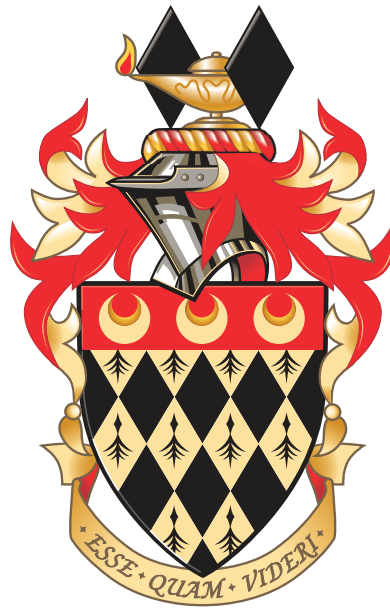# Measurements of Hadronic $t\bar{t}$ Differential Cross Sections with ATLAS and Unfolding with Gaussian Processes

Adam James Bozson

Department of Physics
Royal Holloway, University of London

A thesis submitted to the University of London
for the degree of Doctor of Philosophy

Declaration

I declare that the work presented in this thesis is my own. Where the work of others has been consulted, this is indicated in the text.

Adam James Bozson

September 2019

Abstract

Measurements of $t\bar{t}$ production cross sections at the Large Hadron Collider are presented. Proton–proton collision data with a centre-of-momentum energy of $13\,\mathrm{TeV}$ were collected using the ATLAS detector, totalling an integrated luminosity of $36.1\,\mathrm{fb}^{-1}$. The measurements are performed for the fully hadronic decay channel where the jets can be resolved. This allows all $t\bar{t}$ decay components to be measured along with additional QCD radiation.

The results are presented as absolute and normalised unfolded differential cross sections at particle level, as functions of several kinematic variables. Some cross sections are also unfolded to parton level, and some two-dimensional differential cross sections are presented. These distributions are compared to several Monte Carlo simulations.

Potential uses for Gaussian processes in particle physics are discussed, and a novel method for unfolding with Gaussian processes is introduced. The method is derived and assessed in terms of the unfolded estimators, statistical covariance matrices, and flexibility of the model.

Contents

Figures

Tables

For Paul

Chapter 1

Introduction

The Standard Model of particle physics has a rich history of accurate predictions of the phenomena occurring with high energy at subatomic scales. It describes particles and their interactions through a renormalisable quantum field theory. Despite its fruitful successes, the theory does not describe a range of observations, indicating the existence of a description of the universe beyond the Standard Model.

The *t quark* is the most massive constituent particle of the Standard Model. Studying it could elucidate any discrepancies between theoretical predictions and the data. At the Large Hadron Collider, *pp* collisions with energy $\sqrt{s} = 13\,\text{TeV}$ provide favourable conditions to perform measurements of $t\bar{t}$ production. Differential measurements of $t\bar{t}$ production cross sections can probe many facets of the multiple perturbative models available. In this thesis, a set of measurements of differential $t\bar{t}$ production cross sections in the fully resolved all-hadronic decay channel using the ATLAS detector is presented.

The statistical treatment of such measurements is also considered. In particular, the process of *unfolding* applies deconvolution methods to allow the measured distributions to be directly compared with theoretical predictions and other experimental observations. The measured differential $t\bar{t}$ production cross sections are unfolded and compared with computer simulations. Additionally, Gaussian processes may provide improvements to the traditional collection of unfolding techniques used in particle physics.

Thesis layout

An overview of the theoretical setting is given in Chapter 2. The Standard Model of particle physics is presented along with a discussion of phenomena it fails to describe. Motivation and previous measurements of differential $t\bar{t}$ production cross sections are given, and the case for measurements in the all-hadronic decay channel is made.

In Chapter 3, the experimental apparatus used to perform the measurements – the Large Hadron Collider and ATLAS – are described. A description of the methods used to reconstruct physics analysis objects from the data is given. An analysis of ATLAS data taken from pp collisions is described in Chapter 4 and the results are presented. The measured differential $t\bar{t}$ production cross sections are compared to Standard Model predictions from simulations.

An overview of Gaussian processes and the applications to particle physics is presented in Chapter 5. A novel method of unfolding with Gaussian processes is presented in Chapter 6.

Personal contribution

The ATLAS Collaboration consists of approximately 3000 members from 183 institutes in 38 countries [1]. Some of the work in this thesis is the result of this collaborative research effort and the design, construction, and running of the ATLAS experiment. Here my own individual contributions are outlined.

I worked on the software for the trigger configuration system described in Section 3.2.5. I also provided expertise on the trigger and data acquisition systems in the form of hardware installation, control room shift work, and on-call expertise during operation. For the data analysis in Chapter 4, I played a major role in the analysis team and wrote analysis software and documentation.

I worked equally in collaboration with others on introducing the techniques of Bayesian optimisation with Gaussian processes to the field, in particular for configuring simulation software. This particular technique is discussed in Section 5.4.5. For the method of unfolding

with Gaussian processes described in Chapter 6, I am the principal investigator and lead the research project.

Chapter 2

Theoretical foundations

The mechanics of natural phenomena at the smallest scale can be described with the mathematical framework of *quantum field theory*, of which a few principles are presented in Section 2.1. The Standard Model of particle physics, presented in Section 2.2, is a rigorous and successful quantum field theory with almost a century's history of construction, scrutiny, and experimental testing. A few symmetries lead to a rich structure of field content and interactions. Despite broad-ranging successes, there remain some theoretical failures and experimental discrepancies, some of which are presented and discussed in Section 2.2.5. Modern particle physics research is focused on precision tests of the theory and resolving these issues. Finally, the theoretical motivation behind measurements of $t\bar{t}$ production cross sections is presented in Section 2.3. This forms the foundation for the measurements presented in Chapter 4.

