CHARLES UNIVERSITY

HABILITATION THESIS

Physics of the Top Quark: From Precision Tests of the Standard Model to Searches for New Physics

---

Author:
Mgr. Jiří KVITA, Ph.D.

February 27, 2025

Declaration of Authorship

The author declares that he wrote this habilitation thesis by himself using own simulations or referring to results and figures in which he took part as member of the ATLAS collaboration, or as stated otherwise using properly cited materials.

Signed:

Date:

CHARLES UNIVERSITY

Abstract

Faculty of Mathematics and Physics

Habilitation Thesis

Physics of the Top Quark: From Precision Tests of the Standard Model to Searches for New Physics

by Mgr. Jiří KVITA, Ph.D.

An overview of the physics of the top quark is presented from its discovery to measurements of its key properties with a special focus on the differential cross sections of top-antitop quark pairs at Large Hadron Collider at CERN. The top quark rôle in the electroweak physics and the Higgs sector is outlined with effects of the quantum chromodynamics and of electroweak (EW) interactions in its production and decay. Experimental techniques are described, from physics processes and detector simulations to the data analysis, systematics uncertainties propagation and combination of results. Examples of top quark transverse momentum measurements with the ATLAS experiment are presented, with authors contributions, with implications to the top quark processes modelling and precision tests of the Standard Model. The process of unfolding is discussed withing modern statistical methods, together with a review and update on methods of the top quark kinematic reconstruction. Applications to searches for possible new physics using selected spin-0 and spin-1 dark matter mediator models produced along with top-quark pairs are studied and their significance is evaluated for variables of different sensitivity using own simulations.

Contents

List of Abbreviations

0B2S	**0B**oosted (top-tagged) jets, **2S**emiboosted (*W*-tagged) jets
1B1S	**1B**oosted (top-tagged) jet, **1S**emiboosted (*W*-tagged) jet
2B0S	**2B**oosted (top-tagged) jets, **0S**emiboosted (*W*-tagged) jets
BLUE	**B**est **L**inear **U**nbiased **E**stimator
BSM	**B**eyond the **S**tandard **M**odel
CPU	**C**entral **P**rocessing **U**nit
DM	**D**ark **M**atter
EW	**E**lectro**W**eak
FBU	**F**ully **B**ayesian **U**nfolding
JES	**J**et **E**nergy **S**cale
LHC	**L**arge **H**adron **C**ollider
LO	**L**eading **O**rder
MC	**M**onte **C**arlo
ME	**M**atrix **E**lement
ML	**M**achine **L**earning
NLL	**N**ext-to-**L**eading **L**ogarithm
NLO	**N**ext-to-**L**eading **O**rder
NNLL	**N**ext-to-**N**ext-to-**L**eading **L**ogarithm
NNLO	**N**ext-to-**N**ext-to**L**eading **O**rder
PDF	**P**arton **D**istribution **F**unctions
PS	**P**arton **S**hower
QCD	**Q**uantum **C**hromo**D**ynamics
QED	**Q**uantum **E**lectro**D**ynamics
SM	**S**tandard **M**odel
pQCD	**p**erturbative **Q**uantum **C**hromodynamics

Chapter 1

Introduction

This thesis describes the physics of the heaviest elementary particle known to-date, the top quark, from the perspective of measuring its kinematic spectra at the Large Hadron Collider (LHC) at CERN over a variety of central-mass-energies of proton-proton and proton-lead collisions. Further ideas are presented how the top quark can be used as a tool for exploring and searching for possible physics beyond the Standard Model of elementary particles as for its particle content or interactions. The top quark keeps its attributes as an object of high physics interest, and is being probed by key experiments at LHC, currently the world-leading accelerator at the energy frontier.

The document is structured as follows: we shall provide an overview of the necessary parts of the Standard Model and the physics of the top quark, its production, decay, key properties and a note on its discovery. Next, experimental techniques in high energy physics relevant to the presented author's own work [Auth1] are briefly described, ranging from the events generation and simulation to detection techniques for objects relevant for top quark physics, followed by statistical methods and unfolding [Auth2] of measured spectra.

A brief chapter is devoted to an overview of top quark measurements at the Tevatron collider, followed by the key chapter on selected topics on the physics of the top quark with the ATLAS experiment at the LHC, with highlighted author's direct contributions.

The main focus is the story of the top quark transverse momentum, its measurement, and the theory (in)description of this important variable, the evolution of the data and theory agreement, highlighting different experimental approaches, measurements in various phase space regions and using varied detector objects and concepts. Other spectra important for modelling of this complex final state are also discussed.

This thesis thus documents author's contributions to his key ATLAS analyses at the energy of pp collisions of 7 TeV in [ATLAS1, ATLAS2], followed by the 8 TeV analyses [ATLAS3, ATLAS4] and at 13 TeV as published as [ATLAS5], with additional measurement on the same data sample presented in [ATLAS6] and completed by the measurement using the full 13 TeV dataset [ATLAS7]. A section is devoted also to selected interpretations stemming from these measurements. The recent measurement of the top quarks pair production in the proton-lead collisions with the ATLAS experiment [ATLAS8], with key author's involvement, is also described.

Finally, top quark as a tool for searches for possible new physics beyond the Standard Model is described in the last Chapter, in analyses [Auth3, Auth4] performed using own simulation, with studies at the detector level using Standard Model as well as possible new physics samples, and also testing new signal samples behaviour in unfolding to the particle level.

Chapter 2

The Standard Model

We shall first provide an introduction to the current understanding of particles and their interactions within the Standard Model of interactions, with a brief overview of its particle content, dynamics of the key forces and a sketch of the Higgs mechanism, with indications of what physics may be present beyond. We also briefly discuss relevant experimental aspects of the strong interaction.

2.1 Historical overview

Particle physics studies the fundamental components of matter and their interactions. Ever since the idea of elementary and possibly non-divisible building blocks of matter, the human curiosity has pursuit the exploration of the microcosm, arriving to a more complex picture that ever thought. Hand in hand with the theory, experimental nuclear and particle physicists have gradually unveiled the existence of the electron, proton, neutron, positron, strange particles containing a novel quantity of matter, the s quark. The experimental tools of using the cosmic rays for pions and kaons discovery, and also of a member of the second generation of leptons, the muon, was quickly furbished with artificial particle accelerators, from cyclotrons and linacs to current state-of-the-art superconducting synchrotrons. Dramatic discoveries of the c and b quarks were the reward, followed by the discovery of the third charged lepton τ. Later, large-volume neutrino detectors unveiled the existence of neutrino oscillations proving the non-zero mass of the neutrinos. Along, the quest for hadrons and their excitations, the highly abstract non-Abelian structure of the electroweak (EW) interactions was experimentally confirmed and predicted the phenomenon of weak neutral currents, confirmed as such in neutrino interactions and then also by the creation of the real quanta of the W and Z bosons in 1980s in proton-antiproton collisions. The non-point-like character of the proton and its deep partonic structure has lead to the understanding of one of the most peculiar interaction, the strong force. Finally, the discovery of the scalar boson with properties consistent with the one predicted by the minimal electroweak symmetry breaking mechanism, closed, for the moment, the chapter of the quest for the building blocks and interaction bosons of the Standard Model (SM). The three families of leptons and quarks are an established paradigm of the Standard Model, a very precise theory, which is however thought to be only an approximation to a possible more complex theory beyond, due to astrophysical and astroparticle experimental hints, non-zero neutrino masses as well as theory reasons. The top quark, not only the heaviest fermion of the SM, but actually the heaviest particle among all, may play a special rôle in the SM as well as in possible connections to signs of a new physics. This thesis is dedicated to the studies of this particle, with the emphasis on precision measurements of its properties to constrain the SM or to look for possible effects beyond.

2.2 Standard Model

The Standard Model describes the interactions of the known elementary particles to a high level of precision and in agreement with a plethora of experimental results, explaining rich phenomenology of particles interactions and decay. The path to the discovery of the underlying principles, details of the interactions, their experimental confirmation together with the discovery of the particle content and number of families is an adventure of its own which has taken place since the discovery of the electron.

2.2.1 Particle content of the Standard Model

The neutral and charged leptons and the six quarks as content of the Standard Model are summarized in Table 2.1, together with their charges and masses. Neutrinos and the u, c and t quarks carry the weak isospin $T_3 = \frac{1}{2}$ while charged leptons and the d, s and b quarks carry $T_3 = -\frac{1}{2}$.

Electric charge	Symbol and mass					
0	ν_e		ν_μ		ν_τ	
$-1\|e\|$	e	(0.511 MeV)	μ	(0.1056 GeV)	τ	(1.777 GeV)
$+2/3\|e\|$	u	(2.2 MeV)	c	(1.27 GeV)	t	(172.7 GeV)
$-1/3\|e\|$	d	(4.7 MeV)	s	(0.093 GeV)	b	(4.18 GeV)

TABLE 2.1: Fermionic particle content of the Standard Model with particle masses [1] in brackets: neutral and charged leptons, and quarks; with charges indicated. Not shown are antiparticles to those presented in the table. Mass uncertainties vary, the number of digits corresponds to their precision. Neutrino mass eigenstates are different from their flavour eigenstates, with current upper limits being around 1 eV.

All the fermions take part in weak interactions, charged fermions in electromagnetic interactions, and quarks also in strong interactions. Gravity is not relevant for collider experiments and the nature of its possible unification with the framework of the gauge theories describing the other interactions is not fully understood.

The electroweak interactions are mediated by the massive charged $W^\pm$ and the neutral Z bosons and the massless photon, strong interactions by eight colour-charged massless gluons. An additional complex scalar field is added to the theory in order to generated in a gauge-invariant way mass terms in the SM Lagrangian for the $W^\pm$ and Z bosons. This results in an additional physics scalar appearing in the spectrum, the BEH (often just the Higgs) boson, see below. Table 2.2 shows an overview of the calibration as well as the BEH bosons properties in terms of charge, mass and full decay width.

2.2.2 Standard Model as a gauge theory

The Standard Model of interactions of elementary particles [2] is a quantum field theory with a a specific underlying symmetry group, which generates, requiring the invariance of the form of the Lagrangian under the calibration transformation, the strong and electroweak interactions. The electroweak (EW) sector was proposed by Seldon Glashow, Abdus Salam and Steven Weinberg, see *e.g.* the famous Model of Leptons [3], introducing three vector bosons A^a_μ related to the $SU_L(2)$ group transforming chirally left-handed fermions, and an additional vector boson B_μ related to

Symbol	Name	Interaction	Mass	Width
γ	photon	electromagnetic	0	
$\boldsymbol{W}^{\pm}$	charged vector bosons	weak (charged currents)	80.37 GeV	2.09 GeV
$\boldsymbol{Z}$	neutral vector boson	weak (neutral currents)	91.19 GeV	2.50 GeV
$\boldsymbol{g}$	gluons	strong	0	
$\boldsymbol{H}$	BEH boson (scalar)	Yukawa	125.2 GeV	$3.7^{+1.9}_{-1.4}$ MeV

TABLE 2.2: An overview of calibration bosons of the electroweak and strong interactions, and the BEH boson, with their masses and decay widths. All bosons are neutral with the exception of the $W^{\pm}$ bosons; all are vector (spin-1) particles except for the scalar BEH boson.

the so-called hyper-charge Abelian group $U_Y(1)$. Masses of the physical calibration bosons $W^{\pm}$ and Z are generated by the Higgs mechanism [4, 5, 6] and the mass terms for fermions are obtained by adding the allowed gauge-invariant Yukawa coupling terms.

Quantum chromodynamics is a complex part of the SM in which self interactions of its calibration bosons (eight massless electrically neutral gluons G^a_μ) lead to quarks confinement in hadrons and asymptotic freedom of quarks (oppositely running of the coupling constant as function of the energy of the process). The result is a rich hadron phenomenology, the picture of partons in hadrons and their distribution functions, and the need to define and compute measurable observables, often related to hadronic final states, jets.

As it is not possible to detail all the aspects of the SM within a limited space, we shall briefly summarize the structure of the theory in a condensed form, assuming the reader has been at least partially already exposed to the topic.

The Lagrangian of the Standard Model can be split into parts containing the kinetic terms of vector bosons, including their self-interactions; and the EW, strong, Higgs and its Yukawa interactions with fermions as

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{kin+EW\,self}} + \mathcal{L}_{\mathrm{EW}} + \mathcal{L}_{\mathrm{strong}} + \mathcal{L}_{\mathrm{Higgs}} + \mathcal{L}_{\mathrm{Yukawa}}\,.$$

In a compact form, this can be elegantly expressed as

$$\mathcal{L}_{\mathrm{SM}} = -\frac{1}{4}\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu} + i\bar{\psi}\not{D}_F\psi + |D^\mu\Phi|^2 - V(\Phi) + H\sum_{jk} y_{jk}\bar{\psi}_j\psi_k + \mathrm{h.c.}\,,$$

where $\mathbf{F}$ are all the field strength tensors of SM gauge bosons which can be expressed as commutators of the covariant derivatives as

$$-\frac{1}{4}\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu} = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{2.1}$$

$$F_{\mu\nu} = -\frac{1}{ig}\left[D^{EW}_\mu, D^{EW}_\mu\right], \quad SU_L(2)\ \text{(electroweak)} \tag{2.2}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\nu\,, \quad U_Y(1)\ \text{(electroweak)} \tag{2.3}$$

$$G_{\mu\nu} = -\frac{1}{ig_S}\left[D^S_\mu, D^S_\nu\right], \quad SU_{\mathrm{colour}}(3)\ \text{(strong)}\,. \tag{2.4}$$

The covariant derivatives for fermions differ for leptons and quarks depending on their weak hyper-charge and are different for chiral left and right-handed fermion

fields $\psi_L \equiv \frac{1}{2}(1-\gamma_5)\psi$ and $\psi_R \equiv \frac{1}{2}(1+\gamma_5)\psi$; and can be written as

$$D_\mu^{EW} \equiv \begin{cases} \partial_\mu - ig\frac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu - ig'Y_L B_\mu & \text{for } \psi_L\,, \quad Y_L = -\frac{1}{2} \\ \partial_\mu - ig'B_\mu & \text{for } \psi_R \end{cases} \tag{2.5}$$

$$D_\mu^S \equiv \partial_\mu - ig_S\frac{1}{2}\boldsymbol{\lambda}\cdot\mathbf{G}_\mu \tag{2.6}$$

$$D_\mu^F \equiv D_\mu^{EW} + D_\mu^S \tag{2.7}$$

$$D_\mu \equiv \partial_\mu - ig\frac{1}{2}\boldsymbol{\tau}\cdot\mathbf{A}_\mu - ig'Y_\Phi B_\mu\,, \quad Y_\Phi = -Y_L = \frac{1}{2} \tag{2.8}$$

where $\boldsymbol{\tau}$ and $\boldsymbol{\tau}$ are the Pauli and Gell-Mann matrices, respectively. The relation between the fermion charge and weak isospin is $Q = T_3 + Y$ and only quark fields couple to gluons fields. Expanded, the field strength tensors for the non-Abelian electroweak and strong parts become

$$F_{\mu\nu}^a = \partial_\mu A_\nu - \partial_\nu A_\mu + g f^{abc} A_\mu^b A_\nu^c \tag{2.9}$$

$$G_{\mu\nu}^a = \partial_\mu G_\nu - \partial_\nu G_\mu + g_S \tilde{f}^{abc} G_\mu^b G_\nu^c\,, \tag{2.10}$$

which leads to the calibration bosons self interactions, with f and $\tilde{f}$ being the $SU(2)$ and $SU(3)$ groups structure constants, respectively.

Further in the notation, ψ is a vector of fermion fields of quarks and leptons of different flavours, $\bar{\psi} \equiv \psi^\dagger\gamma^0$ is the Dirac-conjugated bispinor, Φ is the doublet of the complex scalar Higgs field, H is its physics degree of freedom (the Higgs field), D_F and D are the covariant derivatives for fermions and the Higgs field, respectively, which include the kinetic as well as interaction terms with the relevant SM fields of the electroweak $W^\pm$ and Z bosons. More precisely, three A_a^μ spin-1 boson fields of the $SU_L(2)$ group and one B^μ spin-1 boson field of the $U_Y(1)$ group, forming the $SU_L(2)\times U_Y(1)$ group, L denoting the relation to left-handed fermions only, Y standing for the weak hyper-charge. Finally, there are eight spin-1 gluon fields G_a^μ in case of quarks interactions.

The covariant derivatives depend on specific choices of the weak hyper-charge (in order to lead to the desired phenomenology and agreement with experiment) and the fact that only left-handed fermions enter the interaction with the $SU_L(2)$ calibration bosons.

Last, $V(\Phi)$ is the Brout-Englert-Higgs (and also Guralnik–Hagen–Kibble) potential and y_{jk} are the Yukawa coupling constants between fermions of flavours j and k before diagonalization which then gives rise to fermion masses. The diagonalized mass matrices in the quark (the Cabibbo–Kobayashi–Maskawa, CKM, matrix) and neutrino (the Pontecorvo–Maki–Nakagawa–Sakata, PMNS, matrix) sectors each carry one complex phase which is responsible for the violation of the CP symmetry in the Standard Model. More details and the derivation can be found in [2].
In a more expanded form

$$\mathcal{L}_{\text{kin+EW self}} = i\sum_j \bar{\psi}_j \partial^\mu \psi_j - \frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} \tag{2.11}$$

$$\mathcal{L}_{\text{EW}} = \sum_j \bar{\psi}_L^j \left[g' Y_j^L B^\mu + \frac{1}{2} g A_a^\mu \sigma^a \right] \gamma_\mu \psi_L^j \tag{2.12}$$

$$+ \sum_j \bar{\psi}_R^j\, g' Y_j^R B^\mu \gamma_\mu \psi_R^j \tag{2.13}$$

$$\mathcal{L}_{\text{strong}} = -\frac{1}{4} G^a_{\mu\nu} G_a^{\mu\nu} + \frac{1}{2} g_S \sum_q \bar{\psi}_q G^\mu_a \lambda^a \gamma_\mu \psi_q \tag{2.14}$$

$$\mathcal{L}_{\text{Higgs}} = D_\mu \Phi^\dagger D^\mu \Phi + \mu\, \Phi^\dagger \Phi - \lambda \left(\Phi^\dagger \Phi \right)^2 \tag{2.15}$$

$$\mathcal{L}_{\text{Yukawa}} = H \sum_{j,k \in \text{flavours}} y_{jk} \bar{\psi}_j \psi_k + \text{h.c.}\,, \tag{2.16}$$

where index j runs over all quark and lepton flavours, F and G are the field strengths of the electroweak and gluon bosons, respectively, which include also their self interactions; g and g' are the electroweak coupling constants while g_S is the strong coupling constant; and h.c. stands for the Hermitian conjugated terms.

Indices a label the number of fields and generators of the non-Abelian symmetry groups, *i.e.* they run from 1 to 3 for the $SU_L(2)$ part of the EW group, and from 1 to 8 for the $SU_{\text{colour}}(3)$, the strong group, related to the strong charge called colour. The colour indices in the strong part of the Lagrangian are not indicated (should be borne by the λ matrices and quark fields around them). The sums over j and k run over charge and neutral leptons as well as quarks, the sum over q runs over quark bispinors in 8 colour versions; σ_a, λ_a and γ_μ are the three Pauli, eight Gell-Mann and four Dirac matrices, respectively, while μ (where not a Lorentz index) and λ are parameters of the Higgs potential (where not the Gell-Mann matrices).

It is worth noting that the electromagnetic and strong interactions are of a pure vector nature, *i.e.* that the neutral vector photon and gluons fields couple to vector currents of fermions. On the other hand, the charged currents, *i.e.* interactions of fermions with the W bosons are of the vector-minus-axial-vector (V−A) nature in a form which actually maximally violates the parity in these interactions. Interactions of fermions with the Z boson are a more complicated mixture of V and A currents.

The manifestly symmetric Higgs potential does not lead to a straightforward particle interpretation, and the four degrees of freedom of the original complex field doublet are usually recast into the radial $H(x)$ and three phase fields $\pi_a(x)$ as

$$\phi(x) = \frac{1}{\sqrt{2}}\, e^{\frac{i}{v}\tau^a \pi_a(x)} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix},$$

where $v = \frac{\mu}{\sqrt{\lambda}}$ is the location of the minimum of the classical Higgs potential function form in terms of $|\Phi|$. One can "calibrate out" the radial fields π_a, and in such a unitary gauge, only the Higgs fields remains in the theory, while the interactions of the Higgs field with the $SU_L(2) \times U_Y(1)$ gauge bosons, together with their reparameterisation[1] to physical fields of the Z boson and the photon using the Weinberg weak mixing angle $\cos\theta_W = g/\sqrt{g^2 + g'^2}$, $\theta_W \doteq 30^\circ$ as

$$\begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} A^\mu_3 \\ B^\mu \end{pmatrix} \tag{2.17}$$

and the charged W bosons as $W^\pm_\mu = \frac{1}{\sqrt{2}} \left(A^1_\mu \mp i A^2_\mu \right)$ give rise to the masses of the $W^\pm$ and Z bosons

$$m_W = \frac{1}{2} v g\,, \quad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}\,,$$

[1] With the reparameterisation, Higgs boson self interactions also emerge from the "kinetic" term with the covariant derivative.

leading to the famous relation $m_W = m_Z \cos\theta_W$. Numerically the masses are $m_W = 80.377 \pm 0.012\,\text{GeV}$ and $m_Z = 91.1876 \pm 0.0021\,\text{GeV}$ [1]. The mass of the Higgs boson, though, is not predicted by the theory, and is measured by the ATLAS and CMS experiments as $m_H = 125.25 \pm 0.17\,\text{GeV}$ [1]. Together with the relation between the Fermi and the EW weak coupling constants $G_F = \frac{\sqrt{2}g^2}{8m_W^2}$ and its known value of $G_F \doteq 1.1663788(6) \times 10^{-5}\,\text{GeV}^{-2}$ this implies the Higgs field potential minimum, or the vacuum expectation value (VEV), $v/\sqrt{2} \doteq 246\,\text{GeV}$.

All SM Electroweak and Higgs boson (self)interactions are displayed in Figure 2.1 while the strong interactions are visualized in Figure 2.2.

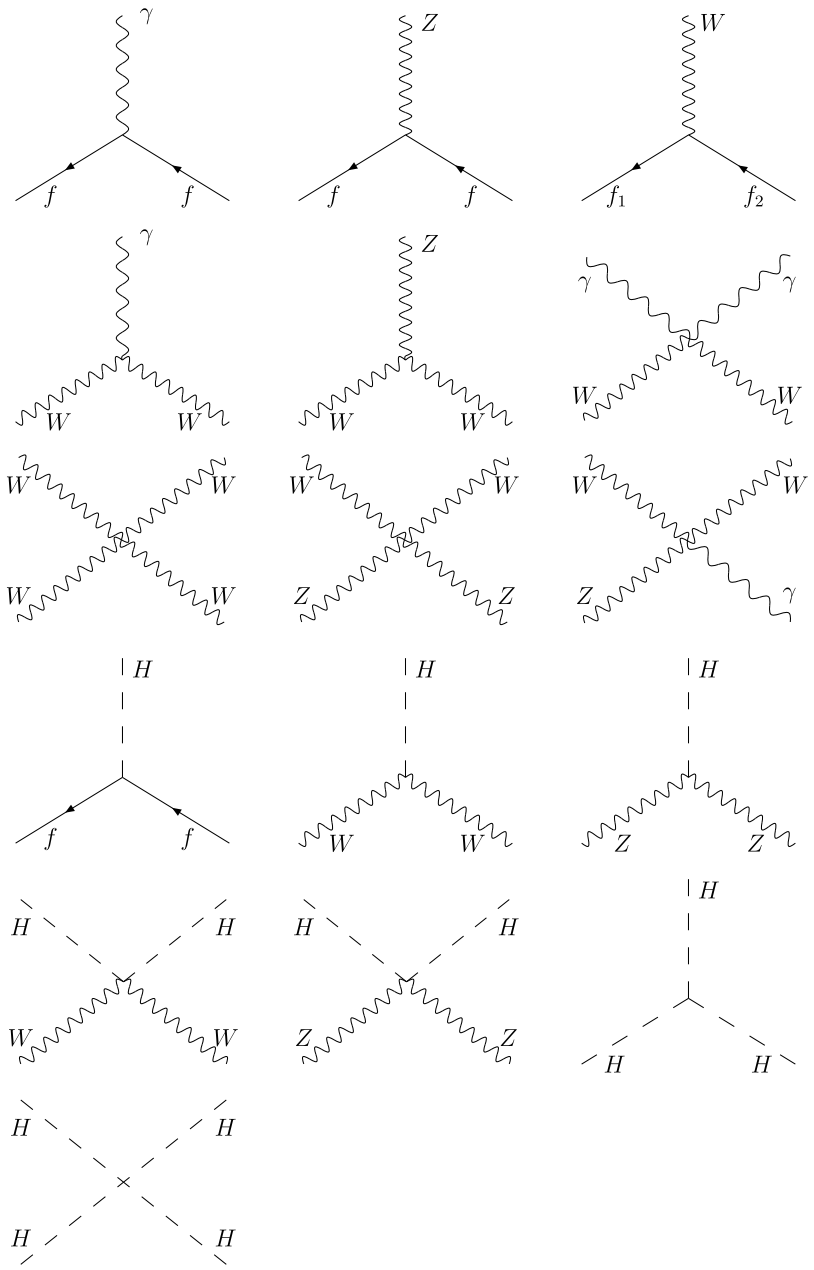


FIGURE 2.1: Electroweak and BEH boson interaction terms of the SM Lagrangian displayed as Feynman diagrams [2].

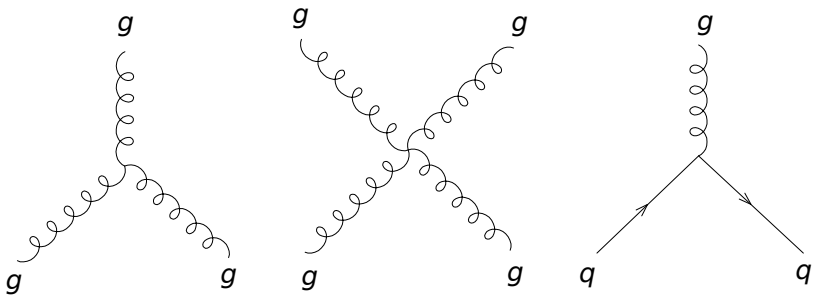


FIGURE 2.2: Strong interaction terms of the SM Lagrangian displayed as Feynman diagrams [7].

2.2.3 Strong interactions and the factorization theorem

The underlying theory of the strong force is the Quantum Chromodynamics (QCD), for a review see *e.g.* [8].

All current experimental evidence points to the existence of the proton structure, *i.e.* of partons, being both charged quarks (and anti-quarks) and neutral gluons. In the current picture of the inelastic strong interactions, it is possible (and in fact necessary) to separate and factorize the short distance physics of the high-energy (large momentum scale and transfer) partons collisions and the long-distance (and therefore low energy) structure of hadrons.

The resulting QCD factorization theorem combines the computable parton-level cross section with the semi-empirical parton distributions functions (PDFs) to predict the cross-sections at the level of colliding hadrons. See *e.g.* [9] for an exposure on the topic or more recent [10, 11].

Besides the strength of the coupling constant, which actually evolves with some typical energy scale of the process Q, evaluating to $\alpha_S(m_Z) \doteq 0.116$, there are many more aspects of the QCD, stemming from the opposite dependence (running) of its coupling constant α_S [12] on the energy scale of the process as compared to that of the electromagnetic force, α, numerically $\alpha(0) \doteq 1/137 \doteq 0.00730$. These are the quarks confinement in hadrons, and the asymptotic freedom, *i.e.* quarks being effectively non-interacting at short distances (large energy or momentum transfer).

The phenomenology of strong interactions is rich due to the fact that quarks (and gluons) readily radiate additional partons in the soft (low-energy) and collinear regime. This leads to the concept of jets as the only meaningful observable final states already at the parton level. Simply put, one cannot ask what is the cross section to produce a certain number of final state partons, but rather what is the cross section to produce a given number of jets above some momentum threshold. It is possible to compute these cross sections from the first principles of the QCD while the non-perturbative parts of the QCD like the hadronization have to be measured and are only phenomenologically modelled.

The functions $f_{i/h}(x, \mu_F)$ describe the probability distribution of resolving parton i in hadron h, with i carrying the fraction x of the original hadron four-momentum, and μ_F being the factorization scale which is effectively the virtuality (μ_F^2 being the four-momentum squared) of the parton entering the hard QCD process. This scale emerges in the evolution of parton distribution functions, separating the partons initial state radiation as part of the evolution of the partons entering the hard process, or the radiation being part of additional partons to be included in the the matrix element which is computable.

Current understanding of the strong force does not allow one to compute the (non-perturbative) PDFs from the first QCD principles (the Lagrangian). However, it is possible to compute their evolution as function of the scale μ_F. Therefore, once the PDFs are measured at a certain scale, they can be predicted and computed at any other scale. This of course also assumes their universality, *i.e.* the measurements of the PDFs *e.g.* at an electron-proton collider do bring information about the universal proton structure, which is then transferable to other collisions.

The cross section for the production of a final state X in the collisions of two hadrons h_1 and h_2 can then be written in terms of the PDFs and the cross-section at the parton level as

$$\mathrm{d}\sigma_{h_1h_2\to X}(s) = \sum_{ij} \int_{\hat{s}_{\mathrm{thr}}}^{s} \Big[f_{i/h_1}(\cdot,\mu_F) \otimes f_{j/h_2}(\cdot,\mu_F) \Big](\hat{s},\mu_F) \cdot \mathrm{d}\hat{\sigma}_{ij\to X}(\hat{s},\mu_R)\,\mathrm{d}\hat{s}\,,$$

where the sum runs over possible combinations of partons i and j to enter the process of interest, the hat denotes the cross-section at the parton level, μ_f is the factorization scale (here set the same for both partons for simplicity), μ_R is the renormalization scale (which regularizes high energy behaviour of the perturbative series), $\otimes$ stands for the Mellin convolution of the two PDFs; $\sqrt{s}$ is the total available energy in the central-mass-system of the h_1 and h_2 hadrons collision, $\sqrt{\hat{s}}$ is the energy of the colliding partonic system and the lower integration limit $\hat{s}_{\mathrm{thr}}$ is the process threshold energy squared. It holds that $\hat{s} = sx_1x_2$ and the parton luminosity dependence on the $\hat{s}$ can be expressed using Dirac's delta function as

$$\Big[f_{i/h_1}(\cdot,\mu_F) \otimes f_{j/h_2}(\cdot,\mu_F) \Big](\hat{s},\mu_F) = \iint\limits_{sx_1x_2\geq\hat{s}_{\mathrm{thr}}} f_{i/h_1}(x_1,\mu_F)\, f_{j/h_2}(x_2,\mu_F)\, \delta(\hat{s}-sx_1x_2)\,\mathrm{d}x_1\mathrm{d}x_2\,.$$

In theory, the summed pQCD results to all orders of the perturbation theory should not depend on the choice of the renormalization scale μ_R. In practice, it does as the theory predicts the cross section only to a finite order of pQCD. Similarly, one has to choose the factorization scale μ_F. The practical choice of the scales are the typical energy of the collision, momentum transfer, masses of the massive particles, their transverse momenta, or some their combination. One can thus distinguish static or dynamic (event-dependent, event kinematic specific) scale choices.

2.2.4 Evolution of parton distribution functions

Elaborating more on the scale dependence of the PDFs, let's consider an example processes of $e^+e^- \to q\bar{q}$ and $e^+e^- \to q\bar{q}g$. It turns out that the emission of the extra gluon factorizes from the $q\bar{q}$ production and leads to soft (small gluon energy) and collinear (emission at a small angle w.r.t. the initial quark) singularities [13] leading to the probability of the soft gluon emission

$$\mathrm{d}\mathcal{S} = \frac{2\alpha_S C_F}{\pi} \frac{\mathrm{d}E}{E} \frac{\mathrm{d}\theta}{\sin\theta} \frac{\mathrm{d}\phi}{2\pi}\,,$$

where C_F is related to the trace of colour Gell-Mann matrices in the $q\bar{q}g$ vertex.

Redefining the variables describing the emitted gluon from E and θ to z and k_{T} defined as $E = p(1-z)$ and $k_{\mathrm{T}} \equiv E\sin\theta$, where p is the momentum of the original

quark before the gluon radiation, this becomes, in the limit $k_\mathrm{T} \approx E\theta$,

$$\mathrm{d}\mathcal{S} = \frac{\alpha_S C_F}{\pi} \frac{\mathrm{d}z}{1-z} \frac{\mathrm{d}k_\mathrm{T}^2}{k_\mathrm{T}^2} .$$

If one were to integrate over the degrees of freedom, two types of singularities, the soft (infrared) singularity $z \to 1$ and the collinear singularities ($k_\mathrm{T} \to 0$) appear which can be isolated and remedied in two separate ways, also depending whether appearing in the initial or final partonic states. Both the soft and collinear singularities in the final state cancel with virtual corrections at the level of cross sections (Kinoshita-Lee-Nauenberg theorem [14, 15]) when adding cross sections of the $2 \to 2$ and $2 \to 3$ processes.

Extending the above to the process of hadrons scattering, a similar collinear singularity in the initial state parton evolution stems from the integration over the additional gluon momentum transverse to the direction of the original quark $\mathrm{d}k_\mathrm{T}^2/k_\mathrm{T}^2$ from zero, where actually the perturbative QCD is not expected to work at such low scales.

Not being interested in nor being able to experimentally resolve the soft or collinear gluon, one can analyze a contribution to the production of final state X with the additional gluon emission from the initial state quark $\mathrm{d}\hat{\sigma}_{q\bar{q}\to Xg}(\hat{s}, p, x, z)$ which depends on the energy of the process and also on the momentum fraction of the quark resolved from the hadron, x, w.r.t. to original hadron momentum p, and also the fraction z of the momentum of the quark entering the hard process, while the gluon carries energy fraction $1-z$. In the integration over one of the PDF functions, the low-k_T (non-perturbative) part of the quark annihilation process can be absorbed into a redefinition of the parton distribution function by defining the lower limit in the integral over the gluon k_T as the factorization scale μ_F

$$\begin{aligned} \mathrm{d}\sigma_{h_1 h_2 \to Xg}(s,p) \quad &\ni \quad \frac{\alpha_S C_F}{\pi} \int\limits_{\mu_F^2}^{Q^2} \frac{\mathrm{d}k_\mathrm{T}^2}{k_\mathrm{T}^2} \int\limits_0^1 \mathrm{d}z \iint \frac{\delta(\hat{s} - szx_1x_2)}{1-z} \, \mathrm{d}\hat{\sigma}_{q\bar{q}\to X}(\hat{s}, zxp) \\ &\times \quad f_{q/h_1}(x_1, \mu_F)\, f_{\bar{q}/h_2}(x_2)\, \mathrm{d}x_1 \mathrm{d}x_2 . \end{aligned}$$

The PDF now depends on μ_F as it contains the integration over gluon transverse momenta below μ_F. The upper limit Q^2 should be smaller than the typical momentum scale of the hard process as we have dealt with emission of additional gluons as a higher order correction to the original process. In this expression we have regularized only the infinity coming from the gluon emission from the initial quark. This can (and should) be done also for the antiquark, or, in general, for any initial state parton. The divergent soft ($z \to 1$) part cancels with a similar divergence of an opposite sign coming from a virtual contribution to the quark propagator in the initial state, and the integral over k_T^2 is then finite.
In this soft limit, one can identify the core of the splitting kernel

$$p_{qq}(z) \approx 2C_F \frac{1}{1-z}$$

which becomes in a more detailed calculation

$$p_{qq}(z) = 2C_F \frac{1+z^2}{1-z}$$

and eventually, considering both virtual and real contributions to the evolution of the partons, the LO QCD quark-to-quark splitting kernel

$$P_{qq}(z) \equiv \left[\frac{1+z^2}{1-z}\right]_+ ,$$

where the plus distribution ensures the cancellation of the divergence with the virtual contribution.

While the PDFs are per se non-perturbative and cannot be calculated from current first QCD principles, their evolution is known, governed by the DGLAP equations [16] derived by G. Altarelli, G. Parisi and V. N. Gribov with L. N. Lipatov. Symbolically, the one for the quark splitting can be written as a convolution

$$\frac{\mathrm{d}}{\mathrm{d}\ln\mu_F^2} f_q(x,\mu_F) = \frac{\alpha_S(\mu_F)}{2\pi}\int_x^1 \mathrm{d}z\, P_{qq}(z)\frac{f_q(x/z,\,\mu_f)}{z} \equiv \frac{\alpha_S(\mu_F)}{2\pi} P_{qq}(\cdot)\otimes f_q(\cdot,\,\mu_f)\,.$$

Assuming all possible parton splitting, this becomes coupled differential equations

$$\frac{\mathrm{d}}{\mathrm{d}\ln\mu_F^2}\begin{pmatrix} f_q \\ f_g \end{pmatrix} = \frac{\alpha_S(\mu_F)}{2\pi}\begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix}\otimes\begin{pmatrix} f_q \\ f_g \end{pmatrix}.$$

The splitting kernels depend on the order of the perturbation theory in which they are computed and are currently known up to the NNLO accuracy. They allow the applicability of the universal parton PDFs, yet measured at low momentum scales *e.g.* at the HERA collider, to the case of LHC for much higher momentum transfers.

2.2.5 Modifications of PDFs in nucleus

Let us also discuss briefly modifications to parton distributions functions (PDFs) in the nuclear environment. Experimental data have brought evidence of various effects on both the structure functions and the PDFs in scattering experiments with ions, see Figure 2.3. In order to compare to those of free nucleons, nuclear modification factors are defined as ratios of structure functions or PDFs in the nucleus to the Z and $A-Z$ weighted sum of the individual proton and neutron structure functions, *e.g.*

$$R_q^A(x,Q^2) \equiv \frac{f_{q/A}(x,Q^2)}{Z f_{q/p}(x,Q^2) + (A-Z) f_{q/n}(x,Q^2)}\,.$$

As nucleons are bound in nucleus, modifications of parton distribution functions can be present, accounting for various mechanisms of probing the nucleus as well as through additional interactions the dense nuclear environment brings, see [17, 18] for reviews. In a brief summary, there are several effects of various origins, see also Figure 2.4:

- Shadowing is a collective effect (due to the presence of other nucleons forming a dense nuclear environment) of decreased PDFs at low x, below about 0.1, and depends only slowly on the momentum transfer (grows with energy).
- Anti-shadowing can be regarded as a compensation effect due to the shadowing and the sum rules of the parton distribution functions, being effectively probability density functions in some sense; occurring in x of about 0.1–0.3.

- At large x (about 0.3–0.7) the EMC effect (named after the European Muon Collaboration experiment) is the effect of the nucleon structure modified by the medium around it, leading to reduction of PDFs at high x, either due to short-range correlations between nucleons or via bound nucleons being larger due to higher partons reduction by the medium and a corresponding larger nucleons size via the uncertainty principle. The EMC effect actually came as a surprise and the first indication of nuclear environment modifications to PDFs.
- Fermi motion of nucleons effectively adds momenta to partons and enhances PDF regions of highest x.

An example of the nuclear modification factor for the gluon distribution in lead is shown in Figure 2.5. It is evident that while the shape extracted using experimental data by the EPPS team is nontrivial, it also bears large systematic uncertainties of more than 10%. Additional experimental data are thus needed to constrain the nuclear PDF uncertainties.

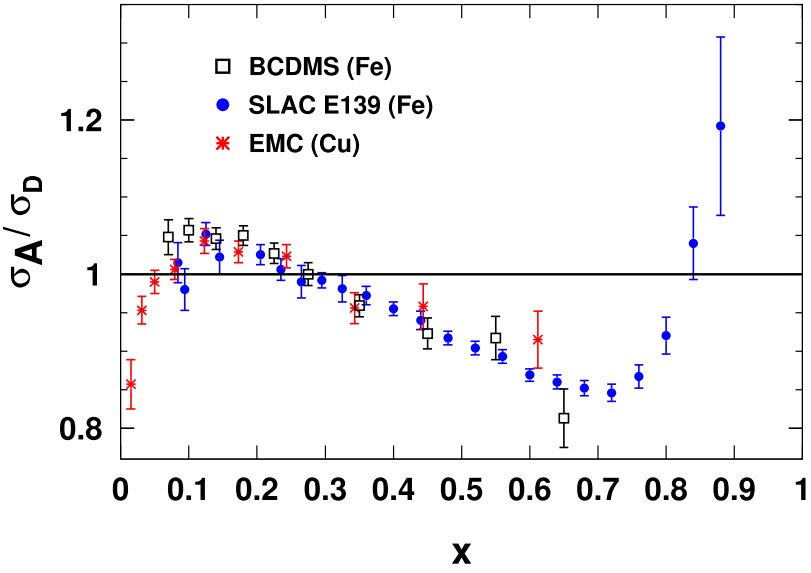


FIGURE 2.3: The ratio of the measured cross sections of scattering of muons on iron or copper to those on deuterium as extracted by the BCDMS, SLAC E139 and the EMC collaborations [19].

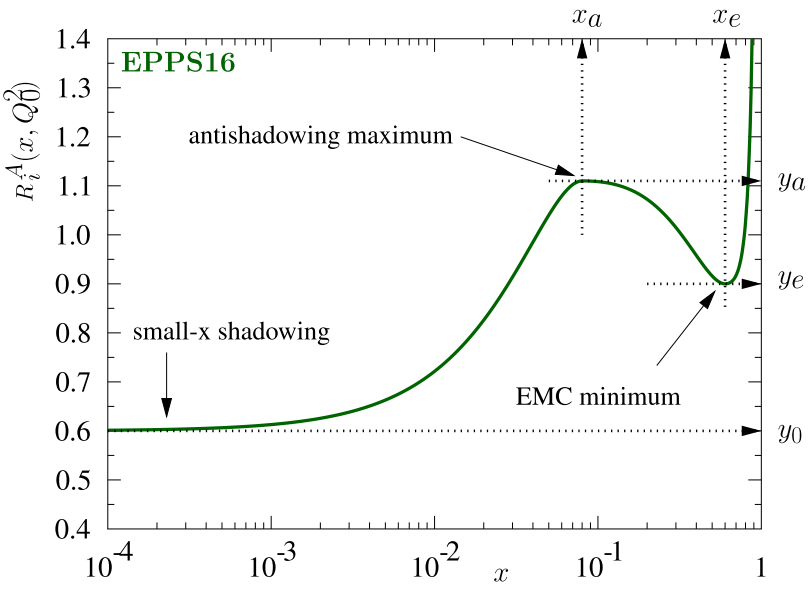


FIGURE 2.4: The fit form for the nuclear modification factors used by the EPPS collaboration [20], showing the shadowing, antishadowing, EMC and Fermi regions in Bjorken x.

