



TECHNISCHE
UNIVERSITÄT
WIEN



DIPLOMARBEIT

Unbinned Energy Correlator Unfolding in the Context of Top Quark Mass Measurements

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

im Rahmen des Studiums

Technische Physik

eingereicht von

Simon Hablas

Matrikelnummer 11821438

ausgeführt am Marietta-Blau-Institut für Teilchenphysik
der Österreichischen Akademie der Wissenschaften

Betreuung:

Privatdoz. DI Dr. Robert Schöfbeck (ÖAW / MBI)

Dr. Dennis Schwarz (DESY)

Wien, November 25, 2025

(Unterschrift Verfasserin)

(Unterschrift Betreuer)



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

This work explores the novel approach of unfolding the threefold energy correlator with generative machine learning methods. In recent investigations energy correlators (EEEC) were found to have a top quark mass sensitive peak which is also accessible analytically. To extract the top quark mass by comparing data with analytic calculations, the measured correlators must be unfolded to correct for detector effects. Normalizing flows with conditional invertible neural networks (CiNN) and conditional flow matching (CFM) are two generative machine learning methods, able to perform unbinned unfolding. In this work, simulated data from boosted $t\bar{t}$ events was used to construct the correlators with emphasis on feasibility with machine learning methods, while allowing the comparison to theory. In contrast to latest studies, the jet p_T dependency of the mass peak, was solved via the construction of a new observable without this dependency. Configurations of the two models were found enabling the unfolding of these correlators. The concept of unfolding EEEEC could be demonstrated and showed promising results. Supplementary investigations regarding the generalization capabilities of the models showed a mass bias towards distributions originating from samples with $m_t = 172.5$ GeV for both models.

Danksagung

Mein herzlicher Dank gilt meinen Betreuern Robert Schöpfbeck und Dennis Schwarz, für ihre wertvolle Unterstützung, ihr offenes Ohr und die angenehme Zusammenarbeit. Ihre fachlichen Anregungen und ihre freundliche, konstruktive Art haben die Arbeit in vielerlei Hinsicht bereichert.

Über diese fachliche Unterstützung hinaus muss ich noch weiteren Personen meinen Dank aussprechen, alle auf ihre Art unerlässlich. Danke Franziska, Ursula, Günther, Marco, David!

Contents

| | |
|--|-----------|
| Abstract | 4 |
| 1 Introduction | 7 |
| 2 Theory | 9 |
| 2.1 Particles in the Standard Model | 9 |
| 2.1.1 Fermions | 9 |
| 2.1.2 Bosons | 10 |
| 2.2 Quantum Electrodynamics | 12 |
| 2.3 Quantum Chromodynamics | 13 |
| 2.4 Weak Interaction | 15 |
| 2.5 Electroweak Unification | 16 |
| 2.6 Higgs Mechanism | 17 |
| 2.7 Top Quark | 18 |
| 2.7.1 Production and Decay | 18 |
| 2.7.2 Top Quark Mass | 19 |
| 2.8 Proton-Proton Collisions | 22 |
| 3 Experimental setup and event reconstruction | 24 |
| 3.1 LHC | 24 |
| 3.2 The CMS Detector | 26 |
| 3.3 Objects Reconstruction and Jet Clustering Algorithms | 28 |
| 3.3.1 Particle Flow | 29 |
| 3.3.2 Jet Clustering | 29 |
| 4 Generative Machine Learning for Unfolding | 31 |
| 4.1 Normalizing Flows and CiNN | 33 |
| 4.2 Masked Affine Autoregressive Transformation | 35 |
| 4.3 Conditional Flow Matching | 36 |
| 5 Unfolding | 38 |
| 5.1 Binned Unfolding | 39 |
| 5.2 Unbinned Unfolding | 41 |

| | | |
|----------|--|-----------|
| 6 | Analysis | 44 |
| 6.1 | Analysis Strategy | 44 |
| 6.2 | Simulated Event Samples | 44 |
| 6.3 | Event Selection | 45 |
| 6.4 | Breit-Wigner Mass Reweighting | 47 |
| 6.5 | Constructing the Energy Correlator | 49 |
| 6.6 | Unfolding with Conditional invertible Neural Networks (CiNN) | 50 |
| 6.6.1 | Model Training | 50 |
| 6.6.2 | Unfolding Performance | 56 |
| 6.6.3 | Generalization and Bias | 60 |
| 6.7 | Unfolding with Conditional Flow Matching (CFM) | 65 |
| 6.7.1 | Model Training | 65 |
| 6.7.2 | Unfolding Performance | 66 |
| 6.7.3 | Generalization and Bias | 69 |
| 6.8 | CiNN and CFM in Comparison | 71 |
| 7 | Conclusions | 73 |
| 8 | Outlook | 75 |
| | References | 76 |

1 Introduction

The top quark (t) is the most massive fundamental particle in the Standard Model of Particle Physics (SM) [1]. According to the most precise current measurement [2] it has a mass of 172.52 ± 0.14 (stat) ± 0.30 (syst) GeV. This mass is of great interest, as it implies a strong coupling to the Higgs boson, which makes it important in the context of the electroweak symmetry breaking and vacuum stability [3]. Furthermore, the precise knowledge of the top quark mass enables self consistency tests of the SM and exploration of beyond SM physics [4]. In order to measure this mass, top quarks must be produced in high quantity. The particle accelerators therefore are required to feature a sufficient center-of-mass energy and luminosity. Both are provided by the Large Hadron Collider (LHC) at CERN. The Compact Muon Solenoid Detector (CMS) is one of the detectors along the LHC and used to perform top quark mass measurements. Multiple methods have been employed to obtain this mass from the measurements of the occurring proton-proton collisions. The most prominent of which is referred to as direct measurement [5] and effectively constructs the four momentum of a top quark by combining the four momenta of its decay products. This method infamously introduces an uncertainty interpreting the mass found by comparison to simulated data [6]. Recently however, novel approaches are discussed and evaluated mitigating this problem [7]. So called energy correlators are using the angular separation of final state particle triplets, to construct an observable with a top quark mass sensitive peak. This method lets one obtain predictions analytically due to its use of factorization of non perturbative physics. The method was proposed by [7] and tested regarding feasibility by [8, 9]. This kind of measurement, as the other, requires the unfolding of the experimentally obtained data. Unfolding in this sense means the process of reversing the influence the detector to measured variables. This task usually is performed by binned algorithms like TUnfold [10]. This work tests two Machine Learning methods to unfold energy correlators in the context of future top quark measurements. On the one hand normalizing flows with conditional invertible neural networks (CiNN) and on the other hand conditional flow matching (CFM). Both were already tested and used in the context of particle physics. The authors of [11, 12] used CiNN and the authors of [13] recently used CFM for unfolding. These ML methods require simulated data to learn the dependency of particle level data from detector level data.

In this work, energy correlators of simulated $t\bar{t}$ events were constructed and used to train models of the two approaches. These generative models were subsequently tested regarding their ability to unfold simulated data. Furthermore, the bias of these models

to their training data and their generalization ability were investigated.

In chapter 2 the theoretical background of the standard model, the top quark and the physics of proton-proton collisions is provided. Chapter 3 gives an introduction to the experimental setup of the LHC and CMS. The function of the two generative machine learning methods is explained in chapter 4. Chapter 5 elaborates on unfolding and different unfolding methods. The event simulation and correlator construction, as well as the unfolding performance and bias investigations are found in the analysis chapter 6.

2 Theory

The Standard Model of Particle Physics (SM) is the currently best-tested theoretical framework describing the known fundamental particles and three of the four fundamental forces: the strong interaction, the weak interaction, and electromagnetism. While the strong interaction is described by Quantum Chromodynamics (QCD), the weak and electromagnetic forces are unified in the electroweak theory. The SM is a relativistic quantum gauge field theory. Since in 2012 the Higgs boson was discovered by the ATLAS and CMS Collaborations [14, 15], all particles predicted by the SM have been experimentally observed. The model provides a rigorous mathematical structure and describes the properties, interactions, and symmetries of elementary particles. This chapter is largely inspired by [16].

2.1 Particles in the Standard Model

A fundamental distinction in particle physics is between particles with half-integer spin (fermions) and those with integer spin (bosons), as established in quantum mechanics. Fermions follow Fermi-Dirac statistics and constitute matter, whereas bosons follow Bose-Einstein statistics and mediate forces. The fundamental fermions and their properties are depicted in Fig. 2.1, the bosons in Fig. 2.2.

2.1.1 Fermions

Fermions are the matter constituents of the SM and have a spin of $1/2 \hbar$. The fermions in the SM are divided into two groups: quarks and leptons. Each group contains three generations of increasing mass, but otherwise identical quantum numbers. [17]

Fermions participating in the strong interaction are called quarks, while those that do not are called leptons. Among leptons, a further distinction is made based on electric charge: the electrically neutral ones are neutrinos, while the charged ones are simply referred to as charged leptons. The quantum numbers of fermions include electric charge and weak isospin. The weak isospin for left-handed fermions is $+\frac{1}{2}$ for up-type quarks and neutrinos, and $-\frac{1}{2}$ for down-type quarks and charged leptons. No fundamental neutral, strongly interacting fermions are known.

The first generation of fermions constitutes the stable matter found in everyday life. For each fermion, an associated antiparticle exists with the same mass but opposite quantum numbers, such as electric charge and baryon/lepton number.

Quarks carry a color charge and therefore participate in the strong interaction, mediated by gluons. They exist in six flavors in two types:

Up-type: up (u), charm (c), top (t) with electric charge $+\frac{2}{3}e$

Down-type: down (d), strange (s), bottom (b) with electric charge $-\frac{1}{3}e$

Each quark has a distinct mass, electric charge, and color charge. Quarks are not observed in isolation [18] but only in color-neutral bound states, which are called hadrons. Mesons are hadrons consisting one quarks and one anti-quarks. Baryons are hadrons with three quarks or three anti-quarks (Anti-Baryon). Masses of quarks range from a few MeV (up and down) to over 170 GeV (top quark)[17]

Charged Leptons are color-neutral and do not participate in the strong interaction. They include electron (e), muon (μ), and tau lepton (τ). Each has an electric charge $-e$ and an increasing mass with generation.

Neutrinos Neutrinos are electrically neutral leptons that interact only via the weak interaction. They exist in the three flavors electron-neutrino (ν_e), muon-neutrino (ν_μ), tau-neutrino (ν_τ). Though initially thought to be massless, neutrino oscillations imply that they have non-zero masses [19]. However, the exact mass hierarchy and absolute mass values remain unknown.

2.1.2 Bosons

Bosons are force carriers and obey Bose-Einstein statistics. They have integer spin and mediate the fundamental interactions of the SM.

Gauge Bosons The interactions in the SM arise from local gauge symmetries, each associated with a set of gauge bosons. The electromagnetic interaction is mediated by the photon (γ), which is a massless particle and therefore enables long-range interactions. The massive W^+ , W^- , and Z bosons, mediate the weak interaction. The W^\pm Bosons have a mass of 80.369 ± 0.00015 GeV and the Z boson has a mass of 91.188 ± 0.0020 GeV [1]. The mediator of the strong interaction are eight massless gluons (g), which carry color charge themselves [20].

Higgs Boson The Higgs boson is a scalar particle (spin-0) arising from the Higgs mechanism [21], which provides mass to the W and Z bosons and to fermions via Yukawa couplings. It was discovered in 2012 as the last particle predicted by the SM [14, 15].

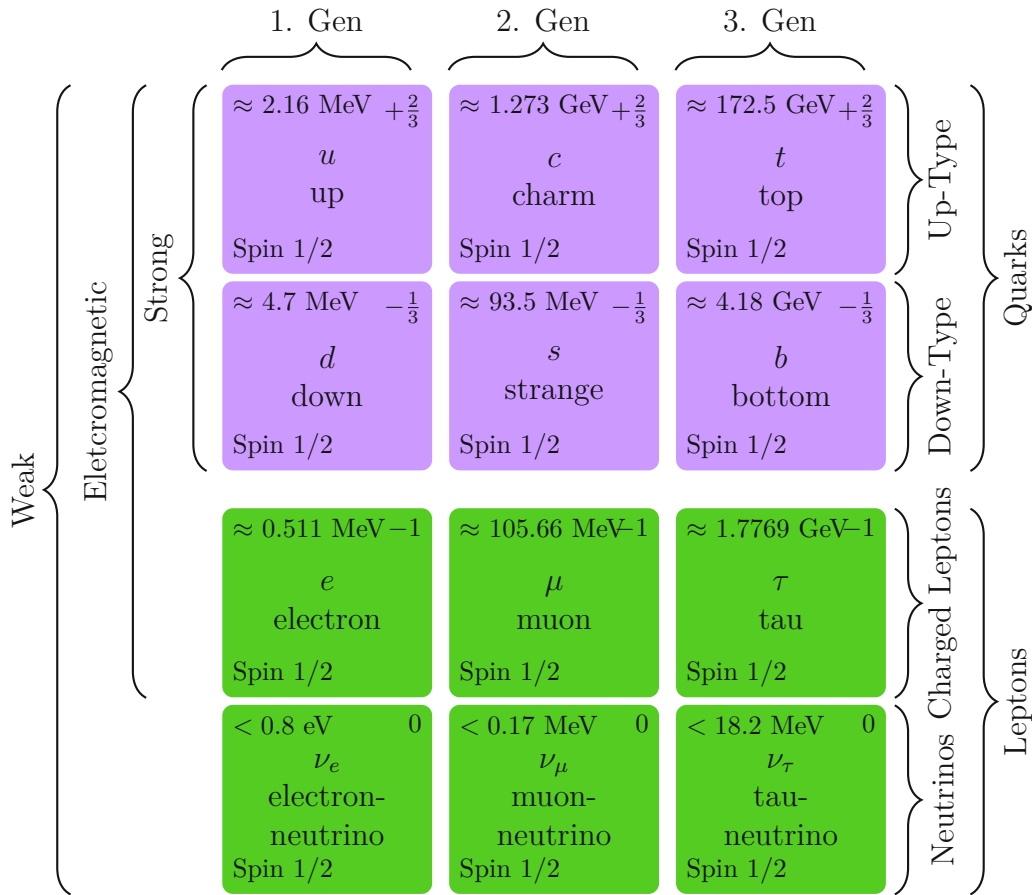


Figure 2.1: Fermions of the Standard Model. In the particle boxes corners, the mass (top left), the charge (top right, in units of e) and the Spin (bottom left) are depicted. The brackets on the left indicate the interactions the particles partake. The brackets on the right indicates their type. Particles of a generation are aligned vertically. Mass values from [1].

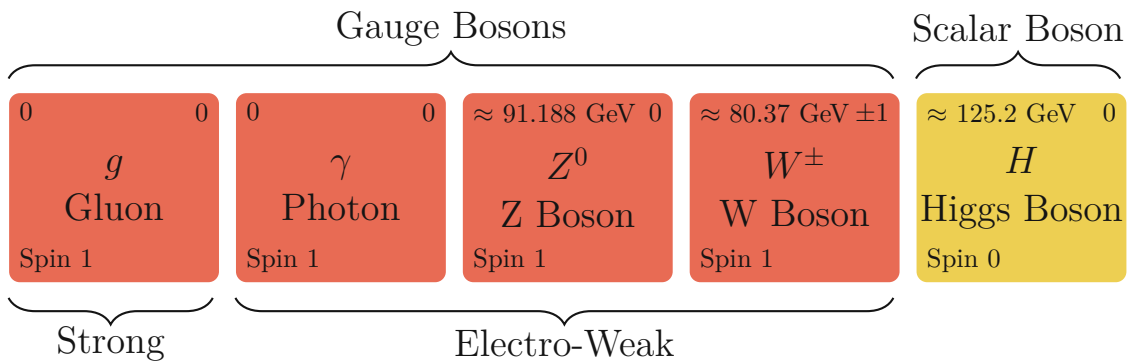


Figure 2.2: In the particle boxes corners, the mass (top left), the charge (top right, in units of e) and the Spin are depicted. The brackets on the top and on the bottom indicate their type and mediating interaction. Mass values from [1].

2.2 Quantum Electrodynamics

Quantum Electrodynamics (QED) describes the interaction of electrically charged particles via the exchange of photons and represents the quantum field theory of the electromagnetic force. As a gauge theory, based on a local U(1) symmetry, it provides a natural starting point for the construction of more complex interactions in the SM. This section outlines the derivation of QED to a fully gauge-invariant Lagrangian. The derivation illustrates how local symmetry requirements constrain the form of interactions.

The starting point are the Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi = \left(-\frac{1}{2m}\nabla^2 + V\right)\Psi \quad |c = \hbar = 1 \quad (2.1)$$

$$\hat{E}\Psi = \hat{H}\Psi \quad (2.2)$$

and the energy-momentum relation of special relativity

$$E^2 = p^2 + m^2. \quad (2.3)$$

Inserting the operators for energy $E \rightarrow i\frac{\partial}{\partial t}$ and momentum $p \rightarrow i\vec{\nabla}$, in the energy-momentum relation, one will obtain the Klein-Gordon [22, 23] equation:

$$\square\Psi + m^2\Psi = 0, \quad (2.4)$$

where $\square = \frac{\partial^2}{\partial t^2} - \nabla^2$ is the D'Alembert-Operator.

However, this leads to problems interpreting ψ^2 as a probability density, for single particles, due to the second order derivatives on the D'Alembert-Operator. Dirac proposed the following equation[24], which solutions would also need to obey the Klein-Gordon equation, to fulfill the relative energy-momentum relation

$$\hat{E}\psi = (\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + \beta \cdot m)\psi. \quad (2.5)$$

This implies, for $\boldsymbol{\alpha}$ and β to be matrices and obey anti commutator relations. Using again the operators for energy and momentum and the matrices γ^μ ¹, which are calculated

¹In the Dirac-Pauli representation, the Dirac matrices are:

from α and β and their constraints, one can write Eq. 2.5 as:

$$(i\gamma^\mu \partial_\mu - m) \psi = 0, \quad (2.6)$$

where $\gamma^0 = \beta$ and $\gamma^i = \beta\alpha^i$. The Lagrangian of the Dirac [25] equation 2.6 is

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi, \quad (2.7)$$

where one requires the Lagrangian to be invariant under local $U(1)$ phase transformations $e^{iq\Lambda(x)}$. This Lagrangian is not invariant under local transformations of the wave functions like $\psi \rightarrow \psi e^{iq\Lambda(x)}$. However observables are and therefore the derivative has to be adjusted accordingly.

$$\partial_\mu \rightarrow \partial_\mu - iqA_\mu = D_\mu, \quad (2.8)$$

where A_μ transforms with $A_\mu \rightarrow A_\mu + \partial_\mu \Lambda(x)$. This leads to the $U(1)$ invariant Lagrangian

$$\mathcal{L} = \bar{\psi} (i\gamma^\mu (\partial_\mu - iqA_\mu) - m) \psi \quad (2.9)$$

$$= i\bar{\psi}\gamma^\mu \partial_\mu \psi + q\bar{\psi}\gamma^\mu A_\mu \psi - m\bar{\psi}\psi. \quad (2.10)$$

In 2.10 one can identify a kinetic term of the fermion ($i\bar{\psi}\gamma^\mu \partial_\mu \psi$), a field A_μ interacting with the fermion via the coupling q and the mass term of the fermion. The field A_μ does not appear in a mass term, is therefore massless, and is considered the photon field. The charge q is the electric charge.

2.3 Quantum Chromodynamics

The approach to derive the strong interaction is similar as for QED. The charge in QCD is called color charge (hence Quantum Chromodynamics) and the symmetry group is $SU(3)$

$$\gamma^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad \gamma^2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix}, \quad \gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

They obey the Clifford Algebra $\{\gamma^\nu, \gamma^\mu\} = \mathbf{1} \cdot 2 \cdot \eta^{\nu\mu}$.

[20]. The Lagrangian should thus be invariant under these $SU(3)$ color-rotation transformations. Again, the partial derivation needs to be adjusted, to ensure this invariance:

$$\partial_\mu \rightarrow \partial_\mu - ig_s G_\mu^k \lambda_k = D_\mu. \quad (2.11)$$

For $U(1)$ in QED only one boson-field is required in the covariant derivative resulting in one type of photon. For QCD and $SU(3)$ there are 8 boson fields G_μ^k , associated with 8 types of boson, the gluons. The Gell-Mann matrices [26] λ_k are the 8 generator matrices of the Lie algebra and g_s is the strong coupling.

Analogous to QED, one can now obtain the $SU(3)$ invariant Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - g_s \bar{\psi}(\gamma^\mu \lambda_k G_\mu^k)\psi - \frac{1}{4} G_{k\mu\nu} G^{k\mu\nu} \quad (2.12)$$

$$= i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - g_s \bar{\psi}(\gamma^\mu \lambda_k G_\mu^k)\psi - \frac{1}{4} G_{k\mu\nu} G^{k\mu\nu}, \quad (2.13)$$

and interpret its terms, as kinetic fermion term, mass term, the gluon-field interaction term and a gluon-gluon interaction term. The latter gluon-gluon interaction is a consequence of the $SU(3)$ not being abelian in contrast to QED.

Confinement The gluon-gluon term in 2.13 implies, that gluons interact not only with quarks but also with each other and that they carry color charge themselves. This at least theoretically suggests bound gluon states [27] but also has further implications. Due to gluon-gluon interactions, the attractive force between two quarks does not decrease with increasing distance. The binding energy increases with distance to the point, where the production of a quark-antiquark pair is possible and energetically favorable. This new pair binds with the remaining quarks to hadrons, leaving no solitary quark behind. This phenomenon is called quark confinement [18] and explains why no free quarks are observed. Hadrons are color neutral, meaning that the colors of the constituents add up to zero or “white”.

Strong Coupling α_s The coupling constant in the strong interaction increases at lower energies and decreases at higher energies [28]. This opposes the electroweak interaction, and this has profound implications. For high energies and small distances, the coupling gets smaller and quarks behave as nearly free particles. This is called asymptotic freedom. For low energies, the coupling gets stronger. Late-state, low-energy hadronization can therefore, no longer be described using perturbation theory, since higher order terms are

no longer suppressed, but increased with α_s^n .

