fig. 27: ATLAS Run 2 muon reconstruction and identification efficiencies for the Loose, Medium, and Tight working points are calculated in both data and MC for (left) $J/\Psi \to \mu\mu$ events as a function of p_T and (right) for $Z \to \mu\mu$ events as a function of η. Above $p_T > 15$ GeV, all working points exceed 90% efficiency for $|\eta| < 2.5$. For $|\eta| > 0.1$, all working points exceed 95% efficiency for $p_T > 10$ GeV. [28]

"tagged") by their originating parton.

Before the jets are tagged to determine their likely originators, they must first be reconstructed as a jet in the first place via a jet-clustering algorithm. The purpose of jet clustering is to aggregate calorimeter topo-clusters into discrete groupings that can then be assigned a (calibrated) energy. The default ATLAS jet-finding algorithm is the anti-k_t algorithm [29], which produces circular hard jets, using distance parameter $R = 0.4$ for "small-R" jets representing quarks and gluons and $R = 1.0$ for "large-R" jets representing massive particles decaying hadronically [30]. Figure 28 shows an illustration of the results of an anti-k_t clustering using $R = 1.0$ on a sample parton-level event.

Once a jet has been clustered, potential pileup contributions are removed via a pileup suppression scheme. First, the jet axis is relocated to the hard-scattering vertex of interest, and then the energy corresponding to the product of the jet area times the event-specific energy density within $|\eta| < 2.0$ is subtracted from the jet. Next, the Jet Energy Scale (JES) is calibrated using MC truth information to calculate the lost energy between a true jet and a reconstructed jet.

Following a jet's calibration is the jet tagging algorithm to identify the origin of the jet. Dedicated algorithms exist for top-quark tagging, W boson tagging, and c-quark/b-quark

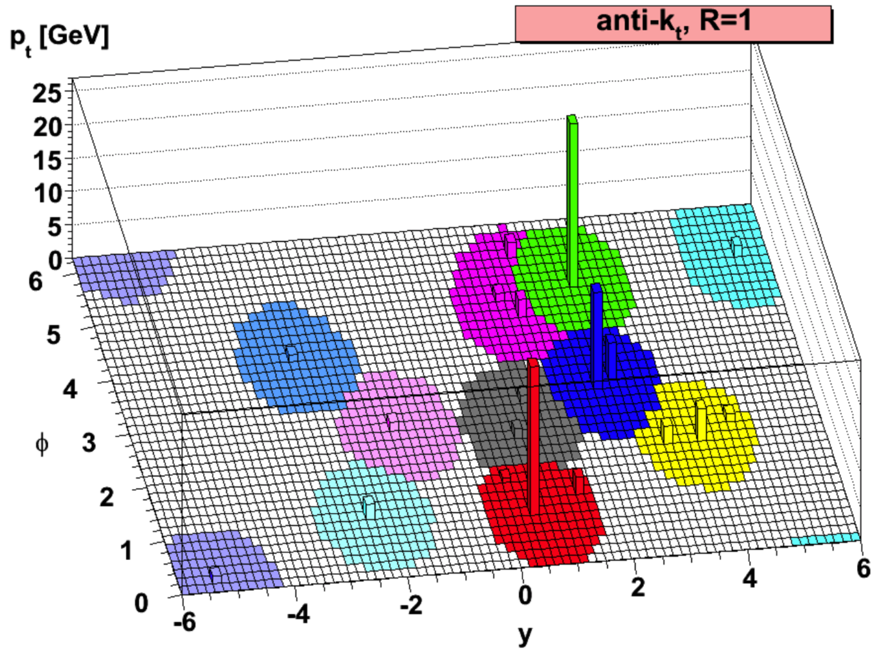


Fig. 28: The clustering results of the anti-k_t jet-finding algorithm for a parton-level event are shown as a function of ϕ, y, and p_T. [29]

tagging, among others ([31], [32]).

2.6.5 Missing E_T

Counterintuitively, one of the crucial physics analysis objects we consider at ATLAS isn't an observable at all. Missing E_T, sometimes referred to as MET, refers to the combined energy in the transverse plane carried out of the ATLAS detector by particles we are unable to detect. Proton-proton beam collisions along the $\hat{z}$ axis are assumed to be head-on, meaning they should take place with exactly zero momentum in the transverse $(\hat{x} - \hat{y})$ plane. Based on energy and momentum conservation laws, we should therefore expect the sum of all the transverse energy in the final state of the particle collision process to also be exactly zero. $\sum E_T^{\text{miss}} \neq 0$ in the final state suggests the presence of a particle, whether a neutrino or perhaps a new particle from a process beyond the Standard Model, that can escape the ATLAS detector without being detected. Neutrinos are expected to escape the ATLAS detector as they are electrically neutral, colorless, and nearly massless, meaning that they will leave neither significant calorimeter deposits nor tracker hits. In fact, neutrinos interact so rarely with regular matter that they usually pass without a trace through the entire Earth's diameter, so we certainly expect them to escape the ATLAS detector.

The process of reconstructing missing E_T requires inputs from the whole detector from fully-reconstructed and calibrated electrons, muons, taus, photons, and jets. There are two contributing terms to the reconstruction: one for the p_T vectors of the hard-scattering objects and one for the p_T vectors of the soft event signals, i.e. charged tracks, associated with the vertices of hard-scattering events but not the objects themselves:

$$E_{x(y)}^{\text{miss}} = - \sum_{i \in \{\text{hard objects}\}} p_{x(y),i} - \sum_{j \in \{\text{soft signals}\}} p_{x(y),j} \tag{62}$$

Given the components E_x^{miss} and E_y^{miss}, E_T^{miss} is constructed as:

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}. \tag{63}$$

The performance of missing E_T reconstruction in data and MC simulation in Run 2 is shown by comparing data and MC as a function of missing E_T in Figure 29.

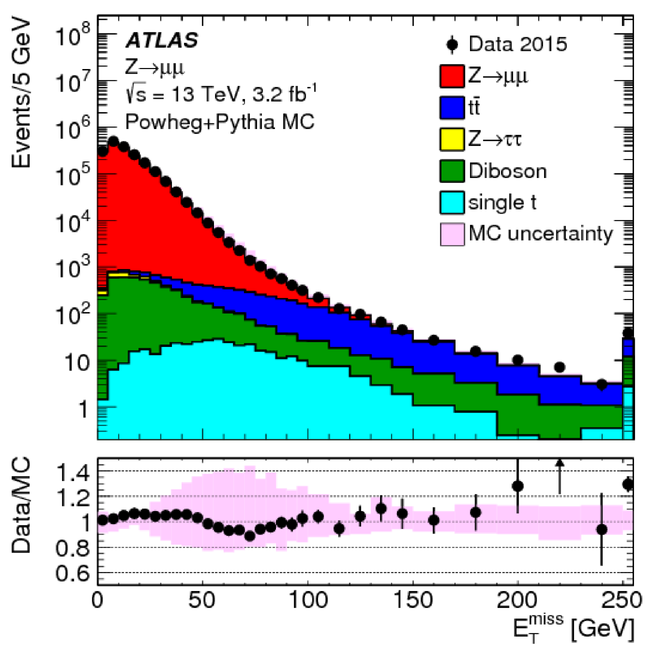


Fig. 29: A distribution of missing E_T for inclusive $Z\mu\mu$ events at $\sqrt{s} = 13$ TeV shows good agreement between data and MC for all relevant samples. [33]

3 Common Machine Learning (ML) Architectures

3.1 Introduction

Machine learning (ML) researchers try to build **machines that learn**. What does this actually mean, and why is it significant? For much of the history of technology, machines were *programmed* to perform specific tasks based on complicated pre-installed logic. A machine that *learns*, however, is a much more subtle thing. This often means that a machine has an internal *model* of the world (or, more realistically, the environment immediately surrounding the task at hand) that it regularly and automatically updates based on feedback from its attempts to perform a task. When this process succeeds, the machine's internal model can not only solve problems it has already seen, but can also generalize pretty well on completely new data.

To achieve this, researchers commonly split a dataset into separate portions and only use one portion for *training*, or fitting, a model to that data. The model is then run on another portion of the data, called *validation* or *testing* data, to test how well it performs on data it has not previously encountered. ML training regimens are structured around the delicate balance between finding an optimal solution for the training data while preserving generalizability, i.e. performance on the validation or test sets.

Though the term "machine learning" was first popularized in 1959 [34], the models developed for this thesis all belong to the category of **deep neural networks** that emerged in more recent decades and has seen an explosion in growth since 2012 that is sometimes called the *deep learning revolution*. Deep neural networks provide plentiful possibilities for learning complex relationships in data without the need for heavy feature engineering by stacking multiple **hidden layers** of artificial neurons, or "units", together with nonlinear activation functions (see Figure 30). These operations allow the model to create abstracted representations of input data. With a large enough hidden layer, this model structure has been shown to be able to theoretically approximate any function to within an arbitrary precision [35]. However, computing and time constraints often mean that training the largest possible network is not feasible, and so many variants of the basic neural network

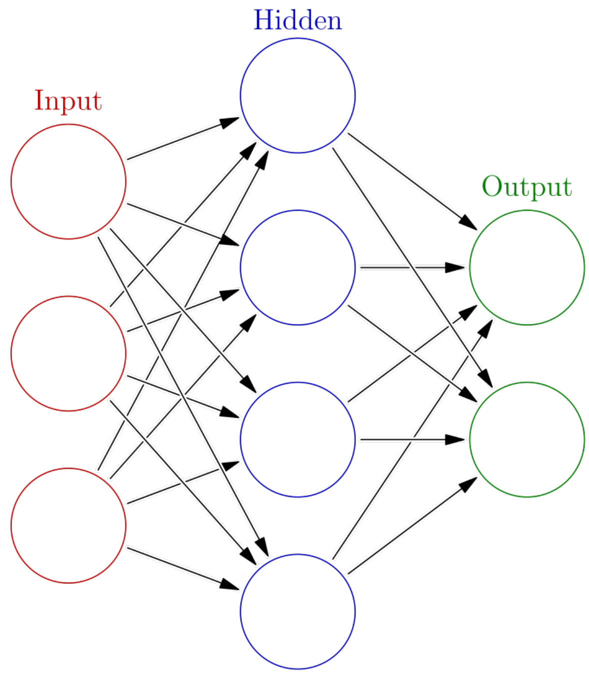


Fig. 30: A diagram of a neural network, with information flowing from left to right, starting with an input layer of size 3, a hidden layer of size 4, and an output layer of size 2. (Public domain image from Wikimedia Commons).

structure have been developed to exploit specific properties of the data such as structural symmetries or sparsity. These variants can also have benefits in terms of interpretability – their latent structures can reveal meaningful patterns in the data.

In the following sections, I will give a brief overview of a variety of neural network architectures that will be relevant for understanding the results shown in later chapters. Important deep neural network models I won't cover here, as they are not relevant to the studies I show later, include Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), transformers, and many others.

3.2 Neural Networks (NNs)

A deep neural network is fundamentally a series of functions applied to a single instance of input data. These functions are a combination of linear and nonlinear in nature, and the specific construction of functions in a ML model is called its **architecture**. The linear functions are called **layers** and the nonlinear functions are called **activation functions**. Usually, activation functions are applied after each layer such that linear and nonlinear functions alternate.

The basic linear operation applied in each layer is:

$$f(h_j) = \sum_i W_{i,j} h_i + b_j \tag{64}$$

where $\vec{h}$ is a vector of the model's representation of the input data, W is a matrix of **weights**, and $\vec{b}$ is a **bias** vector with the same length as $\vec{h}$.

In the first layer of the NN, $\vec{h}$ is simply the input data. To get the updated representation of the data after a pass through the first layer of the model, we apply the weight matrix and then add an offset provided by the bias vector. The weight matrix and bias vector (together called the model's **trainable parameters**) typically begin with random initializations, but are updated during the model's training process to better adapt to the task at hand.

This operation is completely linear, analogous to $y = mx + b$. Stacking purely linear functions together is, unfortunately, not going to allow us to approximate *any* given function to within arbitrary precision. The expressive power of neural networks is significantly improved with the inclusion of *nonlinear* activation functions following each layer. There are countless possible activation functions to choose from, but ones recently popular in the ML field include the Rectified Linear Unit (**ReLU**) and **sigmoid** (σ) activation functions depicted in Figure 31. These functions are defined as:

$$\mathrm{ReLU}(x) = f^+(x) = \max(0, x) \tag{65}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{66}$$

In most of the models I discuss in this thesis, ReLU activation functions are used throughout, while sigmoid functions are often applied in the very last layer of a model if it is meant to be used as a classifier. This is because the ReLU function has several computational benefits during training compared with the more complex sigmoid function, but unlike ReLU, the sigmoid function helpfully maps an initial activation $(-\infty, \infty) \rightarrow [0, 1]$, i.e. an output score.

At this point in the description of a neural network, we have understood how an example of input data is transformed into an abstract representation by passing through multiple

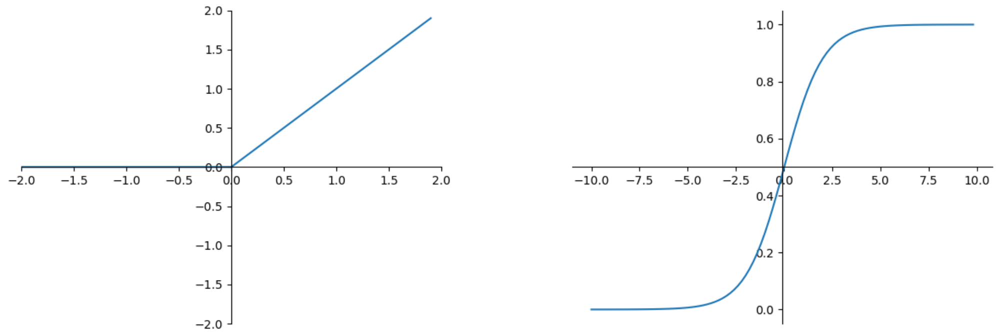


Fig. 31: Common nonlinear activation functions used in modern neural networks. **Left:** the Rectified Linear Unit (ReLU); **Right:** the sigmoid (σ) function.

layers and activation functions. However, no *learning* has happened yet – the flow of information has only been in one direction. A machine learning model must adapt its weight matrices and bias vectors in hopes of learning from its mistakes. We can picture this taking place in three (simplified) main steps:

1. The model receives a piece of input data and transforms it into an output.
2. The quality of the output is measured by comparing it to the correct output value.
3. The model adjusts its internal parameters based on the feedback it received about the quality of its outputs.

Our model is a bit like an idealized student with the following learning process:

1. The student receives a pop quiz, thinks about the questions, and takes their best guess at the answers.
2. The teacher collects the quiz, grades it, and hands it back to the student.
3. The student reflects on their mistakes, paying more attention to the costlier mistakes, and adjusts their thinking process to better prepare for the next quiz.

The quality of a machine learning's output given a specific input is measured with a **loss function**. The specific formulation of a loss function varies significantly based on the

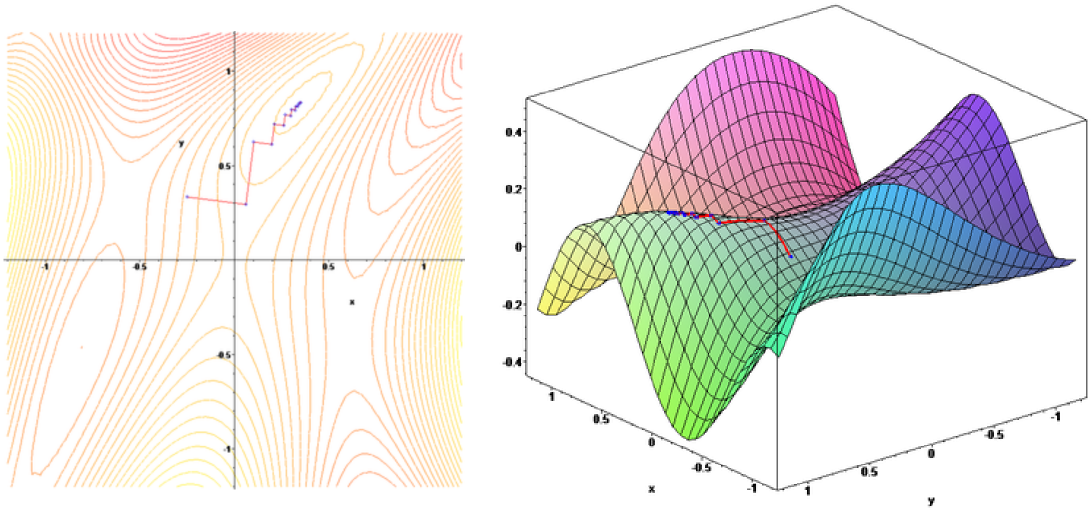


Fig. 32: An illustration of gradient descent in a challenging loss landscape (**left:** in 2D; **right:** in 3D) shows a trajectory towards a local minimum of the loss function. Step sizes decrease as the model gets closer to the local minimum and the gradient decreases. Public domain images from Wikimedia Commons.

task, but a basic example is *mean squared error* (MSE) loss commonly used for regression problems:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{67}$$

where N is the total number of data points, y_i is the model's output, and $\hat{y}_i$ is the true value, i.e. what the model would have guessed if it had perfect knowledge of the problem. A higher loss value, then, indicates that the model's prediction was very far from the target output. Poetically, the loss is also sometimes referred to as "regret".

Once equipped with the computed loss for a specific output (*forward propagation*, in the sense that information is flowing forwards, away from the input data and towards the output data), the model can move on to adjusting its trainable parameters (*backward propagation*, commonly called **backpropagation**, in the sense that information is flowing backwards from the output data and into the model itself). This means calculating the **gradient** of the loss function with respect to the model's trainable parameters θ: $\nabla_\theta \mathcal{L}(\theta)$. To understand the usefulness of the gradient, consider that minimizing the loss function is often a high-dimensional optimization problem not unlike a hiker walking through a complex landscape of hills and valleys (see Figure 32). A hiker trying to quickly descend to the lowest point in the landscape should take a step along the steepest path down from their current location. Likewise, the gradient of the loss function indicates the direction with the steepest increase in the loss value, and consequently $-\nabla_\theta \mathcal{L}(\theta)$ indicates the direction of steepest descent in the loss landscape. This process of taking steps in the direction of steepest descent and adjusting the model's parameters accordingly is called **gradient descent**. With step sizes parameterized by a *learning rate* γ, the parameters θ are adjusted as:

$$\theta \rightarrow \theta' = \theta - \gamma \nabla_\theta \mathcal{L}(\theta) \tag{68}$$

Exactly calculating the gradient is computationally demanding, as it necessitates using the entire training dataset, so many modern training methods incorporate a form of gradient descent that operates on *mini-batches* of data a portion at a time as a way of stochastically approximating the true gradient. This is usually coupled with an **optimizer** function that can dynamically adjust the learning rate γ to improve the training process.

The canonical neural network structure, also known as a Multi-Layer Perceptron (MLP), is a multi-layer **dense** (or "fully-connected") neural network, meaning each node in each layer is connected to every other node in its neighboring layers (as shown in Figure 30). The following sections will describe more advanced variations on this core structure.

3.3 Recurrent Neural Networks (RNNs)

The family of Recurrent Neural Networks (RNNs) are neural networks specially-formulated to process sequential data. This could mean timeseries data such as a stock price or other kinds of ordering, e.g. words in a sentence. They are optimized to retain contextual information about what values preceded a particular value in a sequence. Additionally, RNNs can process input data sequences of arbitrary lengths, though this doesn't necessarily mean that they will perform well on very long sequences.

An RNN retains contextual information via a stored *hidden state* that is updated throughout its training. The updating of this hidden state is regulated by *gates* that control the flow of information via their own activation functions. The RNN model described later in this thesis uses Long Short-Term Memory (LSTM) RNN layers, which are controlled

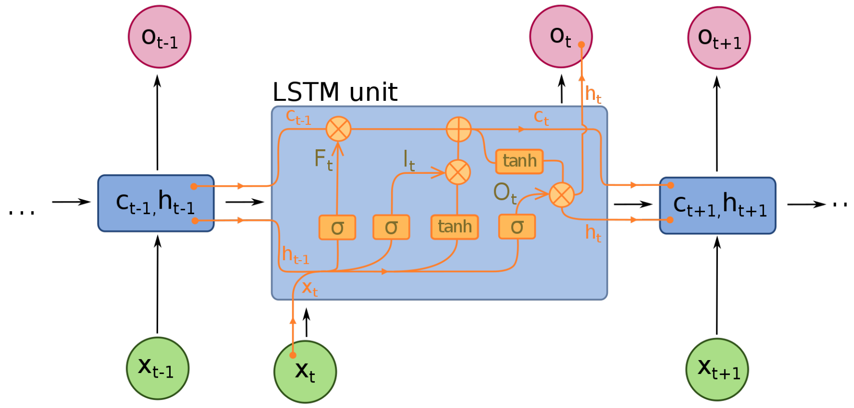


Fig. 33: A diagram of an LSTM unit, with information flowing from bottom to top. An input sequence (x_{t-1}, x_t, x_{t+1}, $\cdots$) is fed into the LSTM cell with input gate I_t, forget gate F_t, and output gate O_t, and the cell outputs a sequence (o_{t-1}, o_t, o_{t+1}, $\cdots$). (Public domain image from Wikimedia Commons).

by three gates: a forget gate, an input gate, and an output gate (see Figure 33). These gates determine what information should be stored for later decisions and what information has become irrelevant and should be removed from the internal hidden state. The basic operations of an LSTM layer are (as a function of the recurrent cell state c_t at time t, the weight matrix W, the bias vector b, and the hidden vector h, with $\circ$ denoting a Hadamard, or element-wise, product):

1. **Forget Gate:** Remove information that is no longer needed.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{69}$$

2. **Input Gate:** Choose what new information to consider.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{70}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{71}$$

3. **Update recurrent hidden state:** Incorporate new information chosen by input gate.

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{72}$$

4. **Output Gate:** Choose what parts of the internal cell state to output.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{73}$$

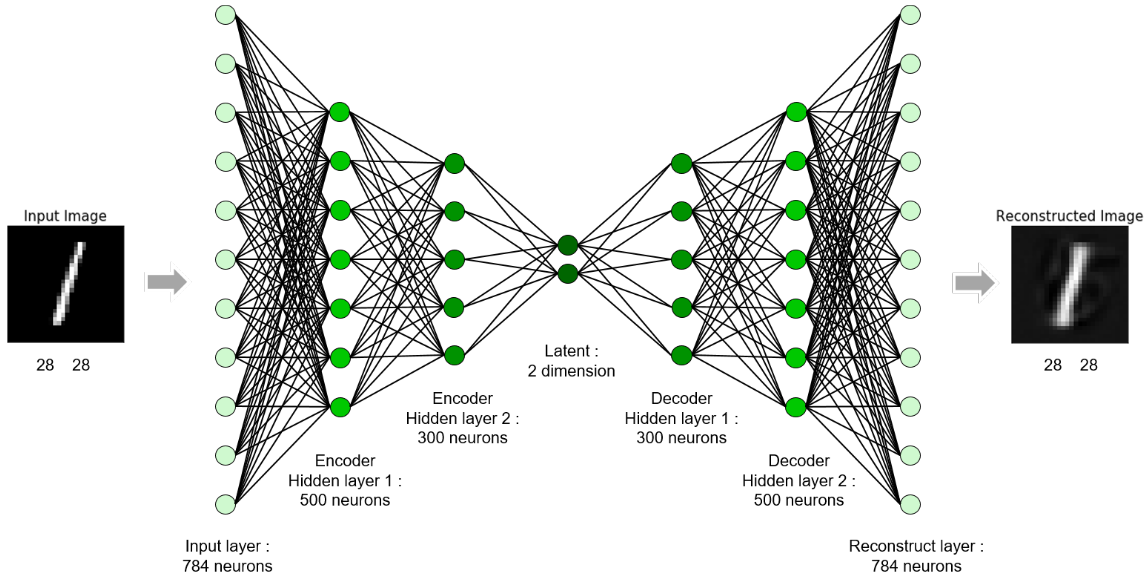


Fig. 34: An example diagram of an autoencoder designed to reconstruct input images of handwritten digits. The input image is split into 784 pixels and fed through the encoder, on the left, until the data has been transformed into an abstract 2-dimensional latent vector. Then, that latent vector is augmented through the decoder until the output matches the dimensions of the input. Once well-trained, the autoencoder's 2-dimensional latent space has learned a useful representation of the input data such that the image is able to be reconstructed fairly well using a starting point of just 2 dimensions instead of the original 784.

5. **Update hidden layer:** Apply output of the LSTM cell to the hidden layer.

$$h_t = o_t \circ \tanh(c_t) \tag{74}$$

3.4 Variational Autoencoders (VAEs)

Autoencoders are a class of neural networks designed such that the outputs closely mirror the inputs. Deep autoencoders typically consist of an **encoder** and a **decoder**, each of which contains several NN layers. The layer connecting the encoder and the decoder contains a latent representation of the data. By adjusting their internal parameters to efficiently imitate input data, autoencoders can learn useful representations of the data that can be exploited for other purposes. For example, autoencoders with a small latent layer in the middle of the network can be used as a dimensionality-reduction technique that is both nonlinear and invertible. Autoencoders designed for dimensionality-reduction usually have a characteristic "bow-tie" shape due to the reduced dimension of the latent layer forming an information bottleneck in the model architecture (see Figure 34).

Variational autoencoders (VAEs) are a special subcategory of autoencoders that are often used as generative models. Unlike a basic autoencoder with single values associated with each of its latent dimensions, variational autoencoders output a probability distribution associated with each of its latent dimensions. Commonly, these probability distributions are assumed to be Gaussian, and are therefore exactly described by vectors of μ (mean) and σ (standard deviation) values. The encoder portion of the VAE is designed to output a μ and σ value for each latent dimension, thereby constructing as many distinct Gaussian distributions as the size of the latent space. Sampling from these distributions then provides the input vector for the decoder. This procedure has the advantage of creating a latent space that is explicitly continuous, whereas a vanilla autoencoder might have large gaps in its latent space not covered by the training data.

The loss function of a VAE has two terms: the **reconstruction loss** and the **Kullback-Leibler (KL) loss**.

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{reconstruction}} + \beta \sum_i \text{KL}(q_i(z|x) \mid\mid \mathcal{N}(0,1)) \tag{75}$$

The reconstruction loss captures the accuracy of the model's outputs compared with the true target values, while the KL loss constrains the i latent distributions learned by the model to resemble unit Gaussian distributions with $\mu = 0$ and $\sigma = 1$. The KL loss term derives from the KL divergence, a quantity capturing the difference between two probability distributions. The KL loss is an important regularizer in the network that helps enforce continuity in the latent space, but it needs to be carefully balanced (by tuning the weight coefficient β) with the reconstruction loss in order to maintain good model performance.

In general, the KL divergence between two probability distributions $p(x)$ and $q(x)$ can be written as:

$$D_{KL}(p \mid\mid q) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{76}$$

Conveniently, when $p(x)$ is Gaussian and $q(x)$ is a unit Gaussian, this can be written in a closed-form expression parametrized by the means ($\vec{\mu}$) and standard deviations (the matrix Σ with $\vec{\sigma}$ along the main diagonal) of $p(x)$:

$$D_{KL}(\mathcal{N}(\vec{\mu},\Sigma)\ ||\ \mathcal{N}(0,\mathbb{I})) = \frac{1}{2}\sum_i(\Sigma + \mu^2 - 1 - \ln(\Sigma)) \tag{77}$$

In practice, Σ is often replaced with $\ln(\Sigma)$ when coding this term in the loss function, as exponentiating is more numerically stable than taking a logarithm.

The "variational" title for VAEs refers to the origin of this construction of the loss function from the application of variational inference to approximate the posterior $p(z|x)$, where z is a latent variable and x is an observation. This is a nice way of summarizing the process of a VAE: we want to understand the distribution of latent variables z that best describes our actual data, x. By Bayes' Theorem, the posterior is expressed as

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}. \tag{78}$$

However, calculating $p(x)$ is onerous and usually intractable, as it requires integrating over every possible configuration of the latent space. We can therefore use notions from variational inference to construct a new, tractable probability distribution $q(z|x)$ that resembles the original posterior $p(z|x)$ as closely as possible. This new distribution will allow us to approximate the intractable posterior. The requirement that $q(z|x)$ resemble $p(z|x)$ as much as possible can be translated into minimizing the KL divergence between these two distributions, i.e.:

$$D_{KL}(q(z|x)\ ||\ p(z|x)) \tag{79}$$

After applying principles from variational inference, we find that minimizing D_{KL} is equivalent to *maximizing* the Evidence Lower Bound (ELBO):

$$E_{q(z|x)} \log p(x|z) - D_{KL}(q(z|x)\ ||\ p(z)), \tag{80}$$

where $E_{q(z|x)} \log p(x|z)$ is the reconstruction likelihood. To translate the task of maximizing the ELBO into the training process, we flip the sign on each term and convert it into a loss term that the model tries to minimize. Thus, we arrive at the loss term described in Equation 75: negating the reconstruction likelihood creates a reconstruction loss term, while the second term adds the KL divergence term.

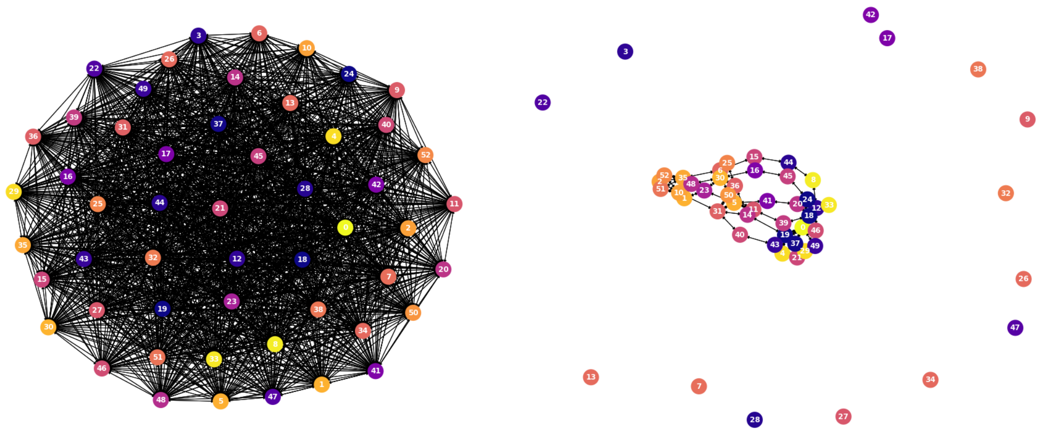


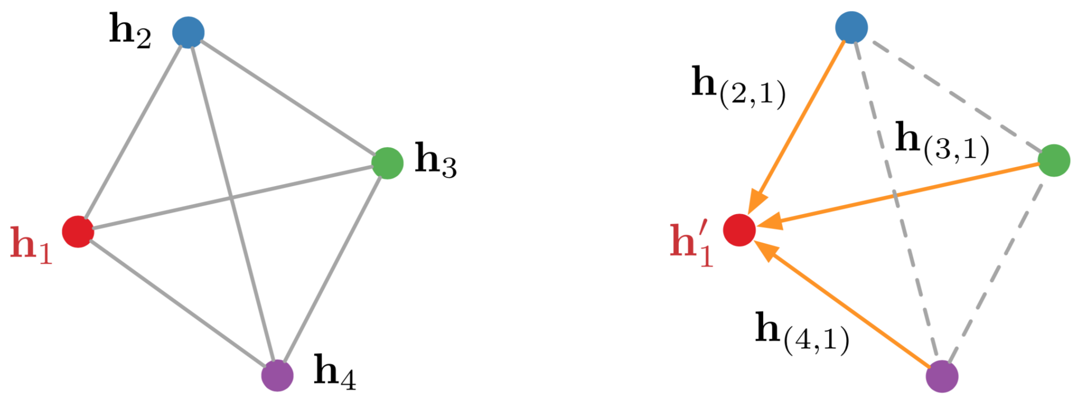
Fig. 35: Two examples of graph-structured data with 53 nodes. **Left:** A fully-connected graph, meaning there are edges connecting every pair of nodes. **Right:** A more sparsely-connected graph, meaning there are only a handful of edges connecting some nodes in the graph.

3.5 Graph Neural Networks (GNNs)

Graph neural networks are a class of neural networks designed to operate on graph-structured data such as social networks, knowledge graphs, or molecular structures. This means that the data takes the form of some finite number of vertices (or **nodes**) connected by links (or **edges**). Graph-structured data is sometimes sparsely-connected (see Figure 35), meaning only some nodes are connected via edges. In a directed graph, there is a specific orientation to these connections (i.e. Node 1 connects to Node 2, but not vice-versa). In an undirected graph, edges connect pairs of nodes in each direction.