Natural units are used throughout, where the speed of light is $c = 1$ and the reduced Planck constant is $\hbar = h/2\pi = 1$. Therefore these factors are not written except where it is instructive to do so. Spacetime coordinates are $x = (t, \mathbf{x})^{\mathsf{T}} = (t, x^1, x^2, x^3)^{\mathsf{T}}$ with the Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$. The Einstein summation convention is used, whereby pairs of repeated indices in a product imply a sum over that index. For notational convenience, the differential operator is written $\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$.

2.1 Mechanics

In mechanics [2], every path through the configuration space of generalised coordinates for some system has an associated quantity called the *action*, given by the functional

$$S[\phi, \partial_\mu \phi] = \int \mathrm{d}^4 x \; \mathcal{L}[\phi, \partial_\mu \phi], \tag{2.1}$$

where $\mathcal{L}$ is the *Lagrangian density* for the system containing the fields $\phi(x) = \{\phi_1(x), \phi_2(x), \ldots\}$, and the integration is over a region in spacetime. The variational principle of stationary action says that the path taken in the classical limit is given by an extremum of S. Such a path obeys the Euler–Lagrange equations [3],

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_a} = 0, \tag{2.2}$$

for each field ϕ_a in ϕ. These give the equations of motion for a classical system.

In Feynman's path integral formulation of quantum mechanics [4, 5], the amplitude for a process is given by the sum over histories where each path has weight $e^{iS/\hbar}$. In the Heisenberg picture the equation of motion associated with the field operator $\phi_a(t, \mathbf{x}) = e^{iHt} \phi_a(0, \mathbf{x}) e^{-iHt}$ is [6]

$$i \, \partial_t \phi_a = [\phi_a, H], \tag{2.3}$$

where the term on the right is a commutator and the Hamiltonian is given by

$$H = \int \mathrm{d}^3 \mathbf{x} \left(\frac{\partial \mathcal{L}}{\partial(\partial_t \phi_a)} (\partial_t \phi_a) - \mathcal{L} \right). \tag{2.4}$$

A similar equation of motion follows for the conjugate momentum operator $\partial_\mu \phi_a$. The Hamiltonian is the generator of translations through time. Naturally, the classical result is recovered for $S \gg \hbar$ [7].

Lagrangian density

In both the classical and quantum mechanical cases, the dynamics of a system are described by the local Lagrangian density. Given this, the propagation of the system through spacetime can be determined by the application of the relevant equations of motion.

In ϕ^4 theory [8], a system containing one real scalar field ϕ has Lagrangian density

$$\mathcal{L} = -\frac{1}{2}\,\partial^\mu \phi\,\partial_\mu \phi \underbrace{-\frac{1}{2}m^2\phi^2 - \frac{1}{4!}\lambda\,\phi^4}_{-V(\phi)}. \tag{2.5}$$

The first term is referred to as the *kinetic term* and only contains spacetime derivatives of the field. The remainder is the negative of the potential $V(\phi)$, which is polynomial in ϕ. Its first term, $\frac{1}{2}m^2\phi^2$, is the mass term, where m is the mass of the field. This is the general form of the Lagrangian density for a renormalisable quantum field theory. Such theories are also subject to other constraints – namely certain symmetries – discussed in the rest of this chapter.

Noether's theorem

A general continuous transformation of the field ϕ_a can be expressed as multiple applications of smaller transformations of the form

$$\phi_a \to \phi_a + \epsilon\,\Delta\phi_a, \tag{2.6}$$

where ϵ is an infinitesimal scalar parameter and $\Delta\phi_a$ is some arbitrary change in the field configuration. Under one such transformation, the Lagrangian density becomes

$$\mathcal{L} \to \mathcal{L} + \frac{\partial\mathcal{L}}{\partial\phi_a}(\alpha\,\Delta\phi_a) + \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right)\partial_\mu(\alpha\,\Delta\phi_a) \tag{2.7}$$

$$= \mathcal{L} + \alpha\,\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\,\Delta\phi_a\right) + \alpha\left[\frac{\partial\mathcal{L}}{\partial\phi_a} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right)\right]\Delta\phi_a \tag{2.8}$$

(the sum over a is performed if there are multiple fields). From Equation 2.2, the expression in the square brackets is equal to zero for classical paths. If the equations of motion are

left unchanged, the transformation is considered a *symmetry* of the system. In this case, the Lagrangian density is invariant under the transformation, up to a divergence:

$$\mathcal{L} \to \mathcal{L} + \alpha\,\partial_\mu \mathcal{J}^\mu. \tag{2.9}$$

Comparing Equations 2.8 and 2.9, a conservation law applies to classical paths:

$$\partial_\mu j^\mu = 0 \qquad \text{for} \quad j^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)}\,\Delta\phi_a - \mathcal{J}^\mu. \tag{2.10}$$

This result is known at *Noether's theorem* [9] for classical fields, and a similar argument follows in quantum field theory [6]. It states that for every continuous symmetry of a system there is a conserved current $j^\mu(x)$. Moreover, Equation 2.10 implies local conservation of the charge j^0, since it follows that

$$\partial_t j^0 = -\nabla \cdot \mathbf{j}, \tag{2.11}$$

and therefore the global charge $Q = \int_{\text{all space}} \mathrm{d}^3\mathbf{x}\, j^0$ is constant in time. These charges will be used later in this chapter to categorise the fields of the Standard Model by their transformation properties under the actions of various symmetry group transformations.