2.4 Weak Interaction

The weak interaction is mediated by three gauge bosons: the charged W^+ , W^- and the neutral Z^0 boson and the symmetry group of the theory is $SU(2)_L$ [29]. In contrast to the photon and gluons, these particles are not massless. They have a mass of $m_W^\pm \approx 80.4$ GeV and $m_Z \approx 91.2$ GeV. Due to the high masses of the mediators, the weak interaction is short-ranged. Only left-handed fermions and right-handed anti-fermions couple to the weak interaction, indicated by the L in the symmetry group. The weak gauge bosons therefore, couple differently to fermions depending on their chirality². This chiral structure poses a fundamental asymmetry that leads to parity violation and was first observed by [30]. Fermions are grouped into weak isospin doublets (only for left-handed particles), e.g.

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L, \quad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad (2.14)$$

while right-handed particles appear as singlets and do not couple to W^\pm bosons. Processes involving the weak interaction can change the flavor of fermions. For instance, a W^- boson can mediate the decay $d \rightarrow u + W^-$, which is a charged current process. However, it is experimentally observed that flavor transitions across generations (e.g. $s \rightarrow u$) also occur [31], although the $SU(2)_L$ theory only allows for flavor changes within isospin doublets. The flavor and mass eigenstates of quarks thus do not coincide. While the up-type quarks are defined to be aligned in both bases per convention, the down-type quarks mix via the Cabibbo-Kobayashi-Maskawa (CKM) [32] matrix V_{CKM} , which can be understood as a change of bases from the mass base to the flavor base:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.15)$$

²**Chirality and Helicity** Helicity is defined as the projection of a particles spin on its momentum. Chirality is a property of spinors, defined via the γ^5 matrix. Using the chiral projector operator $P_R = \frac{1}{2}(1 + \gamma^5)$ and $P_L = \frac{1}{2}(1 - \gamma^5)$ the right and left handed components of a spinor can be obtained. For massless particles, chirality and helicity are identical, but for massive ones, they differ.

The squared modulus $|V_{ij}|^2$ gives the probability of a quark of type j changing into a quark of type i in a weak decay.

As a consequence of the complex phase in V_{CKM} , the weak interaction is the only known interaction in the SM that violates the combined CP symmetry (charge and parity) [33].

2.5 Electroweak Unification

The electroweak theory unifies the electromagnetic and weak interactions. It is based on the symmetry group $SU(2)_L \times U(1)_Y$. The group $SU(2)_L$ gives rise to the three weak bosons W_1, W_2, W_3 , and $U(1)_Y$ leads to a fourth boson B . The charges are weak isospin T and hypercharge Y . The electric charge Q of a particle is then related to these charges via the Gell-Mann-Nishijima relation [26]

$$Q = T_3 + \frac{Y}{2}, \quad Y = 2(Q - T_3), \quad (2.16)$$

where T_3 is the third component of the weak isospin.

Gauge invariance under this symmetry group leads to the following bosonic field combinations [34]:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (2.17)$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (2.18)$$

where θ_W is the Weinberg angle, a free parameter in the theory. The photon field A_μ couples only to electric charge Q , while Z_μ couples to both T_3 and Y . This theory unifies the two interactions, however the resulting gauge bosons are all massless. This is not in line with experimental findings of massive W and Z bosons. A mechanism is needed to break the symmetry and give mass to the weak bosons maintaining gauge invariance. This problem is solved by introducing the Higgs mechanism.

2.6 Higgs Mechanism

The Higgs Mechanism introduces a new complex scalar field [21] $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ to the Lagrangian. This field comes with a potential

$$V(\phi) = \mu(\phi^* \phi) + \lambda(\phi^* \phi)^2,$$

which leaves the Lagrangian invariant under global phase rotations of ϕ . To also fulfill invariance under local $SU(2)_L \times U(1)_Y$ transformations, once again the covariant derivative is adjusted introducing boson fields to the Lagrangian

$$D_\mu = \partial_\mu - i\frac{g}{2}\tau^a W_\mu^a - i\frac{g'}{2}Y B_\mu.$$

If $\phi = 0$, this leads to QED of a charged scalar field. However for $\mu^2 < 0$, the potential forms the “Mexican hat” shape and the vacuum expectation value of ϕ will be nonzero, but take the lowest energy state

$$|\phi| = v = \pm \sqrt{\frac{-\mu^2}{\lambda}}.$$

This breaks the Lagrangian’s symmetry spontaneously. Inserting the expectation value in the Lagrangian, this symmetry breaking leads to mass terms for the weak gauge bosons

$$m_{W^\pm} = \frac{1}{2}gv \quad \text{and} \quad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}, \quad (2.19)$$

while no such term for photon appears, leaving it massless. The angle θ_W is related to the coupling constants via $\tan \theta_W = g'/g$. The Goldstone bosons that would appear due to the broken symmetry are “eaten” [21] by the W^\pm and Z fields to provide their longitudinal polarization components, necessary for a consistent description of massive spin-1 particles. This mechanism gives rise to a new scalar particle: the Higgs boson H . It was discovered in 2012 at the LHC with a mass of 125.09 ± 0.21 (stat.) ± 0.11 (syst.) GeV [14, 15].

Fermions get masses via the Higgs mechanism, with Yukawa couplings [35] which read

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L \phi \psi_R, \quad (2.20)$$

where $m_f = y_f \cdot v / \sqrt{2}$ and y_f is the Yukawa coupling. The values of y_f are free parameters

in the SM and are determined experimentally.

2.7 Top Quark

The top and bottom quarks were postulated by [32] in 1973. While the bottom quark and therefore a third generation of quarks was confirmed at Fermilab [36] in 1977, higher energy colliders were necessary to observe the top quark. It was discovered in 1995 at the Fermilab Tevatron collider [37, 38]. The Tevatron, a $p\bar{p}$ collider experiment with $\sqrt{s} = 1.8$ TeV was used to produce $t\bar{t}$, with a measured cross section of $6.8_{-2.4}^{+3.6}$ pb. A mass of $m_t = 176 \pm 8(\text{stat}) \pm 10(\text{syst})$ GeV was extracted [37]. This makes it the most massive quark, fermion, and elementary particle in the SM. With a lifetime of $0.5 \cdot 10^{-24}$ s [17] it is the only quark that does not hadronize. This section focuses on the production, decay and mass measurements of the top quark.

2.7.1 Production and Decay

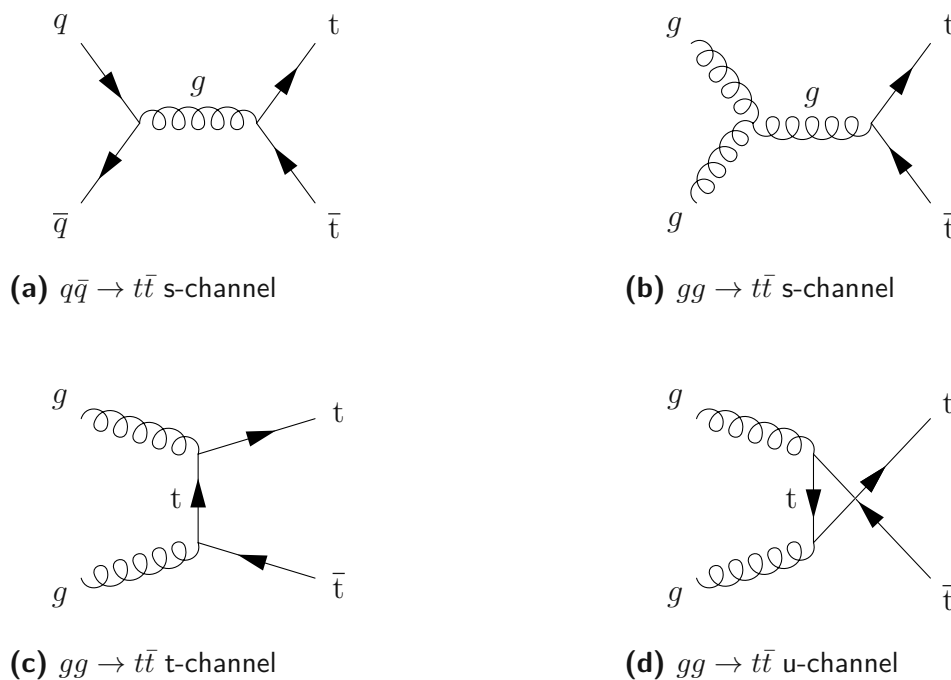


Figure 2.3: Leading order Feynman diagrams for top quark pair production via different channels. (a) Depicts the quark-anti-quark annihilation. (b-d) Depict gluon-gluon fusion. Source [39].

There are two fundamentally different possibilities for the production of top quarks. Either the weak interaction, producing single top quarks through, for instance the decay

of W Bosons, or the strong interactions producing a top anti-top pair, called $t\bar{t}$. Being a weak process, the cross section of single top production, is smaller than for the strong production channels. The latter are displayed in Fig. 2.3.

For a center of mass energy of 13 TeV, a cross section of $\sigma_{t\bar{t}} = 803 \pm 2$ (stat) ± 25 (syst) ± 20 (lumi) pb [40] was measured. At these energies, the processes via gluon-gluon fusion are dominant (90 %) [40].

The top quark predominantly decays via a charged weak process to a bottom quark and a W boson. The CKM matrix element is accordingly high $|V_{tb}|^2 > 0.99$ [41]. The decay of the top quark is typically characterized by the decay of the resulting W boson. There are two options both depicted in Fig. 2.4:

Hadronic The W^+ (W^-) decays to up-type quark (antiquark) and a down-type anti-quark (quark) pair, with a branching ratio of 67% [42].

Leptonic The W^+ (W^-) decays to a antilepton (lepton) neutrino (antineutrino) pair. The branching ratio of this process is 33% .

Events involving $t\bar{t}$ are characterized by the decay of both resulting W bosons: fully hadronic, fully leptonic, lepton+jets. While the high-energy lepton is a good marker to find $t\bar{t}$ events, its neutrino cannot be measured and therefore its kinematics are missing in the analysis. The fully leptonic process features two leptons and therefore has a distinct signature [43]. Nevertheless, the energy of two neutrinos is missing from the measurable decay products, and the branching ratio is low. In the process with two hadronic decays, the result are 6 hadronic jets, which are hard to differentiate from the jets produced by QCD multijet processes in high number. The lepton+jets process has the advantage that the lepton can be used as a marker, while in theory all of the decay products of the hadronically decaying top quark can be reconstructed, leaving no missing energy.

2.7.2 Top Quark Mass

The top quark mass is a key parameter in the SM due to its large value and the resulting strong coupling of the top quark to the Higgs field, making it important to electroweak symmetry breaking. Precise knowledge of the top mass enables consistency checks within the SM, particularly regarding the relation of the masses of the top quark, W boson, and Higgs boson. It also plays a crucial role in indirect searches for new physics, as deviations from predicted values may signal beyond SM phenomena. Additionally, its mass significantly affects the stability of the electroweak vacuum [3] through its effects

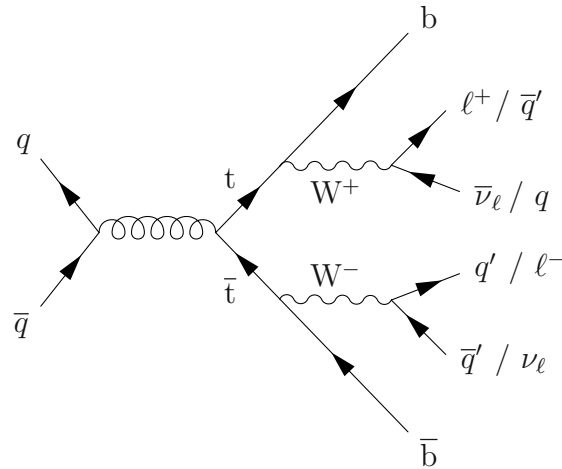


Figure 2.4: Feynman diagram of $q\bar{q}$ annihilation, eventually leading to a $t\bar{t}$ pair. The possible decay modes are depicted. The W^\pm either decays to a quark-anti-quark pair or a lepton-neutrino pair. Source [39]

on the Higgs potential, making it central to understanding the universe’s fundamental structure. [1, 44]. A comprehensive review of top mass measurements is given by [45]. In the following an overview of measurement methods is given.

Interpretation of the top quark mass As elaborated in detail in [6], there are multiple definitions of the top quark mass. The Monte Carlo mass m_t^{MC} is the mass parameter used as input for the generation of events in simulation. This mass is varied in order to find the best match of the simulated distributions to match experimental observables. The *pole mass* m_t^{pole} encodes the position of the pole in the renormalized quark propagator [46].

Parton shower and hadronization are modeled with non-perturbative models for low energies, introducing an uncertainty of up to 1 GeV [47] when relating the Monte Carlo mass to theoretically defined masses like m_t^{pole} .

Direct Measurements The mass of the top quark can be obtained via its decay products. They are identified and combined to reconstruct the invariant mass of the top quark. Although this method is precise with the value of the best single measurement being 171.77 ± 0.37 GeV [5, 45], it bears some difficulties. To obtain the mass, the measured observables, are compared to simulated observables by varying the top mass in the simulation. However, it is not possible to compare this fitted Monte Carlo mass to masses of popular renormalization schemes like the pole mass m_t^{pole} [6]. The used ob-

observables rely on the handling of parton shower and non perturbative low energy QCD effects. Since they are not calculable perturbatively, they are simulated and depend on simulation parameters.

Cross Section Another method of measuring the top quark mass is via the total production cross section $\sigma_{t\bar{t}}$. Since fermion propagators explicitly depend on the particle mass, so does the $t\bar{t}$ cross section. Thus, the top mass can be extracted by comparing the measured cross section to theoretical predictions calculated in perturbative QCD at next-to-leading (NLO), next-to-next-to-leading logarithmic (NNLL), or approximate next-to-next-to-leading order (NNLO) accuracy [48]. This method provides a complementary determination of the top quark mass and is less affected by ambiguities in kinematic reconstruction used in direct measurements [48]. However the result depends on the parton distribution functions, and thus inherits the according uncertainties. A value of $173.4_{-2}^{+1.8}$ GeV was obtained using dileptonic top decays at ATLAS and CMS[45, 49].

Measurements in the Lorentz-boosted regime In [47] it is shown that in certain regimes, observables can be constructed, by factorization theorems, that allow for purely analytical predictions not relying on simulation. The required regimes are boosted top jets. In this regime, of jets with $p_T > 750$ GeV, due to the high collimation, all the decay products fall in one single jet. Therefore, the total jet mass is sensitive to the top quark mass. It can be obtained from the differential cross section of the $t\bar{t}$ process. In [50] this method was used to obtain a value of $m_t = 173.06 \pm 0.84$ GeV. This value was obtained using a fit to simulated $t\bar{t}$ samples, since the measurement uses jets with $p_T > 400$ GeV and in this kinematic regions analytic calculations do not exist yet.

Energy correlators This method uses the substructure of boosted hadronic top decays and also relies on the factorization schemes, separating the hard decay physics from non-perturbative terms, discussed in Sec. 2.7.2. Rather than measuring the jet mass directly, the top quark mass is extracted through energy correlators encoding angular correlations of the energy flow within jets. As shown in [7], specific observables built from jet substructure such as the three-point energy-energy correlator (EEEC) feature a peak representing the top quark. The correlators encode the opening angle of the decay products, which is sensitive to the top quark mass m_t and transversal momentum p_T , allowing to obtain the top mass. Importantly this peak is highly insensitive to non-perturbative effects [7]. These properties make the method robust to comparisons with the theory predictions.

Energy correlators are defined as energy-weighted correlation functions of final-state particles, incorporating angular separations. The three-point correlator is defined as [7]

$$G^{(n)}(\zeta_{12}, \zeta_{23}, \zeta_{31}) = \int d\sigma \hat{\mathcal{M}}^{(n)}(\zeta_{12}, \zeta_{23}, \zeta_{31}), \quad (2.21)$$

where $\hat{\mathcal{M}}^{(n)}$ denotes the measurement operator applied to final states.

At hadron colliders, such as the LHC, the measurement operator can be written as:

$$M_{(\text{pp})}^{(n)}(\zeta_{12}, \zeta_{23}, \zeta_{31}) = \sum_{i,j,k \in \text{jet}} \frac{(p_{T,i})^n (p_{T,j})^n (p_{T,k})^n}{(p_{T,\text{jet}})^{3n}} \delta\left(\zeta_{12} - \hat{\zeta}_{ij}^{(\text{pp})}\right) \delta\left(\zeta_{23} - \hat{\zeta}_{ik}^{(\text{pp})}\right) \delta\left(\zeta_{31} - \hat{\zeta}_{jk}^{(\text{pp})}\right). \quad (2.22)$$

Where the summation is performed over all triplets of particles i , j and k in the boosted jet. Here, the angular variable is defined as[7]:

$$\hat{\zeta}_{ij}^{(\text{pp})} = \Delta R_{ij}^2 = (\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2. \quad (2.23)$$

The variables $\Delta\eta_{ij}$ and $\Delta\phi_{ij}$ are the pseudo-rapidity and azimuthal angle differences between particles i and j as elaborated in section 3.2. The measurement operator therefore effectively is a weighted histogram of ζ_{12} , ζ_{23} , ζ_{31} . For the mass sensitive peak to occur in this histogram, triplets in an approximately equilateral configuration are selected [7–9]. Triplets in these equilateral configurations are sensitive to hard decay dynamics and can be characterized via one value, namely the mean of the larger two sides of the triangle.

2.8 Proton-Proton Collisions

Since the LHC is a proton-proton collider and this work uses its data, this section elaborates on the physics of given inelastic collisions. Protons constitute of the three valence quarks, two up and one down quarks, but also of gluons and sea quarks. Therefore not the whole center of mass energy \sqrt{s} is available for a sub-process involving two partons i and j of the two protons. One has to account for their fractions of the protons momenta x_i and x_j . The energy available is $\sqrt{s'} = \sqrt{s \cdot x_i \cdot x_j}$ [1]. However, these fractions themselves follow a distribution, the Parton Distribution Functions (PDFs). The PDFs $f_i(x, q^2)$ describe the probability of finding a parton i in the proton carrying a momentum fraction x at a momentum transfer scale q^2 . This information is crucial to calculate cross sections dependent of the initial states. The cross section $\sigma_{ij \rightarrow X}$ where i and j are the possible partons leading to a final state X , is the convolution of the PDFs of i and j with

the according cross section summed over all possible i and j [51]

$$\sigma_{pp \rightarrow X} = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, q^2) f_j(x_2, q^2) \sigma_{ij \rightarrow X}(x_1, x_2, q^2).$$

They are typically measured at experiments with electron-proton-collisions, like at HERA. Nonetheless the kinematics of a given collision, depends on its unknown initial parton momenta fractions. The transversal momentum $p_T = \sqrt{p_x^2 + p_y^2}$ of the initial patrons, is due to the collimation of the proton beam known to be virtually zero. Thus the transversal momentum of collision products is an often used observable at CMS.

Pileup and Underlying Event In the analysis of hadron collision data, several additional effects must be taken into account. The crossing of two proton bunches may result in a hard scattering process, such as $t\bar{t}$ production, that triggers the data acquisition system. However, alongside this primary interaction, further processes typically occur. The remaining partons not involved in the hard scatter can also interact and produce detectable signals. This is called the underlying event [52]. Moreover, additional proton-proton collisions within the same bunch crossing, commonly referred to as pileup, contribute to further energy deposits in the detector. These additional interactions are predominantly low-energy (soft) QCD processes and can significantly affect the measurement and reconstruction of the primary event.

3 Experimental setup and event reconstruction

To investigate the properties of the top quark, it must be produced in sufficient quantities, which requires high center-of-mass energies. These are only achievable in large-scale particle colliders. To date, only two facilities have provided such conditions: the Tevatron at Fermilab and the LHC at CERN³. The top quark was first observed in 1995 at the Tevatron using the CDF [37] and DØ[38] detectors. As for today the LHC remains the only operational collider capable of producing top quarks, enabling precision measurements. These measurements are carried out by the ATLAS and CMS experiments, two general-purpose detectors at the LHC. This thesis aims to contribute to the next iteration of measurements and analysis and therefore, this chapter introduces the experiment and detector. It also outlines how particles and jets are reconstructed from the measured data.

3.1 LHC

The Large Hadron Collider (LHC) is a circular particle accelerator located at CERN near Geneva, at the borders of Switzerland and France. The machine is housed in a 27 km long [53] underground tunnel at a depth approximately 100 meters below ground. Since its initial operations in 2008, the LHC has proven successful and increased the center-of-mass energy since. During Run 2 (2016–2018), it operated at a center-of-mass energy of $\sqrt{s} = 13$ TeV [54]. The analysis presented in this thesis is based on simulations of proton-proton collision data from this run. The protons are delivered from hydrogen atoms, with electrons removed. The resulting protons are accelerated in multiple stages: first by the LINAC 2 (used during Run 2), followed by the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), before finally being injected into the LHC ring. The full acceleration system can be seen in Fig. 3.1 . Since head-on experiments are energetically favorable over fixed target experiments, the protons are inserted in both directions.