Graph-structured data is complicated: meaningful connections might exist not only between a node and its neighbors, but also between a node and distant nodes on the graph, yet graphs can be arbitrarily large. Furthermore, there is no spatial locality enforced for graphs: one can contort a graph quite a bit, but as long as its connections remain unchanged, it remains unaltered. Other ML algorithms designed for Euclidean-structured data or sequential data are typically not suitable for use on graphs. Graph Neural Networks (GNNs) are designed specifically to handle some of the nuances necessary for analyzing graph-structured datasets.

A common form of GNN takes a graph-structured input $G = (V, E)$, where V is a set of nodes and E a set of edges, and learns a hidden representation of the graph that is repeatedly



(a) Node representations. **(b)** Message passing step.

Fig. 36: An example of message passing implemented on a fully-connected graph with four nodes. **Left:** Each of the four nodes has a hidden node representation h_i. **Right:** To update the hidden representation h_1 corresponding to Node 1, messages are aggregated from each incoming edge from the neighborhood of Node 1. [36]

updated via a method called **message passing**. These updates happen separately for edges, as a function of nodes x_i, edges $x_{(i,j)}$, and hidden node embeddings h_i:

$$h_{(i,j)} = f_{\text{edge}}(h_i, h_j, x_{(i,j)}) \tag{81}$$

and nodes:

$$h_i' = f_{\text{node}}(h_i, \sum_{j \in \mathcal{N}_i} h_{(j,i)}, x_i). \tag{82}$$

Both f_{edge} and f_{node} are themselves neural networks – usually MLPs.

The edge update step in the message passing algorithm creates a hidden representation of the information contained in each edge of the graph as a function of the values of the nodes connected by that edge. Subsequently, the node update step aggregates the edge messages incoming to a particular node from its neighborhood of immediately-connected nodes $\mathcal{N}_i$. This message passing scheme can be repeated such that the latent graph representation is updated based on information propagated from throughout the whole graph. Additional layers can then be added to e.g. aggregate the hidden node embeddings in order to classify the entire graph structure.

4 ML for Triggering on Hadronic Taus

4.1 Properties of Taus

The Higgs boson decay mode of interest in the signal process of this thesis work involves the most massive leptons: tau leptons, which I'll refer to in the future as simply **taus**. As a Generation III lepton, discovered after the electron and muon, the tau's name was chosen by Martin Perl and his collaborators from the Greek word $\tau\rho\iota\tau\nu$, meaning "third". Perl later received the 1995 Nobel Prize in Physics for the discovery of the tau.

As the reader might recall from Table 1, taus are fermions with electric charge $-e$, just like electrons, but with a mass of 1.78 GeV, exceeding 3,000 electrons. While an electron is stable, the average lifetime of the tau is approximately 3×10^{-13} seconds. This means that even taus moving at relativistic speeds will almost always decay before encountering the ATLAS detector. The average distance traversed by the tau is $d =$ (velocity)(time) $= (v)(\gamma\tau_\tau) = \beta c\gamma\tau_\tau$, where τ_τ is the tau's lifetime, c is the speed of light, γ is the Lorentz factor and $\beta = v/c$. In its rest frame, a tau's proper decay length is 87 μm. A tau produced at a substantial energy of 100 GeV would travel about 5 mm, but the ATLAS detector begins at a radius of about 33.25 mm. The tau's decay products are therefore the actual physics objects we analyze in order to understand the nature of the taus from which those products originated.

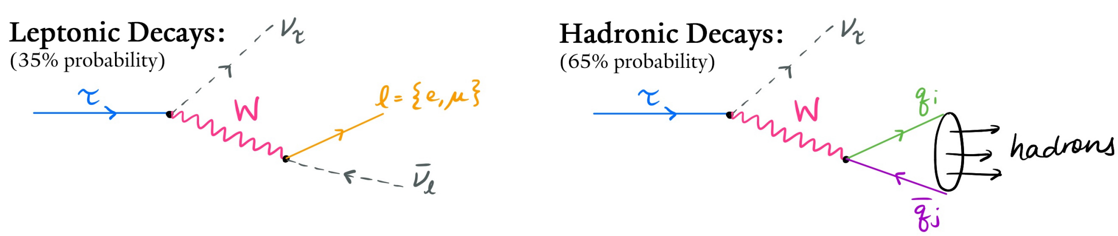


Fig. 37: Sample Feynman diagrams illustrating leptonic vs. hadronic tau decays. The leptonic decay modes include a W boson decaying to a lepton and anti-neutrino pair of the same lepton flavor (either e or μ). The hadronic decay modes include a coupling of the W boson to quarks (q_i and q_j) and, subsequently, a variety of hadrons. All tau decay modes feature a tau neutrino (ν_τ) in their final states in order to preserve lepton flavor.

Approximately 35% of the time, a given tau will decay **leptonically**, i.e. $\tau \to l\bar{\nu}_l\nu_\tau$, where $l = e, \mu$. In other words, the tau decays into a tau neutrino plus an additional lepton (electron or muon) and its corresponding antineutrino, thereby preserving charged lepton flavor, lepton number, and electric charge. This follows the same pattern of the muon, which has one dominant decay mode: $\mu \to e\bar{\nu}_e\nu_\mu$.

The remaining 65% of the time, the tau will decay into collections of mostly hadrons (composite particles consisting of two or more quarks). This special quality makes the tau unique among the leptons, as it is the only one massive enough to decay into hadrons. Though a tau neutrino will always be present in the final state due to lepton flavor conservation, and therefore one of the decay products will always be a lepton, I'll refer to these final states containing hadrons as **hadronic** tau decays for convenience. Example diagrams of each of these two categories of decays may be seen in Figure 37.

The tau's primary hadronic decay modes include 1 or 3 charged hadrons as well as potentially one or more neutral hadrons. Pions are the main hadrons appearing in hadronic tau decays, followed by kaons. This is because pions are mesons combining quarks of the same generation (up and down, i.e. Generation I) while kaons combine quarks between Generations I and II (i.e. up, down, and strange quarks). To understand this effect, we can look at the Cabbibo-Kobayashi-Maskawa (CKM) Matrix in theoretical particle physics, shown in Equation 83, that describes the strengths of flavor-changing weak interactions. Experimental results have confirmed that the quarks have much stronger interaction vertices for interactions within a single generation rather than between generations. Hadronic tau decays into pions, then, are referred to as "Cabibbo-favored", while hadronic tau decays into kaons are "Cabibbo-suppressed."

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.9745 \pm 0.0001 & 0.2245 \pm 0.0004 & 0.0037 \pm 0.0001 \\ 0.2244 \pm 0.0004 & 0.9736 \pm 0.0001 & 0.0421 \pm 0.0008 \\ 0.0090 \pm 0.0002 & 0.0413 \pm 0.0007 & 0.9991 \pm 0.00003 \end{pmatrix} \tag{83}$$

A hadronic decay mode with 1 charged hadron, i.e. one associated charged track, is called a **1-prong tau**, while a hadronic decay mode with 3 charged hadrons and three associated charged tracks is called a **3-prong tau**. Since the tracking portion of the ATLAS

Tau Decay Mode	Leptonic Decay	Hadronic Decay	Prongs	% Total Decays
$\pi^-\pi^0\nu_\tau$		✓	1-prong	25.4941 ± 0.0893
$\mu\bar{\nu}_\mu\nu_\tau$	✓			17.3937 ± 0.0384
$e\bar{\nu}_e\nu_\tau$	✓			17.8175 ± 0.0399
$\pi^-\nu_\tau$		✓	1-prong	10.8164 ± 0.0512
$\pi^-2\pi^0\nu_\tau$		✓	1-prong	9.2595 ± 0.0964
$\pi^-\pi^-\pi^+\nu_\tau$		✓	3-prong	8.9868 ± 0.0513
$\pi^-\pi^-\pi^+\pi^0\nu_\tau$		✓	3-prong	2.7404 ± 0.0710

Table 6: Summary of the primary tau decay modes [2].

detector is only sensitive to charged tracks, we only consider charged tracks for labeling tau decays. 1-prong taus represent 72% of all hadronic tau decays, while 3-prong taus represent 22% of all hadronic tau decays [37]. The most frequent hadronic decay mode is $\tau \to \pi^-\pi^0\nu_\tau$ via the ρ^- resonance, an excited state of a charged pion (see Table 6).

As leptonic tau decays into lighter leptons are largely under the purview of electron and muon reconstruction and identification, tau reconstruction and identification in this thesis refers to **hadronic taus only**. The main sources of background in the pursuit of hadronic tau reconstruction and identification are jets of hadrons initiated by the QCD hadronization processes of gluons and quarks. Additionally, electrons are a key background for 1-prong hadronic taus, as they leave similar signatures in the ATLAS tracker and calorimeters.

4.2 Tau Reconstruction

Hadronic tau candidates require an initial jet seed with $p_T > 10$ GeV and $|\eta| < 2.5$, excluding the crack region between the barrel and endcap calorimeters ($1.37 < |\eta| < 1.52$). Charged tracks falling in a cone of $\Delta R < 0.2$ around the jet's center of mass are considered associated with the jet seed so long as they have a $p_T > 1$ GeV, $|d_0| < 1$ mm, and $|z_0 \sin(\theta)| < 1.5$ mm.[8] Those tracks associated with the jet seed are defined to fall within the **core region** ($\Delta R < 0.2$), while any tracks just outside of the core region are defined to fall in the **isolation region** ($0.2 < \Delta R < 0.4$).

The tau's vertex is chosen as the vertex of the tracks in the core region containing the largest fraction of the tau's momentum. The tau's η and ϕ values are then calculated as the

[8] Recall that d_0 and $|z_0 \sin(\theta)|$ are defined as the points of closest approach from the transverse and longitudinal planes, respectively.

vector sum of the respective η and ϕ values of the tau's constituent TopoClusters within $\Delta R < 0.2$ of the seed jet's center of mass.

4.3 Tau Energy Scale Calibration

The reconstructed tau's mass is defined as exactly zero, meaning that the tau candidate's p_T and E_T are equivalent. Dedicated calibration schemes are applied to each visible tau candidate in order to best match the measured energy deposits from the calorimeter to the true visible energy of the original tau. By referring to a *visible* tau's energy here, we acknowledge that a hadronic tau decay has an invisible contributor to its total energy in its neutrino, but it is often useful for us to only focus on calibrating the energy of the visible tau decay products. First, the energies of the tau candidate's TopoClusters are calibrated using the Local Cluster (LC) calibration scheme designed to calibrate the energy of each cell based on the probability that the entire cluster is hadronic [38]. The sum of these calibarted energies is denoted E_{LC}. Next, there is a correction to the tau candidate's energy designed to subtract contributions to the tau's energy from pileup, i.e. other interactions from the same bunch crossing. This is done by subtracting E_{pileup} from E_{LC}, where E_{pileup} scales linearly with the difference between the event's number of pileup vertices and the average pileup $(N_{\mathrm{PV}} - \langle N_{\mathrm{PV}} \rangle)$ in bins of $|\eta|$. Then, a response calibration $\mathcal{R}$ is calculated as the average of the corrected tau candidate energy distribution $(E_{\mathrm{LC}} - E_{\mathrm{pileup}})/E_{\mathrm{true}}^{\mathrm{vis}}$.

The baseline tau energy scale correction is calculated as:

$$E_{\mathrm{calib}} = \frac{E_{\mathrm{LC}} - E_{\mathrm{pileup}}}{\mathcal{R}(E_{\mathrm{LC}} - E_{\mathrm{pileup}}, |\eta|, n_p)} \tag{84}$$

This information, along with additional tracking and calorimeter information, is then fed into a Boosted Regression Tree (BRT) designed to output the final energy of the hadronic tau candidate.

4.4 Tau Reconstruction in the ATLAS Trigger

The ATLAS tau trigger is optimized to perform the difficult task of separating hadronic tau decays from high-rate quark- or gluon-initiated jet backgrounds of a given energy (see

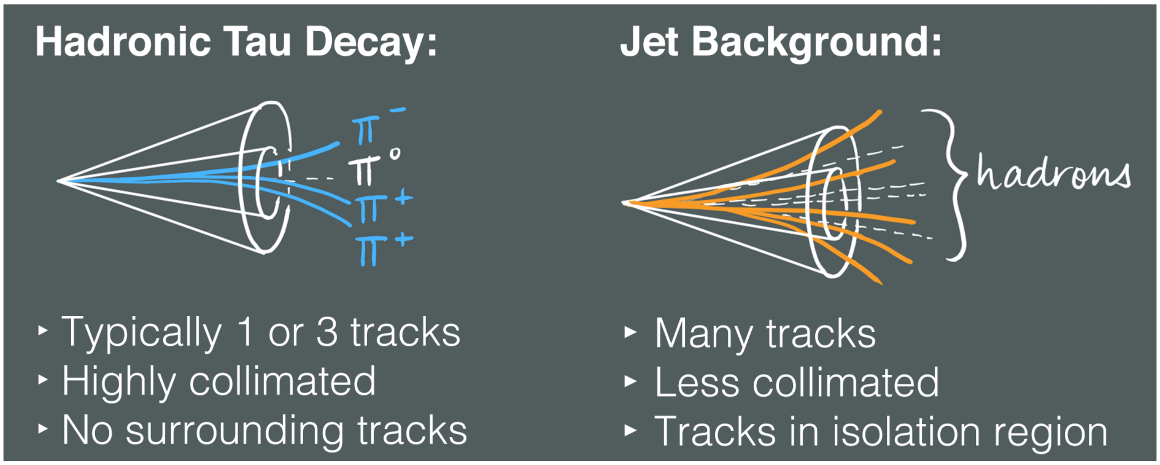


Fig. 38: A diagram of typical signatures of a hadronic tau decay versus a quark- or gluon-initiated jet used to inform hadronic tau identification strategies in the tau trigger. Hadronic taus have constrained numbers of charged tracks from charged hadrons due to charge conservation from the tau parent, and these tracks tend to be collimated in the core region ($\Delta R < 0.2$) surrounding the reconstructed tau's center of mass for tau candidates with the same E_T. Quark- or gluon-initiated jets do not have the same charged track number constraints, and depending on their origin, the tracks can tend to be less collimated than those of a hadronic tau. This means that some charged tracks may appear in the isolation region ($0.2 < \Delta R < 0.4$) surrounding the core.

illustrations in Figure 38). Hadronic tau decays mainly feature track multiplicities of 1 or 3 in the core region ($\Delta R < 0.2$ from the reconstructed tau cluster's center of mass) and no tracks in the isolation region ($0.2 < \Delta R < 0.4$) surrounding the core. Quark- and gluon-initiated jets, on the other hand, can feature many tracks with a more even distribution throughout the core and isolation regions.

Though the tau trigger follows similar methodologies as offline tau reconstruction when possible, it operates under unique constraints. The tau trigger must operate within the latency budget of the trigger and data acquisition system, and it does not have access to the full granularity of the ATLAS detector. The processes of energy calibration, calorimeter clustering and track-finding therefore differ slightly from the offline methods. The ATLAS tau trigger also follows a similar tau identification scheme as is used offline, but special requirements at the trigger level necessitated a dedicated architecture and training scheme (detailed in Section 4.6) [39].

4.4.1 Calibration, Clustering, and Tracking

At Level 1, the ATLAS calorimeter trigger towers have a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and a coverage of $|\eta| < 2.5$. (For comparison, as shown in Figure 23, the middle layer of the EM calorimeter has a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.) The "core" region for a tau seed in the EM calorimeter is defined by a 2×2 square of towers, while the "isolation" region consists of the ring of surrounding trigger towers between $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ and $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ (see Figure 39). After a minimum E_T cut at Level 1 for the core calorimeter towers based on the Level 1 trigger threshold, there is an energy-dependent E_T cut in the isolation region defined up to 60 GeV:

$$E_T^{\text{EM isol}} \leq \frac{E_T}{10\ \text{GeV}} + 2. \tag{85}$$

Calorimeter clustering in the software-based ATLAS High-Level Trigger (HLT) again uses calibrated TopoClusters of calorimeter cells, but they are instead derived from a cone of $\Delta R < 0.2$ surrounding the Level 1 tau seed, which can sometimes differ from the offline seed as Level 1 only has access to lower-granularity information. Energy calibration in the ATLAS HLT for tau candidates resembles the methods used offline, but is customized for the trigger environment. For example, instead of using N_{PV}, or the number of pileup vertices in the event, the average number of interactions per bunch crossing μ is used for the pileup subtraction. An HLT-level calorimeter-only preselection is also applied at this stage, including a minimum p_T cut.

Tracking in the ATLAS HLT is done in two steps. First comes a preselection called the Fast Track Finder (FTF). The FTF searches for a track in a narrow cone ($\Delta R < 0.1$) around the cluster's center of mass and along the full beamline ($|z| < 225$ mm). This establishes the leading track for the tau candidate, and if no track is found, the candidate is rejected. Next, the FTF reconstructs all tracks within a larger angular region ($\Delta R < 0.4$) but a more restricted range along the beamline ($|z| < 10$ mm), as illustrated in Figure 40. The track seeds identified in the FTF process are then passed onto the second stage of tracking: precision tracking. During precision tracking, a more comprehensive tracking is applied to the track seeds identified at the FTF stage.

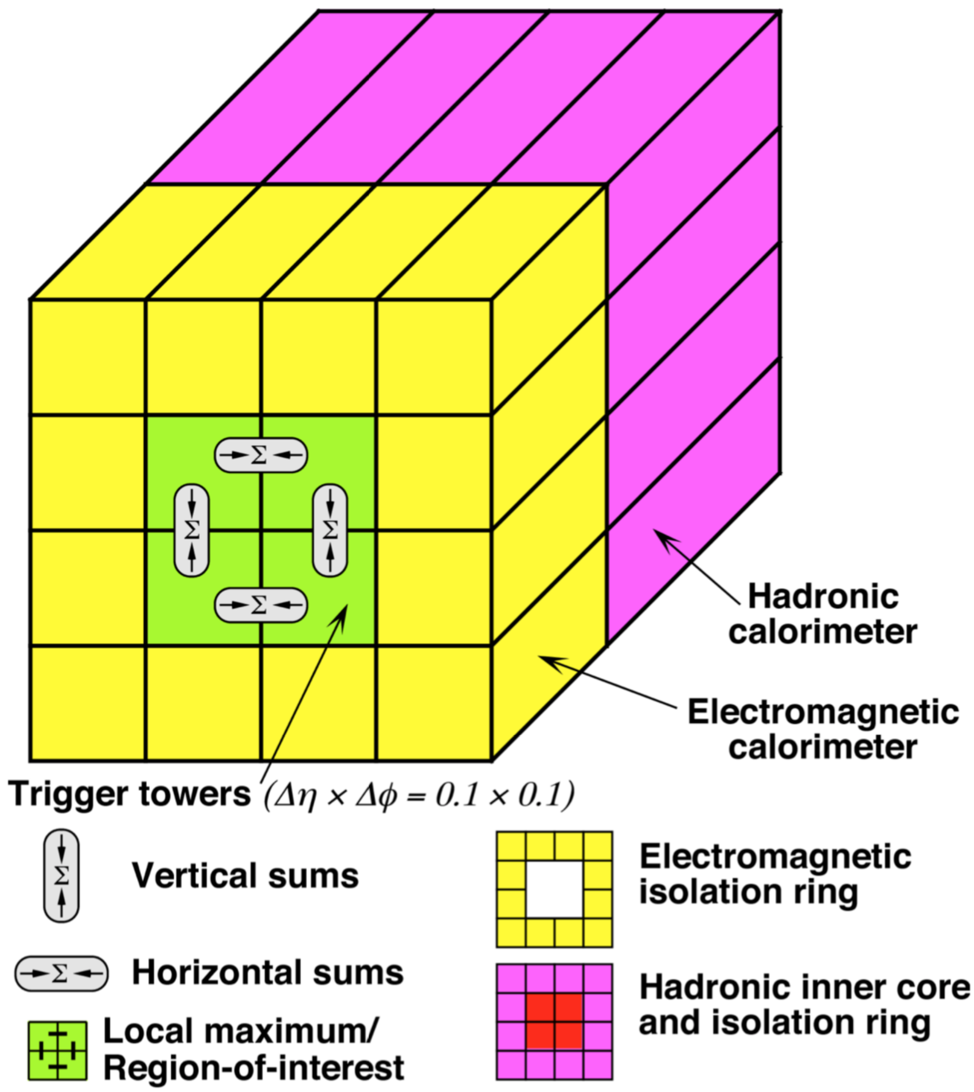


Fig. 39: The Level 1 ATLAS calorimeter towers, with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ and a coverage of $|\eta| < 2.5$, are shown. The core region of a tau candidate at Level 1 of the trigger is shown in lime green, with its surrounding isolation region in the Electromagnetic and Hadronic calorimeters shown in yellow and pink, respectively. [40]

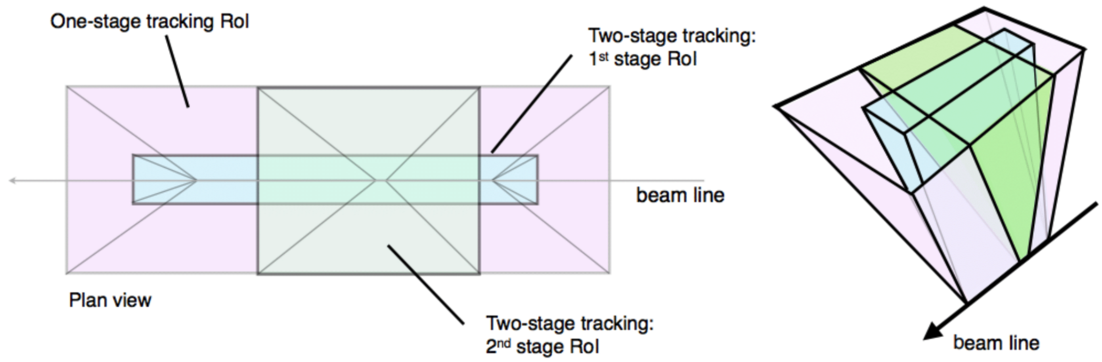


Fig. 40: The geometric ranges searched during the Fast Track Finder (FTF) in the first portion of tracking in the ATLAS HLT are shown. The pink region shows the region of interest (RoI) identified at Level 1. In blue is the first range searched: $\Delta R < 0.1$ and $|z| < 225$ mm. In green is the second range searched: $\Delta R < 0.4$ and $|z| < 10$ mm.

Finally, a cut is applied on the number of tracks identified in the core and isolation regions of the tau candidate. A tau candidate must have between 1 and 3 (inclusive) tracks in the core region and 1 or 0 tracks in the isolation region to pass onto the tau identification stage.

4.5 Online BDT Tau Identification

Note: The word *online* here indicates that the process occurs at the trigger level during live datataking, as opposed to offline analysis that need not occur during datataking.

At the start of Run 2, the online tau identification scheme was based on a Boosted Decision Tree (BDT) trained on a fixed number of high-level tau identification variables detailed in Table 7 [39]. Two separate BDT models were trained for 1-prong and 3-prong taus using simulated $Z \to \tau\tau$ signal events and QCD multijet background events. These events were required to have at least one reconstructed tau candidate that passed the full offline selection, not including the final BDT identification criteria, and a total missing $E_T \leq 20$ GeV, to suppress any events from $W \to l\nu$+jets processes. An independent set of signal and background simulations was used to evaluate the performance of the model.

The trained BDT model maintained an efficiency of 96% for true 1-prong taus and 82% for true 3-prong taus that also passed offline reconstruction as a baseline *medium* working

Variable	Description	1-prong	3-prong
f_{cent}	Central energy fraction	✓	✓
$f^{-1}_{\text{leadtrack}}$	Leading track momentum fraction	✓	✓
R_{track}	Track radius	✓	✓
$f^{\text{track-HAD}}_{\text{EM}}$	Fraction of EM energy from charged pions	✓	✓
$f^{\text{EM}}_{\text{track}}$	Ratio of EM energy to track momentum	✓	✓
$m_{\text{EM+track}}$	Mass of track + EM system	✓	✓
$p_T^{\text{EM+track}}/p_T$	Ratio of track + EM system p_T to tau p_T	✓	✓
$\lvert S_{\text{leadtrack}}\rvert$	Leading track impact parameter significance	✓	
$f^{\text{track}}_{\text{iso}}$	Fraction of p_T from tracks in isolation region	✓	
$\Delta R_{\max}$	Maximum ΔR		✓
S_T^{flight}	Transverse flight path significance		✓
m_{track}	Track mass		✓

Table 7: The high-level input variables used for BDT tau identification in the ATLAS HLT at the start of Run 2, as well as whether they were used for the 1-prong model, the 3-prong model, or both.

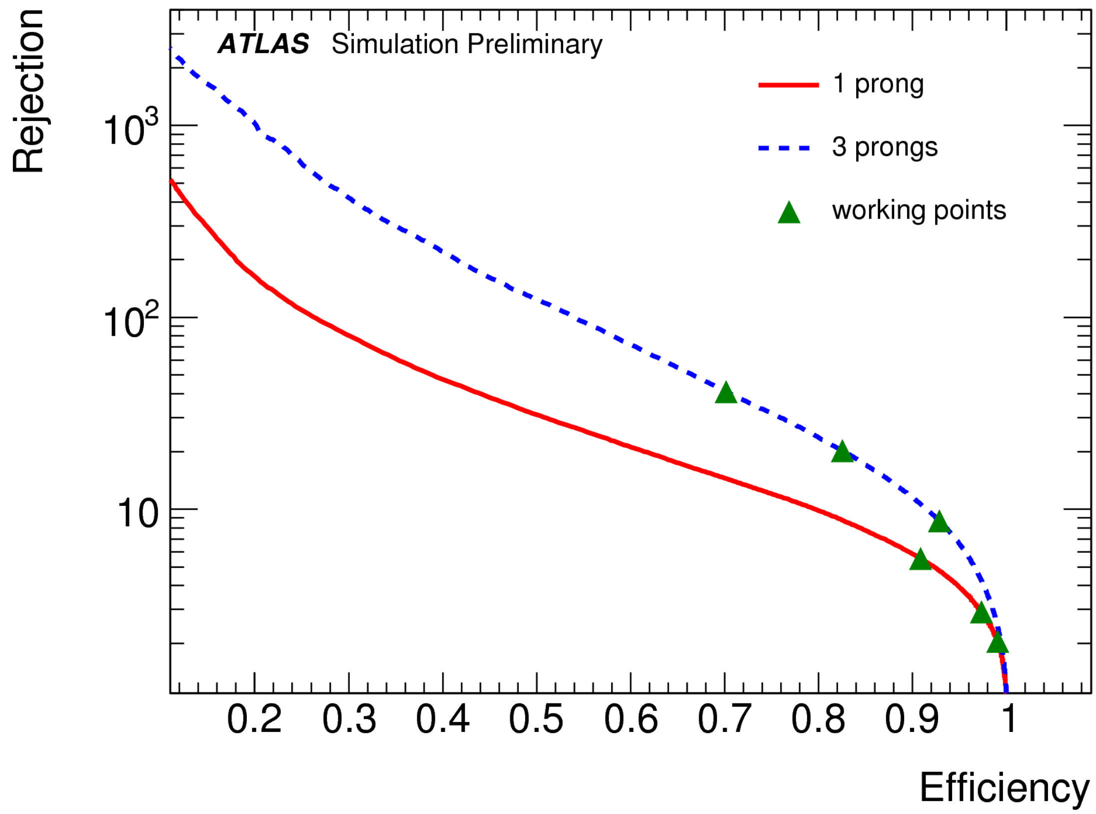


Fig. 41: Background rejection vs. signal efficiency curves for the 1-prong (solid red line) and 3-prong (dashed blue line) tau identification BDT models used in the ATLAS HLT at the beginning of Run 2. The *tight*, *medium*, and *loose* working points are marked with green triangles in order of decreasing rejection for each model. [39]

point, as shown in Figure 41. It operated with an average latency of less than 1 ms, or $\sim 1\%$ of the total execution time of the tau HLT.

4.6 Online RNN Tau Identification

Online RNN tau identification used a very similar architecture as the RNN tau identification scheme designed for offline tau identification in the latter part of Run 2 [41]. I will describe the methodology here and point out key differences from the offline construction.

4.6.1 Data Pre-Processing

MC simulations of $\gamma^* \to \tau\tau$ signal and dijets background (binned in p_T ranges between 0 and 1,800 GeV) were used for training and testing. Samples were converted into MxAOD ROOT formats with trigger-level variables included using the ATLAS internal software `THOR`. Then, these ROOT files were converted into flat Pandas dataframes [42] stored in HDF5 binary file formats using PyROOT, i.e. ROOT's Python-C++ bindings [43], and `root_numpy` [44]. Samples used trigger-level variables from the HLT chain `HLT_tau25_idperf_tracktwoEF`, which provides events with additional trigger-level tracking information saved. The specific MC samples used were produced with Pythia8 and the NNPDF23LO PDF set.

Each of the three RNN models applied the following cuts to tau candidates:

- **0-prong**
  - Reconstructed tau $p_T > 20$ GeV
  - Reconstructed tau $|\eta| < 2.5$
  - True tau $p_T > 20$ GeV (signal only)
  - True tau $|\eta| < 2.5$ (signal only)
  - # of charged tracks = 0
  - # of true charged tracks = 1 or 3 (signal only)
- **1-prong**
  - Reconstructed tau $p_T > 20$ GeV
  - Reconstructed tau $|\eta| < 2.5$
  - True tau $p_T > 20$ GeV (signal only)
  - True tau $|\eta| < 2.5$ (signal only)
  - # of charged tracks = 1 or 3
  - # of true charged tracks = 1 (signal only)
  - Truth-matched (signal only)
- **Multi-prong**
  - Reconstructed tau $p_T > 20$ GeV

- Reconstructed tau $|\eta| < 2.5$
- True tau $p_T > 20$ GeV (signal only)
- True tau $|\eta| < 2.5$ (signal only)
- # of charged tracks = 2 or 3 if tau $p_T < 440$ GeV, otherwise ≥ 2 tracks
- # of true charged tracks = 3 (signal only)
- Truth-matched (signal only)

While the offline RNN tau identification trainings reweighted the background dijet p_T spectrum to match the signal $\gamma^* \to \tau\tau$ p_T spectrum in bins of p_T during training and evaluation, in the online implementation, I reweighted the signal to match the background p_T spectrum in bins of p_T after training to determine same-rejection working points, as the tau trigger rates are dominated by low-p_T jets.

The input variables used in the online RNN trainings are listed below, and were scaled to have a mean of 0 and a standard deviation of 1. In some cases, the logarithm of the variable is also used in its place in order to better capture dynamic differences across multiple orders of magnitude. The variables used are almost the same between online and offline except for `ptIntermediateAxis` (vertex-corrected tau axis starting from the calorimeter cluster center-of-mass), which is replaced with `ptDetectorAxis` (uncorrected tau axis starting from the calorimeter cluster center-of-mass) online, and variables related to the counting of track hits:

- `nInnermostPixelHits` (# of innermost pixel hits) → `nIBLHitsAndExp` (if an IBL hit is expected, use the number of innermost pixel hits; otherwise, use 1)
- `nPixelHits` (# of pixel hits) → `nPixelHitsPlusDeadSensors` (# of pixel hits + # of dead pixel sensors)
- `nSCTHits` (# of SCT hits) → `nSCTHitsPlusDeadSensors` (# of SCT hits + # of dead SCT sensors)

These changes were made to protect against varying detector conditions and unnecessarily eliminating tau track candidates with no track hits in the IBL.