2.1.1 Particles and interactions

Free field theory

When $\lambda = 0$, Equation 2.5 describes a free scalar field with Lagrangian density

$$\mathcal{L}_0 = -\frac{1}{2}\partial_\mu \phi\,\partial^\mu \phi - \frac{1}{2}m^2\phi^2. \tag{2.12}$$

Transition amplitudes can be calculated – for example via the LSZ reduction formula [8, 10] – from n-point time-ordered correlation functions. These are given by the vacuum expectation

values

$$\langle 0|\mathcal{T}\phi(x_1)\phi(x_2)\dots\phi(x_n)|0\rangle_0 = \frac{\displaystyle\int \mathcal{D}\phi\,\phi(x_1)\phi(x_2)\dots\phi(x_n)\,e^{i\int \mathrm{d}^4 x\,\mathcal{L}_0}}{\displaystyle\int \mathcal{D}\phi\,e^{i\int \mathrm{d}^4 x\,\mathcal{L}_0}}, \qquad (2.13)$$

where $\mathcal{T}$ denotes time ordering of the operators and $\mathcal{D}\phi \propto \prod_x \mathrm{d}\phi(x)$ is the functional measure up to a constant normalising factor. In this free field theory, the two-point correlation function can be determined exactly through Gaussian integration as [6]

$$\langle 0|\mathcal{T}\phi(x_1)\phi(x_2)|0\rangle_0 = \int \frac{\mathrm{d}^4 k}{(2\pi)^4}\frac{i}{k^2 - m^2 + i\epsilon}e^{-ik_\mu(x_1^\mu - x_2^\mu)} \equiv \Delta(x_1 - x_2), \qquad (2.14)$$

where ϵ parameterises a small imaginary offset introduced so that the integral converges. This is known as the *Feynman propagator* for a scalar particle.

Similarly, the four-point correlation function reduces to

$$\begin{aligned}
\langle 0|\mathcal{T}\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)|0\rangle_0 &= \Delta(x_1 - x_2)\Delta(x_3 - x_4) \\
&+ \Delta(x_1 - x_3)\Delta(x_2 - x_4) \\
&+ \Delta(x_1 - x_4)\Delta(x_2 - x_3),
\end{aligned} \qquad (2.15)$$

by integration of Gaussian terms. The vacuum expectation value of products of odd numbers of field operators is zero. The general result, known as *Wick's theorem* [11], can be written

$$\langle 0|\mathcal{T}\phi(x_1)\dots\phi(x_{2n})|0\rangle_0 = \sum_{P(1,\dots,2n)} \Delta(x_{P_1} - x_{P_2})\dots\Delta(x_{P_{2n-1}} - x_{P_{2n}}), \qquad (2.16)$$

where the sum is taken over all pairings of $\{1,\dots,2n\}$. This means that n-point correlation functions can be expressed as simple products of the propagator when the Lagrangian density is quadratic in $\phi(x)$ [12].

ϕ^4 interacting theory

For small values of λ, Equation 2.5 is a Lagrangian density for a free scalar field and a perturbative contribution from an interacting term,

$$\mathcal{L} = \mathcal{L}_0 - \frac{\lambda}{4!}\phi^4. \tag{2.17}$$

Then the phasor can be expanded as

$$e^{i\int \mathrm{d}^4 x \,\mathcal{L}} = e^{i\int \mathrm{d}^4 x \left(\mathcal{L}_0 - \frac{\lambda}{4!}\phi^4\right)} \tag{2.18}$$

$$= e^{i\int \mathrm{d}^4 x \,\mathcal{L}_0}\left(1 - i\int \mathrm{d}^4 x \,\frac{\lambda}{4!}\phi^4 + \dots\right). \tag{2.19}$$

Consequently, the vacuum expectation values for this interacting theory can be expressed solely in terms of the propagator of the free-field theory, by Wick's theorem. For example,

$$\langle 0|\mathcal{T}\phi(x_1)\phi(x_2)|0\rangle = \frac{\int \mathcal{D}\phi\,\phi(x_1)\phi(x_2)\dots\phi(x_n)\,e^{i\int \mathrm{d}^4 x \left(\mathcal{L}_0 - \frac{\lambda}{4!}\phi^4\right)}}{\int \mathcal{D}\phi\,e^{i\int \mathrm{d}^4 x \left(\mathcal{L}_0 - \frac{\lambda}{4!}\phi^4\right)}} \tag{2.20}$$

$$= \frac{\int \mathcal{D}\phi\,\phi(x_1)\phi(x_2)\dots\phi(x_n)\,e^{i\int \mathrm{d}^4 x\,\mathcal{L}_0}\left(1 - i\int \mathrm{d}^4 x\,\frac{\lambda}{4!}\phi^4 + \dots\right)}{\int \mathcal{D}\phi\,e^{i\int \mathrm{d}^4 x\,\mathcal{L}_0}\left(1 - i\int \mathrm{d}^4 x\,\frac{\lambda}{4!}\phi^4 + \dots\right)} \tag{2.21}$$

$$= \frac{1}{Z[0]}\left(-i\frac{\delta}{\delta J(x_1)}\right)\left(-i\frac{\delta}{\delta J(x_2)}\right)Z[J]\bigg|_{J=0}, \tag{2.22}$$

where $\frac{\delta}{\delta J(x_j)}$ denotes functional differentiation and the *generating functional* of the correlation functions, with source term $J(x)\phi(x)$, is given by

$$Z[J] \equiv \int \mathcal{D}\phi \,\exp\left[i\int \mathrm{d}^4 x\left(\mathcal{L}_0 - \frac{\lambda}{4!}\phi^4 + J\phi\right)\right] \tag{2.23}$$

$$= \exp\left[-\frac{i\lambda}{4!}\int \mathrm{d}^4 x\left(\frac{\delta}{\delta J(x)}\right)^4\right]\int \mathcal{D}\phi\,e^{i\int \mathrm{d}^4 x\,(\mathcal{L}_0 + J\phi)} \tag{2.24}$$

$$= \exp\left[-\frac{i\lambda}{4!}\int \mathrm{d}^4 x\left(\frac{\delta}{\delta J(x)}\right)^4\right]Z_0[J]. \tag{2.25}$$