Proton beams are grouped into bunches, each containing roughly 10^{11} protons. Up to 2808 bunches can be stored in a beam, circulating in opposite directions in two vacuum-separated beam pipes. Dipole magnets steer the beams, while quadrupole magnets focus them. These magnets are superconducting and are kept at a temperature of 1.9 K using liquid helium to achieve the necessary magnetic field strengths. Their magnetic field

³Conseil Européen pour la Recherche Nucléaire “European Organization for Nuclear Research”

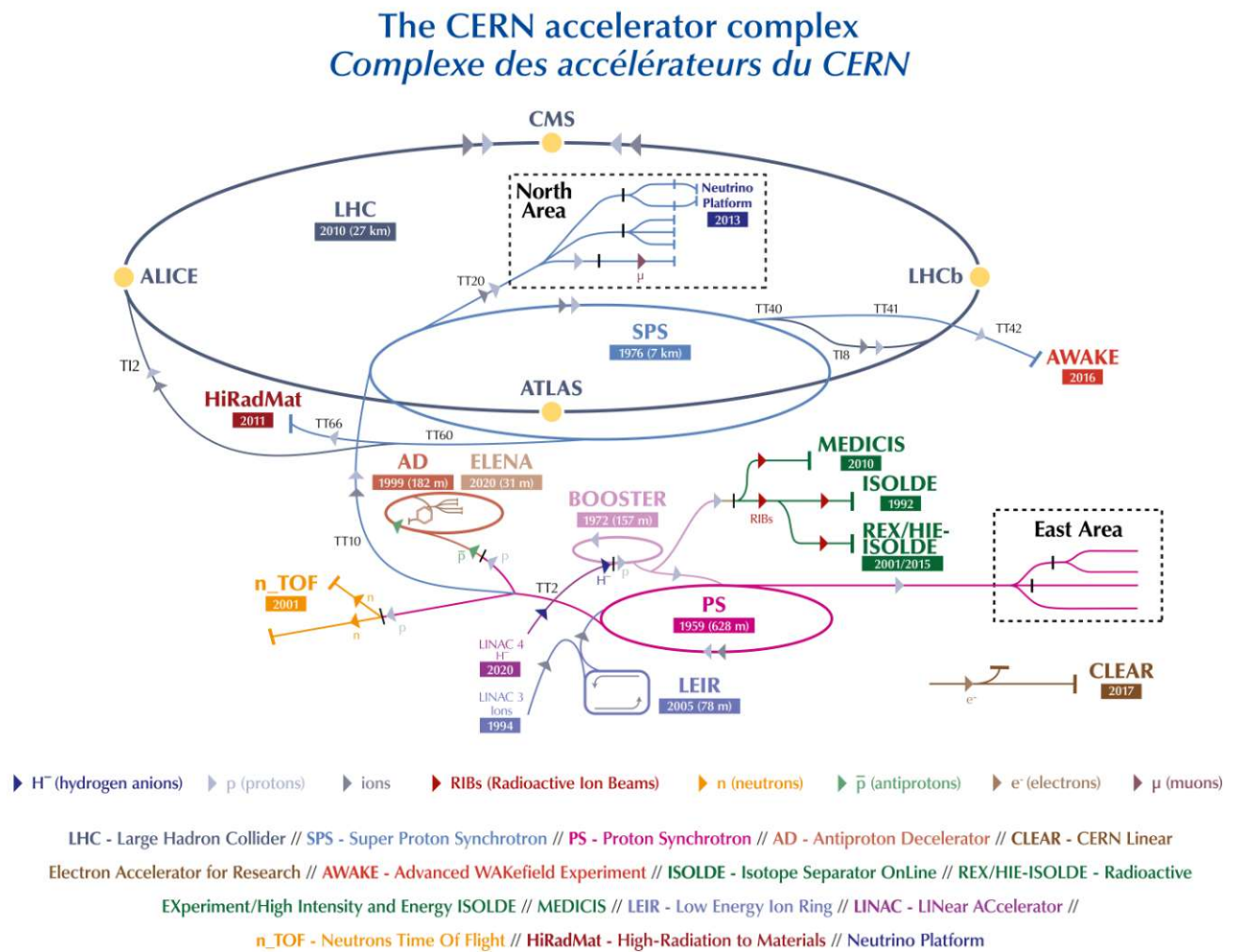


Figure 3.1: Overview of the CERN accelerator complex. LINAC 2 was updated with LINAC 4 in 2020. Succession from LINAC to PSB (Booster), PS and SPS in the bottom center. Figure Source: [55]

strength of the bending dipoles is up to 8.4 T.

Collisions occur every 25 ns, corresponding to a frequency of 40 MHz. Each bunch crossing can produce multiple proton-proton collisions. The number of collisions per crossing depends on the instantaneous luminosity, a measure of collision rate per unit area. It is calculated as:

$$L = \frac{nN_1N_2f}{4\pi\sigma_x\sigma_y}, \quad (3.1)$$

where n is the number of bunches, N_1 and N_2 are the number of protons per bunch, f is the revolution frequency, and σ_x , σ_y are the transverse beam sizes [8].

The LHC reached its design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ in 2016. The total number of events for a given process i can be estimated using the luminosity integrated over time $L_{\text{int}} = \int L dt$ and the cross section σ_i : $N_i = \sigma_i \cdot L_{\text{int}}$.

3.2 The CMS Detector

The Compact Muon Solenoid (CMS) [56] is one of the four major detectors at the LHC (among ATLAS, ALICE and LHCb). It is a general-purpose experiment capable of a wide range of physics studies, including the search for new particles and precision measurements of the SM. With its cylindrical structure measuring 22 m in length and 15 m in diameter, and a total mass of 14,000 tons, CMS combines multiple detector technologies into a layered design. The key components of CMS include (from inside to outside) the tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), the superconducting solenoid magnet, and the muon system.

Coordinate System A right handed Cartesian coordinate system is described in [56] and in use at CMS, it is depicted in Fig. 3.2. The origin is chosen to be at the collision point. The x -axis points toward the center of the LHC ring, the y -axis points upward, and the z -axis follows the beam direction. For angular coordinates, the azimuthal angle ϕ lies in the x - y plane, and the polar angle θ is measured from the beam axis. In practice, the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$ is used instead of θ , as differences in η are invariant under Lorentz boosts along the beam axis. As argued in Sec. 2.8, another common coordinate is the transversal momentum p_T , a projection of the momentum to the x - y plane, also depicted in Fig. 3.2.

Tracking System The tracker is the innermost part of CMS inside of the magnet. It consists of silicon semiconductor detectors, which basically are diodes used in reverse

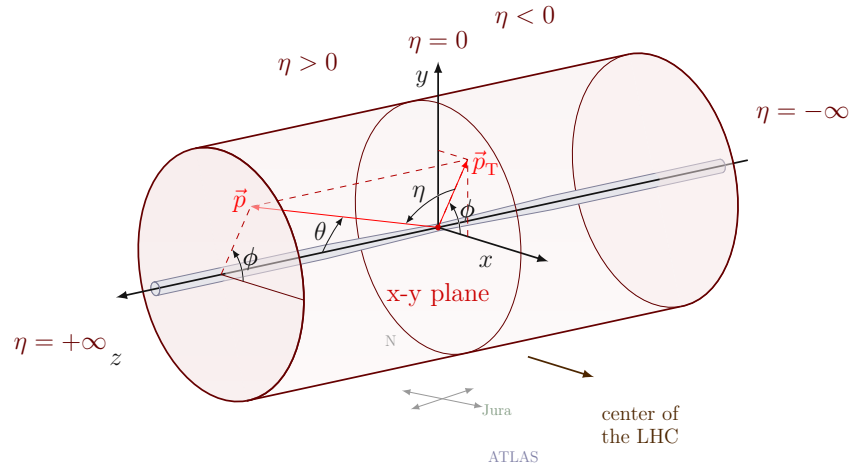


Figure 3.2: CMS Coordinate system. Figure Source: [57]

direction. Charged particles induce free electrons in the diode, which create a current leaving the depleted zone. This current can be amplified and measured. Multiple layers allow the spacial reconstruction of the charged particles trajectory in the tracker, enabling access to the tracks curvature and therefore momentum, due to the magnetic field. The tracker itself consists of multiple subdetectors. The innermost subdetector is the pixel-detector followed by the strip detectors. The strip detector consists of an inner and outer part each separated in a barrel section, meaning a cylinders coaxial with the beam and a disc section for $|\eta|$ up to 2.5.

Electromagnetic Calorimeter (ECAL) The ECAL [56] measures the energy of electrons and photons using scintillating crystals (lead tungstate PbWO_4). When a particle deposits energy in a crystal, it emits light with an intensity proportional to the energy deposit. This light is collected and measured by photo detectors. The ECAL covers up to $|\eta| < 3.0$, with a barrel section ($|\eta| < 1.48$) and two endcaps. The energy resolution is limited by statistics, noise through electronic and pileup and energy independent effects like calibration errors. It was measured by [58] using electron beams. The resolution increases with higher energies until the constant term with 0.3 % dominates. For an electron with 100 GeV, the resolution it is about 5%.

Hadronic Calorimeter (HCAL) Surrounding the ECAL is the HCAL [56], which measures the energy of hadrons. It features an alternating arrangement of dense absorber material (brass) and scintillator layers. The HCAL extends coverage up to $|\eta| < 5.0$, including forward calorimeters for the most extreme angles. Even though its energy res-

olution is lower than the energy resolution of the ECAL, it plays a key role in measuring jets and missing energy.

Solenoid Magnet The eponymous feature of CMS is its large solenoid magnet. It is 12.9 m long with an inner diameter of 5.9 m and generates a magnetic field of 3.8 T. This size allows the tracker, ECAL and HCAL to sit inside the magnet. It's field bends charged particles trajectories, allowing for momentum measurements, and is critical for particle identification. Outside the solenoid, the magnetic field is confined via massive iron yokes, which also forms part of the muon system. Since the magnetic field changes direction between inside and outside the solenoid, the curvature of charged particles also changes accordingly.

Muon System The muon system forms the outermost part of CMS. Since muons penetrate most detector layers, they can be detected in gas-based chambers embedded in the return yoke. The system includes drift tubes (DTs) in the barrel, cathode strip chambers (CSCs) in the endcaps, and resistive plate chambers (RPCs) in both regions. These allow for precise muon identification and momentum reconstruction over the range $|\eta| < 2.4$. [56]

Trigger System Given the LHC's high collision rate (up to 40 MHz) and the resulting data volume, CMS uses a trigger [59] system to reduce the data to a manageable rate. The trigger system uses two steps to successively filter events. The Level-1 Trigger (L1), is a hardware system based on custom electronics and selects potentially interesting events at a rate of about 100 kHz using fast detector information of the muon system and the calorimeters. These are then further filtered by the High-Level Trigger (HLT), a software-based system that performs detailed event reconstruction, reducing the output rate to about 1 kHz [59].

3.3 Objects Reconstruction and Jet Clustering Algorithms

The output of the CMS Detector of an event are in principle the measurements of the single detector subsystem described in 3.2. A crucial step in the analysis and further use of the data is to identify and reconstruct the individual particles that left signals in the detector. This is done with the Particle Flow [60] algorithm.

3.3.1 Particle Flow

The Particle Flow [60] algorithm exploits the distinct signatures of different particle types by combining information from all CMS subdetectors to achieve optimal identification. It proceeds by successively identifying particles based on their characteristic signals across the detector systems, removing associated data at each step. Particle tracks are reconstructed using data from the tracking and muon systems, and, together with the energy deposits in the calorimeters, serve as the primary input for particle identification. Muons and neutrinos are the only particles that can pass through the entire detector without being absorbed. However, only muons leave detectable tracks. They can therefore be identified by their presence in the muon system. Their charge, momentum, and energy are determined from the curvature of their tracks in the tracker and the muon system. Once identified, muon tracks can be removed before the algorithm proceeds to the remaining particles.

Electrons also produce tracks, but only in the inner tracker. Unlike muons, they deposit energy in the ECAL. Hence, tracker tracks that point to energy clusters in the ECAL are identified as electrons. Their properties can similarly be derived from track curvature. After electrons and muons are removed from the data, the algorithm identifies charged hadrons. These appear as tracks associated with energy deposits in both the ECAL and the HCAL. At this stage, all particles that produce tracks have been accounted for. The remaining energy deposits can then be interpreted as neutral particles: clusters in the ECAL correspond to photons, while those in the HCAL are associated with neutral hadrons.

3.3.2 Jet Clustering

The resulting quarks and gluons of a collision in a collider experiments, ultimately manifests itself as a shower of particles, measured in the detector. This shower is created from the initial particles by hadronization. These collimated particles shower are combined to jets, and the properties of those jets are input for analysis. Hence algorithms are required to obtain how the final state particle should be clustered into jets. There are multiple jet clustering algorithms in use, all of which are implemented in the C++ package *FastJet* [61]. Common ones are the k_T , anti- k_T , and Cambridge–Aachen algorithms, which are described here. They only differ in the parameter n . The algorithm begins with a list of all particle flow candidates and their kinematic properties, which are treated as *pseudojets* using the conventions of FastJet. In the first step, a distance measure is calculated for

each pair of pseudojets i and j :

$$d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \cdot \frac{\Delta R_{ij}^2}{R^2}$$

For each pseudojet i , a beam distance is also calculated:

$$d_{iB} = p_{Ti}^{2n},$$

where $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the angular distance between two pseudojets, p_{Ti} is the transverse momentum of pseudo-jet i , R defines the cone size of the clustered jets and n is the order parameter of the algorithm.

After computing all d_{ij} and d_{iB} values, the minimum of all these values is found. There are two possible cases:

1. The minimum is among the d_{iB} . In this case, the corresponding pseudo-jet is added to the final jet list and removed from the list of remaining pseudojets.
2. The minimum is among the d_{ij} . In this case, according pseudojets i and j are combined into a new pseudo-jet by summing their four-momenta.

The algorithm is iterative. It returns to the calculation of all d_{ij} and d_{iB} , and the process is repeated until no pseudojets remain. At this point, all particle flow candidates have been clustered into jets. The k_T algorithm with an $n = 1$ clusters soft particles first, while $n = -1$ (anti- k_T) favors clustering of high p_T particles and the Cambridge–Aachen algorithm with $n = 0$, does not consider, p_T at all and clusters solely based on the angular distance of particles. All these algorithms are infrared and collinear safe (IRC), meaning a particle with small momentum will not change the clustering and splitting particles in collinear parts, also leaves the clustering unchanged.

4 Generative Machine Learning for Unfolding

At its core, machine learning is applied mathematics, specifically applied algebra. Given sample data and an unknown relationship between inputs and observations, machine learning aims to approximate this dependency. By defining a suitable model and a computable metric that quantifies the performance of the model, the *loss*, the model can be optimized to generalize the underlying relationship and make accurate predictions. The model needs to be parametrized and differentiable with respect to these parameters. Learning refers to adjusting a model's parameters to optimize the modeled dependency and thereby minimize the loss function. The loss function accounts for the quality of the model of all given tuples of sample data. It is therefore a function of the total sample data and the models parameters, defined as

$$L = f(\vec{x}, \vec{y}, \theta), \quad (4.1)$$

where \vec{x} are the input variables, \vec{y} are the observations and θ are the model parameters.⁴ The gradient with respect to θ of said loss function $\frac{\partial f}{\partial \theta}$ is then used to minimize the loss function successively via a optimization algorithm of choice. The optimal parameters for the model are therefore

$$\theta^* = \arg \min_{\theta} f(\vec{x}, \vec{y}, \theta). \quad (4.2)$$

Neural Network A neural network (NN) is a parametric nonlinear model that is used to train unknown dependencies. It features a number of input variables in the input layers. These are inputs to the neurons in the next layer. For each node, its input values are multiplied with weights and summed up. This weighted sum is then input to the activation function of the node, which will determine its output and introduces nonlinearity to the NN. The weights in the weighted sum are the parameters θ , giving the NN its flexibility to be trained to minimize a constructed loss function. The outputs of one layer are the input to the next layer. Vital hyperparameters of NNs are its depth, meaning the number of hidden layers, the nodes per layer and the input and output layers, as seen in Fig. 4.1.

⁴While noted as vectors, a component of those vectors is abstract and can reach from one numerical values, to many values or a picture etc. However it is of relevance that not one input-observation-pair is used, but all.

Activation Function The activation function of a neuron determines its output-behavior. The activation function, which will give the output-value of the summation in the neuron above a certain threshold and 0 below, is called Rectifying Linear Unit (ReLU). In back-propagation the gradient of a neural network is calculated. This means the derivative of the activation function needs to be calculable. However, under the threshold-value of a neuron with ReLU activation, not only its output, but also the derivative of the output will be 0.

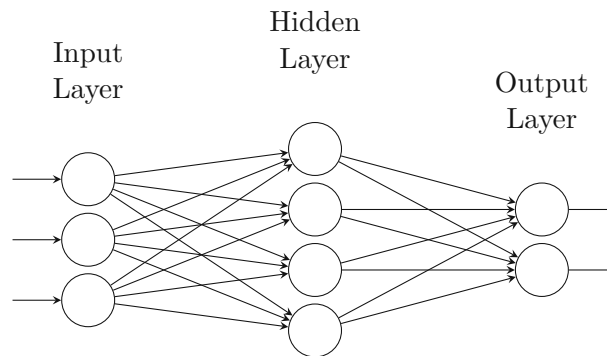


Figure 4.1: Diagram of a simple neural network. It features three input nodes, two output nodes and one hidden layer. The circles in the hidden layer and in the output layer are neurons. Every node from one layer is connected to every node in the next layer, thus it can be described as fully connected.

Optimizer and Learning Rates The optimizer performs the successive optimization of the model's parameter θ . It updates the parameters θ from epoch to epoch to find the minimum in the loss function. The most basic form of an optimizer would be a gradient descent

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(\theta_t), \quad (4.3)$$

where θ_t are the models parameter in the t^{th} epoch, and η is a parameter encoding the learning rate. However more sophisticated optimizers have prevailed. Frequently used optimizers are the adam [62] and adamW[63] algorithms. They feature momenta, meaning they update θ not solely based on the gradient of the loss function, but include gradients of previous epochs. This enables the algorithms to improve the convergence and reduce oscillations in the loss functions of the epochs. They differ in the way they handle regularization [63]. These optimizers, however, still require the learning rate as an additional hyperparameters. Choosing this hyperparameters requires a trade-off between

fast convergence to a global minimum which requires a high learning rate and the ability to find this minimum without leaving it due to too big steps introducing oscillations. This can be optimized by adapting the learning rate during the training from epoch to epoch. In the beginning of the training it is chosen to be high, quickly moving in the direction of the minimum in the loss landscape. Moving to higher epochs, the learning rate should decrease to avoid the said oscillations. Multiple learning rate schedulers are available to obtain this behavior. Typically are decays of the learning rate in the shape of steps or linear, exponential, and cosine functions. All these require the knowledge of the approximate epoch number of convergence, to be properly configured.

The AdamWSchedFree [64, 65] optimizer works without an explicit learning rate, and therefore without the additional hyperparameters of estimated convergence. It does so by incorporating the step size in the parameter update mechanism, also based on previous gradients. Only a starting learning rate is required.

Software implementation Machine learning applications are typically written in Python. There are multiple libraries handling the large matrices and the mathematical operation required for the training. Commonly used frameworks are pytorch [66] and tensorflow [67]. A key feature of these libraries is, that backpropagation, meaning the calculation of the gradient of the loss function with respect to the parameters θ , is automated. Thus, a model can be trained by setting up an architecture, defining a loss-function and iterate over the epochs and training data, in which the gradient is calculated via pytorch. The parameter update is performed by the optimizer. In this work the nflows [68] library was used to model normalizing flows. It is based on pytorch.

4.1 Normalizing Flows and CiNN

A normalizing flow [69] is a series of transformations from one distribution to another. This could be from one Gaussian probability density function to an arbitrary probability density function or, like in the context of this work, a series of transformations from a histogram to a Gaussian probability density function in latent space. Such a series of transformation is illustrated in Fig. 4.2. In each of these transformations, a change of variables is performed. Hence, in each transformation step in this series, one introduces a function to transform a variable a_n to a variable a_{n+1} , like $a_{n+1} = f(a_n)$. To again obtain the same probability $p(a_n)$ using a_{n+1} one has to account for the Jacobian of f , to keep the PDF normalized

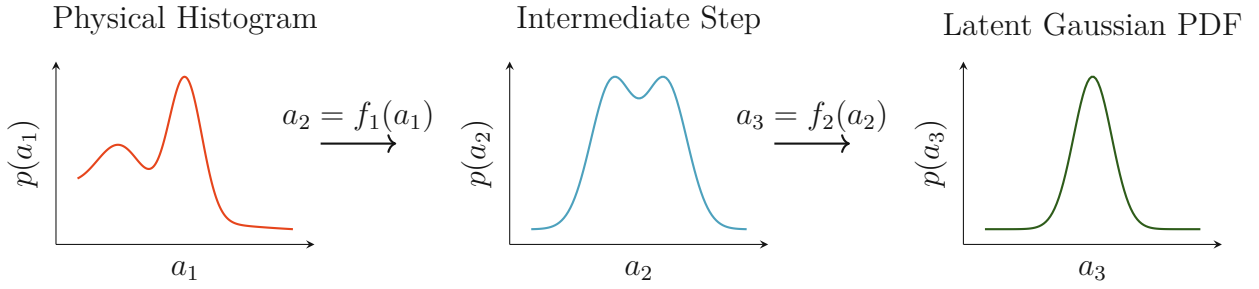


Figure 4.2: Sketch of a normalizing flow in one dimension with two transformations. $p(a_1)$ is obtained from the histogram of the observable. The latent space distribution $p(a_3)$ is defined to be a gaussian.

$$p(a_n) = p(a_{n+1}) \cdot \det \left| \frac{\partial f(a)}{\partial a} \right|. \quad (4.4)$$

Using a series of n transformations from a_0 to a_n where

$$a_n = (f_n \circ f_{n-1} \circ \dots \circ f_2 \circ f_1)(a), \quad (4.5)$$

this reads

$$p(a) = p(a_n) \prod_i^n \left| \det \frac{\partial f_i}{\partial a_{i-1}} \right|, \quad (4.6)$$

and accordingly

$$p(a_n) = p(a) \prod_i^n \left| \det \frac{\partial f_i}{\partial a_{i-1}} \right|^{-1}. \quad (4.7)$$

Using these transformations, one can calculate the probability of a physical value to fit to the latent space distribution and vice versa. The latent space distribution is known exactly, since it is set to a Gaussian of the desired dimension. The transformation is constructed such that it depends on the parameters θ . Subsequently, the probability depends on θ and can be interpreted as a likelihood. The goal is to find a transformation, that maximizes this likelihood for all the given values of the training data in the physical space. Thus, the total likelihood is the product of the single value likelihoods. We choose the loss to be the negative logarithm of the likelihood such that single likelihood fractions

can be summed and the optimization becomes a minimization problem. The gradient of these transformations with respect to the parameters can now be computed in order to determine the optimal parameters using gradient descent.

With a trained model one can now sample from the known latent space and perform the transformations to the physical space, which is thus required to be invertible. Furthermore in the context of unfolding, it is required to feed additional input to the transformations. The normalizing flow should be able to be trained and sampled from under a certain condition in the sense of Bayesian statistics.

4.2 Masked Affine Autoregressive Transformation

Above in 4.1, a series of invertible conditional transformations is mentioned. Masked affine auto-regressive transformations enable invertible conditional transformations based on a neural network. The structure of such a NN is depicted in Fig. 4.3. The key to invertibility lies in the masking within the network, meaning that not every node in a layer is input to every node in the next layer. Masking ensures that each output depends only on the corresponding input and all preceding inputs, enforcing a strictly autoregressive structure. Masking is applied directly to the neural network's weights by effectively setting certain weights to zero. In Fig. 4.3, this is reflected by the fact that the outputs of a depend solely on a, the outputs of b depend on a and b, and the outputs of c on a, b, and c. The underlying transformation are affine transformations and follow

$$\begin{aligned} a_{n+1} &= a_n \cdot s_a + t_a \\ b_{n+1} &= b_n \cdot s_b(a_n) + t_b(a_n) \\ c_{n+1} &= c_n \cdot s_c(a_n, b_n) + t_c(a_n, b_n), \end{aligned} \tag{4.8}$$

where s_i and t_i are determined by the NN. The parameter of the first transformation s_a and t_a have no dependency and are scalar.