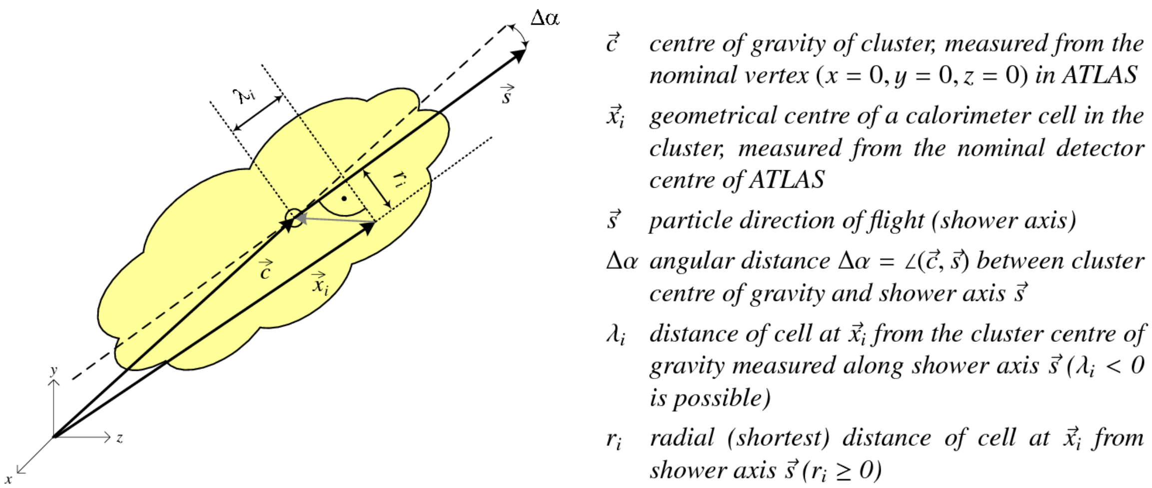


Fig. 42: A diagram of an example calorimeter cluster with its corresponding geometrical cluster moments related to λ_{cluster} and r_{cluster}. [45]

4.6.2 RNN Input Variables

The input variables into the RNN include both low-level and high-level quantities relating to tracking and calorimetry information. The same 12 high-level tau identification variables used as inputs to the BDT (see Table 7) are included. Track-level variables include basic, low-level quantities like p_T of the tau candidate, p_T of the jet seed, points of closest approach to the tau candidate from the transverse and longitudinal planes ($|d_0|$ and $|z_0 \sin(\theta)|$), angular distance from each track to the axis of the tau candidate ($\Delta\eta$ and $\Delta\phi$), and information about the quality of each track based on track hits. Cluster-level variables include basic information such as E_T, p_T of the jet seed, angular distance from each cluster to the axis of the tau candidate ($\Delta\eta$ and $\Delta\phi$), and cluster moments capturing geometric information about the clusters including λ_{cluster}, $\langle\lambda^2_{\text{cluster}}\rangle$, and $\langle r^2_{\text{cluster}}\rangle$. These represent the spread of the cluster along the cluster axis and perpendicular to the cluster axis, as shown in Figure 42.

- **Track-level variables**
  - `pt_log`: Log of track p_T
  - `pt_jetseed_log`: Log of jet seed p_T
  - `d0_abs_log`: Log of track $|d_0|$

- `z0sinThetaTJVA_abs_log`: Log of track $|z_0 \sin(\theta)|$
- `dEta`: $\Delta\eta$ between track and tau axis
- `dPhi`: $\Delta\phi$ between track and tau axis
- `nIBLHitsAndExp`: If a hit in the innermost pixel layer is expected, use the actual number of IBL hits. If not, set the number of IBL hits = 1.
- `nPixelHitsPlusDeadSensors`: Number of pixel hits in the tracker + number of dead sensors
- `nSCTHitsPlusDeadSensors`: Number of SCT hits in the tracker + number of dead sensors

- **Cluster-level variables**
  - `et_log`: Log of cluster E_T
  - `pt_jetseed_log`: Log of jet seed p_T
  - `dEta`: $\Delta\eta$ between cluster and tau axis
  - `dPhi`: $\Delta\phi$ between cluster and tau axis
  - `CENTER_LAMBDA`: λ_{cluster}, the distance from the cell to the cluster center of mass along the cluster axis
  - `SECOND_LAMBDA`: $\langle\lambda^2_{\text{cluster}}\rangle$, the moment of inertia of the cluster along the cluster axis
  - `SECOND_R`: $\langle r^2_{\text{cluster}}\rangle$, the moment of inertia of the cluster perpendicular to the cluster axis

4.6.3 RNN Architecture

Unlike the BDT, which recieved a fixed number of high-level input variables for each tau candidate, the Recurrent Neural Network (RNN) architecture treats the problem of tau identification as a sequence-classification problem. It receives a variable number of tracks and clusters, sorted in descending order in p_T and E_T respectively, associated with each tau along with the fixed set of tau identification variables. This is analogous to tasks encountered in Natural Language Processing (NLP) such as categorizing the sentiment of sentences with a variable number of constituent words in each sentence. This RNN architecture used Long Short-Term Memory (LSTM) layers.

The general RNN architecture is split into three branches: one for track inputs, one for cluster inputs, and one for tau identification variable inputs. The tau identification variables, having a fixed size for each tau candidate, do not need the advantages of the recurrent layers, and instead are fed into a series of fully-connected, or "dense", layers of sizes (128, 128, 16). Track and cluster inputs, on the other hand, are fed separately into branches consisting of two dense layers with shared weights of size 32 units each, thereby forming latent representations of the tracks and clusters. These latent representations are then passed to two recurrent LSTM layers of size 32 for tracks and 24 for clusters that map the input sequences of tracks or clusters into a single latent vector each. The output vectors of each branch are then concatenated together via a Merge layer, and this concatenated state is passed through a final set of three dense layers of sizes (64, 32, 1) to result in a final output size of 1 unit. This final unit, after transformation under a sigmoid activation function, represents the final RNN score for the input tau candidate. ReLU activation functions are used throughout, with the exception of the output layer. The full model has approximately 56,000 trainable parameters.

Though the offline RNN tau identification scheme consists of two separate trained models for 1-prong and 3-prong taus, I trained three different models for "0-prong", 1-prong, and "multi-prong" taus. "0-prong" refers to a true 1-prong tau for which the initial tau seed track has been misidentified, while "multi-prong" refers to a true 3-prong (or higher) tau for which at least one charged track has been poorly reconstructed. These 0-prong taus, at the trigger level, represent an important cause of signal inefficiencies, particularly for taus with low p_T in a high-pileup environment.

4.6.4 Training Details

The number of tracks and clusters used per tau candidate was variable but capped at 10 and 6 respectively, as it was determined that increasing the numbers beyond this point did not yield a significant benefit in terms of performance.

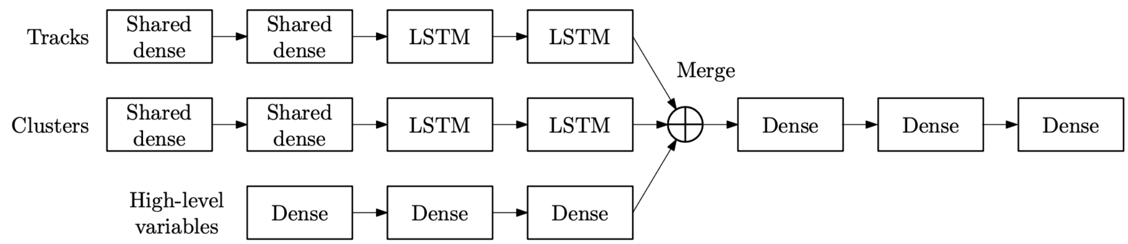


Fig. 43: A diagram of the three branches included in the overall RNN architecture for tracks, clusters, and tau identification variables. [41]

The models were trained using stochastic gradient descent (SGD) with an initial learning rate of 0.01, momentum of 0.9, and minimizing the binary crossentropy loss term. The sample sizes used for online trainings had fewer statistics than the versions used for offline, especially at low p_T:

- **0-prong:** Signal = ∼100,000 events, Background = ∼50,000 events
- **1-prong:** Signal = ∼2,000,000 events, Background = ∼175,000 events
- **2-prong:** Signal = ∼700,000 events, Background = ∼3,000,000 events

Initial samples were split between training/validation samples at a ratio of 80%/20%. Each model was trained until convergence, with early stopping induced after 10 consecutive epochs with no improvement to the validation loss. Trainings took up to approximately 6 hours on an NVIDIA Tesla p100 GPU. Following trainings, the output scores were transformed in order to be invariant across μ (average number of interactions per bunch crossing) and p_T. These models were implemented in `Keras` [46] with a `TensorFlow` backend [47]. Trained models were converted into a format suitable for use in a C++ production framework using `lwtnn` [48].

4.6.5 Performance

These trigger-level RNN tau identification models were implemented in the ATLAS HLT in July 2018 and operated with an average latency at the High-Level Trigger comparable to that of the BDT. Each RNN model saw improved background rejection vs. signal efficiency compared to the 1-prong and 3-prong BDT models previously implemented at the trigger

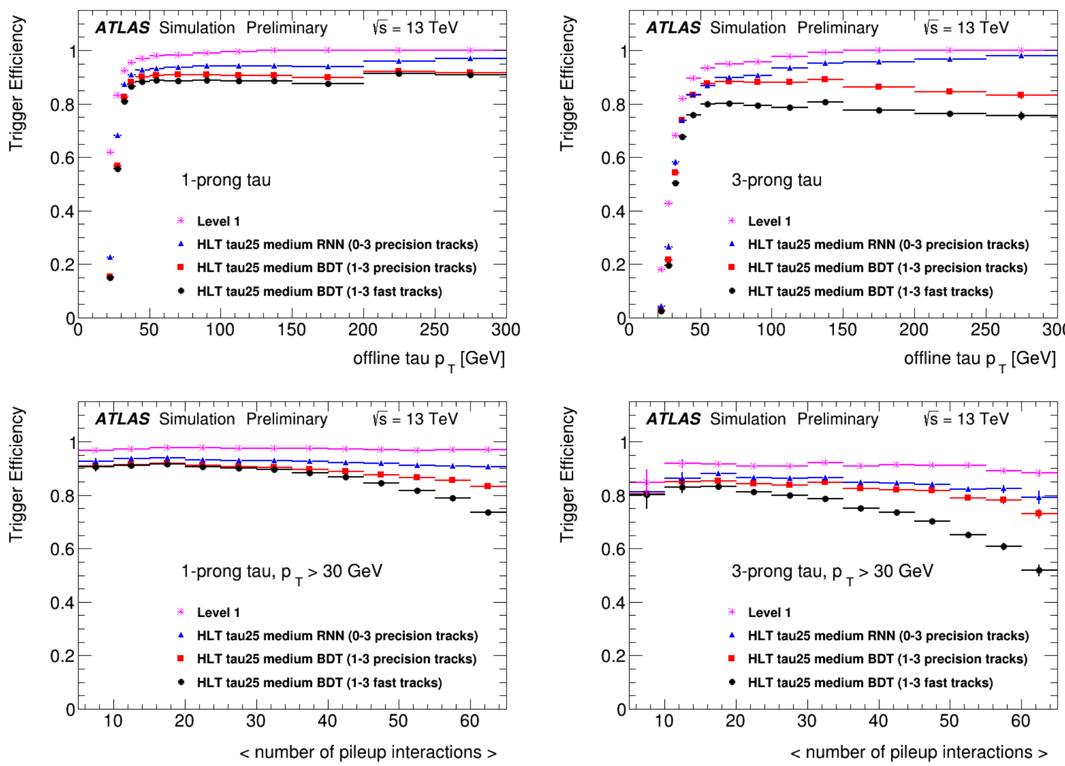


Fig. 44: Trigger efficiencies in MC simulation samples at $s = \sqrt{13}$ TeV for tau triggers with comparable rates. In the left column are efficiencies for 1-prong taus, while in the right column are efficiencies for 3-prong taus. In the top row, the $\hat{x}$-axis is p_T of offline taus passing the *Medium* BDT working point. In the bottom row, the $\hat{x}$-axis is μ, the average number of pileup interactions. The triggers using RNN tau identification (blue triangles) show better performance across each of these criteria compared to triggers using BDT tau identification with precision tracks (red squares) and fast tracks (black circles), even approaching the efficiency of the Level-1 trigger (pink crosses) for large values of offline tau p_T.

level. Overall, particularly due to the recovery of true 1-prong taus with a misreconstructed track thanks to the 0-prong RNN network, the tau trigger increased its background rejection by approximately 35%.

Tau efficiencies at the trigger level are shown versus p_T of offline taus passing the *Medium* BDT identification working point and versus μ (average number of pileup interactions) in Figure 44 for triggers with comparable rates. The triggers using RNN tau identification have excellent efficiencies compared with their BDT equivalents across a wide range of offline tau p_T values.

4.6.6 Future Directions

This RNN method will be the default tau identification algorithm in the ATLAS HLT in Run 3. Future iterations of tau identification for the HL-LHC will have to contend with much higher pileup contamination from quark- and gluon-initiated jets, potentially up to an average pileup level of $\langle\mu\rangle = 200$ [49]. This high-pileup environment will create additional challenges for the HLT tau tracking algorithms and could result in significantly degraded performance for low-p_T tau candidates. Not only will the RNN tau identification models need to be re-trained, potentially on data instead of MC, but they might need to be re-imagined entirely. This could involve innovative Graph Neural Network (GNN)-based tracking models implemented at Level 1 on FPGAs [50], Lorentz-equivariant networks trained on tau candidates [51], or perhaps a more unified approach to tau reconstruction and identification. Without a doubt, advanced machine learning models will be essential to identifying taus at the trigger level throughout the remaining lifetime of the LHC.

5 ML for Dynamic Many-Body Systems

5.1 Motivation

Throughout this thesis, I have deliberately avoided using the term *artificial intelligence*, or AI, in favor of the more constrained vocabulary of *machine learning* (ML). The concept of AI lacks a precise definition even among academics engaged in computer science research, making it a somewhat unhelpful form of scientific jargon. Additionally, depictions of AI in popular media can misleadingly suggest that what a scientist calls "AI" these days has something resembling consciousness or human-like intelligence. The models I describe in this thesis have a level of consciousness and intelligence much more comparable to a microwave oven than to HAL 9000 or R2-D2. That said, we can still attempt to think deeply about what it might mean for a model to exhibit artificial intelligence as a way of inspiring new types of model architectures and research paradigms.

Artificial intelligence is sometimes splintered into more specific pieces: narrow AI and general AI. **Narrow AI** is essentially analogous to what I call machine learning: using software algorithms to analyze and learn patterns from data in a semi-autonomous manner, meaning that the models themselves can automatically adjust to find optimal solutions under certain constraints. Narrow AI models are designed to solve specific tasks within a narrow scope of applicability. **General AI**, on the other hand, refers to a speculative kind of AI model that would exhibit human-like problem-solving skills, reasoning, and even consciousness. What would it take to construct something resembling a human mind out of code? It's a humbling and compelling question, and more importantly, one that should be approached carefully, as it has the potential to cause real-world harm by strengthening preexisting prejudices and injustices in human society.

The field of computational creativity is engaged in the pursuit of General AI by practically considering what it would mean for an algorithm to be creative. Creativity, even in humans, is poorly understood and is often thought of as intuitive or possibly beyond description. There are rich discussions to be had in the intersection between (Narrow or General) AI and creativity that challenge the capabilities of our algorithms just as much as they challenge our own understanding of art and the human mind. For these reasons and more, I find research directions in computational creativity, or broadly what I call **Creative ML**, to be particularly important and stimulating. Just as CERN was founded with a mission of peace, my vision for Creative ML is fundamentally tied to the betterment of humanity. Research in Creative ML can be inspired not only by profit, but by pure curiosity in service of an artistic practice. Additionally, Creative ML is just as interested in lateral progress, i.e. disrupting one research direction by splintering it into many sub-questions, as it is in forward progression towards state-of-the-art metrics along preexisting paths. I believe this impulse is essential for making meaningful progress in ML that is orthogonal to the strong incentives from surveillance capitalism that ultimately benefit corporations and frequently reinforce social disparities based on race, class, gender, and other stratifiers ([52, 53, 54, 55], etc.). Through the lens of Creative ML, we can ask expansive questions at the cutting edge of technological progress without compromising critical conversations about how these advances are shifting power in the real world or if they should exist at all.

Throughout my PhD, I led multiple independent research teams engaged in questions relating to ML and human movement – more specifically, understanding my own dancing and embodied thought process with ML. Using a motion capture studio, I collected data of myself performing solo improvisations, and together with my research collaborators, learned latent representations of my movements in order to extract larger patterns and meaning from that data. Though we have published our methods, code, and data in the spirit of openness and collaboration, these projects really emerged from an introspective impulse. They were designed for my own creative use, in hopes of building tools to help me continue to innovate in my specific movement practice.

These models are clearly instances of Narrow AI, and moreover were designed to be explicitly useful for my own art practice only. I am including them in this thesis, ironically, because of their generality. What began as a deeply personal investigation has clear potential for applications far beyond generating movements that resemble my own. Section 5.2 describes a generative model for time-dependent many-body systems in 3D, while Section 5.3 describes a methodology for discovering categories of interactions and graph structures within those same systems. What is listed here as a method for understanding cross-body interactions could, with just a different dataset, transform into a multi-particle tracking system within the ATLAS detector.

By orienting the models' successes towards my own creative instincts instead of external reward structures, and by training and applying them on my own body alone, I have tried to build these tools to ask big questions while minimizing possible unintended harms. ML tools to estimate 3D human poses from video alone have dramatically improved in recent years, meaning that these kinds of models are becoming available to a much broader audience than those who happen to have access to a state-of-the-art motion capture studio. This availability is especially concerning given recent results showing that even supposedly "anonymized" datasets of human movement can be quickly de-anonymized. This means that a unique individual can be identified out of a pool of more than 500 anonymized participants with $> 95\%$ accuracy using less than 5 minutes of their movement data [56]. I believe we can build ML tools like the ones detailed in this chapter for diffuse applications designed to enrich the lives of individuals, but great care must be taken in this direction to prevent abuses related to privacy, surveillance, and human rights.

I hope this chapter serves as proof that artistic questions can be just as fruitful as traditionally scientific ones when wading through the murky waters of technological research towards something like General AI, and that engaging with these diverse perspectives is essential if we are to truly build AI tools that serve everyone.

5.2 Beyond Imitation: Generative & Variational Choreography with VAEs

The contents of this chapter were adapted from a publication in the proceedings of the 10th International Conference on Computational Creativity [57] in 2019. This work was developed from 2017 - 2019 alongside my collaborators Chase Shimmin, Douglas Duhaime, Ilya Vidrin, and Raymond Pinto, with generous support from the Yale Center for Collaborate Arts & Media.

5.2.1 Introduction

"I didn't want to imitate anybody. Any movement I knew, I didn't want to use." [58] Eminent postmodern dance choreographer Pina Bausch felt the same ache that has pierced artists of all generations – the desire to generate something truly original from within the constraints of your own body.

Recent technologies enabling the 3D capture of human motion as well as the analysis and prediction of timeseries datasets with machine learning have opened provocative new possibilities in the domain of movement generation. This project introduces a suite of configurable machine learning tools to augment a choreographer's workflow.

Many generative movement models from recent publications use Recurrent Neural Networks (RNNs) [59] as their fundamental architecture [60, 61, 62, 63, 64, 65]. Others create methods to draw trajectories through a lower-dimensional space of possible human poses constructed through techniques such as Kernel Principal Component Analysis (KPCA) [66, 67]. In this project, my collaborators and I build upon existing RNN techniques with higher-dimensional datasets and introduce autoencoders [68] of both poses and sequences of poses to construct variations on input sequences of movement data and novel unprompted sequences sampled from a lower-dimensional latent space.

Our models not only generate new movements and dance sequences both with and without a movement prompt, but can also create infinitely many variations on a given input phrase. These methods have been developed using a dataset of my own improvisational dance, recorded using a state-of-the-art motion capture system with a rich density of datapoints representing the human form. With this toolset, we equip artists and movement creators with strategies to tackle the challenge Bausch faced in her own work: generating truly novel movements with both structure and aesthetic meaning.

5.2.2 Context within Dance Scholarship

Dance scholarship, psychology, and philosophy of the past century has increasingly seen movement as embodied thought. Prominent proposals including psychologist Jean Piaget's sensorimotor stage of psychological development, the philosopher Maurice Merleau-Ponty's "phenomenology of embodiment", and Edward Warburton's concept of *dance enaction* have guided us today to view the human body as an essential influencer of cognition and perception [69].

Our vision for the future of creative artificial intelligence necessitates the modeling of not only written, visual, and musical thought, but also kinesthetic comprehension. The application of machine learning to movement research serves not as a mere outsourcing of physical creative expressiveness to machines, but rather as a tool to spark introspection and exploration of embodied knowledge in humans.

Concurrently with this branch of research, choreographers have wrestled with the problem of constructing a universal language of movement. Movement writing systems in use today such as Labanotation, Benesh Choreology, and Eshkol-Wachmann Notation can be effective, but none are as universal as, say, musical notation, and some make culturally-specific assumptions about how human bodies and types of motion should be abstracted and codified [71].

It is not our aim to replace existing methods of dance notation. However, we note the significance of 3D motion-capture techniques and abstract latent spaces in potentially reorienting movement notation away from culturally-centered opinions such as qualities of movement or which segments of the body get to define movement. Rather than gravitating in the direction of defining "universal" movement signifiers, we see this work as more aligned with the expressive figures generated by the visual artist Henri Michaux in an attempt to capture what he called *envie cinétique*, or "kinetic desire" – in other words, the pure impulse to move (see Figure 45). We therefore avoid limiting our generated movement outputs to only physically-achievable gestures, as this would only serve to limit the potential imaginative sparks lying dormant in these sequences.

Ethics in the philosophy of emerging media raise particular questions about how technology impacts what it means to be human, especially given the way constraints and resources of technology affect our embodied dispositions. When we consider the ethical dimensions of choreography in the context of machine learning, one major benefit is the opportunity to reflect on movement habits by observing, interpreting, and evaluating what is generated technologically. Other drawbacks could also emerge: if we ascribe great value to what we see, we may find ourselves in a position where we envy an algorithm's capacity to generate novel choreography. This may in turn lead us to cast judgement on ourselves and doubt our own human-created choreographies. While technology may provide new insights into patterns within dance sequences, it also inevitably leads to normative discussion about what

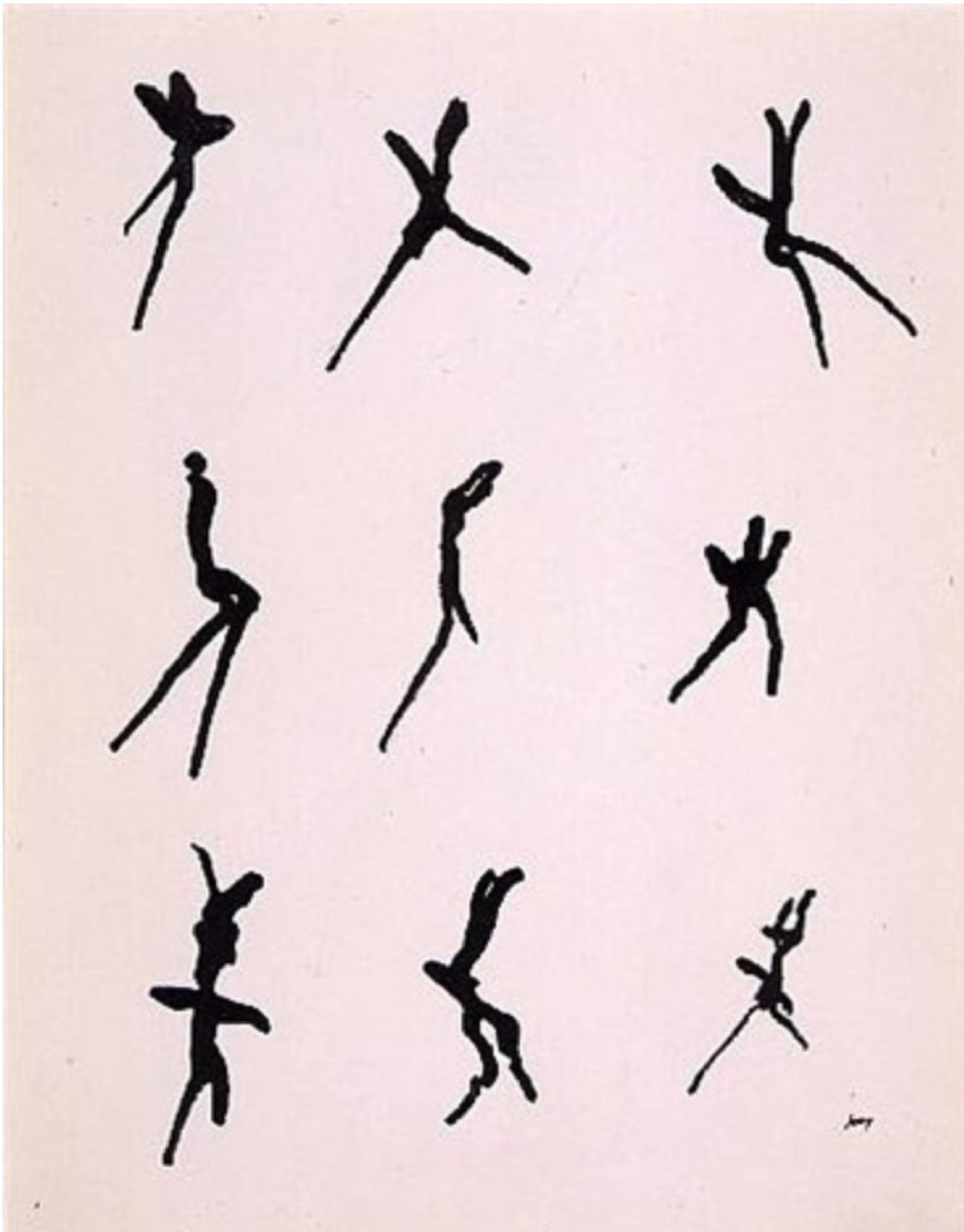


Fig. 45: Henri Michaux's notion of *envie cinétique*, or "kinetic desire", is represented by expressive, personal, and idiosyncratic gestures in calligraphic ink in his series *Mouvements* [70].

it means to choreograph well, or appropriately, or even creatively. This opens the door for fears of replacing our own practice with algorithms that could ostensibly rob us of the opportunity to get better at choreography, or learn to be more creative.

Several prominent choreographers have sought out both motion capture and machine learning tools to augment their practice, from Bill T. Jones and the OpenEndedGroup's 1999 motion capture piece *Ghostcatching* to William Forsythe to Merce Cunningham [72, 73, 74]. Wayne McGregor recently collaborated with Google Arts & Culture to create *Living Archive*, a machine learning-based platform to generate a set of movements using movement data extracted from McGregor's video archives [75].

Our work represents a unique direction in the space of "AI-generated" choreographies, both computationally and artistically. Computationally, we combine high-dimensional and robust 3D motion capture data with existing RNN-based architectures as well as introducing the use of autoencoders for 3D pose and movement sequence generation. Artistically, we deviate from having novel predicted sequences as the only end goal – in addition to this functionality, we grant choreographers the power to finely-tune existing movement sequences to find subtle (or not-so-subtle) variations from their original ideas.

5.2.3 Methods

Training data was recorded in a studio equipped with 20 Vicon Vantage motion-capture cameras and processed with Vicon Shogun software. This data consists of the positions of 53 fixed vertices on a dancer in 3 dimensions through a series of nearly 60,000 temporal frames recorded at 35 fps, comprising approximately 30 minutes of real-time movement. Each frame of the dataset is transformed such that the overall average (x,y) position per frame is centered at the same point and scaled such that all of the coordinates fit within the unit cube. The data was then exported to Numpy array format for visualization and processing in Python, and to JSON format for visualization with the interactive 3D Javascript library `three.js`. The neural network models were constructed using Keras with a Tensorflow backend.

In the following subsections, we describe two methods for generating dance movement in both conditional (where a prompt sequence of fixed length is provided) and unconditional (where output is generated without input) modes. The first method involves a standard approach to supervised training for sequence generation: an RNN is presented with a sequence of training inputs, and is trained to predict the next frame(s) in the sequence. The second method takes advantage of autoencoders to convert either an arbitrary-length sequence of dance movement into a trajectory of points in a low-dimensional latent space, or a fixed-length sequence to a single point in a higher-dimensional latent space.

LSTM+MDN: The model proposed in *chor-rnn* [61] uses RNNs to generate dance from a dataset of 25 vertices captured with a single Kinect device, which requires the dancer to remain mostly front-facing in order to capture accurate full-body data. Our RNN model uses an input layer of size $(53 \times 3 \times m)$ to represent 53 three-dimensional vertices with no rotational restrictions in a prompt sequence of m frames at a time. These sequences are then input to a series of LSTM layers, typically three, followed by a Mixture Density Network [62] that models proposals for the vertex coordinates of the subsequent n frames. The LSTM layers ensure the model is capable of capturing long-term temporal dependencies in the training data, while the MDN layer ensures generated sequences are dynamic and do not stagnate on the conditional average of previous vertex sequences [76]. The network is trained using supervised pairs of sequences by minimizing the negative log likelihood (NLL) of the proposed mixture model.

We also developed a modification of this structure using Principal Component Analysis (PCA) to reduce the dimensionality of the input sequences. This reduces the amount of information that must be represented by each LSTM layer. We then invert the PCA transformation to convert generated sequences in the reduced-dimensional space back into the $(53 \times 3 \times n)$-dimensional space.

The structure of a Mixture Density Network, as laid out in detail in [76], allows us to sample our target predictions from a linear combination of m Gaussian distributions, each multiplied by an overall factor of α_i, rather than from a single Gaussian. The probability density is therefore represented by

$$p\left(\vec{t} \mid \vec{x}\right) = \sum_{i=1}^{m} \alpha_i(\vec{x}) \phi_i\left(\vec{t} \mid \vec{x}\right)$$

where $\vec{x}$ represents our input data, $\vec{t}$ reprents a given predicted output, m represents the total number of Gaussian distributions in the mixture, and c represents the total number of components to predict (here, 53×3 for each timeslice). Each of the Gaussian distributions is modeled as:

$$\phi_i\left(\vec{t} \mid \vec{x}\right) = \frac{1}{(2\pi)^{\frac{c}{2}}\sigma_i(\vec{x})^c} e^{-\frac{|\vec{t}-\vec{\mu}_i(\vec{x})|^2}{2\sigma_i(\vec{x})^2}},$$

where $\vec{\mu}_i(\vec{x})$ and $\sigma_i(\vec{x})$ represent the mean values and variances for each component of the generated output.

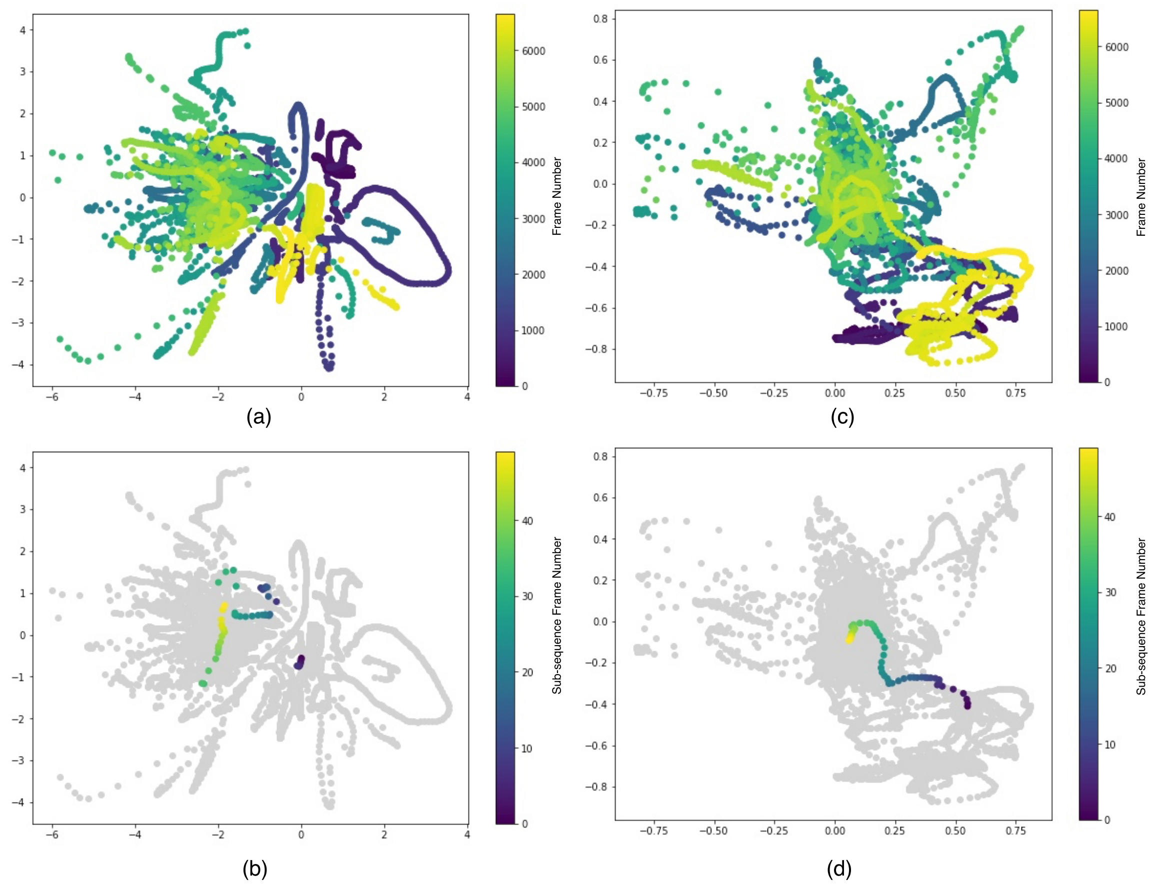


Fig. 46: **(a)** The 2-dimensional latent space of an autoencoder trained on a subset of the full dataset. The frame numbers show the procession of the sequence through time at a frame rate of 35 fps. **(b)** An example sequence of real training data is highlighted in this latent space. Note that its structure is highly noncontinuous. **(c)** The 2-dimensional latent space of an autoencoder trained on the same subset of data as the previous plots, but with the angular orientation of the frames subtracted. **(d)** The same sequence of real training data is highlighted, showing a much smoother and more continuous structure.