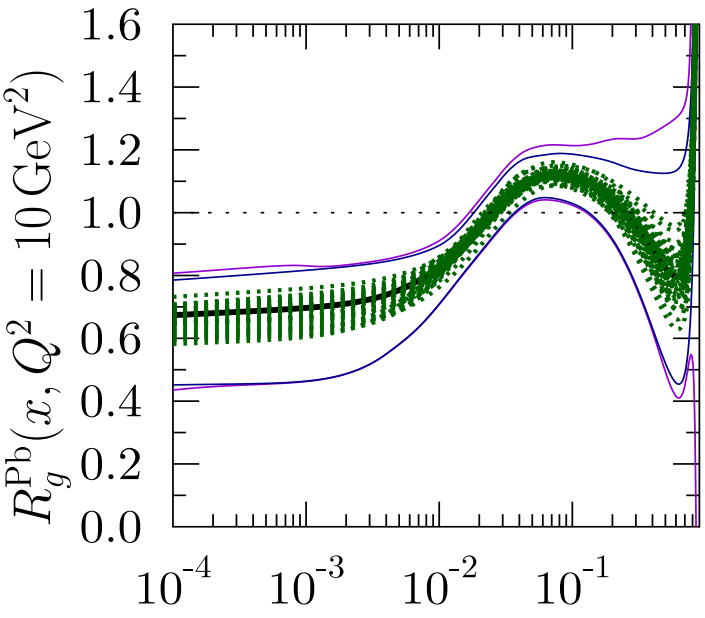


FIGURE 2.5: The EPPS21 modifications to the gluon distribution function [21] in lead at $Q^2 = 10\,\mathrm{GeV}^2$ as function of Bjorken x. The central result is indicated by the thick black curve, the nuclear uncertainty sets by green dotted curve. The blue curve corresponds to the nuclear uncertainties while the purple one to the full uncertainty (nuclear and baseline errors added in quadrature).

2.3 Physics beyond the Standard Model

Starting with the non-zero neutrino masses, the question of whether SM is the final ultimate theory of the elementary particles and their interaction has been around for decades. Other hints are the large number of free parameters in the SM Lagrangian, three different couplings, which however may evolve to a single one at high energies [1] (section 93 of the full review), the hierarchy of particle masses, loop corrections to the Higgs boson mass and the problem of fine-tuning its mass from different loop contributions. A critical analysis of this problem can be found *e.g.* in [22].

Other important hint is the amount of matter-to-antimatter asymmetry observed in our Universe, for which a large violation of a combined CP symmetry [2] is needed, but in fact larger than the CP parameter contained in the quarks CKM mixing matrix while the CP sector of the neutrino mixing matrix still awaits its precise determination at future neutrino experiments. Along, missing signs of the CP violation in the sector of strong interactions (the strong CP problem [23]) is another puzzle of a fine-tuning nature.

Other reasons for the presence of some additional physics beyond the SM (BSM) are the astrophysical and cosmological observations, ranging from the indications of the presence of dark matter in galaxies, in fluctuations of the cosmic microwave background, through the non-zero value of the cosmological constant [24] or the evidence of the inflation period of the Universe expansion at its early stage [25].

Proposed BSM theories range from extended symmetry groups of the SM, extra dimensions, super-symmetry or the string theory. Additional gauge bosons could lead to interactions not yet observed or of other particles beyond the SM, or SM particles coupling to possible new particles. Experimental search for such new features is an active field of activities at current high energy as well as precision high-luminosity physics experiments.

[2] Combined symmetry operation of exchanging particles for antiparticles, and applying a space inversion operation.

Chapter 3

Top quark

In this chapter the theoretical needs for the top quark are presented, with a notion on its rôle in driving a generation of experiments at hadron colliders. Top quark production and main decay channels are presented, with highlights of the interplay of strong and weak interactions, and of differences between the recent hadron colliders. We further discuss top quark properties and opportunities to study QCD and EW interactions using the top quark. The discovery of the top quark completed the era of finding the basic building blocks of strongly interacting matter. We shall provide a short historical overview of the first establishment of the top quark signal at the Tevatron accelerator experiments.

3.1 Theoretical needs for the top quark

The SM predicts several interesting expectations on the number of families of particles. For instance, the number of lepton and quark families should be equal in order to cancel the theoretical ABJ anomaly effects in the quantum field theory [2]. Since the discovery of the τ lepton in 1975 and of the b quark in 1977, the question of the existence of the sixth quark, the top quark, has guided experimental efforts to discover it. The mass of the top quark was not predicted by the theory, and it is in fact one of the many fundamental parameters of the SM. Due to the vacuum expectation value of the Higgs field in the SM ($v/\sqrt{2} \sim 246$ MeV), the question is perhaps not why the top quark is so heavy but why all other fermions are so light.

3.2 Top quark production

At hadron colliders, top quarks are most frequently produced in pairs via the strong interaction. In the picture of the leading order (LO) of the perturbative quantum chromodynamics (pQCD) the quark-antiquark annihilation was the major process at the proton-antiproton ($p\bar{p}$) collider Tevatron while the gluon-gluon fusion dominates in proton-proton (pp) at the LHC. Beyond the LO picture, the $t\bar{t}j$ process can also be initiated by a quark-gluon collision, see Figure 3.1. Table 3.1 shows the fraction of the $q\bar{q} \rightarrow t\bar{t}$ production and total $t\bar{t}$ cross-sections as computed at the LO using the MADGRAPH5 generator for the Tevatron and LHC energies as well as for hypothetical colliders at the same energy but with complementary types of colliding hadrons.

Colliding hadrons	$\sqrt{S}$	$q\bar{q}$ fraction at LO	total LO $t\bar{t}$ cross-section
$p\bar{p}$ (Tevatron)	1.96 TeV	0.89	5.95 pb
pp (hypothetical)	1.96 TeV	0.35	1.05 pb
pp (LHC design)	14 TeV	0.064	562 pb
$p\bar{p}$ (hypothetical)	14 TeV	0.135	608 pb

TABLE 3.1

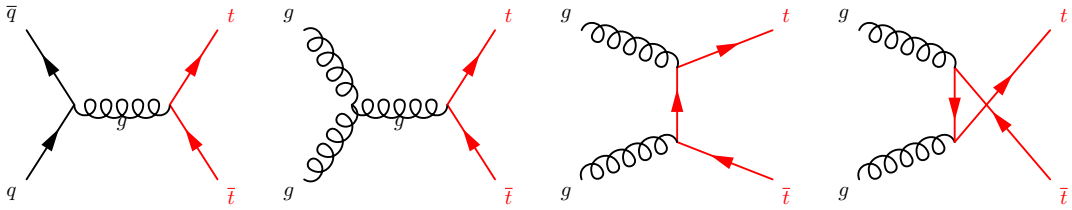


FIGURE 3.1: Dominant leading-order Feynman diagrams for the SM processes of the $t\bar{t}$ pair production. From left to right: the s-channel quark antiquark annihilation and gluon-gluon fusion diagrams, and the t and u-channel gluons fusion diagrams. Created using the *FeynMf* package [26].

3.3 Initial partons flavour fractions in $t\bar{t}$ production

The fractions of the initial $q\bar{q}$, gg and qg or $\bar{q}g$ parton combinations for the Tevatron and LHC colliders as function of the partonic system energy $\hat{s}$ are shown in Figure 3.2. One can also see the threshold of $2m_t \approx 350$ GeV. At Tevatron, the $q\bar{q}$ annihilation was the dominant process for the $t\bar{t}$ production, while at the LHC the gg dominates. The results were obtained using own simulations in MADGRAPH5 at NLO at the respective energies and colliding beam particles.

Distributions of the Bjorken x for the Tevatron and LHC cases are shown in Figure 3.3, demonstrating that while at Tevatron the mean x was about 0.25, it is 0.05 for gluon PDFs at the LHC. The most probable values of x are, however, near the threshold and about twice smaller than the mean. Still, different regions of momentum fractions of protons have been probed at the two colliders.

3.4 Other processes involving the top quarks

Further processes involving top quarks production are the electroweak production of single (anti)top quarks, and associated production of top quarks or top quark pairs with an electroweak or the Higgs boson (tV, $\bar{t}V$, tH, $\bar{t}H$, tVH, $\bar{t}VH$, $t\bar{t}V$, $t\bar{t}\gamma$ *etc.*, with $V = W, Z$). Other more exotic processes involve the production of three or four top quarks, the latter being also recently observed experimentally [28, 29].

3.5 Top quark decay

In the SM at the tree level Feynman diagrams, quarks decay or transform to another type of quark and a W boson via the weak interaction. In cases of the u, d, c, s and b quarks, the associated W boson is highly off-shell, and can decay to a pair of a lepton and a neutrino, or to a pair of quarks if energy is provided by *e.g.* the parent particle containing the original quark. In contrast, the top quark decays almost exclusively

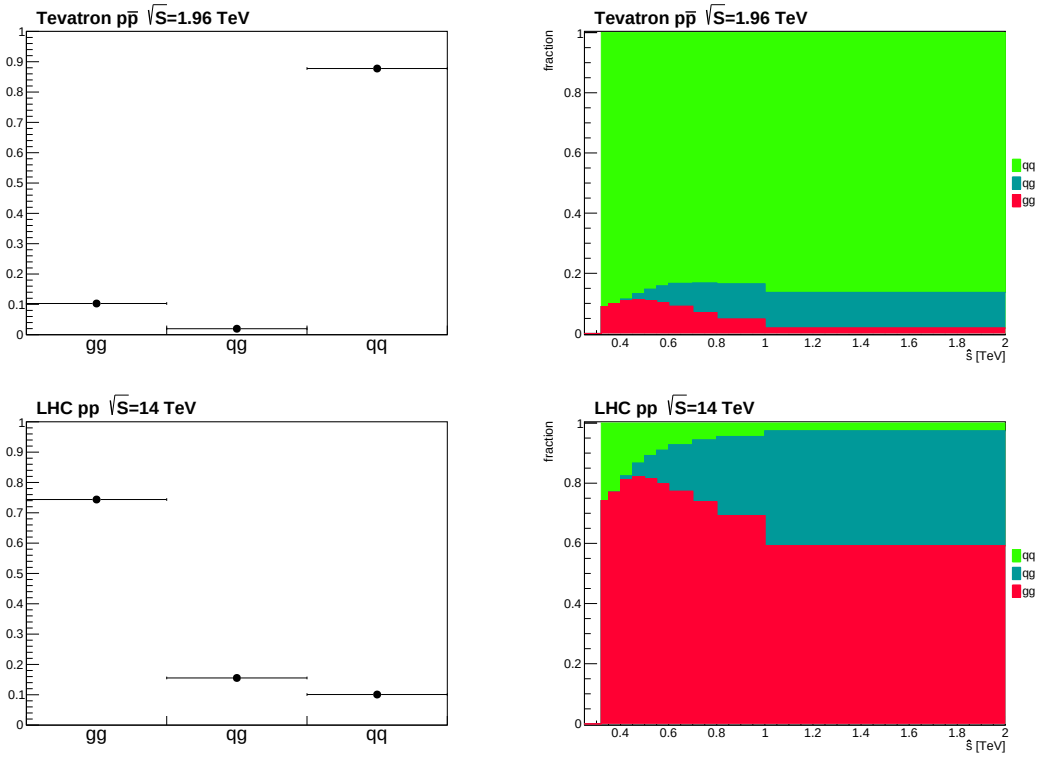


FIGURE 3.2: Flavour fractions of the the gluon-gluon (gg), quark-gluon (qg) or quark-antiquark ($q\bar{q}$) initial parton collisions at the energy of $\sqrt{s} = 1.96\,\text{TeV}$ for $p\bar{p}$ collisions (top) and $\sqrt{s} = 14$ TeV pp collisions (bottom), as generated by MADGRAPH5 at NLO using the `NNPDF23_nlo_as_0119` PDF set [27], inclusively (left) and as function of the energy of the parton-parton collision $\sqrt{\hat{s}} = \sqrt{sx_1x_2}$.

into the down-type quark within the third family, the b quark, and an on-shell W boson which then decays hadronically or leptonically, see Figure 3.5.

3.6 $t\bar{t}$ decay modes

The classification of the $t\bar{t}$ final state decays is based on the way the W bosons decay, see the illustration in Figure 3.6. Starting with the branching fraction of the W boson decay modes to hadrons of 0.6741 [1], this leads to final states and branching fractions as shown in Table 3.2. The leptonic decays in the table also include the τ lepton, experimentally one usually assumes only channels leading to electrons and muons while hadronic tau decay modes need special identification techniques.

$t\bar{t}$ LO final state	channel name	branching fraction
$b\bar{b}\ell\nu_\ell\ell'\nu_{\ell'}$	dilepton	10.6%
$b\bar{b}q\bar{q}'\ell\nu_\ell$	ℓ+jets	43.9%
$b\bar{b}q\bar{q}'q''\bar{q}'''$	all-hadronic	45.4%

TABLE 3.2: Leading order $t\bar{t}$ final states, their naming convention and branching fractions.

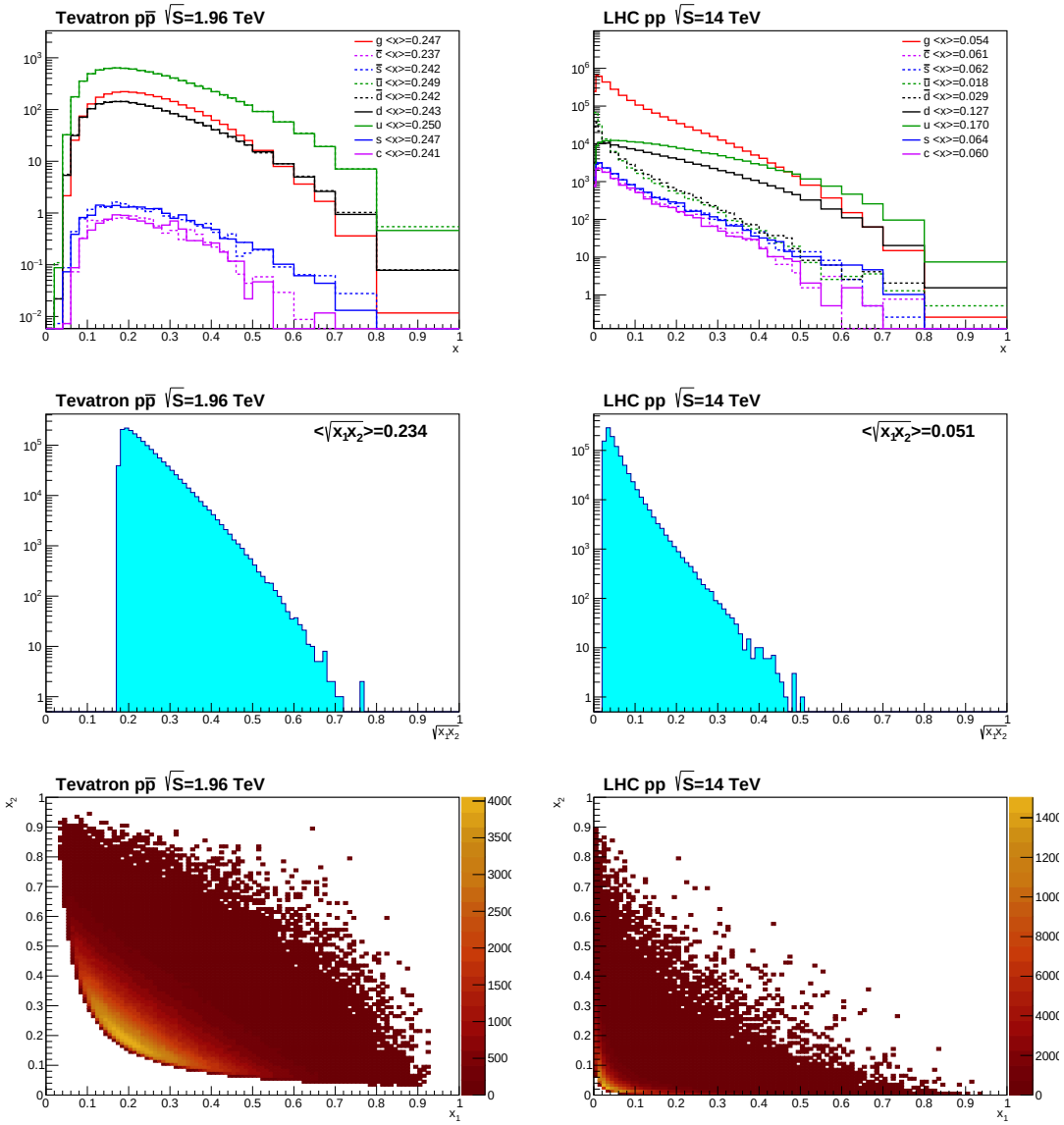


FIGURE 3.3: Top: Histograms of the Bjorken x for different flavours at the energy of $\sqrt{s} = 1.96\,\text{TeV}$ for $p\bar{p}$ collisions (left) and $\sqrt{s} = 14\,\text{TeV}$ pp collisions (right) in the $t\bar{t}$ production, as generated by MADGRAPH5 using the `NNPDF23_nlo_as_0119` PDF set [27]. Quarks (gluons) clearly dominate the Tevatron (LHC) case. Middle: the product of the fractions carried by the two partons entering the hard process; bottom: the scatter plot of the two fractions, showing the hyperbolic kinematic threshold boundary.

3.6.1 Leptonic final states

Cases where the $t\bar{t}$ system decays such that one of the W bosons decays leptonically and the other one hadronically are denoted as the ℓ+jets channel while both W bosons decaying leptonically to an electron or a muon, or to a tau lepton decaying leptonically are called the dilepton channel. These final states, with a different degree of the boost of the hadronic part of the system, are depicted in Figure 3.7.

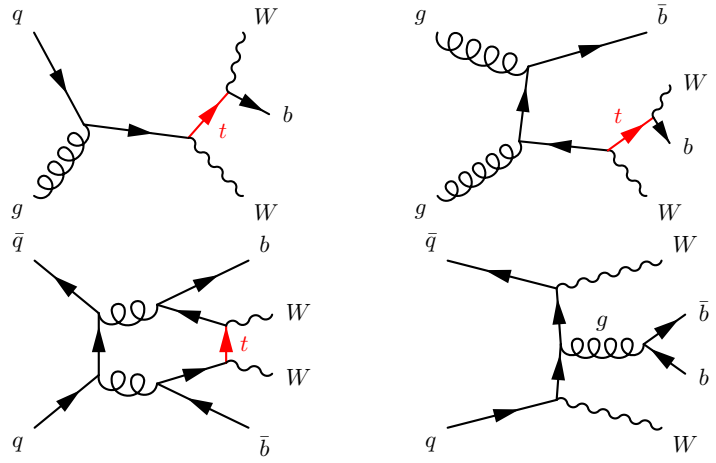


FIGURE 3.4: Example single resonant and non-resonant (w.r.t. the top quark propagator) diagrams leading to the WbW and $WbWb$ final states. Created using the *FeynMf* package [26].

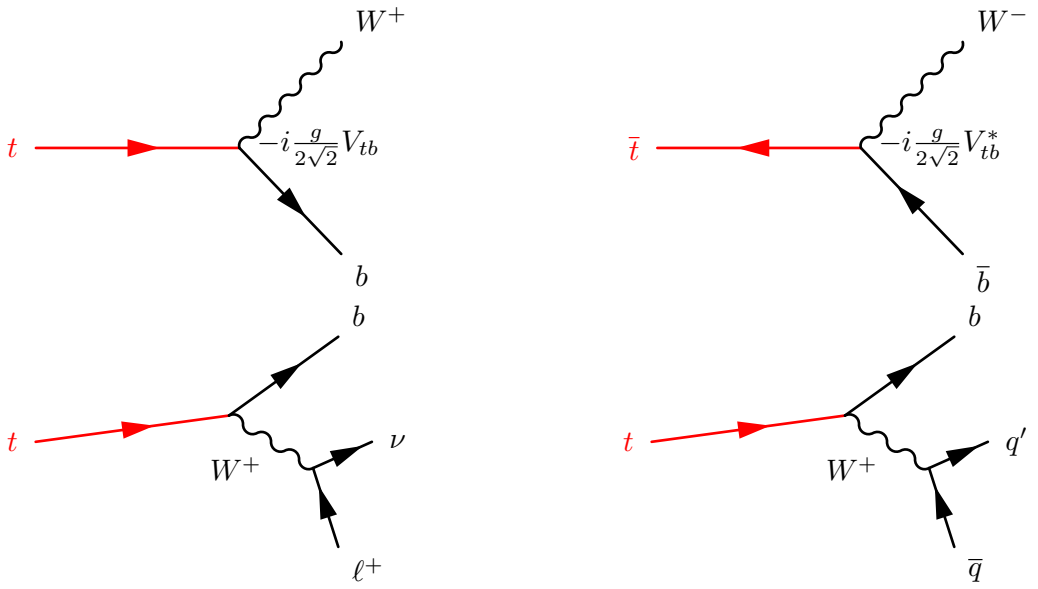


FIGURE 3.5: Top quark and anti-top quark decays (top) and leptonic and hadronic top quark decays (bottom). Created using the *FeynMf* package [26].

3.6.2 Hadronic final states

Of special importance are the hadronic final states of top quarks which we describe below.

In the regime of a moderate transverse momentum of the top quark, its decay products will be angularly separated in the detector. In the case of a hadronic top quark decay, one expects three hadronic final states, jets, of which one should correspond to a b-jet, *i.e.* a jet with reconstructed tracks preferring the signature of a long-lived B hadron. Such a resolved topology of three jets is useful in allowing the reconstruction of invariant masses of the two and three jet systems, with an expected peak structure around the masses of the W and top quark, respectively.

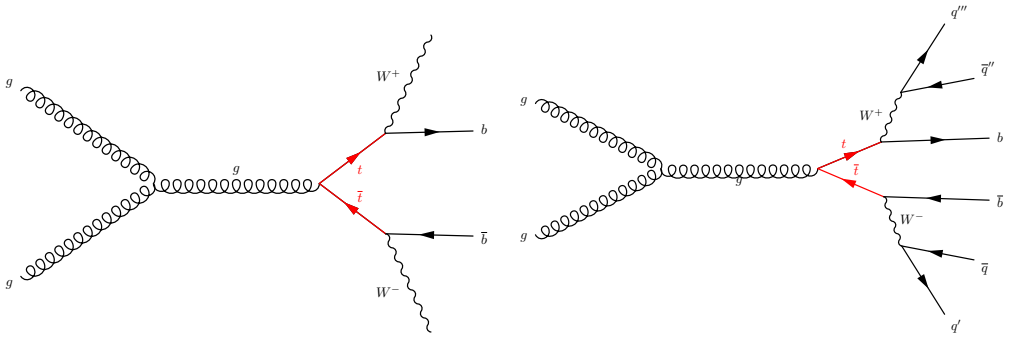


FIGURE 3.6: Decay of the $t\bar{t}$ to $W^+bW^-\bar{b}$ final state (left) and fully decayed $t\bar{t}$ final state to the all-hadronic channel. Created using the *FeynMf* package [26].

In the kinematic regimes of top quarks highly boosted in the transverse plane, jets coming from the hadronically decaying W boson or event the full top quark decay products can manifest in the detector as wide large-radius jets. These can be searched for by dedicated jet algorithms of larger distance parameters than is usually done when reconstructing individual jets in the angularly resolved case.

Such boosted hadronic W and top jets can be present both in the ℓ+jets as well as allhadronic final states. The different types of hadronic final states based on the degree of the boost and thus also the angular separation of the jets is illustrated in Figure 3.8. For a visualization of what the events look like in current detectors, event displays of the hadronic boosted final states from the parameterized DELPHES simulation are shown in Figure 3.9.

3.7 Higher order processes

There are more complex processes going beyond this simple picture of the $t\bar{t}$ creation and decay. Top quarks and b-quarks can radiate gluons, as governed by the strong interactions. Processes beyond the naïve separation of the top quark production and decay are called beyond the top quark narrow-width approximation (NWA). Different higher order QCD as well as electroweak (EW) corrections can further modify the kinematics and in general add further and more complex Feynman diagrams of different kind for different decay modes of the W bosons. At the hard process stage, additional jets can be produced be hard emissions [30] of the initial or final state partons in the matrix element (ME) corresponding to the diagram, with the boundary of what emissions should be included in parton evolution or in the Feynman diagram of the hard process being somewhat arbitrarily and reflected in the evolution of the parton distributions functions. This is partially illustrated in Figure 3.10.

In practice, further final state parton radiation is modelled by dedicated parton shower (PS) generators like PYTHIA or HERWIG and double counting in the phases of these extra emissions must be accounted for in the complex matching of PS to ME. See Chapter 4.1 for more discussion.

3.8 Top quark properties

Here we shall provide a list of key top quark properties and their rôle in verification and tests of the Standard Model.

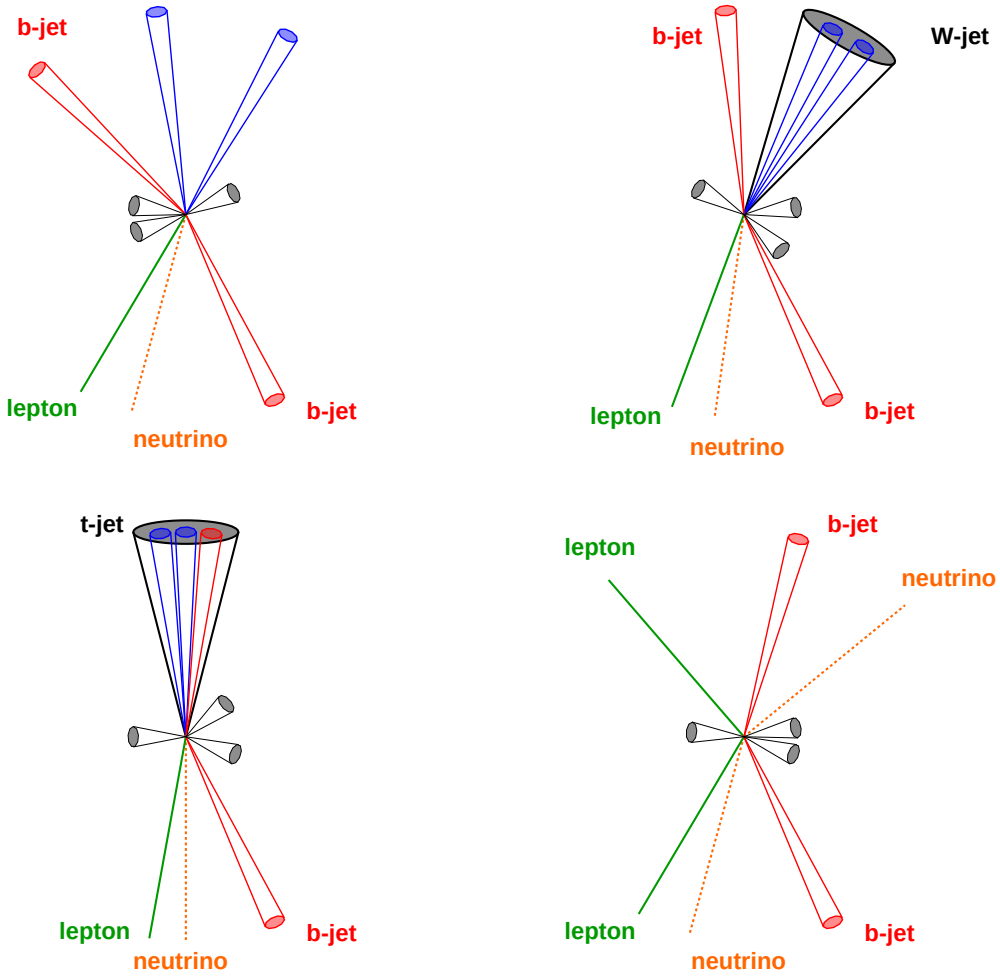


FIGURE 3.7: Cartoons of the ℓ+jets topologies with the resolved, semiboosted and boosted hadronic parts of the $t\bar{t}$ decay; and the (resolved) dilepton event. We do not show cases where, due to a large boost, one can also consider angular merging of the leptons and the b-jets. Smaller black cones represent additional jet activity in the event.

3.8.1 The mass of the top quark

Top quark mass being larger than that of the W bosons makes the top quark special in its weak decay to an on-shell W boson and a b quark, whereas the W boson can decays to a pair of a quark and another antiquark for flavours u, d, s, c and b.

Top quarks weak decay actually proceeds with such a large decay amplitude that its life time resembles that of resonances decaying via a strong interaction. In fact, top quark decay width of about 1.4 GeV is of the same order as the natural width of the Z and W bosons. From purely dimensional reasons, the top quark decay width is proportional to $G_F m_t^3$, leading to the mean life time of $4.7 \cdot 10^{-25}$ s. Therefore, while the decay process is mediated by the weak interaction, the resulting mean life time is actually comparable to that of particles decaying due to strong interactions. Strictly speaking, as top quark is not stable, it cannot be regarded as an asymptotic state in the scattering theory. There are therefore ambiguities regarding the definition of the top quark mass. One possibility is the pole mass, *i.e.* the mass where

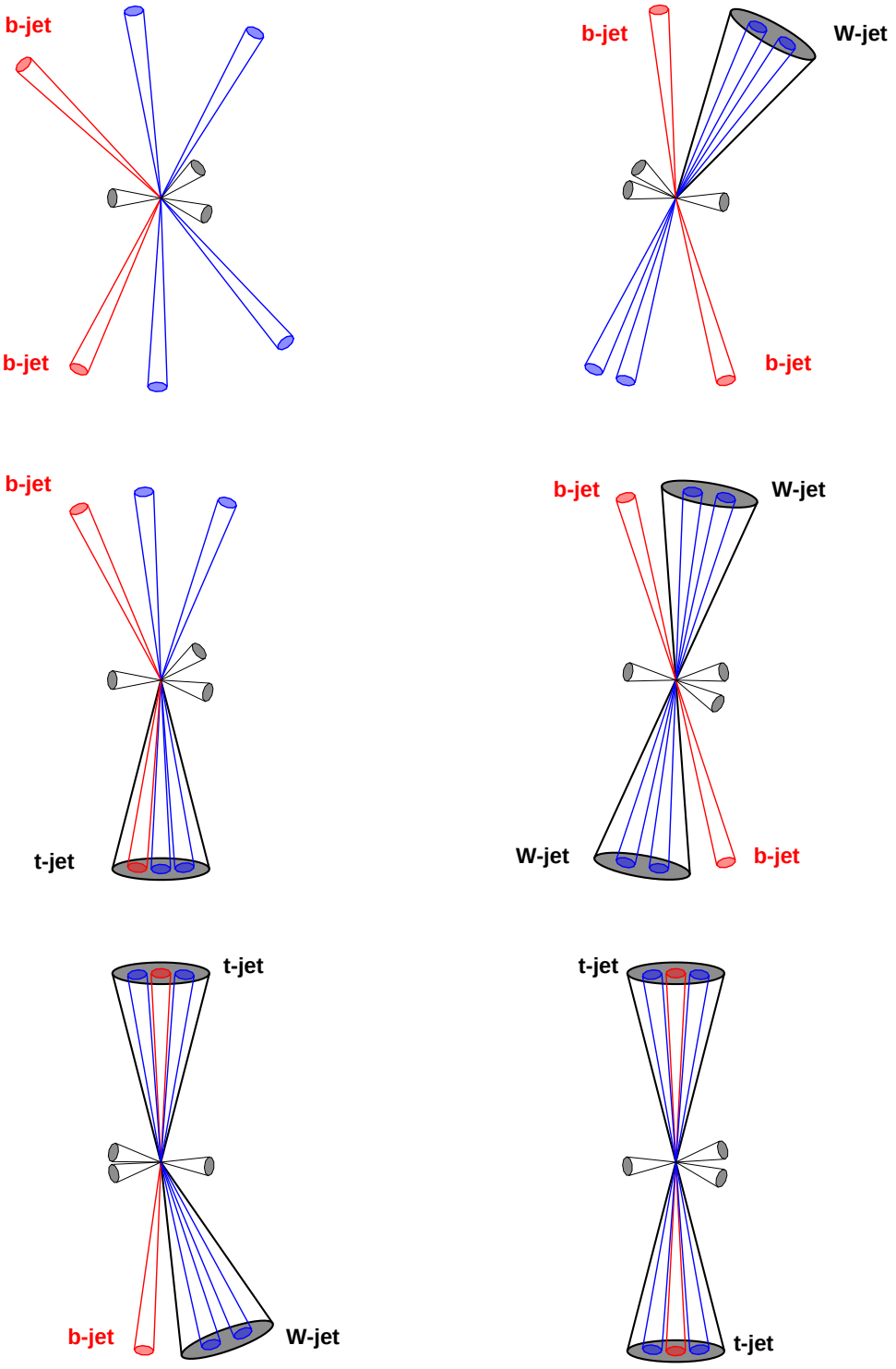


FIGURE 3.8: Cartoons of the allhadronic $t\bar{t}$ event topologies, from top to bottom and left to right: fully resolved (2R), resolved-semibosted (1R1S), resolved-boosted (1R1B), semiboosted-semiboosted (0B2S), boosted-semiboosted (1B1S) and boosted-boosted (2B0S). Smaller black cones represent additional jet activity in the event.

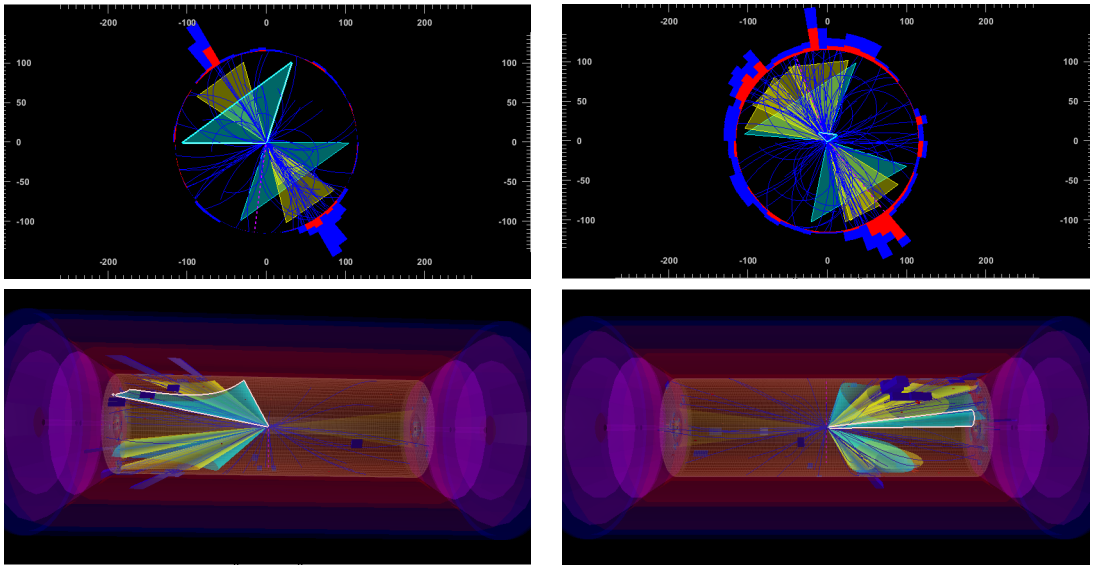


FIGURE 3.9: DELPHES event display of a boosted $t\bar{t}$ event (left) and a more complex final state, namely additional jet activity (right), from the Z' sample with $m_{Z'} = 1\,\text{TeV}$. Small-R jets are the yellow cones, large-R jets in light blue.

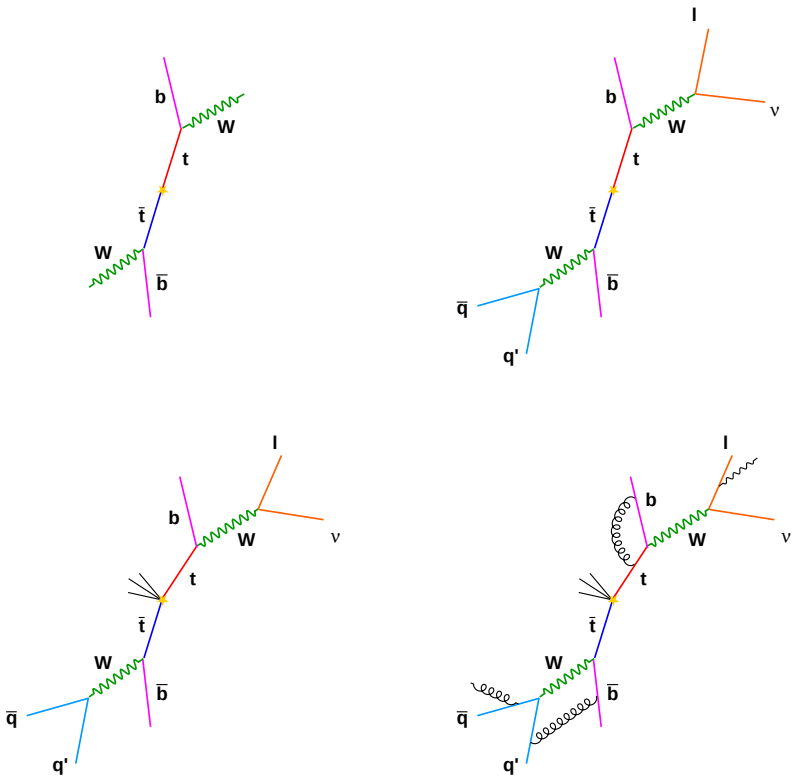


FIGURE 3.10: Cartoon of complexities and modelling issues in $t\bar{t}$ pQCD production, illustrating higher-level corrections, the weak top quark decay, corrections beyond the narrow width approximation.

the top quark propagator is maximal. Other option is the so-called running mass, renormalized parameter of the Lagrangian, which then depends on the renormalization scheme and suffers from ambiguities. The top quark mass can be obtained from experiment as the kinematic peak peak in mass-sensitive observables, or even from the cross-section which depends on its value. Various techniques with complementary sensitivities to systematic uncertainties, especially to the one in the jet energy scale, have covered a full landscape of many measurements especially at the LHC. The latest overview of the measurements of the top quark mass are summarized in Figure 3.11.

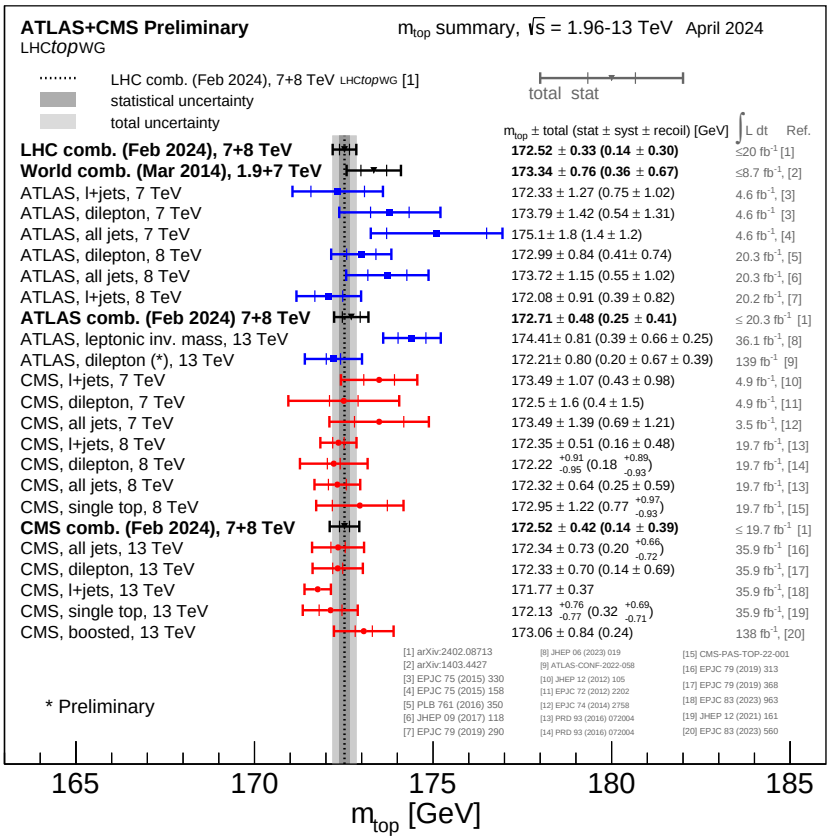


FIGURE 3.11: Individual measurements as well as the combinations on the mass of the top quark from the LHC experiments [31].

3.8.2 Mass consequences

Due to its short life time, much of the properties of the top quarks are imprinted into its decay products. For example, its mass can be determined from the decay products observed in the detector, including hadronic final states (jets). Also, the spin correlation and entanglement between the two top quarks when produced in pairs are preserved in terms of angular correlations of the decay products, and are not affected by hadronization effects.

Top quark thus decays before a typical hadronization time, and consequently does not form hadrons. There is very probably no toponium, a hypothetical $t\bar{t}$ bound

state analogous to set of states as observed in bound systems of $c\bar{c}$ or $b\bar{b}$ quarks, although some hints of its creation could be deduced from the region of low $t\bar{t}$ invariant mass [32].

3.8.3 Further properties

We provide a short overview of other top quark processes which can be measured at a hadron collider. The charge of the top quark in the SM is expected to be $2/3$ of the elementary charge. Experimentally the charge is accessible either via a direct measurement of the charges of leptons and (in a more complex way) of a jet, or via the cross section measurements of the $t\bar{t}\gamma$ process. The top quark charge measurement is important to confirm the structure of the $SU_L(2)$ doublet also in the third family of quarks. Other measurements include the W boson helicity fractions in top quark decays, spin correlations in the $t\bar{t}$ system [33], quantum entanglement in $t\bar{t}$ decays [34], testing the lepton universality in $t\bar{t}$ decays [35, 36] and more. In this sense, the top quark is a both QCD and EW laboratory of the SM.

3.8.4 Rôle in the electroweak symmetry breaking

The top quark is the only fermion with mass close to the vacuum-expectation value of the Higgs field. This not only suggests possible connections to the mechanism of the spontaneous electroweak symmetry breaking, but possibly also to new physics.

Top top quark large mass also amounts to important corrections to the Higgs boson mass via loop diagrams. Similarly, virtual corrections with a top quark contribute largely to rare decays of mesons, forbidden at a tree-level.

Corrections to the Higgs boson potential lead to the question of the EW vacuum stability [37] with the important rôle of the precise knowledge of the top quark mass.

Other experimental searches include the CP violation in top quark production or decays (not observed yet [1], Section 13 and *e.g.* [38]) or coupling to Higgs [39], searches for anomalous flavour changing neutral currents in top quark decays (*e.g.* by searching for the $tq\gamma$ vertex) or in coupling to the Higgs boson by checking the tHq coupling [40, 41]. Operators involving the top quark are an active field of study within effective field theories (EFT) [42].

3.8.5 Top quark as a calibration tool

The presence of the b-jets in the $t\bar{t}$ decay allows experimentalists to calibrate the efficiency of the b-jets identification (b-tagging) using data. The efficiency can be compared to the one in simulation, and an important scale-factor (the ratio between the data and simulation efficiencies) can be extracted to be applied to other $t\bar{t}$ decay channels or other processes involving top quarks or b-jets, in order to minimize the uncertainty in the description of this efficiency in simulation. This leads to more precise measurements and more stringent search results.

3.8.6 Top quark as background

Top quark is a background to many physics processes at the LHC, from rare SM processes like $t\bar{t}H$ or $t\bar{t}b\bar{b}$ productions, to searches for new physics.

3.9 Higher orders and interference with single top final states

Beyond the leading order (LO) diagrams of the $t\bar{t}$ production leading to the $WbWb$ final states, the situation becomes more complex as other processes start to interfere with the $t\bar{t}$ production diagrams, see Figure 3.4 for example of Feynman diagrams of the processes.