This structure enables the inversion of the transformations at each step of the flow without inverting the NN, as one can see from Eq. 4.8. However, in this case no dependency of a on c can be modeled. This is dealt with a permutation of the input variables between the transformations in the flow.

The NN can be conditioned using additional inputs. These can be understood in a way, that the NN provides a transformation *given* a certain condition. This condition, is required in training, and also in sampling allowing to use this setup as an unfolding

technique, as it is introduced in Sec. 5.

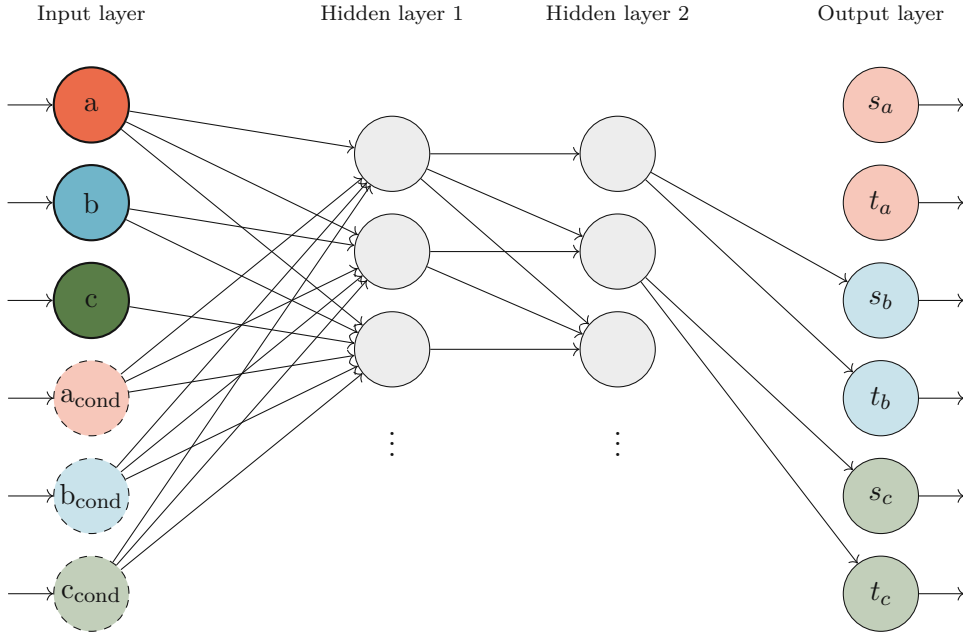


Figure 4.3: Structure of a Neural Network featuring masking, three inputs, three conditions and two outputs per input. The masking is visible by the connections. The NN is not fully connected, as there are no connections, from the b and c inputs to the a output. This reflects Eq. 4.8. s_a and t_a are not connected, as they are independent of the NN.

4.3 Conditional Flow Matching

Conditional Flow Matching (CFM) [70] assumes a continuous transition of an arbitrary distribution to a gaussian latent space. The method is described in detail and applied in [13, 71, 72]. The said transition can be thought of as a flow starting at $t = 0$ in the physical space, to the latent space at $t = 1$. This corresponds to a density $p(x, t)$

$$p(x, t) \rightarrow \begin{cases} p_{\text{data}}(x) & t \rightarrow 0 \\ p_{\text{latent}}(x) & t \rightarrow 1 \end{cases}.$$

A velocity field is introduced, which describes the transition from the physical space $x_{t=0} = x_0$ to the latent space $x_{t=1} = x_1$ at a time t

$$v(x, t) = \frac{dx}{dt}, \quad (4.9)$$

where x is the linear interpolation from x_0 to x_1 :

$$x(t, x_0, x_1) = x_0 \cdot (1 - t) + x_1 \cdot t. \quad (4.10)$$

The velocity $v(x, t)$, therefore, depends on the physical space quantity x_0 , which one wants to avoid. However, the neural network v_θ can be trained to this dependency with a mean square error training [71] to obtain θ^*

$$\theta^* = \operatorname{argmin}_\theta \langle (v_\theta(x_t, t) - v(x, t, x_0, x_1))^2 \rangle \quad (4.11)$$

$$= \operatorname{argmin}_\theta \langle (v_\theta(x_t, t) - (x_1 - x_0))^2 \rangle. \quad (4.12)$$

For each training data point x_0 , t is sampled from $\mathcal{U}(0, 1)$ ⁵, x_t is calculated by the linear interpolation and x_1 is random noise, sampled from the latent distribution.

Once the velocity field is obtained, it can be used to generate physical samples from latent samples, solving the ordinary differential equation (ODE)

$$x_0 = x_1 + \int_1^0 v_\theta(x, t) dt, \quad (4.13)$$

which is the integral version of Eq. 4.9. The starting point $x(t = 1) = x_1$ is sampled from the latent space.

To use this method for unfolding, one needs to condition this velocity field on a separate input. This is done by additional inputs to the Neural Network describing the velocity field during training and sampling. The training of v_θ then becomes

$$\operatorname{argmin}_\theta \langle (v_\theta(x_t, t, \text{cond}) - (x_1 - x_0))^2 \rangle, \quad (4.14)$$

respectively the sampling becomes

$$x_0 = x_1 + \int_1^0 v_\theta(x, t, \text{cond}) dt. \quad (4.15)$$

⁵Uniform Distribution [0, 1]

5 Unfolding

In collider experiments, direct comparison of experimentally measured values with analytically calculated values is not possible. This is due to the effects of the detector. One method to compare the prediction to the measurement, would be to apply a detector simulation to the predicted values. The other method is unfolding the measured observables. Unfolding refers to the method of correcting for detector effects, enabling a comparison to different experiments and with analytic predictions. This section introduces the concept of unfolding, outlines different unfolding methods, and presents general terminology.

Particle Level This is what one would generally define as truth values or distributions, after parton shower and hadronization. The domain of particles before those processes occur is called *parton level* and for $t\bar{t}$ decays would be the top quarks and W bosons decay products. In the context of experiments, particle level, refers to the physical values of the final state particles before they are measured. Of course these are not accessible without the detector. In the context of simulated data, this refers to the properties of particles and jets, after parton shower and hadronization, but before the detector simulation was performed.

Detector Resolution The finite resolution of the detector leads to a smearing of the measured values with respect to the truth values. This means that a value might shift from one histogram bin to another [10].

Detector Simulation The C++ based framework GEANT4 [73] provides a MC based simulation of the interaction of particles with matter. The software features particle tracking and detector responses in the CMS detector. Therefore, it can simulate particles in the detector and what the latter would measure under given particle level data.

Detector Level Observables at the detector level are the ones, that actually were measured and reconstructed. In the context of simulation, this refers to the values of observables, after the GEANT4 detector simulation and the object reconstruction, described in Sec. 3.3, were performed.

Unfolding is the process of obtaining **particle-level** observables based on **detector-level** observables. Various unfolding methods and algorithms exist. Some are operating on binned data, while others use unbinned approaches. In CMS and ATLAS analyses,

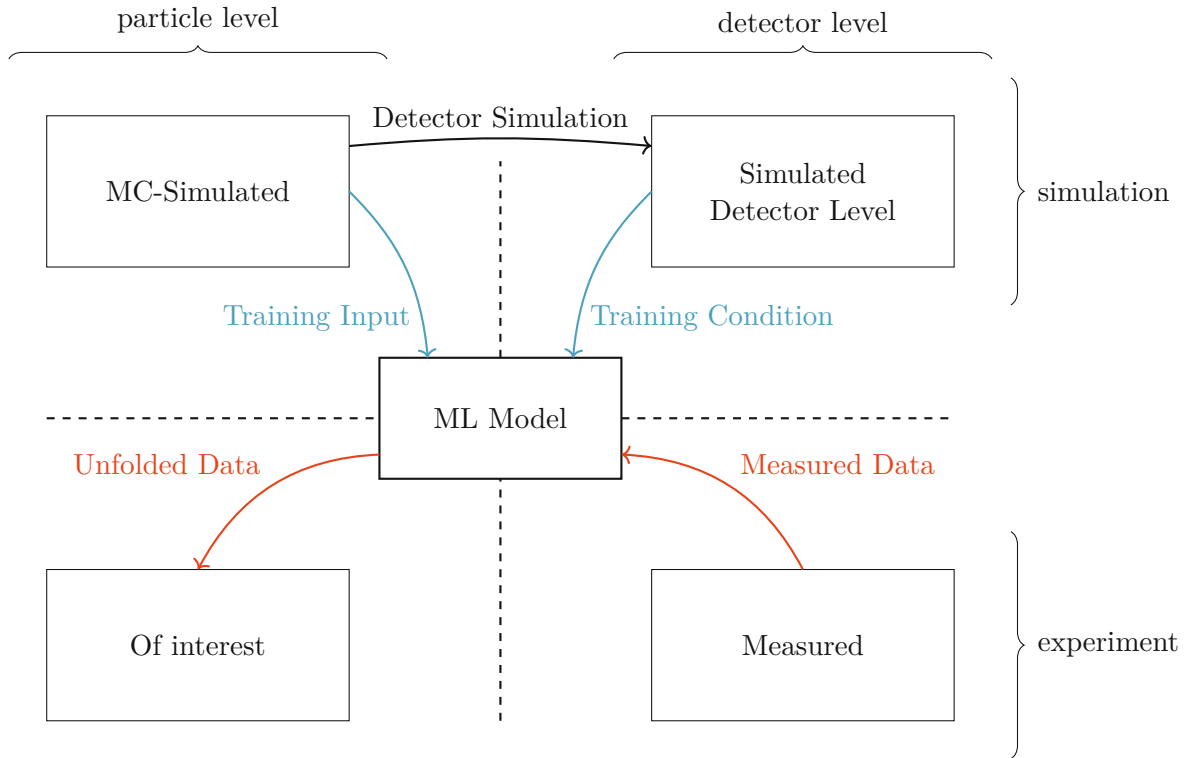


Figure 5.1: Data flow diagram of unfolding. The presented unfolding strategies all use simulated data to obtain the particle level truth from measured detector level data. In the diagram, a machine learning model is displayed, as in this thesis machine learning models are presented.

binned unfolding methods are typically used as the standard approach, whereas in this work unbinned methods are presented. The following presents a selection of both types, along with a discussion of their respective properties. All methods rely on simulated particle and detector-level data. Figure 5.1 schematically depicts the relation between particle-level, detector-level, and the measured and unfolded distributions in the unfolding procedure.

5.1 Binned Unfolding

In the discretized approach of binned unfolding one bins the truth data on particle level \mathbf{T} and the measured data on detector level \mathbf{M} . They then obey

$$\mathbf{M} = \mathbf{R}\mathbf{T}, \quad (5.1)$$

where the matrix \mathbf{R} is the detector response.

One could invert \mathbf{R} to obtain a particle level distribution from the detector level distribution. The inversion of this matrix, is unstable due to the noise and the counting uncertainty of the data and leads to oscillations in the solutions [10, 74–76]. There is a multitude of methods to solve this problem via constraints and regularization.

Iterative Bayesian Unfolding

The response matrix provides the conditional probability that an event originating in truth-level bin j is reconstructed in detector-level bin i , denoted as $P(i | j)$. The inverse probability $P(j | i)$, represents the probability that an event observed in detector bin i originated from truth bin j . $P(j | i)$ is not directly accessible but required to perform the unfolding measured data. However it can be computed using Bayes' theorem :

$$P(j | i) = \frac{P(j) P(i | j)}{P(i)}. \quad (5.2)$$

Where

- $P(i | j)$ is given by the response matrix, usually obtained from Monte Carlo simulation,
- $P(j)$ is the prior probability distribution of the truth observable on particle level, the distribution of question,
- $P(i) = \sum_k P(i | k) P(k)$ is the total probability for an event to be reconstructed in bin i , accessible via the measurement.

The iterative Bayesian unfolding method [76] starts with an initial guess for the prior $P(j)$ and uses it to compute the posterior $P(j | i)$. This posterior is then used to reassign the measured data back to the truth bins and update the estimate of the truth-level distribution. The updated distribution becomes the new prior in the next iteration. This is repeated until convergence or a defined number of iterations is reached, serving as regularization.

TUnfold

This algorithm makes use of Tikhonov regularization and an optional area constraint [10]. One might solve Eq. 5.1 for \mathbf{T} with a least square fit: $\arg \min |\mathbf{RT} - \mathbf{M}|^2$. The Tikhonov

regularization introduces a penalty term to that least square fit: $\arg \min |\mathbf{RT} - \mathbf{M}|^2 + |\lambda \mathbf{LT}|^2$. By choice of the regularization matrix \mathbf{L} , the bin yields, the first, or second order derivatives in T can be suppressed, making the unfolding stable. Another term in the least square fit is optional and constraints the area of the truth values to the measured values.

SVD

In this method a singular value decomposition of the detector response matrix is performed $\mathbf{R} = \mathbf{USV}^T$ [75]. The matrix \mathbf{S} contains the singular values of the decomposition in its diagonal. Small singular values correspond to ill-conditioned components of the response matrix. By truncating these small singular values, the new response matrix can be inverted to obtain the truth data.

These binned algorithms theoretically could be applied to data of higher dimensions, however the dimension of the response matrix would also grow with the data dimension, making this unfeasible, as the introduction of additional dimensions may lead to sparsely populated bins. The following two approaches work with unbinned data, and with any number of input data dimension. Both of them use machine learning techniques.

5.2 Unbinned Unfolding

In unbinned unfolding, the observables are not binned, but their distributions are interpreted as probability density functions. The measured detector level distribution $m(y)$ is then the physical particle level $t(y)$ distribution convoluted with the detector response function $r(y|x)$ [12]

$$m(y) = \int r(y|x)t(x)dx, \quad (5.3)$$

which is the continuous version of Eq. 5.1. In this sense, unfolding is the inversion of $r(y|x)$ to finally obtain $t(x)$.

OmniFold

OmniFold [77] uses two classifiers and event weights to iteratively reconstruct the particle-level distribution from measured data. The required inputs are again simulated events

with paired observables on particle and detector level, as well as the measured observables on detector level.

In the first step, a classifier is trained using machine learning to distinguish simulated detector-level events from measured detector-level events. The classifier outputs the probability $D(x)$ that a given event x originates from the measured data. From this, the density ratio between data and simulation can be estimated as

$$w_{\text{det}}(x) = \frac{p_{\text{data}}(x)}{p_{\text{sim}}(x)} \approx \frac{D(x)}{1 - D(x)}.$$

This ratio quantifies how under- or overrepresented an event is in the simulation relative to the real data, and is used to assign a weight to each simulated detector-level event so that the weighted distribution matches the measured data.

In the second step, these weights are transferred to the corresponding simulated particle-level events. A second classifier is then trained to distinguish the reweighted particle-level simulation from the original (unweighted) particle-level simulation. This classifier provides a new density ratio, which is used to reweight the particle-level simulation. The process is then iterated to progressively approximate the true particle-level distribution that best explains the measured detector-level data.

CiNN

As introduced in section 4.1, this method features a model consisting of a series of transformations each governed by a conditional invertible neural network. It learns to transform each particle-level event from the simulated data to the latent space distribution given the condition of the corresponding detector-level event observables. Once the training is completed, the model has learned the particle levels conditional dependency on the detector level. This invertible model can then be used to unfold. This means sampling from the latent space, given the condition of a measured detector-level event as input to the transformation, and applying the backward transformation to the physical space. The outputs are the observables of an unfolded particle-level event.

This method was used in [11, 12], where an iterative approach was chosen featuring classifiers and reweighting according to phase space density similar to Omnifold.

CFM

This method as described in 4.3, is similar to the CiNN approach, in that it also implicitly learns the detector response and uses the detector level data as a condition to tweak the transformation from the physical space to a latent space. Unfolding with the trained model is in principle also similar, as it means sampling from the latent space and applying a transformation conditioned with the detector level observables to obtain the unfolded particle level observables. However the NN are not invertible per se and a ODE needs to be solved to invert the transformation. CFM is the continuous version of a normalizing flow in the sense, that the trained velocity field provides a conditional transformation for every point between the physical and latent space. CFM was featured in [13] also in the context of unfolding top decays.

6 Analysis

This part of the thesis examines the performed analysis of unfolding energy correlator top mass measurements with the two approaches. Starting with Sec. 6.1 the Analysis Strategy is introduced. The simulation and selection of $t\bar{t}$ events that are used to train and test the unfolding model are explained in Sec. 6.2. A description of the implementation of two unfolding model follows in Sec. 6.7 and Sec. 6.7.

6.1 Analysis Strategy

The scope of this work is the demonstration of the unfolding of observables needed for a top quark mass measurement using energy correlators. These machine learning methods were trained and validated using simulated events from the CMS experiment corresponding to the data collected in 2018 in RunII of the LHC. Following the work of [8] and for the reasons described in Sec. 2.7.1 this analysis is based on simulated data of $t\bar{t}$ events in the lepton+jets channel. Only highly boosted top quarks are selected, such that the decay products of the hadronic decay are collimated into a single jet. The decay products of the hadronically decaying top quarks are identified, selected and matched to each other on particle level and detector level. These matched particles are used to calculate the energy correlators on both the particle and detector level, resulting in a list of energy correlators and energy correlator weights, on both particle and detector level with a one-to-one correspondence. This pairs of correlator data combined with the total jet momentum $p_{T, \text{jet}}$ of the underlying event are further filtered and consequently used as the training input of the two machine learning models. After a converged training of the models, orthogonal simulated data is used to sample particle level data given the datasets detector level data, effectively validating the unfolding method by comparing the unfolded results with the particle-level truth.

To investigate the models generalization ability, the validation data is modified to correspond to a simulated sample originating from $t\bar{t}$ events with a different top quark mass using Breit-Wigner-Reweightings.

6.2 Simulated Event Samples

In this work, simulated data, that corresponds to Run II in 2018 of the CMS experiment at the LHC is used to train the unfolding model. The simulation of this data, is performed in three steps:

Event Generation The calculation of the hard scattering events leading to the production of a $t\bar{t}$ pair, requires the calculation of the corresponding matrix element under the considerations of the incoming protons parton distribution functions. These calculations are performed by employing Monte Carlo simulations. The center-of-mass energy is $\sqrt{s} = 13$ TeV. In the context of this work, POWHEG [78] is used. It calculates the matrix elements up to next-to-leading order (NLO) terms in QCD.

Parton Shower and Hadronization PHYTHIA [79] is used to simulate hadronization and parton shower of the decay products of the top quark. Pythia uses a p_T -ordered shower and the Lund string model for hadronization. The outputs are the final state particles on particle level. For the simulated samples in this work, the CP5 tuning was used.

Detector Simulation The interactions of the final state particles with the CMS detector are simulated using Geant4 [73]. Geant4 models the passage of particles through the detector material, including all relevant electromagnetic and hadronic interactions. The output of this step corresponds to the reconstructed observables on detector level as they would be seen by CMS.

The dataset therefore includes the final state particles on particle level and detector level. The dataset is stored and manipulated with the particle physics software ROOT [80].

Three different samples with a m^{MC} of 172.5 GeV, 171.5 GeV, and 173.5 GeV were used in this analysis.

6.3 Event Selection

This section describes, how the matching particles (particle level and detector level) were obtained for each $t\bar{t}$ event in the simulated data.

Jet Selection and Matching For a hadronically decaying parton level top quark, the particle level AK8 jet with the smallest relative angular distance ΔR is identified. Only jets with a $p_{T, \text{jet}} > 400$ GeV are further used, to select jets, where the decay products fall within one jet of $R = 0.8$. In the next step, the detector level AK8 jet with the smallest angular distance ΔR to the found particle level jet is identified and also required to have a $p_{T, \text{jet}} > 400$ GeV.

Particle Selection and Matching Once the closest particle level and detector level jets are found, the jet constituents are filtered and prepared for matching. Constituents at the particle level are matched with constituents at the detector level by their angular distance and similarity in p_T . Matches are made if the constituents have $\Delta R < 0.05$. If more than one detector level constituent match via this criterion, the one with the closest p_T is matched.

Figure 6.1 depicts the number of jet constituents for different selection parameters. The number of triplets grows with the number of jet constituents cubed. As the computational time and the required resources for the unfolding model training grow with the number of triplets, a selection of the jet constituents is required. As mentioned in Sec. 3 and [8], only charged particles are suitable for a good momentum reconstruction, as neutral particles do not interact with the tracker subdetectors of CMS. Thus, in this analysis only charged particles are used for the triplet construction. The number of jet constituents was further reduced with the employed selection of the minimal transversal momentum p_T of the single constituents $p_T > p_{T \text{ min}}$. This can be done because it was found that the soft particles do not change the correlator distributions [8]. The peak in the number of jet constituents, for the neutral and charged particles Fig. 6.1 (red), lies between 50 and 75 constituents per jet. For only charged particles (blue) it lies around 30 constituents per jet. By selecting charged particles with $p_T > 5 \text{ GeV}$, the peak moves to 0 to 25 constituents per jet. By using charged particles with $p_T > 5 \text{ GeV}$ (green), the number of triplets is reduced by roughly two orders of magnitude, to about 1% of the number obtained when including all neutral and charged constituents (red).