Autoencoder Methods: Unlike the RNN methods described above, autoencoders can learn features of the training data with a less directly supervised approach. The input and output layers are identical in dimensionality, while the intermediate layer or layers are of a reduced dimension, creating a characteristic bottleneck shape in the network architecture. The full network is then trained to replicate the training samples as much as possible by minimizing the mean-squared error loss between the input and the generated output. The network therefore learns a reduced dimensionality representation of "interesting" features in an unsupervised manner that can be exploited in the synthesis of new types of movement.

While a well-trained autoencoder merely mimics any input data fed into it, the resulting network produces two useful artifacts: an *encoder* that maps inputs of dimension $(53 \times 3 \times m)$ to a $(d \times m)$-dimensional space $(d < 159)$ and a *decoder* that maps $(d \times m)$-dimensional data back into the original dimensionality of $(53 \times 3 \times m)$. This allows us to generate new poses and sequences of poses by tracing paths throughout the $(d \times m)$-dimensional latent space which differ from those found in the training data.

While there are many other dimensional reduction techniques for data visualization, such as PCA, UMAP, and t-SNE [77, 78, 79], a significant advantage of autoencoders is that they learn a nonlinear mapping to the latent space that is by construction (approximately) invertible. Some differences between these other dimensionality-reducing techniques are illustrated in Figure 47.

In principle, autoencoders can be used to synthesize new dance sequences by decoding any arbitrary trajectory through the latent space. We prioritize continuity and smoothness of paths in the latent space when possible, as this allows human-generated abstract trajectories (for example, traced on a phone or with a computer mouse) a greater likelihood of creating meaningful choreographies. These qualities of trajectories in the latent space are most prevalent in PCA and our autoencoder methods (see Figure 47). However, as PCA is a linear dimensionality-reduction method, it is far more limited in ability to conform to the full complexity of the realistic data manifold compared to autoencoder methods.

The autoencoders' latent spaces do tend to produce mostly continuous trajectories for real sequences in the input data. This continuity can be greatly enhanced by subtracting out angular and positional orientation of the dancer, as shown in Figure 46. Removing these

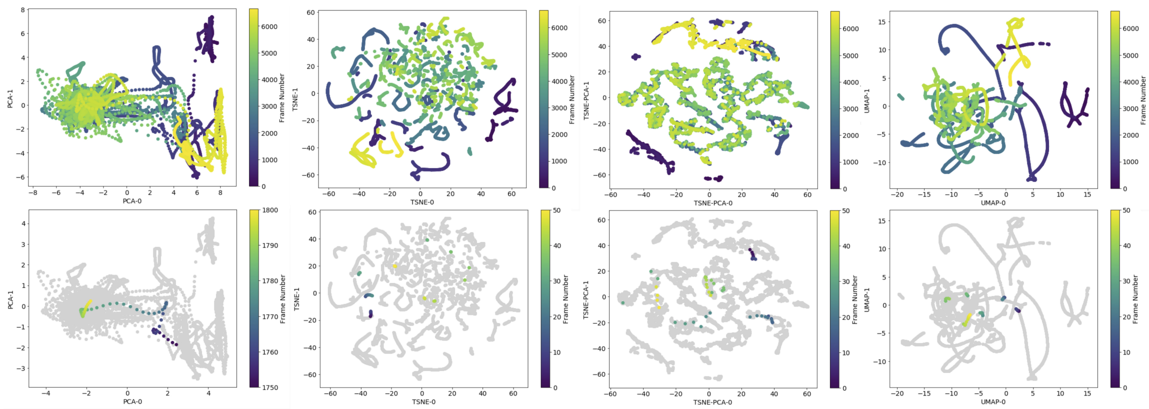


Fig. 47: A variety of 2D latent spaces are compared across multiple linear and nonlinear dimensionality-reduction techniques (excluding autoencoders): **(a)** PCA, **(b)** t-SNE, **(c)** t-SNE following PCA, and **(d)** UMAP. The top row shows full latent spaces for a subset of the training data, while the bottom row highlights the same example sequence of 50 frames in each space. All but PCA show a very segmented and discontinuous path for the sequence across the latent space. Our autoencoder techniques (see Figure 46) are comparable to PCA in terms of continuity of the paths in latent space, but have a much higher capacity to learn complex, nonlinear relationships than PCA alone.

dimensions of variation further reduces the amount of information that must be stored by the autoencoder and allows it to create less convoluted mappings of similar poses regardless of the overall spatial orientation of the dancer.

However, absent a deliberate human-selected trajectory as an input, it is *a priori* unclear how to select a meaningful trajectory, i.e., one that that corresponds to an aesthetically or artistically interesting synthetic performance.

In order to address this limitation, and to give some insight into the space of "interesting" trajectories in the latent space, we take another approach in which a second autoencoder is trained to reconstruct fixed-length sequences of dance poses by mapping each sequence to a single point in a high-dimensional latent space. Moreover, we train this network as a Variational Autoencoder (VAE) [80] which attempts to learn a latent space whose distribution is compatible with a $(d \times m)$-dimensional Gaussian. Sampling from this latent space results in unconditionally-generated sequences that are realistic and inventive (see Figure 48). For each sampling, we look at a single point in the latent space corresponding to a fixed-length movement sequence. Within the scope of this project, we do not attempt to impose any continuity requirements from one sampling to the next. Latent space points are chosen approximately isotropically. This creates a complementary creative tool to our previously-described traditional autoencoder for poses. We anticipate that choreographers and researchers could draw continuous paths through the latent space of poses to generate new movements as well as sample from the VAE latent space to generate new movement phrases and/or variations on existing phrases.

With both standard and variational autoencoders trained to replicate single poses and sequences of poses respectively, we introduce some techniques for taking a given input phrase of movement and generating infinitely many variations on that phrase. We define "variation" to mean that the overall spirit of the movement be preserved, but implemented with slightly different timing, intensity, or stylistic quality.

After identifying a desired dance phrase from which to create variations, we identify the sequence of points in the latent space representing that sequence of poses. We first constructed trajectories close to the original sequence by adding small sinusoidal perturbations to the original sequence. This created sequences resembling the original phrase, but with

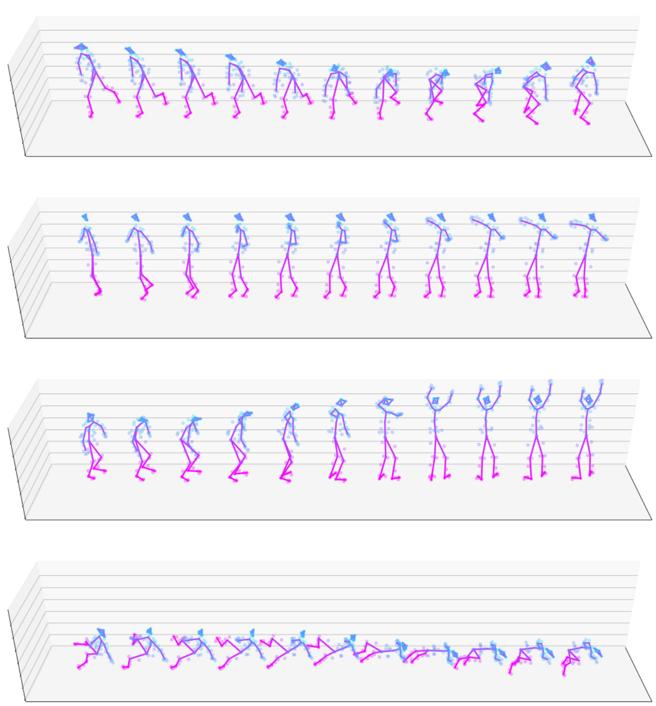


Fig. 48: Unconditionally-sampled sequences from the VAE.

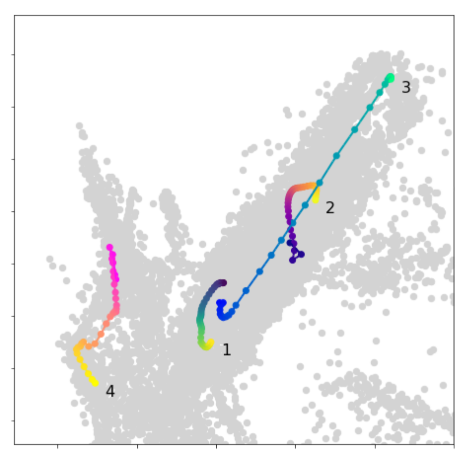


Fig. 49: Unconditionally-sampled sequences from the VAE projected into the latent space of the pose autoencoder (1 = top-most sequence; 4 = bottom-most sequence). Trajectories begin at darker colors and end at lighter colors.

an oscillatory frequency that was apparent in the output. This frequency could be tuned to the choreographer's desired setting, if the oscillatory effect is desired. However, we also sought out a method that constructed these paths in a less contrived manner.

For a VAE trained on sequences of poses, each point in the latent space represents an entire sequence of a fixed length m. We can construct variations on the input sequence by adding a small amount of random noise to the latent point and then applying the decoder to this new point in the latent space. This creates a new generated variation on the original sequence, with a level of "originality" that scales with the amount of noise added. Since the VAE's latent space has been constrained to resemble a Gaussian distribution, we can sample frequently from the latent space within several standard deviations of the origin without observing highly unphysical output sequences. Sampling within less than approximately 0.5σ tends to give very subtle variations, usually in timing or expressiveness in the phrase. Sampling within approximately 1 to 2σ gives more inventive variations that deviate further from the original while often preserving some element of the original, e.g. a quick movement upwards of a limb or an overall rotational motion. Sampling within 3 to 4σ and higher can produce myriad results ranging from no motion at all to extreme warping of the body to completely destroying the sense of a recognizeable human shape.

The relationship between these two latent spaces – that of the pose autoencoder and that of the sequence VAE – may be exploited to gain insight into the variations themselves. Points in the VAE latent space directly map to trajectories in the pose autoencoder space. By introducing a slight amount of noise to the point in the VAE latent space corresponding to a desired input sequence, we may decode nearby points to construct trajectories in pose space that are highly related to the original input sequence. Examples of variations from reference sequences are shown in Figure 50.

5.2.4 Results and Discussion

Both the RNN+MDN and autoencoded outputs created smooth and authentic-looking movements. Animations of input and output sequences for various combinations of our model parameters may be viewed here: `http://www.beyondimitation.com`.

Training the RNN+MDN with a PCA dimensionality reduction tended to improve the quality of the generated outputs, at least in terms of the reconstruction of a realistic human body. We used PCA to transform the input dataset into a lower-dimensional format that explains 95% of its variance. This transformation of the training data shortened the training time for each epoch by up to 15%, though test accuracy was not significantly affected. The output resulted in a realistic human form earlier in the training than without the application of PCA. In the future, we may also investigate nonlinear forms of dimensionality reduction to further improve this technique.

The architectures used for the RNN+MDN models included 3 LSTM layers with sizes varying from 32 to 512 nodes. They took input sequences of length m ranging from 10 to 128 and predicted the following n frames ranging from 1 to 4 with a learning rate of 0.00001 and the Adam optimizer [81].

The final architecture for the pose autoencoder comprises an encoder and a decoder each with two layers of 64 nodes with LeakyReLU activation functions with $\alpha = 0.2$ and compiled with the Adam optimizer. The pose autoencoder takes inputs of shape (53×3) and maps them into a latent space of 32 dimensions. Training this over 80% of our full dataset with a batch size of 128 and a learning rate of 0.0001 produced nearly-identical reconstructions of frames from the remaining 20% of our data after about 50 epochs. We also trained a modification of this architecture with a data augmentation technique that added random offsets between $[0, 1]$ to the $\hat{x}$ and $\hat{y}$ axes. This did not yield a significant advantage in terms of test accuracy, however, so we did not use it for our latent space explorations.

The final architecture for the sequence VAE also comprises an encoder and a decoder, each with 3 LSTM layers with 384 nodes and 1 dense layer with 256 nodes and a ReLU activation function, where 256 represents the dimensionality of the latent space. The model was compiled with the Adam optimizer. The VAE maps inputs of shape $(53 \times 3 \times l)$, where l is the fixed length of the movement sequence, to the $(256 \times l)$-dimensional latent space and then back to their original dimensionality. We used input sequences of length $l = 128$, which corresponds to about 4 seconds of continuous movement. We augmented our data by rotating the frames in each batch by a randomly-chosen $\theta \in [0, 2\pi]$. The VAE was trained with a learning rate of 0.001, a Kullback-Leibler weight = 0.0001, and a Mean Squared Error

(MSE) loss scaled by the approximate resolution of the motion capture data for about 1 day on a CuDNN-enabled GPU.

Sampling from the latent space of standard and variational autoencoders for both poses and sequences provided a rich playground of generative movements. We are particularly interested in the dynamic range provided by these tools to create variations on input sequences: by increasing the magnitude of the perturbation of the latent sequence to be decoded, choreographers can decide how 'creative' the outputs should look. By opting for either a standard or variational autoencoder, choreographers can sample from latent spaces with a bit more or a bit less similarity in the movements themselves to the training data. Adding sinusoidal perturbations as well as generating stylistically-related variations by exploiting the relationship between these two latent spaces proved effective and compelling methods for creating choreographic variations. The subtlety and smoothness with which we can vary input sequences using the VAE also underscores that the model is truly generating new outputs rather than memorizing the input data.

These methods have already been effective at sparking choreographic innovation in the studio. They center the choreographer's embodied knowledge as something to be modeled and investigated – not just as a compendium of possible bodily positions, but as a complex and high-dimensional landscape from which to sample movements both familiar and foreign. Movements throughout these abstract landscapes can be constructed in a variety of ways depending on the application. For example:

- For a choreographer seeking primarily to document their practice, training these models allows them to save not only the physical motions captured in the motion capture data, but also their *potential* for movement creation as approximated by a well-trained model. Different models may be saved from various periods of their practice and compared or re-explored indefinitely.

- For a choreographer looking to construct a new piece out of their own typical patterns of movement, sampling from within 1σ in the VAE latent space can generate multiple natural-looking phrases that can then be stitched together in the studio to create a cohesive piece. They could also prompt new sequences of arbitrary length following from existing choreography via the RNN+MDN model.

- For a choreographer who wants to understand and perhaps break out of their typical movement patterns, analyzing the latent space of the pose autoencoder can be instructive. Visualizing trajectories through the space can inform what areas lie unexplored. Drawing continuous paths through the latent space can then construct new phrases that might otherwise never emerge from an improvisation session.

- A choreographer might also use these methods to support teaching movements to others. By comparing trajectories in the same latent space, students can track their mastery of a given movement sequence.

Future technical work to develop these methods will include the investigation of non-linear, invertible data-reduction techniques as a form of pre-processing our inputs, other neural network-based models designed to work with timeseries data such as Temporal Convolutional Networks, and more sophisticated methods for sampling from latent spaces.

Feedback from other choreographers who used our interactive models also indicated that it would be interesting to extend our current dataset with additional data focused on the isolation of certain regions of the body and/or modalities of movement. Another possible next step in extending this work includes exploring latent spaces of multiple dancers. While only solo dances were captured for these studies, the Vicon system can readily accommodate multiple simultaneous dancers, which could allow us to explore the generation of duets and group choreographies.

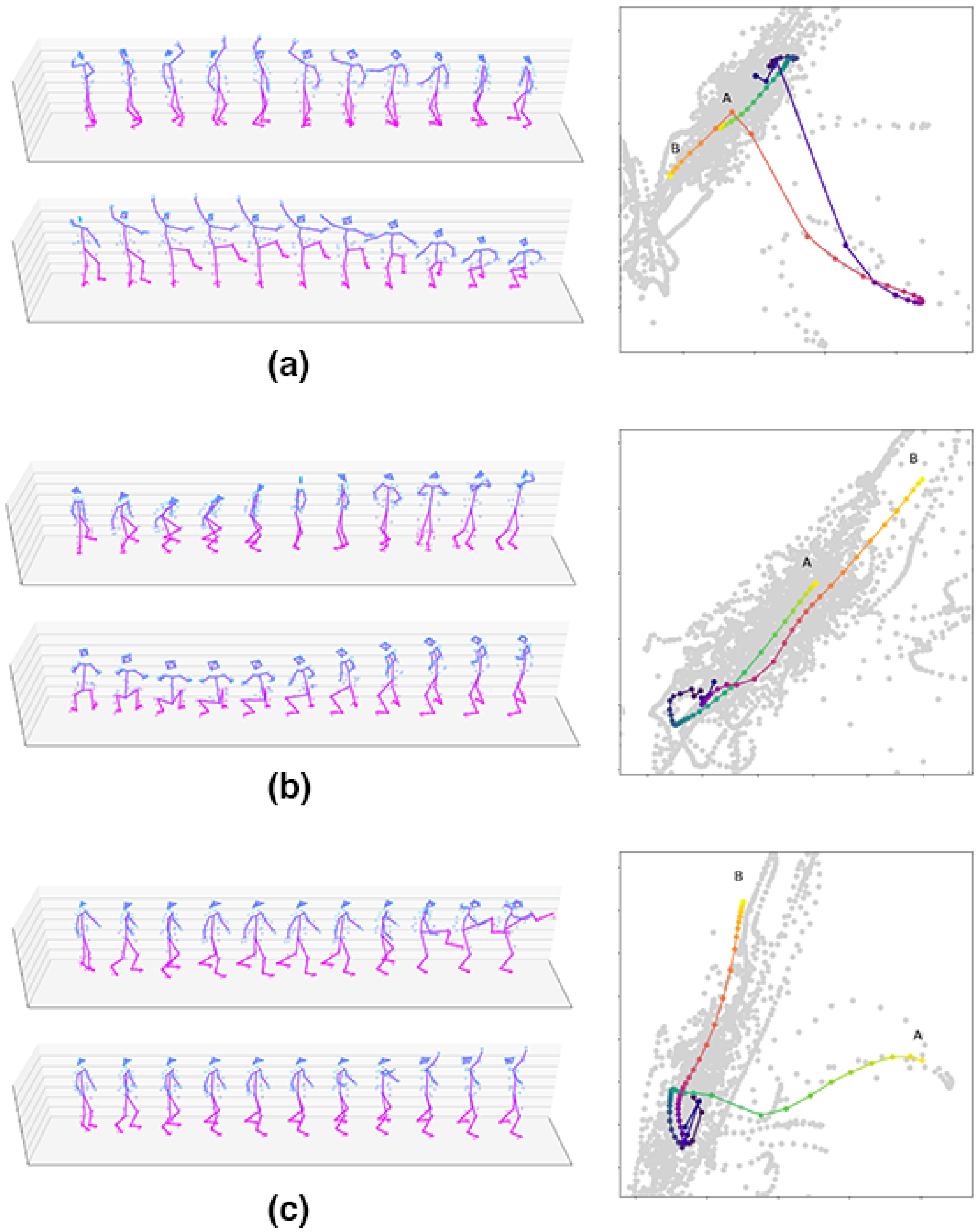


Fig. 50: In the left column, a reference input sequence (above, A) and a generated variation sequence (below, B) with 0.5σ noise added to the input's representation in latent space are shown, both with lengths of 32 frames (time progressing from left to right). In the right column, Reference (A) and generated (B) variation sequences projected into the pose autoencoder space. Trajectory colors go from dark to light over time. Observed modifications to the sequences include: (a) while the reference sequence includes a rotation, the generated variation removes the spin, while the movements of the left arm are synchronous in both cases; (b) the generated variation preserves the rising motion but adds a rotation; and (c) the reference sequence features a kick, while the variation instead translates this upward motion into the arms, rather than the feet.

5.3 Choreo-Graph: Learning Latent Graph Representations of the Dancing Body

The contents of this chapter were adapted from a publication in the proceedings of the NeurIPS 2020 Workshop on Machine Learning for Creativity and Design [82]. This work was developed in 2020 during a summer internship with Intel's AI lab alongside my collaborators Santiago Miret, Somdeb Majumdar, and Marcel Nassar.

5.3.1 Introduction

This project introduces, to the best of my knowledge, the first use of Graph Neural Networks (GNNs) for the generation and analysis of choreographic data.

Extending the work in [57] just described in Section 5.2, we apply GNNs to impose an explicit graph structure to the latent space encodings of movement sequences based on the Neural Relational Inference (NRI) model [83]. These latent graphs represent the body as nodes and a variational topology of edges, with positions over time included as node features. This method not only generates future movements, but also augments the analysis by highlighting compelling learned categories of connections and identifying a small subset of important interactions within the dancer's body. The creative implications of this work are somewhat different than my earlier work using VAEs: rather than only focusing on movement generation, this work results in an introspective analysis on the interconnectivity of the moving body.

The results shown here are from our custom implementation of the NRI model using the GNN library Pytorch Geometric [84]. Our code, data, and animations of our generated outputs and edge types can be found at `https://github.com/mariel-pettee/choreo-graph`.

5.3.2 Neural Relational Inference

NRI is a GNN-based model designed to analyze the dynamic evolution of many-body systems for which the underlying interaction structure between particles is unknown. Based on a VAE, the model learns to categorize the edges of a fully-connected input graph into a finite number of learned edge types. It uses a neural network-based encoder to convert the input graph data into a discrete probability distribution over edge types, and then applies an RNN-based decoder to predict the future system dynamics given those edge type predictions (see Figure 51). The encoder and decoder are jointly trained with a Gumbel-Softmax sampling of edge types in the latent graph embedding space.

The Gumbel-Softmax approximation allows differentiable sampling during training, which can then become categorical during inference runs after training. The loss function consists of two components: (1) the negative log-likehood (NLL) reconstruction loss and (2) the Kullback-Leibler (KL) loss reflecting how much the latent edge type probability distributions resemble the desired prior distribution. Our experiments used a sequence length of 49 timesteps, of which the final 10 timesteps are predicted, and four total edge types, including a non-edge as the first type.

5.3.3 Visualization of Edge Types

Following training of our model, we use our test set to create animations of the generated sequences as well as the categories of edges learned by the model. Each learned edge type defines a category of directed edges within the body that describe similar physical dynamics between two nodes. By ranking the edges by their normalized log-probabilities within each edge type, we can isolate the most indicative edges for each learned category, revealing distinct latent cross-body interactions that would be otherwise invisible (see Figure 52). The first edge type is coded as a non-edge in order to encourage sparse graph representations. This ranking strategy can reduce a dataset of nearly 3,000 fully-connected, directed edges into fewer than 100 edges total that still capture the main kinematic features and edge categorizations of the motion.

5.3.4 Context within Contemporary Dance & Future Directions

Building on post-modern and contemporary dance's rich history of generating choreography via algorithms, games of chance, and other technologies [72, 73, 75, 85, 86], this application of GNNs to dance opens new pathways for generating movements for both human and animated bodies with techniques that many dancers already use. The discovery of various edge types yields new insight into a dancer's movements, revealing their dominant cross-body interactions and challenging them to focus on their least important connections for future improvisations as a way of resisting more comfortable patterns of movement generation. AI-generated dance could also be fine-tuned to fit various aesthetic preferences by varying the strengths of these edge connections.

In our experiments, we find that the important learned edges often correlate with tenets of the dancer's classical and contemporary dance training. In the example in Figure 52, category (c) reflects the balletic notion of the opposition of arms and legs. This example also has distinct edge types representing how the movement of the left hand influences the upper torso and head (b) as well as how the left foot influences the movements of the hands (d). This notion of cause-and-effect, of one body part initiating a movement that ripples to a distant corner of the body, is suggestive of Ohad Naharin's movement vocabulary *Gaga* [87], which emphasizes the generative potential of various "engines" within the body. Through the lens of Gaga, the body is highly connective, and regular movement practice can help a dancer become more sensitive to these granular self-interactions.

Interestingly, most of the learned edge types in our experiments do not follow the physical connectivity of the skeleton. The model's emphasis on the extremities as initiators makes sense in the context of the dancer's many years of ballet training, which emphasizes extension through the hands and feet. Contemporary and post-modern dance, including Gaga, can tend to focus more on contractions of the abdominals and the hips than of movements in the body's distal extremeties. We could also potentially interpret the lack of edges to and from the core of the body as a relative dearth of datapoints from the motion capture suit in this area.

Future work in this direction will include a description of the qualities of the learned interaction types, as well as an ablation study to determine how much predictive information is contained in the sparsest edge types. We hope that our application of GNNs to the body can introduce dancers to a new set of self-interactions to investigate as a complement to their own embodied knowledge and perceptions.

5.3.5 Additional Technical Details

- **Dataset:** We use the same motion capture dataset used in [57], which contains approximately 50,000 timesteps of (x,y,z) positions for 53 points on the body. This corresponds to around 30 minutes of improvisational choreography of my own body, drawing from my extensive background in ballet and contemporary dance. Of the 53 points on my body, about 10% are on the head, 30% on the arms, 8% on the hips, 37% on the legs and feet, and the remaining 15% on the chest and back. We also calculate velocities for each timestep in the dataset. For training the GNN, we convert the dataset into batches of fully-connected graphs for sequences of 49 timesteps in a sliding window fashion. The first 70% of the batches are designated as training batches, the next 15% are used for validation during training, and the remaining 15% are used for testing following the completion of the training.

- **Training:** The model used, which embeds the node features to 256 dimensions and contains hidden layers of size 256 as well, has 1.7 million trainable parameters. The Adam optimizer was used with learning rate 0.00005 and a weight decay coefficient of 0.0005. The model was trained on a GPU until convergence for approximately 12 hours.

5.3.6 Ethical Considerations

These results should be understood as a personal investigation into one dancer's movement generation process, not as a general declaration about what body connections are significant for all dancers or bodies. The dataset reflects the unique movement practice of a single dancer and therefore cannot represent the full diversity of human bodies or global movement practices. By only using motion capture data captured from my own body, we ensured the data was sourced consensually, with the explicit goal of using it to train machine learning models for creative insight.

5.4 Physics Applications

Though the methods described in this chapter were developed for a particular dataset of my own movements, the techniques are generic, and have several interesting connections to recent problems in ML for physics. Generative models such as VAEs could be used in the future for studies concerning the creation of simulated LHC data without the significant computational demands of the current MC generators. In contrast with Generative Adversarial Networks (GANs) that have been applied towards this problem, VAEs are often faster, more stable, and easier to train [88]. Rather than generating e.g. images of jets [89], the methods described here could be adapted to consider timeseries datasets or dynamic object trajectories such as particle tracks. GNNs are also of great interest for track reconstruction in the HL-LHC [90], and the particular focus on using GNNs for learning interaction edges has seen compelling results in extracting physically-meaningful relationships from edge types on simulations and even real dark matter data [91].

By applying our GNN model to particle track information, we can extract the most important interaction edges between tracker hits, investigate the interaction types learned by the model, and predict the future trajectories of the tracks. By imposing a sparse prior distribution on the edge types, we can require that the majority of edges are classified as a non-edge, potentially allowing for more efficient, information-rich representations of the underlying particle interactions themselves. I am particularly curious to use this model to reframe trigger-level and offline particle identification, whether for taus vs. QCD-initiated jets or for classifying many different particle types at once, as a question of learning latent time-dependent interaction graph structures.

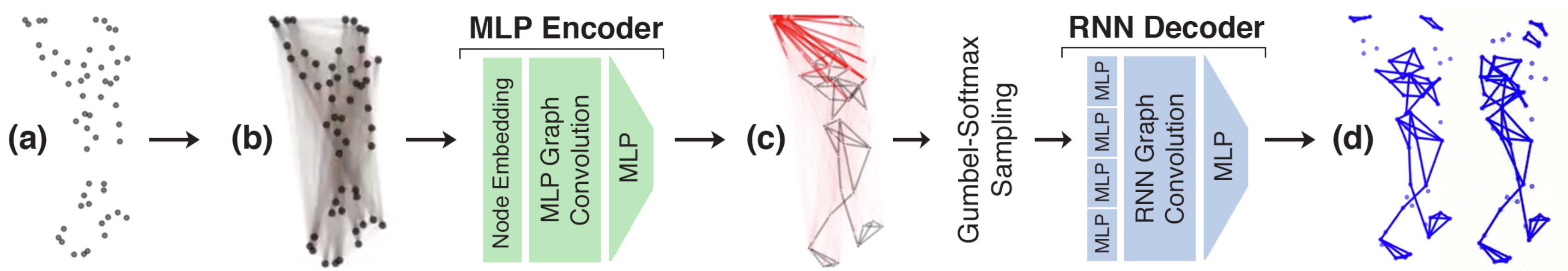


Fig. 51: NRI: a) Input point-cloud data; b) Fully-connected graph; c) Latent edges; d) Predictions

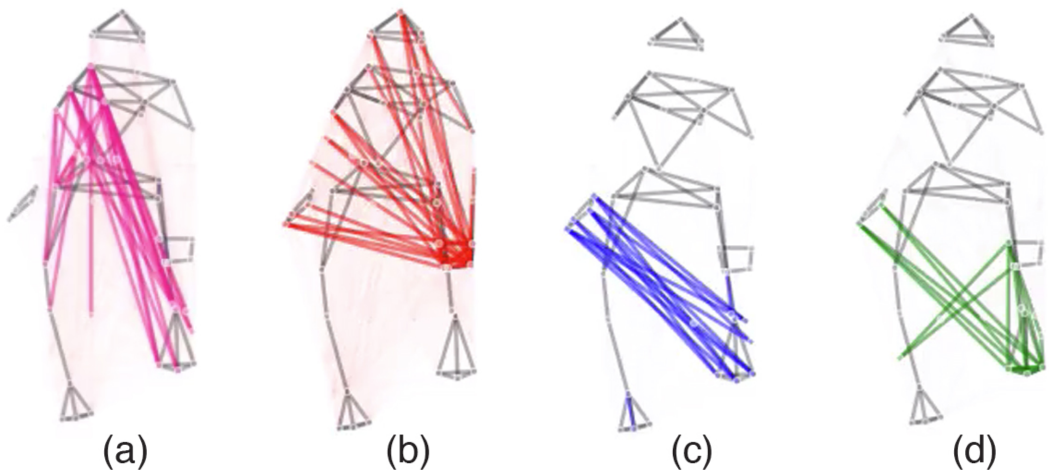


Fig. 52: The top 1% of edges, ranked by the normalized magnitude of each edge's log-probability, for the four learned edge types on the same batch of test data. (a) represents a non-edge, while the other edge types show directed edges going primarily from (b) the left hand to the upper body, (c) right hand to left foot, and (d) left foot to both hands.