Now the evaluation of the correlation functions for the interacting theory can be carried out through functional differentiation of the free field partition function, given by Equation 2.25 with [6]

$$Z_0[J] = Z_0[0] \exp\left[-\frac{1}{2}\int \mathrm{d}^4x\,\mathrm{d}^4y\,J(x)\Delta(x-y)J(y)\right]. \tag{2.26}$$

Taylor expanding Equation 2.25 in powers of λ, $Z[J]$ is given by the perturbation series

$$Z[J] = Z_0[J] - \overbrace{\frac{i\lambda}{4!}\int \mathrm{d}^4x \left(\frac{\delta}{\delta J(x)}\right)^4\bigg|_{J=0} Z_0[J]}^{\text{Leading order}}$$

$$\underbrace{-\frac{1}{2}\left(\frac{\lambda}{4!}\right)^2\int \mathrm{d}^4x\,\mathrm{d}^4y \left(\frac{\delta}{\delta J(x)}\right)^4\left(\frac{\delta}{\delta J(y)}\right)^4\bigg|_{J=0} Z_0[J] + \dots}_{\text{Next-to-leading order}}. \tag{2.27}$$

Feynman rules

Scattering amplitudes can be calculated using the Lehmann–Symanzik–Zimmermann (LSZ) reduction formula [8, 10]. For an initial state containing n particles with momenta $\{k_1, \dots, k_n\}$ transitioning to the final state containing m particles with momenta $\{p_1, \dots, p_m\}$, the amplitude is given by

$$\langle p_1, \dots, p_m | S | k_1, \dots, k_n \rangle =$$
$$\prod_{j=1}^{n}\left\{\int \mathrm{d}^4x_j\, ie^{ik_j \cdot x_j}\left(\partial_{x_j}^2 + m^2\right)\right\} \times \prod_{l=1}^{m}\left\{\int \mathrm{d}^4y_l\, ie^{-ip_l \cdot y_l}\left(\partial_{y_l}^2 + m^2\right)\right\}$$
$$\times \langle 0 | \mathcal{T}\phi(x_1)\dots\phi(x_n)\phi(y_1)\dots\phi(y_m) | 0 \rangle, \tag{2.28}$$

where $k \cdot x \equiv k_\mu x^\mu$ and $\partial_{x_j}^2 = [\partial_{x_j}]_\mu [\partial_{x_j}]^\mu$ (no sum over j) with $[\partial_{x_j}]_\mu \equiv \frac{\partial}{\partial x_j^\mu}$. The *S-matrix* operator contains a trivial (non-interacting) part and an interacting part written $S = \mathbf{1} + iT$ [6]. This formula applies for unbound massive particles, and it allows for the scattering amplitude to be expressed in terms of correlation functions. For a theory with a perturbative interaction

potential, these correlation functions can be written as sums of multiples of the free-field propagator, as illustrated by Equations 2.22, 2.26, and 2.27 for the example of scalar ϕ^4 theory.

Taking $2 \to 2$ scattering as a simple example, the amplitude is then given by

$$
\begin{aligned}
\langle p_1, p_2 | S | k_1, k_2 \rangle = {}& \int \mathrm{d}^4 x_1 \, i e^{i k_1 \cdot x_1} \left(\partial_{x_1}^2 + m^2 \right) \int \mathrm{d}^4 x_2 \, i e^{i k_2 \cdot x_2} \left(\partial_{x_2}^2 + m^2 \right) \\
& \times \int \mathrm{d}^4 y_1 \, i e^{-i p_1 \cdot y_1} \left(\partial_{y_1}^2 + m^2 \right) \int \mathrm{d}^4 y_2 \, i e^{-p_2 \cdot y_2} \left(\partial_{y_2}^2 + m^2 \right) \\
& \times \langle 0 | \mathcal{T} \phi(x_1) \phi(x_2) \phi(y_1) \phi(y_2) | 0 \rangle,
\end{aligned}
\tag{2.29}
$$

where the four-point correlation function is, by Wick's theorem and Equations 2.26 and 2.27,

$$
\langle 0 | \mathcal{T} \phi(x_1) \phi(x_2) \phi(y_1) \phi(y_2) | 0 \rangle = \frac{(-i)^4}{Z[0]} \frac{\delta}{\delta J(x_1)} \frac{\delta}{\delta J(x_2)} \frac{\delta}{\delta J(y_1)} \frac{\delta}{\delta J(y_2)} Z[J] \bigg|_{J=0}
\tag{2.30}
$$

$$
= -i\lambda \frac{Z_0[0]}{Z[0]} \int \mathrm{d}^4 x \, \Delta(x_1 - x) \Delta(x_2 - x) \Delta(x_3 - x) \Delta(x_4 - x) + \dots.
\tag{2.31}
$$

The Klein–Gordon wave operator acts on the propagator from Equation 2.14 as

$$
i \left(\partial_x^2 + m^2 \right) \Delta(x - y) = i \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \left(\partial_x^2 + m^2 \right) \frac{i}{k^2 - m^2 + i\epsilon} e^{-i k_\mu (x^\mu - y^\mu)}
\tag{2.32}
$$

$$
= \int \frac{\mathrm{d}^4 k}{(2\pi)^4} e^{-i k_\mu (x^\mu - y^\mu)}
\tag{2.33}
$$

$$
= \delta^{(4)}(x - y),
\tag{2.34}
$$

where $\delta^{(4)}(x - y)$ is the Dirac delta in four dimensions and its Fourier transform [13] is used to obtain the last line. Applying this to the substitution of Equation 2.31 into Equation 2.29, the contribution of interactions to the scattering amplitude is given by

$$
\langle p_1, p_2 | iT | k_1, k_2 \rangle = -i\lambda \, (2\pi)^4 \, \delta^{(4)}(k_1 + k_2 - p_1 - p_2).
\tag{2.35}
$$