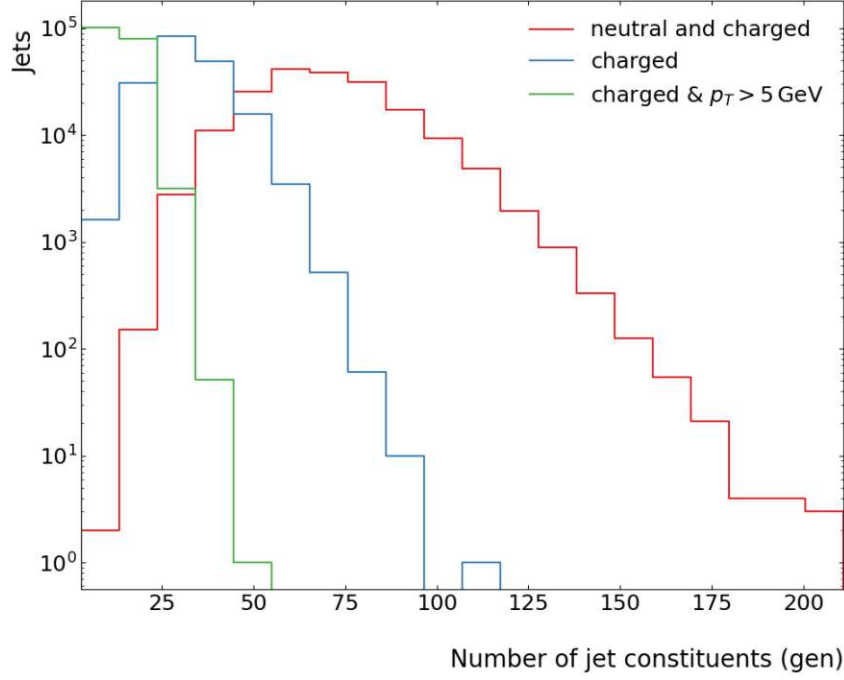


Figure 6.1: Comparison of the used jet constituents with different selections. The histograms displays the number of jets with a given number of jet constituents. The red curve displays the distribution using neutral and charged constituents. For the blue curve only charged particles were selected. For the green curve also only charged particles were selected and only jet constituents were considered with $p_T > 5$ GeV.

6.4 Breit-Wigner Mass Reweighting

The unfolding methods presented in this work aim to unfold measured data and use the unfolded data to determine the mass of the top quark. Therefore their ability to generalize needs to be tested with input data corresponding to distributions of varying top masses. Breit-Wigner mass reweighting is one way to obtain these test datasets. The Monte Carlo event generation produces $t\bar{t}$ events with the top mass following the Breit-Wigner distribution [8]

$$f_{\text{BW}}(e) = \frac{k}{(e^2 - m_{\text{top}}^2)^2 + m_{\text{top}}^2 \Gamma^2}, \quad (6.1)$$

where

$$k = \frac{2\sqrt{2} m_{\text{top}} \Gamma \gamma}{\pi \sqrt{m_{\text{top}}^2 + \gamma}}, \quad \gamma = \sqrt{m_{\text{top}}^2 (m_{\text{top}}^2 + \Gamma^2)},$$

e denotes the mass of one events top quark, $m_{top} = 172.5$ GeV is the MC-mass input of the generator, and $\Gamma = 1.3$ GeV is the decay width.

The constructed observables therefore also feature this underlying distribution. Since the generated top mass is accessible in the simulated data, this means one can change the underlying distribution and reweight each event to a new distribution. This is done by effectively applying a weight to the simulated events depending on their generated top mass. The weight can be obtained by dividing the new desired distribution by the generated distribution

$$w_{BW}(e) = \frac{(e^2 - m_{top}^2)^2 + m_{top}^2 \Gamma^2}{k} \cdot \frac{k'}{(e^2 - m_{top}'^2)^2 + m_{top}'^2 \Gamma^2}, \quad (6.2)$$

where m_{top}' and k' are the values of the desired distribution, a Breit-Wigner distribution with a new mass parameter m_{top}' .

In Fig. 6.2, results of this reweighting are depicted. The plots validates the reweighting, showing the shift of the mass peaks in the simulated top quark mass distributions.

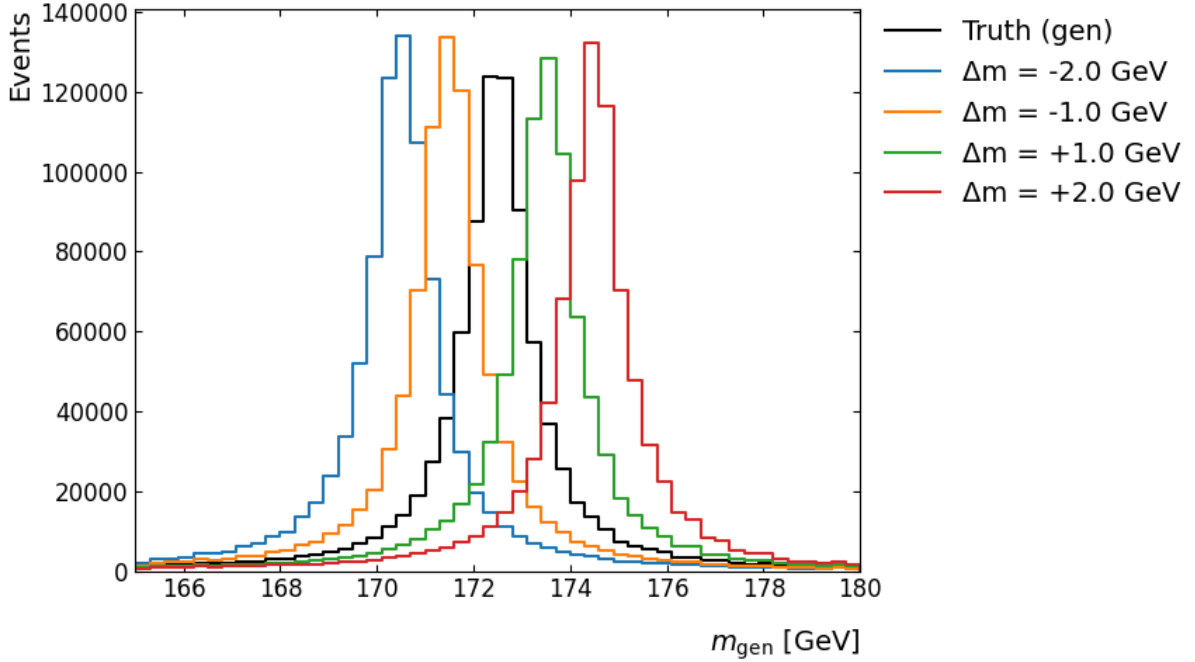


Figure 6.2: Breit Wigner reweighted distributions of the generated top quark mass. The original distribution is displayed in black, while the shifted distributions are color coded. They are reweighted to masses of 170.5 GeV, 171.5 GeV, 173.5 GeV and, 174.5 GeV.

6.5 Constructing the Energy Correlator

The energy correlators were constructed based on the matched particles. To calculate the three point energy correlator ζ , all combinations of three particles are considered. For each triplet the pairwise angular separations

$$\zeta_{ij} = (\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2$$

are calculated. These ζ_{ij} are then sorted ascending to $\zeta_{\text{short}} \leq \zeta_{\text{medium}} \leq \zeta_{\text{long}}$, and are transformed to

$$\zeta_x = \frac{\zeta_{\text{medium}} + \zeta_{\text{long}}}{2}, \quad \zeta_y = \zeta_{\text{long}} - \zeta_{\text{medium}} \quad \text{and} \quad \zeta_{\text{short}}. \quad (6.3)$$

The triplets are filtered with respect to their short side and asymmetry

$$\zeta_{\text{short}} > 0.1, \quad \zeta_{\text{long}} - \zeta_{\text{medium}} < \left(\frac{172.5}{p_{T,\text{jet}}} \right)^2. \quad (6.4)$$

These conditions are applied to select triplets which form equilateral triangles, which are required for the top mass sensitive peak to arise [7, 8].

To observe the EEEEC observable described in 2.22, the correlator weight is yet to be calculated. For the exponent $n = 1$ it is

$$w = \frac{p_{T,1} \cdot p_{T,2} \cdot p_{T,3}}{(p_{T,\text{jet}})^3}. \quad (6.5)$$

As depicted in Fig. 6.3 (lower left) the distribution of the transverse momentum $p_{T,\text{jet}}$ of the selected top quark falls exponentially. The lower boundary of $p_{T,\text{jet}} > 400$ GeV is in use to maintain the boosted regime. Thus, most of the jets have a $p_{T,\text{jet}}$ around 400 GeV. Since the peak position in the correlator distribution changes with $(m_t/p_{T,\text{jet}})^2$, the work of [8] bins the correlator distribution in $p_{T,\text{jet}}$ and studies a bin close to the threshold of 400 GeV. In this work a different approach was chosen. The peaks dependency on $p_{T,\text{jet}}$ was compensated by multiplying each ζ with the according $p_{T,\text{jet}}^2$ and dividing by the constant 172.5^2 , to work in the same magnitude of values. This introduces a new observable $\zeta \rightarrow \zeta' = \zeta \cdot \left(\frac{p_{T,\text{jet}}}{172.5} \right)^2$ removing the dependency on $p_{T,\text{jet}}$. The peaks of different p_T are aggregated to one single non-smeared peak. This peak is shown in Fig. 6.3 (lower right). In each of the plots in Fig. 6.3, the four color coded distributions differ in the applied selection. There are three variants of the minimal jets constituents p_T and an

additional variant, which also features a minimal correlator weight selection introduced in Sec. 6.5. Figure 6.3 (top right) depicts the distributions of ζ' . One can see the impact of the different selections on the total number of events. The shape of the distributions with the lower filter for the constituents p_T (0 GeV and 1 GeV), differs from the other two selections ($p_T > 5$ GeV, with and without weight filter), especially right to the peak. However, the shape of the weighted ζ' , due to the correlator weight, does not feature this difference right to the peak.

Correlator Selections Implied by the construction of the correlator weight in Eq. 6.5, the filter on the jet constituents minimal traversal momentum is effectively a filter on the correlator weight. This is evident by the reduction of triplets with a low correlator weight in the first bins in the correlator weight subplot in Fig. 6.3. To further reduce the high number of low weight triplets, the correlator weights were explicitly required to have a weight $w_{\text{corr}} > 1.5 \cdot 10^{-5}$. This is visible by the sharp edge in the weight subplot, Fig. 6.3, of the according selection. The lower right subplot depicts the constructed weighted correlator ζ' . With the chosen selection of the triplets asymmetry and the usage of ζ' , a clear peak is visible. This plot also shows, as already observed in [8], that the chosen selection on the correlator weight, does not shift the peak, but in fact sharpens it. Since the peak is located as about $\zeta' = 3.5$, and has fallen at about $\zeta' = 10$, another filter was introduced to the training data leaving the peak unchanged but reducing the amount and the span of the triplets in regards to ζ' . Only triplets were used with $\zeta' < 20$ on both the detector and particle level.

The observed peak was further investigated as presented in Fig. 6.4. In this plot, the share of the weights in each bin of the peak are displayed. There is no visible shift of weights within the bins of ζ' , meaning the share of weights appears to be the same in each bin. Furthermore from this plot it is evident, that small weights, in total only have a small contribution to the bins, thus it is beneficial to focus on larger weights in the training.

6.6 Unfolding with Conditional invertible Neural Networks (CiNN)

6.6.1 Model Training

As mentioned in the previous chapters, the input data to the training are ζ' , $p_{T, \text{jet}}$, and the correlator weight w of the selected events and triplets on both particle level and detector level. The unfolding models therefore have three inputs and three conditions. As usual

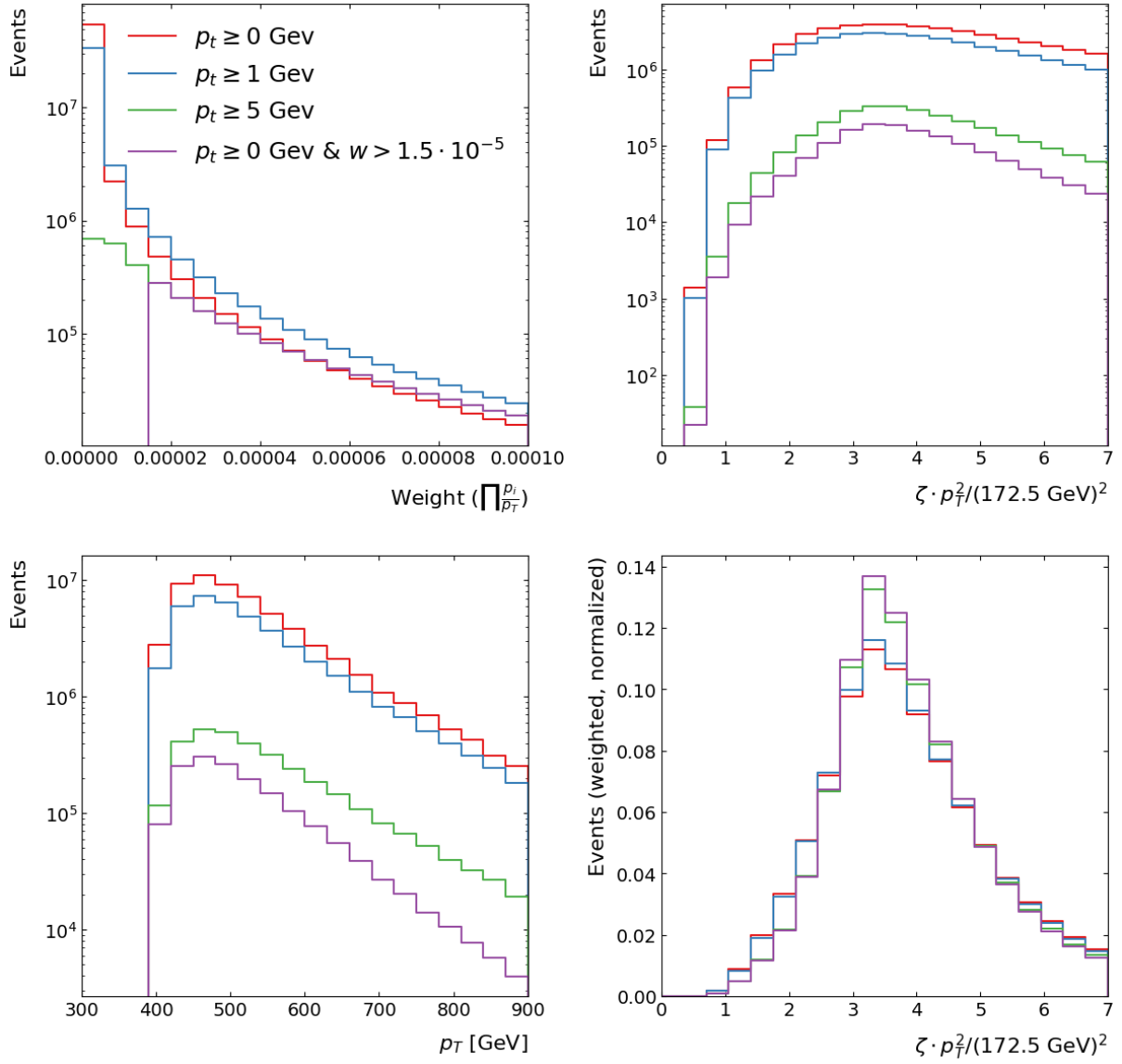


Figure 6.3: Comparison on the particle level for different selections of the minimal constituent p_T and minimal correlator weight in the correlator weight (upper left), ζ' in log scale (upper right), jet p_T (lower left), and weighted ζ' weighted (lower right). For the p_T selections of 0 and 1 GeV the total number of triplets was limited to 10^8 .

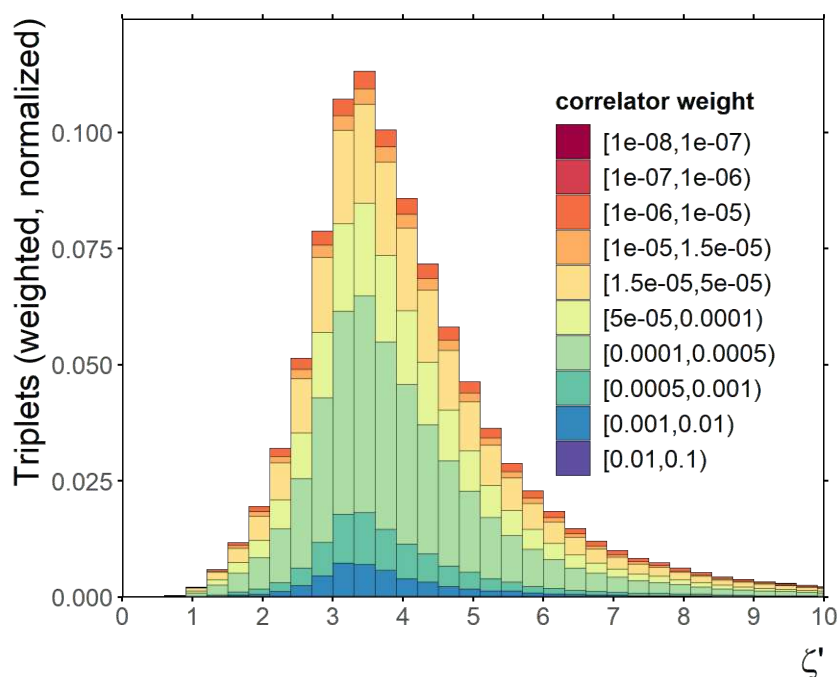


Figure 6.4: Peak in the histogram of ζ' (weighted) on particle level, showing the shares of correlator weights in each bin of ζ' by color encoding. In each bin, red sections showcase the contributions of small correlator weight, while blue sections showcase the contributions of large correlator weights.

in ML applications, these values are transformed, as the large range of magnitude in the inputs may lead to numerical problems. The values are normalized in the first step before a logit transformation and a standardization are applied. These transformed values are then fed batch wise epoch per epoch to the model to train it. The batch wise training reduces the required memory as it gets successively optimized with a part of the training data. The loss function of the training and validation data set were monitored to indicate the status and progress of the training.

In addition to the dataset simulated for $m_t^{MC} = 172.5$ GeV, two more datasets were used during the training. The additional data was constructed with datasets simulated for $m_t^{MC} = 171.5$ GeV and $m_t^{MC} = 173.5$ GeV and underwent the same selections as described in Sec. 6.3. The training data consists of equal shares of the three datasets. The combination of samples with different m_t already gives the models the information how to unfold events with different top quark masses.

Of the resulting data, 500 000 triplets were used in the training.

The computational results presented were obtained using the CLIP cluster⁶.

Optimizer and Learning Rate Scheduler The setup of the optimizer (and Learning Rate Scheduler) is important to reach convergence of a model. A variety of different combination were tested. The Adam and AdamW optimizers were used with a constant learning rate, which behaved poorly in terms of the found minimum and convergence in general, as oscillations in the loss and sampled results were present. To mitigate these problems, successively more complex setups were used. Of the used learning rate setups described in Sec. 4, Cosine Annealing gave the best results, however still lacked a stable convergence. Finally the scheduler free AdamW optimizer led to a stable convergence, rendering further optimization of the learning rate obsolete.

Table 6.1 lists, the hyperparameter of the optimizer which were found suitable for the training.

Model Optimization With the setup of the optimizer mentioned above, the hyperparameters of the model were varied to investigate their impact on the training and model performance. The three varied hyperparameters of the model are the number of transformations in the normalizing flow, the number of hidden layers in the NN modeling of the transformations, and the number of nodes in each of the hidden layers. The used hyperparameters are listed in Table 6.2. The loss as a function of the epoch number is

⁶<https://clip.science>

| Parameter | Value |
|------------------------|--------------------|
| Starting Learning Rate | 1×10^{-3} |
| Weight Decay | 0.01 |
| Warmup Steps | 1000 |
| Weight LR Power | 1.0 |
| Betas | (0.9, 0.999) |
| Epsilon (ϵ) | 1×10^{-8} |

Table 6.1: Optimizer hyperparameters using the Schedule-Free AdamW algorithm.

displayed in Fig. 6.5. In the plot, four different model configurations are displayed. The most complex of which (yellow, 3 transformations, 4 hidden layers, 256 nodes per hidden layer), shows clear symptoms of an overfitted model, as the validation loss (dashed) reaches its minimum at around 50 epochs, before it strongly rises again indicating insufficient modeling performance on the validation dataset. On the other hand, the least complex model of the displayed (red, 6 transformations, 4 hidden layers, 64 nodes per hidden layer) shows signs of a too simple model and underfitting, as the training loss is clearly higher as for the other models. The same can be said about the curve in blue reaching convergence at about 170 epochs. The green loss (6 transformations, 4 hidden layers, 128 nodes per layer), also features slight overfitting, but since the rising of the validation loss is significantly smaller, this model is a promising candidate for a generalizing model. The models with 4 transformations, 3 hidden layers and 64/128 nodes per hidden layer interestingly, could not be brought to convergence.

| Transformations | Nodes | Hidden Layers | Parameters |
|-----------------|-------|---------------|------------|
| 3 | 64 | 2 | 54,162 |
| 3 | 64 | 4 | 105,618 |
| 6 | 64 | 4 | 108,324 |
| 6 | 64 | 4 | 211,236 |
| 3 | 128 | 2 | 206,610 |
| 3 | 128 | 4 | 407,826 |
| 6 | 128 | 2 | 413,220 |
| 6 | 128 | 4 | 815,652 |
| 3 | 256 | 2 | 806,418 |

Table 6.2: CiNN hyperparameter variation.

Early stopping was employed to the training, reducing the training time, resources, and unnecessary training bias. This can be seen in Fig 6.5, the comparison of the loss function of models of different hyperparameters. For a difference in the loss from one epoch to next, smaller than 10^{-5} the training stops.

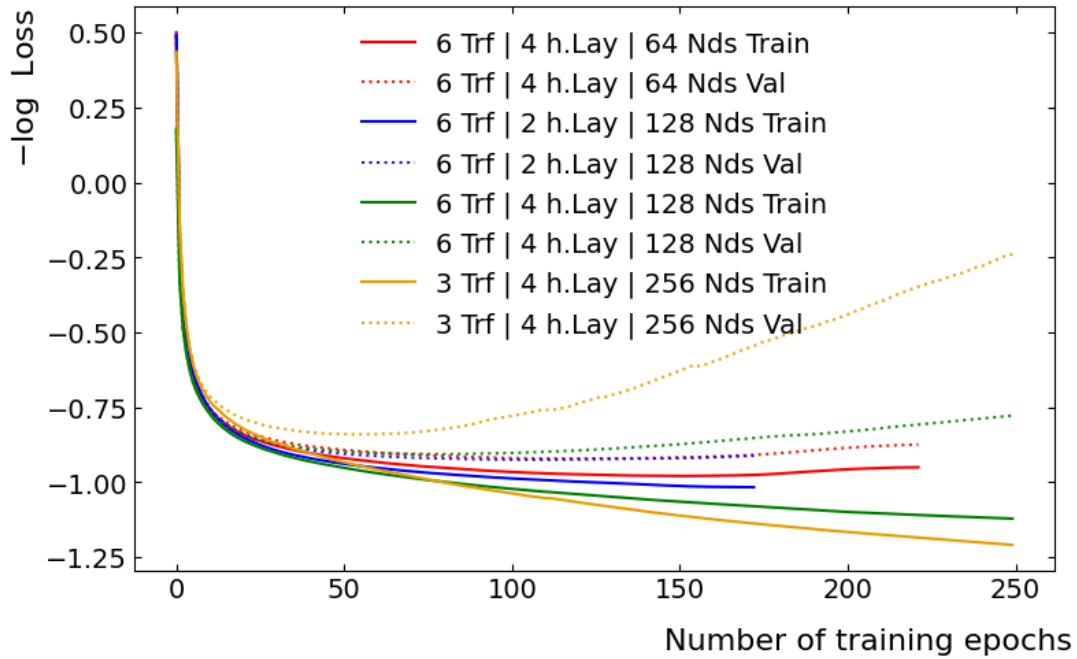


Figure 6.5: Losses for different CiNN setups. The solid lines show the loss of the training data, while the dotted lines display the loss of the validation dataset.