6 The $V \to$ leptons, $H \to \tau\tau$ Analysis Strategy

6.1 Motivation

One of the primary mandates of the LHC is to investigate whether the origin of the masses of most elementary particles is actually the result of electroweak symmetry breaking and the Higgs mechanism. Since the discovery of the Higgs boson in 2012, LHC physicists have been feverishly investigating its properties to understand if this new particle aligns with our theoretical expectations for the Higgs boson from the Standard Model. This is a monumental task, as we can study the Higgs boson from a variety of angles based on its many combinations of production channels and decay channels. As we saw in Table 4, there are four primary ways we expect to produce a 125 GeV Higgs boson at the LHC at $\sqrt{s} = 13$ TeV:

- Gluon-gluon fusion (ggF): ∼88% of total Higgs boson production cross-section (σ_H)
- Vector boson fusion (VBF): ∼7% of σ_H
- Associated production with a vector boson (VH): ∼4% of σ_H
- Associated production with 2 top quarks (ttH): $< 1\%$ of σ_H

The Standard Model also provides predictions for the branching ratios of the Higgs boson as it decays into other particles (shown in Table 5). However, this does not necessarily correlate with the most likely observations of 125 GeV Higgs decays at the LHC, as some decay channels are more difficult to detect than others. For example, $H \to \gamma\gamma$ has a relatively small branching ratio, with less than 1% of expected Higgs boson decays, yet it was one of the first Higgs boson decay channels observed at the LHC due to its clean experimental signature. The primary Higgs boson decay channels targeted during Run 2 at the LHC are:

- Higgs decaying to two photons ($H \to \gamma\gamma$)
- Higgs decaying to two Z bosons ($H \to ZZ^*$)
- Higgs decaying to two W bosons ($H \to WW^*$)
- Higgs decaying to two tau leptons ($H \to \tau\bar{\tau}$)
- Higgs decaying to two b quarks ($H \to b\bar{b}$)

Of the 20 possible combinations of these primary production and decay modes, only three have been excluded from the latest combination measurements of the properties of the Higgs boson (see Figure 53): ggF, $H \to b\bar{b}$; VH, $H \to WW^*$; and VH, $H \to \tau\bar{\tau}$. If the Higgs boson is indeed responsible for lepton masses via the Yukawa couplings, the most massive lepton – the tau – is our main probe of Higgs interactions with the lepton sector. The VH, $H \to \tau\bar{\tau}$ process is therefore an important missing piece of our overall description of the Higgs boson based on the decay modes we have best access to at the LHC.[9]

Run 2 at the LHC is a particularly interesting time for the VH, $H \to \tau\tau$ analysis: for the first time, we might have enough data to detect evidence of this interaction. The Run 1 VH, $H \to \tau\tau$ analysis was limited in its statistics for this rare channel and was only able to set upper limits on the overall VH, $H \to \tau\tau$ cross-section [93]. This current analysis work, however, uses about seven times the total integrated luminosity from the Run 1 analysis at the LHC. It also benefits from the increased average number of bunch crossings per event and higher center-of-mass energy of Run 2 ($\sqrt{s} = 8$ GeV $\to \sqrt{s} = 13$ GeV). Thanks to these improved experimental conditions, both the VH production mode [94] and the $H \to \tau\tau$ decay mode [95] have separately been observed by the ATLAS Experiment in recent years with $> 5\sigma$ significance. Additionally, since this analysis targets specifically $V \to$ leptons, $H \to \tau\tau$, we benefit from high trigger efficiencies for electrons and muons in particular for selecting our events. By using electron and muon triggers exclusively, we do not risk biasing our $H \to \tau\tau$ decay with the use of tau triggers. Evidence of this channel in Run 2 would be a significant milestone in our study of the Higgs boson with the potential to improve the precision of the combined $H \to \tau\tau$ measurement and our understanding of the Higgs boson as a whole.

6.2 Summary of Analysis Strategy

The $V \to$ leptons, $H \to \tau\tau$ analysis is composed of four signal regions:

- $W \to l\nu_l$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$
- $W \to l\nu_l$, $H \to \tau_{\text{had}}\tau_{\text{had}}$

[9] Note that I will omit the bar ($\bar{\tau}$) when writing out $H \to \tau\bar{\tau}$ for the remainder of this chapter, both for simplicity and to avoid implying that we require each of our leptonic/hadronic tau decays to be a particular sign, rather than the more general requirement that the pair of taus is opposite-sign.

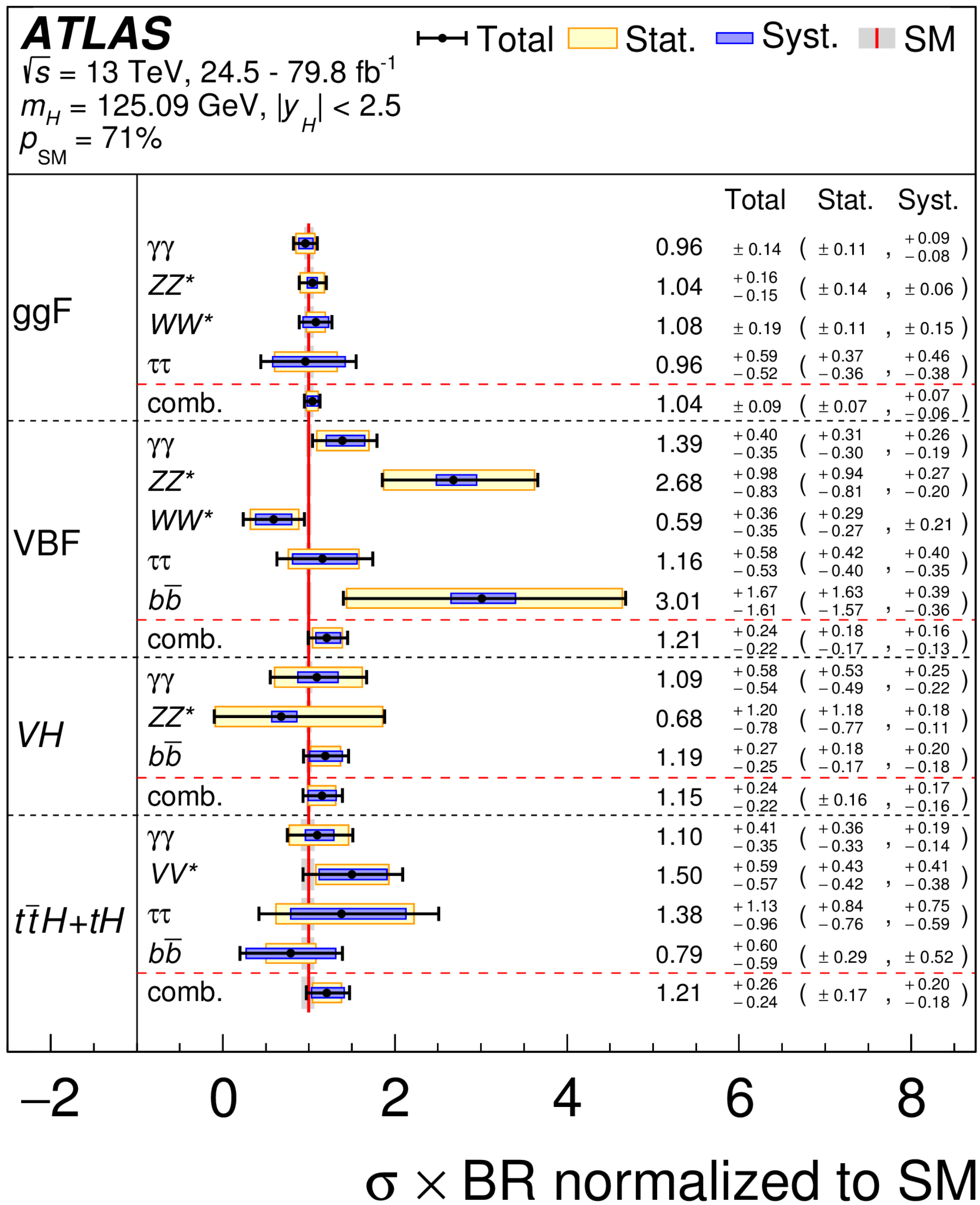


Fig. 53: The cross-section σ times branching ratio (BR) for each combination of Higgs boson production and decay modes included in the overall Higgs boson measurements using Run 2 data at $\sqrt{s} = 13$ TeV are shown, normalized to their SM expected values. The black error bars show the total uncertainties in the measurements, while the yellow boxes show statistical and blue boxes show systematic uncertainty values. Notably absent from this list are three primary Higgs boson modes, including VH, $H \to \tau\bar{\tau}$. [92]

- $Z \rightarrow ll$, $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$
- $Z \rightarrow ll$, $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$

In each of these cases, $l \equiv \{e, \mu\}$. Interactions with the Higgs boson decaying to two leptonically-decaying taus, i.e. $H \rightarrow \tau_{\text{lep}}\tau_{\text{lep}}$, are excluded from our analysis to avoid overlapping with other important Higgs decay channels such as $H \rightarrow ZZ^* \rightarrow llll$. Due to the multiple neutrinos (i.e. sources of missing E_T) present in the final state of each signal region from the tau decays, it is not possible to make a fully-reconstructed measurement of the Higgs boson mass. Each signal region therefore uses a particular technique for estimating or constraining the Higgs boson mass (described in Section 6.8). The ZH categories use a technique called the Missing Mass Calculator (MMC), while the WH categories use a quantity called Late-Projected Transverse Mass ($M_{2\text{T}}$).

Across the four signal categories, there are two possible sources of non-signal events that we must account for: **irreducible** and **reducible** background events. Irreducible events are those in which all final-state particles have been correctly reconstructed and identified and match the final-state configuration for one of our signal regions, but the event itself is not a $V \rightarrow$ leptons, $H \rightarrow \tau\tau$ process. These events are typically diboson events, e.g. WZ ($W \rightarrow l\nu$, $Z \rightarrow \tau\tau$) and ZZ ($Z \rightarrow ll$, $Z \rightarrow \tau\tau$), where a $Z \rightarrow \tau\tau$ decay mimics a $H \rightarrow \tau\tau$ decay. The Run 1 analysis used a Monte Carlo (MC) subtraction technique to estimate the contributions of these events [93]. For Run 2, I have developed a neural network technique trained on MC that exploits the kinematic differences between diboson and signal events for this separation. Reducible backgrounds, on the other hand, are a variety of events that enter into our signal region because of one or more misidentified objects and/or correctly-identified objects that originated from a non-prompt process, e.g. conversion electrons produced from photon interactions in the ATLAS detector. These types of events, called **fakes**, are dominated by the contribution from quark- and gluon-initiated jets misidentified as taus, but also have contributions from fake electrons and muons. Electron, muon, and tau fake factors are calculated separately using a data-driven technique called the Fake Factor Method. These fake factors allow us to estimate the contribution of fake objects in our signal region based on how many fake objects are measured in a fake-enriched region in data ($Z \rightarrow ll$ + jets).

6.2.1 Data and Monte Carlo Sample Productions

The data used for this analysis corresponds to the full Run 2 LHC dataset, i.e. years 2015 - 2018 of datataking, deemed good for physics. This amounts to an integrated luminosity of 139 fb^{-1} of data. Table 8 summarizes the simulated processes incorporated into this analysis. Both data and MC samples were produced using the HIGG4D1 ("lep-lep") and HIGG4D2 ("lep-had") derivations to correspond with the ZH and WH analysis regions.

Process	MC Generator + UEPS	PDF Set	Perturbative Order
$W \to l\nu$, $H \to \tau\tau$	Powheg+Pythia8	NNPDF30NLO	NLO
$Z \to ll$, $H \to \tau\tau$	Powheg+Pythia8	NNPDF30NLO	NLO
$ggF\ H \to \tau\tau$	Powheg+Pythia8	NNPDF30NNLO	NNLO
VBF $H \to \tau\tau$	Powheg+Pythia8	NNPDF30NLO	NLO
ttH, $H \to \tau\tau$	Powheg+Pythia8	NNPDF23LO	LO
Diboson	Sherpa 2.2.2	NNPDF30NNLO	NNLO
Triboson	Sherpa 2.2.2	NNPDF30NNLO	NNLO
W+jets	Sherpa 2.2.1	NNPDF30NNLO	NNLO
Z+jets	Sherpa 2.2.1	NNPDF30NNLO	NNLO
t	Powheg+Pythia8	NNPDF30NLO	NLO
$t\bar{t}$	Powheg+Pythia8	NNPDF30NLO	NLO

Table 8: Information on the Monte Carlo generators used to produce the major simulated processes incorporated into this analysis in Run 2, including the process name, names of the MC generator and the model of the underlying event with hadronization and parton showering (UEPS), the corresponding PDF set, and the perturbative order in QCD to which the cross-section has been calculated (NLO = Next-to-leading order; NNLO = Next-to-next-to-leading order.)

The HIGG4D1 ("lep-lep") derivation requires one of the following reconstruction-level criteria for an event to be included:

- Two electrons with $p_T^e > 13$ GeV and passing medium ID
- One electron with $p_T^e > 13$ GeV and passing medium ID + one muon with $p_T^\mu > 13$ GeV with good reconstruction quality
- One muon with $p_T^\mu > 13$ GeV + one muon with $p_T^\mu > 9$ GeV, each with good reconstruction quality

The HIGG4D2 ("lep-had") derivation requires one of the following reconstruction-level criteria for an event to be included:

- One electron passing medium ID + one hadronic tau with electric charge $|q| = 1$ and one or three charged tracks, passing either $p_T^e > 22$ GeV & $p_T^\tau > 18$ GeV or $p_T^e > 15$ GeV & $p_T^\tau > 23$ GeV

- One muon with good reconstruction quality + one hadronic tau with electric charge $|q| = 1$ and one or three charged tracks, passing either $p_T^\mu > 18$ GeV & $p_T^\tau > 18$ GeV or $p_T^\mu > 12$ GeV & $p_T^\tau > 23$ GeV

6.2.2 Analysis Software

After producing the HIGG4D1 and HIGG4D2 derivations for our MC and data samples, we apply a software called the `xTauFramework` for producing ROOT TTree output files for further analysis. The ATLAS-internal `xTauFramework` is collectively maintained and updated by several ATLAS members and is used for other tau-related analyses such as the ggF/VBF $H \to \tau\tau$ coupling measurement as well as searches for lepton flavor-violating processes involving taus in the final state. It is written primarily in C++ and is responsible for calibrating and sorting physics objects, calculating scale factors and systematic uncertainty variations, and applying overlap removal. Additionally, we require events to contain at least one reconstructed tau and one reconstructed electron or muon passing overlap removal, as this is the minimal set of shared objects across all four of our signal regions. From input xAODs or derivation (DxAOD) files, the `xTauFramework` outputs ROOT files called MxAODs, or "mini xAODs", as ~ 1 GB-size files suitable for interactive analysis.

With these ROOT files in hand, we then use a custom (ATLAS-internal) Python-based analysis framework called `vhtautau` to apply our selections, calculate event-level and object-level weights, and make our fake rate measurements. Our software benefits from `uproot` [96], a package designed for pure Python-based ROOT I/O, to convert our variable-length vectors of particles in the ROOT file structure into fixed-length Pandas [42] DataFrames stored in an HDF5 binary file format.

6.3 Event Selection

6.3.1 Trigger Selection

One of the advantages of the search for $H \to \tau\tau$ in the leptonic VH production mode is the guaranteed presence of electrons or muons in our final states of our signal events. We therefore optimize our analysis to take advantage of the relatively clean and efficient electron and muon triggers and suppress background from multijet pileup events. We use the same trigger chains as the $H \to \tau_{\text{lep}}\tau_{\text{lep}}$ channel in the ggF/VBF SM $H \to \tau\tau$ combination analysis: single-electron triggers, single-muon triggers, $e + \mu$ triggers, $e + e$ triggers, and $\mu + \mu$ triggers. Because each of these triggers is not maximally efficient at the minimum p_T threshold (see Figures 54 and 55), we apply p_T cuts on our triggered objects that exceed the minimum trigger thresholds by 2 GeV.

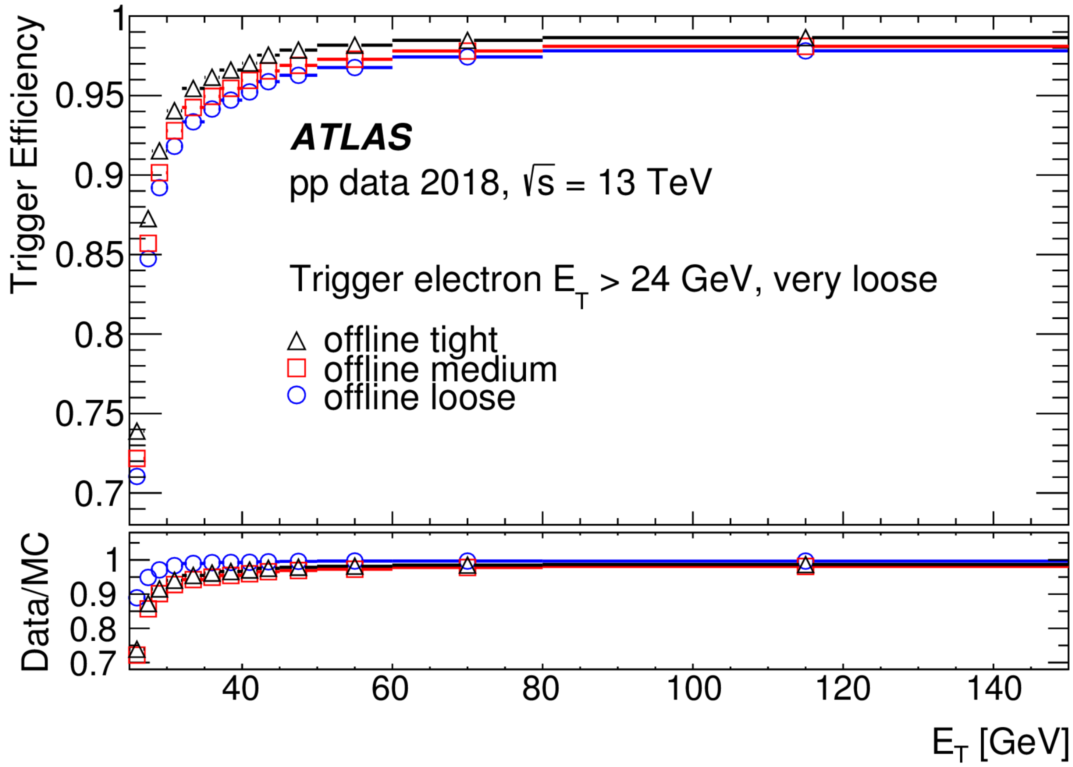


Fig. 54: Efficiency of the `HLT_e24_lhvloose_nod0` trigger versus offline electron E_T for 2018 datataking. A steep turn-on curve in efficiency starting at the minimum trigger p_T threshold demonstrates the benefit of applying a slightly higher p_T requirement on our triggered objects. [27]

Trigger Type	2015 Threshold(s) [GeV]	2016-2018 Threshold(s) [GeV]
Single electron	$p_T^e > 24$	$p_T^e > 26$
Single muon	$p_T^\mu > 20$	$p_T^\mu > 26$
Electron + muon	$p_T^e > 17$, $p_T^\mu > 14$	$p_T^e > 17$, $p_T^\mu > 14$
Symmetric di-electron	$p_T^e > 12$	$p_T^e > 17$
Symmetric di-muon	$p_T^\mu > 14$	$p_T^\mu > 14$
Asymmetric di-muon	$p_T^{\mu_1} > 20$, $p_T^{\mu_2} > 8$	$p_T^{\mu_1} > 22$, $p_T^{\mu_2} > 8$

Table 9: Run 2 minimum trigger thresholds. Analysis-level cuts are 2 GeV higher than each of these thresholds in order to avoid the relatively degraded efficiency of the trigger at its minimum threshold.

6.3.2 Overlap Removal

During the ATLAS reconstruction and identification process for each physics object, it is sometimes the case that a single object will be reconstructed as multiple categories of objects. For example, an electron candidate might also be reconstructed as a 1-prong tau. If these different reconstructed objects overlap, i.e. their calorimeter clusters or tracks fall within the same region of ΔR, then an overlap removal process is initiated in order to assign each reconstructed object a single label. Broadly, this results in the preference of electrons and muons over taus and the preference of taus over jets. The standard overlap removal process in ATLAS Run 2 compares particles in the following order:

1. Choose electrons over taus (within $\Delta R < 0.2$)
2. Choose muons over taus (within $\Delta R < 0.2$)
3. Choose muons over electrons (if the object has a muon-like calorimeter signature); otherwise, prefer electrons to muons
4. Choose electrons/muons over photons (within $\Delta R < 0.4$)
5. Choose electrons/muons over jets (within $\Delta R < 0.2$); otherwise, choose jets over electrons/muons (within $\Delta R < 0.4$)
6. Choose taus over jets (within $\Delta R < 0.2$)
7. Choose jets over photons (within $\Delta R < 0.4$)

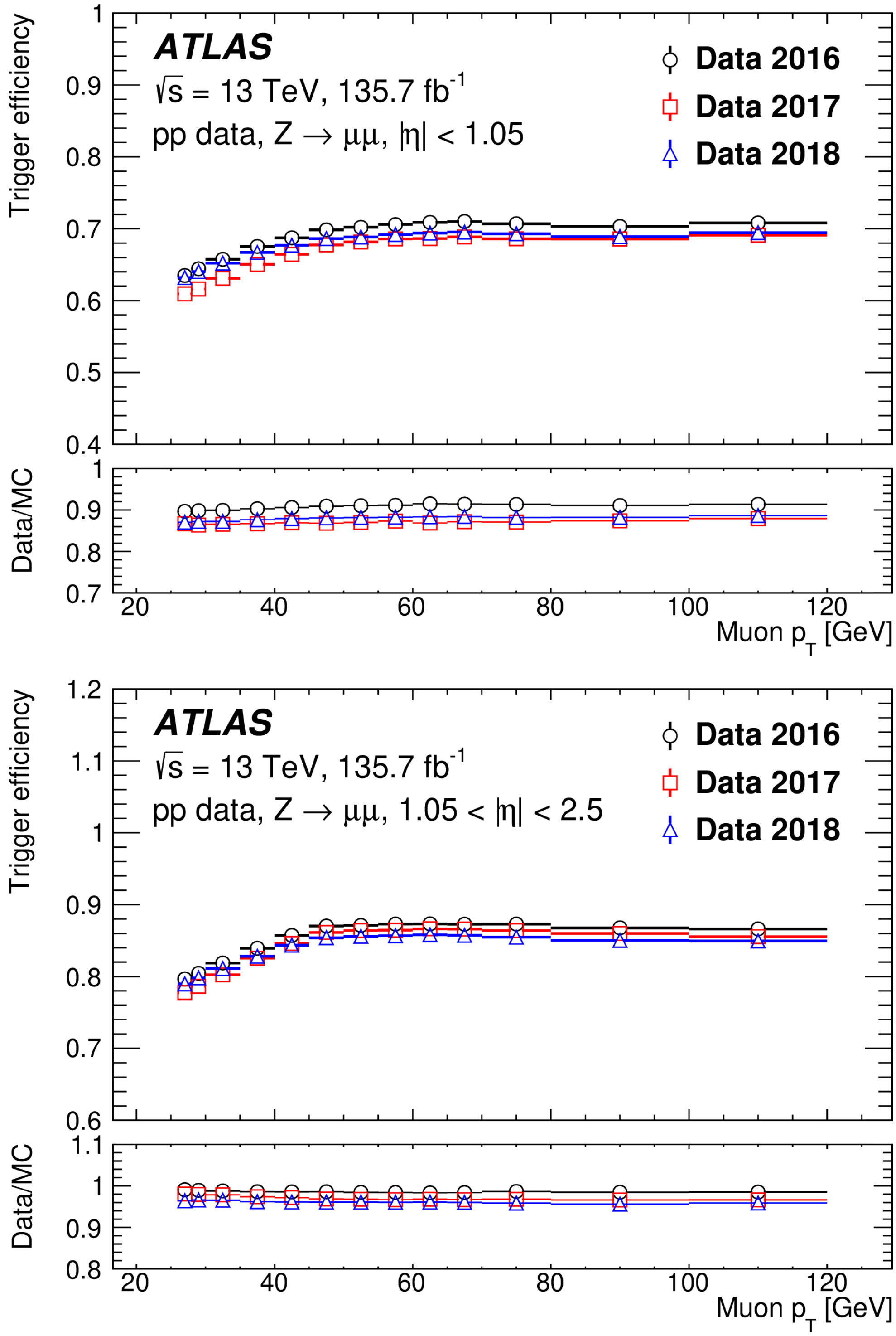


Fig. 55: Efficiency of passing the single muon trigger HLT_mu26_ivarmedium or HLT_mu50 as a function of muon p_T in the barrel (above, $|\eta| < 1.05$) and endcap regions (below, $1.05 < |\eta| < 2.5$). [28]

6.3.3 Object Criteria

The following documents the specific object criteria we require for physics objects used in the VH, $H \to \tau\tau$ analysis in Run 2. In many cases, they closely map onto similar requirements from the Run 1 analysis, but overall they reflect the latest recommendations based on optimizations and changes in object reconstruction and identification in Run 2. This analysis does not make explicit selections using photons, missing E_T, or jets, though missing E_T is used to separate our WH and ZH categories when calculating our fake rates and it is also an important input into our Higgs mass variables (MMC and $M_{2\mathrm{T}}$). Several of our analysis regions also include b-jet vetos, which implicitly places object requirements on jets.

- **Electrons:** Electrons are required to pass a minimum p_T threshold of 13 GeV, have an $|\eta| < 2.47$, have good reconstruction quality, and pass the Loose likelihood working point for electron identification (corresponding to an average of 93% efficiency for 2015-2017 datasets) [27]. For an electron to qualify as part of one of our signal categories, however, it must pass the Tight working point (80% efficient). Electrons additionally must pass overlap removal as described in the previous subsection. Electrons must also pass an isolation criterion called `FCLoose`, short for "Fixed Cut Loose", defined for both calorimeter-level and track-level isolation [97]. At the calorimeter level, the calorimeter cluster associated with the electron candidate must satisfy

$$E_T^{\mathrm{iso}}(\Delta R < 0.2)/E_T < 0.2,$$

  meaning the combined transverse energies of the calorimeter topo-clusters in a cone of $\Delta R < 0.2$ around the electron candidate must be less than 20% of the total transverse energy of the electron candidate. Similarly, at the track level, the electron candidate must satisfy

$$p_T^{\mathrm{iso}}(\Delta R^{\mathrm{var}} < 0.3)/p_T < 0.15,$$

  meaning the scalar sum of good-quality electron tracks in a cone of a p_T-dependent ΔR value around the electron candidate must be less than 15% of the total p_T of the electron candidate. Good-quality tracks are defined as those with $p_T > 1$ GeV and with longitudinal impact parameter $|z_0 \sin(\theta)| < 3$ mm, meaning the distance of closest approach in the longitudinal plane to the track. The varying parameter ΔR^{var} has a maximum value of 0.2 and decreases as a function of p_T in increments of 10 GeV/p_T.

- **Muons:** Muons are required to pass a minimum p_T threshold of 9 GeV, have an $|\eta| < 2.5$, have good reconstruction quality, and pass the Loose muon identification working point (corresponding to over 98% efficiency) [28]. For a muon to qualify as part of one of our signal categories, however, they must pass the Tight working point (between 90-93% efficient, depending on the p_T of the muon). They additionally must pass overlap removal as described in the previous subsection. Muons must also pass a track-based isolation criterion called `TightTrackOnly_FixedRad`, defined as

$$p_T^{\text{varcone30}} < 0.06 \cdot p_T^{\mu},$$

meaning the scalar sum of the muon p_T within a cone of $\Delta R = \min(10 \text{ GeV}/p_T^{\mu}, 0.3)$ must be less than 6% of the total muon candidate p_T for $p_T^{\mu} < 50$ GeV. For muon candidates $p_T^{\mu} > 50$ GeV, a fixed radius of $\Delta R < 0.2$ is used instead to help reject hadronic activity. Tracks included in this isolation requirement must have a minimum track $p_T > 1$ GeV.

- **Taus:** Taus are required to pass a minimum p_T threshold of 20 GeV and have an $|\eta|$ between 0 and 2.5, excluding the crack region between the barrel and endcap calorimeters ($1.37 < |\eta| < 1.52$). They must also have an absolute electric charge $|q| = 1$ and either 1 or 3 charged tracks. The baseline tau requirements for our analysis do not require a specific RNN tau ID working point, but instead make the cut `JetRNNSigTransMin` > 0.01, meaning taus must not fall in the bottom 1% of the RNN tau ID score distribution that has been flattened vs. μ and p_T. For a tau to qualify as part of one of our signal categories, however, it must pass the Medium RNN ID working point (75% efficient for 1-prong taus and 60% efficient for 3-prong taus). Taus must also pass the standard overlap removal process.

- **Jets:** Jets are required to pass a minimum of p_T threshold of 20 GeV and fall within

$|\eta| < 4.5$. Jets with $p_T < 60$ GeV and $|\eta| < 2.4$ must also pass a Jet Vertex Tagger (JVT) cut, a multivariate algorithm designed to help suppress pileup events [98]. Jet cleaning criteria are also applied to minimize jets coming from non-prompt processes. We also apply a b-tagging requirement to our jets with a fixed efficiency of 85%.

6.4 $W \rightarrow l\nu$, $H \rightarrow \tau_{\mathbf{lep}}\tau_{\mathbf{had}}$ Analysis Region

In this and the three following subsections, I will detail the selection requirements placed on our four analysis regions. Each of these is structured with an initial **preselection** stage that forms the baseline cuts for all subcategories under that same analysis region. From there, additional cuts are applied to form the final **signal region** analysis region. Some of these signal region cuts are altered or negated to form one or more **control regions** that aid primarily in the validation of our background estimation methods.

These cuts are based on the selection criteria used for the Run 1 analysis, but have been restructured for consistency across the four analysis regions such that what we refer to as "preselection" in each case is truly a baseline for both the signal and control regions. They have also been updated to include the latest recommendations for object reconstruction and identification in Run 2. A summary table of all signal selections in the Run 2 analysis is depicted in Table 10.

When applying our trigger selections, only one trigger is associated with an event, and if multiple triggers are fired for one event, the trigger is chosen with the following order of preference: (1) single muon; (2) single electron; (3) di-muon; (4) di-electron; (5) electron + muon.

These selections are subject to change for the Run 2 analysis as we continue to establish our anticipated background levels. For example, the thresholds on the b-jet veto, $|p_T|$ sums, ΔR requirements, etc. may need adjustments to suppress additional background events in our analysis regions. We may also introduce additional control regions to take advantage of increased statistics throughout.

Category	Selections
$W \to l\nu_l$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$	Taus: • Exactly one τ_{had} passing medium RNN ID • τ_{had} $p_T > 25$ GeV • τ_{had} has opposite charge as other leptons Leptons: • Exactly two tight, isolated leptons (electrons or muons) • Same electric charge Other: • $\lvert p_T \rvert$ of both leptons + $\lvert p_T \rvert$ of $\tau_{\text{had}} > 80$ GeV • Veto b-jets
$W \to l\nu_l$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$	Taus: • Exactly two τ_{had} passing medium RNN ID • $p_T > 20$ GeV for each τ_{had} • Opposite electric charges • Sum of $\lvert p_T \rvert > 100$ GeV • $0.8 < \Delta R(\tau^1_{\text{had}}, \tau^2_{\text{had}}) < 2.8$ Leptons: • Exactly one isolated lepton l ($l = e, \mu$) • $m_T(l, E_T^{\text{miss}}) > 20$ GeV Other: • Veto b-jets
$Z \to ll$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$	Taus: • Exactly one τ_{had} passing medium RNN ID • τ_{had} $p_T > 20$ GeV • Opposite electric charge as the lepton from the Higgs boson • Sum of $\lvert p_T \rvert$ from both taus > 60 GeV Leptons: • Exactly three leptons l ($l = e, \mu$) • Two of these must be a same-flavor, opposite-sign particle-antiparticle pair with invariant mass between 80 and 100 GeV
$Z \to ll$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$	Taus: • Exactly two τ_{had} passing medium RNN ID • $p_T > 20$ GeV for each τ_{had} • Opposite electric charges • Sum of $\lvert p_T \rvert > 88$ GeV Leptons: • Exactly one particle-antiparticle pair of electrons or muons • Combined invariant mass between 60 and 120 GeV

Table 10: Run 2 preselections + signal selections for the four leptonic VH analysis regions. These have been adapted from the Run 1 analysis for the new Run 2 analysis environment, but have not been fully optimized.

6.4.1 Preselection

$W \to l\nu$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ events must fire one of the trigger types summarized in Table 9. Following trigger selection, we require exactly two reconstructed, isolated leptons passing the *Tight* identification working point and exactly one reconstructed hadronic tau passing the *Medium* RNN identification working point. The hadronic tau is additionally required to pass $p_T^\tau > 25$ GeV.

6.4.2 Signal Selection

Following preselection, we further define our signal selection in this category by requiring that the two selected leptons have the same sign (to avoid selecting $Z \to +$ 1 jet events) and the hadronic tau has the opposite sign. All three objects then must pass a requirement on their combined p_T to help reduce multijet backgrounds, which often consist of low-p_T jets: $p_T^{l_1} + p_T^{l_2} + p_T^\tau > 80$ GeV. Finally, we apply a b-jet veto, meaning we require that signal events have no jets with the signature of a jet initialized by a b-quark, to reduce $t\bar{t}$ backgrounds.