In the leading-order picture, the single top production is physics background to the $t\bar{t}$ production, and the $t\bar{t}$ process is a physics background to the single top production. But in fact, the differentiation between $t\bar{t}$ and singly-produced top quarks is a construct valid only at the LO of the perturbation theory, while the addition of higher-order processes allows to discuss single, double and non-resonant production of $\bar{b}W^+bW^-$.

Thus the naïve experimentalists picture of Feynman diagrams being somehow the way the process happen is not valid as interference effects are actually in place. Asking what way the $\bar{b}W^+bW^-$ final state was reached is as good as the question what slit a photon traversed in the double-slit scattering experiment. Nevertheless, by selecting events with different kinematics, phase-space enriched in processes close to a square of a single amplitude (*e.g.* of the doubly-resonant $t\bar{t}$ diagram) can be obtained. In practice, the usage of separately generated $t\bar{t}$ and single top processes means an interference term must be added to either of the two, or neglected, in which case an estimate or a systematic uncertainty related to its removal must be evaluated. This procedure has been termed as diagram removal (DR) or diagram subtraction (DS) [43] techniques.

3.9.1 Driving the experiment

The advances in the field of accelerating techniques have enabled experimental particle physicists to look for unstable particles of large masses, not present in ordinary matter around us. These particles have played important roles in the expansion of the Universe at its very early stages. Current and recent hadron colliders likes the Large Hadron Collider (LHC) at CERN or Tevatron at Fermilab have contributed to key discoveries of the particle content, unveiling details of particles interactions, using versatile and complex experiments like those at CERN (ATLAS, CMS, ALICE and LHCb at the LHC, and ALEPH, DELPHI, Opal and L3 at the Large Electron Positron collider, LEP) or at Fermilab (the CDF and DØ experiments). These modern particle experiments are the result of an intense development of detection techniques, each collaboration consisting of hundreds and thousands of researches from around the world.

Top quark turned out to be so heavy that it had eluded discovery on many accelerators until the Tevatron era. In fact, the expected dominant decay to a W boson and a b quark has driven the progress in the development of silicon tracking detectors which allow the reconstruction of the displaced secondary vertices of the decays of long lived B-hadrons. The story of the top quark is thus a fascinating interplay of developments in the theory, phenomenology, analysis methods and detector techniques.

3.10 Discovery of the top quark at Tevatron

The first super-conducting-magnets synchrotron Tevatron was build in 1980s with the main motivation to reach the top quark and the Higgs boson discoveries. During the Run I programme of the accelerator at the centre-of-mass energy of $p\bar{p}$ collisions

of 1.8 TeV, the top quark was observed by both the CDF and DØ experiments in 1995 [44, 45], completing the picture of 6 quarks as elementary matter constituents as for hadrons.

The Monte Carlo generator used at that time was ISAJET [46]. The expected dominant decay of the top quark $t \to W^+ b$ had driven the development of highly pixelized silicon based solid state tracking detectors with high spacial resolution in order to identify secondary decays vertices from the decay of the relatively long-lived B mesons, with $c\tau$ distance factors of about 0.5 mm. This approach was adopted by the CDF collaboration already in its Run I design, where a-four-layer silicon strip tracking and vertexing [47]) detector was immersed in a central-solenoid magnetic field on the axis of the multi-wire central tracking chamber. A 4.8σ excess was observed, with a mass distribution shown in Figure 3.12, measuring the mass of the top quark to be 176 GeV.

In contrast, the DØ detector at Run I did not posses a central magnetic filed, but relied on dedicated muon chambers to tag semileptonic B decays, and included a compensating uranium-doped hadronic calorimeter for jets energy measurements. On the other hand, the DØ experiment employed a compensating uranium calorimeter and a precision muon system (allowing for soft muon tagging of b-jets in semileptonic B decays). The first top quark mass estimate as measured in the DØ discovery paper was almost 200 GeV, see Figure 3.13, observing 17 events with the expected background of about 4, constituting a 4.6σ excess.

During the Tevatron Run II period, with the increased centre-of-mass energy of $p\bar{p}$ collisions to 1.96 TeV, more detailed properties of the top quark were measured, such as precise measurement of its mass, but also observation of weak interaction induced production of single top quarks [48]. The single-top quark discovery employed one of the first massive usage of the technique of boosted decision trees in high energy physics (neural networks used already at Run I) allowing to identify signal-rich phase spaces and observing the processes, yield of which was otherwise smaller than the statistical uncertainty in the preselected number of events in data.

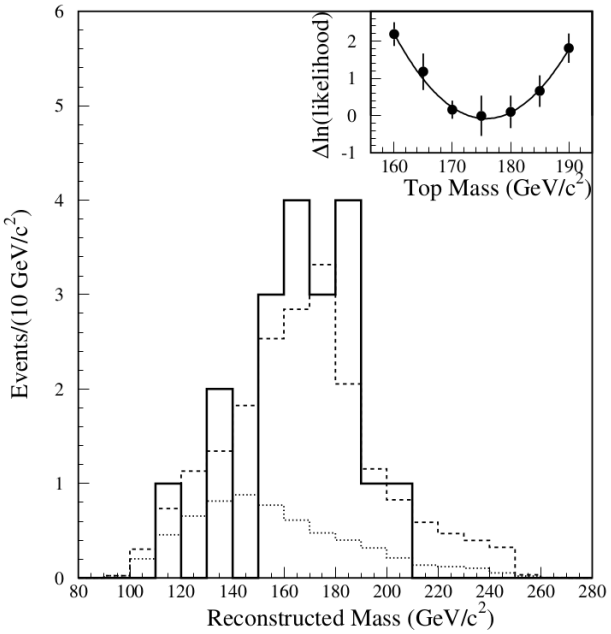


FIGURE 3.12: The reconstructed top quark mass as measured by the CDF experiment in the discovery paper [44], with the expected backgrounds and signal shapes determined by the dotted and dashed lines, respectively. The inset shows the negative likelihood fit to the top quark mass.

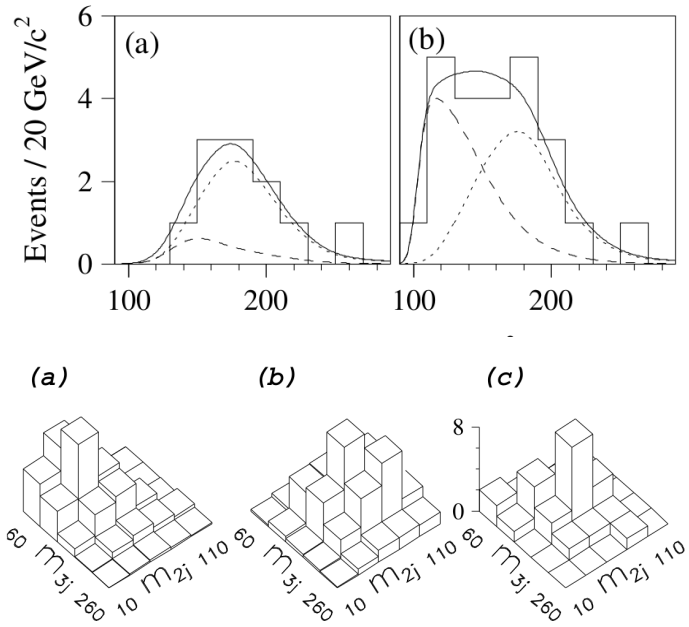


FIGURE 3.13: Reconstructed top quark mass as measured by the DØ experiment for a standard (top left) and loose (top right) selection, and the invariant masses of the two and three jets for background, signal and data events (bottom, from left to right), as published in the discovery publication [45].

Chapter 4

Experimental techniques

In this chapter we shall provide an exposure on simulation in high energy physics and on experimental techniques important for the study of differential spectra measurements in the $t\bar{t}$ system. Since the author is member of the experiment, this thesis will focused on the ATLAS detector as an example case of object identification, reconstruction and analysis methods. We finish with the description of the unfolding procedure as the key technique in correcting spectra for detector effects.

4.1 Simulations in high energy physics

4.1.1 Current picture of high energy physics

As described in Section 2.2.3, the current picture of high energy collisions of hadrons is based on the possibility to factorize perturbative (computable) and non-perturbative (hard to compute, often based on phenomenological models) parts of the problem, see illustration in Figure 4.1. This is closely related to the nature of the underlying theory of the strong interactions, the quantum chromodynamics, where quarks become asymptotically free at very high energies, *i.e.* short distances, while at low-momentum transfer, *i.e.* large distances, the running coupling constant of QCD grows, and quarks are confined in observable hadrons.

Consequently, also the simulation steps are split into the generation of events within the hard process, corresponding to a matrix element provided by the theory, followed by the parton shower, hadronization, decays and optionally the detector simulation.

4.1.2 Event generators

Generators of simulated events are an indispensable tool for both precision measurements and searches at particle colliders. They provide expected kinematic distributions, predict amounts and shapes of backgrounds, help to select events of interest, validate and design experimental methods, provide correction factors, measure selection efficiencies and more.

Such generators are often based on statistical methods of Monte Carlo sampling of a multidimensional phase space using random numbers.

Generators can be those providing the core matrix element (ME) and the corresponding hard process as function of kinematics of the outgoing primary particles, often partons, but also vector bosons W, Z, direct photons, the Higgs boson or hypothetical particles corresponding to processes beyond the Standard Model. Generators can also be specific in simulating and generating only additional QCD or QED shower processes in particles already generated by ME generators.

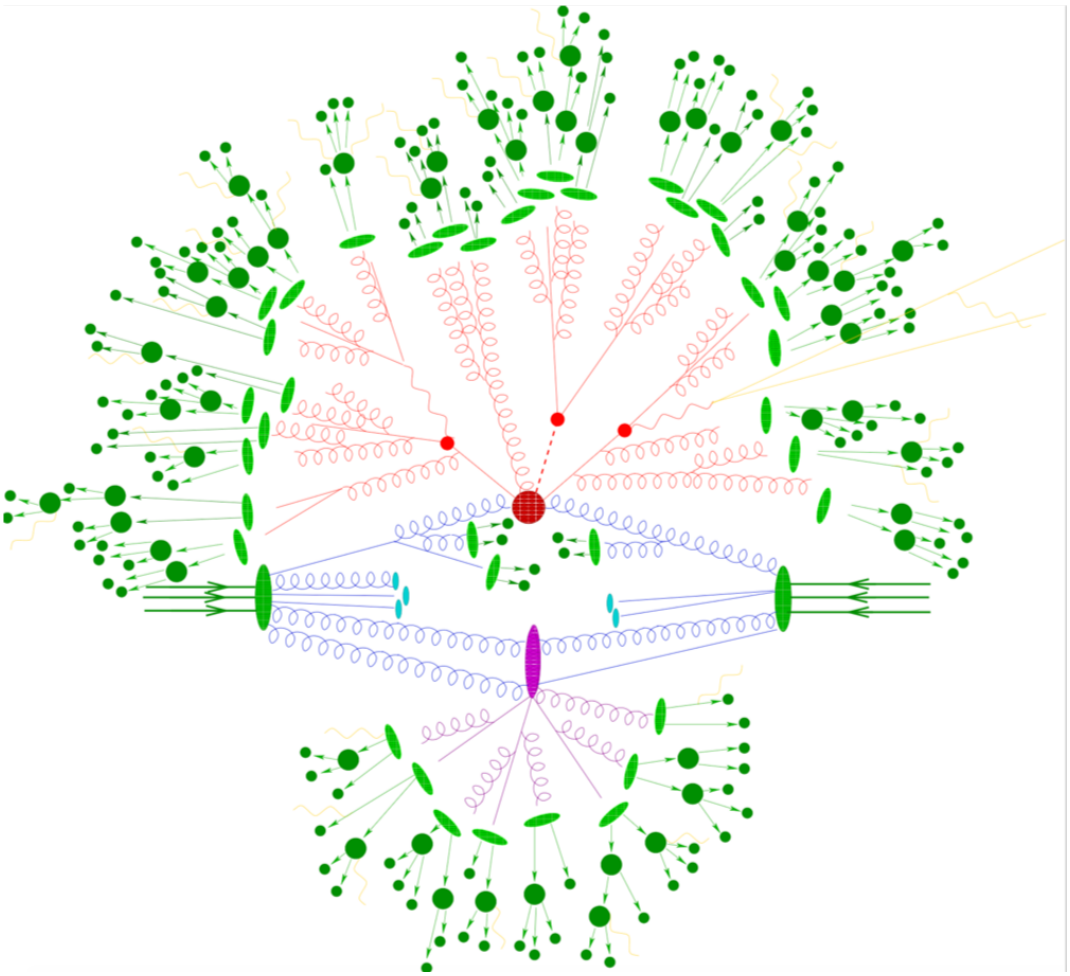


FIGURE 4.1: A cartoon of a high energy collision event by Sherpa [49, 50] authors, depicting the resolution of protons to partons via parton distribution functions (green large ellipses), the hard process stage (dark red circle), underlying event from initial hadron remnants (light blue), multiple parton scattering (purple), initial and final state radiation, showering, primary partons decay (light red), hadronization (light green small ellipses) and further particles decay (dark green).

An overview of the important features of current HEP event generators can be found *e.g.* in [51]. The PYTHIA8 generator [52] is a leading-order (LO) generator, *i.e.* providing physics amplitudes and cross sections at the tree-level of Feynman diagrams, and providing a model of a parton shower (PS). It is thus often combined with more developed NLO generators and used as solely the PS generator. Another LO generator with a different PS model is HERWIG7 [53, 54]. Generators for the hard process most used in practice have been *e.g.* the LO-multileg generator ALPGEN [55], the NLO-precision POWHEG [56, 57, 58] and another NLO one MADGRAPH5 [59] (also known as AMC@NLO).

The parton shower generators partially encompass effects of additional orders in QCD by developing the partons produced by the ME via a shower of collinear and soft partons where the QCD is perturbative. Such a region of phase space of the additional radiated partons is preferred by the nature of QCD. In the shower development, the splitting kernels of QCD are used to define Sudakov form factors

which stand for the probability of no emission of an additional parton within a given transverse momentum or virtuality range. Effectively, the exponential form of the Sudakov form factor stands for the probability of no emission of any order, thus using the factor one effectively sums up to all orders in this soft and collinear regime. In essence, the Sudakov factor is an exponential of the probability of no parton emission which we have encountered along with the splitting kernels (see Section 2.2.4). The infinitesimal probability for no parton emission with $k_T \in (Q_1, Q_2)$ is then

$$\mathrm{d}P_{\text{no emission}} \approx 1 - \int_{Q_1^2}^{Q_2^2} \frac{\mathrm{d}k_T^2}{k_T^2} \int_{z_1}^{z_2} \mathrm{d}z \, \frac{\alpha_S(k_T)}{\pi} \, P(z) \,,$$

leading to the exponential of the form

$$\mathrm{d}P_{\text{no emission}} \equiv \Delta(Q_1, Q_2, z_1, z_2) = \exp\left[- \int_{Q_1^2}^{Q_2^2} \frac{\mathrm{d}k_T^2}{k_T^2} \int_{z_1}^{z_2} \mathrm{d}z \, \frac{\alpha_S(k_T)}{\pi} \, P(z) \right] .$$

The derivative of the Sudakov factor is the probability distribution for the radiated parton momentum, and most of the MC generators proceed by using it to choose randomly the hardest emission, followed by the PS to smaller angles and/or energies. The two main PS generators, HERWIG and PYTHIA, consider angular and p_T-ordered showers, respectively. The p_T-ordered shower is also the case for the SHERPA generator [60] with its own PS model while the Vincia PS generator [61] is a relatively recent mixed approach.

Matching of matrix elements to parton showers

Beyond the LO, a problem arises with combining fixed-order matrix element calculations with parton showers. In order to avoid double counting or empty parts of the phase space in terms of the extra jets being produces as hard emissions in the ME or via the PS, a matching scheme must be applies, see *e.g.* [62, 63] for an overview, or the two main techniques dubbed as MLM [55] and CKKW [64].

An important consequence of the matching procedure are negative or zero event weights which compensate for the double counting of events in the phase space where an additional parton can be generated via the ME or by the PS.

The already 10 years long standing NLO matching [65] is being gradually replaced by emerging new methods allowing matching of NNLO pQCD matrix elements with PS, leading to NNLO generators like MiNNLO$_{\text{PS}}$ [66].

The two main NLO generators mostly used these days for LHC experiments are the AMC@NLO and POWHEG.

The AMC@NLO generator resolves the overlap of events in the phase space by setting negative weights to events where a hard emission is further showered to a point which would lead to a double counting. This is a known feature as the NLO cross-sections are not expected to be positive-definite in the phase space locally. In practice, one has to consider longer convergence to the result of desired statistical uncertainty.

The POWHEG method does not incorporate the negative weights. For generating and matching the NLO ME with PS it features the h_{damp} parameter, which needs to

be tuned, and which effectively controls the hard emission by a damping factor

$$\frac{h_{\text{damp}}^2}{h_{\text{damp}}^2 + p_{\text{T}}^2}$$

which is used to suppress the extra hard high-p_{T} emission of a parton. The h_{damp} value of ∞ means no suppression and should lead to a result closer to NNLO one while a finite value regularizes the high-p_{T} making it closer to the NLO. Various values of the h_{damp} parameter have been used in the case of modelling the $t\bar{t}$ production, with experimental data preferring finite values of the order of the mass of the top quark.

4.2 Detector simulation

It is not within a scope of this document to provide details on the simulations of current HEP detectors, we shall only briefly discuss some of their key aspects and requirements.

Detailed simulations of particles passing active as well as passive components of individual particle detectors are an important tool used in the design process of building complex particle detectors. Requirements have to be met on the angular, spacial, momentum and energy resolutions, objects identification efficiencies and purities, radiation hardness, time resolution *etc*.

The current software packages which are mostly used and which encompass our best knowledge of the detailed processes of particles interaction with matter are the GEANT4 [67, 68, 69] and FLUKA [70, 71] packages. They contain parametrizations of relevant processes ranging from nuclear interactions to ionization, bremsstrahlung, Cherenkov and transition radiation, photon-specific processes and more. These tools find their applications also in medical physics, astroparticle physics, cosmic research and industry.

It is thus clear that the spectra as obtained using realistic detectors are measured on a limited phase-space and more importantly they are "smeared", *i.e.* affected by the finite detector resolution. It is the task of the unfolding procedure to correct such detector-level spectra for these effects in order to obtain ideally an experiment-independent result which can be compared between experiments and to theory predictions of today and to come.

More on the infrastructure related to a detailed ATLAS simulation can be found *e.g.* in [72] for the full simulation or in [73] for the parameterized fast simulation with simplified calorimeter shower response simulation.

For even faster studies, a simplified SW package for a fully parameterized detector simulation DELPHES [74] has served the HEP community for fast phenomenological or detector development studies with the ability to simulate approximately current, past or future detectors based on their publicly available measured or expected object resolution and identification efficiencies. The framework enables one *e.g.* to roughly compare ATLAS and CMS expected performance and more, via predefined detector "card files". As an input, only the simulated particle-level information in the HEPMC format [75] is required, allowing one to study any SM or BSM physics process and its phenomenology, detector signature, selection efficiency optimization, ability to kinematically reconstruct a resonance of interest *etc*.

4.3 Objects identification in current HEP detectors

A typical high energy physics multipurpose detector consists of layers of sub-detectors targeting identification of different objects. These are, starting radially from the interaction point, tracks of charged particles, their reconstruction, charge determination, mass identification and momentum measurement as provided at ATLAS by the Inner Detector (ID). Next is the energy measurement in electromagnetic (EM) and hadronic (HAD) calorimeters, followed by dedicated devices for muon identification and momentum measurement. This general concept is depicted in Figure 4.2 on the example sketch of how the ATLAS experiment identifies objects in different dedicated parts of the detector.

Combining measurements from the sub-detectors, objects identification is possible. Photons leave isolated energy deposits in the EM calorimeter, and should not contain angularly close hadronic activity or tracks. Electrons are identified as EM clusters with a matched track from the ID. Muons muons penetrate most of detector layers with only small energy deposits even in calorimeters, leaving ideally an isolated signal in dedicated muon spectrometers (MS), with the MS track matching a one in the ID. The increasingly populated and studied high top quark p_{T} region has lead to the need for p_{T}-dependent isolated lepton definition, resulting in so-called gradient isolation [76, 77].

Neutrinos are observed only indirectly as the missing transverse energy or momentum, defined as the negative sum of energies or momentum in the transverse direction. They are an important signature especially of leptonic W decays.

Hadronic final states, jets, are the consequence of the QCD where partons readily radiate soft (low energy) and collinear (w.r.t. the original parton directions) partons. As one cannot experimentally distinguish energy deposits in the calorimeter coming from a single energetic parton or from a collimated collection of partons, the right observables are jets, *i.e.* cross-sections related to their spectrum and multiplicity, not to that of partons.

Jets are reconstructed by a clustering algorithm from jet constituents, stable particles at the particle level, calorimeter clusters or a combination of tracking and calorimeter information (particle flow jets). The algorithm clusters close particles according to some metrics. Older and theoretically unstable algorithms used to cluster particles based on their angular separation. Current widely used (anti)k_t algorithms use both angles as well as relative transverse momenta of the constituents, which are gradually merged until a jet is declared.

For precision measurements, the critical task is the calibration of the reconstructed jet energies and masses to the particle level. The complex procedure is described in [78], starting with the correction for additional pp interactions (pile-up) bringing energy to the jet in form of an offset, continuing with η dependent calibrations, the main response correction and finally data-to-simulation in situ corrections.

At current large LHC experiments both small and large radius jets are used, aiming to capture jets from QCD processes or EW decays, or jets coming from hadronic decays of massive particles with a large transverse momentum (boosted jets). Mass of the large jets then highly corresponds to the mas of the primary particle, and the large jet substructure also reflects the details of the decay, *e.g.* the number of jets should be closely correlated to the number of hadronic decay products of the particle. Dedicated W, Z, t and H taggers target the decays $W \to q\bar{q}'$, $Z \to q\bar{q}$, $t \to bW \to bq\bar{q}'$ and $H \to b\bar{b}$ etc.

The crucial technique in the top quark physics is the b-jets tagging, *i.e.* identification of jets as coming from the decay of B-hadrons which contain the b-quarks

dominantly from the top quark decays. Numerous techniques employ the expected signature of a displaced vertex due to decays of long-lived B-hadrons, but also other track and vertex properties, often combined to a multivariate or machine learning-based classifier, see *e.g.* [79]. Typically one needs to optimize between the desired large real and low fake b-jet tagging efficiencies.

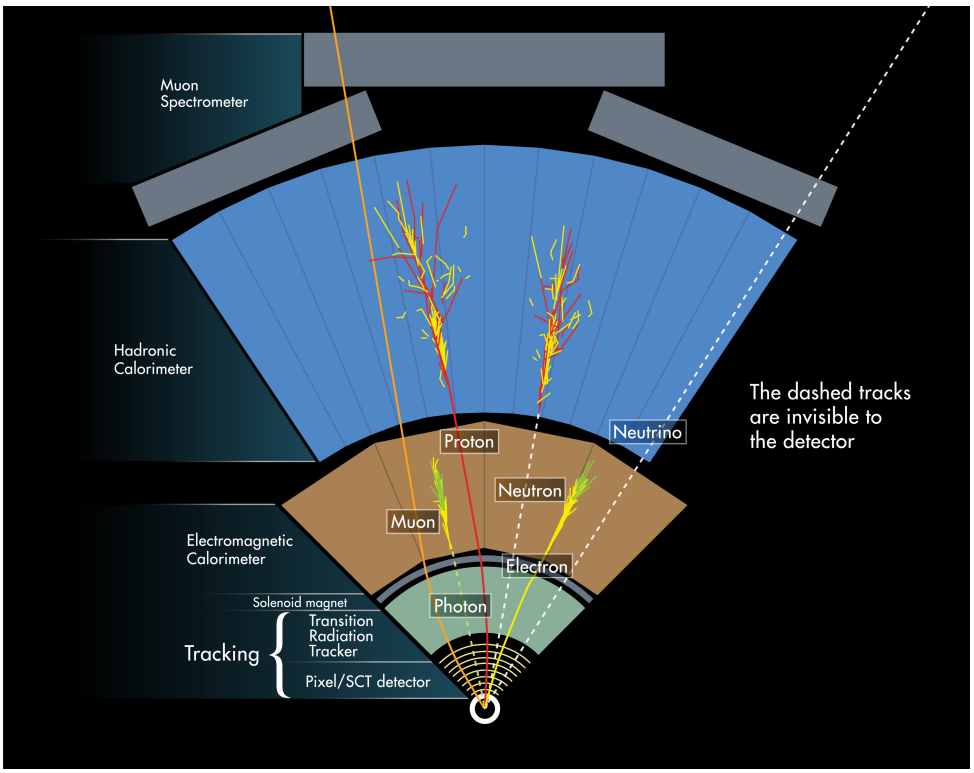


FIGURE 4.2: A sketch of identification principles of particles in the ATLAS experiment [80], starting radially with tracking detectors, followed by electromagnetic and hadron calorimeters and finally the muon chambers, with particles of different kinds leaving signals in specific sub-detectors.

4.4 Statistical methods

4.4.1 Cross-section combination

In the 7 TeV ATLAS parton-level analysis [ATLAS2], the Best Linear Unbiased Estimator (BLUE [81], for a review see [82]) method was used for the combination of cross-sections in the e+jets and μ+jets channels. For this purpose, a covariance matrix between the two channels must be formed, taking into account the correlated or uncorrelated systematic uncertainties between the channels, all evaluated after the unfolding. Absolute cross-sections have been combined, with the relative (normalized) computed after the combination. The caveat is that for largely correlated measurements and combinations of the absolute cross-sections, the covariance matrix is dominated by large systematic uncertainties and the matrix inversion can be

problematic. In this way, covariance matrices corresponding to correlated and uncorrelated uncertainties are build as follows

$$\mathrm{Cov}_{\mathrm{corr}} \equiv \begin{pmatrix} \sigma_e^2 & \sigma_e\sigma_\mu \\ \sigma_e\sigma_\mu & \sigma_\mu^2 \end{pmatrix} \quad \mathrm{Cov}_{\mathrm{uncorr}} \equiv \begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_\mu^2 \end{pmatrix}, \tag{4.1}$$

and their sum defines the total covariance, a needed input into the BLUE method. The BLUE then defines the channel weights for their combination as a row vector

$$\mathbf{w} \equiv \left[\mathcal{U}^\top \mathrm{Cov}_{\mathrm{total}}^{-1} \mathcal{U}\right]^{-1} \mathcal{U}^\top \mathrm{Cov}_{\mathrm{total}}^{-1} \tag{4.2}$$

with $\mathcal{U}$ being the unit row vector, allowing to define the resulting combined quantity (number of events or cross-section) in each bin x_j and its uncertainty δx_j as

$$x_j \equiv \mathbf{w}_j \cdot \begin{pmatrix} x_j^e \\ x_j^\mu \end{pmatrix}, \quad \text{and} \quad \delta x_j \equiv \sqrt{\mathbf{w}_j \cdot \mathrm{Cov}_{\mathrm{total}}^j \cdot \mathbf{w}_j^\top}. \tag{4.3}$$

4.4.2 Cross-sections ratio

An interesting question arises in measurements of a ratio of two cross sections. The application can be a cross-section ratio for a given process at different collider energies, a ratio between cross sections of two different processes at the same collider energy and data set, or a ratio of a cross section of a given process in nuclear (AA or pA) collisions to the one in a reference pp environment. In all these cases, the benefit is a partial cancellation of systematic uncertainties which can be, depending on the case, the luminosity uncertainties, theory or modelling uncertainties or systematic uncertainties if evaluated using the same or similar methods. For a ratio of two quantities $z = \frac{x}{y}$ the uncertainties propagation can be expressed as split in those uncorrelated or partially correlated, ρ being an additional or effective correlation coefficient between the systematics in x and y, as

$$\frac{\sigma_z^2}{z^2} = - 2\rho \left(\frac{\sigma_x^{\mathrm{syst,corr}}}{x}\right)\left(\frac{\sigma_y^{\mathrm{syst,corr}}}{y}\right) + \left(\frac{\sigma_x^{\mathrm{syst,uncorr}}}{x}\right)^2 + \left(\frac{\sigma_y^{\mathrm{syst,uncorr}}}{y}\right)^2 \tag{4.4}$$

$$+ \left(\frac{\sigma_x^{\mathrm{stat}}}{x}\right)^2 + \left(\frac{\sigma_y^{\mathrm{stat}}}{y}\right)^2 + \left(\frac{\sigma_x^{\mathrm{syst,corr}}}{x}\right)^2 + \left(\frac{\sigma_y^{\mathrm{syst,corr}}}{y}\right)^2. \tag{4.5}$$

This formula can be verified using "toy" experiments, in which x_i and y_i are computed in each i-th experiment by smearing the central values x_0 and y_0 by random numbers drawn from Gaussian distributions as follows:

- for uncorrelated uncertainty sources like the statistical or uncorrelated systematics, different random numbers are drawn for x and y;
- for correlated uncertainties, a single random number is drawn to smear x and y coherently.

This can be expressed as follows, with λ_i being random numbers drawn from standard normal distribution, *i.e.* $\lambda \sim \mathcal{G}(\lambda|0,1)$

$$x_i = x_0 + \sum_j \lambda_j^{(x)} \sigma_j^{\text{x,uncorr}} + \sum_k \lambda_k \sigma_k^{\text{x,corr}} \tag{4.6}$$

$$y_i = y_0 + \sum_j \lambda_j^{(y)} \sigma_j^{\text{y,uncorr}} + \sum_k \lambda_k \sigma_k^{\text{y,corr}} . \tag{4.7}$$

The event-by-event ratio $z_i = x_i/y_i$ is then computed in each pseudo-experiment, an histogram of z is filled and its width extracted and compared to the standard deviation computed using the analytical formula. This test is illustrated in Figure 4.3 where the "toy" and analytical-based uncertainties in z are plotted in the same picture, as a function of the weight (a fraction) given to the correlated systematic uncertainty. As expected, larger fraction of uncertainties treated as correlated leads to their partial cancellation in the ratio and a small total uncertainty in the ratio.

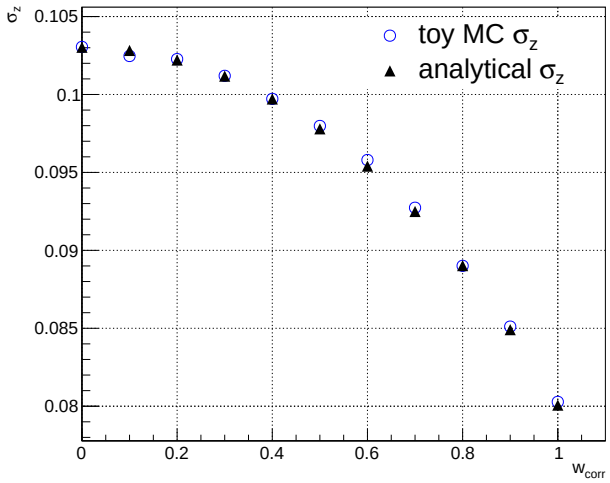


FIGURE 4.3: Evolution of the total uncertainty in a model ratio variable depending on the weight (fraction) of correlated uncertainties in the numerator and denominator, evaluated using 100k statistical toys (open circles) compared to the result of an analytical formula (filled triangles).

4.4.3 Binned likelihood fit of the signal strength

A fit to extract the amount of a signal process in terms of its yield, cross-section or the signal strength (see below) can be done using a binned sensitive variable, *i.e.* an histogram where the data distribution is compared to the prediction composed of usually several background sources and a signal template. Background templates can be of simulation or data-driven origin, or a mixture or a combination of both, *e.g.* taking a shape of a background from simulation while its normalization adjusted in a control selection region (CR) in data, or data from a control region scaled using simulation or some transport procedure to predict the yield of the background in the signal selection region (SR) ideally dominated by signal, or at least with a non-trivial signal-to-background ratio. The sensitive observable should have power to

separate signal and background. It can be a single kinematic variable, a function of several, or an output from some multivariate discriminant.

One is often interested in a single physics parameter of interest, the signal strength μ, a single parameter scaling coherently the amount of signal w.r.t. *e.g.* the SM expectation value in all bins of the prediction. The likelihood function is usually constructed as follows

$$\mathcal{L}(\mu, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \mathbf{d}, \mathbf{S}, \mathbf{B}) \equiv \prod_{i=1}^{N_{\text{bins}}} \mathcal{P}(d_i|\mu S_i + B_i) \cdot \mathcal{P}(h_i|\gamma_i h_i) \cdot \prod_{j=0}^{N_{\boldsymbol{\theta}}} \mathcal{G}(\theta_j|0, \sigma_j)\,, \tag{4.8}$$

where

- $\mu = \frac{\sigma_{\text{observed}}}{\sigma_{\text{theory}}}$ is the signal strength;
- $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ are vectors of additional ("nuisance") parameters (NPs, representing systematic uncertainties shifts) and their uncertainties;
- $\boldsymbol{\gamma}$ are bin-by-bin scaling factors allowing to let the total prediction $S_i + B_i$ in each bin fluctuate according to the Poisson distribution with the statistical uncertainty modified by the γ factor;
- $\mathbf{d}$; $\mathbf{S}$ and $\mathbf{B}$ are the binned data; and signal and background binned prediction templates; and $h_i = S_i + B_i$;
- $\mathcal{P}(k|\nu)$ and $\mathcal{G}(x|x_0, \sigma)$ are the Poisson and Gauss probability distribution functions, respectively.

This likelihood function is simultaneously minimized for the best signal strength $\hat{\mu} \equiv \mu_{\text{observed}}$ as well as for the shifts of the nuisance parameters $\boldsymbol{\theta}$, allowing to change *e.g.* some background normalization and/or, in a more involved way, modify also shape variations of both background and signal *e.g.* due to variations of the jet energy scale or other detector, calibration or modelling sources. The γ factor allow the prediction in each bin vary according to its statistical fluctuations.

The fitted signal strength is then used to calculated the fitted cross section, calculated as $\sigma_{\text{observed}} = \mu\, \sigma_{\text{theory}}$. In principle, any σ_{theory} would be possible to be used in the procedure, resulting, in case of a stable fit, in a different fitted μ but still leading to the same extracted observed cross-section.

4.4.4 Absolute and normalized cross-sections

Here we will go through the advantages and disadvantages of measuring the absolute or relative spectra in terms of the physics message and presentation of the results.

In case of the relative cross-sections, systematics correlated across the bins of variable X cancel with the overall systematics shift of the measured integrated cross-section, leading to a more precise shape measurement.

An issue arises when trying to perform a χ^2 test between the measured and theory cross sections in the case of normalized spectra. The normalization condition removes one degree of freedom and the following reduced expression needs to be computed

$$\chi^2 = (\mathbf{d} - \mathbf{t})^{\text{T}}_{N_{\text{bins}}-1} \cdot \text{Cov}^{-1}_{N_{\text{bins}}-1} \cdot (\mathbf{d} - \mathbf{t})_{N_{\text{bins}}-1}\,, \tag{4.9}$$

where $\mathbf{d}$ and $\mathbf{t}$ are the number of events in data and theory in given bin and Cov is the combined measurement and theory covariance matrix, allowing for uncertainties in both be correlated across bins. The subscript $N_{\text{bins}} - 1$ denotes that a single

row of an arbitrary index and the corresponding column of the same index in the $\mathrm{Cov}_{N_{\mathrm{bins}}-1}$ matrix is removed, leading to a $(N_{\mathrm{bins}}-1)\times(N_{\mathrm{bins}}-1)$ sub-matrix derived from the full covariance matrix, the same performed for the row or column element of the data-theory difference vector or its transpose. Such sub-matrix of the covariance matrix is invertible, resulting in a meaningful χ^2 be computed which does not depend on the choice of the discarding index.

However, the overall cross-section still caries an important information and should also be extracted, in order to compare to different generators or to predict cross-section magnitude in various kinematic regions for other analyses using this particular final state.

4.4.5 Propagation of systematic uncertainties

Systematic uncertainties influence the modelling of the backgrounds, properties of reconstructed objects (energies, momenta and angles of leptons jets etc.), and modelling of the $t\bar{t}$ signal which is needed to derive corrections applied to data. All this affects the measured quantity, be it the integrated cross section or differential spectra.

Four-momenta of objects of physics candidates on leptons or jets are typically calibrated and while the absolute energy and mass scales of objects are highly desired, it is in fact the difference of objects properties between data and simulation which is more important and desired to be minimal. In short, kinematic distributions in simulated events at the detector level and in real data can be off w.r.t. particle level in terms of an absolute mass scale, but the data ratio to the prediction at the detector level should be close to unity, signifying that the two "worlds" of that of data and simulation are at the same footing.

In high energy physics, the effects of systematic uncertainties are propagated through the full chain of the analysis. In case of an absolute calibration of object properties, an alternative set of events with altered corrections applied is created, and the effect is propagated to the distributions of interest and finally to unfolded observables, extracted cross-sections etc.

The procedure of systematics evaluation is performed either by changing data or by altering the simulation, usually, the latter.

In unfolding, two choices exist as for propagating systematic uncertainties through the full unfolding procedure. One can keep the data distribution and unfold it with varied corrections, including the 1D ones like acceptance and efficiency, but also the migration matrix. Or one can keep the corrections fixed to their best known nominal versions and alter the unfolding input, which can be systematically varied data or simulation.

Statistical disadvantage of the first propagation model done by varying the corrections is clear especially in the case of smearing the 2D migration matrix, where large fluctuations in low populated regions can lead to unstable results. It is thus usually preferred to evaluate each systematic by unfolding varied, and usually simulated, spectra, by nominal corrections, and assign the resulting relative uncertainty to the unfolded data. The sum-in-quadrature of the independent statistical variations then defines the total systematic uncertainty.

Alternatively, one can also generate pseudo-experiments for unfolding, correlated over the kinematic bins, by creating typically hundreds of spectra with random admixture of systematics. By drawing for each systematics shape variation source "syst" a random number $\lambda_{\mathrm{syst}} \sim \mathcal{G}(\lambda|0,1)$ one can create, in the j-th bin of each of the detector-level spectrum R, an i-th pseudo-spectrum by combining contribution

from every systematic source as

$$R_j^{(i)} = R_j \sum_{\text{syst}} \left(1 + \lambda_{\text{syst}}^{(i)} S_j\right)$$

where S is the relative systematic shape change w.r.t. the nominal shape of the spectrum. This way, correlations across bins are preserved.

On the other hand, statistical uncertainties due to finite MC samples size are best evaluated coherently by smearing the migration matrix, resulting in coherent fluctuations in projections to the particle and detector axes which are then used to define the acceptance and efficiency corrections.

In the evaluation of individual systematics uncertainties in the unfolded distribution, it is often not reliable to propagate the single systematics source through the full unfolding machinery. This is because the random admixture of systematics described above may behave differently to the case of assuming only one systematics each time. Therefore, a subtraction technique is applied in such cases (and also in likelihood fits with systematics profiling) where the uncertainty is evaluated first for the case of allowing all systematics in the procedure and then for the case of removing a single particular one. The effect of the single systematics is then defined by a subtraction in quadrature of the uncertainty, in the case of unfolding in each bin j, as

$$\sigma_{\text{syst}} \equiv \sqrt{\sigma_{\text{all}}^2 - \sigma_{\text{all}\neg\,\text{syst}}^2}$$

and similarly to define a covariance matrix accounting for individual systematics uncertainties for the BLUE method to combine measurements.

4.4.6 Statistically correlated spectra

A technique of bootstrapping has been actively applied at the ATLAS experiment according to [83] in order to construct statistically-varied versions (replicas) of histograms of any dimension, by repeatedly filling an arbitrary number (typically 100–1000) of replicas by the same events but weighted by random weights drawn from the Poisson distribution of mean of 1, *i.e.* as $w \sim \text{Poisson}(1)$. This means that each event is used on average once in each replica, but allows for a multiple, yet weighted, spectra be generated, in order to study effects of statistical fluctuations based directly on the measured data spectra, without the need for extensive MC simulations. Setting the initial random seed to a defined combination of the run and event number of the collision event, one can create statistically correlated replicas over the replica number, over spectra, or even across analyses. Similarly, the bootstrap technique applied to MC events allows to determine the statistical significance of systematic uncertainties and allow one to possibly drop or smoothen them.

4.4.7 The BUMPHUNTER algorithm

The BUMPHUNTER algorithm [84] is a statistically robust model-independent search for a bump in a spectrum, a possible sign of the presence of a new physics or a deviation from the prediction of a finite accuracy. It evaluates the probability of the observed spectrum of not containing a bump in a bin range of a selected window size of a 1D distribution. More specifically, the algorithm finds the interval of the largest incompatibility of data with the background-only hypothesis. The advantage of the algorithm is that it does not rely on pseudo-experiments to determine the probability

density function for the background hypothesis in each bin but uses an analytical formula to compute the test statistics, independently of a signal model.

The BUMPHUNTER (BH) probability of the data compatibility with the background-only hypothesis (p-value, p_{val}) is the smallest, by constructing BH tests statistics

$$t \equiv -\log p_{\text{val}}^{\text{min}}$$

which grows with the discrepancy. In detail, the p-value for the given area, standing for the probability of observing greater or equal excess of the data events above the expected background, can be computed using the normalized incomplete gamma function as

- $p_{\text{val}} \equiv \Gamma(d, b)$ for the case when the number of data events d is larger than the expected number of background events b;
- $p_{\text{val}} \equiv 1 - \Gamma(d+1, b)$ otherwise;

where

$$\Gamma(x, y) \equiv \frac{1}{\Gamma(x)} \int_0^y \zeta^{x-1} \mathrm{e}^{-\zeta} \, \mathrm{d}\zeta \,, \quad \Gamma(x) \equiv \int_0^\infty \zeta^{x-1} \mathrm{e}^{-\zeta} \, \mathrm{d}\zeta \,. \tag{4.10}$$

The function appears in the replacement for the sum of Poisson weight factors determining the probability of observing d events or more as the sum of Poisson terms for a distribution with mean b as

$$\sum_{k=d}^{\infty} \frac{b^k}{k!} \mathrm{e}^{-b} = \Gamma(d, b) \,. \tag{4.11}$$

We shall come back to the application of the BH test statistics in Section 8.2.2.

4.5 Unfolding

Unfolding is the procedure of correcting the observed spectra at the detector level for a finite resolution and efficiency effects. The procedure yields a detector-independent result usually at the particle or parton level, *i.e.* at a defined theoretical stage (after or before hadronization) and phase-space (often called fiducial). The measured spectra are affected not only by random fluctuations (smearing) of measured energy, momentum and angles of objects used to define a particular kinematic observable, but also by combinatorial effects, where, at the detector level, objects not originating from the top quark decay are selected and picked for the top quark four-momentum estimate, or objects are wrongly assigned for reconstruction of different parts of the final state.

We will focus on methods of reconstructing top quark kinematics in $t\bar{t}$ events in the ℓ+jets channel, leaving out dedicated methods of top quark reconstruction dealing with larger ambiguities. We will come back to the all-hadronic decay modes later in Chapter 8.

The study of extending the Fully Bayesian Unfolding with the regularization option has been performed elsewhere [Auth2]. Here, we describe the basic concepts of the unfolding in the context of physics spectra as used in the ATLAS analyses presented in Chapter 7.