Loss Weighting As it can be seen in Fig. 6.3 (top left) the logarithmic histogram of the correlator weight w decreases approximately linearly with the correlator weights. In the loss function, this exponential decay, favors the fit of low weights, by orders of magnitude. This may lead to poorly fitted high weights, which significantly impact the weighted . This can be counteracted by introducing a triplet wise weight in the loss function. The loss function is proportional to the sum of the single triplet loss terms l_i . The loss weight $a(w_i)$ is multiplied to each of these terms based on their correlator weight, turning this sum into a weighted sum

$$\sum_i l_i \rightarrow \sum_i l_i a_i(w).$$

The distribution of the correlator weights follows

$$N(w) \propto e^{-k \cdot w}.$$

With the loss weight in place, this impacts the loss with

$$N_{loss}(w) \propto e^{-k \cdot w} \cdot a(w),$$

where $a(w)$ is defined in a manner that leaves N_{loss} as constant as possible.

This method was applied to the training of the CiNN models. However, it led to problems regarding high correlator weight terms, evident by outliers in the distribution of the weighted correlators ζ' . To avoid the dominance of triplets with high correlator weights and subsequent high loss weights, an upper limit for the loss weight was applied and successively reduced when it did not show the expected result. Finally this method was not introduced to the training, since it did not improve the results and was obsolete as more suitable hyperparameters, optimizer and learning rate schedulers were used.

6.6.2 Unfolding Performance

A disjoint validation dataset of the same selection parameters as the training dataset with a MC generator mass of $m_t^{MC} = 172.5$ GeV was unfolded using the model. This means, the detector level observables were used as the condition while sampling from the generative model. Meaning, for each triplet in the validation dataset unfolded observables were generated. Since the dataset also includes the according particle level observables, they can be compared to the unfolded results. In Fig. 6.6 this comparison is depicted, while the data is also shown on the detector level. The three subplots displays the three unfolded variables. The constructed observable ζ' weighted, is depicted in Fig. 6.7. As this method is unbinned, the binning in these plots merely helps visualization and comparison. Hence, the binning is arbitrary and does not depend on the unfolding. Most importantly, the unfolded data broadly follows the truth particle data for ζ' in Fig. 6.6 (upper right), the correlator weight in Fig. 6.6 (upper left) and jet p_T in Fig. 6.6 (lower left). This is a first indicator for a working unfolding. Furthermore, the construction of the weighted ζ' in Fig. 6.7 on the unfolded domain also leads to a peak very similar to the truth level peak. One can see in the upper left subplot if Fig. 6.6, that the weights can be successfully replicated over at least three magnitudes, before deviations are visible. The same is true for $p_{T, \text{jet}}$. However the unfolded distribution features a residue in the bin below 400 GeV, which can be understood as a result of the transformations limitation

on the sharp edge of the $p_{T, \text{jet}}$ distribution introduced through the p_T cut. For ζ' the behavior in the tails is not relevant, as the peak in the weighted correlator is located around the value 3.5 and the difference to the weighted distribution only is the weight of the single correlator values in their bin. While the statistical Poisson uncertainty of a histogram bin i with N_i events in the bin is $\sigma_i = \sqrt{N_i}$, the error bars in this plot show the statistical uncertainty of the weighted histogram. For a bin i with the content $n_i = \sum_{k \in \text{Bin } i} w_k$, the statistical uncertainty is

$$\sigma_i = \sqrt{\sum_{k \in \text{Bin } i} w_k^2}. \quad (6.6)$$

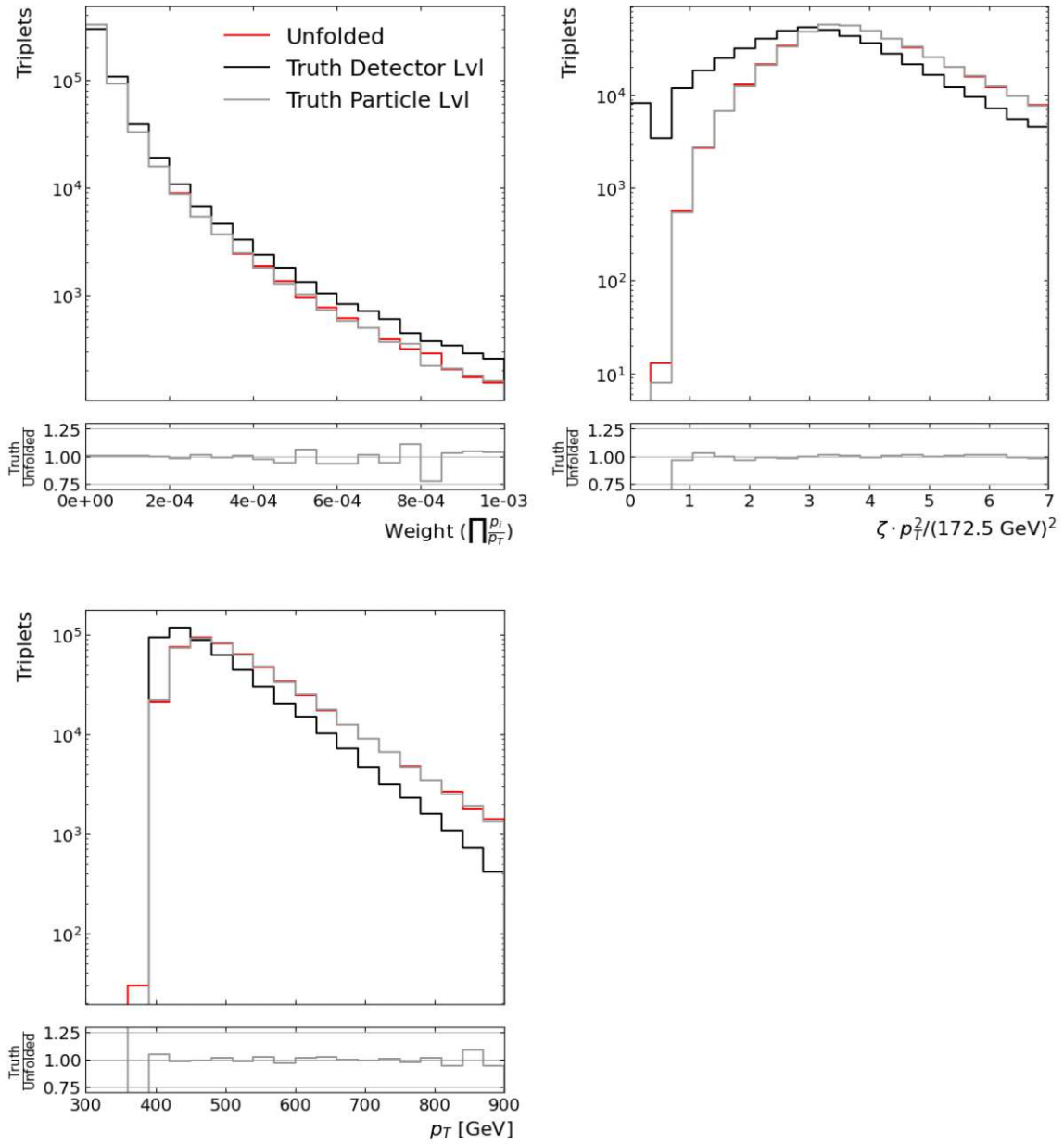


Figure 6.6: Comparison of the detector level (black), particle level (grey), and unfolded (red) distributions in the correlator weight (upper left), ζ' in log scale and not weighted (upper right), jet p_T (lower left), and weighted ζ' weighted (lower right). The unfolding was performed using a CiNN model with 6 transformations, 2 hidden layers, and 128 hidden nodes.

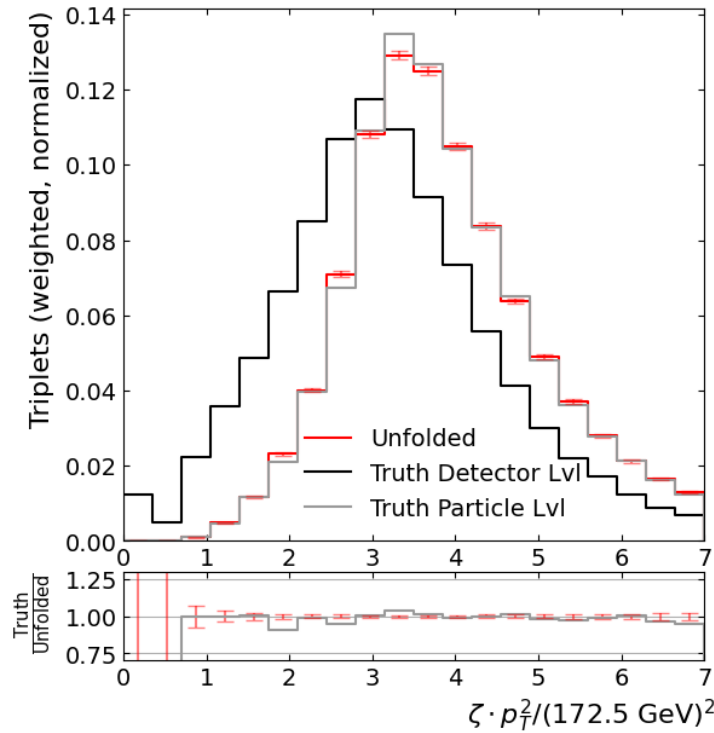


Figure 6.7: Comparison of the detector level (black), particle level (grey), and unfolded (red) distributions in the weighted ζ' (lower right). The unfolding was performed using a CiNN model with 6 transformations, 2 hidden layers, and 128 hidden nodes.

For most bins in Fig. 6.6 (lower right), the ratio ratio of the true and unfolded distribution lies within the statistical uncertainties of the unfolded values and, in general, has a flat profile. This indicates that the model is able to unfold the weighted peak in ζ' at detector level with high fidelity to the corresponding peak at particle level. Given the clear difference between the distributions at particle level (grey in Fig. 6.6, lower right) and detector level (black), this represents a successful correction of the detector effects.

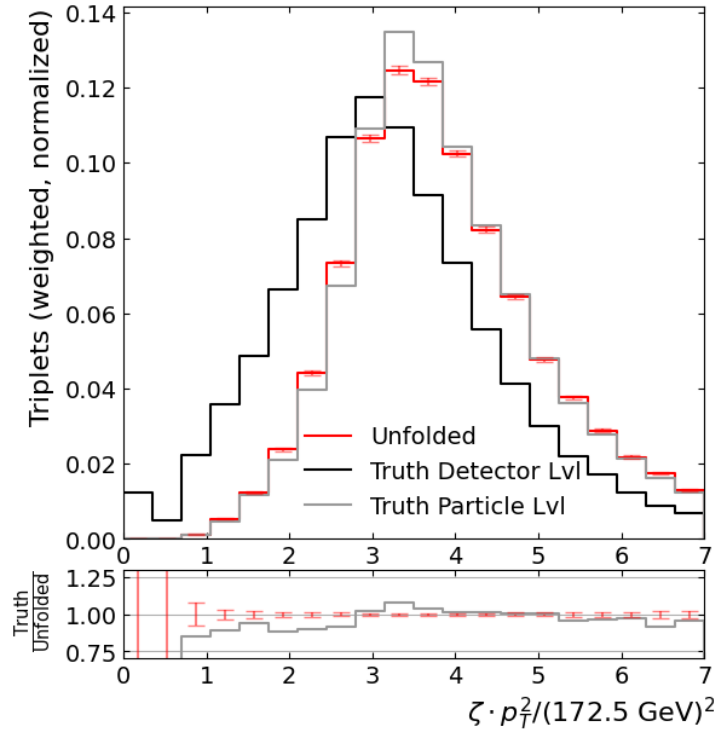


Figure 6.8: Comparison of the detector level (black), particle level (grey), and unfolded (red) distributions in the weighted ζ' (lower right). The unfolding was performed using a CiNN model with 3 transformations, 2 hidden layers, and 64 hidden nodes.

In Fig. 6.8, the results of a model with lower expressivity is depicted. One can see, that the peak in ζ' depicted in Fig. 6.8 is modeled less accurately. This is also the case for the sharp edge in the p_T distribution, while the weight is still modeled well, for multiple magnitudes.

6.6.3 Generalization and Bias

Data of different underlying mass distributions were unfolded, to explore the models generalization capabilities. These tests validate, if the unfolded data still has the structure of a top mass dependent peak. Figure 6.9 depicts the unfolding results obtained by the usage of Breit-Wigner reweighting. The unfolded particle level distributions, obtained using validation data with $m_t^{MC} = 172.5 \text{ GeV}$, were reweighted by the mass of the generated top quark. These reweighted unfolded results were compared to the true particle level of the validation datasets. A second validation method uses validation data of the datasets with $m_t^{MC} = 171.5 \text{ GeV}$ and $m_t^{MC} = 173.5 \text{ GeV}$. The results of this method are depicted in Fig. 6.10, which also compares the unfolded results to the truth for different m_t .

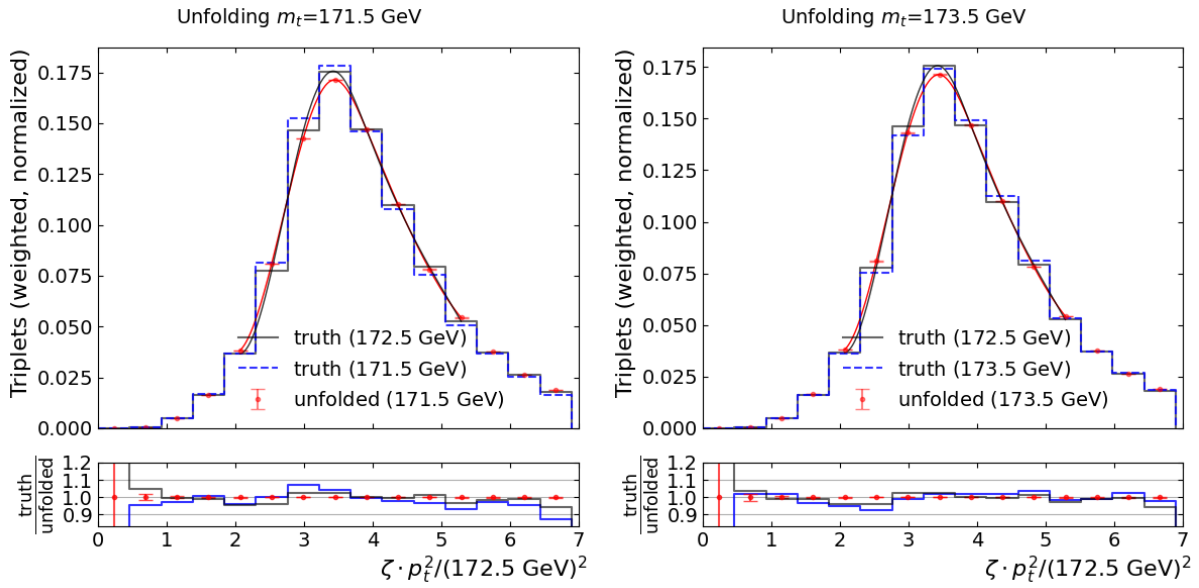


Figure 6.9: CiNN unfolding performance for reweighted data corresponding to a mass hypothesis of 171.5 GeV and 173.5 GeV, obtained by Breit-Wigner reweighting. In the plot, the peak in ζ' is depicted on particle level. The reweighted unfolded distribution (red) is compared to the truth for 172.5 GeV (grey) and the reweighted truth (blue). The unfolding was performed using a CiNN model with 3 transformations, 2 hidden layers, and 64 hidden nodes.

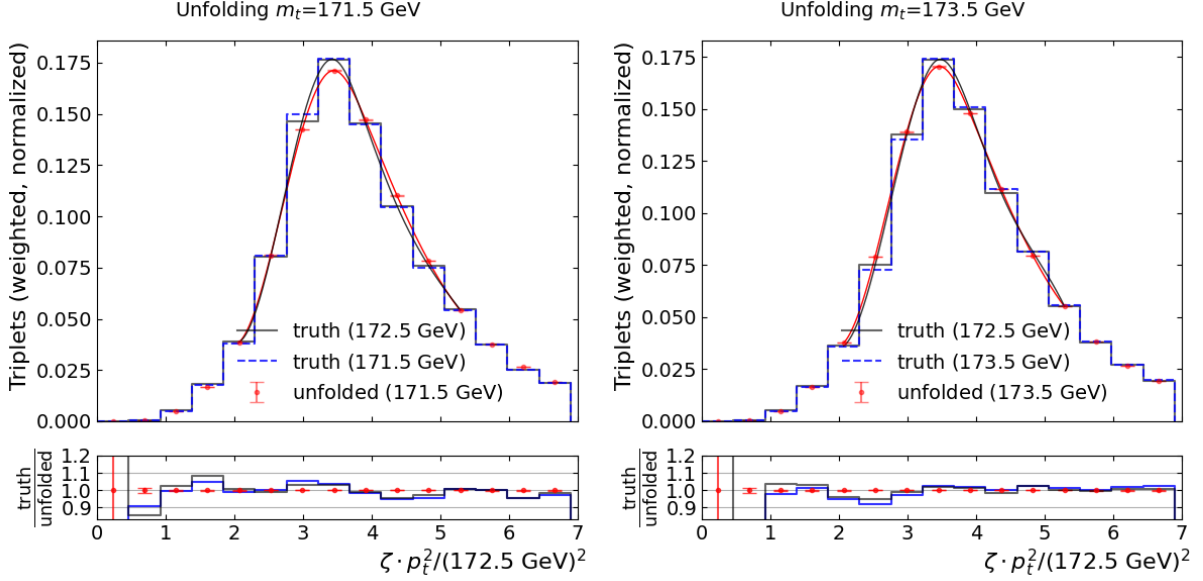


Figure 6.10: Unfolding performance for validation data originating from samples with $m_t^{MC} = 171.5$ GeV and 173.5 GeV. In the plot, the peak in ζ' is depicted on particle level. The plot shows the unfolded distribution (red) and the truth of the according dataset (blue). To identify the expected shift in regards to the training data, the particle level truth is depicted for $m_t^{MC} = 172.5$ GeV (grey). The unfolding was performed using a CiNN model with 3 transformations, 2 hidden layers, and 64 hidden nodes.

Both the unfolding and then reweighting (Fig. 6.9) and usage of varying datasets (Fig. 6.10), are difficult to interpret. Firstly it can be said, that statistical uncertainties are unlikely to impact these results, as the errors in the single sample plots are already small and the unfolded distribution was obtained via 100 samples. Secondly, the peak position only shifts to small amounts as a function of m_t . This is due to the construction of ζ' , which allows to use all triplets regardless of their $p_{T, \text{jet}}$ in the same unfolding model and enhance statistical precision but also smears the correlator peak. Nevertheless the markers of the unfolded values do hardly change for either of the presented cases. This is a clear indicator of a bias and weak generalization.

To further validate the generalization, a χ^2 fit was performed. Therefore the truth distribution was sequentially reweighted to distributions of different mass, and for each distribution the bin wise difference of the unfolded and reweighted truth histograms was calculated to obtain the test value χ^2 . These values were then plotted for the different reweighted truth distributions to display which distributions has the best fit. In other words, the unfolded results were reweighted and compared to distributions of varying

mass hypotheses. Where as in Eq. 6.6, the error is calculated assuming independent bins, this in fact is not the case, as small changes might lead to migration between bins introducing correlations. To handle these bin to bin correlations in the χ^2 -fit, for each detector level triplets 100 particle level triplets were unfolded, effectively sampling 100 times. This allows for the handling of correlations via the covariance matrix

$$c_{ij} = \frac{1}{N_{\text{samples}}} \sum_{n=1}^{N_{\text{samples}}} (u_i^{(n)} - \bar{u}_i) (u_j^{(n)} - \bar{u}_j), \quad (6.7)$$

where the sum runs over the $N_{\text{samples}} = 100$ unfolded samples, u_i^n is the bin content of the i^{th} bin and \bar{u}_i is the mean content of the i^{th} bin. With this matrix c one can calculate χ^2 with

$$\chi_m^2 = (\mathbf{u} - \mathbf{t}_m)^\top \mathbf{c}^{-1} (\mathbf{u} - \mathbf{t}_m),$$

where \mathbf{t}_m , is the truth distribution reweighted to the mass m . To obtain χ^2 , only bins, close to the peak were chosen, since their content is most sensitive to the top quark mass.