The motivation for the b-tagged jet veto comes from the CKM matrix. The top quark must decay through a W boson and a bottom-type quark (d, s, or b), but approximately 90% of the time, we can expect it to decay into a W boson and b quark due to the relative dominance of the $|V_{tb}|$ element of the CKM matrix compared to the other elements involving top quarks. Finally, to calculate the $M_{2\text{T}}$ lower bound on the Higgs boson mass, we choose the lepton with the lowest reconstructed p_T as the lepton associated with the Higgs boson.

6.4.3 Control Regions

- $Z \to \tau\tau$ Control Region
  - Following preselection, we further define the $Z \to \tau\tau$ control region by imposing a b-jet veto and requiring that the two leptons have opposite sign, therefore negating one of the signal selection criteria and increasing the likelihood that the two leptons came from the same neutral parent particle (i.e. the Z). We then require that the invariant mass of the dilepton pair falls between [60,120] GeV – a wide window around the Z mass peak. These selections optimize for

$Z \to \tau_{\text{lep}}\tau_{\text{lep}}$+1 jet background events for which the jet has been misidentified as a tau.

- $t\bar{t}$ Control Region
  - Following preselection, we further define the $t\bar{t}$ control region by requiring at least 1 b-tagged jet and requiring that the two leptons be opposite-sign. We expect that a top quark pair decaying through two W bosons and two b quarks will yield opposite-sign leptons from the opposite-sign W decays.

6.5 $W \to l\nu$, $H \to \tau_{\text{had}}\tau_{\text{had}}$ Analysis Region

6.5.1 Preselection

$W \to l\nu$, $H \to \tau_{\text{had}}\tau_{\text{had}}$ events must fire one of the *single-lepton* trigger types summarized in Table 9. Following trigger selection, we require exactly one reconstructed, isolated leptons passing the *Tight* identification working point and exactly two reconstructed hadronic tau passing the *Medium* RNN identification working point.

6.5.2 Signal Selection

Following preselection, we further define our signal selection in this category by requiring that the two hadronic taus have opposite sign. We then apply a b-jet veto, i.e. requiring that there are no b-tagged jets in the event, to reduce $t\bar{t}$ backgrounds. Next, we require that the two hadronic taus fall within the following range in angular separation ΔR: $0.8 < \Delta R(\tau_0, \tau_1) < 2.8$. We then apply a cut on the transverse mass m_T between the lepton and the missing transverse energy E_T in the event: $m_T(l, E_T^{\text{miss}}) > 20$ GeV, targeting the reduction of $Z \to \tau\tau$ background events. Finally, we require that the sum of the p_T of each hadronic tau is greater than 100 GeV to suppress multijets, W+jets, and Z+jets events that contribute low-p_T jets that may be misreconstructed as taus.

6.5.3 Control Regions

- W+jets Control Region

- This control region marks the dominant background in the $W \to l\nu$, $H \to \tau_{\text{had}}\tau_{\text{had}}$ channel. To target W+jets events, we require same-sign hadronic taus, thereby negating one of the signal region criteria, and apply the transverse mass cut $m_T(l, E_T^{\text{miss}}) > 60$ GeV to avoid selecting Z+jets events.

- $Z \to \tau_{\text{lep}}\tau_{\text{had}}$ + jets Control Region
  - This control region first applies a b-jet veto to avoid selecting $t\bar{t}$ events, then requires that the transverse mass $m_T(l, E_T^{\text{miss}}) < 40$ GeV, helping to isolate Z events. Lastly, it selects for events with $M_{2\text{T}} < 60$ GeV.

- $t\bar{t}$ Control Region
  - This control region requires at least one b-tagged jet for the same reasons explained in the $W \to l\nu$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ control region section.

- Same-Sign Taus Control Region
  - This control region applies a b-jet veto and requires that the two hadronic taus be same-sign. Unlike the other control regions, it does not target a particular background source, but provides an orthogonal region in phase space to our signal region to analyze our background compositions.

- $M_{2\text{T}}$ Sideband Control Region
  - This control region, like the Same-Sign Taus control region, does not target a particular background source. It applies a b-jet veto and requires that the Late-Projected Transverse Mass ($M_{2\text{T}}$) of the Higgs parent particle falls in either of the mass variable sideband regions: $M_{2\text{T}} < 60$ GeV or $M_{2\text{T}} > 120$ GeV.

6.6 $Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ Analysis Region

6.6.1 Preselection

$Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ events must fire one of the trigger types summarized in Table 9. Following trigger selection, we require exactly three reconstructed, isolated leptons passing the *Tight* identification working point and exactly one reconstructed hadronic tau passing the *Medium* RNN identification working point. Among the three leptons, we require that there exist at least one pair of same-flavor, opposite-sign leptons. If there is only one such pair, we associate that pair with the two leptons coming from the Z boson. If more than one such pair exists, we choose the pair with the closest invariant mass to the Z mass to associate with the $Z \to ll$ decay. Lastly, we require that the calculated MMC value for the event is strictly positive (MMC > 0) to eliminate any events for which the MMC algorithm did not converge.

6.6.2 Signal Selection

Following preselection, we further define our signal selection in this category by requiring that the hadronic tau and the lepton associated with the leptonic tau decay have opposite sign, a characteristic of a $H \to \tau\tau$ event. The lepton associated with the leptonic tau decay is chosen as the remaining lepton not included in the same-flavor, opposite-sign lepton pair. Furthermore, we require that the same-flavor, opposite-sign lepton pair has an invariant mass between 80 and 100 GeV, in order to constrain the pair to have a mass within approximately 10 GeV of the Z boson mass (91 GeV). Lastly, we require that the sum of the p_T of the hadronic tau and of the lepton associated with the leptonic tau decay is greater than 60 GeV, to reduce lower-p_T multijet events.

6.6.3 Control Regions

- Same-Sign Taus Control Region
  - In addition to preselection cuts, this control region applies a selection requiring that the hadronic tau and the lepton associated with the leptonic tau decay have the same sign. This negates one of the signal selection criteria.
- MMC Sideband Control Region
  - This control region applies a requirement in addition to the preselection that the hadronic tau and lepton associated with the leptonic tau decay be opposite-sign

and that the MMC of this pair fall within one of two sidebands of the MMC signal region of interest: MMC < 80 or MMC > 160 GeV.

6.7 $Z \to ll$, $H \to \tau_{\mathbf{had}}\tau_{\mathbf{had}}$ Analysis Region

6.7.1 Preselection

$Z \to ll$, $H \to \tau_{\text{had}}\tau_{\text{had}}$ events must fire one of the trigger types summarized in Table 9. Following trigger selection, we require exactly two reconstructed, isolated leptons passing the *Tight* identification working point and exactly two reconstructed hadronic taus passing the *Medium* RNN identification working point. We require that the pair of leptons have the same flavor but opposite sign, and associate the pair with the two leptons coming from the Z boson. Lastly, we require that the MMC calculated using the two hadronic taus is strictly positive (MMC > 0) to eliminate any events for which the MMC algorithm did not converge.

6.7.2 Signal Selection

Following preselection, we further define our signal selection in this category by requiring that the two hadronic taus have opposite sign and that $p_T^{\tau_1} + p_T^{\tau_2} > 88$ GeV. These cuts target $H \to \tau\tau$ events and suppress $Z +$ jets backgrounds. Furthermore, we require that the same-flavor, opposite-sign lepton pair has an invariant mass between 60 and 120 GeV, in order to constrain the pair to have a mass within approximately 30 GeV of the Z boson mass (91 GeV).

6.7.3 Control Regions

- Same-Sign Taus Control Region
  - In addition to preselection cuts, this control region applies a selection requiring that the two hadronic taus have the same sign. This negates one of the signal selection criteria.
- MMC Sideband Control Region

– This control region applies a requirement in addition to the preselection that the two hadronic taus be opposite-sign and that the MMC of the hadronic tau pair fall within one of two sidebands of the MMC signal region of interest: MMC < 80 or MMC > 160 GeV.

6.8 Higgs Boson Mass Reconstruction

Reconstructing the mass of a resonance such as a boson decaying to multiple taus is difficult due to the presence of a neutrino in the final state associated with each tau in the decay. Since neutrinos are not detectable with the ATLAS detector, we can only infer their characteristics based on missing transverse energy (E_T^{miss}). This is an event-level quantity, not an object-level quantity, and therefore it is not obvious how to make judgments about how many neutrinos were present in the event, what their energies were, and what directions they went in.

Following the reduction of our real and fake backgrounds, the amount of signal observed is determined via a fit to the reconstructed mass of particles assumed to have come from the Higgs boson in each of the four final state categories. Given that there are between two and four neutrinos in each final state, however, it is necessary to arrive at this reconstructed mass via an additional calculation. In ZH events with either two or three final-state neutrinos, we use the Missing Mass Calculator (MMC) method [99] to generate a most-likely Higgs boson mass. In WH events, we can no longer assume that the missing E_T is entirely associated with the Higgs boson and its tau decay products, as the W boson will introduce an additional neutrino to the final state. For this reason, instead of using MMC, we use the Late-Projected Transverse Mass ($M_{2\text{T}}$) method to generate an event-by-event lower bound on the reconstructed Higgs boson mass [100].

6.8.1 Missing Mass Calculator (MMC)

Without a more sophisticated strategy, we could approach the challenge of calculating the estimated mass of the Higgs boson in a $Z \to ll$, $H \to \tau\tau$ event by simply calculating the mass of the Higgs boson based on the masses of its visible decay products and adding the missing E_T. This method is suboptimal because it does not take the geometric configuration of the neutrinos into account, and it is often the case that two taus produced in a $H \to \tau\tau$ decay are produced almost back-to-back, meaning their missing E_T will partially cancel out. This will produce a wide distribution for the $H \to \tau\tau$ mass that is difficult to separate from other background sources of ditau decays.

Another method used for this problem is the collinear mass approximation, which assumes that each neutrino is approximately angularly aligned with its associated visible tau decay products. While this method yields a reasonable mass resolution for boosted $H \to \tau\tau$ decays, the collinear assumption is not valid for many $H \to \tau\tau$ events.

The Missing Mass Calculator (MMC) method provides a more precise manner of calculating a most-likely parent particle mass when that parent also decays into multiple sources of missing E_T for all $H \to \tau\tau$ event topologies [99]. It uses a likelihood minimization strategy on simulated distributions of kinematically-allowed configurations of the neutrinos to choose a most-likely geometry of the event, then calculates a most-likely parent mass.

A system of four equations must be solved to exactly calculate the ditau mass:

$$E_{T_x}^{\text{miss}} = \vec{p}_{\text{miss}_1} \sin(\theta_{\text{miss}_1}) \cos(\phi_{\text{miss}_1}) + \vec{p}_{\text{miss}_2} \sin(\theta_{\text{miss}_2}) \cos(\phi_{\text{miss}_2}) \tag{86}$$

$$E_{T_y}^{\text{miss}} = p_{\text{miss}_1} \sin(\theta_{\text{miss}_1}) \sin(\phi_{\text{miss}_1}) + p_{\text{miss}_2} \sin(\theta_{\text{miss}_2}) \sin(\phi_{\text{miss}_2}) \tag{87}$$

$$M_{\tau_1}^2 = m_{\text{miss}_1}^2 + m_{\text{vis}_1}^2 + 2\sqrt{p_{\text{miss}_1}^2 + m_{\text{vis}_1}^2}\sqrt{p_{\text{vis}_1}^2 + m_{\text{vis}_1}^2} - 2p_{\text{miss}_1}p_{\text{vis}_1}\Delta\theta_1 \tag{88}$$

$$M_{\tau_2}^2 = m_{\text{miss}_2}^2 + m_{\text{vis}_2}^2 + 2\sqrt{p_{\text{miss}_2}^2 + m_{\text{vis}_2}^2}\sqrt{p_{\text{vis}_2}^2 + m_{\text{vis}_2}^2} - 2p_{\text{miss}_2}p_{\text{vis}_2}\Delta\theta_2 \tag{89}$$

$E_{T_x}^{\text{miss}}$ and $E_{T_x}^{\text{miss}}$ refer to the $\hat{x}$ and $\hat{y}$ components of the event missing E_T, while M_{τ_1} and M_{τ_2} refer to the rest masses of the two taus. m_{miss_1} and m_{miss_2} refer to the missing invariant mass from the additional neutrino associated with a leptonic tau decay, while $\vec{p}_{\text{miss}_1}$ and $\vec{p}_{\text{miss}_2}$ are the missing momenta of the two tau neutrinos. m_{vis_1} and m_{vis_2} refer to the visible invariant mass of the tau decay products, while $\vec{p}_{\text{vis}_1}$ and $\vec{p}_{\text{vis}_2}$ are the visible momenta of each tau's decay products. The visible and invisible components of the tau decay products

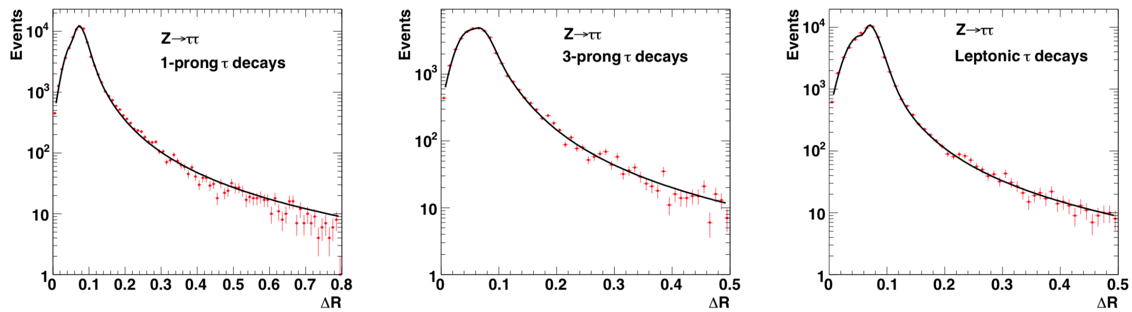


Fig. 56: Example simulated ΔR distributions, along with their fits, are shown for a given value of p_τ. These distributions correspond to a (left) 1-prong hadronic tau decay, (middle) 3-prong hadronic tau decay, and (right) leptonic tau decay. [99]

each have their own θ and ϕ. Lastly, $\Delta\theta$ indicates the angle between each tau's visible and invisible decay product momenta.

For $Z \to ll$, $H \to \tau_{\text{had}}\tau_{\text{had}}$ events, there is no additional neutrino from a leptonic tau decay, so m_{miss_1} and m_{miss_2} are set to 0. There are then 6 unknowns (3 each from the missing momenta of the neutrino involved in each of the two hadronic tau decays, $\vec{p}_{\text{miss}_1}$ and $\vec{p}_{\text{miss}_2}$), and therefore this system of equations is underconstrained. However, given two more inputs, say ϕ_{miss_1} and ϕ_{miss_2}, the masses could be exactly calculated. The MMC algorithm therefore calculates the ditau mass exactly for several pairs of (ϕ_{miss_1},ϕ_{miss_2}) in a grid in kinematically-allowed phase space. Each of these masses is then weighted by a probability drawn from fits of each of the simulated ΔR distributions: $\mathcal{P}(\Delta R_1, p_{\tau_1}) \times \mathcal{P}(\Delta R_2, p_{\tau_2})$ (see Figure 56). The values of (ϕ_{miss_1},ϕ_{miss_2}) are chosen based on which pair minimizes the negative log-likelihood:

$$\mathcal{L} = -\log(\mathcal{P}(\Delta R_1, p_{\tau_1}) \times \mathcal{P}(\Delta R_2, p_{\tau_2})). \tag{90}$$

For $Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ events, there is an additional unknown introduced by a m_{miss} variable from the additional neutrino associated with the leptonic tau decay. In this case, the grid search occurs in three dimensions instead of two: (ϕ_{miss_1},ϕ_{miss_2},m_{miss}), where m_{miss} is sampled uniformly across all kinematically-allowed values.

6.8.2 Late-Projected Transverse Mass (M_{2T})

While the MMC technique is particularly effective for the $Z \to ll$, $H \to \tau\tau$ channels, it is not as suitable for application to our $W \to l\nu$, $H \to \tau\tau$ channels. This is because the Z boson has no invisible decay products in our chosen final states, while the W boson will decay into a lepton and neutrino. This neutrino not only adds further unknowns to our already underconstrained system of equations described in Equations 86 - 89, but it also adds a source of missing E_T to the event that is not associated with the Higgs boson decay products. This means that the assumption underlying the formulation of Equations 86 - 89 – that the Higgs boson decay is the only source of missing E_T – is no longer valid.

Instead, we use the Late-Projected Transverse Mass ($M_{2\mathrm{T}}$) method [100] for our Higgs boson mass variable in the $W \to l\nu$, $H \to \tau\tau$ channels. Unlike MMC, which provides a most-likely Higgs boson mass, the $M_{2\mathrm{T}}$ method provides an event-by-event lower bound for the Higgs boson mass. The term "late-projected" refers to the ordering of the two primary operations involved in this process: (1) combining the desired momentum vectors via a summation and (2) projecting them into the transverse plane. "Late-projected" therefore indicates that the projection happens *after* the momentum summation. Unlike the combination of *space-like* vectors such as p_T, for which this order of operations has no effect on the final result, combining *time-like* quantities such as E_T or m_T in the transverse plane does require careful attention to this ordering.

One might wonder why we bother calculating the transverse mass, m_T, rather than attempting to calculate the full invariant mass. This is because the incoming momenta of the proton beams colliding at the LHC is not exactly known. However, the incoming *transverse* momentum, in the $\hat{x} - \hat{y}$ plane orthogonal to the beamline, is known to be almost exactly 0. It is therefore advisable to design our analysis such that is invariant to changes in the proton beam p_z, or momentum along the beamline.

$M_{2\mathrm{T}}$ is a specific case of the generic method $M_{N\mathrm{T}}$, where N refers to the number of parent particles involved as well as the number of mass inputs that parameterize the invisible sector. For our use case, we have 2 parent particles: the Higgs boson and the W boson. In the first stage of $M_{2\mathrm{T}}$, the four-vectors of the visible and invisible decay products of each parent are separately combined. This results in four (1+3)-dimensional four-vectors: P_1^μ and Q_1^μ, representing the combined visible and invisible momenta of Parent #1, and P_2^μ and Q_2^μ, representing the combined visible and invisible momenta of Parent #2. The combined visible and invisible vectors for Parent #1, for example, take the form:

$$P_1^\mu = \begin{pmatrix} E_1 \\ p_{1x} \\ p_{1y} \\ p_{1z} \end{pmatrix} = \begin{pmatrix} E_1 \\ \vec{p}_{1T} \\ p_{1z} \end{pmatrix} \text{; where } E_1 = \sum_{i \,\in\, \text{visible daughters}} \sqrt{M_i^2 + (\vec{p}_{iT})^2 + p_{iz}^2} \tag{91}$$

$$Q_1^\mu = \begin{pmatrix} \tilde{E}_1 \\ q_{1x} \\ q_{1y} \\ q_{1z} \end{pmatrix} = \begin{pmatrix} \tilde{E}_1 \\ \vec{q}_{1T} \\ q_{1z} \end{pmatrix} \text{; where } \tilde{E}_1 = \sum_{i \,\in\, \text{invisible daughters}} \sqrt{\tilde{M}_i^2 + (\vec{q}_{iT})^2 + q_{iz}^2} \tag{92}$$

Once these combined vectors have been constructed, they are then projected into the transverse plane. This yields four (1+2)-dimensional vectors indexed by α. The transverse-projected versions of the (1+3)-dimensional vectors shown above are:

$$p_{1T}^\alpha = \begin{pmatrix} e_{1T} \\ \vec{p}_{1T} \end{pmatrix} \text{; where } e_{1T} = \sqrt{M_1^2 + (\vec{p}_{1T})^2} = \sqrt{E_1^2 - p_{1z}^2} \tag{93}$$

$$q_{1T}^\alpha = \begin{pmatrix} \tilde{e}_{1T} \\ \vec{q}_{1T} \end{pmatrix} \text{; where } \tilde{e}_{1T} = \sqrt{\tilde{M}_1^2 + (\vec{q}_{1T})^2} = \sqrt{\tilde{E}_1^2 - q_{1z}^2} \tag{94}$$

The transverse masses of Parents #1 and #2 may then be constructed as:

$$\mathcal{M}_{1T} \equiv \sqrt{(e_{1T} + \tilde{e}_{1T})^2 - (\vec{p}_{1T} + \vec{q}_{1T})} \tag{95}$$

$$\mathcal{M}_{2T} \equiv \sqrt{(e_{2T} + \tilde{e}_{2T})^2 - (\vec{p}_{2T} + \vec{q}_{2T})} \tag{96}$$

Since we expect the Higgs boson to be more massive than the W boson, we choose the heaviest of these two masses, but with an additional constraint that the sum of the invisible momenta, $\sum_i \vec{q}_{iT}$, must equal the total missing $\vec{p}_T$ in the event. The final $M_{2\mathrm{T}}$ variable is therefore chosen as the minimum value of the more massive parent particle's transverse mass subject to this constraint:

$$M_{2\mathrm{T}} = \min_{\sum \vec{q}_{iT} = \not{\vec{p}}_T} [\max\{\mathcal{M}_{1T}, \mathcal{M}_{2T}\}] \tag{97}$$

For our analysis, we additionally require that the invariant mass of the W boson decay products be as close as possible to the W boson rest mass.

Performing this minimization in full would require optimizing across a 9- or 12-dimensional phase space for the $W \to l\nu$, $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ and $W \to l\nu$, $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channels. To reduce to the dimensionality of this problem, the collinear approximation is used, meaning that neutrinos are assumed to be aligned with their affiliated visible tau decay products. This allows neutrino momenta to be calculated via

$$\vec{p}_\nu = \left(\frac{1}{x} - 1\right)\vec{p}_{\mathrm{vis}}, \text{ where } x = \frac{p_{\mathrm{vis}}}{p_{\mathrm{vis}} + p_\nu}. \tag{98}$$

These constraints and approximations reduce a previously 9- to 12-dimensional optimization problem into one parametrized by just three variables:

1. $q_{z,\nu}$, the $\hat{z}$-momentum of the neutrino from the W decay
2. x_1, the fraction of τ_1 momentum that goes into visible decay products
3. x_2, the fraction of τ_2 momentum that goes into visible decay products

By construction, the actual mass of the heaviest parent particle must be greater than or equal to $M_{2\mathrm{T}}$, and the actual mass forms an upper bound for this quantity. Additionally, by using the "late-projected" rather than "early-projected" formulation of transverse mass, we take advantage of the fact that $M_{2\mathrm{T}} = M_2$. In other words, this method uses transverse information, but is equivalent to using (1+3)-dimensional vectors throughout instead. M_2 is not exactly the same as the true invariant mass of the most massive parent particle, as it is a function of the combined visible and invisible momenta and not the *individual* visible and invisible momenta. Given our ignorance of the actual neutrino kinematics, however, this method does a good job of constraining our Higgs boson mass with a reasonable number of parameters.

7 Background Estimation for $V \to$ leptons, $H \to \tau\tau$

7.1 Reducible Background Estimation

The largest source of background events for this analysis come from events for which one or more reconstructed hadronic taus or leptons is actually a misidentified jet or non-prompt lepton. These events are primarily from $Z \to \tau\tau$ and $t\bar{t}$ processes. We call these events "fakes" and use a data-driven background estimation strategy called the Fake Factor Method to predict the quantity of these events expected in our signal region [101].

7.1.1 $Z \to ll$ + Jets Selection

The fake-enriched analysis region uses $Z \to ee$ + jets and $Z \to \mu\mu$ + jets MC samples. It is designed to reflect our analysis region closely without overlapping with any of our actual analysis phase space. It therefore uses the same trigger selection as in our signal regions. Additionally, it requires that there exists at least one pair of light leptons l, where $l = \{e, \mu\}$, that are same-flavor, opposite-sign, pass the *Medium* identification working point, and pass their respective isolation criteria. The electron, muon, and tau fake rates are calculated using additional electron, muon, or tau objects in the event in addition to this baseline selection.

7.1.2 The Fake Factor Method

Our analysis' signal regions are composed of objects (electrons, muons, and taus) that we pass through various selection cuts until we choose a subset of these objects. These objects, however, are fundamentally a mix of correctly- and incorrectly-identified electrons, muons, and taus. For explanatory purposes, I will only focus on fake taus here, though the method can be straightforwardly expanded to include multiple possible fake categories.

We can express the number of what we call "selected taus" in our analysis regions as:

$$N_{\text{selected}} = \epsilon N_{\text{real taus}} + r N_{\text{fake taus}}, \tag{99}$$

where ϵ represents the selection efficiency for true taus (assumed to be 100% for this method) while r represents the selection efficiency of fake taus. r is also called the **fake rate**. In other words, we can assume that some portion of the taus we select in our analysis will actually be fakes, and it is useful to state this explicitly in order to anticipate our expected fake contributions.

The fake rate r is calculated in data in the fake-enriched $Z \to ll$ + jets region described above. It is calculated individually for 1-prong WH and ZH events as well as 3-prong WH and ZH events. Following the calculation of the fake rate r for a particular analysis region, the **fake factor** f is then used to extrapolate the number of selected fakes from our fake-enriched region to our signal region:

$$N^{\text{VH data}}_{\text{selected taus}} = N^{\text{VH MC}}_{\text{selected real taus}} + f(N^{Z\to ll\text{ + jets data}}_{\text{anti-selected fake taus}} - N^{Z\to ll\text{ + jets MC}}_{\text{anti-selected true taus}}), \tag{100}$$

where f is defined as:

$$f = \frac{r}{1-r}. \tag{101}$$

The fake factor f provides an extrapolation factor that is then applied to the number of anti-selected taus measured in the fake-enriched $Z \to ll$ + jets region in data following a subtraction (calculated in the $Z \to ll$ + jets region in MC) of any real taus that were labeled as anti-taus in the fake-enriched region. This extrapolation factor allows us to estimate the contribution of fake taus to our selected VH signal regions in data. While the fake rate r expresses the ratio of the number of selected taus to the total number of reconstructed tau candidates in the fake-enriched region, the fake factor f expresses the ratio of the number of selected taus to the number of anti-selected taus in the fake region. It is therefore the fake factor f that we apply to the number of anti-selected tau candidates in the fake region in Equation 100 in order to get a projected contribution for our signal region.

The actual implementation of the Fake Factor Method in our analysis is complicated by the fact that multiple types of objects can be faked (electrons, muons, and taus) and that there are three (four) selected objects in our WH (ZH) signal regions that could potentially be faked in each event. Since we don't have complete information as to whether or not a given tau candidate in data is real or fake, we construct the mathematical expression of the Fake Factor Method based on observable quantities like the number of selected or anti-selected events. These expressions take into account that while there are a fixed number of selected objects for a given analysis region, there are many potential objects in the event that could fake a given selected object.

The 1-object case of the Fake Factor Method, relevant for an analysis with one selected object in its final state, results in the following expression:

$$N_{\bar{S}} = fN_A, \tag{102}$$

where $N_{\bar{S}}$ represents the number of objects entering the signal region as a result of the fake factor extrapolation and N_A is the number of anti-selected objects measured in the fake region. Each additional anti-selected object requires an additional fake factor. This is a simplified expression encapsulating a potentially infinite sum of possibilities of which one object out of the arbitrarily many available objects is selected for a given event. In each of these cases, f could potentially be different depending on the type and kinematics of the selected object.

The 2-object case, for an analysis with 2 selected final-state objects, results in:

$$N_{\bar{S}\bar{S}} = fN_{SA} - ffN_{AA}. \tag{103}$$

This extends the 1-object case by including a subtraction term associated with scenarios in which the two final-state objects are anti-selected instead of selected. Again, f will vary for each term in the summation based on the selected objects in each case.

Finally, we can move on to the 3-object and 4-object cases, which are the ones actually relevant for our WH and ZH signal regions. These follow the convention that terms are positive with an odd number of fake factors and negative with an even number of fake factors:

$$N_{\bar{S}\bar{S}\bar{S}} = fN_{SSA} - ffN_{SAA} + fffN_{AAA} \tag{104}$$

$$N_{\bar{S}\bar{S}\bar{S}\bar{S}} = fN_{SSSA} - ffN_{SSAA} + fffN_{SAAA} - ffffN_{AAAA} \tag{105}$$

The full derivations of these expressions are tedious, and I recommend referring to the thorough treatment in [102] for more detail.

7.1.3 Tau Fake Rates

The tau fake rate r is defined as the ratio of the number of selected taus to the total number of reconstructed tau candidates in the $Z \to ll$ + jets fake-enriched region in data. Reconstructed tau candidates must pass all of the selections described in Section 6.3.3 except for the Medium RNN ID score requirement. Selected taus must pass all of these selections, including the Medium RNN ID score requirement.

Preliminary tau fake rate measurements, along with statistical uncertainties, are reported in Tables 11 - 14 for approximately 20 million events passing our $Z \to ll$ + jets selection in Run 2 data. 1-prong tau fake rates are significantly higher than 3-prong tau fake rates, reflecting the fact that 3-prong taus are more difficult for a jet to fake. Tau fake rates are currently binned in p_T and $|\eta|$, though other parameterizations are being explored for the final analysis result.

Tau Fake Rate	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.37$	$1.37 < \|\eta\| < 2.5$
$20 < p_T < 25$ GeV	0.152 ± 0.002	0.150 ± 0.002	0.124 ± 0.002
$25 < p_T < 30$ GeV	0.151 ± 0.003	0.152 ± 0.003	0.131 ± 0.002
$30 < p_T < 35$ GeV	0.145 ± 0.003	0.133 ± 0.004	0.119 ± 0.003
$35 < p_T < 40$ GeV	0.133 ± 0.004	0.125 ± 0.005	0.109 ± 0.004
$40 < p_T < 60$ GeV	0.123 ± 0.003	0.115 ± 0.004	0.094 ± 0.003
$p_T > 60$ GeV	0.105 ± 0.004	0.097 ± 0.005	0.069 ± 0.004

Table 11: Preliminary WH, $H \to \tau\tau$ fake rates for 1-prong taus.

Tau Fake Rate	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.37$	$1.37 < \|\eta\| < 2.5$
$20 < p_T < 25$ GeV	0.029 ± 0.001	0.033 ± 0.001	0.023 ± 0.001
$25 < p_T < 30$ GeV	0.030 ± 0.001	0.032 ± 0.002	0.027 ± 0.001
$30 < p_T < 40$ GeV	0.025 ± 0.001	0.027 ± 0.001	0.024 ± 0.001
$p_T > 40$ GeV	0.020 ± 0.001	0.021 ± 0.001	0.015 ± 0.001

Table 12: Preliminary WH, $H \to \tau\tau$ fake rates for 3-prong taus.

7.1.4 Fake Factor Validation

The modeling of the Fake Factor Method is assessed using closure tests, meaning a comparison of the projected distributions of tau fakes stacked on top of MC estimates with

Tau Fake Rate	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.37$	$1.37 < \|\eta\| < 2.5$
$20 < p_T < 25$ GeV	0.169 ± 0.002	0.166 ± 0.002	0.137 ± 0.002
$25 < p_T < 30$ GeV	0.169 ± 0.003	0.164 ± 0.003	0.143 ± 0.003
$30 < p_T < 35$ GeV	0.151 ± 0.003	0.160 ± 0.004	0.122 ± 0.003
$35 < p_T < 40$ GeV	0.137 ± 0.004	0.134 ± 0.006	0.118 ± 0.005
$40 < p_T < 60$ GeV	0.117 ± 0.003	0.115 ± 0.004	0.099 ± 0.004
$p_T > 60$ GeV	0.089 ± 0.005	0.074 ± 0.006	0.072 ± 0.005

Table 13: Preliminary ZH, $H \to \tau\tau$ fake rates for 1-prong taus.