The terms that make up the correlation functions can be represented by graphs, *Feynman diagrams* [14], composed of edges and vertices. This allows for the calculation of many more complicated terms involved in the expansion of higher-order correlation functions. For each

such diagram, the *momentum-space Feynman rules* for a particular theory are used. Each element of the diagram has an associated multiplicative factor[1]:

$$\text{Vertex} \qquad \qquad = \quad -i\lambda; \qquad \qquad (2.36)$$

$$\text{Propagator} \quad \xrightarrow{k} \quad = \quad \frac{i}{k^2 - m^2 + i\epsilon}; \qquad (2.37)$$

$$\text{Loop} \qquad \qquad = \quad \int \frac{\mathrm{d}^4 k}{(2\pi)^4}. \qquad \qquad (2.38)$$

Every vertex conserves momentum – the sum of momenta flowing into it must be zero. Since there are often multiple ways to draw the same diagram, each is divided by its *symmetry factor*.

The action of functional differentiation of $Z[J]$ in this framework is to remove a source J and fix the propagator to an external point; it becomes a *leg* and the diagram is now *connected*. The partition function, $Z[0]$, contains only unconnected vacuum 'bubble' diagrams, and the effect of the denominator in Equations 2.13 and 2.31 is to remove all such diagrams from the calculation of the scattering amplitude. Similarly, loop corrections to any propagators in diagrams do not contribute to the interacting part of the S-matrix operator. Therefore, the contribution due to interactions is given by the sum of all fully-connected amputated Feynman diagrams for the process under consideration [6]. Since the example term considered here is the only such diagram to first order in λ in this theory, the operator on the left-hand side of Equation 2.35 is written as iT. The unwritten terms in the expansion of the 4-point correlation function in Equation 2.31 are for diagrams which contribute to the non-interacting part of the S-matrix.

---

[1]Feynman diagrams in this thesis were drawn with *TikZ-Feynman* [15].

Cross sections

The Feynman rules can be used to calculate the cross section for a process. For $2 \to 2$ particle scattering, the differential cross section in the centre-of-momentum frame is given by [6]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{|\mathcal{M}|^2}{64\pi^2 E_{\mathrm{CM}}^2}, \tag{2.39}$$

where $E_{\mathrm{CM}} = 2\sqrt{k^2 + m^2}$ is the collision energy. The *matrix element* $\mathcal{M}$ is defined by

$$\langle p_1, \ldots, p_m | iT | k_1, \ldots, k_n \rangle = i\mathcal{M} (2\pi)^4 \delta^{(4)}(k_1 + \ldots + k_n - p_1 - \ldots - p_m), \tag{2.40}$$

and it is the sum of all fully-connected amputated diagrams. There is only one first-order interacting diagram in scalar ϕ^4 theory for $2 \to 2$ scattering, so the matrix element is given by

$$i\mathcal{M} = \quad \begin{matrix} p_1 \\ k_1 \quad \times \quad p_2 \\ k_2 \end{matrix} \quad = -i\lambda. \tag{2.41}$$

Substituting this into Equation 2.40 gives the same result obtained in Equation 2.35, but from application of the Feynman rules. The differential cross section is therefore

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{\lambda^2}{64\pi^2 E_{\mathrm{CM}}^2}. \tag{2.42}$$

The matrix element has no angular dependence in this case. The total cross section is obtained by integrating over full solid angle (4π) and dividing by two [6], since the two final-state particles are indistinguishable:

$$\sigma_{\mathrm{total}} = \frac{\lambda^2}{32\pi E_{\mathrm{CM}}^2}. \tag{2.43}$$

If the scalar ϕ^4 theory detailed here provided an accurate description for subatomic interactions, a collider experiment would be able to measure the cross section and thus determine a value

for the parameter λ. At the time of writing, however, the best description is given by a richer, more complicated theory with many more parameters – the Standard Model.

2.2 The Standard Model

The Standard Model is a quantum field theory with multiple interacting fields and symmetries. There is a history of excellent agreement between the predictions of the Standard Model and experimental data from particle physics experiments. For example, a prediction of the dimensionless fine-structure constant $\alpha = e^2/4\pi\epsilon_0\hbar c$ – including $12\,672$ tenth-order Feynman diagrams [16] – agrees with the currently accepted measured value [17] within $0.25\,\text{ppb}$ (parts per billion), with theoretical and experimental uncertainties of $0.25\,\text{ppb}$ and $0.23\,\text{ppb}$, respectively.

The Lagrangian for this theory demonstrates invariance under spacetime transformations of the Poincaré group. This is the isometry group of Minkowski spacetime in 1+3 dimensions, so it corresponds to the symmetries of special relativity. Such transformations act on coordinates as

$$x^\mu \to \Lambda^\mu{}_\nu \, x^\nu + a^\mu. \tag{2.44}$$

The spacetime interval between events, $\Delta s^2 = \Delta t^2 - \Delta\mathbf{x}^\mathsf{T}\Delta\mathbf{x}$ is preserved under these transformations. The Poincaré group can be decomposed into translations (a^μ) and the subgroup of proper, orthochronous Lorentz transformations which leave the metric tensor invariant ($\Lambda^\mathsf{T}\eta\,\Lambda = \eta$), composed of three-dimensional rotations under $SO(3)$ and linear velocity boosts. Under these symmetries, the theory conserves 4-momentum.