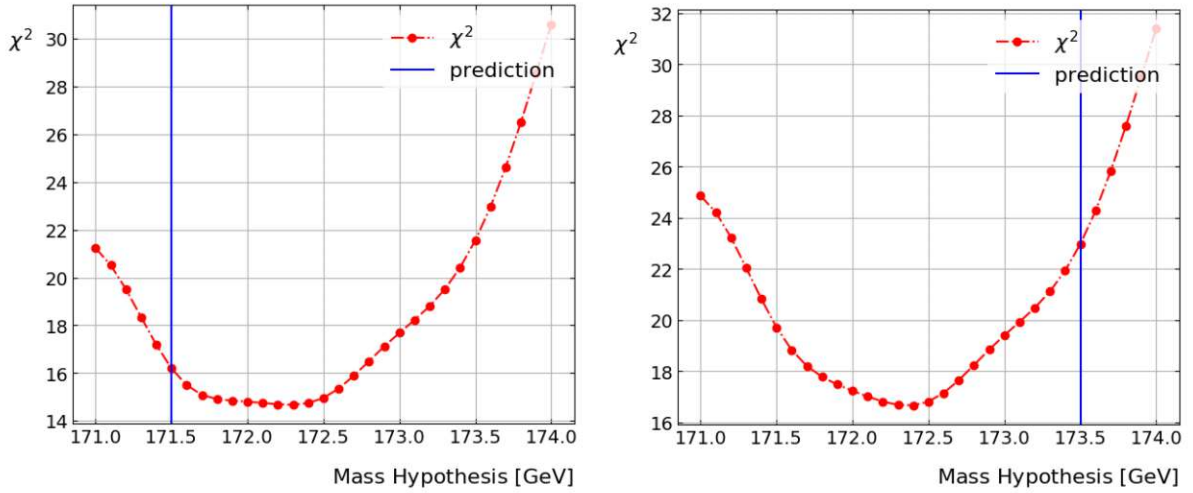


Figure 6.11: χ^2 fit using a CiNN model (6 transformations | 4 hidden layers per NN | 128 nodes per hidden layer). The unfolded distribution was sampled 100 times and reweighted to a distribution of the mass marked *prediction* (blue).

The minimum of the curves in Fig. 6.11 does not deviate significantly from 172.5 GeV. Nevertheless, the plot illustrates that the bias problem is difficult to interpret, as the prediction remains within the expected statistical uncertainty.

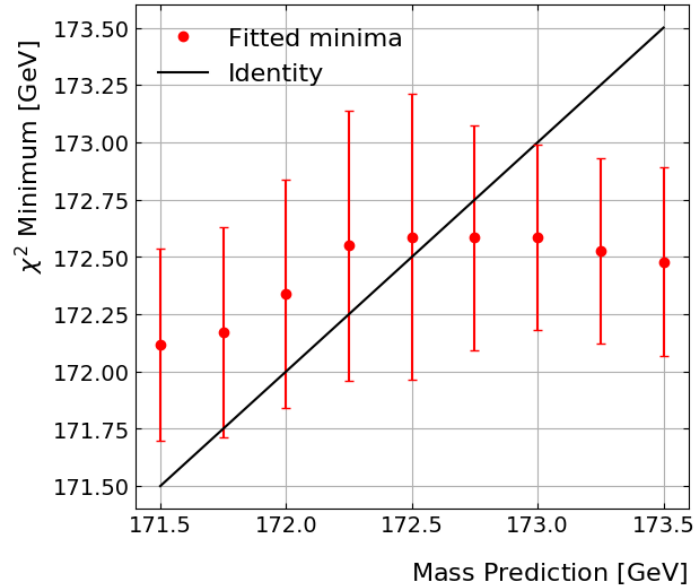


Figure 6.12: Linearity test of the fitted minimum in χ^2 test. The markers (red) display the minimum for the quadratic fit of χ^2 test for a given prediction. The depicted uncertainty is obtained from the curvature of the according quadratic fit. The diagonal line (black) depicts the optimal result for this test. The unfolding was performed using a CiNN model (6 transformations | 4 hidden layers per NN | 128 nodes per hidden layer).

The unfolding of the sample with $m_t = 171.5$ GeV leads to a minimum consistent with its true value within statistical uncertainties. In the unfolding of a sample with $m_t = 173.5$ GeV a clear bias towards the central value of $m_t = 172.5$ GeV is visible. However, these studies have to be interpreted with caution since only statistical uncertainties from the unfolding are considered and missing uncertainties and their bin-to-bin correlation can significantly alter the χ^2 values.

The χ^2 fit was performed for multiple predictions. This enables the test of the linearity of the minimum of the χ^2 . The minimum was obtained employing the fit of a parabola to the minimum of the χ^2 curve. The minimum of this fit for a given prediction is plotted in Fig. 6.12 (red). The uncertainties are obtained from the width of the fitted parabola for $\chi^2 = \chi_{\min}^2 + 1$. The plot shows results similar to the results depicted in Fig. 6.11. For mass predictions $m_{\text{pred}} < 172.5$ GeV, a linear behavior and thus sensitivity to the top mass is evident. For higher mass predictions the top quark mass bias can be observed.

6.7 Unfolding with Conditional Flow Matching (CFM)

For the training of the CFM model, the same data sets were used. The CFM was configured for the three inputs and three conditions, before the batch wise and epoch wise training performed. After the training, similar to the CiNN unfolding, the unfolded distributions are obtained by sampling from the latent space and performing the backward transformation under the given detector level data. This transformation requires the solution of the differential equation Eq. 4.9, which here was performed by the `pytorch` [81] function `odeint`, which internally uses the Dormand-Prince method ⁷.

6.7.1 Model Training

As proven reliably, the optimizer and optimizer hyperparameters described in Sec. 6.6.1 were also used for the CFM. However, different hyperparameters were in use for a parameter sweep and model optimization. They are listed in Tab. 6.3.

| Hidden Layers | Nodes | Parameters |
|---------------|-------|------------|
| 3 | 64 | 13,187 |
| 6 | 64 | 25,667 |
| 3 | 128 | 50,947 |
| 6 | 128 | 100,483 |
| 3 | 256 | 200,195 |
| 6 | 256 | 397,571 |

Table 6.3: CFM hyperparameter variation.

⁷The Dormand-Prince, or RK5 is a Runge Kutta family method for the stepwise numerical solution of differential equations, which uses multiple function calls to obtain the next time step.

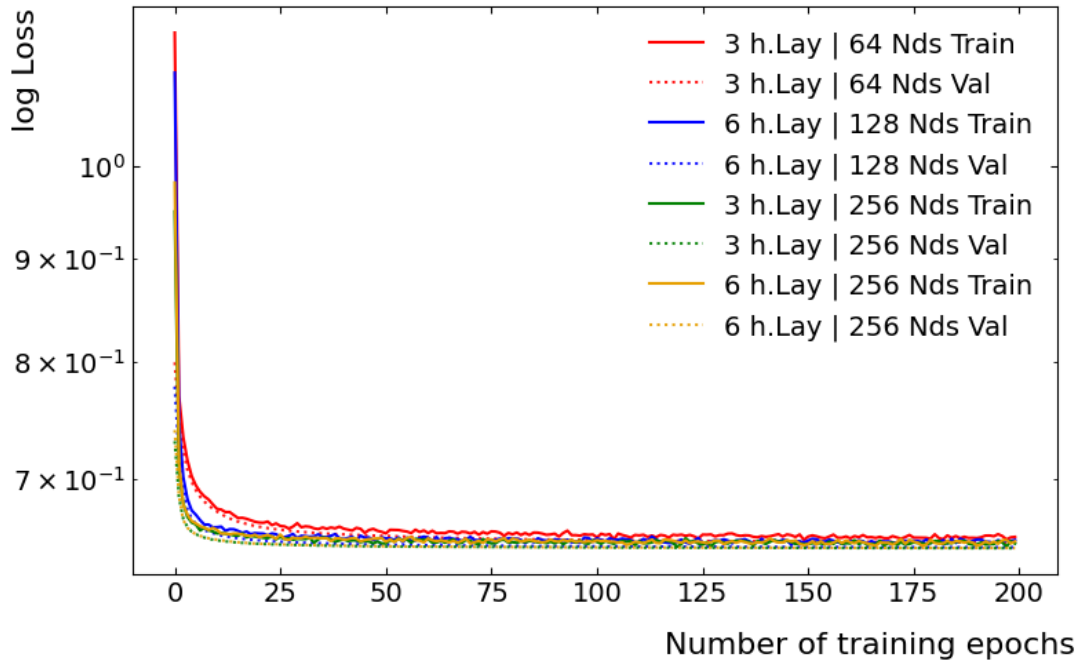


Figure 6.13: Losses of different CFM Model. The solid line shows the loss of the training data, while the dotted line displays the loss of the validation dataset.

The quality of the training with different hyperparameters is discussed with the help of Fig. 6.13, displaying loss curves of different model configurations. The loss values can however not be directly compared to those of the CiNN training in Fig. 6.5, since the loss is constructed differently. Surprisingly, the loss functions do not change significantly with the used hyperparameters as it was the case for the CiNN models. However, one has to keep in mind, that the loss function Eq. 4.12 for CFM, does not describe the likelihood of a compatibility of the model prediction and true values, but the fit of the vector field performing the transformation.

6.7.2 Unfolding Performance

Figure 6.15 depicts the unfolded observables in comparison to the particle level truth. It shows a successful unfolding, as the unfolded distribution widely matches the particle level truth in all the three unfolded observable. The constructed observable ζ' weighted, is depicted in Fig. 6.14, and importantly also indicates a successful unfolding. This model also shows good capabilities of reconstructing the particle level distributions, over multiple magnitudes in the histograms. Even the sharp edges of the p_T selections in 6.15 (lower left), are shaped immaculately. In contrast to the shape of the losses however, the

unfolding results do differ with varying hyperparameters, as displayed in Fig. 6.16, where the reconstruction of the peak, is deficient, evident by the deviation from the unfolded distribution to the truth particle level distribution for $\zeta' \approx 3 - 4$.

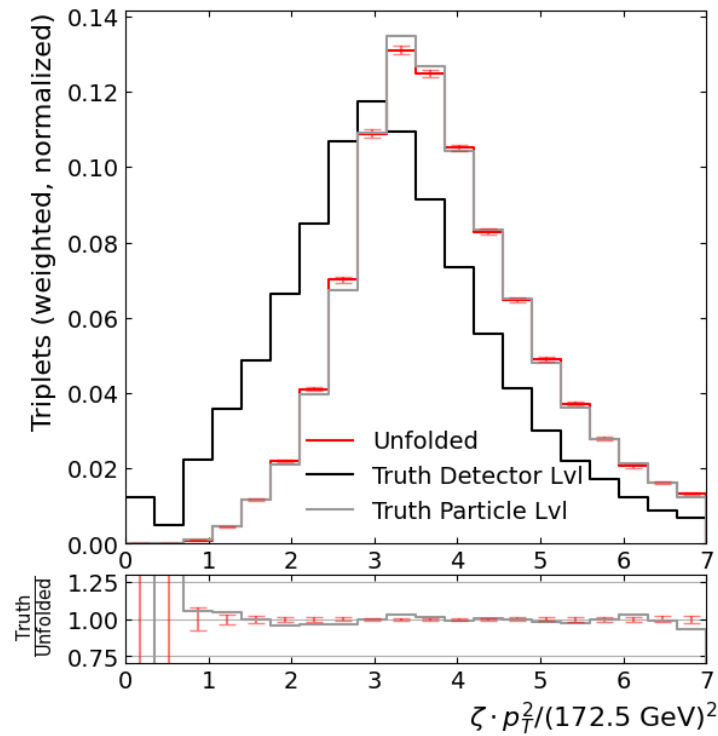


Figure 6.14: Comparison of the detector level (black), particle level (grey), and unfolded (red) distributions in the weighted ζ' (lower right). The unfolding was performed using a CFM model with 6 hidden layers and 256 nodes per hidden layer.

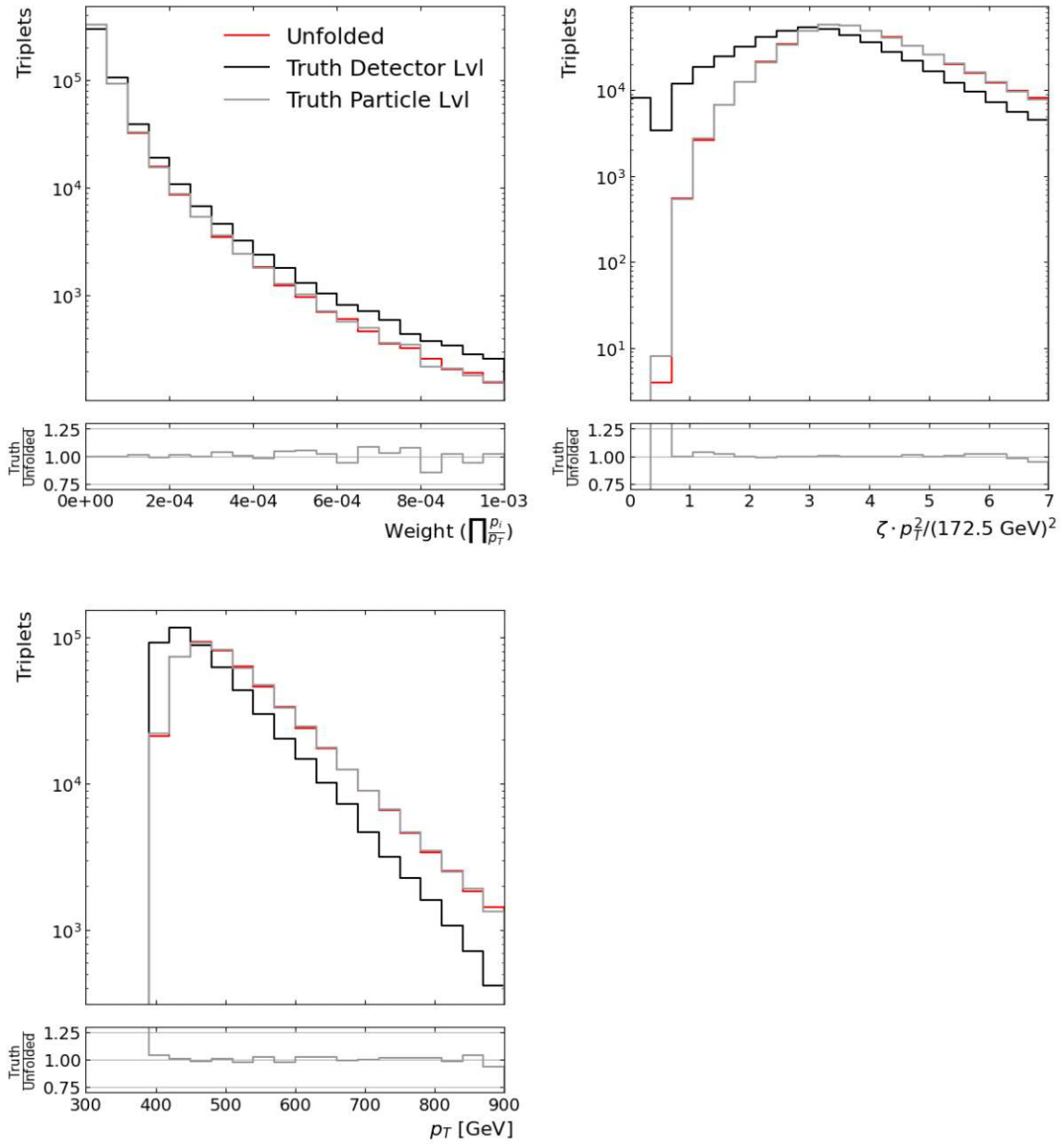


Figure 6.15: Comparison of the detector level (black), particle level (grey), and unfolded (red) distributions in the correlator weight (upper left), ζ' in log scale and not weighted (upper right), jet p_T (lower left), and weighted ζ' (lower right). The unfolding was performed using a CFM model with 6 hidden layers and 256 nodes per hidden layer.

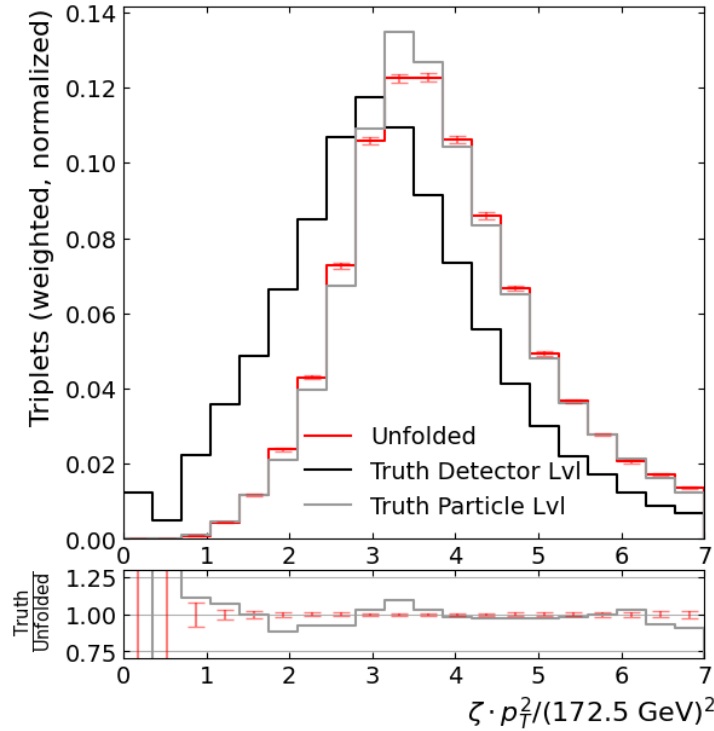


Figure 6.16: Comparison of the detector level (black), particle level (grey), and unfolded (red) distributions in the weighted ζ' (lower right). The unfolding was performed using a CFM model with 3 hidden layers and 64 nodes per hidden layer.

6.7.3 Generalization and Bias

Similar to Sec. 6.6.3, the bias towards the top quark mass of 172.5 GeV is investigated for the CFM unfolding using a χ^2 method. The following plots were also obtained using reweighted and simulated data of varying mass distributions. The χ^2 fit follows the same principles as explained above, however due to the resource consuming unfolding, the data was unfolded 50 times, for the calculation of the covariance matrix.

The generalization performance for the CFM model is depicted in Fig. 6.17. The plots show the unfolded distribution for validation data of samples using $m_t^{MC} = 171.5 \text{ GeV}$ and $m_t^{MC} = 173.5 \text{ GeV}$. It mainly shows the same results as for the CiNN model, as slight shifts in the bins and the splines are visible.

The χ^2 values in Fig. 6.18 show a comparable picture as in the CiNN model. While the unfolding for a mass of 171.5 GeV (left) tends to unfold back to its truth, a clear bias is visible for unfolding the sample with $m_t = 173.5 \text{ GeV}$ (right). However, only the statistical uncertainties of sampling multiple times from the model are considered here.

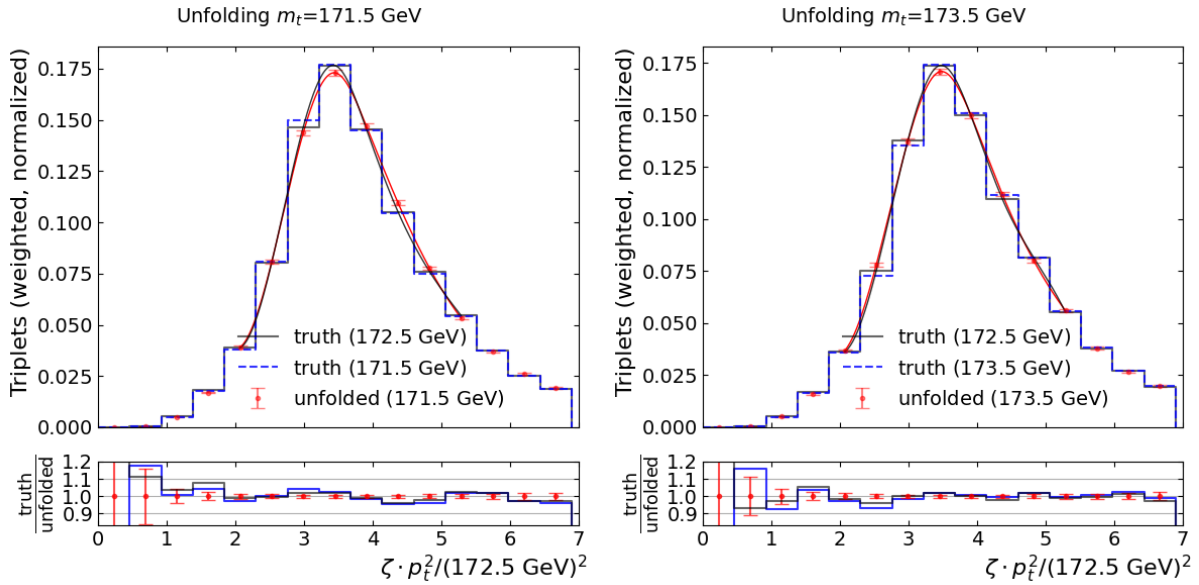


Figure 6.17: CFM unfolding performance for validation data originating from samples with $m_t^{MC} = 171.5$ GeV and 173.5 GeV. In the plot, the peak in ζ' is depicted on particle level. To identify the expected shift in regards to the training data, the particle level truth is also depicted for 172.5 GeV. The splines constructed through the unfolded distribution and the truth for $m_t^{MC} = 172.5$, serve comparison purposes.

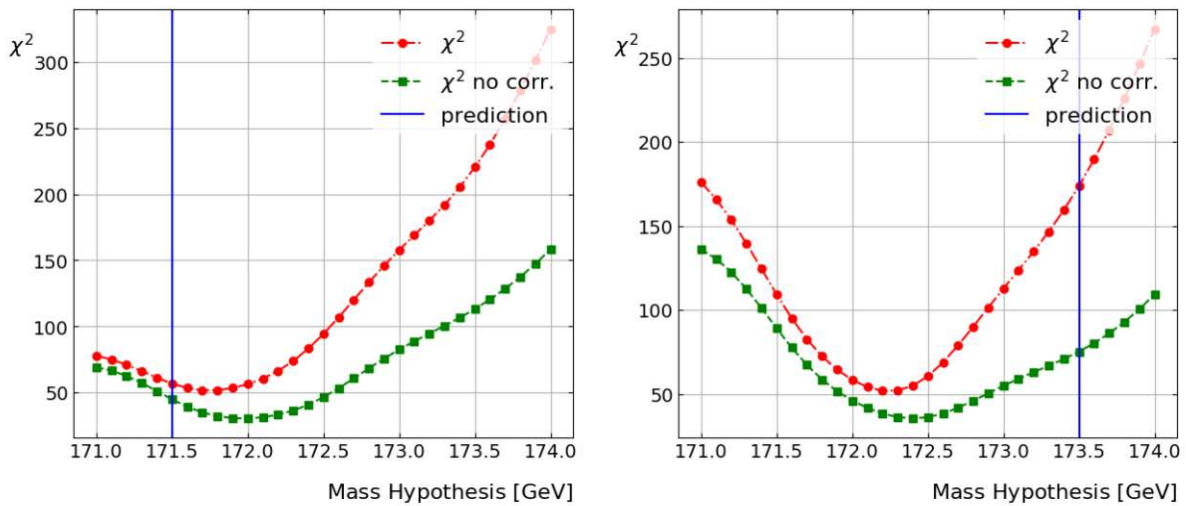


Figure 6.18: χ^2 fit using a CFM model (6 hidden layers 256 nodes per hidden layer). The unfolded distribution was sampled 50 times and reweighted to a distribution of the mass marked *prediction* (blue). χ^2 is depicted for the full covariance matrix (red), and without correlations using only, the diagonal of the covariance matrix (green).