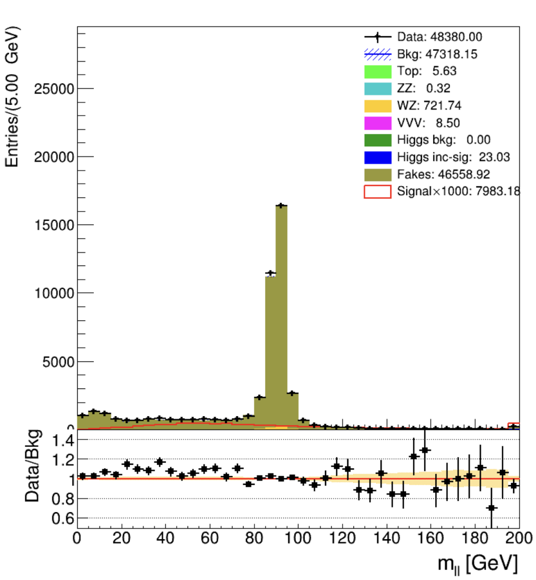
Tau Fake Rate	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.37$	$1.37 < \|\eta\| < 2.5$
$20 < p_T < 25$ GeV	0.029 ± 0.001	0.034 ± 0.001	0.027 ± 0.001
$25 < p_T < 30$ GeV	0.033 ± 0.001	0.038 ± 0.002	0.033 ± 0.001
$30 < p_T < 40$ GeV	0.029 ± 0.001	0.032 ± 0.002	0.022 ± 0.001
$p_T > 40$ GeV	0.016 ± 0.001	0.017 ± 0.002	0.015 ± 0.001

Table 14: Preliminary ZH, $H \to \tau\tau$ fake rates for 3-prong taus.

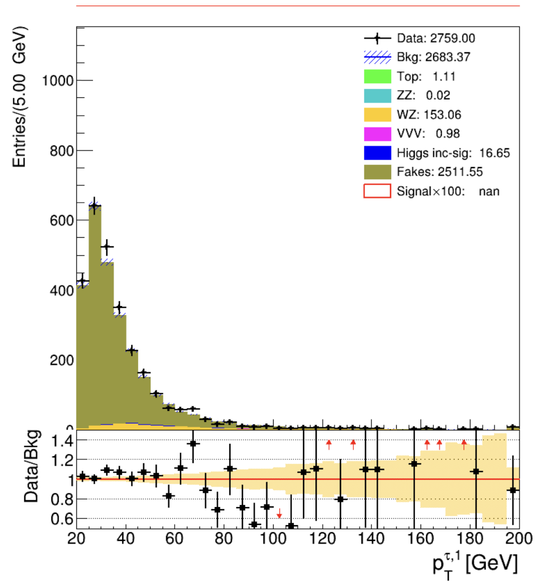
the data for our analysis regions. If the fake model is performing well, the MC+fakes histograms should agree within uncertainties with our data distributions. These closure tests are plotted only in the preselection and control regions, not our final analysis regions, to avoid "unblinding," or revealing our final result before we are prepared to finalize our analysis. Example preliminary closure tests for each of our four analysis preselection regions are shown in Figure 57.

7.1.5 Tau Fake Factor Systematic Uncertainty

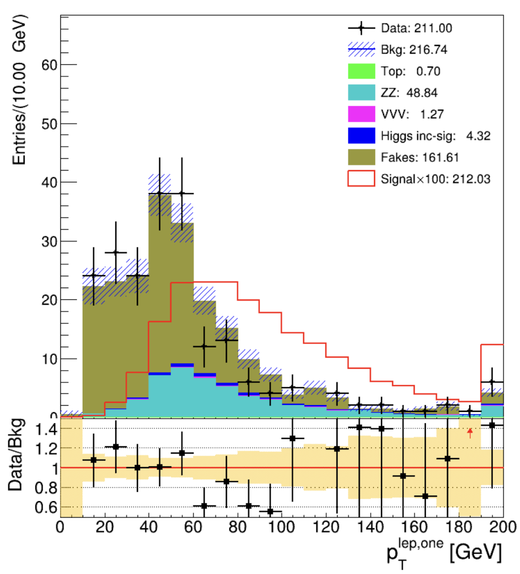
The dominant systematic uncertainty associated with the tau fake factor calculation is in the composition of quark-initiated versus gluon-initated jets in the fake-enriched $Z \to ll$ + jets region compared with our VH analysis regions. The tau fake factor is particularly sensitive to the ratio of quark jets to gluon jets, and unfortunately there's no guarantee that this composition will be equivalent between our VH analysis regions and our $Z \to ll$ + jets region in which we measure our fake rates. Furthermore, the exact composition cannot be exactly determined in data, as we don't have access to the true origin of each jet in a given event. I have investigated the Run 1 method for calculating the tau fake factor systematic uncertainty and have begun evaluating its suitability for the Run 2 analysis. These studies are ongoing, and it is possible that we will select an alternative procedure for our final calculation.



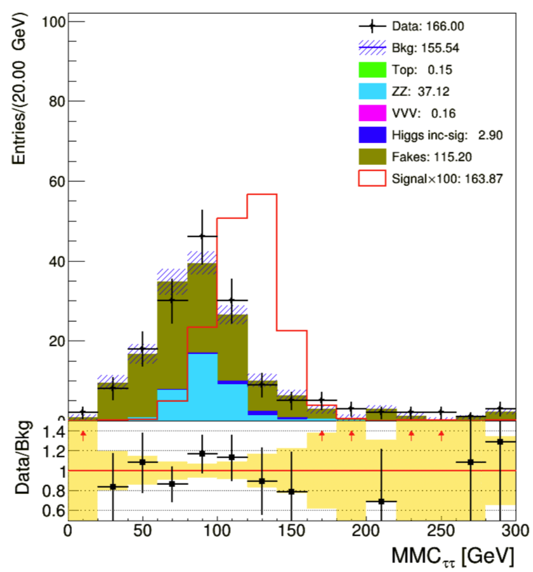
(a) $W \to l\nu_l$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$ (Dilepton mass)



(b) $W \to l\nu_l$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$ (Lead tau p_T)



(c) $Z \to ll$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$ (p_T of leading lepton from same-flavor, opposite-sign pair)



(d) $Z \to ll$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$ (Ditau MMC)

Fig. 57: Preliminary closure tests validating our Fake Factor Method at the preselection level in each of our four signal categories for several different distributions. MC contributions at preselection are stacked with the Fake Factor estimate of fake contributions, and this combination is compared with the actual data distributions. In each of our four analysis regions, these closure tests indicate that our Fake Factor Method is performing well at predicting the contributions from fakes in these regions. This data represents 36.2 fb^{-1} of the Run 2 dataset, and the final validation plots will include even higher statistics.

The Run 1 method for calculating the tau fake factor systematic uncertainty relies on the composition of gluon and quark jets varying with different minimum cuts on the event missing E_T. The strategy is as follows:

1. Apply missing E_T cuts ($E_T^{\text{miss}} > 20$ GeV and $E_T^{\text{miss}} > 30$ GeV) to both MC and data samples in the fake-enriched region ($Z \to ll$ + jets).

2. Calculate the gluon jet fractions r_1 and r_2 in each MC subsample.

3. Calculate the tau fake rates $FR(r_1)$ and $FR(r_2)$ in each data subsample.

4. Calculate the extrapolated values for a fully gluon-dominated fake rate (FR_g) and a fully quark-dominated fake rate (FR_q) by inverting the following relationship:

$$\begin{pmatrix} FR(r_1) \\ FR(r_2) \end{pmatrix} = \begin{pmatrix} r_1 & 1 - r_1 \\ r_2 & 1 - r_2 \end{pmatrix} \begin{pmatrix} FR_g \\ FR_q \end{pmatrix} \tag{106}$$

   to get the relationship:

$$\begin{pmatrix} FR_g \\ FR_q \end{pmatrix} = \frac{1}{r_1(1-r_2) - r_2(1-r_1)} \begin{pmatrix} 1 - r_2 & -(1-r_1) \\ -r_2 & r_1 \end{pmatrix} \begin{pmatrix} FR(r_1) \\ FR(r_2) \end{pmatrix} \tag{107}$$

5. Calculate the fake rate as a function of gluon jet fraction $FR(r)$ using:

$$FR(r) = r \cdot FR_g + (1-r) \cdot FR_q \tag{108}$$

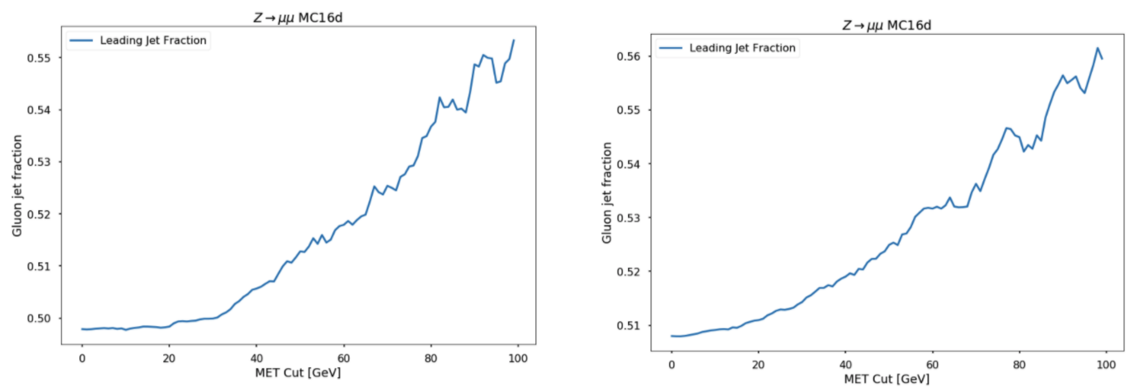
6. Vary r between $r_{\text{nom}}/2$ and $2r_{\text{nom}}$, where r_{nom} is the nominal gluon fraction measured in the fake region.

7. Treat these differences in the fake rate as the systematic uncertainty on the fake rate measurement.

An initial study into the efficacy of this method for the Run 2 analysis environment using approximately 50 fb^{-1} of data revealed that the systematic uncertainties for 3-prong taus were reasonable, but were rather large for 1-prong taus, as shown in Table 15. The nominal fake rate for 1-prong and 3-prong taus, with uncertainties, was measured to be 0.136 ± 0.09 and 0.029 ± 0.0069, respectively. Distributions for the gluon jet fraction and

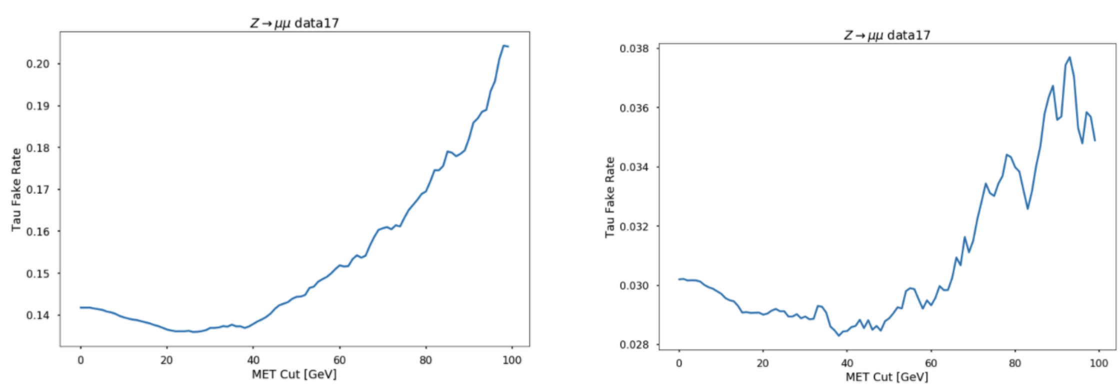
Systematic Uncertainty	$r_{\mathrm{nom}}/2$	$2r_{\mathrm{nom}}$
1-prong taus	-47%	94%
3-prong taus	16%	-32%

Table 15: The systematic uncertainty for 1-prong and 3-prong tau fake rates from a preliminary investigation into the efficacy of the Run 1 method for calculating tau fake factor systematic uncertainty is reported. The 1-prong systematic uncertainties are large enough to merit an investigation of other methods in addition to the Run 1 technique. In Run 1, the largest tau fake factor systematic uncertainties ranged from approximately -15% to 30%.



(a) 1-prong **(b)** 3-prong

Fig. 58: The fraction of leading jets that are gluon-initiated vs. quark-initiated as a function of minimum missing E_T cut for events with a 1-prong or 3-prong leading tau.

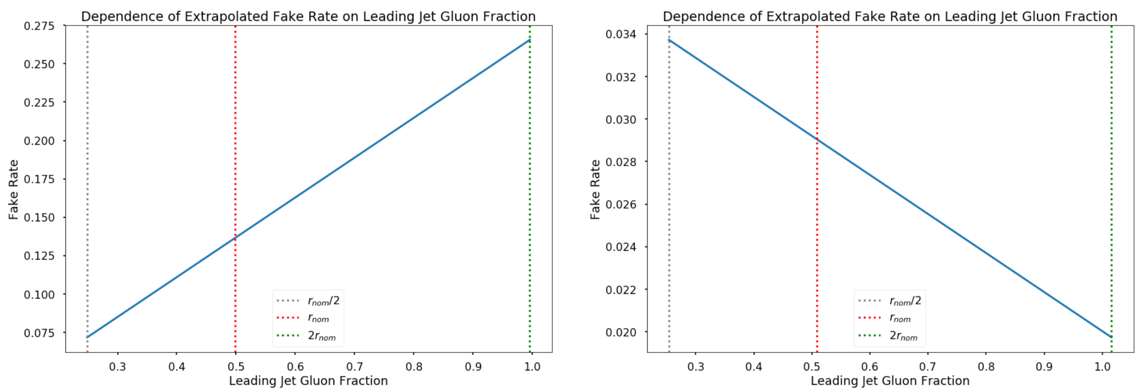


(a) 1-prong **(b)** 3-prong

Fig. 59: The tau fake rate as a function of minimum missing E_T cut, measured in a subset of the Run 2 dataset in the fake-enriched region.

tau fake rates as a function of missing E_T cuts are shown in Figures 58 and 59, while the dependence of the extrapolated fake rate on the gluon jet fractions is shown in Figure 60.

The combined $H \to \tau\tau$ analysis is currently developing a tool for calculating this systematic uncertainty using a template fit method that determines the estimated quark/gluon jet composition in a sample via an interpolation of the tau jet width. Though this tool is not yet recommended for official use, it provides an interesting future direction for potential alternative strategies for this calculation.



(a) 1-prong **(b)** 3-prong

Fig. 60: The extrapolated tau fake rate as a function of gluon jet fraction for the 1-prong and 3-prong cases. As in Run 1, the 1-prong and 3-prong cases are anti-correlated. The three vertical lines correspond to $r_{\text{nom}}/2$, r_{nom}, and $2r_{\text{nom}}$, from left to right.

7.2 Irreducible Background Estimation

There are other physics processes beyond the four signal categories considered in this analysis that can result in the exact same final-state physics objects we seek. When this happens – each hadronic tau and lepton has been correctly reconstructed and identified, but the origin process was not a VH event – we refer to the event as an irreducible background. These are primarily diboson events such as WZ and ZZ for which the Z decays to two taus, mimicing the $H \to \tau\tau$ decay process. In Run 1, the contributions of these backgrounds were estimated from Monte Carlo simulation and subtracted from the final counts in each signal region.

7.2.1 Motivation

Given the increased signal statistics in Run 2, a more sophisticated strategy for estimating the contributions of these backgrounds is warranted. The Run 1 analysis also included a preliminary study using a Boosted Decision Tree (BDT) architecture trained on each signal category to separate WH signal from WZ backgrounds and ZH signal from ZZ backgrounds. Though it was not used in the final analysis, the initial study showed promising results for such a structure – a potential combined improvement in the upper limit of 30%.

Using the input variables of this BDT as a guide, I implemented four separate neural network architectures for each signal region. Using a neural network instead of a BDT allowed more freedom to design layers inspired by the physics of the network during the training process.

7.2.2 Neural Network Architecture

For each of the four analysis regions, I trained a baseline neural network with a single fully-connected hidden layer of 10 nodes and ReLU activation followed by an output layer of 1 node with a sigmoid activation. The performance of this model set the baseline for comparisons when creating the optimized NN architecture.

The optimized NN architecture for each model consisted of two initial transformation layers followed by three fully-connected hidden layers of 128 nodes each, each with a ReLU activation function. The output layer consisted of a single node with a sigmoid activation function.

The intial transformation layers had the following purposes:

1. **Global ϕ Offset:** Add a global ϕ offset to the ϕ input variables, to help the network learn that events can be globally-rotated in ϕ without affecting the classification of the physics event. This layer is only activated during training time, not during model inference.

2. **Angular Encoding:** Split each ϕ input variable into $\sin(\phi)$ and $\cos(\phi)$, to help the network learn that $\phi + 2\pi n \to \phi$ for $n \in \mathbb{Z}$.

Each model's optimized architecture had approximately 32,000 trainable parameters.

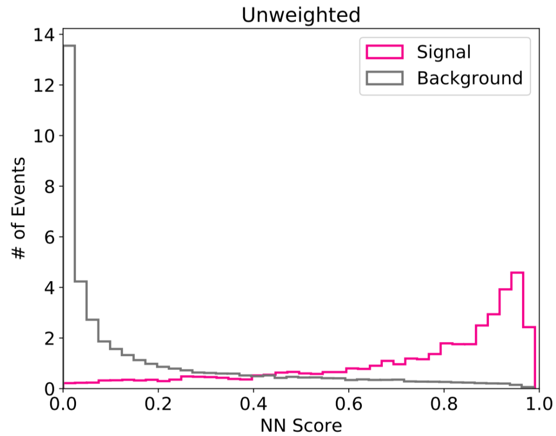
7.2.3 Training Details

70% of each overall simulated dataset was dedicated to the training dataset, while the remaining 30% was split equally between the validation and test datasets. As diboson background statistics sometimes significantly outnumbered the VH signal statistics, class weights were used to balance the training process and account for the imbalance in statistics between the two classes on the training dataset only. MC weights were not used during the training process, though events with negative MC weights (approximately 5% of all events) were excluded from the training dataset. MC weights were then used for the evaluation of the NN performances, however. Training datasets ranged in total size from approximately 50,000 - 200,000 events. Each model was trained using Stochastic Gradient Descent (SGD) with momentum = 0.9 as an optimizer.

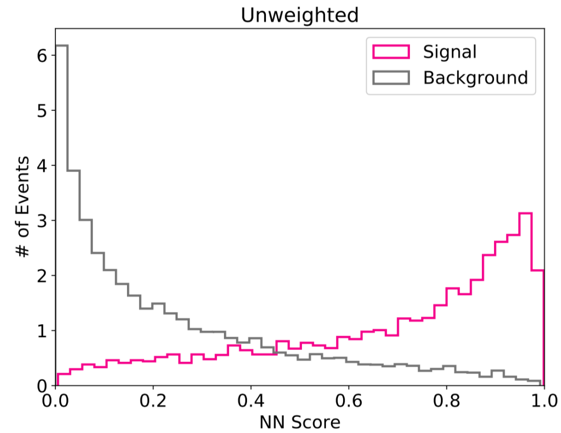
Following training of each network, the optimal cut on the NN score to separate signal from background was determined by choosing the cut that would maintain signal efficiency above 80% while simultaneously maximizing the MC-weighted significance ($S/\sqrt{B}$) calculated in a wide window of 50 GeV - 150 GeV in MMC or $M_{2\mathrm{T}}$ for ZH and WH events, respectively. The performance of each network is described in detail in the following sections and summarized in Table 16, Figure 61, and Figure 62.

7.2.4 WH LepHad Performance

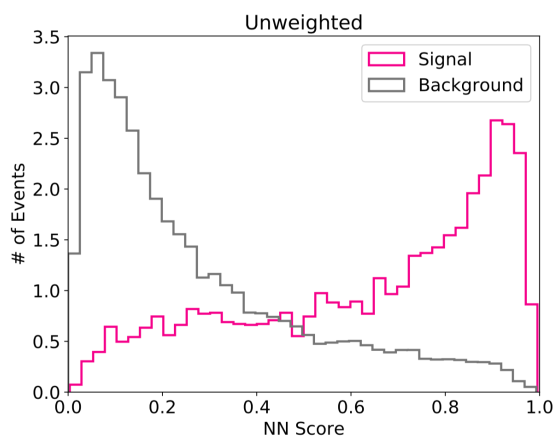
The $W \to l\nu$, $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ channel had the highest statistics out of the four analysis regions. The training dataset consisted of 28,375 signal events and 200,917 WZ background events. To ameliorate the discrepancy in sizes between signal and background training sets, class weights were used to weight the signal 7.1 times more than the background during training. The optimized neural network architecture had an AUC (Area Under Curve) score of 0.915. The optimal neural network score cut was placed at `nn_score` > 0.41 to increase the significance ($S/\sqrt{B}$) at preselection from 0.29 to 0.45, a relative increase of 56%. This cut yielded a signal efficiency of 83.6% with a background rejection of 79%, i.e. a rejection factor of 4.76. Crucially, the mean of the signal $M_{2\mathrm{T}}$ distribution only shifted by 0.6% following this score cut, meaning the neural network is not severely biasing our final fit quantity. The diboson background, on the other hand, shifted by -25%.



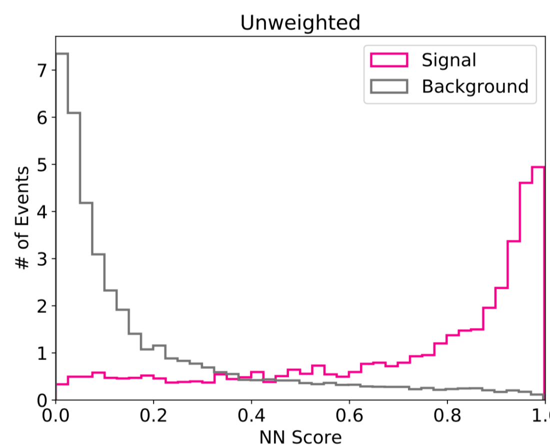
(a) $W \to l\nu_l$ & $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$



(b) $W \to l\nu_l$ & $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$



(c) $Z \to ll$ & $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$

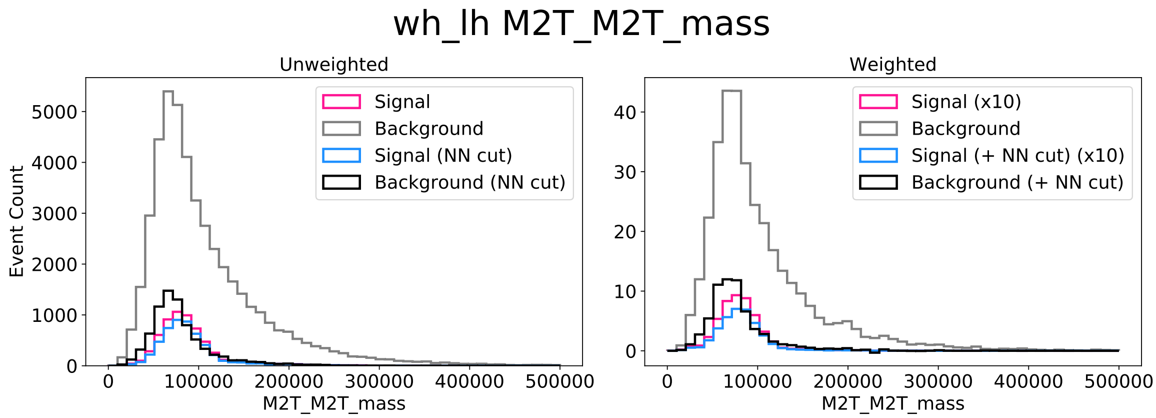


(d) $Z \to ll$ & $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$

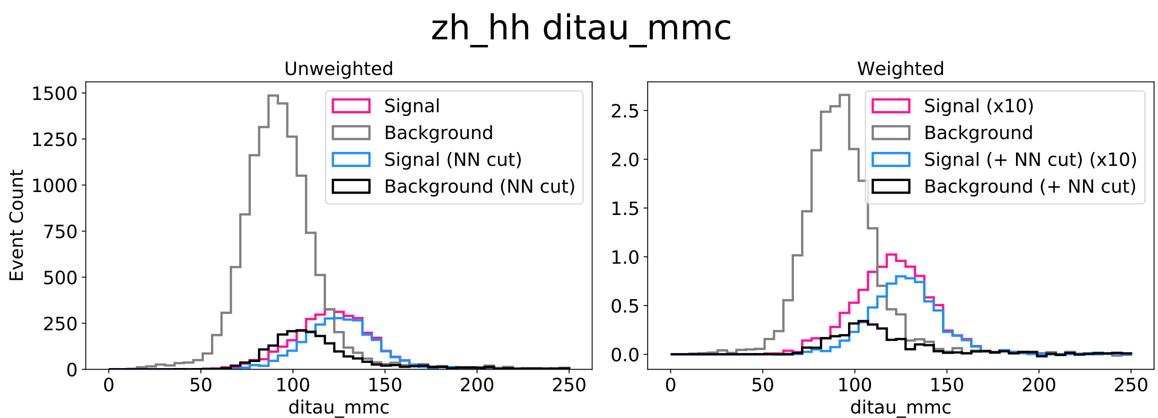
Fig. 61: Unweighted & normalized plots of test signal and background MC sorted by the neural network output score.

7.2.5 WH HadHad Performance

The $W \to l\nu$, $H \to \tau_{\mathrm{had}}\tau_{\mathrm{had}}$ channel had a training dataset consisting of 19,873 signal events and 41,955 background events. Class weights were used to weight the signal 2.1 times more than the background during training. The optimized neural network architecture had an AUC (Area Under Curve) score of 0.877. The optimal neural network score cut was placed at `nn_score` > 0.42 to increase the significance ($S/\sqrt{B}$) at preselection from 0.38 to 0.53, a relative increase of 40%. This cut yielded a signal efficiency of 80.3% with a background rejection of 79.4%, i.e. a rejection factor of 4.85. The mean of the signal $M_{2\mathrm{T}}$ distribution shifted by 3.0% following this score cut, introducing a minimal bias into our



(a) $W \to l\nu_l$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$ $M_{2\text{T}}$ [MeV]



(b) $Z \to ll$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$ MMC [GeV]

Fig. 62: Example unweighted and MC-weighted distributions of $M_{2\text{T}}$ (above, in the WH LepHad channel) and MMC (below, in the ZH HadHad channel) for signal and background before (in pink and grey) and after (in blue and black) the cut on the neural network output score show a large decrease in background events with relatively few signal events lost by comparison. The peaks of the signal mass distributions move very little following the neural network score cut and are also still fairly distinct from the peaks of the diboson backgrounds. The signal has been given an additional weight factor of 10X in the weighted histograms for the purposes of visibility during plotting.

final fit quantity, while the background distribution shifted by -55.4%.

7.2.6 ZH LepHad Performance

The $Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ channel had a training dataset consisting of 20,236 signal events and 71,640 background events. Class weights were used to weight the signal 3.5 times more than the background during training. The optimized neural network architecture had an AUC (Area Under Curve) score of 0.837. The optimal neural network score cut was

Region	ROC AUC	`nn_score` cut	$\Delta\frac{S}{\sqrt{B}}$	Sig. Efficiency	Bkg. Rejection
WH LepHad	0.915	0.41	+56%	83.6%	4.76
WH HadHad	0.877	0.42	+40%	80.3%	4.85
ZH LepHad	0.837	0.26	+24%	86.3%	2.63
ZH HadHad	0.899	0.41	+91.6%	80%	6.62

Table 16: A summary of the preliminary performance of the four diboson neural networks.

placed at `nn_score` > 0.26 to increase the significance ($S/\sqrt{B}$) at preselection from 0.23 to 0.29, a relative increase of 24%. This cut yielded a signal efficiency of 86.3% with a background rejection of 62.0%, i.e. a rejection factor of 2.63. The mean of the signal MMC distribution shifted by 2.1% following this score cut, introducing a minimal bias into our final fit quantity, while the background distribution shifted by 10%.

7.2.7 ZH HadHad Performance

The $Z \to ll$, $H \to \tau_{\text{had}}\tau_{\text{had}}$ channel had a training dataset consisting of 15,215 signal events and 51,508 background events. Class weights were used to weight the signal 3.4 times more than the background during training. The optimized neural network architecture had an AUC (Area Under Curve) score of 0.899. The optimal neural network score cut was placed at `nn_score` > 0.41 to increase the significance ($S/\sqrt{B}$) at preselection from 0.20 to 0.38, a relative increase of 91.6%. This cut yielded a signal efficiency of 80% with a background rejection of 84.9%, i.e. a rejection factor of 6.62. The mean of the signal MMC distribution shifted by 3.2% following this score cut, introducing a minimal bias into our final fit quantity, while the background shifted by 21.3%.

8 From Run 1 to Run 2

8.1 ATLAS Results from Run 1

The Run 1 search for this process, detailed in [93], considered 20.3 fb^{-1} of LHC data at center-of-mass energy $s = \sqrt{8}$ TeV. The observed signal strength of this process, defined as $\mu = \sigma/\sigma_{SM}$, was found to be 2.3 ± 1.6. This signal strength is consistent with Standard Model expectations for a 125 GeV Higgs boson, but is not yet significant enough to claim a discovery of this Higgs boson process. We are, however, able to set an upper limit on the signal strength of this process at $\mu < 5.6$ with a 95% confidence level.

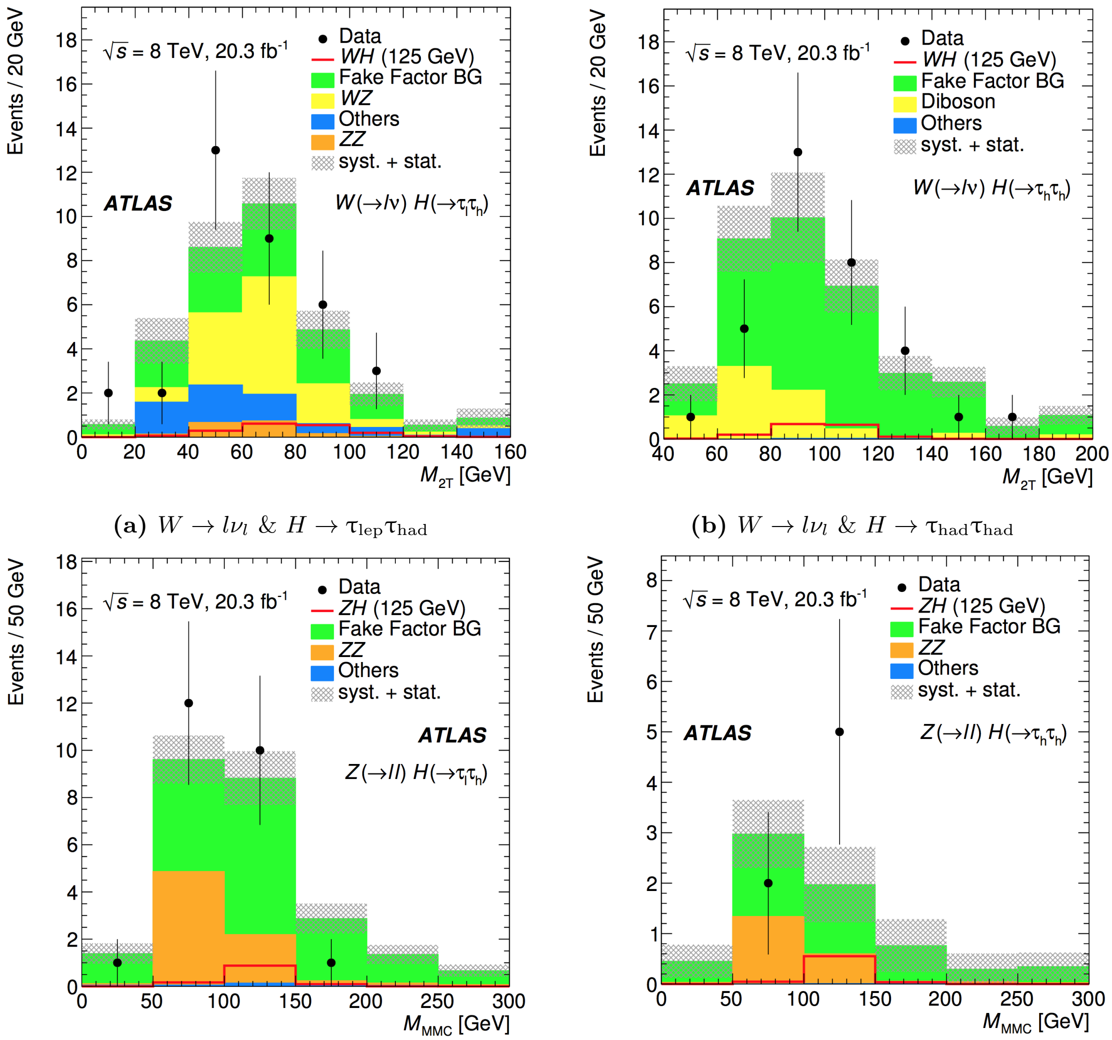
Distributions of expected real and fake backgrounds as well as actual observed data as a function of each reconstructed Higgs boson mass variable from Run 1 are shown in Figure 63. Expected and observed event counts from the Run 1 analysis are summarized in Table 17. Expected and observed significances for each of the four signal regions in Run 1 are shown in Table 18.