4.5.1 Parton level

In top quark physics, unfolding to the parton level is motivated by the fact that the top quark is very heavy and its fast decay prevents it from hadronization, as well it suppresses QCD radiation from top quarks. The sum of four-momenta of top quark decay products observed in detector should then closely correspond to the original parton top quark momentum. Originally at the DØ experiment and then also at the CMS experiment, an iterative χ^2-based fitting algorithm *HitFit* has been used to find the best jet permutations and to fit the $t\bar{t}$ event kinematics. The constrain is a χ^2-like expression

$$\chi^2 \equiv \frac{(m_{\ell\nu b} - m_{jjb})^2}{\sigma_t^2} + \frac{(m_{jj} - m_W)^2}{\sigma_W^2} + \frac{(m_{\ell\nu} - m_W)^2}{\sigma_W^2} \,.$$

used also in top quark mass measurements, or, in order to improve the resolution of other observables, in a form which fixes the top quark mass to a known measured value using

$$\chi^2 \equiv \frac{(m_{jjb} - m_t)^2}{\sigma_t^2} + \frac{(m_{\ell\nu b} - m_t)^2}{\sigma_t^2} + \frac{(m_{jj} - m_W)^2}{\sigma_W^2} + \frac{(m_{\ell\nu} - m_W)^2}{\sigma_W^2}$$

The resolutions σ_t and σ_W are actually computed on an event-by-event basis from resolutions of individual objects which build up the top quark or W candidates four-vectors and these resolutions need to be extracted from simulation. This approach, however, relies on Gaussian resolutions of objects between the detector and the parton or particle level.

At the ATLAS experiment, a kinematic likelihood fitter (KLFitter, [85]) has been widely used to either fit the $t\bar{t}$ event kinematics starting from the detector level objects, or to select the jets permutations with the best resulting likelihood. The likelihood which is maximized in terms of the assumed partonic top kinematics is defined as

$$\begin{aligned} \mathscr{L}(\hat{E}^\ell, \hat{E}_i^{\mathrm{q}}, \hat{p}_{x,y,z}^{\nu} | E^\ell, E_i^{\mathrm{q}}, E_{x,y}^{\mathrm{miss}}) \equiv \mathrm{TF}(\hat{E}^\ell, E^\ell) \cdot \left(\prod_{i=1}^{4} \mathrm{TF}(\hat{E}_i^{\mathrm{jet}}, E_i^{\mathrm{q}}) \right) \\ \cdot \mathrm{TF}(E_x^{\mathrm{miss}} | \hat{p}_x^{\nu}) \cdot \mathrm{TF}(E_y^{\mathrm{miss}} | \hat{p}_y^{\nu}) \\ \cdot \mathrm{BW}(m_{jj} | m_W) \cdot \mathrm{BW}(m_{\ell\nu} | m_W) \\ \cdot \mathrm{BW}(m_{jjb} | m_t) \cdot \mathrm{BW}(m_{\ell\nu b} | m_t) \,, \end{aligned} \tag{4.12}$$

where TF are the transfer function, *i.e.* resolution functions indicating the detector smearing effects when going from the parton (quark q, lepton) to detector levels (jets, lepton) for objects energies and angles, and BW are the Breit-Wigner functions, standing for the expected resonant shapes of the masses of the W bosons and top quarks in the event. The top quark four-momenta can then be constructed either directly from the fitted object kinematics, or by using the fitted objects assignments only while using the original detector-level objects momenta to derive further kinematic variables. The disadvantage is the need for the knowledge of the parton-to-detector level transfer functions which may be model and matching dependent, and in order to use this approach possibly also at the particle level, a separate set of TFs needs to be derived.

In a similar manner, the Ideogram method [86] developped at the DØ [87] and further pursuit at the CMS experiment also relies on the expected Breit-Wigner resonant profiles of the masses of the two W boson and top quark candidates, assuming

separate shapes for the signal and background distributions of varied fractions.

The aforementioned likelihood fits have pioneered the simultaneous fits to the physics parameters of interest (like to mass of the top quark) together with "nuisance parameters" (NPs) which allow for shifts of systematics uncertainties. Originally, such an auxiliary (not of primary physics interest) parameter was a single additional jet energy scale factor which translated into the shift of the measured hadronic W boson and top quark masses. This technique has later proliferated into multidimensional profile likelihood fits with applications varying from the Higgs boson discovery to cross-section fits and event unfolding of differential spectra. The common feature is the possibility to diminish the total uncertainty in the physics parameter of interest by allowing systematic variations shifts of background or signal shapes to better describe the data.

Related to the nature of parton level measurements, however, the major question remains of what a top quark at the parton level is or how to define it across various Monte Carlo (MC) generators, theories at different pQCD precision *etc*. The efficiency of events to be reconstructed in the detector, related to the original full phase-space parton cross section, is very small, often at the level of a few per-cent, and thus the corresponding correction factor is huge. When evaluated for different $t\bar{t}$ models, the procedure leads to large systematic uncertainties, related to the still poorly known hadronization process.

4.5.2 Particle level

In contrast, unfolding to the particle level has been widely used in many SM analyses. The idea is that a parton is not a well-defined object in QCD for light quarks or gluons, partons radiate further partons, and the physically meaningful are predictions and measurements involving jets. Even parton-level jets are not a useful concept for comparison to experiment as hadronization of partons to observable particles (mesons, baryons, and leptons from their prompt decays) is still a strongly model-dependent process. One can then measure various cross-sections of various jets defined using stable particles which can enter the detector and response of which is better known. These can be *e.g.* jets from pure QCD processes, such which accompany electroweak bosons; jets from top quark or Higgs boson decays *etc*.

The problem of defining top quarks at the particle level is obvious: top quarks decay before hadronization, and it is their decay products leptons, jets and neutrinos which are formed and enter the detector. The idea of a pseudo-top quarks was developped at the ATLAS experiment to define objects (pseudo top quarks [88]) with a high correlation of their four-momenta to the kinematics of the original parton top quarks, yet using objects at the particle or detector levels.

In order to remove the combinatorial effects, a matching correction requiring a good angular correspondence of objects forming the pseudo-tops at the particle and detector levels can be applied, and the unfolding procedure then corrects mainly for the physics objects resolutions, meaning it deals with a more diagonal migration matrix.

4.5.3 The pseudo-top concept

In short, the pseudo-top algorithm defines the four-momenta of objects with a strong correlation to the original top quark partons. The primary focus is the $t\bar{t}\to\ell$+jets channel where one W boson decays hadronically and the other one leptonically. Such final state thus contains an isolated charged lepton (usually an electron or a

muon), neutrino (carrying away momentum, observed as missing transverse energy), at least four jets, of which two should be associated with b-quarks. The algorithm can be roughly summarized as follows:

- Identify two jets consistent with the hypothesis of being of the origin from a b-quark. Such b-jets can be tagged by a dedicated algorithm at the detector level, or by an association to B-hadrons at the particle level.
- Select two non-b-jets as best candidates to form candidates of jets coming from the hadronic decay of the W boson. These can be the highest-p_{T} jets or those giving their invariant mass m_{jj} closest to m_W; angularly closest *etc*.
- Reconstruct the leptonically decaying W boson by setting the missing transverse energy as the neutrino transverse momentum, and computing is longitudinal momentum from the constrain on the invariant mass of the lepton-neutrino pair to be the mass of the W boson.
- Add the four-momentum of the leptonic W candidate to the four-momentum of the b-jet angularly closest to the lepton, to define the four-momentum of the leptonically decaying pseudo-top quark candidate.
- Add the four-momentum of the remaining b-jet to the four-momentum of the hadronic W candidate to obtain the four-momentum of the hadronically decaying pseudo-top quark candidate.

Methods of the reconstruction of the kinematics of the $t\bar{t}$ pairs in the ℓ+jets channel in the resolved regime have been studied and compared in [Auth1]. The choices of the light (non-b-tagged) jets for the hadronic W reconstruction were compared, with the "old" option of taking the highest-p_{T} ones proved to lead to high-mass tails in both the hadronic pseudo-W and pseudo-top reconstructed masses compared to the option of choosing jets with their invariant mass closest to m_W, see Figure 4.4, as well as to high-p_{T} tails in their p_{T} spectra. While this may not come as a surprise, it can also lead to "sculpting" effects, *i.e.* artificial peaks in backgrounds without a true hadronically decaying W boson, and the choice is subject of a check for individual analyses. In order to reconstruct the longitudinal component of the neutrino (p_z) coming from the leptonically decaying W boson, the most frequent condition $m_{\ell\nu} = m_W$ is imposed and of its two solutions in general usually the more central one, as for the neutrino rapidity, has been taken. This condition can however be changed to the requirement of closest masses of the reconstructed pseudo-top quarks $m_{b\ell\nu}$ and m_{bjj}. It turns out the latter condition leads to a sharper peak of a hypothetical Z' resonance at the particle level, although the effect is largely smeared away at the detector level; alternatively one can use the condition $m_{b\ell\nu} = m_{bjj}$ to compute the neutrino p_z, see Figure 4.5.

4.5.4 General unfolding formula

The measured absolute differential cross section as function of observable X in bin i can be defined as

$$\frac{\mathrm{d}\sigma}{\mathrm{d}X_i} \equiv \frac{N_i}{\epsilon_i \, \mathcal{L} \, \mathrm{BR} \, \Delta_i}\,, \tag{4.13}$$

where N_i is the number of (unfolded) events observed in the i-th bin of observable X, $\mathcal{L}$ is the integrated luminosity, Δ_i is the width of bin i, ϵ_i is the efficiency of events to be found in particle or parton level bin i of variable X. Optionally one can correct

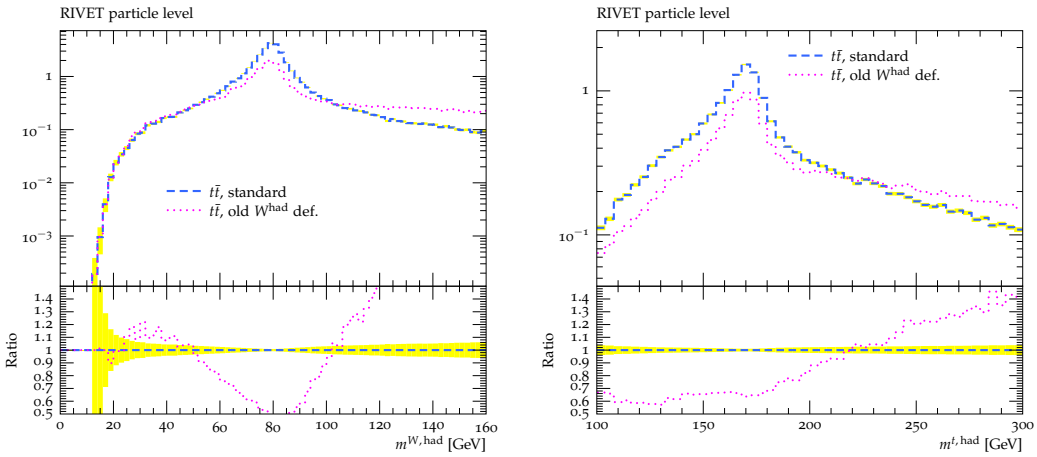


FIGURE 4.4: The particle-level hadronic pseudo-W candidate mass (left) and the hadronic pseudo-top quark candidate mass (right) for different choices of the light jets to form the hadronic pseudo-W: as the pair of jets with invariant mass closest to m_W (dashed, standard), or as the pair of highest-p_T non-b-tagged jets (dotted, "old"). Ratios to the standard choice are in lower panels, the yellow band indicates the statistical uncertainty in the denominator. Plots taken from [Auth1].

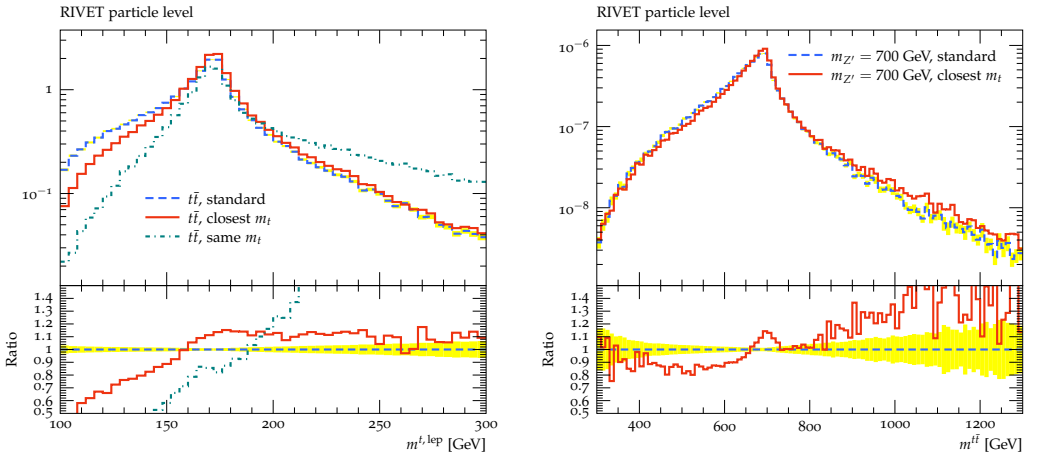


FIGURE 4.5: The particle-level distributions: (left) leptonic pseudo-top candidate mass for different choices of the neutrino p_z solution: the standard choice (dashed), "closest m_t" (solid); and "same m_t" (dot-dashed); and (right) $t\bar{t}$ invariant mass distribution for the hypothetical Z' boson of mass of 700 GeV decaying to a $t\bar{t}$ pair for the different choices of the neutrino p_z solution based on the $m_{\ell\nu} = m_W$ condition: the standard choice (dashed) and "closest m_t" (solid). Ratios to the standard option are in lower panels, the yellow band indicates the statistical uncertainty in the denominator. Plots taken from [Auth1].

for the branching fraction BR in which the process is measured to the inclusive decay case.
Usually, the unfolding procedure extracts N_i as

$$N_i = \sum_j \mathcal{M}_{ij}^{-1} \left[f_j^{\text{acc}} \left(D_j^{\text{det}} - B_j^{\text{det}} \right) \right] , \tag{4.14}$$

where $\mathcal{M}^{-1}$ stands for the matrix inversion or a regularized unfolding method starting with the number of data (background) events D_j^{det} (B_j^{det}) in the detector-level bin j; last, f_j^{acc} is the acceptance correction removing events which were reconstructed at the detector level in bin j but did not pass the particle-level selection and should thus not be considered. Thus, background estimation as well as the acceptance and efficiency corrections are key ingredients needed for correctly unfolded results.

Normalized (sometimes also called relative) differential cross sections as function of an observable X in bin i are then defined as

$$\frac{1}{\sigma} \frac{\mathrm{d}\sigma}{\mathrm{d}X_i} , \quad \text{where} \quad \sigma \equiv \sum_{i=1}^{N_X} \frac{\mathrm{d}\sigma}{\mathrm{d}X_i} , \tag{4.15}$$

and have the advantage of a partial cancellation of systematics uncertainties which are largely correlated across kinematic bins of the variable X. The case of a full correlation across bins of any kinematic variable X is the data luminosity uncertainty, which then does not play a role in normalized differential cross-sections (it may enter in background subtraction via normalization of simulation-based backgrounds to luminosity, in combination with backgrounds extracted from data control regions, not normalized to luminosity).

4.5.5 Unfolding methods

Traditional unfolding methods

We present an overview of unfolding methods which have been used in the past in high energy physics experiments, summarizing briefly their essence, advantages and disadvantages.

- Bin-by-bin unfolding is simple correction based on a simulated spectrum, correcting individual bin contents by a factor derived as a ratio between the simulated and detector levels. A very good agreement between the spectra in data and simulation is needed.

- Ansatz unfolding is an analytical method which folds a parameterized analytical spectrum with a resolution function, comparing the result to the one observed in data to optimize the ansatz parameters. Precise modelling of the resolutions is needed.

- Iterative folding is a method of folding an estimate of a spectrum at the particle or parton level with known detector resolution in terms of a migration matrix, stopping after agreement is reached at some defined condition, like stability *w.r.t.* to previous iteration.

- Direct matrix inversion is an unstable method sensitive to numerical instabilities. Although leading to the unbiased result, it suffers from a large variance [89].

- Regularized Singular Value Decomposition (SVD) [90] has been a popular regularized unfolding method, with the ambiguity of defining which eigenvalue are small so that not to be considered, serving effectively as a regularization parameter.
- Iterative Bayesian unfolding [91, 92] has been heavily used in HEP experiments, with the caveat of defining the number of iterations where to stop the algorithm.
- Dynamically stabilized method of data unfolding [93] is a more recent method trying to preserve possible new features in observed spectra *w.r.t.* the simulation.
- Likelihood-based unfolding [94] allows for a simultaneous fit to systematics uncertainties via the technique of profiling. The danger is the possible of interplay of a possible real data deviations from a theory and its possible "absorption" to a shape allowed by some systematics.

FBU unfolding

Among other methods, the Fully Bayesian Unfolding method [95] stands out for several reasons. It naturally encompasses many possible regularization choices by the definition of a prior. For a flat prior it is equivalent to the likelihood-based unfolding, yet it provides a full multidimensional posterior which can be used to infer the unfolded mean, median, or fitted value as the unfolded result, but allowing also a full covariance matrix be computed. In addition, straightforward modifications of the likelihood allow for systematics profiling, i.e. nuisance parameters fitting, providing possible in situ and process-dependent evaluation of systematic shifts w.r.t. nominal individual objects calibrations or efficiencies.

The disadvantage is high CPU time needed to sample the truth phase-space which may become important in pseudo-experiment or when unfolding many systematic-varied samples. On the other hand, systematics can be naturally implemented into the likelihood formula using known systematic shift constrains, treating them as nuisance parameters similarly to standard likelihood fits.

Based on the Bayes theorem one can write the (conditional) probability of inferring parameters of interest $\boldsymbol{\theta}$ given the measured data $\boldsymbol{x}$ as

$$P(\boldsymbol{\theta}|\boldsymbol{x}) = P(\boldsymbol{x}|\boldsymbol{\theta})\,\frac{P(\boldsymbol{\theta})}{P(\boldsymbol{x})}\,,$$

with

$$P(\boldsymbol{x}) \equiv \int P(\boldsymbol{x}|\boldsymbol{\theta})P(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$$

standing for the integration of all possible hypotheses $\boldsymbol{\theta}$ leading to the observed data using the prior information $P(\boldsymbol{\theta})$.

Reformulating this to the problem of unfolding, the (posterior) probability for having T_i "truth" (*e.g.* particle or parton level) events in truth bin i given the events observed in data at the detector level $\boldsymbol{D}$ can be expressed as

$$P(\boldsymbol{T}|\boldsymbol{D}) = P(\boldsymbol{D}|\boldsymbol{T})\,\frac{P(\boldsymbol{T})}{\int P(\boldsymbol{D}|\boldsymbol{T})P(\boldsymbol{T})\,\mathrm{d}\boldsymbol{T}}\,.$$

The denominator is hard to define as one would have to integrate over all possible truth possibilities. On the other hand, one can also search for the maximum of the function

$$P(\boldsymbol{T}|\boldsymbol{D}) \sim P(\boldsymbol{D}|\boldsymbol{T})\, P(\boldsymbol{T})\,, \tag{4.16}$$

where $P(\boldsymbol{T})$ is some (subjective or motivated) prior information for the truth spectrum. This can be a measurement of $\boldsymbol{T}$ from a previous experiment, but it can also naturally encode the information on its other desired properties, like smoothness, be it expressed by a condition imposed on the curvature, entropy *etc*, and can be regarded as a regularization condition.

The expression $P(\boldsymbol{D}|\boldsymbol{T})$ is the likelihood of observing the $\boldsymbol{D}$ spectrum given the particle level spectrum $\boldsymbol{T}$

$$P(\boldsymbol{D}|\boldsymbol{T}) \equiv \prod_j f(D_j|\hat{D}_j(\boldsymbol{T}))\,,$$

with f being the assumed underlying probability density function, most often Poissonian $f(k|\mu) = e^{-\mu}\mu^k/k!$ or Gaussian $f(x|\sigma,\mu) = e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$. The expected data count $\hat{D}_j$ in detector-level bin j can be expressed using the migration matrix $\mathcal{M}_{ji}$ and the assumed truth distribution T_i in truth-level bins i as folding

$$\hat{D}_j \equiv \sum_i \mathcal{M}_{ji} T_i\,.$$

Maximizing Eq. 4.16 as a function of the truth distribution is the process of unfolding. Employing techniques of sampling the truth space by Markov Chain Monte Carlo methods [96] one can in fact obtain empirically sampled binned posterior distribution for each T_i.

An extension to the FBU unfolding was also proposed, which, besides explicitly implementing several regularization terms, also suggests to sample the standard FBU likelihood and the regularized term part separately, allowing one to select the regularization term strength ex post and not treat it as another degree of freedom, which saves the CPU time [Auth2].

Modern ML unfolding methods

In recent years, modern approaches to the unfolding based on Machine Learning (ML) have emerged, *e.g.* the Omnifold [97] technique. These are promising yet CPU extensive methods, and while a subject to over-training, they offer a way to unfold all kinematics variables at once, being in fact event-by-event unfolding approaches. This enables one to unfold multidimensional distributions at once.

4.5.6 Author's contributions

Author of this thesis is the only author of study [Auth1] which explores and extends the pseudo-top algorithm for novel constrains for measurements as well for searches for new physics, like $t\bar{t}$ resonances.

Author of this thesis is one of the co-authors of the study [Auth2] extending the FBU unfolding for motivated priors and sampling the likelihood and prior terms separately.

Chapter 5

Top quark measurements with the DØ Experiment at Fermilab

The CDF and DØ experiments were the two multipurpose experiments at the Tevatron collider at Fermilab. The author of this thesis was a member of the DØ experiment, and a summary of several key measurements, with the emphasis on the pioneering top quark transverse momentum, will be presented.

5.1 Inclusive cross-section

Using MADGRAPH5, the $t\bar{t}$ cross-section in $p\bar{p}$ collisions at 1.96 TeV at the NLO level is extracted to be about 6.0 pb. More precisely, the inclusive cross section for the production of $t\bar{t}$ pairs in proton-antiproton ($p\bar{p}$) collisions was predicted to be $7.35^{+0.28}_{-0.33}$ pb [98] the NNLO+NNLL level.

The legacy experimental Tevatron combination result [99] of the two main experiments CDF and DØ is $\sigma_{t\bar{t}} = 7.60 \pm 0.41$ pb, for the top-quark mass of $m_t = 172.5$ GeV, in good agreement with the best precision predictions. The assumed mass affects the final states products momenta and thus the selection efficiency. The cross-section is both a test of the pQCD dynamics at higher orders but also of the gluon parton distributions functions. Other interesting quantities have been derived from cross-section measurements or their ratios, like the branching fraction of $t \to Wb$ to $t \to Wq$ [100] *etc*.

5.2 First top quark transverse momentum measurements

The measurements of the top quark p_{T} spectrum were pioneered already at Tevatron. From the Run-I dataset, only the detector-level spectrum was shown as depicted in Figure 5.1. Such first glimpses of the top quark transverse momentum spectrum at the detector level, *i.e.* affected by the finite detector resolution, soon after its discovery in mass measurements DØ collaboration is shown in Figure 8 in [101] or Figure 49 in [102].

The first top quark transverse momentum measurement, fully corrected for detector effects, was carried out at the DØ experiment at Tevatron with a limited statistics [103] and utilized unfolding to the parton level.

A preliminary Run-I DØ result is shown in Figure 5.2 while the larger dataset result in Figure 5.3. A χ^2 kinematic fitter *HitFit* [104, 87] was employed to reconstruct the kinematics of top quarks in the ℓ+jets channel. This first differential spectra were unfolded from the detector to the full phase space parton level using the SVD unfolding technique [90]. The systematic uncertainty was dominated by the jet energy calibration and the unfolding procedure sensitivity to the underlying model

spectrum, evaluated by a reweighting procedure of the parton level spectrum and the observed effect in the unfolded data spectrum. Both the leptonic and hadronic top quark p_{T} values were used and both e+jets and μ+jets channels were unfolded together at once. The unfolding formula in this case reads

$$\frac{\mathrm{d}\sigma_{t\bar{t}\to\ell+\mathrm{jets}}}{\mathrm{d}{p_{\mathrm{T}}}^{i}} \equiv \frac{\sum_j \mathrm{Unfold}_{ij}\left[D_j - B_j\right]}{\mathcal{L}\cdot \mathrm{BR}_{\ell+\mathrm{jets}}\cdot\epsilon\cdot\mathcal{A}_j\cdot\Delta{p_{\mathrm{T}}}^{j}}\,, \tag{5.1}$$

with D_k (B_j) being the number of measured data (estimated background) events in bin j of the top quark p_{T} spectrum; $\mathcal{L}$ is the integrated luminosity of the data sample, $\mathrm{BR}_{\ell+\mathrm{jets}}$ is the branching ratio of $t\bar{t}$ to ℓ+jets, ϵ is the overall (integral) efficiency of events to be reconstructed given the detector-level selection, $\mathcal{A}_j$ is the differential acceptance (relative efficiency binned in the spectrum of interest) and $\Delta{p_{\mathrm{T}}}^{j}$ is the bin width. Finally, "Unfold" stands for regularized unfolding from the detector-level bins j back to the parton-level bin i in the top quark p_{T}.

Both absolute and relative cross-sections were extracted, with the relative one encoding the shape information effectively stripped off by bin-correlated systematics, allowing for a more precise shape comparison.

The full statistics analysis of the Tevatron DØ data [105] is shown in Figure 5.3. The kinematic reach of these first Tevatron measurements is 400 GeV in top quark transverse momentum, with comparable statistical and systematic uncertainties. The data were unable to distinguish more closely different simulation models or theory approaches. The available generators at that time were the LO PYTHIA and multi-leg tree level LO generator ALPGEN. Emerging at that time was the MC@NLO generator [64], the first fixed-order NLO pQCD calculation with scale uncertainties [106] and approximate NNLO one, *i.e.* a NLO one with soft gluon emissions summed up in the next-to-next-to-next-to-leading logarithm (NNNLL) [107].

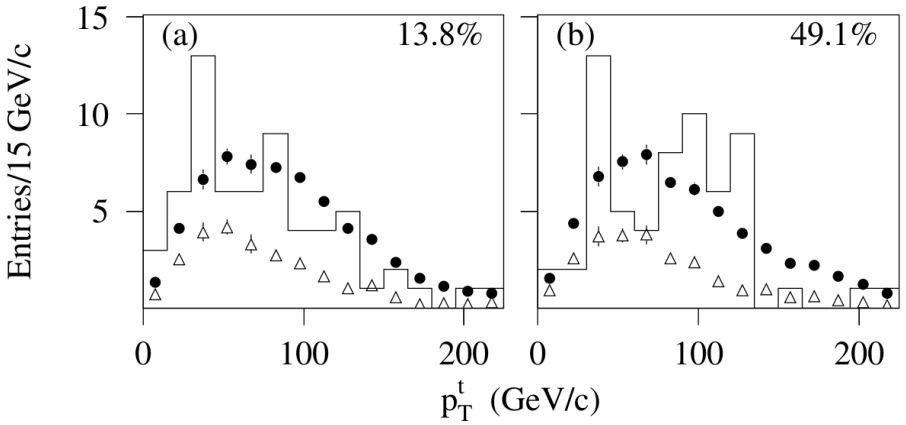


FIGURE 5.1: Detector-level distribution of the top quark transverse momentum (two entries per event) measured by the DØ detector (histogram) at Fermilab's Tevatron Run I [102] with the top quark mass (a) as a free parameter and (b) fixed in the fit. Open triangles: the expected background. Full circles: the expected signal and background.

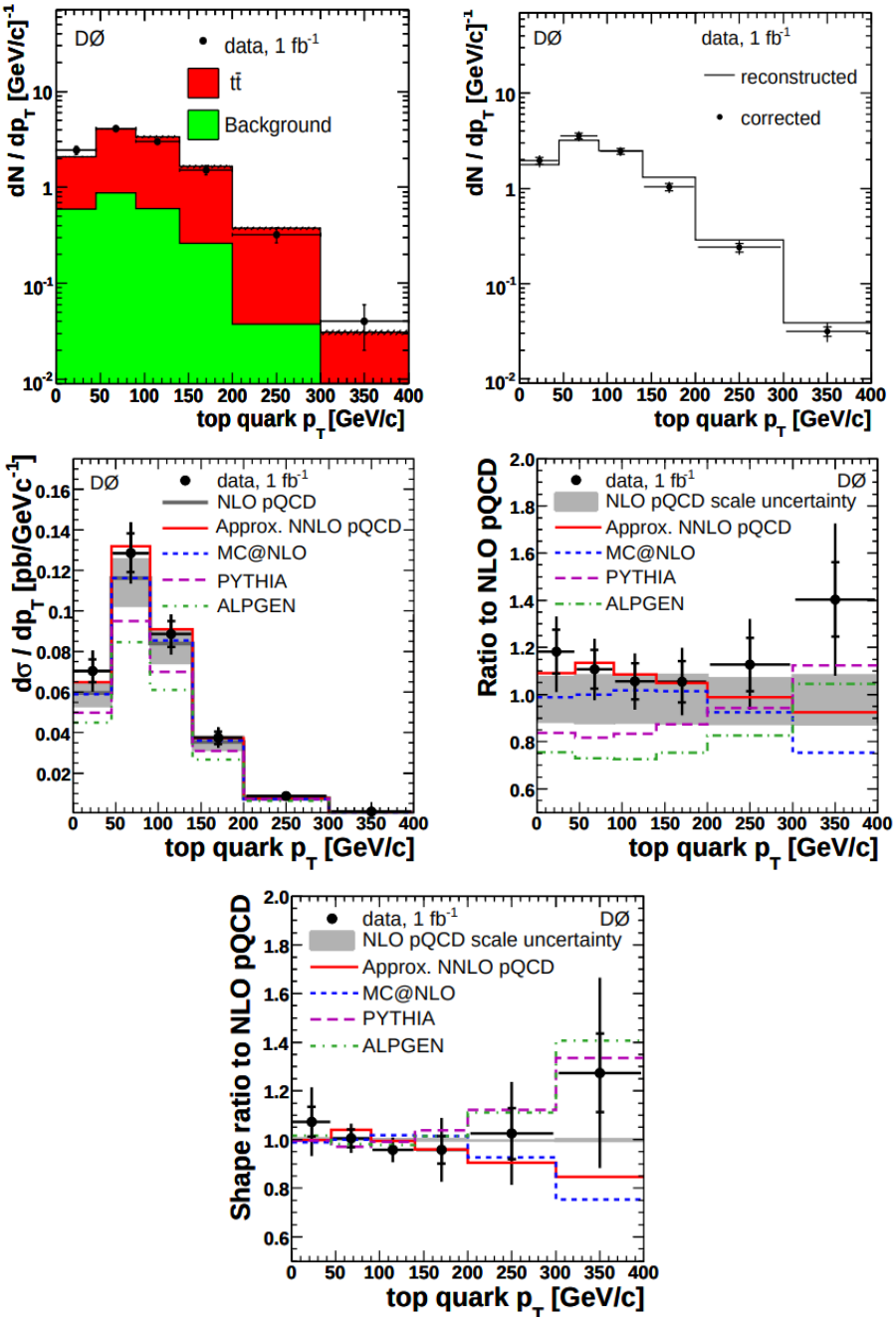


FIGURE 5.2: The top quark transverse momentum distribution in the first analysis by the DØ experiment [103] at the detector level (top left), compared between the background-subtracted detector-level and unfolded (corrected) spectra (top right), fully corrected data compared to predictions from PYTHIA, ALPGEN and MC@NLO generators, and (middle) also to the theory pQCD calculation at the NLO precision (solid black) as well as an approximate NLO+NNLL pQCD prediction (red line). Ratio of the relative cross-section is shown in the bottom plot.

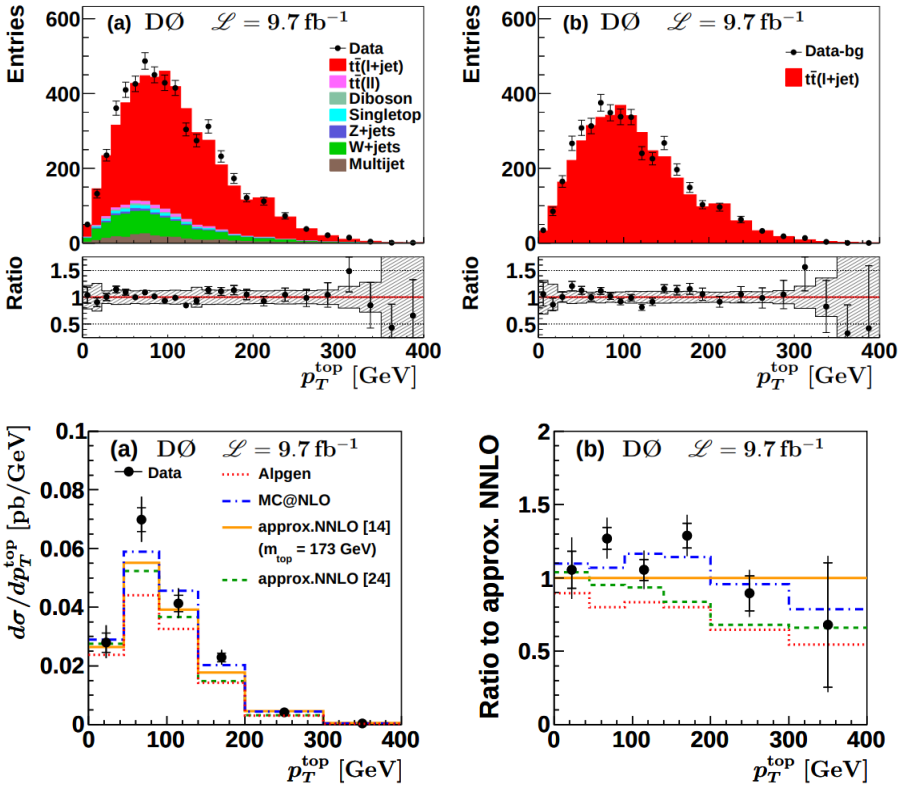


FIGURE 5.3: The final dataset top quark p_{T} measurement by the DØ collaboration [105], showing the data comparison to the sum of the expected backgrounds and signal (top left), the background-subtracted detector-level data compared to signal only (top right) and the absolute differential cross-section as function of the top quark p_{T} (bottom left) and its ratio to the approximate NNLO prediction (bottom right).

5.3 Top quark charge asymmetry at Tevatron

Due to the nature of the collisions at the Tevatron, the fact of colliding protons and antiprotons defines a distinct axis w.r.t. which the top or anti-top quark rapidity can be measured [108]. The forward-backward asymmetry in the preferred angle (and therefore rapidity) if the produced (anti)top quarks is a phenomenon originating in the $q\bar{q}$ only initial state due to an interference of the tree-level s-channel and box diagrams [109]. Experimentally, at Tevatron, (anti)top quarks are produced preferably in the direction of the original (anti)quark entering the collision.

5.4 Further kinematic variables at Tevatron

We shortly discuss other variables of interest which test the QCD or electroweak nature of the top quark production or decay. The helicity of the W boson from the top quark decay was measured in a good agreement to predictions and is a test of the $V - A$ nature of the top quark decay. Due do the fast top quark decay, spin correlations between the top quarks are not spoiled by the hadronization process and they are imprinted into the spins of the top quark decay products. The out-of-plane

momentum p_{out} variable was studied in author's Ph.D. thesis [110] and was implemented based on older QCD studies [111]. Originally, this variable was designed to measure the intrinsic parton transverse momentum (intrinsic k_{T}) but in the $t\bar{t}$ system it turned out to be a useful indication of the additional radiation, especially when introduced in ATLAS analyses [ATLAS3] and which was then used to tune MC parameters to better describe data as for additional jet activity in $t\bar{t}$ events [112]. This has lead to constrains to MC modelling of the $t\bar{t}$ process with impact on diminishing the systematic uncertainties *e.g.* in top quark mass measurements.

Chapter 6

Top quarks pairs at the LHC

With the onset of the Large Hadron Collider at CERN, the physics energy frontier shifted back from the USA to Europe, and proton-proton, but also proton-nucleon and nucleon-nucleon, collisions have been studied at increasing energies of the collider. The two major multipurpose experiments ATLAS and CMS are accompanied by two more specialized experiments ALICE (for the exploration of the nucleon collisions) and LHCb (for the study of flavour physics, most notably of the b quark).

It is impossible to describe in fullest the experimental particle physics programmes at these detectors, ranging from high-p_{T} physics to diffractive and elastic collisions, from the physics of the BEH boson and top quark to baryon and meson physics; and from the discovery of rare Standard Model processes and their precision measurements to searches for physics beyond. All these achievements and still ongoing endeavours have only been possible with simultaneous developments and tireless efforts on the frontiers in computing and hardware development, and in areas of steady improvements and understanding of the detector and objects calibration and identification techniques, statistical tools as well as progress on the theory side of the particle physics.

6.1 Top quark re-discovery at the LHC

Top quark serves its rôle as a "standard candle process" at the LHC as one of the main processes to be re-discovered at the LHC start up.

This included the CMS [113] and ATLAS [114] re-discovery analyses in dilepton channels, followed by more precise inclusive as well as differential cross-section measurements, and detailed studies of top quarks properties, ranging from spin correlations to top quarks charge asymmetry.

Current analyses (at the start of LHC Run 3) at the energy of 13.6 TeV) are no longer statistical-error dominated in precision, while systematic uncertainties can be evaluated in data control regions, depending less on modelling and simulation uncertainties. Also, the top quark process is being explored in measurements of differential top quark spectra to constrain modelling systematics in the $t\bar{t}$ process itself, enabling, in turn, to diminish $t\bar{t}$-related systematic uncertainties *e.g.* in top quark mass measurements.

Combined with the growing LHC luminosities this leads to thousands of top quark events recorded by each of the main LHC experiments, ATLAS and CMS, but also by the LHCb [115].

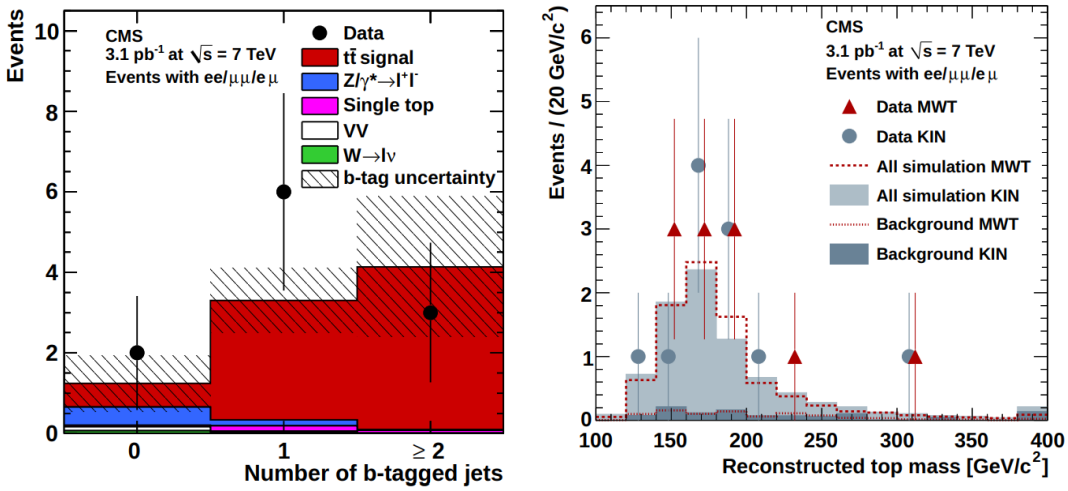


FIGURE 6.1: Expected event composition compared to measured data binned in the number of b-tagged jets (left) and the distribution of the reconstructed top quark mass (right) with a clear evidence for the $t\bar{t}$ signal, as analyzed by the CMS collaboration in its first 7 TeV data in 2012 [113].

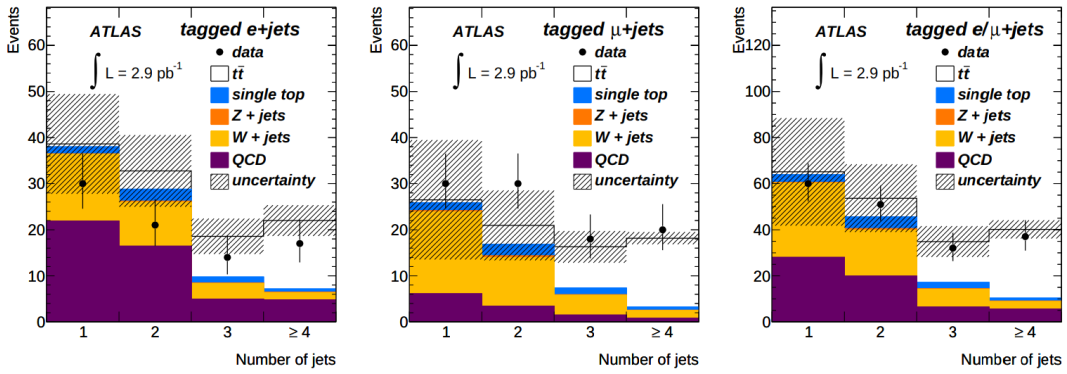


FIGURE 6.2: Expected event composition compared to measured data binned in the number of jets after the b-tagging requirements, with clear indications of the $t\bar{t}$ signal in the highest jet multiplicity, as analyzed by the ATLAS experiment in its first 7 TeV data in 2010 [114].

6.2 The total $t\bar{t}$ cross section

The top quark pair cross-section of 6–8 pb at the Tevatron is dwarfed by the cross section of about 930 pb at the LHC [98] energy of 14 TeV, opening doors to highly precise measurement of its production, properties, and the level of agreement with the Standard Model. The $t\bar{t}$ cross section is measured as compared in Figure 6.3, together with the theory curves for $p\bar{p}$ and pp collisions.

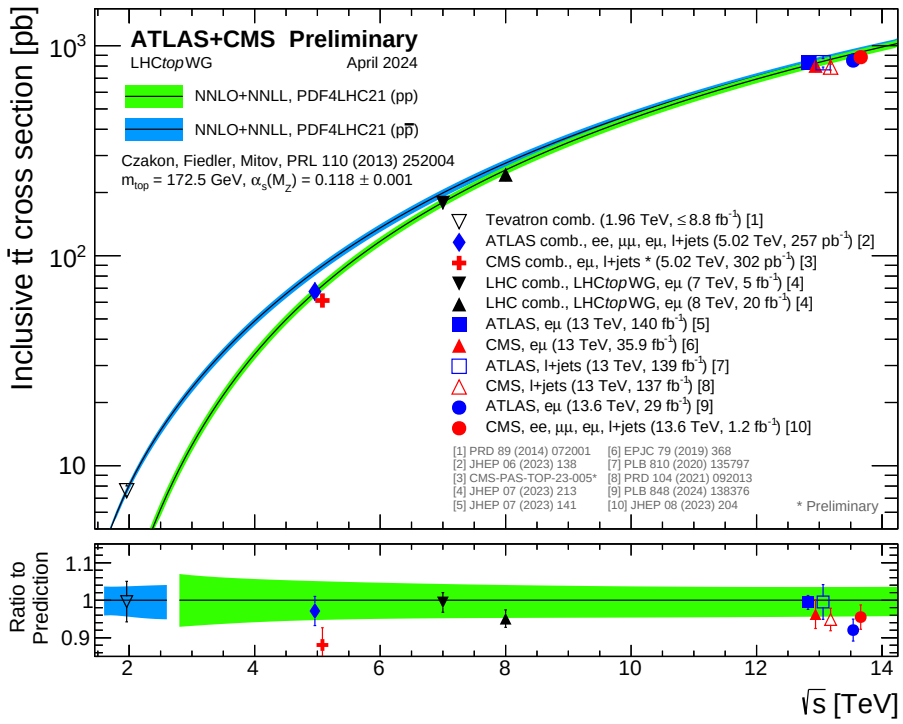


FIGURE 6.3: Dependence of the theoretical $t\bar{t}$ production cross section on the energy of the pp collisions at the LHC as well as of the $p\bar{p}$ collisions at Tevatron, with measurements by the DØ, CMS and ATLAS experiments shown for comparison. Adopted from the LHC Top WG [31].