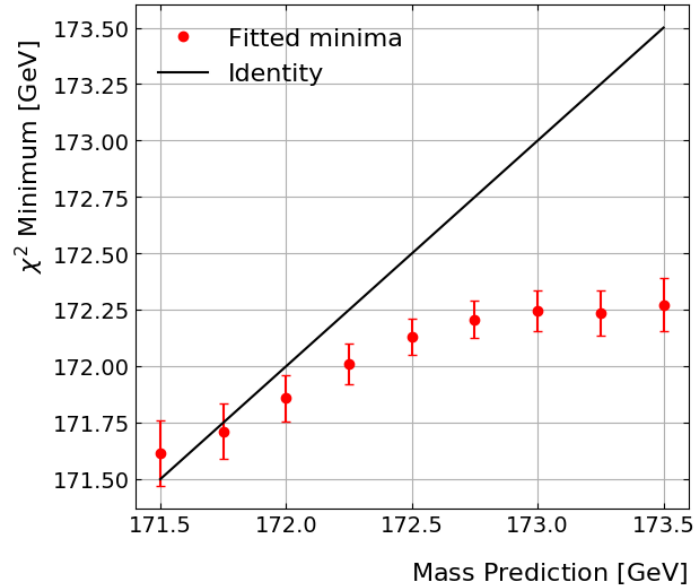


Figure 6.19: Linearity test of the fitted minimum in χ^2 test. The markers (red) display the minimum for the quadratic fit of χ^2 test for a given prediction. The depicted uncertainty is obtained from the curvature of the according quadratic fit. The diagonal line (black) depicts the optimal result for this test. The unfolding was performed using a CFM model (6 hidden layers 256 nodes per hidden layer).

The systematic uncertainties that arise during the unfolding with the trained models are unknown and could have a significant impact on the χ^2 fit. In order to test the influence of the bin-to-bin correlations, the χ^2 values are also calculated using a covariance matrix, where all off-diagonal values are set to 0. Overall, smaller χ^2 values are obtained but with the minimum in a similar position.

The linearity test of the χ^2 fit minimum was performed in accordance to Sec. 6.6.3 and using the full covariance matrix. Also similar results are obtained. For predictions $m_{\text{pred}} < 172.5$ GeV a linear behavior if the fitted minimum can be seen in Fig. 6.19. For higher predictions, a bias towards 172.25 GeV can be seen. These results are also in accordance with the results depicted in Fig. 6.18.

6.8 CiNN and CFM in Comparison

For both setups a model was found that generates unfolded distributions that agree with the particle level truth. Samples with a different top quark mass can be unfolded as well and agree with its true particle level distributions but a bias towards the mass parameter

used in the training could not be ruled out indefinitely. The numbers of free parameters in the ML models are in a similar range, of about 400 000 for the best unfolding models of both types.

A major difference between the two models from the machine learning point of view, is the fact, that the CFM consists of one NN in every case, whereas normalizing flows chain the output of multiple CiNNs. This has consequences for the training and its duration. The typical training duration for a CiNN-normalizing flow model was about 18 hours, and for CFM in the range of 2 hours. However this effect is inverse for the sampling from the models. While CFM requires the solution of integral equations for each single generated data row, the generation for CiNN only require the transformation via the autoregressive structure and usage of the NN. Sampling the same data took about 1 hour for the CiNN and 15 hours for the CFM. All these calculations were performed on the CLIP cluster on CPU nodes.

7 Conclusions

This thesis aimed to perform the novel investigation of the feasibility of unbinned unfolding of energy correlator observables using generative machine learning techniques. To observe the required energy correlators, simulated boosted $t\bar{t}$ lepton+jets events were used. The used observable is the weighted threefold energy correlator, which encodes the triplet wise angular separation of particles in the collimated boosted jet which form nearly equilateral configurations. The $p_{T, jet}$ sensitivity of the correlator peak, is mitigated with the use of $\zeta' = \zeta \cdot \left(\frac{p_{T, jet}}{172.5}\right)^2$. Since $p_{T, jet}$ still carries relevant information and is required to construct ζ' after the unfolding, it was included as part of the unfolding observables. Consequently, the unfolded variables were chosen as ζ' , the correlator weight w , and $p_{T, jet}$. To reduce the resulting data to a more manageable and meaningful dataset, selections were applied. Only jets of hadronically decaying top quarks with $p_{T, jet} > 400$ GeV were used. For each boosted jet, the triplets of jet constituents with $p_T > 5$ GeV and the resulting correlator weight of $w > 1.5 \cdot 10^{-5}$ were selected. The according observables were constructed on both particle level and detector level and matched to obtain triplet wise pairs.

Consequently the two generative ML methods using CiNN and CFM were trained with this data. For both models a setup for the architecture and model training were found. With the optimizer `AdamWSchedFree` both models could be trained to convergence. A hyperparameter scan led to insights regarding the necessary complexity. For the CiNN model 6 transformations, 2 hidden layers, and 128 hidden nodes were identified as suitable hyperparameters. For the CFM model 6 hidden layers and 256 nodes per hidden layer were identified as suitable hyperparameter.

After the training of the models, their capabilities were tested by unfolding the validation data obtained with the same selections but orthogonal to the training data. The detector-level observables served as the condition in the sampling of the unfolded particle-level distribution. The distributions of the three unfolded and the constructed observable ζ' weighted, follows the distribution of the truth data nearly identically for both methods. This marks a successful unfolding using $\zeta', w, p_{T, jet}$ and generative models. Especially given the shift in the mass sensitive peak on the detector level relative to the particle level.

Further investigations of the generalization and the bias of the model were conducted. Therefore the unfolding was applied to samples with a deviating top quark mass. These samples were obtained from simulated samples with a MC mass of $m_t^{MC} = 173.5$ GeV and $m_t^{MC} = 171.5$ GeV and also with the help of Breit-Wigner reweighting of the unfolded

data. These investigations, indicated a bias towards $m_t = 172.5$ GeV and a poor top quark mass sensitivity of the unfolded ζ' peak.

In this work, 500 000 triplet data pairs were employed for model training and validation. The CiNN model completed training within a few hours allowed for fast sampling once trained. The CFM model exhibited shorter training times but significantly longer sampling durations, as each generated sample requires numerical integration along the flow trajectory. This computational overhead limits the practical scalability of CFM for large datasets, whereas the CiNN proved more efficient for iterative unfolding and parameter studies.

8 Outlook

In this work, two unfolding methods have been established that show promising performance. Proven models of this type will ultimately enable the unfolding of measured correlator observables, allowing direct comparison and fitting to analytically obtained distributions. This makes it possible to fully exploit the advantages of correlators in the interpretation of the extracted top quark mass.

The advantages of the unbinned methods presented here could also be applied to other kinds of measurements. In particular, cross section measurements may benefit from such approaches, as they often suffer from binning and the associated regularization and bias problems.

The main limitation identified for the use as an unfolding technique is the residual bias towards the training data, which requires further study. A possible solution is the use of extended training data. In this work, only simulations with mass parameters of 171.5, 172.5, and 173.5 GeV were employed, leading to a dominance of the 172.5 GeV sample. A flatter mass distribution in the training data could help reduce this bias. In [13], an alternative strategy was introduced. The median of the reconstructed top quark masses in each training batch was added as a feature during training, which mitigated the mass bias.

Nevertheless, the presented analysis already demonstrates the feasibility of employing multi-dimensional unbinned unfolding techniques for future jet substructure measurements.

References

- [1] “Review of Particle Physics”. In: *Physical Review D* 110 (2024). DOI: 10.1103/physrevd.110.030001.
- [2] ATLAS and CMS Collaborations. “Combination of Measurements of the Top Quark Mass from Data Collected by the ATLAS and CMS Experiments at $\sqrt{s} = 7$ and 8 TeV”. In: *Physical Review Letters* 132 (2024). DOI: 10.1103/physrevlett.132.261902.
- [3] S. Alekhin, A. Djouadi, and S. Moch. “The top quark and Higgs boson masses and the stability of the electroweak vacuum”. In: *Physics Letters B* 716 (2012), pp. 214–219. DOI: 10.1016/j.physletb.2012.08.024.
- [4] J. Erler and M. Schott. “Electroweak precision tests of the Standard Model after the discovery of the Higgs boson”. In: *Progress in Particle and Nuclear Physics* 106 (2019), pp. 68–119. DOI: 10.1016/j.pnpnp.2019.02.007.
- [5] CMS Collaboration. “Measurement of the top quark mass using a profile likelihood approach with the lepton+jets final states in protonproton collisions at $\sqrt{s} = 13$ ”. In: *The European Physical Journal C* 83 (2023). DOI: 10.1140/epjc/s10052-023-12050-4.
- [6] A. H. Hoang. “What Is the Top Quark Mass?” In: *Annual Review of Nuclear and Particle Science* 70 (2020), pp. 225–255. DOI: 10.1146/annurev-nucl-101918-023530.
- [7] J. Holguin, I. Moul, A. Pathak, and M. Procura. “New paradigm for precision top physics: Weighing the top with energy correlators”. In: *Phys. Rev. D* 107 (11 2023), p. 114002. DOI: 10.1103/PhysRevD.107.114002.
- [8] M. Kettner. “Systematic Analysis of a Top Quark Mass Measurement using Energy Correlators”. Master’s thesis. Vienna: Vienna University of Technology, 2024.
- [9] J. Holguin, I. Moul, A. Pathak, M. Procura, R. Schöfbeck, and D. Schwarz. “Using the W Boson as a Standard Candle to Reach the Top: Calibrating Energy-Correlator-Based Top Mass Measurements”. In: *Phys. Rev. Lett.* 134 (2025), p. 231903. DOI: 10.1103/j4sp-fcmd. arXiv: 2311.02157 [hep-ph].
- [10] S. Schmitt. “TUnfold, an algorithm for correcting migration effects in high energy physics”. In: *Journal of Instrumentation* 7 (2012), T10003–T10003. DOI: 10.1088/1748-0221/7/10/t10003.

- [11] M. Backes, A. Butter, M. Dunford, and B. Malaescu. “An unfolding method based on conditional invertible neural networks (cINN) using iterative training”. In: *SciPost Phys. Core* 7 (2024), p. 007. DOI: 10.21468/scipostphyscore.7.1.007. arXiv: 2212.08674 [hep-ph].
- [12] M. Backes, A. Butter, M. Dunford, and B. Malaescu. “Event-by-event comparison between machine-learning- and transfer-matrix-based unfolding methods”. In: *Eur. Phys. J. C* 84 (2024), p. 770. DOI: 10.1140/epjc/s10052-024-13136-3. arXiv: 2310.17037 [physics.data-an].
- [13] L. Favaro, R. Kogler, A. Paasch, S. Palacios Schweitzer, T. Plehn, and D. Schwarz. “How to Unfold Top Decays”. In: *SciPost Phys. Core* 8 (2025), p. 053. DOI: 10.21468/SciPostPhysCore.8.3.053. arXiv: 2501.12363 [hep-ph].
- [14] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.
- [15] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.
- [16] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [17] Particle Data Group et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2022 (2022). DOI: 10.1093/ptep/ptac097.
- [18] K. G. Wilson. “Confinement of quarks”. In: *Phys. Rev. D* 10 (8 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445.
- [19] M. C. Gonzalez-Garcia and Y. Nir. “Neutrino masses and mixing: evidence and implications”. In: *Reviews of Modern Physics* 75 (2003), pp. 345–402. DOI: 10.1103/revmodphys.75.345.
- [20] M. Y. Han and Y. Nambu. “Three Triplet Model with Double SU(3) Symmetry”. In: *Phys. Rev.* 139 (1965). Ed. by T. Eguchi, B1006–B1010. DOI: 10.1103/PhysRev.139.B1006.
- [21] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [22] O. Klein. “Quantentheorie und fünfdimensionale Relativitätstheorie”. In: *Zeitschrift für Physik* 37 (1926), pp. 895–906. DOI: 10.1007/bf01397481.

- [23] W. Gordon. “Der Comptoneffekt nach der Schrödingerschen Theorie”. In: *Zeitschrift für Physik* 40 (1926), pp. 117–133. DOI: 10.1007/bf01390840.
- [24] P. A. M. Dirac. *The Principles of Quantum Mechanics*. 4th ed. Oxford, UK: Oxford University Press, 1958.
- [25] P. A. M. Dirac. “The quantum theory of the electron”. en. In: *Proc. R. Soc. Lond. A Math. Phys. Sci.* 117 (1928), pp. 610–624.
- [26] M. Gell-Mann. “Symmetries of Baryons and Mesons”. In: *Physical Review* 125 (1962), pp. 1067–1084. DOI: 10.1103/physrev.125.1067.
- [27] V. MATHIEU, N. KOICHELEV, and V. VENTO. “THE PHYSICS OF GLUEBALLS”. In: *International Journal of Modern Physics E* 18 (2009), pp. 1–49. DOI: 10.1142/s0218301309012124.
- [28] A. Deur, S. J. Brodsky, and G. F. de Téra mond. “The QCD running coupling”. In: *Progress in Particle and Nuclear Physics* 90 (2016), pp. 1–74. DOI: 10.1016/j.pnpnp.2016.04.003.
- [29] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22 (1961), pp. 579–588. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [30] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. “Experimental Test of Parity Conservation in Beta Decay”. In: *Physical Review* 105 (1957), pp. 1413–1415. DOI: 10.1103/physrev.105.1413.
- [31] N. Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Physical Review Letters* 10 (1963), pp. 531–533. DOI: 10.1103/physrevlett.10.531.
- [32] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49 (1973), pp. 652–657. DOI: 10.1143/ptp.49.652.
- [33] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138.
- [34] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.
- [35] H. Yukawa. “On the Interaction of Elementary Particles I”. In: *Proc. Phys. Math. Soc. Jap.* 17 (1935), pp. 48–57. DOI: 10.1143/PTPS.1.1.

- [36] Herb et al. “Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions”. In: *Physical Review Letters* 39 (1977), pp. 252–255. DOI: 10.1103/physrevlett.39.252.
- [37] CDF Collaboration. “Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab”. In: *Physical Review Letters* 74 (1995), pp. 2626–2631. DOI: 10.1103/physrevlett.74.2626.
- [38] D0 Collaboration. “Observation of the Top Quark”. In: *Phys. Rev. Lett.* 74 (14 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632.
- [39] I. Neutelings. *Top quark pair production & decay Feynman diagrams (tt)*. <https://wiki.physik.uzh.ch/cms/latex:feynman:ttbar>. Last modified 5 February 2023. 2023.
- [40] CMS Collaboration. “Measurement of the $t\bar{t}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at 13TeV”. In: *The European Physical Journal C* 79 (2019). DOI: 10.1140/epjc/s10052-019-6863-8.
- [41] K. Olive. “Review of Physics”. In: *Chinese Physics C* 40 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.
- [42] CMS Collaboration. “Precision measurement of the W boson decay branching fractions in proton-proton collisions at 13 TeV”. In: *Physical Review D* 105 (2022). DOI: 10.1103/physrevd.105.072008.
- [43] G. Aad et al. “Measurement of the top quark pair production cross section in pp collisions at $\sqrt{s} = 7$ TeV in dilepton final states with ATLAS”. In: *Phys. Lett. B* 707 (2012), pp. 459–477. DOI: 10.1016/j.physletb.2011.12.055. arXiv: 1108.3699 [hep-ex].
- [44] C. Campagnari and M. Franklin. “The discovery of the top quark”. In: *Reviews of Modern Physics* 69 (1997), pp. 137–212. DOI: 10.1103/revmodphys.69.137.
- [45] “Review of top quark mass measurements in CMS”. In: *Physics Reports* 1115 (2025). CMS physics results from the first decade of LHC data, pp. 116–218. DOI: <https://doi.org/10.1016/j.physrep.2024.12.002>.
- [46] A. J. Lee, C. J. Monahan, R. R. Horgan, C. T. H. Davies, R. J. Dowdall, and J. Koponen. “Mass of the b quark from lattice NRQCD and lattice perturbation theory”. In: *Physical Review D* 87 (2013). DOI: 10.1103/physrevd.87.074018.

- [47] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart. “Extracting a short distance top mass with light grooming”. In: *Physical Review D* 100 (2019). DOI: 10.1103/physrevd.100.074021.
- [48] D0 Collaboration. “Determination of the pole and $\bar{M}S$ masses of the top quark from the $t\bar{t}$ cross section”. In: *Physics Letters B* 703 (2011), pp. 422–427. DOI: <https://doi.org/10.1016/j.physletb.2011.08.015>.
- [49] ATLAS and CMS Collaborations. “Combination of inclusive top-quark pair production cross-section measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV”. In: *Journal of High Energy Physics* 2023 (2023). DOI: 10.1007/jhep07(2023)213.
- [50] CMS Collaboration. “Measurement of the differential $t\bar{t}$ production cross section as a function of the jet mass and extraction of the top quark mass in hadronic decays of boosted top quarks”. In: *The European Physical Journal C* 83 (2023). DOI: 10.1140/epjc/s10052-023-11587-8.
- [51] M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. DOI: 10.1201/9780429503559.
- [52] R. Field. “The Underlying Event in Hadronic Collisions”. In: *Annual Review of Nuclear and Particle Science* 62 (2012), pp. 453–483. DOI: 10.1146/annurev-nucl-102711-095030.
- [53] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.
- [54] J. T. Boyd. *LHC Run-2 and Future Prospects*. 2020. DOI: 10.48550/ARXIV.2001.04370.
- [55] *CERNs accelerator complex*. URL: <https://www.home.cern/science/accelerators/accelerator-complex>.
- [56] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3 (2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004.
- [57] *CMS Coordinate System TikZ*. URL: https://tikz.net/axis3d_cms/.
- [58] Q. Ingram. “Energy resolution of the barrel of the CMS Electromagnetic Calorimeter”. In: *Journal of Instrumentation* 2 (2007), P04004–P04004. DOI: 10.1088/1748-0221/2/04/p04004.

- [59] CMS Collaboration. “Performance of the CMS high-level trigger during LHC Run 2”. In: *Journal of Instrumentation* 19 (2024), P11021. DOI: 10.1088/1748-0221/19/11/p11021.
- [60] F. Beaudette. “The CMS Particle Flow Algorithm”. In: (2014). DOI: 10.48550/ARXIV.1401.8155.
- [61] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet user manual: (for version 3.0.2)”. In: *The European Physical Journal C* 72 (2012). DOI: 10.1140/epjc/s10052-012-1896-2.
- [62] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2014. DOI: 10.48550/ARXIV.1412.6980.
- [63] I. Loshchilov and F. Hutter. *Decoupled Weight Decay Regularization*. 2017. DOI: 10.48550/ARXIV.1711.05101.
- [64] A. Defazio, X. A. Yang, H. Mehta, K. Mishchenko, A. Khaled, and A. Cutkosky. *The Road Less Scheduled*. 2024. DOI: 10.48550/ARXIV.2405.15682.
- [65] Facebook Research. *Schedule-Free Optimisation in PyTorch*. https://github.com/facebookresearch/schedule_free. Accessed: 2025-08-21. 2025.
- [66] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. DOI: 10.48550/ARXIV.1912.01703.
- [67] M. A. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015.
- [68] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. *nflows: normalizing flows in PyTorch*. Version v0.14. 2020. DOI: 10.5281/zenodo.4296287.
- [69] D. J. Rezende and S. Mohamed. *Variational Inference with Normalizing Flows*. 2015. DOI: 10.48550/ARXIV.1505.05770.
- [70] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. *Flow Matching for Generative Modeling*. 2022. DOI: 10.48550/ARXIV.2210.02747.

- [71] N. Huetsch, J. Mariño Villadamigo, A. Shmakov, S. Diefenbacher, V. Mikuni, T. Heimel, M. J. Fenton, K. T. Greif, B. Nachman, D. Whiteson, A. Butter, and T. Plehn. “The landscape of unfolding with machine learning”. In: *SciPost Physics* 18 (2025). DOI: 10.21468/scipostphys.18.2.070.
- [72] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. *Flow Matching for Generative Modeling*. 2022. DOI: 10.48550/ARXIV.2210.02747.
- [73] S. Agostinelli et al. “Geant4 a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [74] G. Cowan. “A survey of unfolding methods for physics”. In: *Conf. Proc. C 0203181* (2002). Ed. by M. R. Whalley and L. Lyons, pp. 248–257.
- [75] A. Höcker and V. Kartvelishvili. “SVD approach to data unfolding”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 372 (1996), pp. 469–481. DOI: 10.1016/0168-9002(95)01478-0.
- [76] G. DAgostini. “A multidimensional unfolding method based on Bayes theorem”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-x.
- [77] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler. “OmniFold: A Method to Simultaneously Unfold All Observables”. In: *Physical Review Letters* 124 (2020). DOI: 10.1103/physrevlett.124.182001.
- [78] *The POWHEG BOX*. 2022. URL: <https://powhegbox.mib.infn.it/> (visited on 05/29/2025).
- [79] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, C. T. Preuss, T. Sjöstrand, P. Skands, M. Utheim, and R. Verheyen. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. 2022. DOI: 10.48550/ARXIV.2203.11601.

- [80] R. Brun, F. Rademakers, P. Canal, A. Naumann, O. Couet, L. Moneta, V. Vassilev, S. Linev, D. Piparo, G. GANIS, B. Bellenot, E. Guiraud, G. Amadio, Wverkerke, P. Mato, TimurP, Matev Tadel, Wlav, E. Tejedor, J. Blomer, A. Gheata, S. Hageboeck, S. Roiser, Marsupial, S. Wunsch, O. Shadura, Anirudha Bose, CristinaCristescu, X. Valls, and R. Isemann. *root-project/root: v6.18/02*. 2019. DOI: 10.5281/ZENODO.3895860.
- [81] R. T. Q. Chen. *torchdiffeq*. <https://github.com/rtqichen/torchdiffeq>. Accessed: 2025-09-11. 2018.