Signal Region	Obs.	Signal	Σ Backgrounds	Fake Factor	Diboson
$W \to l\nu$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$	35	1.95 ± 0.05	32.4 ± 1.9	13.1 ± 1.3	13.54 ± 0.35
$W \to l\nu$, $H \to \tau_{\text{had}}\tau_{\text{had}}$	33	1.84 ± 0.04	35.5 ± 2.7	28.1 ± 2.4	7.4 ± 1.2
$Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$	24	1.14 ± 0.03	24.6 ± 1.5	17.1 ± 1.5	7.28 ± 0.16
$Z \to ll$, $H \to \tau_{\text{had}}\tau_{\text{had}}$	7	0.64 ± 0.02	6.8 ± 1.2	4.7 ± 1.2	2.09 ± 0.09

Table 17: Expected and observed event counts from the Run 1 analysis. Background events not listed explicitly are mostly $t\bar{t}$ events and contribute primarily to the sum of the backgrounds in the $W \to l\nu$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$ signal region. Only statistical uncertainties are given.

Signal Region	Expected significance	Observed significance
$W \to l\nu$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$	0.36σ	0.44σ
$W \to l\nu$, $H \to \tau_{\text{had}}\tau_{\text{had}}$	0.32σ	0.60σ
$Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$	0.28σ	0.29σ
$Z \to ll$, $H \to \tau_{\text{had}}\tau_{\text{had}}$	0.32σ	1.38σ

Table 18: Expected and observed significances for each of the four signal regions in Run 1.



(a) $W \to l\nu_l$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$

(b) $W \to l\nu_l$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$

(c) $Z \to ll$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$

(d) $Z \to ll$ & $H \to \tau_{\text{had}}\tau_{\text{had}}$

Fig. 63: Expected signal, expected backgrounds, and observed data from Run 1 in each of the four main analysis categories as a function of the reconstructed Higgs boson mass variables. "Others" refers primarily to $t\bar{t}$ events.

8.2 Changes in Run 2

This analysis strategy is heavily drawn from the previous Run 1 analysis, with a few notable exceptions. In Run 1, the analysis category $W \to l\nu_l$ & $H \to \tau_{\text{lep}}\tau_{\text{had}}$ required exactly one electron and one muon in the final state. Our analysis newly includes final states with same-flavor leptons from both the W and Higgs bosons. We also newly include muon fake factors in addition to tau and electron fake factors in Run 2. We have also introduced an update to our irreducible background estimation technique in using my MC-trained neural networks for separating true signal from diboson backgrounds instead of the

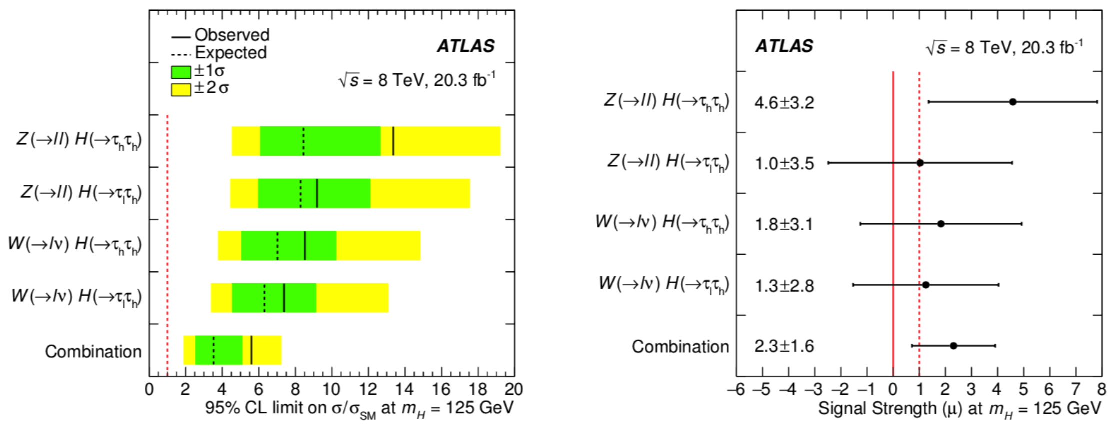


Fig. 64: The combined Run 1 VH, $H \to \tau\tau$ measurements. On the left, the 95% confidence-level (CL) upper limits for σ/σ_{SM} for each of the four signal regions are shown, along with the combined upper limit for all four channels. Expected upper limits are shown in dotted lines, while observed upper limits are shown in solid lines. The lime green regions mark the $\pm 1\sigma$ significance interval, while the yellow regions mark the $\pm 2\sigma$ significance interval. On the right, the measured signal strength $\mu = \sigma/\sigma_{SM}$ for $m_H = 125$ GeV is shown for each signal region individually and combined. The dashed vertical red line indicates a perfectly Standard Model-like result. Each of the four categories shows a measurement of the signal strength μ that is consistent, within error bars, with the Standard Model prediction. [93]

Run 1 subtraction MC subtraction method.

In addition to these analysis strategy updates, the Run 2 dataset brings many changes from the Run 1 analysis environment, including an increase of center-of-mass energy from $\sqrt{s} = 8$ TeV to $\sqrt{s} = 13$ TeV as well approximately a seven-fold increase in integrated luminosity, around double the average number of pileup events, and consequently higher trigger thresholds for each of our objects.

8.3 Expected Event Counts

Calculating our expected event counts for the Run 2 analysis involves not only a scaling based on the amount of total integrated luminosity considered, but also a scaling based on the increase in signal cross-section associated with a higher center-of-mass energy in Run 2 versus Run 1. Scaling up the total integrated luminosity from the Run 1 (20.3 fb^{-1}) to Run 2 (139 fb^{-1}) analysis yields a factor of approximately 6.85. Scaling up the increase in signal cross-section based on the increased center-of-mass energy yields an additional factor of approximately 1.95 for the WH channels and 2.10 for the ZH channels. In total, we might expect the WH channels to increase in counts by about a factor of **13.36** and the ZH channels to increase in counts by about a factor of **14.39**. However, scaling our Run 1 observed counts up by these factors neglects many other subtleties in our analysis, from increased lepton trigger p_T thresholds to additional final states in the analysis to a potentially improved final significance in each signal region due to the diboson neural network.

Taking a different approach, and instead calculating the theoretical number of WH and ZH events we expect to see based on cross-sections and branching ratios, we can get an upper bound on the number of expected events in our analysis regions:

$$N_{\text{VH events}} = \mathcal{L}_{\text{int}} \cdot \sigma_{\text{VH}} \cdot \text{BR}(V \to \text{leptons}) \cdot \text{BR}(H \to \tau_{\text{lep}}\tau_{\text{had}} + H \to \tau_{\text{had}}\tau_{\text{had}}) \tag{109}$$

These values are approximately:

- $\mathcal{L}_{\text{int}} = 139\ \text{fb}^{-1}$
- Inclusive σ_{WH} at $\sqrt{s} = 13$ TeV $= 1.380$ pb
- Inclusive σ_{ZH} at $\sqrt{s} = 13$ TeV $= 0.8696$ pb
- $\text{BR}(W \to l\nu) = 0.324$
- $\text{BR}(Z \to ll) = 0.101$
- $\text{BR}(H \to \tau_{\text{lep}}\tau_{\text{had}}) = \text{BR}(H \to \tau\tau) \cdot \text{BR}(\tau_{\text{lep}}\tau_{\text{had}}) = 0.06272 \cdot 0.455 = 0.0285$
- $\text{BR}(H \to \tau_{\text{had}}\tau_{\text{had}}) = \text{BR}(H \to \tau\tau) \cdot \text{BR}(\tau_{\text{had}}\tau_{\text{had}}) = 0.06272 \cdot 0.4225 = 0.0265$

Combining these values, we get the following approximate upper bounds on our signal categories:

- $W \to l\nu$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$: 1,774 events
- $W \to l\nu$, $H \to \tau_{\text{had}}\tau_{\text{had}}$: 1,647 events
- $Z \to ll$, $H \to \tau_{\text{lep}}\tau_{\text{had}}$: 348 events
- $Z \to ll$, $H \to \tau_{\text{had}}\tau_{\text{had}}$: 324 events

Of course, these counts do not incorporate acceptance information from the perspective of inefficiencies in our detector coverage, our TDAQ system, our particle identification schemes, or our analysis cuts. In Run 1, the signal acceptance for the WH channels was 1.9%, while the signal acceptance for the ZH channels was 5.3%.

8.4 Current Analysis Status

The Run 2 VH, $H \to \tau\tau$ analysis is in excellent shape and is nearing its completion. The analysis has hit major milestones such as producing full MC and data samples with the proper object selections and overlap removal applied corresponding to the Run 2 dataset, creating a fully Python-based standalone analysis framework, measuring fake rates for electrons, muons, and taus, implementing the fake factor method and producing preliminary closure test plots in each of our analysis preselection categories and control regions, calculating the tau fake factor systematic uncertainty using the Run 1 methodology, and developing and evaluating the NN-based irreducible diboson background separation method. We have implemented custom analysis methods for re-calculating the MMC for WH events following our object selections and are developing a similar method for M_{2T} as well.

Remaining milestones on the horizon for this analysis include the optimization of our analysis selection cuts alongside the integration of the diboson NN score cut, the possible re-parameterization of the fake rate measurements to optimize fake factor modeling, a finalized calculation of the tau fake rate systematic uncertainty with potentially a new methodology, and a variation of our full analysis across several systematic uncertainty parameters.

We could not be more excited or honored to be able to soon share our upcoming results on this still-undetected rare process that will contribute to humanity's fundamental understanding of the Higgs boson.

Appendix: Individual Contributions

- RNN Tau ID in the Tau Trigger
  - I produced training samples with trigger-level information, adapted the offline tau CP's NN architecture designed by Christopher Deutsch [41] for use with trigger-level variables, and iterated to find the optimal performance of the RNN for HLT-level taus under the guidance of my supervisor, Bertrand Martin dit Latour. Ultimately, I trained three separate neural networks for 0-prong, 1-prong, and multi-prong taus that were implemented in the ATLAS HLT in summer 2018.
- Implementation of $M_{2\mathrm{T}}$
  - I created a custom C++/ROOT-based module to calculate $M_{2\mathrm{T}}$ and contributed it to the `xTauFramework` repository.
- Sample processing & analysis framework
  - I was a major contributor to the production of ntuple-level samples for our analysis via the `xTauFramework`.
  - I helped pioneer the shared Python/uproot/Pandas-based framework `vhtautau` for our analysis group and re-optimized our analysis several times for significant improvements in speed and computing requirements.
- Tau fake rate & systematic uncertainty
  - I calculated the first round of tau fake rates for our analysis and measured the tau fake rate systematic uncertainty using the Run 1 method based on missing E_T cuts.
  - I also used the experimental version of a new fake tau tool to calculate fake rates using a template fit parameterized by p_T and jet width.
- Diboson NN

- I designed (with the exception of the two ϕ-translation layers in the neural network suggested by my colleague Chase Shimmin), trained, optimized, and evaluated the performance of the four separate NN models for separating signal and diboson MC.

- HLeptons Trigger Contact
  - I served as a liaision between the trigger groups and the HLeptons group, which primarily focuses on the $H \to \tau\tau$ analyses, from 2017-2021.
  - I performed several validation studies of proposed new triggers to measure the trigger acceptance and quantify subsequent efficiency gains for the $H \to \tau\tau$ analysis regions under various kinematic selections.
  - My studies provided sufficient motivation to include several new triggers in the Run 3 trigger menu to benefit the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ analysis.

- Contributions to upgrade physics
  - I performed studies to understand the acceptance gains based on maximum $|\eta|$ cuts on leading and sub-leading jets in the $H \to \tau_{\text{had}}\tau_{\text{had}}$ analysis for VBF triggers.
  - I co-wrote and edited *Expected Performance of the ATLAS Detector at the High-Luminosity LHC* for the CERN Yellow Report [49].

- Choreo VAE
  - I gathered and led an independent research team to collaboratively develop a VAE model trained on motion capture data of my own movements. I contributed the studies of the alternative dimensionality-reduction strategies such as t-SNE, PCA, and UMAP.

- Choreo GNN
  - I designed and implemented an original GNN based on the equations described in the NRI paper [83] using graph-based libraries in Pytorch Geometric. I trained and evaluated the GNN on my own movements.

- Service to the collaboration
  - I have participated in three years of advocacy trips to Capitol Hill to meet with congressional offices to discuss the importance of high-energy physics research and STEM funding in general.
  - I have served several times as a reviewer for the NeurIPS Machine Learning for the Physical Sciences Workshop as well as for the Women in Machine Learning (WiML) Workshop.
- Science outreach
  - I have given a number of public talks about particle physics, including: winning the Windy City Physics Slam at ICHEP 2016, a televised interview with PBS Chicago, a speaker for Yale's *Science in the News* delivering scientific talks to the broader New Haven community, and on social media as a featured speaker for Randi Zuckerberg's STEM outreach initiative.
  - I have also written a long-form opinion piece making a case for a future circular collider beyond the LHC for Yale's *Distilled* magazine and an essay on how physicists view the nature of reality for *Sightline Arts*.

References

[1] Sylvia Berryman. "Democritus". In: *Stanford Encyclopedia of Philosophy* (Aug. 2004). URL: `https://plato.stanford.edu/entries/democritus`.

[2] P.A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020), p. 083C01. DOI: `10.1093/ptep/ptaa104`.

[3] Kerry McKenzie. "On the Fundamentality of Symmetries". In: *Philosophy of Science* 81.5 (2014), pp. 1090–1102. DOI: `10.1086/677697`. eprint: `https://doi.org/10.1086/677697`. URL: `https://doi.org/10.1086/677697`.

[4] Eugene P. Wigner. “The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959”. In: *Communications on Pure and Applied Mathematics* 13.1 (1960), pp. 1–14. DOI: https://doi.org/10.1002/cpa.3160130102. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160130102. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160130102.

[5] M. Tanabashi et al. “Status of Higgs Boson Physics”. In: *Phys. Rev. D* 98.030001 (2018). URL: http://pdg.lbl.gov/2018/reviews/rpp2018-rev-higgs-boson.pdf.

[6] ATLAS Collaboration. “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment”. In: *Eur. Phys. J.* C76.1 (2016), p. 6. DOI: 10.1140/epjc/s10052-015-3769-y. arXiv: 1507.04548 [hep-ex].

[7] Rolf Heuer. *A celebration of science for peace*. URL: https://home.cern/news/opinion/cern/celebration-science-peace.

[8] CERN. *The Synchrocyclotron*. URL: https://home.cern/science/accelerators/synchrocyclotron.

[9] CERN. *The Birth of the Web*. URL: https://home.cern/science/computing/birth-web.

[10] *CERN Images Gallery*. URL: https://home.cern/resources/image/cern/views-cern-images-gallery.

[11] Mike Perricone. *Hitting the Broad Size of a (Classified) Barn*. 2006. URL: https://www.symmetrymagazine.org/article/february-2006/hitting-the-broad-side-of-a-classified-barn.

[12] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *The European Physical Journal C* 75.5 (May 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3397-6. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3397-6.

[13] Anaïs Schaeffer. *New schedule for CERN's accelerators and experiments*. 2020. URL: https://home.cern/news/news/accelerators/new-schedule-cerns-accelerators-and-experiments.

[14] *Run 2 Luminosity Public Results*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[15] ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1999. URL: https://cds.cern.ch/record/391176.

[16] M. Barnett M. Cirilli. *First collisions in ATLAS*. URL: http://atlasexperiment.org/news/2009/first-collisions-in-ATLAS.html.

[17] Joao Pequenao and Paul Schaffner. *A computer generated image representing how ATLAS detects particles*. Jan. 2013. URL: https://cds.cern.ch/record/1505342.

[18] F. Hugging. "The ATLAS Pixel Detector". In: *IEEE Transactions on Nuclear Science* 53 (June 2006), pp. 1732–1736. DOI: 10.1109/TNS.2006.871506. eprint: physics/0412138.

[19] M Capeans et al. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: https://cds.cern.ch/record/1291633.

[20] The ATLAS Experiment. *The Inner Detector*. URL: https://atlas.cern/discover/detector/inner-detector.

[21] Michaela Mlynarikova. "Performance of the ATLAS Hadronic Tile calorimeter". In: *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. arXiv: 1709.00100 [physics.ins-det]. URL: https://inspirehep.net/record/1621271/files/arXiv:1709.00100.pdf.

[22] ATLAS Collaboration. *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*. Tech. rep. CERN-LHCC-2017-017. ATLAS-TDR-026. Geneva: CERN, Sept. 2017. URL: http://cds.cern.ch/record/2285580.

[23] Giordano Cattani and the RPC group. "The Resistive Plate Chambers of the ATLAS experiment: performance studies". In: *Journal of Physics: Conference Series* 280.1 (2011), p. 012001. URL: http://stacks.iop.org/1742-6596/280/i=1/a=012001.

[24] The ATLAS collaboration. "Operation of the ATLAS trigger system in Run 2". In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10004–P10004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/10/p10004. URL: http://dx.doi.org/10.1088/1748-0221/15/10/P10004.

[25] G. Aad et al. "The ATLAS Simulation Infrastructure". In: *The European Physical Journal C* 70.3 (Sept. 2010), pp. 823–874. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1429-9. URL: http://dx.doi.org/10.1140/epjc/s10052-010-1429-9.

[26] D Buskulic et al. "Measurement of α_s from scaling violations in fragmentation functions in e^+e^- annihilation". In: (Jan. 1995).

[27] G. Aad et al. "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data". In: *Journal of Instrumentation* 14.12 (Dec. 2019), P12006–P12006. ISSN: 1748-0221. DOI: 10.1088/1748-0221/14/12/p12006. URL: http://dx.doi.org/10.1088/1748-0221/14/12/P12006.

[28] ATLAS Collaboration. *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*. 2020. arXiv: 2012.00578 [hep-ex].

[29] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[30] Steven Schramm. "ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects". In: *EPJ Web Conf.* 182 (2018). Ed. by Y. Aharonov, L. Bravina, and S. Kabana, p. 02113. DOI: 10.1051/epjconf/201818202113.

[31] M. Aaboud et al. “Performance of top-quark and W W -boson tagging with ATLAS in Run 2 of the LHC”. In: *The European Physical Journal C* 79.5 (Apr. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6847-8. URL: http://dx.doi.org/10.1140/epjc/s10052-019-6847-8.

[32] *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. ATL-PHYS-PUB-2017-013. Geneva: CERN, July 2017. URL: https://cds.cern.ch/record/2273281.

[33] M. Aaboud et al. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 78.11 (Nov. 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6288-9. URL: http://dx.doi.org/10.1140/epjc/s10052-018-6288-9.

[34] A. L. Samuel. “Some Studies in Machine Learning Using the Game of Checkers”. In: *IBM Journal of Research and Development* 3.3 (1959), pp. 210–229. DOI: 10.1147/rd.33.0210.

[35] Kurt Hornik. “Approximation capabilities of multilayer feedforward networks”. In: *Neural Networks* 4.2 (1991), pp. 251–257. ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(91)90009-T. URL: https://www.sciencedirect.com/science/article/pii/089360809190009T.

[36] Thomas Kipf. “Deep Learning with Graph-Structured Representations”. PhD thesis. University of Amsterdam, 2020. URL: https://hdl.handle.net/11245.1/1b63b965-24c4-4bcd-aabb-b849056fa76d.

[37] *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. ATLAS-CONF-2017-029. Geneva: CERN, May 2017. URL: http://cds.cern.ch/record/2261772.

[38] T Barillari et al. *Local Hadronic Calibration*. Tech. rep. ATL-LARG-PUB-2009-001-2. ATL-COM-LARG-2008-006. ATL-LARG-PUB-2009-001. Due to a report-number conflict with another document, the report-number ATL-LARG-PUB-2009-001-2 has

been assigned. Geneva: CERN, June 2008. URL: https://cds.cern.ch/record/1112035.

[39] *The ATLAS Tau Trigger in Run 2*. Tech. rep. ATLAS-CONF-2017-061. Geneva: CERN, July 2017. URL: http://cds.cern.ch/record/2274201.

[40] K. Benslama et al. *Tau trigger at the ATLAS experiment*. 2008. arXiv: 0810.0465 [hep-ex].

[41] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*. Tech. rep. ATL-PHYS-PUB-2019-033. Geneva: CERN, Aug. 2019. URL: http://cds.cern.ch/record/2688062.

[42] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[43] Rene Brun et al. *root-project/root: v6.18/02*. Version v6-18-02. Aug. 2019. DOI: 10.5281/zenodo.3895860. URL: https://doi.org/10.5281/zenodo.3895860.

[44] Noel Dawe et al. *scikit-hep/root_numpy: 4.8.0*. Version 4.8.0. Dec. 2018. DOI: 10.5281/zenodo.2538152. URL: https://doi.org/10.5281/zenodo.2538152.

[45] G. Aad et al. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5004-5. URL: http://dx.doi.org/10.1140/epjc/s10052-017-5004-5.

[46] François Chollet et al. *Keras*. https://keras.io. 2015.

[47] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[48] Daniel Hay Guest et al. *lwtnn/lwtnn: Version 2.11.1*. Version v2.11.1. Dec. 2020. DOI: 10.5281/zenodo.4310003. URL: https://doi.org/10.5281/zenodo.4310003.

[49] *Expected performance of the ATLAS detector at the High-Luminosity LHC*. Tech. rep. ATL-PHYS-PUB-2019-005. Geneva: CERN, Jan. 2019. URL: https://cds.cern.ch/record/2655304.

[50] Aneesh Heintz et al. *Accelerated Charged Particle Tracking with Graph Neural Networks on FPGAs*. 2020. arXiv: 2012.01563 [physics.ins-det].

[51] Alexander Bogatskiy et al. *Lorentz Group Equivariant Neural Network for Particle Physics*. 2020. arXiv: 2006.04780 [hep-ph].

[52] S. Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019. ISBN: 9781610395700. URL: https://books.google.com/books?id=lRqrDQAAQBAJ.

[53] Joy Buolamwini and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Ed. by Sorelle A. Friedler and Christo Wilson. Vol. 81. Proceedings of Machine Learning Research. New York, NY, USA: PMLR, 23–24 Feb 2018, pp. 77–91. URL: http://proceedings.mlr.press/v81/buolamwini18a.html.

[54] Tolga Bolukbasi et al. *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. 2016. arXiv: 1607.06520 [cs.CL].

[55] C. Garvie et al. *The Perpetual Line-up: Unregulated Police Face Recognition in America*. Georgetown Law, Center on Privacy & Technology, 2016. URL: https://books.google.com/books?id=uAYjngAACAAJ.

[56] Mark Roman Miller et al. "Personal identifiability of user tracking data during observation of 360-degree VR video". In: *Scientific Reports* 10.1 (Oct. 2020), p. 17404. ISSN: 2045-2322. DOI: 10.1038/s41598-020-74486-y. URL: https://doi.org/10.1038/s41598-020-74486-y.

[57] Mariel Pettee et al. "Beyond Imitation: Generative and Variational Choreography via Machine Learning". In: *arXiv preprint arXiv:1907.05297* (2019).

[58] Luke Jennings. *Obituary: Pina Bausch*. https://www.theguardian.com/stage/2009/jul/01/pina-bausch-obituary-dance. 2009. (Visited on 06/30/2009).

[59] Alex Graves. *Generating Sequences With Recurrent Neural Networks*. https://arxiv.org/abs/1308.0850. Aug. 2013.

[60] John McCormick et al. "Emergent Behaviour: Learning from an Artificially Intelligent Performing Software Agent". In: *21st International Symposium on Electronic Art*. Vancouver, Canada, Apr. 2015. URL: https://eprints.qut.edu.au/102401/.

[61] Luka Crnkovic-Friis and Louise Crnkovic-Friis. *Generative Choreography using Deep Learning*. https://arxiv.org/abs/1605.06921. May 2016.

[62] Omid Alemi, Jules Françoise, and Philippe Pasquier. *GrooveNet: Real-Time Music-Driven Dance Movement Generation using Artificial Neural Networks*. https://www.researchgate.net/publication/318470738_GrooveNet_Real-Time_Music-Driven_Dance_Movement_Generation_using_Artificial_Neural_Networks. Aug. 2017.

[63] Zimo Li et al. "Auto-Conditioned LSTM Network for Extended Complex Human Motion Synthesis". In: *CoRR* abs/1707.05363 (2017). arXiv: 1707.05363.

[64] Alexander Berman Valencia James. "Learning as Performance: Autoencoding and Generating Dance Movements in Real Time". In: *Computational Intelligence in Music, Sound, Art and Design* (2018), pp. 256–266. URL: https://www.researchgate.net/publication/323592777_Learning_as_Performance_Autoencoding_and_Generating_Dance_Movements_in_Real_Time.

[65] Dimitrije Marković and Nebojša Malešević. "Adaptive Interface for Mapping Body Movements to Sounds". In: *Computational Intelligence in Music, Sound, Art and Design*. Ed. by Antonios Liapis, Juan Jesús Romero Cardalda, and Anikó Ekárt. Cham: Springer International Publishing, 2018, pp. 194–205. ISBN: 978-3-319-77583-8.

[66] Alexander Berman and Valencia James. "Learning as Performance: Autoencoding and Generating Dance Movements in Real Time". In: *EvoMUSART*. 2018.

[67] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. "Nonlinear Component Analysis as a Kernel Eigenvalue Problem". In: *Neural Computation* 10.5 (1998), pp. 1299–1319. DOI: 10.1162/089976698300017467. URL: https://doi.org/10.1162/089976698300017467.

[68] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[69] Edward C. Warburton. "Of Meanings and Movements: Re-Languaging Embodiment in Dance Phenomenology and Cognition". In: *Dance Research Journal* 43.2 (2011), pp. 65–83. ISSN: 01497677, 1940509X. URL: http://www.jstor.org/stable/23266966.

[70] Carrie Noland. *Agency and Embodiment: Performing Gestures/Producing Culture*. Harvard University Press, 2009. Chap. Inscription as Performance: Henri Michaux and the Writing Body.

[71] Brenda Farnell. "Movement Writing Systems". In: *The World's Writing Systems* (1996).

[72] Paul Kaiser Bill T. Jones and Shelley Eshkar. *Ghostcatching*. http://openendedgroup.com/artworks/gc.html. 1999.

[73] Lisa Naugle. *Reflections on Dancing with William Forsythe*. http://dance.arts.uci.edu/lnaugle/files/mocap/forsythe/index.html. 2000.

[74] Paul Kaiser Marc Downie and Shelley Eshkar. *Loops*. http://openendedgroup.com/artworks/loops.html. 2001.

[75] Daphne LePrince-Ringuet. *Google's latest experiment teaches AI to dance like a human*. https://www.wired.co.uk/article/google-ai-wayne-mcgregor-dance-choreography. 2018.

[76] Christopher M. Bishop. "Mixture Density Networks". In: *Neural Computing Research Group Report* (Feb. 1994). URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/bishop-ncrg-94-004.pdf.

[77] Karl Pearson. "On lines and planes of closest fit to systems of points in space". In: *Philosophical Magazine* (1901), pp. 559–572.

[78] L. McInnes, J. Healy, and J. Melville. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". In: *arXiv e-prints* (Feb. 2018). arXiv: 1802.03426 [stat.ML].

[79] L.J.P. van der Maaten and G.E. Hinton. "Visualizing High-Dimensional Data Using t-SNE". In: *Journal of Machine Learning Research* (2008), pp. 2579–2605. URL: https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf.

[80] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, and Ole Winther. "Autoencoding beyond pixels using a learned similarity metric". In: *CoRR* abs/1512.09300 (2015). arXiv: 1512.09300. URL: http://arxiv.org/abs/1512.09300.

[81] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. eprint: arXiv:1412.6980.

[82] Mariel Pettee et al. "Choreo-Graph: Learning Latent Graph Representations of the Dancing Body". In: *NeurIPS*. Workshop on Machine Learning for Creativity and Design. 2020.

[83] Thomas Kipf et al. "Neural Relational Inference for Interacting Systems". In: *arXiv preprint arXiv:1802.04687* (2018).

[84] Matthias Fey and Jan E. Lenssen. "Fast Graph Representation Learning with PyTorch Geometric". In: *ICLR Workshop on Representation Learning on Graphs and Manifolds*. 2019.

[85] Bill T. Jones. *Body Movement Language: AI Sketches with Bill T. Jones*. https://billtjonesai.com/. 2019.

[86] Francisco Sagasti. "Information Technology and the Arts: The Evolution of Computer Choreography during the Last Half Century". In: *Dance Chronicle* 42.1 (2019), pp. 1–52. DOI: 10.1080/01472526.2019.1575661. eprint: https://doi.org/10.1080/01472526.2019.1575661. URL: https://doi.org/10.1080/01472526.2019.1575661.

[87] Deborah Friedes Galili. “Gaga: Moving beyond Technique with Ohad Naharin in the Twenty-First Century”. In: *Dance Chronicle* 38.3 (2015), pp. 360–392. DOI: 10.1080/01472526.2015.1085759. eprint: https://doi.org/10.1080/01472526.2015.1085759. URL: https://doi.org/10.1080/01472526.2015.1085759.

[88] Kosei Dohi. *Variational Autoencoders for Jet Simulation*. 2020. arXiv: 2009.04842 [hep-ph].

[89] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. “Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis”. In: *Computing and Software for Big Science* 1.1 (Sept. 2017). ISSN: 2510-2044. DOI: 10.1007/s41781-017-0004-6. URL: http://dx.doi.org/10.1007/s41781-017-0004-6.

[90] Xiangyang Ju et al. *Graph Neural Networks for Particle Reconstruction in High Energy Physics detectors*. 2020. arXiv: 2003.11603 [physics.ins-det].

[91] Miles Cranmer et al. *Discovering Symbolic Models from Deep Learning with Inductive Biases*. 2020. arXiv: 2006.11287 [cs.LG].

[92] G. Aad et al. “Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment”. In: *Phys. Rev. D* 101 (1 Jan. 2020), p. 012002. DOI: 10.1103/PhysRevD.101.012002. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.012002.

[93] ATLAS Collaboration. “Search for the Standard Model Higgs boson produced in association with a vector boson and decaying into a tau pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *Phys. Rev.* D93.9 (2016), p. 092005. DOI: 10.1103/PhysRevD.93.092005. arXiv: 1511.08352 [hep-ex].

[94] *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*. Tech. rep. ATLAS-CONF-2018-036. Geneva: CERN, July 2018. URL: https://cds.cern.ch/record/2630338.

[95] *Cross-section measurements of the Higgs boson decaying to a pair of tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2018-021. Geneva: CERN, June 2018. URL: `https://cds.cern.ch/record/2621794`.

[96] Jim Pivarski et al. *scikit-hep/uproot: 3.12.0*. Version 3.12.0. July 2020. DOI: `10.5281/zenodo.3952728`. URL: `https://doi.org/10.5281/zenodo.3952728`.

[97] G. Aad et al. "Performance of electron and photon triggers in ATLAS during LHC Run 2". In: *The European Physical Journal C* 80.1 (Jan. 2020). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-019-7500-2`. URL: `http://dx.doi.org/10.1140/epjc/s10052-019-7500-2`.

[98] G. Aad et al. "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 79.11 (Nov. 2019). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-019-7450-8`. URL: `http://dx.doi.org/10.1140/epjc/s10052-019-7450-8`.

[99] A. Elagin et al. "A New Mass Reconstruction Technique for Resonances Decaying to di-tau". In: *Nucl. Instrum. Meth.* A654 (2011), pp. 481–489. DOI: `10.1016/j.nima.2011.07.009`. arXiv: `1012.4686 [hep-ex]`.

[100] A. J. Barr et al. "Guide to transverse projections and mass-constraining variables". In: *Phys. Rev.* D84 (2011), p. 095031. DOI: `10.1103/PhysRevD.84.095031`. arXiv: `1105.2977 [hep-ph]`.

[101] Thomas P. S. Gillam. "Identifying fake leptons in ATLAS while hunting SUSY in 8 TeV proton-proton collisions". PhD thesis. Cambridge U., 2015. DOI: `10.17863/CAM.16619`. URL: `https://www.repository.cam.ac.uk/handle/1810/247906`.

[102] Savannah Jennifer Thais. "Utilizing Electrons in the Search for Associated Higgs Boson Production with the ATLAS Detector: Higgs Decaying to a Tau Pair and Vector Boson Decaying Leptonically". Presented 14 Jun 2019. 2019. URL: `https://cds.cern.ch/record/2686964`.