6.3 Kinematic variables of the $t\bar{t}$ system

Studying in detail spectra of the produced top quarks or of the $t\bar{t}$ system allows to impose stringent tests on the theory prediction to describe these variables, as well as to test how realistic various systematic variations are, allowing one to constrain the uncertainties in the $t\bar{t}$ modelling, helping thus other measurements where top pair production is a background process.

The variables frequently studied in the $t\bar{t}$ system are the mass and rapidity of the top quarks (m^t, y^t), their transverse momenta, be it the leading and sub-leading, their average, or those of the leptonically or hadronically decaying ones; and similar variables for the $t\bar{t}$ system: rapidity $y^{t\bar{t}}$, mass ($m^{t\bar{t}}$ and transverse momentum ($p_{\mathrm{T}}^{t\bar{t}}$). An important angular variable can be composed from the two top quarks, the $\cos\theta^*$ being the angle between a top quark and the z axis in a frame where the $t\bar{t}$ system has zero momentum along the z axis. Of interest is also the angle between the top

quark three-momenta in the transverse plane, $\Delta\phi^{t\bar{t}}$, sensitive to additional QCD radiation, and the ratio of top quark transverse momenta. Variables of interest are also transverse momentum of the $t\bar{t}$ system ($p_{\mathrm{T}}^{t\bar{t}}$) and the out-of-plane momentum p_{out} which is based on the direction of one of the top quarks defining a plane together with the z axis direction, to which the momentum of the other top quark is projected, therefore defined as

$$p_{\mathrm{out}} \equiv \vec{p}^{\,t,2} \cdot \frac{\vec{p}^{\,t1} \times \hat{z}}{|\vec{p}^{\,t,1} \times \hat{z}|}\,, \quad \text{and } 1 \leftrightarrow 2\,.$$

This variable, brought to the LHC analyses by the author from his Tevatron experience, turned out to be useful for being sensitive to additional QCD radiation along with the $t\bar{t}$ pair production, allowing to constrain models of initial and final state radiation.
Additional variables studied in presented ATLAS analyses are

$$y_{\mathrm{boost}}^{t\bar{t}} \equiv \frac{1}{2}\left|y^{t2} + y^{t1}\right| \tag{6.1}$$

$$\chi^{t\bar{t}} \equiv \exp\left|y^{t2} - y^{t1}\right|\,. \tag{6.2}$$

These are sensitive to final state radiation, the boost of the $t\bar{t}$ system and thus also to parton distribution functions; and to new physics via their sensitivity to the production angle in the central-mass-system. A classical variable in jet physics are eigenvalues of the normalized momentum tensor defined using the three-momenta of objects $\mathcal{O}$ in the event (usually jets) as

$$\mathcal{M}_{ij} \equiv \frac{\sum_{\mathcal{O}} p_i^{\mathcal{O}} p_j^{\mathcal{O}}}{\sum_{\mathcal{O}} (\vec{p}^{\,\mathcal{O}})^2} \tag{6.3}$$

where i, j run over the three space indices. The matrix eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are used to define aplanarity $\mathcal{A} \equiv \frac{3}{2}\lambda_1$ and sphericity $\mathcal{S} \equiv \frac{3}{2}(\lambda_1 + \lambda_2)$.

All the aforementioned variables can be studied and compared to predictions of various precision, ideally as measure spectra corrected for detector effects back to the particle or parton levels, and both as absolute or as normalized spectra (studying only the shapes of these variables).

Chapter 7

Top quark physics with the ATLAS experiment

Measuring the shapes of $t\bar{t}$ spectra is an important input to analyses where the $t\bar{t}$ events constitute a background. It is not always possible, *e.g.* due to low number of observed events, to construct $t\bar{t}$-dominated control regions in data. Some analyses at LHC do use simulated $t\bar{t}$ events to model this background, and reweighting of the top quark p_{T} spectrum to the one obtained in data is sometimes employed.

At the LHC, the kinematics reach is much extended compared to Tevatron, top quark pairs are no longer produced close to threshold and therefore are not much at rest. Higher collision energies also allow for more and harder jets be produced along with the $t\bar{t}$ pair, leading to non-trivial p_{T} of the $t\bar{t}$ system and thus a possible imbalance in p_{T} of the two top quarks.

We will focus on measurements in the ℓ+jets channel where the full event kinematics is reasonably reconstructable and the branching ratio leads to large event yields. One of the most interesting story is the one of the measurements of top quark transverse momenta at the LHC. We will discuss key analyses at the increasing energies of proton-proton collisions as evolved with the LHC run periods, and we will highlight main physics lessons learned, also illustrating the development of the complexity of the analyses techniques together with advances in theory.

7.1 Top quark physics at 7 TeV – pioneering the precision and the pseudo-top concept

At the LHC energy of 7 TeV, corresponding to the data recorded in 2010–2011, two analyses [ATLAS2, ATLAS1] measured the $t\bar{t}$ differential cross-sections unfolded to the parton level and the $t\bar{t}$ events kinematics was reconstructed using a kinematic likelihood fitter [85], KLFitter. Similarly to the χ^2 fitter used at the DØ, resolution functions, here called transfer functions, are needed in order to define the allowed realistic smearing of objects four-momenta between the parton and detector levels.

While the first paper studied only unfolded shapes of few selected spectra and compared to NLO predictions (in case of $m^{t\bar{t}}$ also to the available calculation at the NLO+NNLL level), the second analysis involved more spectra and their relative differential cross-sections compared to several NLO generators as well as different parton distribution functions. Cross-sections were measured separately in the e+jets and μ+jets channels and then combined using methods outlined in Section 4.4.1. In addition, a full correlation matrix was published in order to allow for further private PDF fits using all spectra simultaneously, setting a high standard in top quark differential spectra measurements. The matrix was evaluated using the bootstrap technique as described in Section 4.4.6.

Interpretation

For the top quark p_{T} spectrum measured at the 7 TeV, see Figure 7.1. Most notably, one can observe the softness of the top quark p_{T} spectrum in data above 250 GeV in comparison to any choice of PDFs or NLO generator settings, be it a fixed order QCD calculation like that of MCFM [116] or a NLO generator with the subsequent parton shower performed by PYTHIA or HERWIG. This was the first significant indication of this important observation, meaning, among other things, that events populating the boosted regime of very high top quark transverse momenta, are to be expected less frequently in data to come. The measured differential spectra were then used to tune the h_{damp} parameter of the POWHEG generator to better match the top quark p_{T} shape, as described in [112, 117], serving thus subsequent simulation efforts and their usage to better describe data to come, also at higher LHC energies.

The first pseudo-top measurement

Another important analysis by the ATLAS collaboration at 7 TeV was the one which introduced the pseudo-top quark concept in an analysis of absolute differential spectra corrected to the particle level [118]. The novel concept of forming top quark candidates from final state objects like leptons, jets and missing transverse energy, allows to define such pseudo-top quarks both at the particle and detector levels, avoiding the parton level, definition of which is a subject to larger theoretical ambiguities. The unfolded spectra were then compared to then leading generators, namely the LO multi-leg generator ALPGEN and NLO generator POWHEG, see Figure 7.2. The conclusions was again the observation of a hint of softer top quark p_{T} spectrum above 300 GeV, *i.e.* again a deficit of events at large top quark transverse momenta.

Author's contributions

The author contributed to the development of the analysis framework for the first publication with early LHC data [ATLAS1]. He then became one of the two responsible contact editors (together with dr. Lorenzo Bellagamba) and member of the analysis team of the paper of the full dataset at 7 TeV, the analysis [ATLAS2]. He contributed to the core analysis SW, wrote the core statistical code to combine the e+jets and μ+jets channels using the BLUE method, worked on the unfolding stress tests with an artificially injected bump at the parton level of the spectra, supervised MSc. student Peter Berta, evaluated the efficiency and acceptance corrections, binned the approximate NNL predictions to the analysis bins, worked on unfolding closure tests, the summation of systematic uncertainties, plotting, running the analysis SW. A summary of the pseudo-top concept is presented and also further extended in study by the author of this thesis [Auth1].

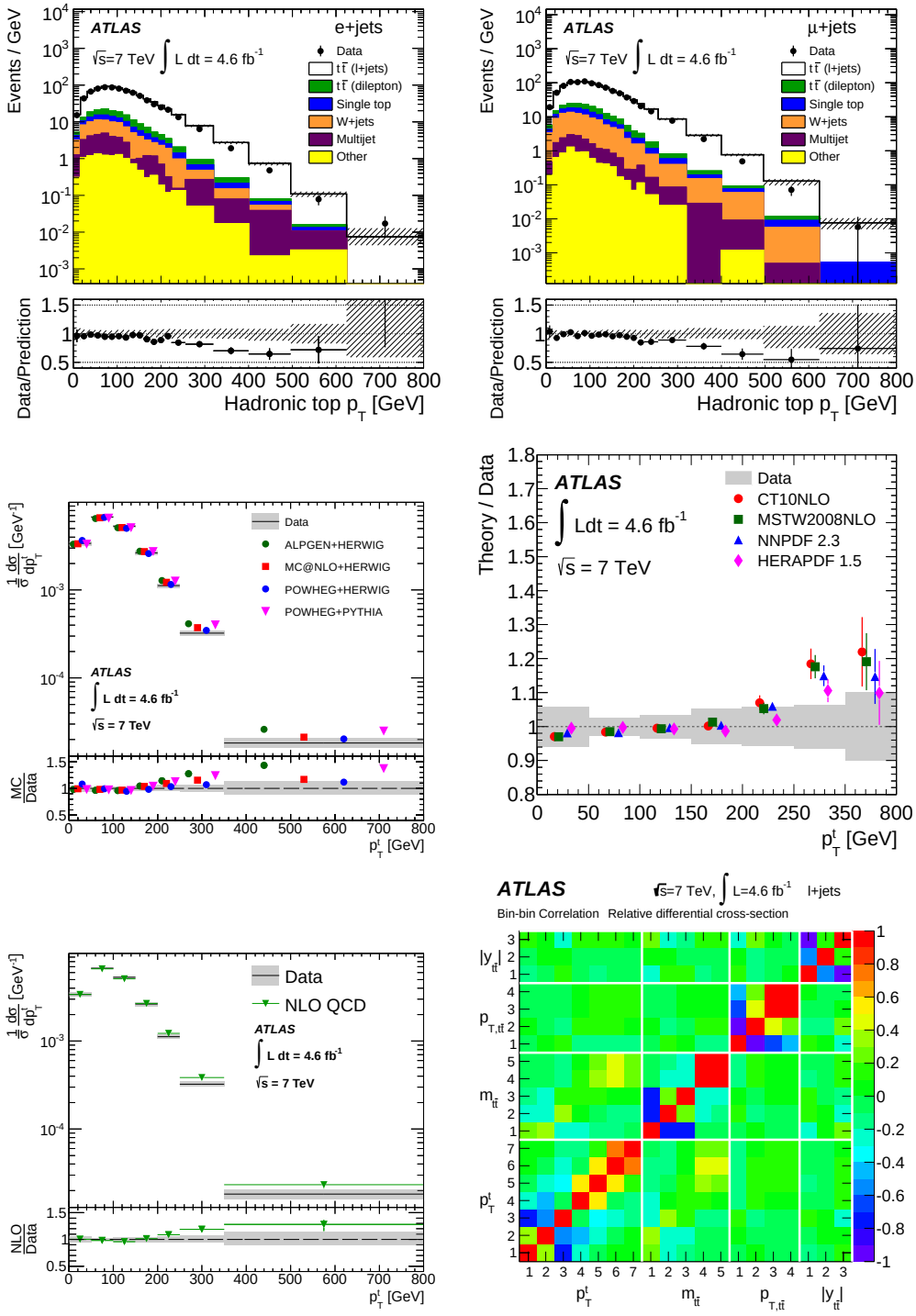


FIGURE 7.1: Top: the detector level top quark transverse momentum in the e+jets (top left) and μ+jets (top right) channels. Middle: comparison of the spectra unfolded the parton full phase-space level to different NLO MC predictions (left) and to different PDF choices using the MCFM NLO prediction (right). Bottom: NLO/data ratio (left), and the correlation matrix between the studied spectra, evaluated using pseudo-experiments (right). Plots taken from the ATLAS parton-level analysis at $\sqrt{s} = 7$ TeV [ATLAS2].

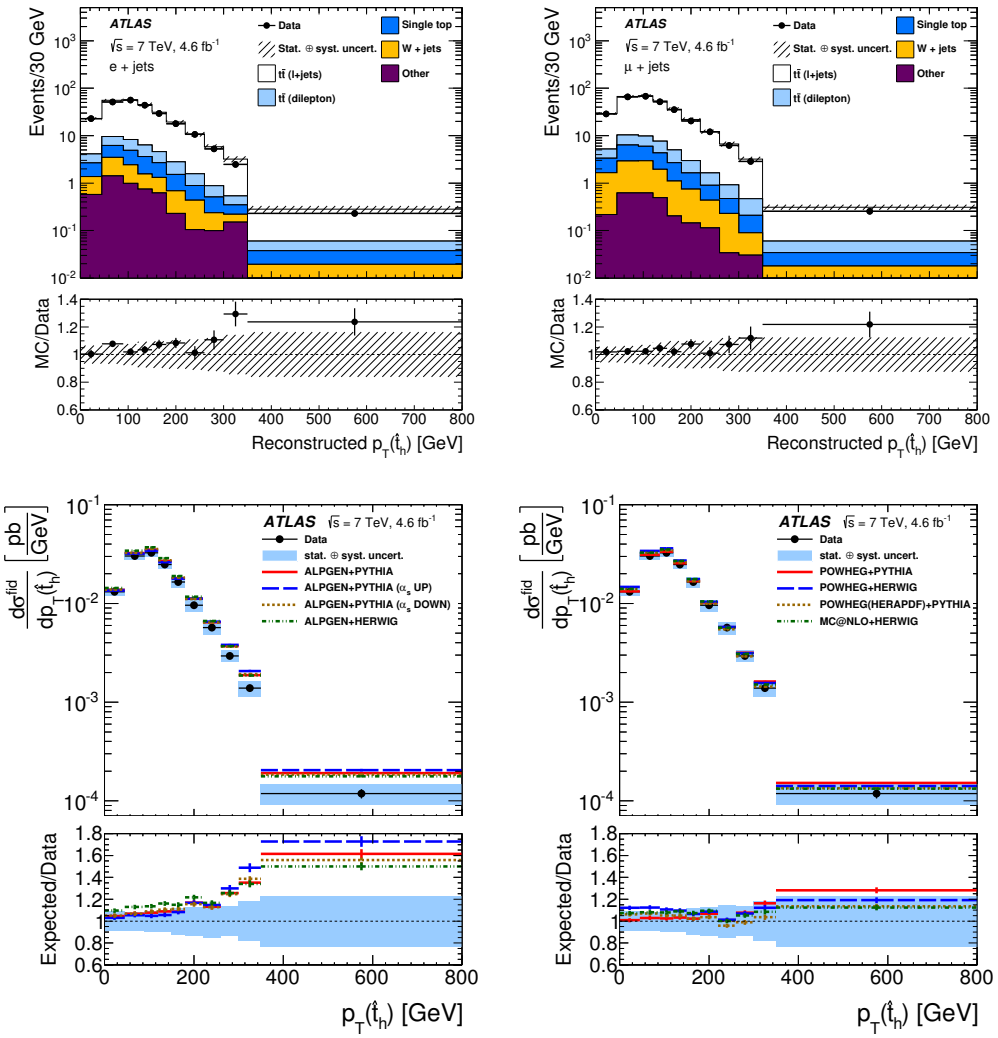


FIGURE 7.2: The detector-level distributions in the e+jets (top left) and μ+jets (top right) channels as function of the hadronic pseudo-top quark p_{T}, compared to the prediction with POWHEG+PYTHIA8 as the signal model. Bottom: unfolded particle-level spectra compared to the ALPGEN generator and its various modelling variations (left) and to the POWHEG and MC@NLO generators (right), interfaced to different parton shower models. Plots taken from the particle-level ATLAS analysis at $\sqrt{s} = 7$ TeV [118].

7.2 Top quark transverse momentum under pressure: 8 TeV analyses in resolved and boosted regimes

At the LHC energy of 8 TeV, corresponding to the run period of 2012 and the integrated luminosity of 20.3 fb^{-1}, the $t\bar{t}$ differential cross-sections were obtained by unfolding e+jets and μ+jets channels together due to their close efficiencies and phase spaces. Emphasis was put on models comparison and the possibility that several sensitive spectra could potentially constrain modelling of the $t\bar{t}$ final state for analyses where top quarks are a background to rarer SM processes or to new physics. Using the pseudo-top concept by ATLAS, spectra were corrected to the particle level and also to the full phase space (parton) level in a publication dedicated to the resolved regime of the hadronically decaying top quark [ATLAS3]

The top quark p_{T} reach in the fiducial (particle-level) analysis touched for the first time the 1 TeV limit, while the parton level full phase space analysis was limited to 0.5 TeV due to larger migrations between the detector and parton levels and larger uncertainties stemming from modelling uncertainties; and from the fact that a larger correction is needed to translate the spectrum from the detector to the full phase space level.

However, the upgraded LHC energy and larger delivered luminosity also allowed for the first time the possibility to observe the phase space of large transverse momenta of top quarks, which can manifest in the detector as a highly boosted (in the transverse direction) hadronic final states, large-radius jets, with a distinct substructure allowing one to tag them as coming from the top quarks. This was the focus of another ATLAS analysis [ATLAS4] which measured the hadronically decaying top quark p_{T} spectra on the interval of 300–1200 GeV, *i.e.* passing the 1 TeV limit.

In these analyses, spectra were unfolded in a single ℓ+jets channel, *i.e.* with the e+jets and μ+jets channels combined at the detector level, and using combined corrections and migration matrix. The resulting spectra of the top quark p_{T} as measured by ATLAS at 8 TeV in the resolved regime and corrected to the particle level using the pseudo-top concept, and to the full phase space parton level, see Figure 7.3. The boosted regime results with spectra also unfolded to both the particle and parton (full phase space) levels are shown in Figure 7.4. Results in the resolved regime have comparable statistical and systematics uncertainties whereas results in the the boosted regime are limited by statistical uncertainties.

Interpretation

The important feature is the presentation of normalized differential cross sections, leading to a more precise shape measurement due to cancellation of bin-correlated systematics as detailed in Section 4.4.4. One can conclude that the feature of the soft top quark p_{T} spectrum observed at 7 TeV persists also in 8 TeV data at both the parton full phase space as well as the particle level.

With the advance of theory calculation, a comparison of the resolved full phase space parton-level result to the approximate NNLO [119] and approximate NNNLO (dubbed as N^{3}LO) [107] predictions was possible, as seen in Figure 7.3 bottom right. These calculations are based on the resummation of additional soft gluon emissions beyond the fixed NLO or NNLO pQCD picture. They therefore contain additional processes but covering only a part of the higher order (and higher parton multiplicity) phase space. However, the aNNNLO prediction for the top quark p_{T} does not agree with data better than the NLO generator combined with a parton shower;

and the aNNLO predictions is softer than data at high top quark p_{T}. Nevertheless, they are important checks of various theories at different precision and a glimpse to higher orders and very useful for the experiment as they provided binned predictions for observables like that of the top quark p_{T}.

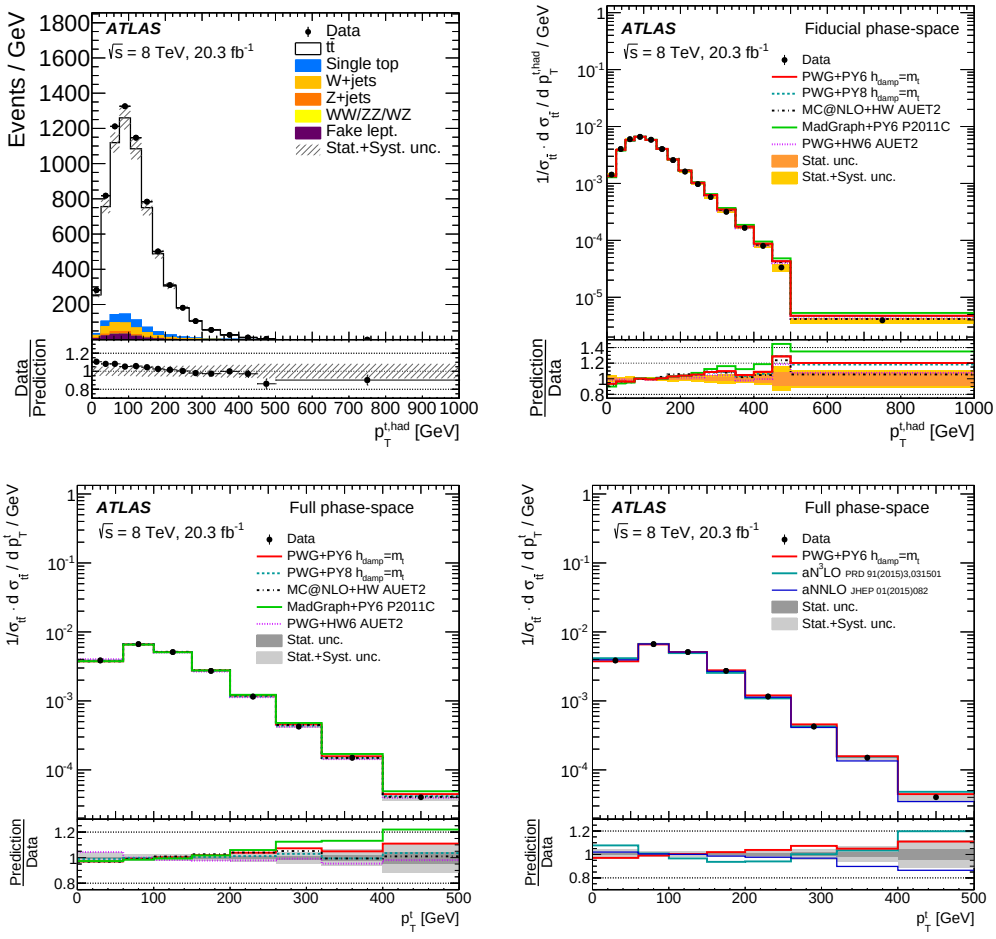


FIGURE 7.3: Results of the unfolded top quark p_{T} in the resolved regime by the ATLAS experiment at $\sqrt{s} = 8\,\mathrm{TeV}$ [ATLAS3]. Top left: detector-level hadronic pseudo-top quark p_{T} spectrum, top right: unfolded to the particle level and compared to NLO generators; bottom left: similar NLO comparison for the parton (full phase space) level spectrum, and the comparison of the parton-level result to the approximate NNLO and NNNLO predictions.

Author's contributions

Author of this thesis was a member of the analysis core team of the analysis [ATLAS3]. He worked on corrections to particle and parton level (efficiency, acceptance, matching), binning optimization, development of the core analysis framework and unfolding. He introduced the p_{out}, $y^{t\bar{t}}_{\mathrm{boost}}$, $z_{t\bar{t}}$ and $\chi^{t\bar{t}}$ variables as new ATLAS observables. He ran the analysis framework, optimized the pseudo-top algorithm, evaluated the migration matrices. He provided students supervision, also to his PhD student Josef

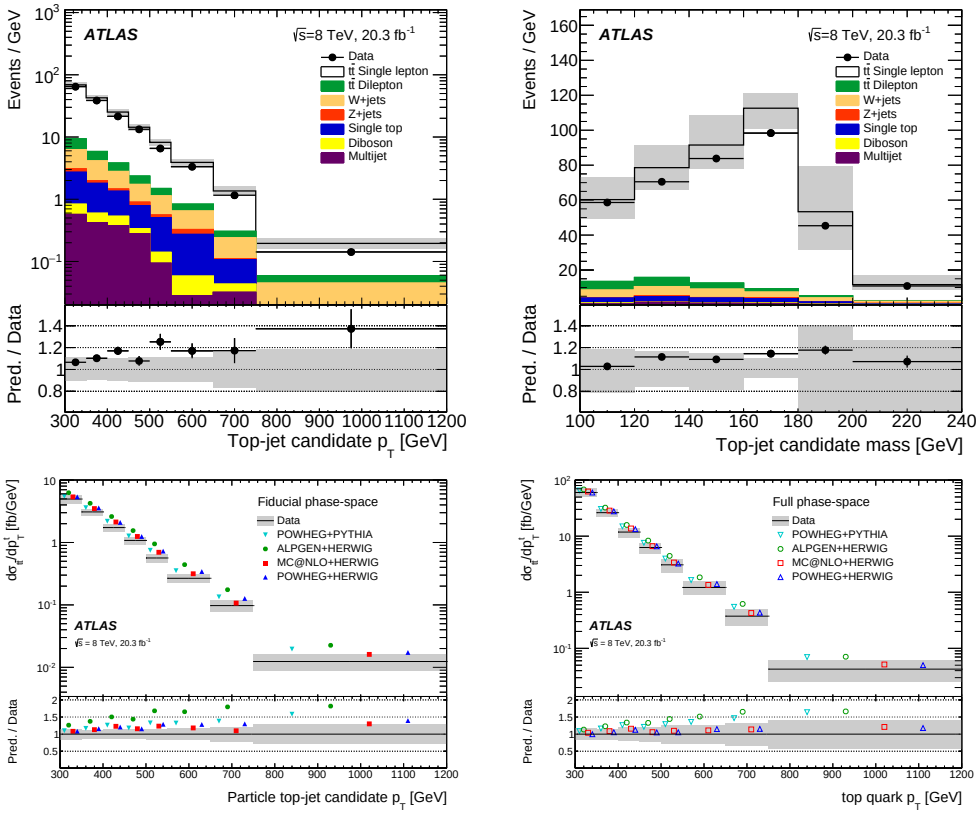


FIGURE 7.4: Detector-level (top) and unfolded (bottom) spectra in the boosted regime by the ATLAS experiment at $\sqrt{s} = 8\,\text{TeV}$ [ATLAS4]. The detector-level hadronic top-jet p_{T} spectrum (top left) and its reconstructed mass (top right); results unfolded to the particle (bottom left) and parton full phase space level (bottom right) as absolute cross sections as function of the hadronic top quark p_{T}.

Pácalt who worked on corrections to particle and parton levels and binning optimization. The author of this thesis also contributed to the boosted analysis [ATLAS4] via the supervision of PhD student Peter Berta.

7.3 The first 13 TeV analyses in resolved and boosted regimes

The initial 13 TeV measurement of $t\bar{t}$ spectra used the data recorded in 2015 corresponding to the luminosity of 3.2 fb^{-1} recorded in 2015 extracting both resolved and boosted results [ATLAS5] with the pseudo-top concept fully explored at the particle level and the boosted regime opening up with the larger $t\bar{t}$ cross-section and kinematic reach at the new energy frontier. Although corresponding to two distinct internal ATLAS notes, both analyses were performed under a single analysis by ATLAS [ATLAS5]

Slight optimization of the pseudo-top concept has taken place in the resolved analysis while in the boosted analysis, the selection criteria have been loosened from the traditional $t\bar{t}$ resonance search selection which usually requires back-to-back top quarks topology in the transverse plane to a more general phase space, in order to allow the study and test the top quark production in more extended kinematic regions.

For this 13 TeV preliminary dataset, example of top quark p_{T} related results are show in Figure 7.5. In addition, the consistency of the results of the top quark p_{T} measured in the resolved and boosted regimes (and therefore also in different phase-spaces) was checked in terms of their ratio to the same NLO generator prediction, showing a good agreement, as seen in the bottom plot of Figure 7.5.

Interpretation

Again, softer top quark p_{T} spectrum is clearly observed in both regimes, with the results uncertainties being dominated by systematic ones in the resolved regimes, and with comparable statistical and systematic ones in the boosted regime. None of the NLO generators with various underlying NLO models nor different parton shower generators and their tunes can describe well the measured spectrum, with the exception of a new tune of the POWHEG+PYTHIA8 where in the POWHEG settings the h_{damp} parameter is set to $1.5 \cdot m_t$ (but still, in reasonable agreement mostly in the resolved regime only).

Measurements with additional jets

Further measurements were performed as function of additional jets present in the event [ATLAS6], namely in jet multiplicities of 4, 5 and $\geq$ 6. Again, also the p_{out} variable was explored, together with the transverse momentum of the $t\bar{t}$ system, and in the $\geq 6j$ region several predictions are shown to be strongly disfavoured by data, leading also to constrains of the additional QCD radiation modelling uncertainties, with consequences for other later analyses including the $t\bar{t}$ processes, see Figure 7.6. The two variables sometimes give contradictory messages for the same model. This can be partially explained by previous theory thoughts on using different dynamic scales for predicting different kinematic variables [120].

Author's contributions

The author of this thesis was one of the contact editors and member of the analysis team of the paper [ATLAS5]. He contributed to efficiencies extraction and their $t\bar{t}$ modelling systematics; he was responsible for differential corrections evaluation and their model dependence study. He served as one of the two responsibles and co-editors (together with dr. Véronique Boisvert) of the two internal notes and of the publication. As for the analysis published as [ATLAS6], this analysis was enabled

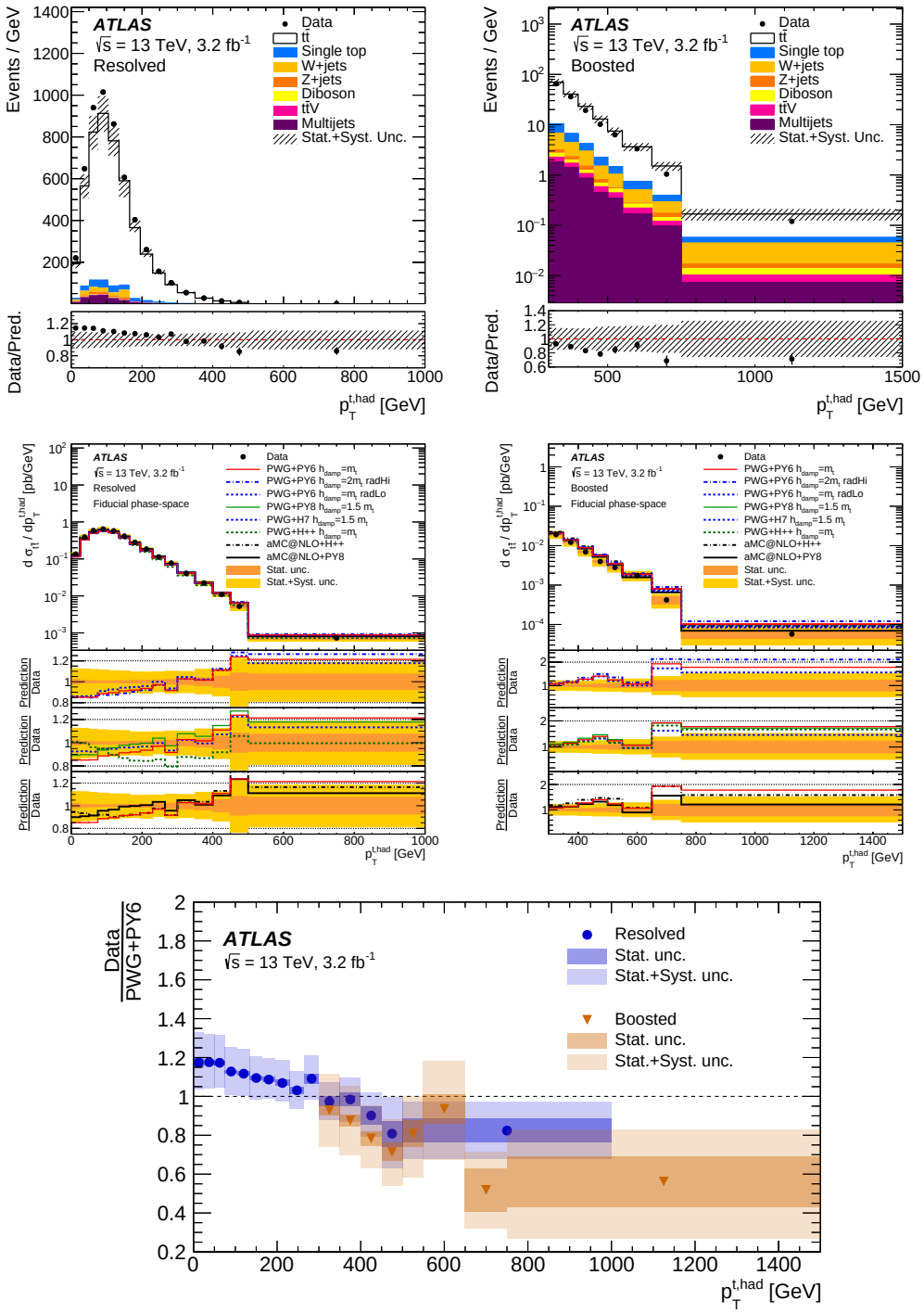


FIGURE 7.5: The first ATLAS analysis of the top quark differential spectra in both resolved (left) and boosted (right) regimes [ATLAS5] at the energy of $\sqrt{s} = 13\,\text{TeV}$ showing the results for the top quark p_{T} at the detector (top) and unfolded the particle (middle) levels. In the bottom plot, the consistency between the two regimes is checked in terms of the ratio to the same NLO generator prediction.

by the previous analysis [ATLAS5] and the author was the analysis team member, contact person for the inclusive analysis; and provided supervision and contributed

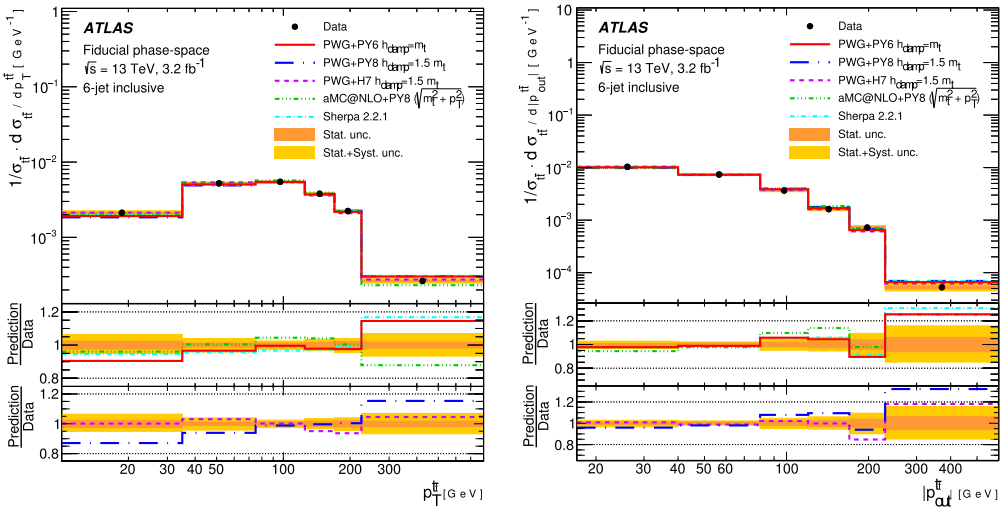


FIGURE 7.6: The differential cross-sections measured by ATLAS as function of the transverse momentum pf the $t\bar{t}$ system ($p_{\mathrm{T}}^{t\bar{t}}$, left) and the absolute value of the out-of-plane momentum of the top quarks (p_{out}, right) in the ≥ 6 jet region at the energy of $\sqrt{s} = 13\,\mathrm{TeV}$, showing different degrees of agreement to various predictions [ATLAS6].

to the internal note.

7.4 Updated 13 TeV ℓ+jets analysis – 1D and 2D spectra

The analysis using about an order of magnitude larger luminosity of than that of the first 13 TeV dataset, amounting to 36 fb^{-1}, corresponds to the data taken between 2015–2016 [ATLAS7] and allowed to study in more detailed the $t\bar{t}$ modelling to provide better constraints for future analyses. The increasing statistics has brought the possibility to measure and unfold also double differential spectra in both resolved and boosted topologies at both parton and particle levels [ATLAS5]

Examples of the 13 TeV spectra, both 1D and 2D, are show in Figure 7.7 for the particle level and in Figure 7.8 for the parton full phase space level.

The results are now systematics limited in precision in both resolved and boosted regimes, challenging further improvements for the need of finding ways to improve systematics uncertainties. The analysis of the full 13 TeV ATLAS dataset is ongoing.

Interpretation

While the NLO generators tuned to the initial 13 TeV analysis perform reasonably well in describing the top quark p_{T} in data, important deviations can be found when this spectrum is also binned in different ranges of the $t\bar{t}$ system invariant mass, see bottom parts of the two figures.

However, one can observe that for the full phase space, the available NNLO prediction [121, 120, 122] does help to describe the double differential spectrum of top quark transverse momentum in bins of the $t\bar{t}$ system mass, as can be seen in bottom of Figure 7.8. Also, the exact novel NNLO prediction describes the data better than the NLO generator, at least in the resolved regime, with an indication of a tighter agreement in the last bin in the boosted case. This is an important message confirming the importance and also correctness of higher order calculations to describe the data. As the experiments gather more data with more LHC collisions delivered, 2D spectra allow to look into less populated yet interesting and important corners of the $t\bar{t}$ kinematics phase space.

Author's contributions

Author of this thesis was a member of the analysis core team of [ATLAS7]. He contributed to the implementation of the bootstrap method preserving statistical correlations between various spectra. He evaluated the efficiency and acceptance corrections and the migration matrices for both 1D and 2D spectra, and their model dependence. He supervised the PhD student Petr Baroň who was granted the exceptional ATLAS co-authorship of the paper, contributing to studies of systematic uncertainties in modelling the single-top and Wt backgrounds.

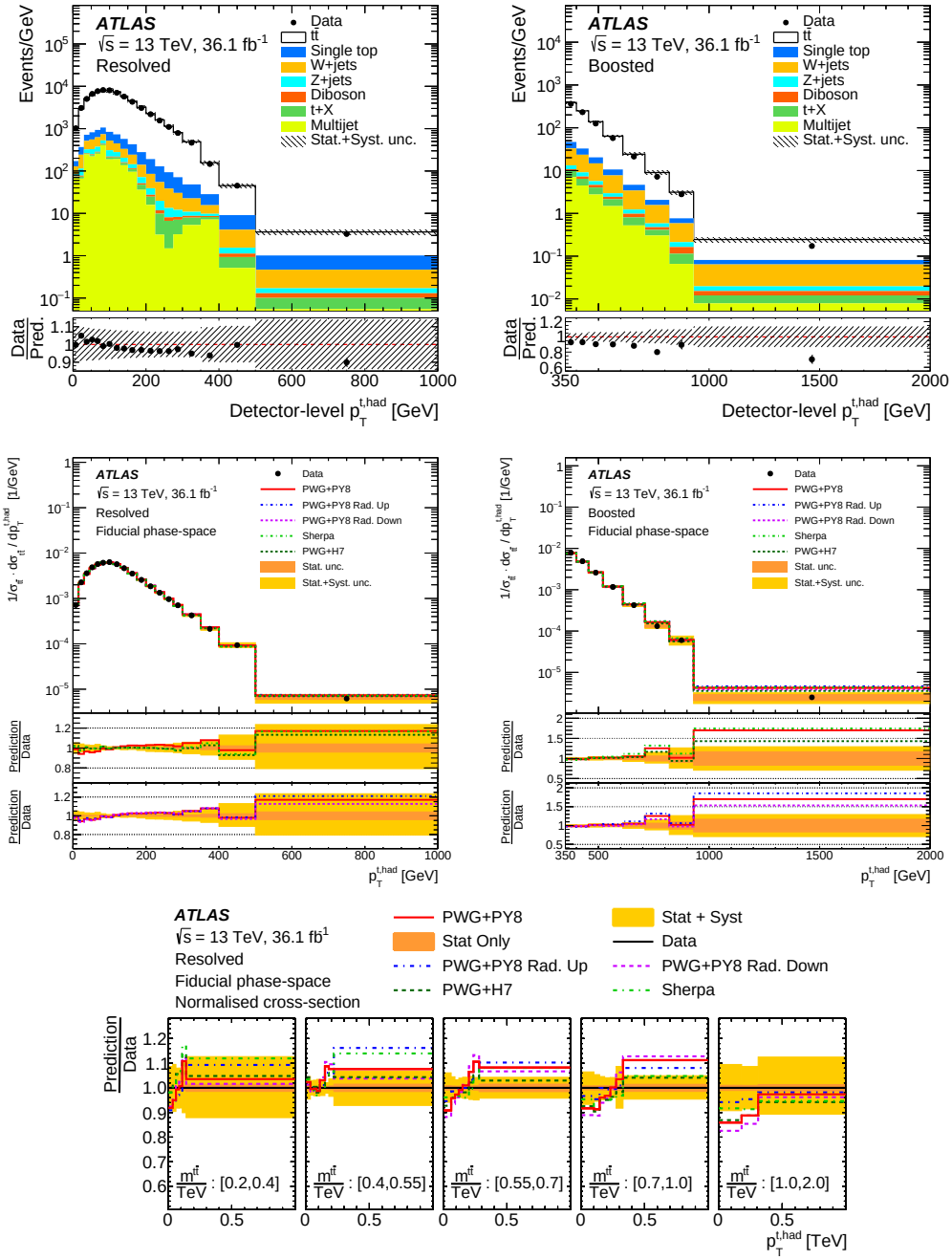


FIGURE 7.7: Results of the ATLAS analysis [ATLAS7] at the energy of $\sqrt{s} = 13$ TeV with a $\sim 10\times$ larger dataset showing spectra of the hadronically decaying top quark in both resolved (left) and boosted (right) regimes at the detector (top) and unfolded to the particle levels (middle). An example of a two-dimensional distribution is shown at the bottom, presenting the top quark p_{T} spectrum in various bins of the $t\bar{t}$ system invariant mass, up to 2 TeV, compared to various NLO generator predictions.