The fields of the Standard Model have definite transformation properties under the Poincaré symmetry group, and they can be defined in terms of group representations. There is a correspondence for the Lorentz subgroup (the *spin double cover*), $SO^+(1,3) \cong SL(2,\mathbb{C}) \cong SU(2) \times SU(2)$ [18]. This permits a representation in terms of ordered pairs of half-integers, $(\mathbf{m}, \mathbf{n})$, corresponding to the fields' embeddings in the two $SU(2)$ subalgebras [19].

A scalar field has no spacetime coordinates and has the singlet representation $(\mathbf{0}, \mathbf{0})$. The Higgs field, φ, is the only scalar field in the Standard Model. The fundamental representation, $(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0})$ represents left-handed chiral bispinors, where $\frac{\mathbf{1}}{\mathbf{2}}$ is the fundamental representation of $SU(2)$ – two-component complex vectors. The Standard model contains two left-handed spinor fields, q and l. Related by Hermitian conjugation, right-handed chiral spinors have the representation $(\mathbf{0}, \frac{\mathbf{1}}{\mathbf{2}})$. There are three such fields in the Standard Model, e, u, and d. The final representation is $(\frac{\mathbf{1}}{\mathbf{2}}, \frac{\mathbf{1}}{\mathbf{2}})$, corresponding to four-component vector fields. The gauge vector fields B, W, and G have this form. Since the Lagrangian remains invariant under Lorentz transformations, it only contains combinations of fields which together transform as a singlet. For example, these can take the form of a left-handed spinor coupled to a right-handed antispinor, e.g. $\bar{q}\slashed{D}q$, or the contraction of vector fields, e.g. $B_{\mu\nu}B^{\mu\nu}$.

For Λ connected to the identity, the Lorentz transformations are continuous, i.e. they can be parametrised by an arbitrary value α. There also exist discrete transformations that are disconnected from the identity, for example $P = \mathrm{diag}(1, -1, -1, -1)$ and $T = \mathrm{diag}(-1, 1, 1, 1)$, known as the *parity* and *time reversal* operators, respectively. Additionally, the *charge conjugation* operator C replaces a particle with its antiparticle, e.g. $q \to \bar{q}$. Since the Standard Model differentiates between left- and right-handed chiral fields (based on experimental observations [20]), it is not symmetric under P and parity is not a conserved quantity. The combined CP is a near symmetry, but measurements show it is broken [21, 22] and the Standard Model therefore allows for *CP-violation*. However, CPT is an exact symmetry of the Standard Model [23–25] by construction. A direct consequence of CPT symmetry is the spin statistics theorem [26–28], which gives the (anti)commutation relations of the field operators. Particles with half-integer spin (i.e. those corresponding to fields that transform as spinors) obey Fermi–Dirac statistics, whereas those with integer spin (scalar and vector fields) obey Bose–Einstein statistics. Particles are therefore labelled as *fermions* (and obey the Pauli exclusion principle [29]) for half-integer spin and *bosons* for integer spin.

2.2.1 Gauge sector

In addition to the spacetime symmetry of the Poincaré group, the Standard Model has an internal symmetry under local transformations of the gauge product group,

$$\mathcal{G}_{\mathrm{SM}} = SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{2.45}$$

This means that spacetime-dependent transformations of the fields,

$$\phi_a \to M_a{}^b \, \phi_b \tag{2.46}$$

leave the Lagrangian unchanged for $M \in \mathcal{G}_{\mathrm{SM}}$. The internal gauge symmetry is realised through modification of the partial differentiation operator to the *gauge covariant derivative*, $\partial_\mu \to D_\mu$, which involves the gauge vector fields. This gives rise to interactions between the fields in the Standard Model, which depend on the fields' charges, $\{C, L, Y\}$, under each symmetry group.

The decomposition of $\mathcal{G}_{\mathrm{SM}}$ in Equation 2.45 is considered here as a combination of two frozen symmetries: $SU(3)$ of the *strong* interaction and $SU(2) \times U(1)$ of the *electroweak* interaction.

$SU(3)$

For invariance of the terms in the Lagrangian under transformations of the $SU(3)$ symmetry group, the gauge covariant derivative is

$$D_\mu = \partial_\mu - i g_S T_a G^a_\mu, \tag{2.47}$$

where g_S is the strong coupling constant and $T_a = \lambda_a/2$ are the eight generators of the fundamental (triplet) representation of $SU(3)$, with the 3×3 traceless Gell-Mann matrices λ_a [30]. Eight gauge vector fields G^a_μ are necessarily introduced, corresponding to massless *gluons*.

The non-abelian nature of $SU(3)$ leads to gluon self-couplings in the Standard Model. The Lie algebra is given by

$$[T^a, T^b] = if^{abc}T^c \tag{2.48}$$

with the totally antisymmetric structure constants f^{abc}. Defining the gluon field strength tensor as $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_S f^{abc} G^b_\mu G^c_\nu$, the contribution to the Lagrangian density due to these fields is given by

$$\mathcal{L}_G = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{2.49}$$

$$= -\frac{1}{4}(\partial_\mu G^a_\nu - \partial_\nu G^a_\mu)(\partial^\mu G^\nu_a - \partial^\nu G^\mu_a)$$

$$- g_S f^{abc}(\partial_\mu G^a_\nu)G^\mu_b G^\nu_c - \frac{1}{4} g_S^2 (f^{eab}G^a_\mu G^b_\nu)(f^{ecd}G^\mu_c G^\nu_d). \tag{2.50}$$
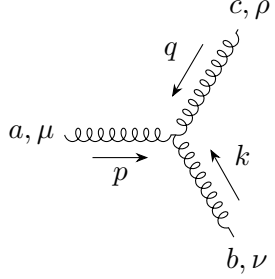
The first line of Equation 2.50 is the free-field (i.e. $g_S \to 0$) kinetic term. By analogy to Section 2.1.1, the free-field gluon propagator is given by [6]