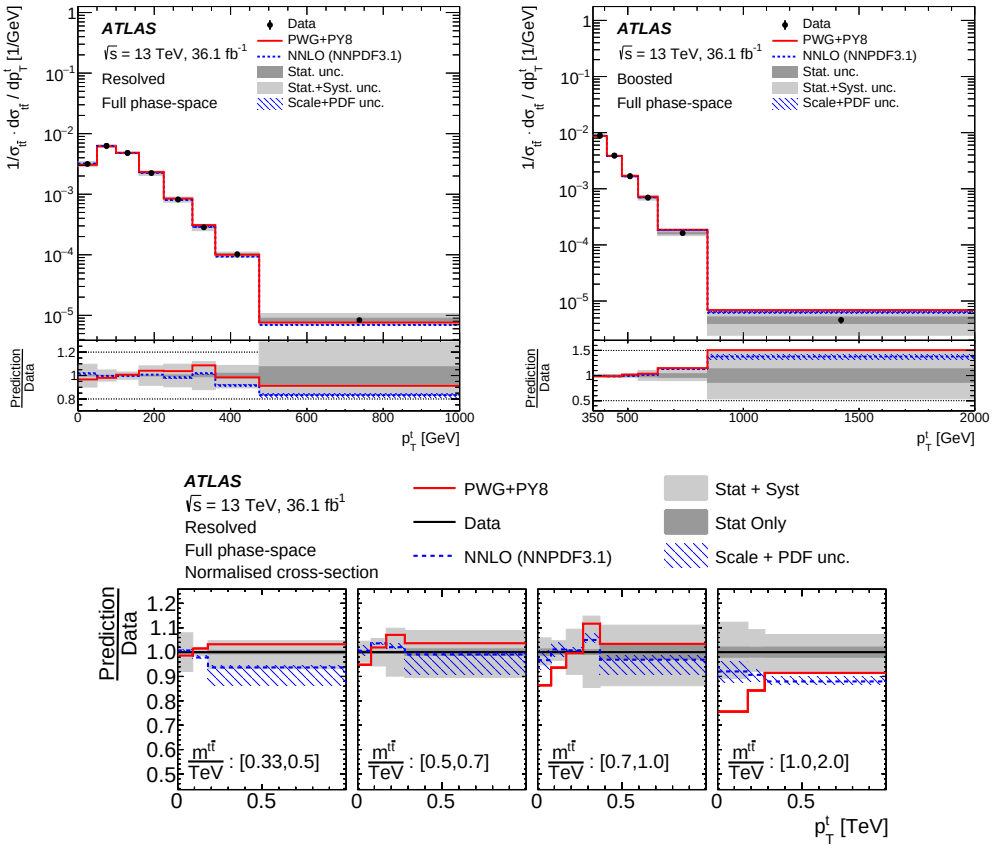


FIGURE 7.8: Results of the ATLAS analysis [ATLAS7] at the energy of $\sqrt{s} = 13$ TeV with a $\sim 10\times$ larger dataset, here showing spectra of the hadronically decaying top quark in both resolved (left) and boosted (right) unfolded to the full phase space parton level (top). An example of a two-dimensional distributions is shown at the bottom, presenting the top quark p_{T} spectrum in various bins of the $t\bar{t}$ system invariant mass, up to 2 TeV, compared to both a NLO generator and NNLO theory prediction.

7.5 Interpretations and the experiment/theory interplay

7.5.1 Monte Carlo generators tuning

Results described in previous sections have been successfully used in various ways in order to improve the modelling of $t\bar{t}$ processes for analyses where top quark constitutes a background, or for modelling the process itself, in order to improve *e.g.* precision of efficiencies extracted using simulations.

One application is the top quark p_{T} reweighting in order the simulation matches the data. The measured variables can be used to tune free parameters in $t\bar{t}$ MC generators in order to better describe data, or so that the variations of these parameters encompass the data; correlations between variables must be taken into account. For example, the p_{out} variable was used to tune QCD radiation systematics controlled by the choice of renormalization and factorization scales and of the h_{damp} parameter in POWHEG, see the analysis [112] which presents MC tuning efforts summary.

7.5.2 Comparison of ATLAS and CMS top quark p_{T} results

An important comparison between the two independent experiments of ATLAS and CMS was made possible by selecting the same binning as provided by the early NNLO+NNLL prediction, see Figure 7.9 where the measured top quark transverse momentum spectra at 8 TeV are compared in the resolved regime on an interval up to 500 GeV. It is observed that the top quark p_{T} spectrum as measured by ATLAS is slightly harder compared to the CMS one and both experiments happen to bracket both the NLO generators as well as the NNLO theory prediction, leaving the 8 TeV state-of-the art not fully understood.

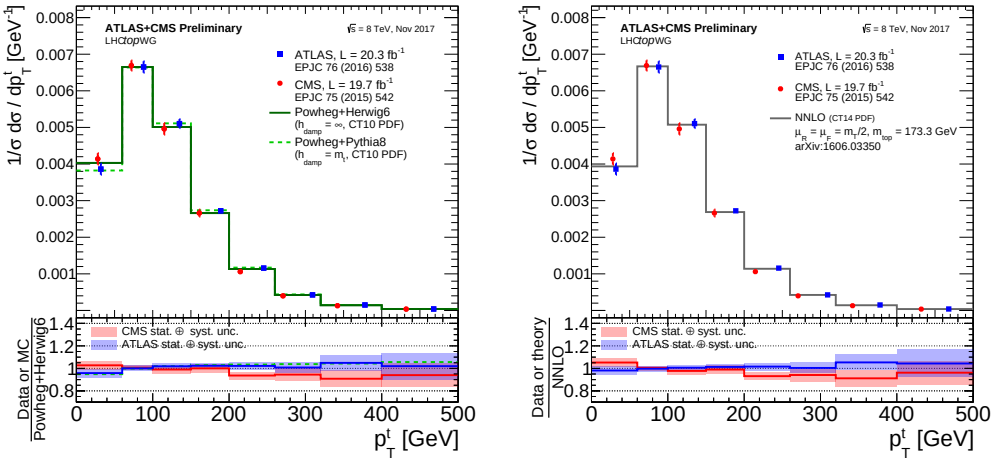


FIGURE 7.9: Comparison of ATLAS and ATLAS unfolded cross-sections as function of the top quark p_{T} to the NLO generators (left) and to the NNLO prediction [31] (right).

7.5.3 Electroweak corrections

An interesting consideration is the rôle and magnitude of electroweak interaction as corrections to the top quark p_{T} spectrum. These have been evaluated at 8 TeV at the LO in QCD, amounting to a negative slope resulting to a drop from zero to about

-13% at $p_{\mathrm{T}}^{t} = 1\,\mathrm{TeV}$, see Figure 7.10, as computed in [123]. However, just these EW effects cannot describe the data excess at low p_{T}^{t} as seen w.r.t. to the NLO generator predictions, *i.e.* the magnitude of the data-to-prediction slope which varies from $+10\%$ to -10% across the full p_{T}^{t} range of $[0, 1]\,\mathrm{TeV}$.

7.5.4 The NNLO theoretical differential predictions

The key and long-awaited theoretical predictions were the differential spectra at the NNLO [120]. The original work performs a comprehensible comparison of multiple 1D differential spectra using various dynamic scales and different PDF sets. The common feature is the behaviour of the NNLO/NLO ratio which exhibits a falling dependence from about $+10\%$ down to -10% to -20% at $p_{\mathrm{T}}^{t} = 1\,\mathrm{TeV}$, thus correcting the NLO predictions into the right direction to better match the observed shape and magnitude of this differential cross-section in data, see example in Figure 7.11

This has truly closed the important and long-standing question of what had been the reason the top quark transverse momentum non-description by NLO predictions: adding another order on pQCD seems to make constitute a remedy. In the boosted regime, with $p_{\mathrm{T}}^{t} > 1\,\mathrm{TeV}$, the main ambiguities affecting the slope of the p_{T}^{t} distribution, are those of the PDF choice.

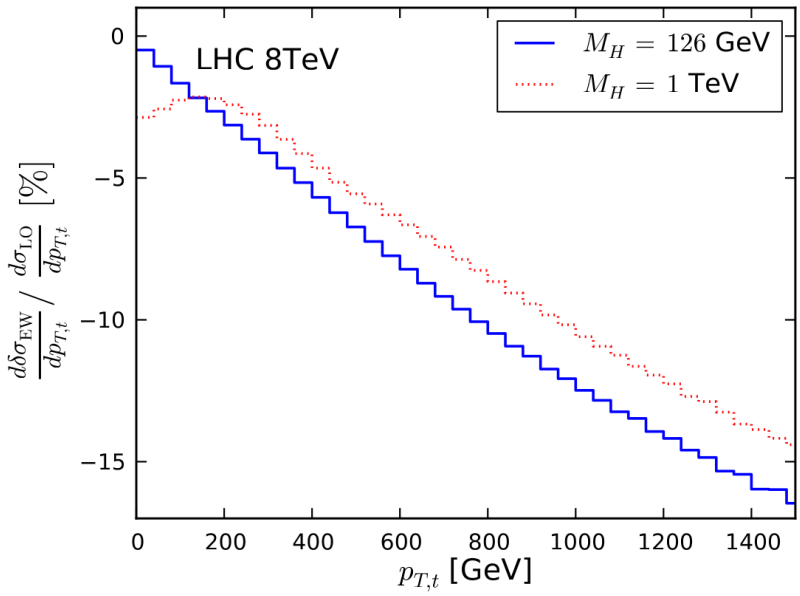


FIGURE 7.10: Electroweak correction to the $t\bar{t}$ differential cross-section as function of the top quark transverse momentum [123] for the two cases of a realistic mass of the BEH boson of 126 GeV (blue) and an example also for a hypothetical case of 1 TeV (red).

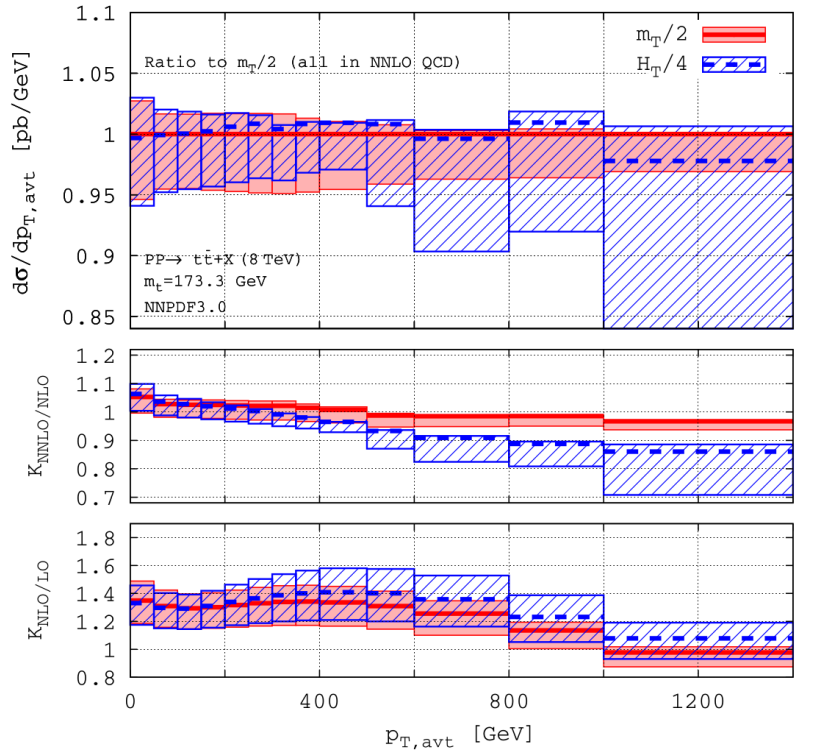


FIGURE 7.11: Ratios of the cross-section as function of the top quark transverse momentum [120]. Top: NNLO / NNLO ratios using dynamic scales of $m_{\mathrm{T}}^{t}/2 = \sqrt{m_t^2 + {p_{\mathrm{T}}^{t}}^2}/2$ and $H_{\mathrm{T}}/4 = (m_{\mathrm{T}}^{t} + m_{\mathrm{T}}^{\bar{t}})/4$. Middle: The NNLO / NLO prediction ratios. Bottom: NLO/LO ratios.

7.6 Top quark in non-standard environments

7.6.1 Top quark pairs in heavy-ion collisions

Top quarks production in heavy-ion collisions delivered by the LHC campaigns with colliding proton-lead or lead-lead beams offer an opportunity to study top quarks production in a low pile-up environment (heavy ion collisions are busy enough in each hadron collision, and so the number of collisions per bunch crossing, μ, is kept low on purpose) and in a different environment of the accompanying nuclear matter. Top quark is thus becoming a probe to nuclear effects in heavy ion collisions.

The CMS experiment has observed the $t\bar{t}$ production in proton-lead [124] at the energy-per-nucleon of 8.16 TeV as well as found the evidence in lead-lead [125] collisions at 5.02 TeV per nucleon, in ℓ+jets and dilepton channels, respectively. In detail, the lead ions being collided at LHC have proton and nucleon numbers $Z = 82$ and $A = 208$. For the Run 2 LHC optics with the nominal beam energy $E_b = 6.5$ TeV this leads to a spectacular energy carried by each of the lead ions Z times of the proton beam, amounting to about 533 GeV, and the central-mass-system energy of two such lead ions exceeding 1 PeV. However, as neutrons are also being carried along within the lead nuclei, the collision energy per nucleon is smaller than in the case of pp collisions at the same momentum. For the four-momenta of the colliding nuclei one can write in the relativistic limit

$$P_p = (E_b, 0, 0, Eb)\,, \quad P_{\mathrm{Pb}} = Z(E_b, 0, 0, -E_b) \tag{7.1}$$

and for average momentum of the nucleon in lead ion $P_{\mathrm{N}} = P_{\mathrm{Pb}}/A$ the energy of the proton-nucleon system is

$$S_{pN} = \left(P_p + P_N\right)^2 = E_b^2 \left[\left(1 + \frac{Z}{A}\right)^2 - \left(1 - \frac{Z}{A}\right)^2\right] \tag{7.2}$$

$$\sqrt{S_{pN}} = 2E_b\sqrt{\frac{Z}{A}}\,. \tag{7.3}$$

In case of the proton-lead collisions this results in $P_N = 2.56$ TeV and $\sqrt{S_{pN}} = 8.16$ TeV for $E_b = 6.5$ TeV, compared to the previous LHC campaign giving $\sqrt{S_{pN}} = 5.02$ TeV for $E_b = 4$ TeV.

Typical Bjorken x for the threshold $t\bar{t}$ production assuming $x_1 = x_2 = x$ is about twice larger compared to the pp case, amounting to $x \approx \frac{4m_t^2}{\sqrt{S}} \approx 0.04$, *i.e.* in the onset of the anti-shadowing nuclear PDF region as described in Section 2.2.5, making the top pair production a highly attractive final state to test the nuclear modification factors.

Other differences of the proton-lead collisions to pp ones are as follows. The pN system is boosted in the proton direction, leading to an average rapidity shift via Lorentz transformation by about ± 0.465 The $t\bar{t}$ cross-section is expected to be enhanced approximately by a factor of A, with possible PDF modification due to nuclear effects (parton distribution function possibly modified by the neighbouring nucleons). The expected (theory) cross-section at the NNLO+NNLL level is about 56 nb, depending on the choice of the PDF set, while CMS measured 45 ± 8 nb.

For lead-lead collisions, one obtains in a similar way

$$P_N = Z/A(E_b, 0, 0, Eb)\,, \quad P'_N = Z/A(E_b, 0, 0, -E_b) \tag{7.4}$$

$$S_{NN} = \left(P_N + P'_N\right)^2 = 4E_b^2 \left(\frac{Z}{A}\right)^2 \tag{7.5}$$

$$\sqrt{S_{NN}} = 2E_b \frac{Z}{A}\,. \tag{7.6}$$

amounting to the energy of the nucleon-nucleon collisions $\sqrt{S_{NN}} = 5.13\,\text{TeV}$ for $E_b = 6.5\,\text{TeV}$ and $\sqrt{S_{NN}} = 5.36\,\text{TeV}$ for $E_b = 6.8\,\text{TeV}$.

For the proton-oxygen ($Z/A = 1/2$) run this gives $\sqrt{S_{NN}} = 9.62\,\text{TeV}$ and for oxygen-oxygen collisions simply $\sqrt{S_{NN}} = 2E_b$ holds. Xenon-xenon collisions taken in 2017 with the isotope $A = 129$ ($Z = 54$) operated at $\sqrt{S_{NN}} = 5.44\,\text{TeV}$.

7.6.2 Top quark pairs in pPb collisions with ATLAS

The ATLAS experiment has observed the $t\bar{t}$ pair production in proton-lead collisions in both the dilepton and ℓ+jets channels [ATLAS8] completing the observation of this process in pPb collisions at the LHC, with a crucial involvement of the author of this thesis, who initiated the analysis at ATLAS. An example of an event display in the dilepton channel is shown in Figure 7.12. The physics opportunities are in the possibility to study the heavy-ion collisions using a new process, at a large scale and for larger parton momenta fractions. The top quark pairs can be used as a new probe to study the nuclear collisions environment and its evolution. Effects like the jet quenching could be evaluated using different decay channels while both can be used to search for possible modifications in the production mechanism via nuclear modifications of the parton distribution functions, especially the gluon ones.

The reconstructed masses of the hadronic W and top candidates are shown in Figures 7.13–7.14, separately in the e+jets and μ+jets and $1b$ and $\geq 2b$ signal fit regions.

The cross-section was extracted using a likelihood fit. The physics parameter of interest, the $t\bar{t}$ signal strength w.r.t. the SM prediction, $\mu_{t\bar{t}}$, was fit along with more than a hundred of nuisance parameters controlling the systematic uncertainties, allowing them to vary in the fit, see the brief method description in Section 4.4.3. The nuisance parameters shifts are shown in Figure 7.15.

This results in the total of six individual channels divided into the cases where one or two and more b-jets were selected, with a further split into the dilepton and e+jets and μ+jets channels, are shown in Figure 7.16. The fit variable was the sum of the transverse momenta of jets and selected isolated leptons in the event, $H_{\text{T}}^{j\ell}$; its pre-fit and post-fit distributions shown in Figures 7.17–7.18.

Interpretation

The extracted cross-section is about twice as precise compared to the CMS result and still compatible to the extrapolated pp measurement by ATLAS and with the theory 7.16. This is the first time the $t\bar{t}$ process is observed in the proton-lead collisions also in the dilepton channel. The measurement opens new paths to using top quarks as a probe in heavy-ion physics and brings the possibility to test possible modifications to nucleon parton distributions functions in the nucleus. The establishment of the $t\bar{t}$ signal can mark the beginning of differential measurements for this process in this new and yet unexplored environment. The resulting measured cross-section is measured to be

$$\sigma_{t\bar{t}} = 58.1 \pm 2.0\ (\text{stat.})\ {}^{+4.8}_{-4.4}\ (\text{syst.})\ \text{nb}\,,$$

i.e. of a total uncertainty of 9%.

The nuclear modification factor is defined as a ratio of the measured cross section to the one obtained from a proton-proton measured cross-section by a simple scaling by A, i.e.

$$R_{p\mathrm{A}} = \frac{\sigma_{t\bar{t}}^{p+\mathrm{Pb}}}{A_{\mathrm{Pb}} \cdot \sigma_{t\bar{t}}^{pp}},$$

where the numerator is measured at the energy $\sqrt{s_{\mathrm{NN}}} = 8.16\,\mathrm{TeV}$ while the denominator is the ATLAS and CMS experiments combined result in pp collisions at $\sqrt{s} = 8$ TeV [126], further extrapolated to the proton-lead nucleon-nucleon collision energy. All uncertainties in these two measurements are conservatively assumed to be uncorrelated. The nuclear modification factor is measured as

$$R_{p\mathrm{A}} = 1.090 \pm 0.039\ (\mathrm{stat.})\ ^{+0.094}_{-0.087}\ (\mathrm{syst.}) = 1.090 \pm 0.100\ (\mathrm{tot.})\,.$$

Thus the measured value is found to be consistent with unity within the uncertainty. A comparison of the the measured $R_{p\mathrm{A}}$ to various theory predictions is in Figure 7.16.

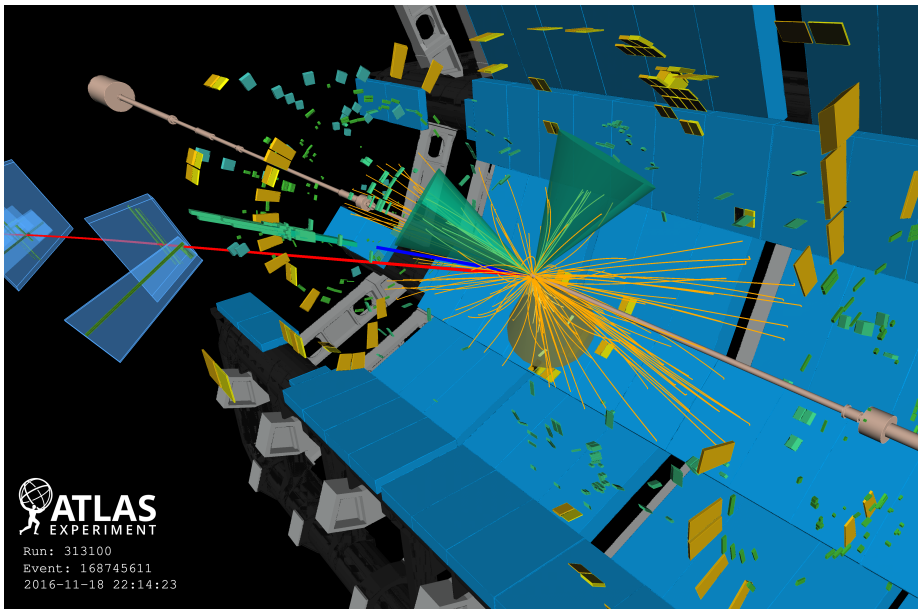


FIGURE 7.12: An event display with the $t\bar{t}$ event candidate in the dilepton channel in ATLAS pPb data [ATLAS8]. Colored are the electron and muon candidates, jets, with a highlighted jet identified as a b-jet.

Further measurements

Finally, for completeness, the ATLAS experiment has recently also observed the $t\bar{t}$ production in lead-lead collision in the dilepton channel [127] using final states with an isolated electron and muon at the energy of the nucleon-nucleon system $\sqrt{s_{NN}} = 5.02\,\mathrm{TeV}$ in data from 2015 and 2018. The resulting measured cross-section

$$\sigma_{t\bar{t}} = 3.6^{+1.0}_{-0.9}\ (\mathrm{stat.})\ ^{+0.8}_{-0.5}\ (\mathrm{syst.})\ \mu\mathrm{b}\,,$$

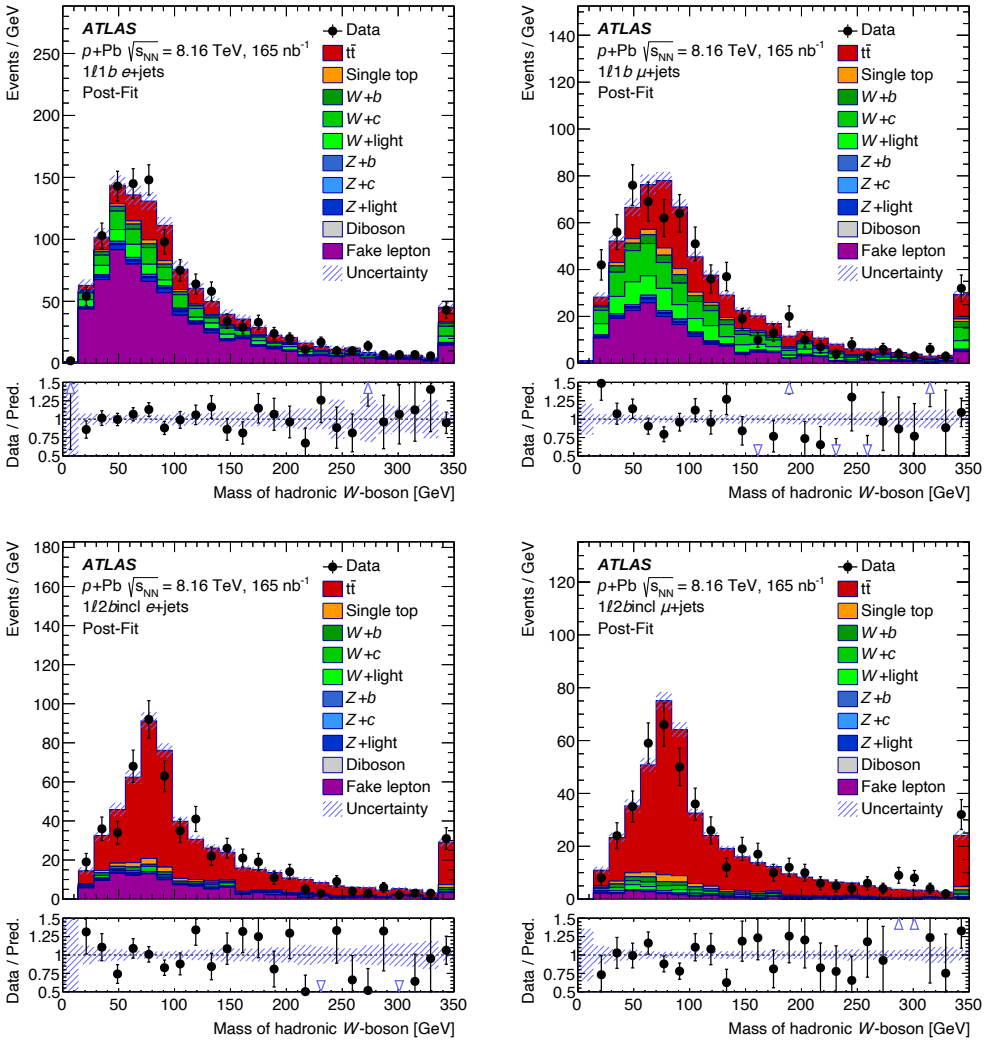


FIGURE 7.13: Invariant masses of the reconstructed hadronic W pseudotop candidates in the e+jets (left) and μ+jets (right) in the $= 1b$ and $\geq 2b$ bins in ATLAS pPb collisions in the ℓ+jets channel [ATLAS8]. Data is compared to the expected backgrounds, with the hatched band indicating the full systematic uncertainty in the prediction. The last bin includes the overflow events.

i.e. of a total relative uncertainty of 31% is in a good agreement with theory prediction at the NNLO+NNLL level [128] at $\sqrt{S_{NN}} = 5.02\,\text{TeV}$ scaled by A^2

$$\sigma_{t\bar{t}} = 2.95^{+0.08}_{-010}\ (\text{scale})\ ^{+0.10}_{-0.09}\ (m_t) \pm 0.21\ (\text{PDF} + \alpha_S)\ \mu\text{b}\,,$$

based on the Top++ programme [98].

Author's contributions

Author of this thesis initiated this measurement [ATLAS8] within the ATLAS collaboration between the Top Quark and Heavy Ion physics groups. He acted as the analysis co-coordinator, under dr. Iwona Grabowska-Bołd, the main analysis framework

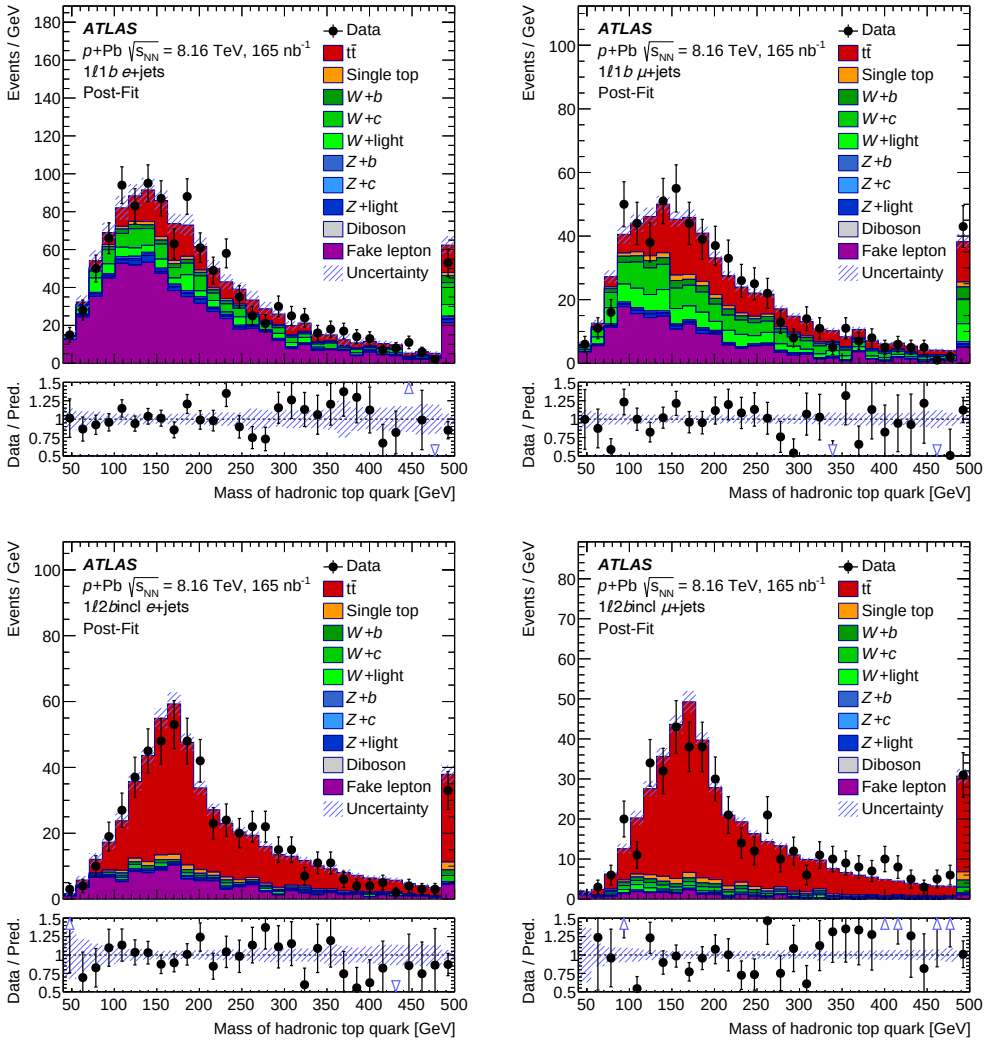


FIGURE 7.14: Invariant masses of the reconstructed hadronic pseudo-top candidates in the e+jets (left) and μ+jets (right) in the $= 1b$ and $\geq 2b$ bins in ATLAS pPb collisions in the ℓ+jets channel [ATLAS8]. Data is compared to the expected backgrounds, with the hatched band indicating the full systematic uncertainty in the prediction. The first (last) bin includes the underflow (overflow) events.

developper, internal note co-editor. He designed the signal and control regions, contributed to the pseudo-top algorithm optimization, the selection and optimization of the fit variable and separation power between various H_{T} variables; he worked on systematics evaluation and studies including $t\bar{t}$ signal modelling and PDF uncertainties; checks of the signal efficiency, correlation among variables, lepton p_{T} reweighting studies, design of fake lepton background shape systematics binned in the W transverse mass; design of the single top background, post-fit control plots, yield tables, the idea of events matching between different jet collections, statistical studies for ratios and systematics cancellation; and to the write-up of both the internal note as well as the final publication.

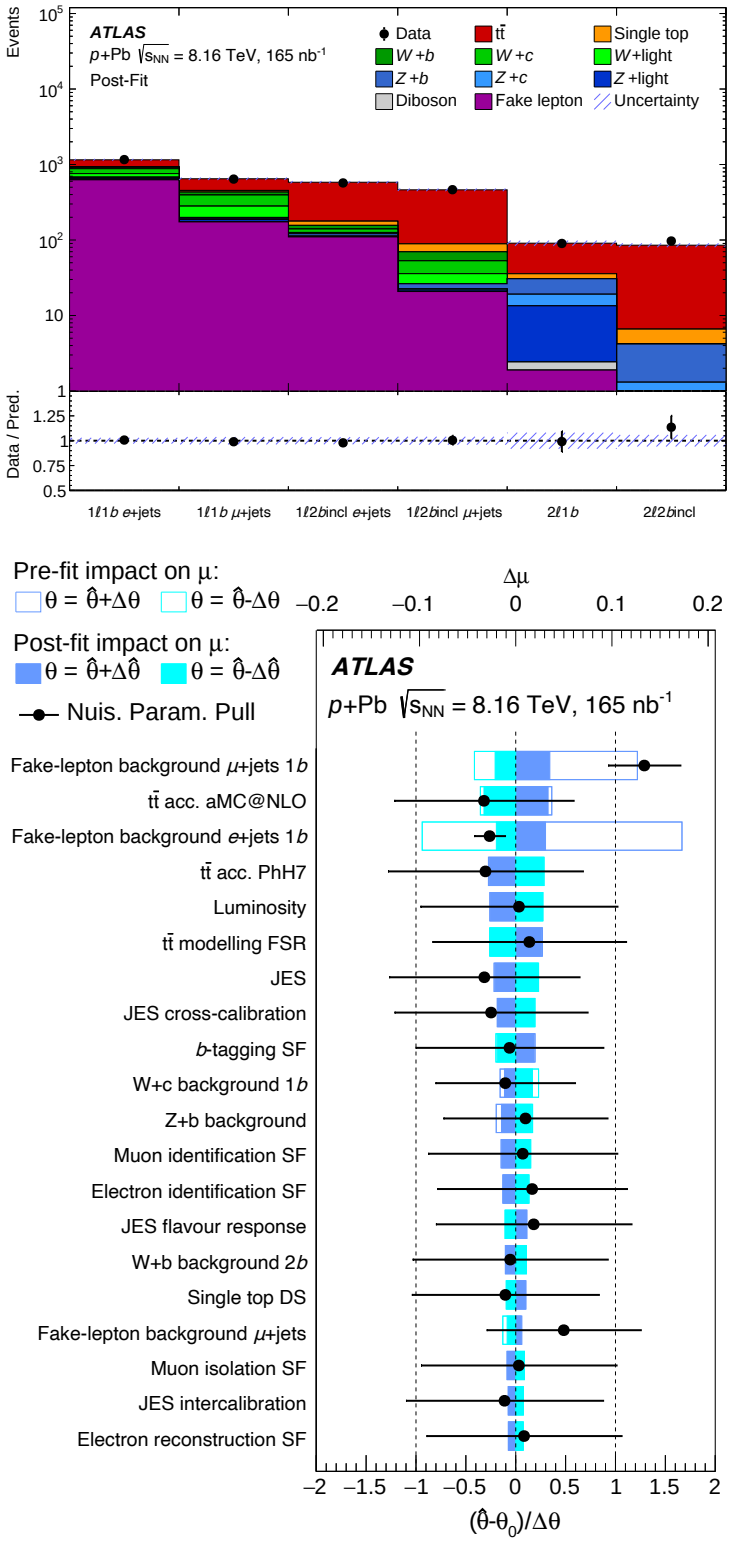


FIGURE 7.15: Top: the post-fit integrated yields in the six fit regions [ATLAS8]; Bottom: the post-fit pulls of the systematics-controlling "nuisance parameters" (points) and their pre-fit and post-fit impact on the fitted signal strength $\mu_{t\bar{t}}$.

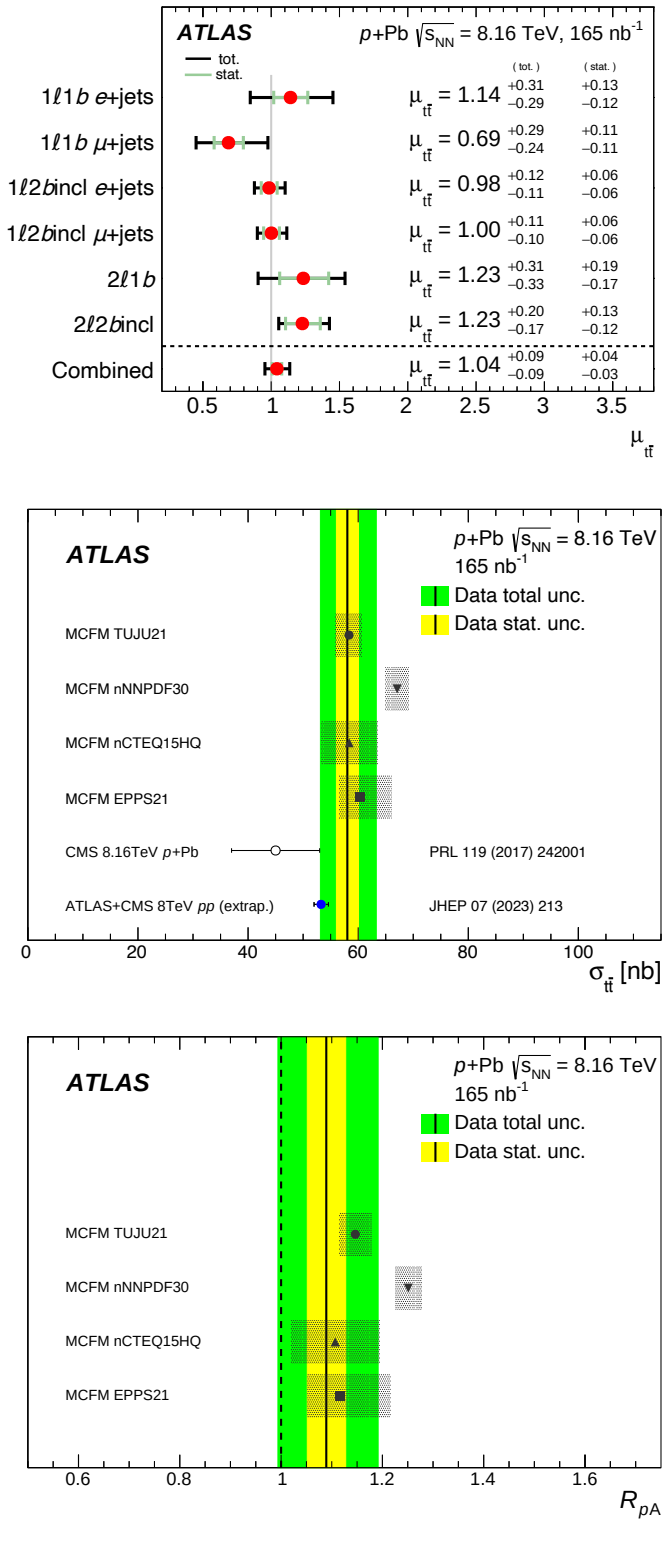


FIGURE 7.16: Fit results for the signal strength $\mu_{t\bar{t}}$ in ATLAS pPb collisions in individual channels (top) and full-fit $\mu_{t\bar{t}}$ transformed to the measured $t\bar{t}$ cross-section compared to the CMS results, various theory predictions as well as to the 8 TeV pp result extrapolated to the 8.16 TeV energy and by the A factor of the lead nucleus [ATLAS8] (middle); and the ratio to the proton-proton ATLAS results (bottom).

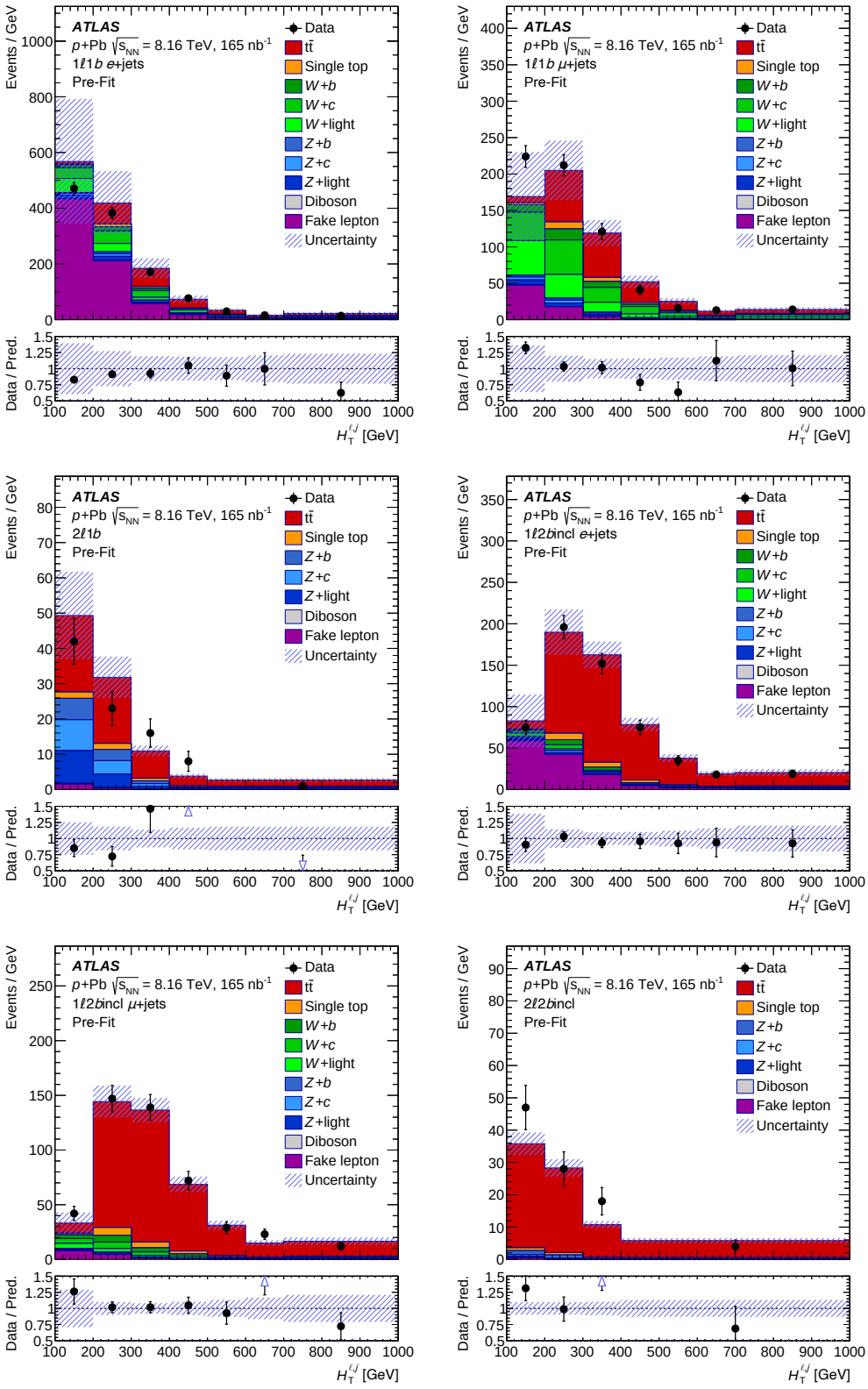


FIGURE 7.17: Pre-fit distributions of the fit variable $H_{\mathrm{T}}^{j\ell}$ in the 6 fit regions in the $1b$ (top) and $\geq 2b$ (bottom) cases [ATLAS8]. The total pre-fit uncertainty band is indicated. The first (last) bin includes the underflow (overflow) events.

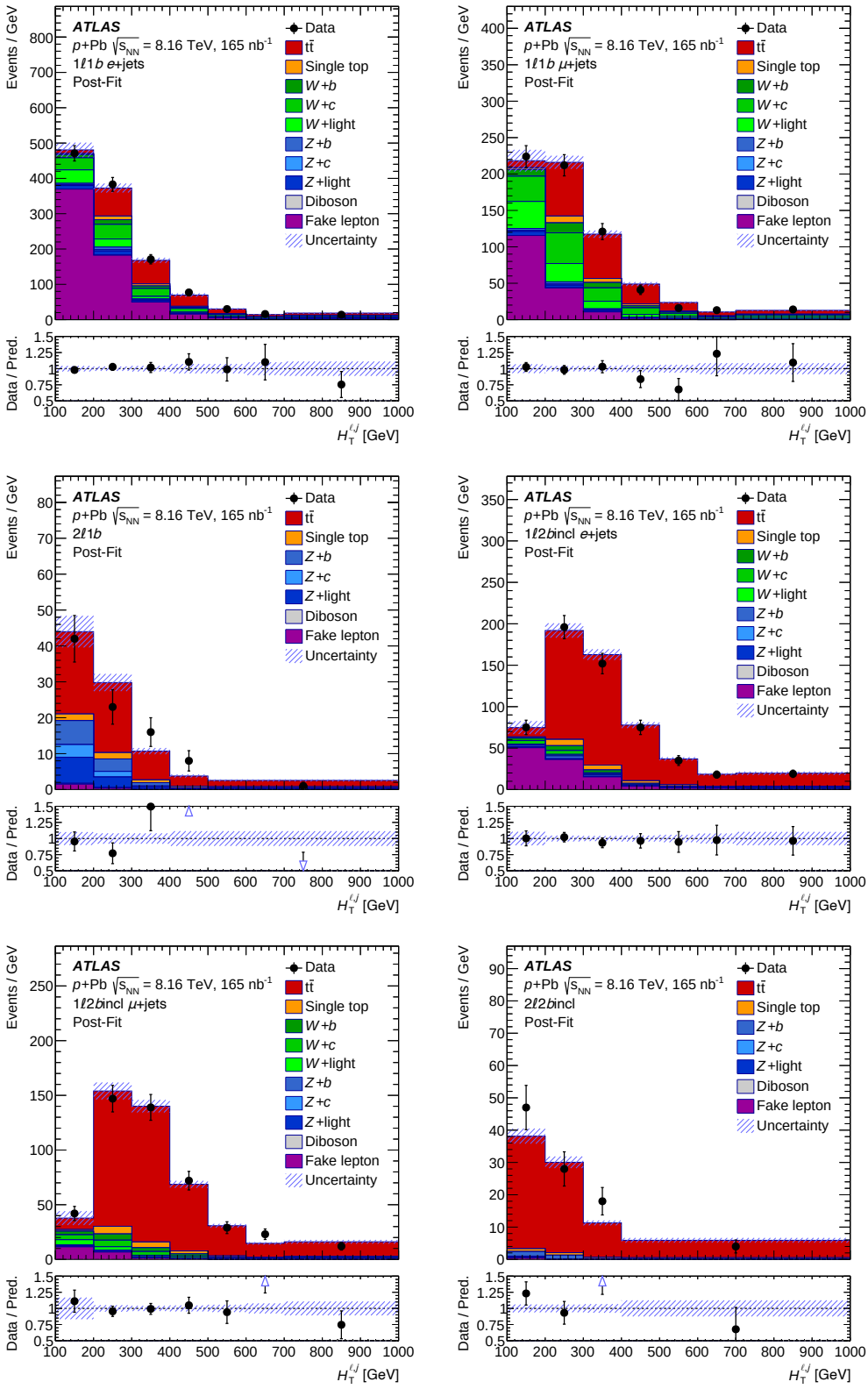


FIGURE 7.18: Post-fit distributions of the fit variable $H_{\mathrm{T}}^{j\ell}$ in the 6 fit regions in the $1b$ (top) and $\geq 2b$ (bottom) cases [ATLAS8]. The total post-fit (therefore reduced) uncertainty band is indicated. The first (last) bin includes the underflow (overflow) events.

7.6.3 Top quark pairs at 5 TeV

Top quark pairs production cross section in pp collisions was also measured at the energy of 5.02 TeV by both the ATLAS [128] and CMS [129] collaborations, adding an important point to the graph of this cross section as function of $\sqrt{S}$. This special LHC run serves as a reference one for the heavy ion collisions campaigns, and is again interesting by its low pile-up environment.

7.6.4 Top quark production in forward rapidity regions

The LHCb experiment is fully instrumented only towards one side of the interaction region in a forward region of rapidities has also observed top quarks pair production [115] allowing to check the theory prediction in this non-standard phase space. The measured $t\bar{t}$ cross section at 7 and 8 TeV is of the order of 200 and 300 fb, in agreement with theory predictions of 180 and 300 fb, respectively, all with uncertainties of tens of per-cent.

7.7 Perspectives with LHC Run 3 and beyond

The LHC Run 3 started in September 2022 at the unprecedented energy of pp collisions of 13.6 TeV. The first ATLAS publication from Run 3 actually concerns the ratio of the top quarks pair cross section to the Z boson cross section [130] while the CMS experiment measured the first preliminary $t\bar{t}$ pair cross-section [131]. Run 3 will continue till middle 2026 to be followed by the Long Shutdown 3 of the LHC complex before the High Luminosity LHC phase will start in 2030, expected to last till 2041.

Top quark measurements in heavy ion collisions currently merely establish the presence of the $t\bar{t}$ signal. What awaits the discovery in this environment are the electroweak single top processes of the single top production. With more statistics, doors will open to differential measurements in proton-lead or lead-lead collisions which will not only open the possibilities to study the nuclear modifications to parton distribution functions, but in addition will allow to measure transverse momenta of leptonically and hadronically decaying top quarks which could carry imprints of interactions of top quarks and their decay products with the dense matter in collisions of nuclei; and shed more light on the phenomenon of jet quenching.

On the side of ATLAS and CMS comparison, a join effort on a common NLO Monte Carlo generator with the same settings is ongoing in order to compare unfolded cross-sections using the same underlying model for corrections at both experiments. Unification procedures are being set up in order to define common systematics uncertainties treatment between the two experiments.

From the theory side, NNLO $t\bar{t}$ generators are approaching applicability in full simulation chains of the main detectors at the LHC. They will be important not only in more stringent test of the pQCD in $t\bar{t}$ production, but also for modelling top quark processes where they are background to possible new physics phenomena.

Forward physics activities of both ATLAS and CMS experiments can provide a complementary view on the top quarks physics in cases where one or both protons remain intact in the collisions. While simple observations of these processes would constitute important tests of the diffractive physics picture, factorization and production mechanisms, top quarks could also couple to new physics with a potential to be experimentally more prominent in exclusive diffractive processes compared to inelastic proton-proton collisions.

Chapter 8

Searches for new physics

Presence of new physics at the LHC can express itself as signal in yet unexplored channels or detector signatures, as modification to mass or other spectra beyond current statistical precision or may be awaiting beyond current kinematic reach. Even in such a case, modification of tails of distributions may reveal the signs of such new physics, and in such a region (as illustrated in Figure 8.1 (top), a global approach of effective field theories (EFT) is useful in describing and allowing to interpret what effective terms in a Lagrangian could be responsible for the observed effect while the terms may be a consequence of several possible BSM theories. Also, interference effects between a SM and BSM signals leading to the same final states are possible and can lead to modified sensitivity to the BSM signal [132].

8.1 Motivation for searches in $t\bar{t}$ final states

Search for the $t\bar{t}$ resonances have been performed by both the ATLAS ([133, 134]) and CMS ([135, 136, 137, 138]) experiments, setting mass limits on BSM Z' particles of various scenarios. As the Large Hadron Collider (LHC) at CERN has provided a large dataset and current upgrade works aim increasing the center-of-mass energy as well as the luminosity, boosted topologies in hadronic final states are becoming heavily used tools in both measurements and searches at the LHC. They provide a handle on processes at large momentum transfer, with a chance of observing possible BSM signals. The aim of the study presented in this Chapter is to identify the behaviour of selected models in terms of their effect on modification of kinematic spectra, their behaviour in the unfolding procedure, and as models to identify observables sensitive to the presence of such BSM signals.

8.2 Study of selected benchmark models

From the plethora of available models of BSM physics models coupling to top quarks, we consider the following models and final states:

- an additional vector boson Z' which couples to quarks [139, 140, 141] can produce a $t\bar{t}$ pair via the s-channel process of the quark-antiquark annihilation. The mass of the Z' particle was selected as 1 TeV, with its width being vanishing, according to the default model parameters. This thus constitutes a model of a narrow vector $t\bar{t}$ resonance;

- a scalar resonance y_0 [140] which can be produced within a triangle loop initiated by a gluon-gluon fusion. More details on this particular model can be found in [142, 143, 144, 145, 146, 147, 148, 149]. The mass of the y_0 particle

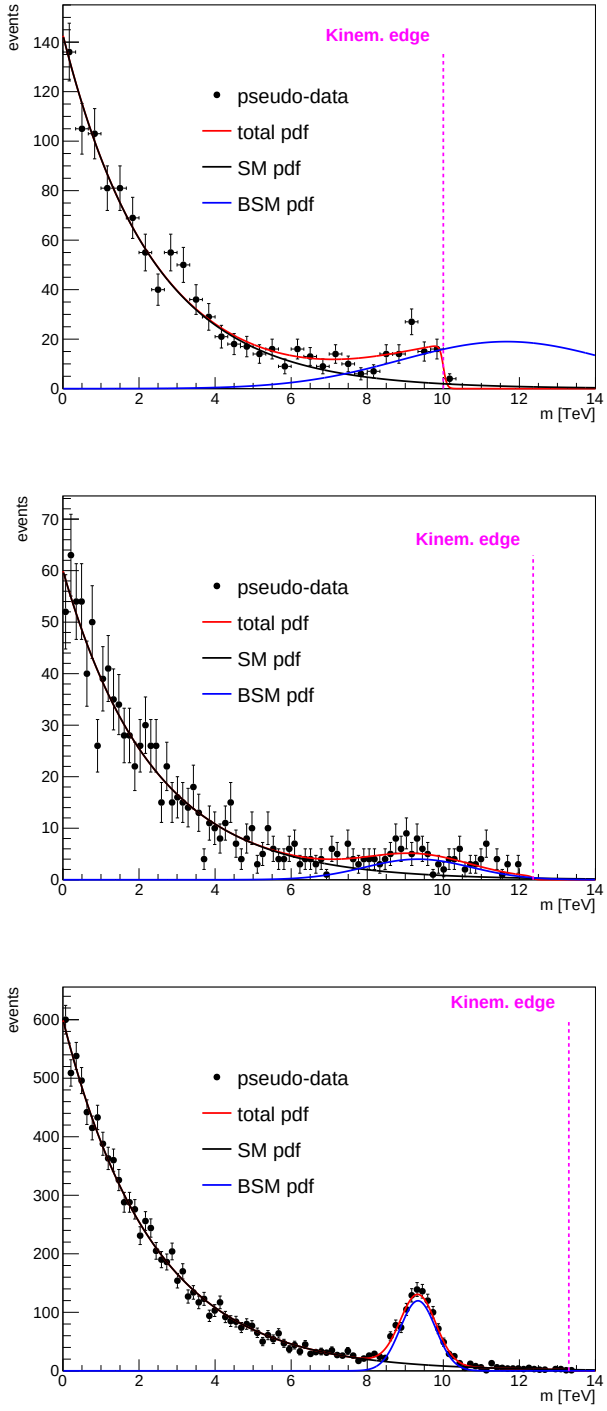


FIGURE 8.1: Illustration of possible new physics appearance as a resonance in some invariant mass spectrum using a generally falling SM and a peaking BSM probability distribution function. Top: new physics beyond the kinematic reach affects tails of distributions. Middle: new physics is within the kinematic reach but at the edge of the statistical observation. Bottom: fully observed peak is possible by both statistical and kinematics considerations. The kinematic edge, the energy reach of the hypothetical accelerator, is modelled by a sigmoidal function. Pseudo-data were generated according to the total $= (1 - f_{\text{sig}})\,\text{SM} + f_{\text{sig}}\,\text{BSM}$ probability distribution function.

was also selected as 1 TeV, and its width as 100 GeV, constituting a model of a larger-width scalar $t\bar{t}$ resonance;

- the process of an associated production of a fermionic dark matter candidates pair in association with a $t\bar{t}$ pair, with the DM pair coming from the decay of the y_0 scalar; the DM particle χ_D carrying-away energy and momentum which directly translates to a missing transverse energy. The mass of the DM particle was selected as 100 GeV in order to allow a higher energy threshold and thus more boosted regimes, and large missing transverse energy as an experimental signature.

Example Feynman diagrams for these processes are shown in Figure 8.2.

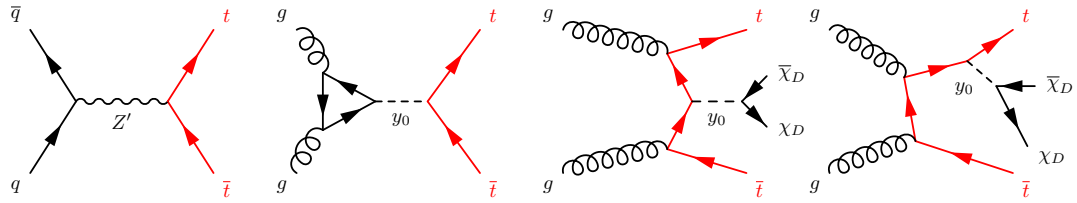


FIGURE 8.2: Example leading-order Feynman diagrams for selected BSM processes of the $t\bar{t}$ pair production. Created using the *FeynMf* package [26].

A whole landscape of variables is available to be explored, with different sensitivity to different BSM models, constituting a multidimensional world where different and surprising levels of separation between SM and BSM signal are possible

In next we shall focus on the all-hadronic decay modes and their boosted jet signatures, as such jets may stem from the production of a heavy resonance. Besides the search for a bump in the distribution of the $t\bar{t}$ pair invariant mass we shall study more angular as well as momentum related variables and will evaluate their sensitivity to reveal the presence of a BSM signal. As described in Section 3.6.2 and in Figure 3.8, we shall consider the 0B2S, 1B1S and 2B0S topologies, corresponding to the presence of two semiboosted top quarks (*i.e.* two W-tagged hadronic jets); one boosted and one semiboosted top quark (*i.e.* one W-tagged and one top-tagged hadronic jet); and two boosted top quarks (*i.e.* two top-tagged hadronic jets); respectively.

8.2.1 Analysis strategy and overview

In the following analysis using own simulations, models of a scalar and vector resonant production of the $t\bar{t}$ pair at 14 TeV c.m.s. energy of the pp system are studied, using the MADGRAPH5 simulation and DELPHES parameterized detector simulation with the ATLAS card. We shall only briefly describe the objects and event selections in different topologies based on the number of W and top-tagged jets in the events. More details can be found in [Auth4].

Events considered in analysis are simulated proton-proton collision events at $\sqrt{s} = 14$ TeV and reconstructed by DELPHES ATLAS detector simulator. The DELPHES card was modified to allow both R (0.4) and large-R (1.0) jets be reconstructed at both particle and detector levels using the FastJet algorithm [150], removing the DM particles from both detector jet clustering as well as from the particle jets. Custom jet energy scale was applied to large-R jets at the detector level.

The event selection requirements can be summarized as follows:

- require at least two large-R jets with $p_{\mathrm{T}} > 80\,\mathrm{GeV}$ and $|\eta| < 2.0$, to enhance the semiboosted and boosted regimes;
- preselect small-R jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ to use jets in well-understood parts of the calorimeter;
- require at least two small-R jets tagged as b-jets, in order to suppress non-$t\bar{t}$ backgrounds.

The widely-used jet subjettiness variables [151] were employed together with a simple selection on the large-R jet mass top select jets coming from the hadronic decays of the W boson or a top quark, tagged as W or top jets. More details can again be found in the preliminary study [Auth4]. Next, the event classification is attempted in this order: the boosted-boosted (2B0S), boosted-semiboosted (1B1S) or semiboosted-semiboosted (0B2S) based on the number of the W and top tagged jets.

8.2.2 Additional kinematic variables and BUMPHUNTER 2D extension

Additional variables compared to the "classical" ones presented in Section 6.3 have been proposed, being global event variables like the invariant (visible) mass of the four-vector sum of the selected large-R jets in the event ($m^{\mathrm{vis}}_{\mathrm{sum\,J}}$), the sum of p_{T}'s of small-R jets (H^{j}_{T}) and the missing transverse energy ($E^{\mathrm{miss}}_{\mathrm{T}}$), and the sum of the two. Further, dimensional-less variables relative to the mass of the $t\bar{t}$ system as well as relative to the geometrical mean of the two top quarks transverse momenta were also constructed, in order to diminish the dependence on the absolute jet energy scale calibration which constitutes one of the largest experimental uncertainty. In addition, 2D variables combining the aforementioned 1D variables were also studied. It turned out, for example, that the p_{out} variable has a negligible correlation to $m^{t\bar{t}}$ while it exhibits large shape change especially in tails for models with additional DM particle produced.

The BUMPHUNTER algorithm (see Section 4.4.7) and its extension to a 2D was then tested using these detector-level distributions over pseudo-experiments using the ATLAS Bootstrap method [83] on a plethora of kinematic variables (a total of 253), their ratios and 2D spectra performance tested in [Auth4], resulting in several variables showing interesting capabilities to separate a BSM signal from the $t\bar{t}$ background, and also being robust to systematic uncertainties, with implications to LHC $t\bar{t}$ analyses and searches.

8.2.3 Shape comparison of BSM models

The following discussion is partially based on [Auth4] where a more complex analysis including systematics uncertainties effects is presented. As an example a selection of variables shape comparisons for the 1B1S topology is presented. The most useful variable for searching for a resonant $t\bar{t}$ production is naturally the invariant mass of the $t\bar{t}$ system, $m^{t\bar{t}}$. The shape comparison shows a clear peak for the vector (Z') as well as the scalar (y_0) models, see Figure 8.3. In contrast, the DM model leading to the associated production of a $t\bar{t}$ pair and a pair of DM particles ($\chi_D\bar{\chi}_D$) leads not only to an enhancement of the missing transverse energy where such a signal is often looked for, but also to a more prominent tail in the out-of-plane momentum distribution (p_{out}), see Figure 8.4.

The spin of the resonance decaying to a pair of top quarks can be determined from the $\cos\theta^*$ distribution, as provided in Figure 8.5, as it leads to a slightly flatter

spectrum for the scalar y_0 case but to a more peaked shape for the case of a Z' production. Note that different effects in peak and tail from different models are present in the p_{out} distribution.

Shapes of other variables are also modified for such a signal, *e.g.* the top quark transverse momentum in Figure 8.6, exhibits a broader peak rather than just a slope change. Figure 8.7 shows the various models shapes as function of the missing transverse energy with the striking signature of additional invisible particles produced along with the $t\bar{t}$ pair. Similar observations were found also in the 2B0S and 0B2S event topologies.

The Z' model was generated at the NLO accuracy in pQCD, with additional jets allowed to be produced by both the hard process or by the parton shower. This leads to some of the energy of the process not being accounted for in the invariant mass of the $t\bar{t}$ pair as reconstructed from the two leading hadronic large-R jets, in contrast to the scalar y_0 model which includes a loop already in the production, not allowing a consistent addition of additional jets in the final state. Thus, the $t\bar{t}$ invariant mass for the scalar-mediated $y_0 \to t\bar{t}$ process has a larger tail to higher masses compared to the vector-mediated $Z' \to t\bar{t}$ process. This also translated to a slightly softer top quark p_{T} spectrum for the vector model.

8.2.4 BSM models at the detector level

For the purpose of creating a realistic mixed sample of a selected SM background (associated production of a W boson and b-jets), the $t\bar{t}$ sample as well as a new physics signal, detector-level distributions were created by mixing the samples based on their cross section, but also scaling the signal in order to make it more prominent to study its shape and unfolding performance. We present examples in the semi-boosted topology, *i.e.* in the case where one large jet is of mass and substructure consistent with the hypothesis of coming from the hadronic t decay while another large-R jet coming from the hadronic W decay. Figures 8.8 and 8.9 show examples of the stacked $t\bar{t}$, non-$t\bar{t}$ as well as BSM signal models compared with pseudo-data composed of the same models but from statistically independent samples. Binned 1D significance of a signal over the total $t\bar{t}$ and non-$t\bar{t}$ background is evaluated at the detector level and is shown as a red (black) curve in the lowest pad without (with) taking into account the uncertainties of the background.

Figure 8.8 shows the $t\bar{t}$ pair invariant mass distribution, with the expected enhancement around the generated mass of the y_0 scalar of 1 TeV; and also the transverse momentum of the top quarks for the same model, showing a broader enhancement. Figure 8.9 shows a similar comparison for the missing transverse energy, compared between the resonant $y_0 \to t\bar{t}$ model and the one with the $t\bar{t}$ production in association with the produced new physics scalar y_0 decaying to a pair of invisible DM particles of mass of 100 GeV. The signal significance is large across the $E_{\text{T}}^{\text{miss}}$ for the former while substantial only in the highest-$E_{\text{T}}^{\text{miss}}$ bin for the latter model.

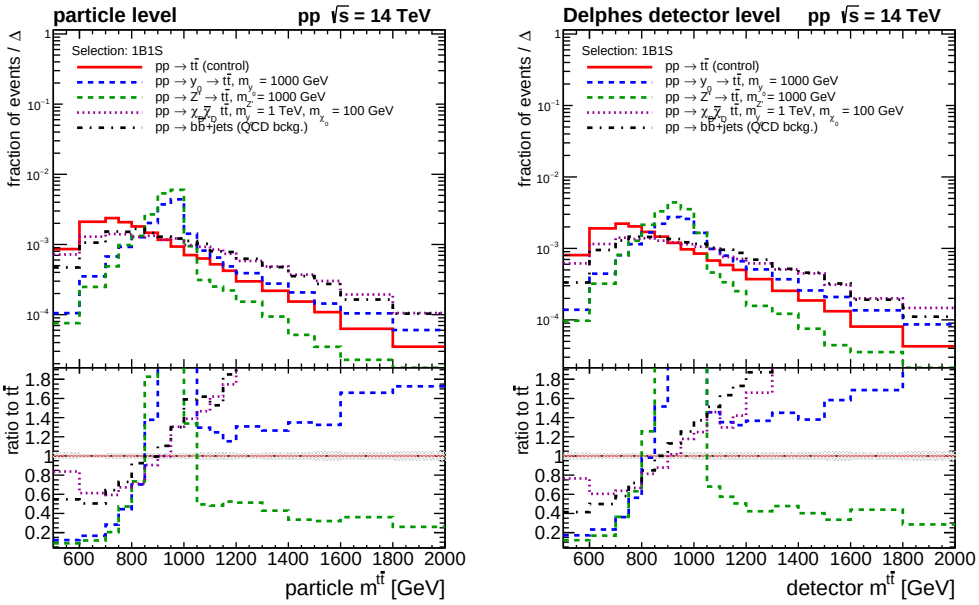


FIGURE 8.3: Shape comparison for the invariant mass of the $t\bar{t}$ pair in the boosted-semiboosted topology. Left: particle level, right: DELPHES ATLAS detector level. A clear peak of a hypothetical resonance can be seen for the scalar (y_0) and vector (Z') BSM models [Auth4].

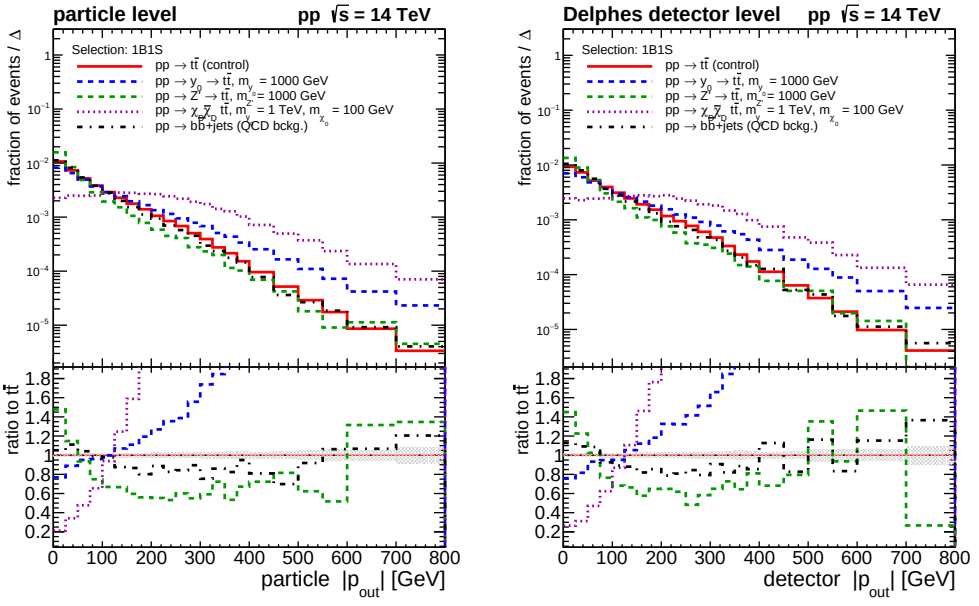


FIGURE 8.4: Shape comparison for the absolute value of the out-of-plane momentum $|p_{\text{out}}|$ in the boosted-semiboosted topology. Left: particle level, right: DELPHES ATLAS detector level. A large modification can be seen for the associated production of a $t\bar{t}$ with an undetected pair of non-interacting DM particles $\chi_D\bar{\chi}_D$ [Auth4].

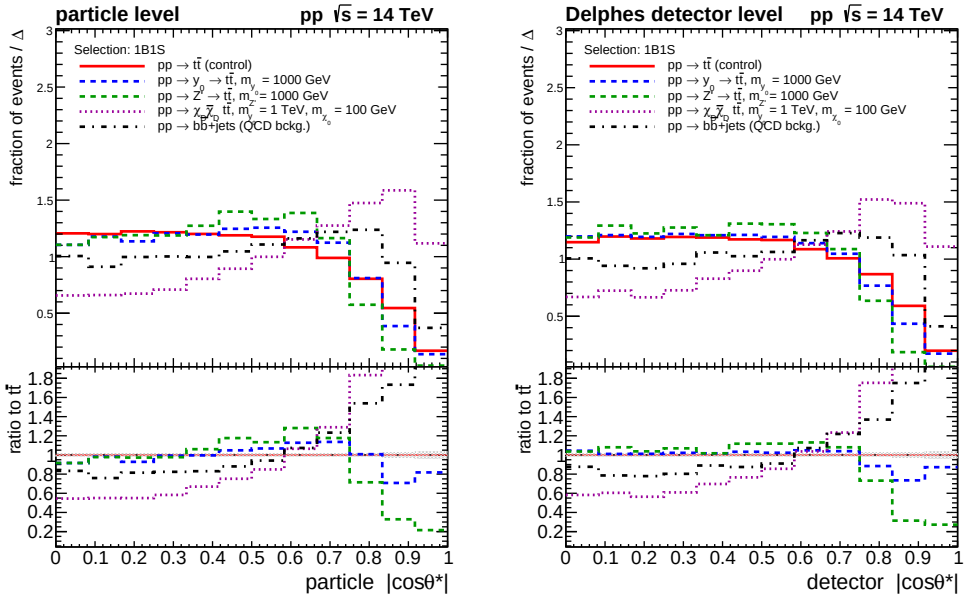


FIGURE 8.5: Shape comparison for the absolute value of the $\cos\theta^*$ variable in the boosted-semiboosted topology. Left: particle level, right: DELPHES ATLAS detector level. A more uniform distribution is observed for the scalar resonance y_0 decaying to a pair of top quarks while the process with additional DM particles pair production peaks in the more forward region [Auth4].

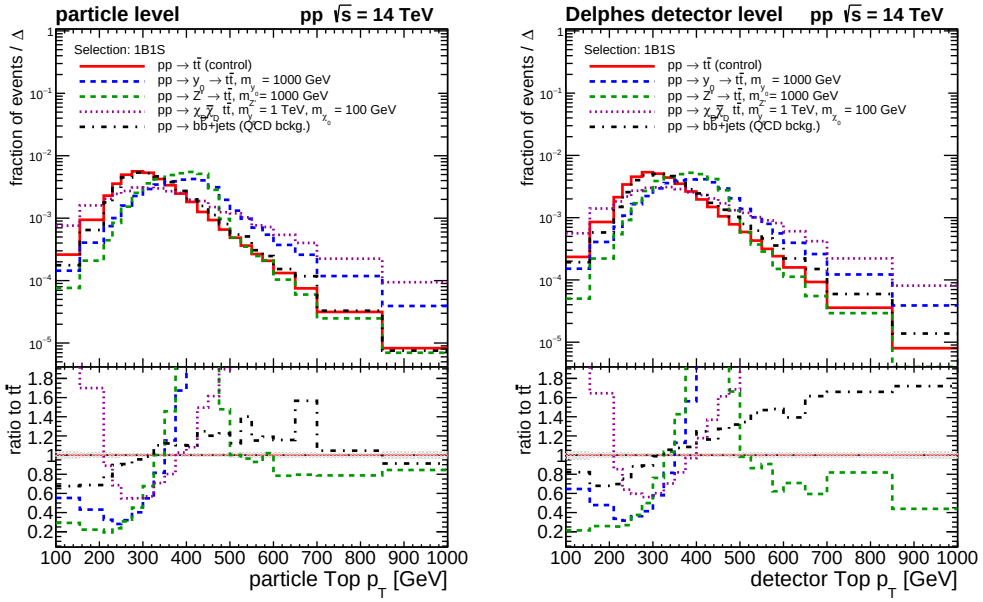


FIGURE 8.6: Shape comparison for the p_{T} of the hadronic top quark candidate in the boosted-semiboosted topology. Left: particle level, right: DELPHES ATLAS detector level. The top quark p_{T} is harder for all BSM models [Auth4].

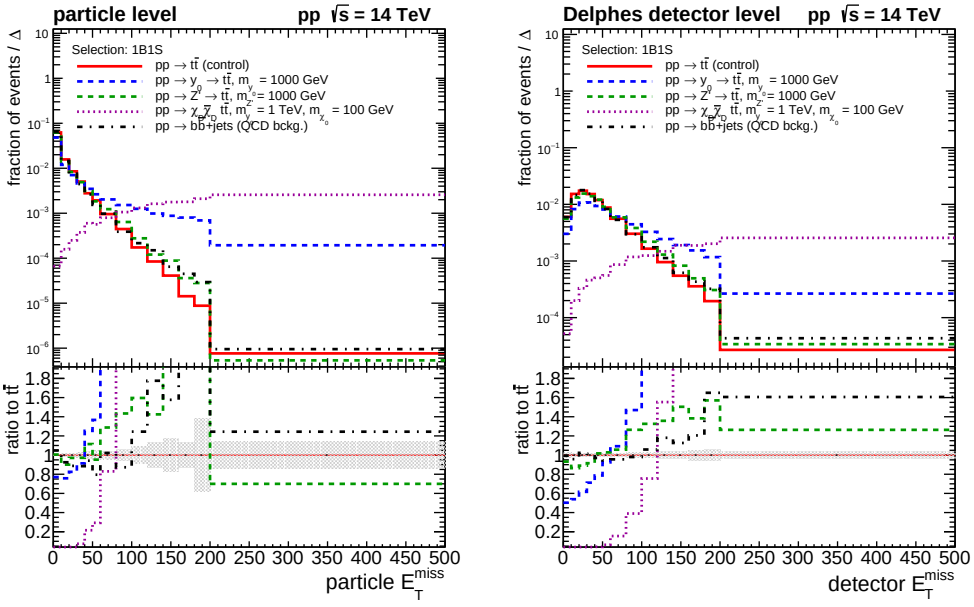
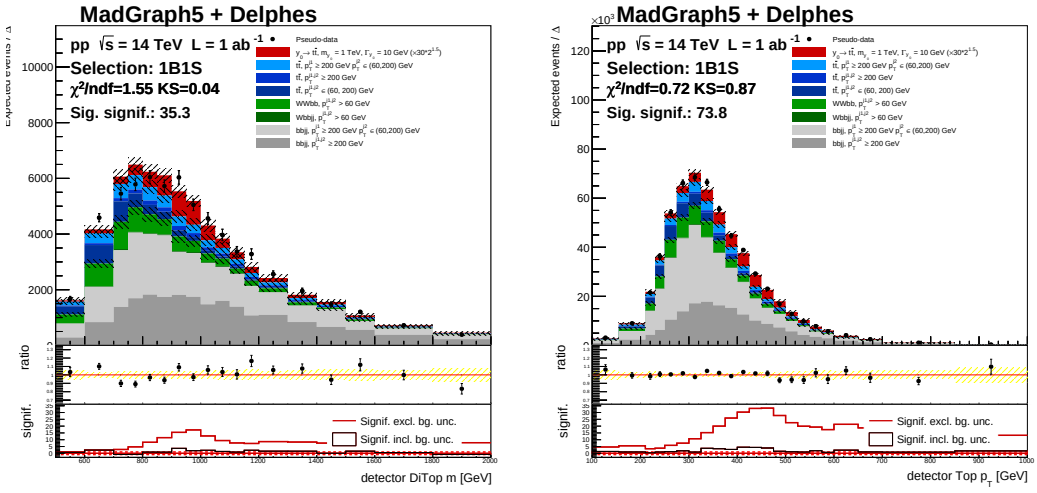
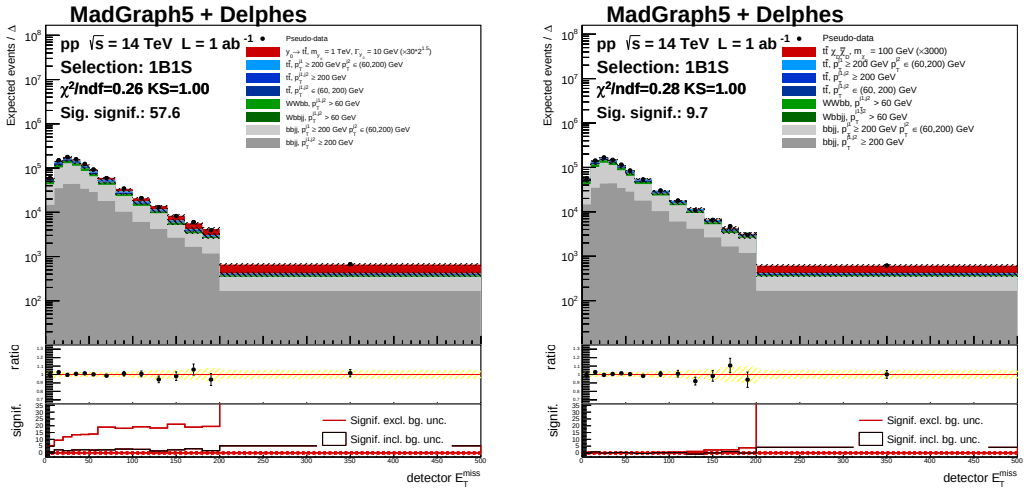


FIGURE 8.7: Shape comparison for the missing transverse energy in the boosted-semiboosted topology. Left: particle level, right: DELPHES ATLAS detector level. A large deviation towards high $E_{\mathrm{T}}^{\mathrm{miss}}$ is observed for the model with an associated production with an invisible pair of the DM particles [Auth4].



$y_0 \to t\bar{t}$ $y_0 \to t\bar{t}$

FIGURE 8.8: Stacked invariant mass of the $t\bar{t}$ pair (left) and the top quark transverse momentum (right) at the DELPHES detector level in the 1B1S topology for the $y_0 \to t\bar{t}$ signal model with both the pseudo-data/prediction ratio and the signal significance shown in lower pads [Auth4].



$y_0 \to t\bar{t}$ $t\bar{t} + y_0, y_0 \to \chi_D\bar{\chi}_D$

FIGURE 8.9: Stacked $E_{\mathrm{T}}^{\mathrm{miss}}$ at the DELPHES detector level in the 1B1S topology for the $y_0 \to t\bar{t}$ and the $y_0 \to \chi_D\bar{\chi}_D$, $m_{\chi_D} = 100\,\mathrm{GeV}$ with both the pseudo-data/prediction ratio and the signal significance shown in lower pads [Auth4].

8.2.5 Identifying regions of best signal-to-background separation

In Figure 8.10, selected pairs of variables illustrate the advantage of computing the 2D BUMPHUNTER score (see [Auth4] for details) instead of just the 1D scores. For the $y_0 \to t\bar{t}$ scalar resonance model, the (visible) mass of the four-vector sum of larger-R jets $m^{\text{vis}}_{\text{sum J}}$ performs similarly in signal separation to backgrounds (dominate by the SM $t\bar{t}$ production) as the scalar sum of small-R jets transverse momenta (H^j_{T}); but together the 2D extended version of the BUMPHUNTER algorithm finds a 2D area where the significance is much higher, resulting in a more powerful separation.

Similarly, for the model with the associated production of a $t\bar{t}$ pair together with a scalar resonance y_0 decaying to a pair of invisible DM candidate particles, the expected strong variables are those including the missing transverse energy, either *per se*, or in a sum with the H^j_{T} variable. But in this case, an interesting effects is encountered when the two variables are paired to a 2D variable: the signal appears to populate an area along but visibly away of the diagonal, well separated from the $t\bar{t}$ and other backgrounds. This allows the 2D version of the BUMPHUNTER algorithm to identify this area as the most sensitive one to observe the signal, or, as the algorithm works, to declare the largest incompatibility with the background-only hypothesis, of much stronger sensitivity then the individual 1D variables, although they are highly correlated.

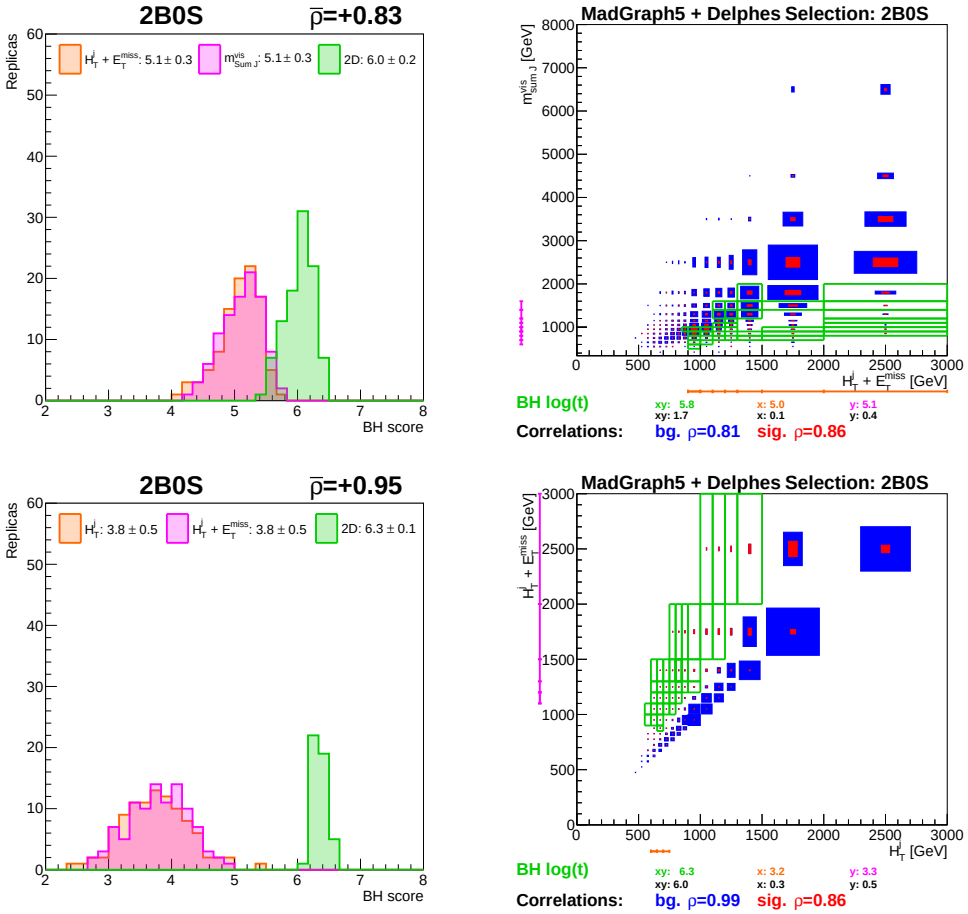


FIGURE 8.10: Left: a histogram of 1D (orange, purple) and 2D (green) BUMPHUNTER scores over pseudo-experiments for the visible mass of the sum of large jets and $H_{\mathrm{T}}^{j} + E_{\mathrm{T}}^{\mathrm{miss}}$ variables using the y_0 signal model (top) and for the H_{T}^{j} and $H_{\mathrm{T}}^{j} + E_{\mathrm{T}}^{\mathrm{miss}}$ variables using the $t\bar{t} + y_0$, $y_0 \to \chi_D \bar{\chi}_D$ (bottom) signal model. Right: The green boxes indicate the best 2D (xy) BUMPHUNTER area while the vertical purple and horizontal orange bars indicate the bin ranges of the best 1D BUMPHUNTER areas for variables on the y and x axes, respectively. The blue (red) boxes area is proportional to the number of total signal+background (signal) events. The 2D approach clearly identifies an area where a higher significance is found compared to the individual 1D variables. The colored xy, x and y labels indicate the BUMPHUNTER 1D and 2D results scores. The black xy, x and y labels indicate the simple and less powerful 1D and 2D χ^2 scores of compatibility to the background-only hypothesis. [Auth4]

8.2.6 Unfolding and BSM

As a last topic, we explore the following natural question to ask: what is the performance of unfolding in the case of presence of new physics beyond the Standard Model. Can presence of a new signal be inferred from an unfolded spectrum? Does unfolding bring new information or precision? In this Section we provide several studies related to these questions.

As outlined in Section 4.5.4, the necessary unfolding ingredients are the binned corrections, namely the acceptance and efficiency, examples of which are shown in Figure 8.11 for the case of the $t\bar{t}$ pair invariant mass in a private simulation using the DELPHES parameterized simulation and focusing on the hadronic $t\bar{t}$ final states as described in Section 8.2.

The key input is the distribution subtracted for backgrounds not coming from the $t\bar{t}$ processes, as shown in Figure 8.12.

Spectra which play important rôles in unfolding comparisons are shown in Figure 8.13, where the pseudo-data spectrum subtracted for the non-$t\bar{t}$ processes is shown. Also shown is the spectrum resulting from the application of the acceptance correction, and finally a particle-level spectrum multiplied by the finite selection efficiency. The latter two spectra are at the same footing as for the phase space given by the finite acceptance and efficiency, refer to Eq. 4.14. Still, they show a visible shape difference since the two spectra still refer to different levels, being the detector and particle, respectively.

This manifestly visualizes the need for unfolding which aims to correct for this shape difference from the detector to the particle level. Also shown is the response matrix (more precisely the normalized migration matrix) between the particle and detector bins where in each particle-level bin the fraction of events migrating to various detector bins is shown.

The unfolded FBU posteriors for the $m^{t\bar{t}}$ spectrum in each bin are shown in Figure 8.14, showing a good Gaussianity and compatibility (closure) to the particle-level spectrum (indicated by vertical read lines) within 1–2 standard deviations.

The comparison of the particle-level $m^{t\bar{t}}$ spectrum and the fully corrected pseudo-data spectrum is shown in Figure 8.15. There is a very good agreement between the particle-level spectrum (composed of both the SM $t\bar{t}$ a BSM y_0 scalar) and the fully unfolded pseudo-data of the same original detector-level composition. Also shown is the particle level spectrum of the $t\bar{t}$ only part, *i.e.* without the peaking BSM component, constituting a falling background-like spectrum. As will be shown in next, the visually sharper peak of the BSM contribution visible in the unfolded spectrum does not necessarily imply improved peak significance after unfolding.

The (expected) unfolding closure test showing a good agreement of the unfolded $t\bar{t}$+BSM spectrum as a ratio over the particle level version of the $t\bar{t}$+BSM spectrum is shown in Figure 8.16 (left). Similarly, the (expected) unfolding non-closure showing a disagreement of the unfolded $t\bar{t}$+BSM spectrum as a ratio over the particle level version of the $t\bar{t}$ spectrum only is shown in Figure 8.16 (right).