$$\langle 0|\mathcal{T}G^a_\mu(x_1)G^b_\nu(x_2)|0\rangle_0 = \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{-i\delta^{ab}\eta_{\mu\nu}}{k^2 + i\epsilon} e^{-k_\rho(x_1^\rho - x_2^\rho)}. \tag{2.51}$$
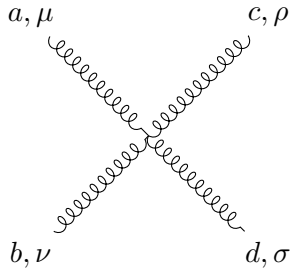
Therefore the corresponding momentum-space Feynman rule is

$$a, \mu \quad \overset{k}{\text{\textemdash}} \quad b, \nu \quad = \quad \frac{-i\delta^{ab}\eta_{\mu\nu}}{k^2 + i\epsilon}. \tag{2.52}$$

The last two terms of Equation 2.50 give the cubic and quartic couplings, respectively, with the momentum-space Feynman rules given by [6]

$$= g_S f^{abc} \Big[\eta^{\mu\nu}(k-p)^\rho + \eta^{\nu\rho}(p-q)^\mu + \eta^{\rho\mu}(q-k)^\nu \Big]; \tag{2.53}$$

$$= \begin{aligned} -ig_S^2 \Big[& f^{abe} f^{cde} (\eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\mu\sigma}\eta^{\nu\rho}) \\ & + f^{ace} f^{bde} (\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\sigma}\eta^{\nu\rho}) \\ & + f^{ade} f^{bce} (\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma}) \Big]. \end{aligned} \tag{2.54}$$

$SU(2) \times U(1)$

Under $SU(2) \times U(1)$ group transformations, the gauge covariant derivative is

$$D_\mu = \partial_\mu - ig\tau_a W_\mu^a - ig'\frac{Y}{2}B_\mu, \tag{2.55}$$

where $\tau_a = \sigma_a/2$ are the three generators of the fundamental representation of $SU(2)$ with the Pauli matrices σ_a [31] and scalar Y is the *weak hypercharge*. Four gauge vector fields are introduced, W_μ^1, W_μ^2, W_μ^3, and B_μ, with coupling constant g for the W_μ^a fields and g' for the B_μ field.

Again, the non-abelian nature of the $SU(2)$ symmetry group gives rise to interactions between the gauge fields. The Lie algebra is given by

$$[\tau^a, \tau^b] = i\epsilon^{abc}\tau^c, \tag{2.56}$$

where the structure constant ϵ^{abc} is the antisymmetric Levi-Civita tensor. U(1) is abelian, and therefore has no self-interaction. The contribution to the Lagrangian is

$$\mathcal{L}_{\text{EW}} = -\frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{2.57}$$

where $W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc}W^b_\mu W^c_\nu$ and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. $\mathcal{L}_{\text{EW}}$ contains a term

$$\mathcal{L}_{\text{EW cubic}} = -\frac{g}{2}\left(\partial_\mu W^a_\nu - \partial_\nu W^a_\mu\right)\epsilon^{abc}W^\mu_b W^\nu_c, \tag{2.58}$$

leading to cubic couplings between the W^a_μ fields. There is also a quartic term which gives four-point couplings [32].

Up to this point, the theory contains a description of four massless electroweak gauge fields. However, the Standard Model contains three massive bosons associated with the weak interaction along with the massless photon of the electromagnetic interaction. Mass terms are included in the potential of the Lagrangian density (c.f. Equation 2.5), and can take multiple forms. The most direct approach gives mass to the vector field W_μ via the term

$$\frac{1}{2}m^2\, W_\mu W^\mu. \tag{2.59}$$

While this contribution to the Lagrangian is invariant to Lorentz transformations, it is not symmetric under transformations of the $SU(2) \times U(1)$ electroweak gauge group. Instead the electroweak gauge fields acquire a mass, while the gauge invariance is preserved, through the *Higgs mechanism.*

Electroweak symmetry breaking

A scheme due to Anderson [33], Brout and Englert [34], Guralnik, Hagen and Kibble [35, 36], Higgs [37–39], and 't Hooft [40] introduces another field, φ, to the Standard Model Lagrangian. Through spontaneous symmetry breaking, the electroweak W^a_μ and B_μ fields mix and acquire masses, with the weak and electromagnetic interactions precipitating out [41–43].

The φ field can be represented as a complex $SU(2)$ doublet, scalar under Lorentz transformations:

$$\varphi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix}. \tag{2.60}$$

It is subject to a quartic potential of the form

$$V(\varphi) = -\mu^2 \varphi^\dagger \varphi + \lambda \left(\varphi^\dagger \varphi \right)^2, \tag{2.61}$$

with $\mu^2 > 0$ and $\lambda > 0$. The field has a non-zero vacuum expectation value, which may be freely chosen to be

$$\langle \varphi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.62}$$

with $v = \sqrt{\mu^2 / \lambda}$. This vacuum solution spontaneously breaks the now hidden $SU(2)$ symmetry but maintains invariance under $U(1)$ transformations, giving the symmetry breaking pattern $SU(2)_L \times U(1)_Y \to U(1)_Q$. The kinetic contribution to the Lagrangian from this field in the vacuum state is given by

$$\mathcal{L}_{K0} = (D^\mu \varphi)^\dagger (D_\mu \varphi) \big|_{\varphi = \langle \varphi \rangle} \tag{2.63}$$

$$= \left| \left(\partial_\mu - ig\tau^a W_\mu^a - \frac{1}{2} ig' B_\mu \right) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 \tag{2.64}$$