The question is whether the unfolding procedure can lead to a more precise spectrum determination *e.g.* in terms of a sharper BSM physics sample peak. To this, the key quantity is the magnitude of uncertainties after unfolding, while background uncertainties may also play an important rôle. This is illustrated in Figure 8.17 (left) with the detector-level significances including as well as excluding the non-$t\bar{t}$ physics background uncertainties are displayed as filled and line histograms, respectively. Excluding the background uncertainties, or knowing the background precisely, *i.e.* in the limit of an infinite statistics, can lead to several times higher

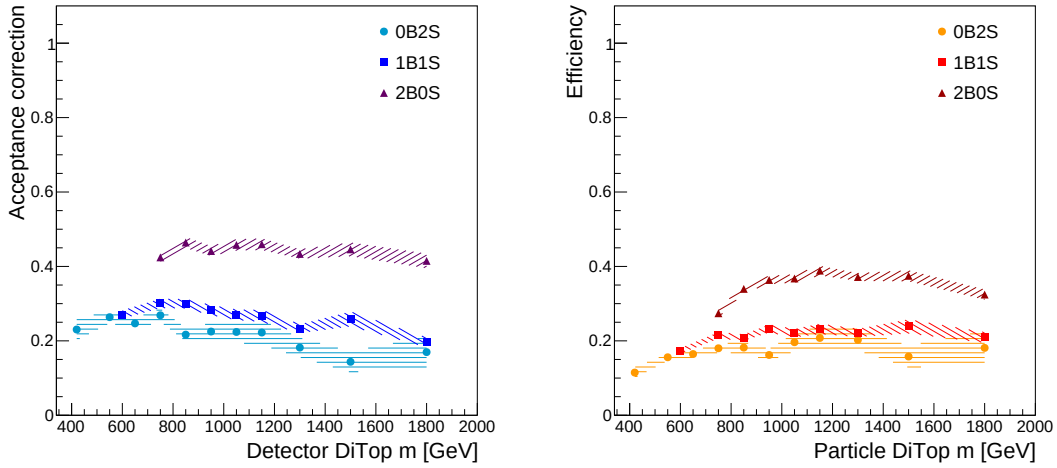


FIGURE 8.11: The acceptance (left) and efficiency (right) corrections and their statistical uncertainties indicated by hatcjed bands over variously boosted topologies as function of the invariant mass of the $t\bar{t}$ pair, derived on the simulated $t\bar{t}$ sample in model pp collisions at $\sqrt{S} = 14\,\text{TeV}$ using MADGRAPH5 and DELPHES packages.

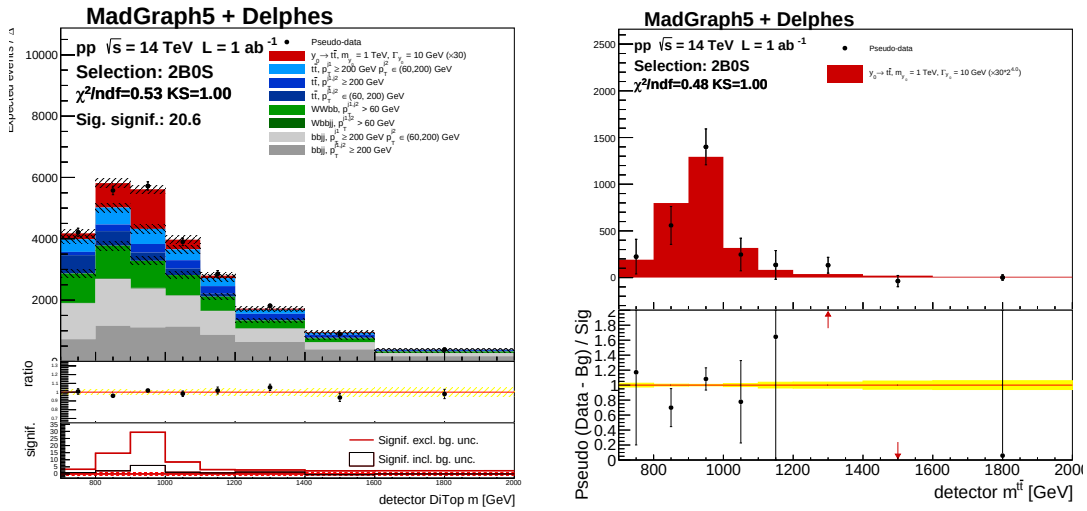


FIGURE 8.12: Left: Stacked invariant mass of the $t\bar{t}$ pair at the DELPHES detector level in the boosted-boosted topology for $y_0 \to t\bar{t}$ process with both the pseudo-data/prediction ratio and the signal significance shown in lower pads. Both significances without (filled) and without (line) the background uncertainties taken into account are shown. Right: the background-subtracted distribution, with the BSM signal left as the only sample, showing its shape as well as residuals in comparison to the statistically independent sample.

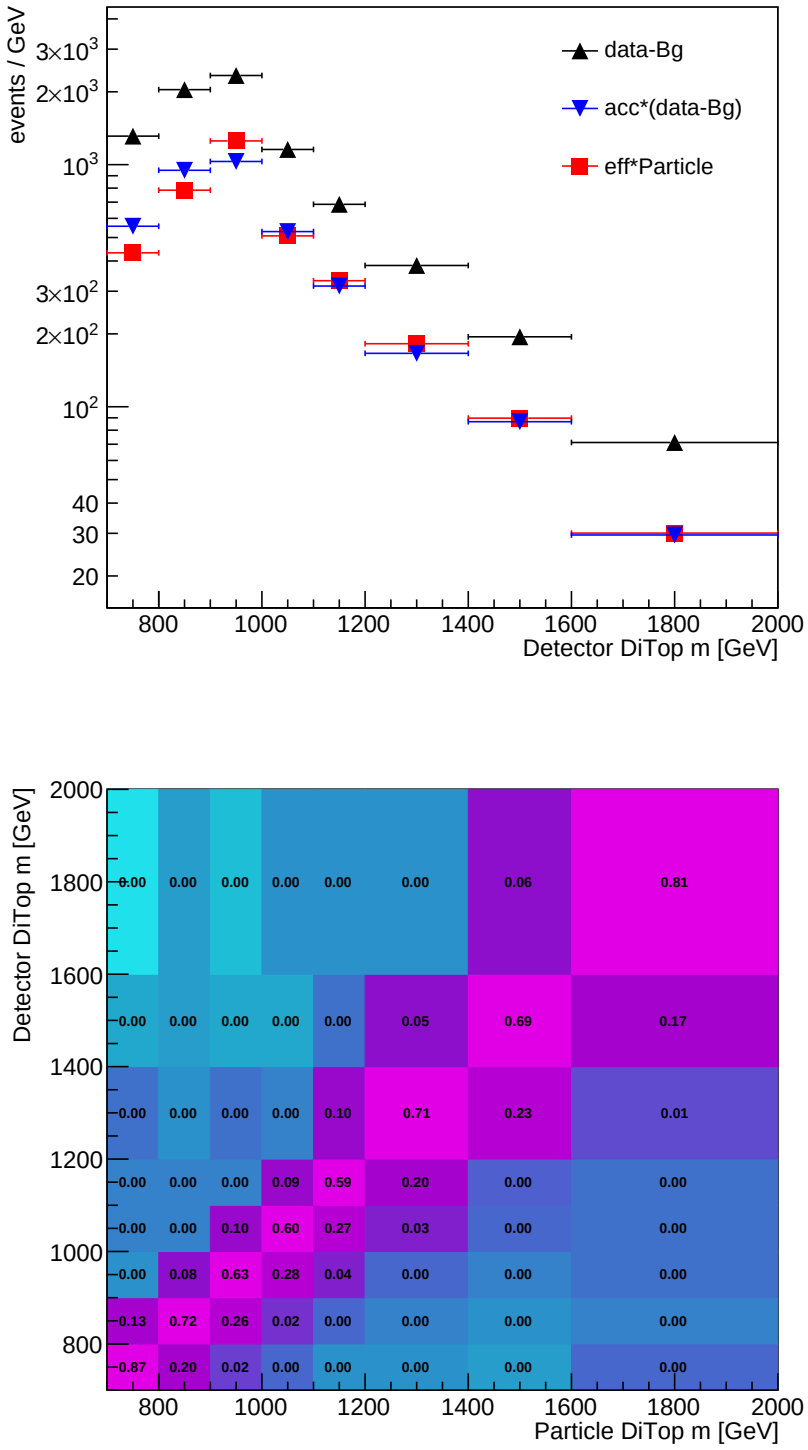


FIGURE 8.13: Top: the pseudo-data $m^{t\bar{t}}$ spectrum in the boosted-boosted topology subtracted for the non-$t\bar{t}$ processes (black), the spectrum resulting from subsequent application of the acceptance correction (blue) and the particle-level spectrum multiplied by the finite selection efficiency (red), showing different spectra shapes at the detector and particle levels in the same selection phase-space (blue and red). Bottom: the migration matrix showing the fraction of events in the detector-level bin originating in given particle-level bin.

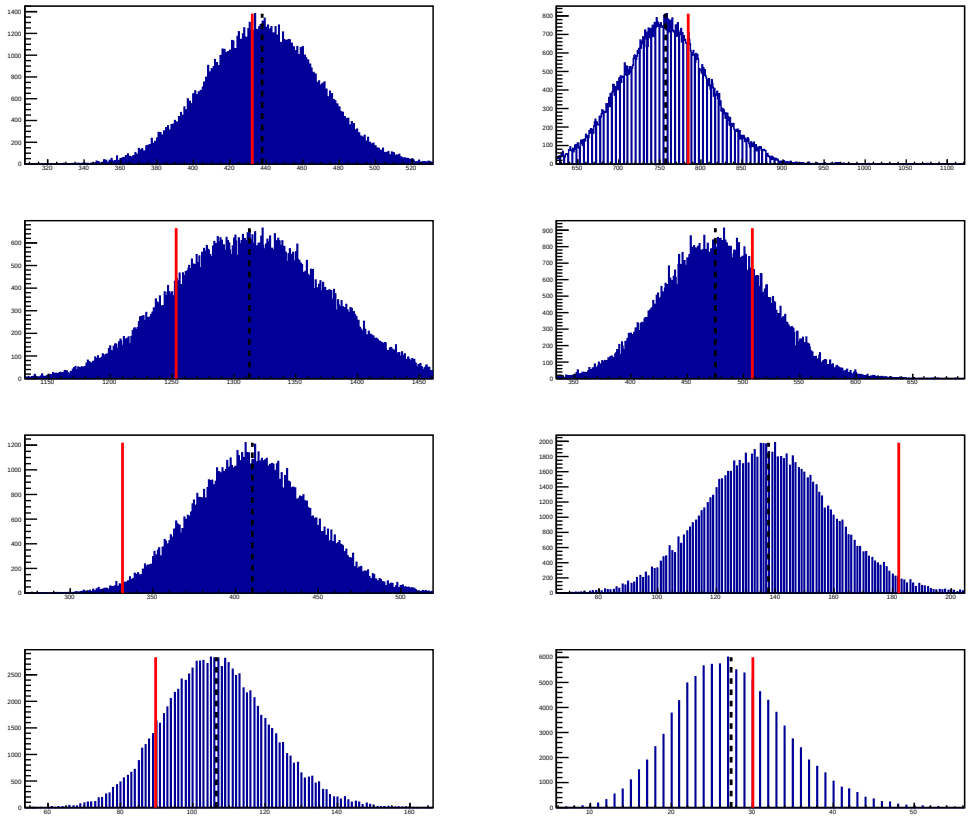


FIGURE 8.14: The unfolded FBU posteriors of the $m^{t\bar{t}}$ spectrum in each bin, with the vertical dashed black lines indicating the mean of the posterior and the solid red lines showing the number of events at the particle-level in each bin.

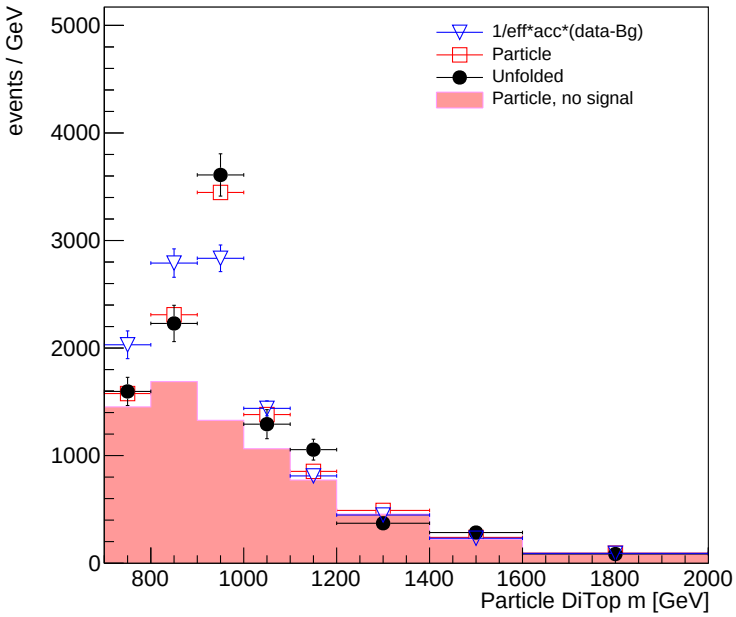


FIGURE 8.15: The unfolding closure test for the $m^{t\bar{t}}$ spectrum in the boosted-boosted regime showing a good agreement of the efficiency corrected unfolded $t\bar{t}$+BSM spectrum and the corresponding particle level one of the same sample composition. Also shown is the background-subtracted detector level pseudo-data spectrum, with both efficiency and acceptance corrections applied, indicating the different spectrum shape before the unfolding. Also shown is the particle level spectrum without the BSM signal (filled).

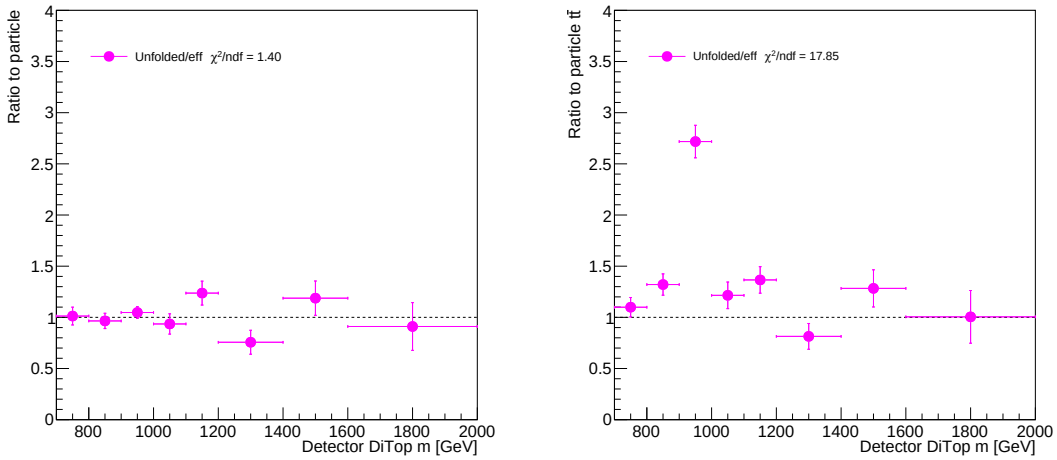


FIGURE 8.16: Left: The unfolding closure test showing a good agreement of the efficiency corrected unfolded $t\bar{t}$+BSM $m^{t\bar{t}}$ spectrum in the boosted-boosted regime as a ratio over the particle level one of the same sample composition. Right: The unfolding non-closure showing a disagreement of the unfolded $t\bar{t}$+BSM spectrum as a ratio over the particle level version of the $t\bar{t}$ spectrum only, showing the expected bump due to the presence of the BSM signal. Also shown are the goodness of fit values of the χ^2 over the number of bins.

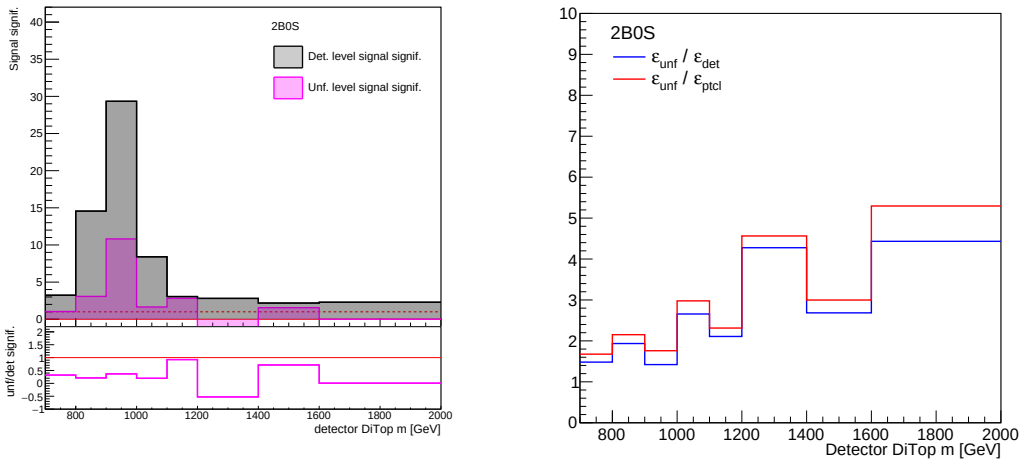


FIGURE 8.17: Left: The detector (gray) and unfolded (pink) BSM signal significances in the $m^{t\bar{t}}$ spectrum in the boosted-boosted regime. Right: the ratio of the relative uncertainties of the unfolded to the detector (blue) and particle (red) levels, demonstrating the increase of the statistical uncertainty after unfolding.

significance. But there are always systematics uncertainties related to backgrounds, too.

The FBU unfolded posterior exhibit larger variances then the detector-level spectrum due to uncertainties in the acceptance and efficiency corrections, and additional statistical uncertainties coming from the unfolding procedure itself. This is illustrated in Figure 8.17 (right) where the ratios of the uncertainties of unfolded spectrum to those of the original detector-level as well as the underlying particle-level distribution are shown, in both cases being around of two or higher and stemming from the unfolding procedure itself. A more detailed study was also presented in [Auth3] in the ℓ+jets channel. While statistical uncertainties in corrections could be remedied by simply generating more events, in practice large additional systematics uncertainties come from $t\bar{t}$ modelling systematics which cannot be omitted nor simply decreased. These are naturally present at both detector and particle levels.

On the other hand, correlations injected using unfolding techniques with regularization could lead to a narrowing of posteriors [Auth2], but the simple binned significance would not be a good measure of the peak significance. One would have to perform a statistical test over the full spectrum taking into account the bin correlations as described in Section 4.4.4.

In conclusion, the unfolding procedure is proved not to lead automatically to a higher significance of a hypothetical new physics signal over the SM $t\bar{t}$ production. This still motivates physics searches at the detector level using dedicated generated signals, while the unfolded spectrum is still the best precision tool for comparison to SM pQCD predictions, or to possibly indicate broader changes in observed spectra like a change of a slope or modified tails in distributions.

Chapter 9

Conclusions

From the Tevatron era, with pioneering measurements of top quark pairs cross-section, top quark mass, the observation of the single-top quark production, and the onset of differential cross sections, the analysis techniques have been sharpened and largely extended for the plethora of top quarks produced in the LHC experiments.

A fruitful interplay of theoretical and experimental efforts have lead to the understanding of the top quark p_{T} puzzle in past 10 years, with the current understanding being the missing NNLO effects.

Tireless Monte Carlo generator tuning efforts succeeded in constraining $t\bar{t}$ modelling for corresponding systematics for LHC analyses where top quarks are a background. Important inputs into the MC tuning have been the precision differential cross section measurements.

The detailed measurements of top quark kinematics in terms of multidimensional spectra offer the possibility not only to further constrain and challenge our understanding of the perturbative Quantum Chromodynamics at higher orders but also to see or constrain glimpses of new physics at the high energy frontier, with the physics of the heaviest fermion of the Standard Model to-day.

Top quarks production in non-standard environments like in proton-lead or lead-lead collisions, in forward regions of LHCb or in diffractive regimes, offer further prospects in understanding the dynamics of this special quark, with the possibility of a future lepton collider energy to be tuned to the process of an associated production of the Higgs boson and a top quark pair, in order to measure the top quark Yukawa coupling to the Higgs boson, and use large top-Higgs coupling as a source of the Higgs bosons in future Higgs factories.

In search for possible new physics in the top quark sector of the Standard Model, we conclude that besides the expected bumps in the $t\bar{t}$ mass spectrum, the presence of a BSM signal can possibly be observed also in other spectra. Importantly, a combination of 1D and 2D variables can reveal unexpected regions where signal significance can be enhanced.

Unfolding of the $t\bar{t}$ spectra with an admixture of artificial BSM signals still leads to a distinguishable peaks of new physics signals above the $t\bar{t}$ background but the significances of the peaks can be diluted by further uncertainties coming from the unfolding procedure, stemming from the finite precision of the unfolding corrections and background uncertainties, and the width of unfolding posteriors. The desired effect of sharper peaks after the unfolding correction for the detector resolution thus may not be always present. Still, unfolding multi-dimensional spectra of top quarks is a very useful tool for tuning Monte Carlo generators and for the era of precision top quark physics in the light of emerging NNLO generators.

Author's private contributions

[Auth1] Kvita J. Study of methods of resolved top quark reconstruction in semileptonic $t\bar{t}$ decay. *Nucl. Instrum. Meth. A*, 900:84–100, 2018. [Erratum: Nucl. Instrum. Meth. A: 1040, 167172 (2022)].

[Auth2] Baron P. and Kvita J. Extending the Fully Bayesian Unfolding with Regularization Using a Combined Sampling Method. *Symmetry*, 12:2100, 2020.

[Auth3] Pácalt J. and Kvita J. Study of semi-boosted top quark reconstruction performance on the line shape of a $t\bar{t}$ resonance. *Int. J. Mod. Phys. A*, 37(16):2250110, 2022.

[Auth4] Kvita J. Boosted and semi-boosted all-hadronic $t\bar{t}$ reconstruction performance on kinematic variables for selected BSM models using a 2D extension of the BumpHunter algorithm. *International Journal of Modern Physics A*, 39(11n12):2430002, 2024.

Author's selected ATLAS publications

[ATLAS1] Georges Aad et al. Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 73(1):2261, 2013.

[ATLAS2] Georges Aad et al. Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Phys. Rev. D*, 90(7):072004, 2014.

[ATLAS3] Georges Aad et al. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 76(10):538, 2016.

[ATLAS4] Georges Aad et al. Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions using the ATLAS detector. *Phys. Rev. D*, 93(3):032009, 2016.

[ATLAS5] M. Aaboud et al. Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *JHEP*, 11:191, 2017.

[ATLAS6] Morad Aaboud et al. Measurements of differential cross sections of top quark pair production in association with jets in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *JHEP*, 10:159, 2018.

[ATLAS7] Georges Aad et al. Measurements of top-quark pair differential and double-differential cross-sections in the ℓ+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 79(12):1028, 2019. [Erratum: Eur.Phys.J.C 80, 1092 (2020)].

[ATLAS8] Georges Aad et al. Observation of $t\bar{t}$ production in the lepton+jets and dilepton channels in p+Pb collisions at $\sqrt{s_{\mathrm{NN}}}$ = 8.16 TeV with the ATLAS detector. *JHEP*, 11:101, 2024.

Bibliography

[1] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[2] Jiří Hořejší. Fundamentals of electroweak theory. https://arxiv.org/abs/2210.0452, 2002.

[3] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[4] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[5] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[7] Michelangelo L Mangano. Introduction to QCD. 2000. CERN-OPEN-2000-255.

[8] Franz Gross et al. 50 Years of Quantum Chromodynamics. 12 2022.

[9] F. Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. 1984.

[10] Jiří Chýla. Quarks, partons and Quantum Chromodynamics. 2004.

[11] Peter Skands. Introduction to QCD. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pages 341–420, 2013.

[12] Alexandre Deur, Stanley J. Brodsky, and Guy F. de Teramond. The QCD Running Coupling. *Nucl. Phys.*, 90:1, 2016.

[13] Gavin P. Salam. Elements of QCD for hadron colliders. In *2009 European School of High-Energy Physics*, 11 2010.

[14] T. D. Lee and M. Nauenberg. Degenerate systems and mass singularities. *Phys. Rev.*, 133:B1549–B1562, Mar 1964.

[15] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[16] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298–318, 1977.

[17] Gunther Piller and Wolfram Weise. Nuclear deep inelastic lepton scattering and coherence phenomena. *Phys. Rept.*, 330:1–94, 2000.

[18] M. Klasen and H. Paukkunen. Nuclear PDFs After the First Decade of LHC Data. 11 2023.

[19] Vadim Guzey, Lingyan Zhu, Cynthia E. Keppel, M. Eric Christy, Dave Gaskell, Patricia Solvignon, and Alberto Accardi. Impact of nuclear dependence of $R = \sigma_L/\sigma_T$ on antishadowing in nuclear structure functions. *Phys. Rev. C*, 86:045201, 2012.

[20] Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado. EPPS16: Nuclear parton distributions with LHC data. *Eur. Phys. J. C*, 77(3):163, 2017.

[21] Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado. EPPS21: a global QCD analysis of nuclear PDFs. *Eur. Phys. J. C*, 82(5):413, 2022.

[22] Sabine Hossenfelder. Screams for explanation: finetuning and naturalness in the foundations of physics. *Synthese*, 198(Suppl 16):3727–3745, 2021.

[23] Anson Hook. TASI Lectures on the Strong CP Problem and Axions. *PoS*, TASI2018:004, 2019.

[24] Rachel Bean. TASI Lectures on Cosmic Acceleration. https://arxiv.org/abs/1003.4468, 2010.

[25] James M. Cline. TASI Lectures on Early Universe Cosmology: Inflation, Baryogenesis and Dark Matter. *PoS*, TASI2018:001, 2019.

[26] Thorsten Ohl. Drawing Feynman diagrams with Latex and Metafont. *Comput. Phys. Commun.*, 90:340–354, 1995. http://osksn2.hep.sci.osaka-u.ac.jp/~taku/osx/feynmp.html.

[27] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and et al. Parton distributions with lhc data. *Nuclear Physics B*, 867(2):244–289, Feb 2013.

[28] Georges Aad et al. Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 80(11):1085, 2020.

[29] Georges Aad et al. Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *JHEP*, 11:118, 2021.

[30] Kirill Melnikov, Andreas Scharf, and Markus Schulze. Top quark pair production in association with a jet: Qcd corrections and jet radiation in top quark decays. *Physical Review D*, 85(5), Mar 2012.

[31] LHC Top Working Group. Lhctopwg summary plots. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots, 2021.

[32] Benjamin Fuks, Kaoru Hagiwara, Kai Ma, and Ya-Juan Zheng. Signatures of toponium formation in lhc run 2 data. *Physical Review D*, 104(3), Aug 2021.

[33] Georges Aad et al. Observation of spin correlation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Phys. Rev. Lett.*, 108:212001, 2012.

[34] Georges Aad et al. Observation of quantum entanglement with top quarks at the ATLAS detector. *Nature*, 633(8030):542–547, 2024.

[35] Georges Aad et al. Test of the universality of τ and μ lepton couplings in W-boson decays with the ATLAS detector. *Nature Phys.*, 17(7):813–818, 2021.

[36] Georges Aad et al. Precise test of lepton flavour universality in W-boson decays into muons and electrons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 84(10):993, 2024.

[37] S. Alekhin, A. Djouadi, and S. Moch. The top quark and Higgs boson masses and the stability of the electroweak vacuum. *Phys. Lett. B*, 716:214–219, 2012.

[38] María Moreno Llácer. Search for CP violation in single top quark events with the ATLAS detector at LHC. https://cds.cern.ch/record/1735252, 2014. Presented 2014.

[39] John Ellis, Dae Sung Hwang, Kazuki Sakurai, and Michihisa Takeuchi. Disentangling Higgs-Top Couplings in Associated Production. *JHEP*, 04:004, 2014.

[40] Search for flavour-changing neutral current interactions of the top quark and the Higgs boson in events with a pair of τ-leptons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 8 2022.

[41] Search for flavour-changing neutral-current couplings between the top quark and the photon with the ATLAS detector at $\sqrt{s} = 13$ TeV. 5 2022.

[42] Cen Zhang and Scott Willenbrock. Effective-Field-Theory Approach to Top-Quark Production and Decay. *Phys. Rev. D*, 83:034006, 2011.

[43] Stefano Frixione, Eric Laenen, Patrick Motylinski, Bryan R. Webber, and Chris D. White. Single-top hadroproduction in association with a W boson. *JHEP*, 07:029, 2008.

[44] F. Abe, H. Akimoto, A. Akopian, M. G. Albrow, S. R. Amendolia, D. Amidei, J. Antos, C. Anway-Wiese, S. Aota, G. Apollinari, and et al. Observation of top quark production in $p\bar{p}$ collisions with the collider detector at fermilab. *Physical Review Letters*, 74(14):2626–2631, Apr 1995.

[45] S. Abachi, B. Abbott, M. Abolins, B. S. Acharya, I. Adam, D. L. Adams, M. Adams, S. Ahn, H. Aihara, J. Alitti, and et al. Observation of the top quark. *Physical Review Letters*, 74(14):2632–2637, Apr 1995.

[46] Frank E. Paige, Serban D. Protopopescu, Howard Baer, and Xerxes Tata. ISAJET 7.37: A Monte Carlo event generator for p p, anti-p p, and e+ e- reactions. 4 1998.

[47] S. Tkaczyk et. al. The cdf silicon vertex detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 342(1):240 – 250, 1994.

[48] V. M. Abazov et al. Observation of Single Top Quark Production. *Phys. Rev. Lett.*, 103:092001, 2009.

[49] Urs Achim Wiedemann. HIP and HEP. *PoS*, ICHEP2020:046, 2021.

[50] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[51] J. M. Campbell et al. Event Generators for High-Energy Physics Experiments. In *2022 Snowmass Summer Study*, 3 2022.

[52] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[53] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J. C*, 58:639–707, 2008.

[54] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.

[55] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[56] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[57] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[58] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[59] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[60] Enrico Bothmann et al. Event Generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[61] Peter Skands and Rob Verheyen. Multipole photon radiation in the Vincia parton shower. *Phys. Lett. B*, 811:135878, 2020.

[62] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002.

[63] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, Andreas Schalicke, and Steffen Schumann. Matching parton showers and matrix elements. In *HERA and the LHC: A Workshop on the Implications of HERA for LHC Physics: CERN - DESY Workshop 2004/2005 (Midterm Meeting, CERN, 11-13 October 2004; Final Meeting, DESY, 17-21 January 2005)*, pages 288–289, 2005.

[64] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.

[65] Stefan Hoeche, Frank Krauss, Marek Schonherr, and Frank Siegert. QCD matrix elements + parton showers: The NLO case. *JHEP*, 04:027, 2013.

[66] Pier Francesco Monni, Paolo Nason, Emanuele Re, Marius Wiesemann, and Giulia Zanderighi. MiNNLO$_{PS}$: a new method to match NNLO QCD to parton showers. *JHEP*, 05:143, 2020. [Erratum: JHEP 02, 031 (2022)].

[67] J. Allison et al. Recent developments in geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225, 2016.

[68] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[69] J. et al. Allison. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278, 2006.

[70] A Ferrari, Paola R Sala, A Fassò, and Johannes Ranft. *FLUKA: A multi-particle transport code (program version 2005)*. CERN Yellow Reports: Monographs. CERN, Geneva, 2005.

[71] T.T. Böhlen et al. The FLUKA Code: Developments and Challenges for High Energy and Medical Applications. *Nuclear Data Sheets*, 120:211–214, 2014.

[72] G. Aad et al. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010.

[73] Georges Aad et al. AtlFast3: the next generation of fast simulation in ATLAS. *Comput. Softw. Big Sci.*, 6:7, 2022.

[74] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.

[75] Andy Buckley, Philip Ilten, Dmitri Konstantinov, Leif Lönnblad, James Monk, Witold Pokorski, Tomasz Przedzinski, and Andrii Verbytskyi. The HepMC3 event record library for Monte Carlo event generators. *Comput. Phys. Commun.*, 260:107310, 2021.

[76] Morad Aaboud et al. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV. *Eur. Phys. J. C*, 79(8):639, 2019.

[77] Georges Aad et al. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV. *Eur. Phys. J. C*, 76(5):292, 2016.

[78] M. Aaboud et al. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96(7):072002, 2017.

[79] Georges Aad et al. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019.

[80] Joao Pequenao and Paul Schaffner. How ATLAS detects particles: diagram of particle paths in the detector. 2013.

[81] Louis Lyons, Duncan Gibaut, and Peter Clifford. How to combine correlated estimates of a single physical quantity. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 270(1):110–117, 1988.

[82] Luca Lista. Combination of measurements and the BLUE method. *EPJ Web Conf.*, 137:11006, 2017.

[83] Gerhard Bohm and Günter Zech. *Introduction to Statistics and Data Analysis for Physicists; 3rd revised*. Verlag Deutsches Elektronen-Synchrotron, Hamburg, 2017.

[84] Georgios Choudalakis. On hypothesis testing, trials factor, hypertests and the BumpHunter. In *PHYSTAT 2011*, 1 2011.

[85] Johannes Erdmann, Stefan Guindon, Kevin Kroeninger, Boris Lemmer, Olaf Nackenhorst, Arnulf Quadt, and Philipp Stolte. A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework. *Nucl. Instrum. Meth. A*, 748:18–25, 2014.

[86] Albert M Sirunyan et al. Measurement of the top quark mass with lepton+jets final states using p p collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 78(11):891, 2018. [Erratum: Eur.Phys.J.C 82, 323 (2022)].

[87] V. M. Abazov et al. Measurement of the Top Quark Mass in the Lepton + Jets Channel using the Ideogram Method. *Phys. Rev. D*, 75:092001, 2007.

[88] Georges Aad et al. Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS detector. *JHEP*, 06:100, 2015.

[89] G. Cowan. *Statistical Data Analysis*. Oxford science publications. Clarendon Press, 1998.

[90] Andreas Höcker and Vakhtang Kartvelishvili. Svd approach to data unfolding. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 372(3):469–481, Apr 1996.

[91] G. D'Agostini. A multidimensional unfolding method based on bayes' theorem. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 362(2):487–498, 1995.

[92] G. D'Agostini. Improved iterative bayesian unfolding, 2010.

[93] Bogdan Malaescu. An Iterative, dynamically stabilized method of data unfolding. 7 2009.

[94] Riccardo Di Sipio. Eikos: a Bayesian unfolding method for differential cross-section measurements. 8 2018.

[95] Georgios Choudalakis. Fully Bayesian Unfolding. https://arxiv.org/abs/1201.4612v4, 2012.

[96] Christian Robert and George Casella. A short history of markov chain monte carlo: Subjective recollections from incomplete data. *Statistical Science*, 26(1), feb 2011.

[97] Anders Andreassen, Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. OmniFold: A Method to Simultaneously Unfold All Observables. *Phys. Rev. Lett.*, 124(18):182001, 2020.

[98] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[99] Timo Antero Aaltonen et al. Combination of Measurements of the Top-Quark Pair Production Cross Section from the Tevatron Collider. *Phys. Rev. D*, 89(7):072001, 2014.

[100] V. M. Abazov et al. Precision measurement of the ratio $\mathrm{B}(t \to Wb)/\mathrm{B}(t \to Wq)$ and Extraction of V_{tb}. *Phys. Rev. Lett.*, 107:121802, 2011.

[101] A. P. Heinson. Top Quark Physics at the DZero Experiment, 1996.

[102] B. et al. Abbott. Direct measurement of the top quark mass by the DØ Collaboration. *Phys. Rev. D*, 58, 07 1998.

[103] V. M. Abazov et al. Dependence of the $t\bar{t}$ Production Cross Section on the Transverse Momentum of the Top Quark. *Phys. Lett. B*, 693:515–521, 2010.

[104] Scott Stuart Snyder. Measurement of the Top Quark Mass at D0. 5 1995.

[105] Victor Mukhamedovich Abazov et al. Measurement of differential $t\bar{t}$ production cross sections in $p\bar{p}$ collisions. *Phys. Rev. D*, 90(9):092006, 2014.

[106] Valentin Ahrens, Andrea Ferroglia, Matthias Neubert, Ben D. Pecjak, and Li Lin Yang. Renormalization-Group Improved Predictions for Top-Quark Pair Production at Hadron Colliders. *JHEP*, 09:097, 2010.

[107] Nikolaos Kidonakis. NNNLO soft-gluon corrections for the top-quark p_T and rapidity distributions. *Phys. Rev. D*, 91(3):031501, 2015.

[108] V. M. et al. Abazov. Forward-backward asymmetry in top quark-antiquark production. *Physical Review D*, 84(11), Dec 2011.

[109] Roman Lysák. Charge asymmetry in top quark pair production. *Symmetry*, 12(8), 2020.

[110] J. Kvita. *Measurement of Differential Cross-Sections in the ttbar -> l+jets Channel*. PhD thesis, Charles U., 2009.

[111] Apanasevich et. al. Evidence for parton k_T effects in high- p_T particle production. *Phys. Rev. Lett.*, 81:2642–2645, Sep 1998.

[112] Studies on top-quark Monte Carlo modelling for Top2016. Technical report, CERN, Geneva, 2016. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-020.

[113] Vardan Khachatryan et al. First Measurement of the Cross Section for Top-Quark Pair Production in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 695:424–443, 2011.

[114] Georges Aad et al. Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 71:1577, 2011.

[115] Roel Aaij et al. First observation of top quark production in the forward region. *Phys. Rev. Lett.*, 115(11):112001, 2015.

[116] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011.

[117] Comparison of Monte Carlo generator predictions to ATLAS measurements of top pair production at 7 TeV. Technical report, CERN, Geneva, 2015. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-002.

[118] Georges Aad et al. Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS detector. *JHEP*, 06:100, 2015.

[119] Marco Guzzi, Katerina Lipka, and Sven-Olaf Moch. Top-quark pair production at hadron colliders: differential cross section and phenomenological applications with DiffTop. *JHEP*, 01:082, 2015.

[120] Michal Czakon, David Heymes, and Alexander Mitov. Dynamical scales for multi-TeV top-pair production at the LHC. *JHEP*, 04:071, 2017.

[121] Michal Czakon, David Heymes, and Alexander Mitov. High-precision differential predictions for top-quark pairs at the LHC. *Phys. Rev. Lett.*, 116(8):082003, 2016.

[122] Stefano Catani, Simone Devoto, Massimiliano Grazzini, Stefan Kallweit, and Javier Mazzitelli. Top-quark pair production at the LHC: Fully differential QCD predictions at NNLO. *JHEP*, 07:100, 2019.

[123] J. H. Kühn, A. Scharf, and P. Uwer. Weak Interactions in Top-Quark Pair Production at Hadron Colliders: An Update. *Phys. Rev. D*, 91(1):014020, 2015.

[124] Albert M Sirunyan et al. Observation of top quark production in proton-nucleus collisions. *Phys. Rev. Lett.*, 119(24):242001, 2017.

[125] Albert M Sirunyan et al. Evidence for Top Quark Production in Nucleus-Nucleus Collisions. *Phys. Rev. Lett.*, 125(22):222001, 2020.

[126] ATLAS, CMS Collaborations. Combination of inclusive top-quark pair production cross-section measurements using ATLAS and CMS data at $\sqrt{s}$ = 7 and 8 TeV. *JHEP*, 07:213, 2023.

[127] Georges Aad et al. Observation of top-quark pair production in lead-lead collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV with the ATLAS detector. 11 2024.

[128] Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 5.02$ TeV with the ATLAS detector. 7 2022.

[129] Armen Tumasyan et al. Measurement of the inclusive $\mathrm{t\bar{t}}$ production cross section in proton-proton collisions at $\sqrt{s}$ = 5.02 TeV. *JHEP*, 04:144, 2022.

[130] Measurement of the $t\bar{t}$ cross-section and $t\bar{t}/Z$ cross-section ratio using LHC Run 3 pp collision data at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV. Technical report, CERN, Geneva, 2022. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-070.

[131] First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13.6$ TeV. Technical report, CERN, Geneva, 2022.

[132] Abdelhak Djouadi, John Ellis, Andrey Popov, and Jérémie Quevillon. Interference effects in $t\bar{t}$ production at the LHC as a window on new physics. *JHEP*, 03:119, 2019.

[133] Morad Aaboud et al. Search for heavy particles decaying into a top-quark pair in the fully hadronic final state in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Rev.*, D99(9):092004, 2019.

[134] Morad Aaboud et al. Search for heavy particles decaying into top-quark pairs using lepton-plus-jets events in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J.*, C78(7):565, 2018.

[135] Serguei Chatrchyan et al. Search for Resonant $t\bar{t}$ Production in Lepton+Jets Events in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 12:015, 2012.

[136] Albert M Sirunyan et al. Search for heavy resonances decaying to a top quark and a bottom quark in the lepton+jets final state in proton–proton collisions at 13 TeV. *Phys. Lett. B*, 777:39–63, 2018.

[137] Albert M Sirunyan et al. Search for resonant $\mathrm{t\bar{t}}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:031, 2019.

[138] CMS Collaboration. Search for a heavy resonance decaying to a top quark and a w boson at $\sqrt{s} = 13$ TeV in the fully hadronic final state, 2021.

[139] Duhr C. FeynRules Implementation of Abelian Higgs Model. 2011. https://feynrules.irmp.ucl.ac.be/wiki/HiddenAbelianHiggsModel.

[140] Neil D. Christensen and Claude Duhr. FeynRules - Feynman rules made easy. *Comput. Phys. Commun.*, 180:1614–1641, 2009.

[141] James D. Wells. How to Find a Hidden World at the Large Hadron Collider. 2008.

[142] Olivier Mattelaer and Eleni Vryonidou. Dark matter production through loop-induced processes at the LHC: the s-channel mediator case. *Eur. Phys. J. C*, 75(9):436, 2015.

[143] Mihailo Backović, Michael Krämer, Fabio Maltoni, Antony Martini, Kentarou Mawatari, and Mathieu Pellen. Higher-order QCD predictions for dark matter production at the LHC in simplified models with s-channel mediators. *Eur. Phys. J. C*, 75(10):482, 2015.

[144] Matthias Neubert, Jian Wang, and Cen Zhang. Higher-Order QCD Predictions for Dark Matter Production in Mono-Z Searches at the LHC. *JHEP*, 02:082, 2016.

[145] Goutam Das, Celine Degrande, Valentin Hirschi, Fabio Maltoni, and Hua-Sheng Shao. NLO predictions for the production of a spin-two particle at the LHC. *Phys. Lett. B*, 770:507–513, 2017.

[146] Sabine Kraml, Ursula Laa, Kentarou Mawatari, and Kimiko Yamashita. Simplified dark matter models with a spin-2 mediator at the LHC. *Eur. Phys. J. C*, 77(5):326, 2017.

[147] Andreas Albert et al. Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for dark matter mediators in visible and invisible decay channels and calculations of the thermal relic density. *Phys. Dark Univ.*, 26:100377, 2019.

[148] Chiara Arina, Mihailo Backović, Jan Heisig, and Michele Lucente. Solar γ rays as a complementary probe of dark matter. *Phys. Rev. D*, 96(6):063010, 2017.

[149] Y. Afik, F. Maltoni, K. Mawatari, P. Pani, G. Polesello, Y. Rozen, and M. Zaro. DM+$b\bar{b}$ simulations with DMSimp: an update. In *Dark Matter at the LHC 2018: Experimental and theoretical workshop*, 11 2018.

[150] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.

[151] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011.