$$= \frac{v^2}{8} \left| \begin{pmatrix} gW_\mu^1 - igW_\mu^2 \\ -gW_\mu^3 + g'B^\mu \end{pmatrix} \right|^2 \tag{2.65}$$

$$= m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} m_Z^2 Z_\mu Z^\mu, \tag{2.66}$$

where in the last line $W^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right)$ and $Z_\mu = \cos\theta_w W_\mu^3 - \sin\theta_w B$ with $\cos\theta_w = g/\sqrt{g^2 + g'^2}$ and $\sin\theta_w = g'/\sqrt{g^2 + g'^2}$. Here the massless electroweak gauge fields have mixed and acquired masses $m_W = gv/2$ and $m_Z = m_W / \cos\theta_w$. The fourth mixed field, $A_\mu = \sin\theta_w W_\mu^3 + \cos\theta_w B_\mu$ is massless and is identified as the electromagnetic vector gauge field, with particle γ, of the surviving $U(1)$ symmetry.

Rewriting the covariant derivative given by Equation 2.55 in terms of these definite-mass fields gives

$$D_\mu = \partial_\mu - \frac{ig}{\sqrt{2}} \left(\tau^+ W_\mu^+ + \tau^- W_\mu^- \right) - \frac{ig}{\cos\theta_w} \left(\tau^3 - \sin\theta_w Q \right) Z_\mu - ieQA_\mu, \tag{2.67}$$

where $\tau^\pm = \tau^1 \pm i\tau^2$, $Q = \tau^3 + Y/2$, and $e = g\sin\theta_w$.

The quadratic contributions to $\mathcal{L}_{\text{EW}}$ in Equation 2.57 lead to the Feynman propagators for the electroweak gauge bosons [6], given (in the Feynman–'t Hooft gauge [44]) by

$$\mu \;\xrightarrow{k}\; \nu \;\;=\;\; \frac{-i\eta_{\mu\nu}}{k^2 - m_W^2 + i\epsilon}; \tag{2.68}$$

$$\mu \;\xrightarrow{k}\; \nu \;\;=\;\; \frac{-i\eta_{\mu\nu}}{k^2 - m_Z^2 + i\epsilon}; \tag{2.69}$$
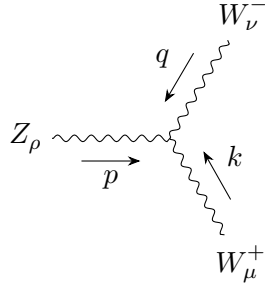
$$\mu \;\xrightarrow{k}\; \nu \;\;=\;\; \frac{-i\eta_{\mu\nu}}{k^2 + i\epsilon}. \tag{2.70}$$

The interaction between the gauge fields is found by rewriting Equation 2.58 as
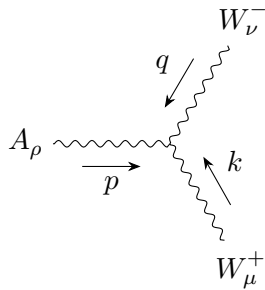
$$\begin{aligned}
\mathcal{L}_{\text{EW cubic}} = ig \Bigg\{ & \left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+ \right) W^{-\mu} \left(\cos\theta_w Z^\nu + \sin\theta_w A^\nu \right) \\
& - \left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^- \right) W^{+\mu} \left(\cos\theta_w Z^\nu + \sin\theta_w A^\nu \right) \\
& + \frac{1}{2} \left[\cos\theta_w \left(\partial_\mu - \partial_\nu \right) Z^\nu - \sin\theta_w \left(\partial_\mu - \partial_\nu \right) A^\nu \right] \left(W^{+\mu}W^{-\mu} - W^{-\mu}W^{+\nu} \right) \Bigg\}
\end{aligned} \tag{2.71}$$

$$\begin{aligned}
= \; & ig\cos\theta_w \left[\left(\partial_\mu W_\nu^+ - \partial_\nu W^{+\mu} \right) W^{-\mu}Z^\nu - \left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^- \right) W^{+\mu}Z^\nu \right. \\
& \left. \qquad\qquad + \frac{1}{2} \left(\partial_\mu Z^\nu - \partial_\nu Z^\nu \right) \left(W^{+\mu}W^{-\mu} - W^{-\mu}W^{+\nu} \right) \right] \\
& + ie \left[\left(\partial_\mu W_\nu^+ - \partial_\nu W^{+\mu} \right) W^{-\mu}A^\nu - \left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^- \right) W^{+\mu}A^\nu \right. \\
& \left. \qquad\qquad + \frac{1}{2} \left(\partial_\mu A^\nu - \partial_\nu A^\nu \right) \left(W^{+\mu}W^{-\mu} - W^{-\mu}W^{+\nu} \right) \right].
\end{aligned} \tag{2.72}$$

The first two lines of Equation 2.72 give the cubic W^+W^-Z coupling. The last two lines give the cubic $W^+W^-\gamma$ coupling. These lead to the momentum-space Feynman rules [6],
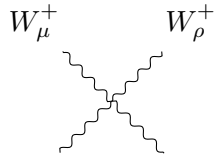
$$= ig\cos\theta_w\Big[\eta_{\mu\nu}(k-q)_\rho + \eta_{\nu\rho}(q-p)_\mu + \eta_{\rho\mu}(p-k)_\nu\Big];\qquad(2.73)$$
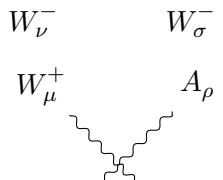
$$= ie\Big[\eta_{\mu\nu}(k-q)_\rho + \eta_{\nu\rho}(q-p)_\mu + \eta_{\rho\mu}(p-k)_\nu\Big].\qquad(2.74)$$
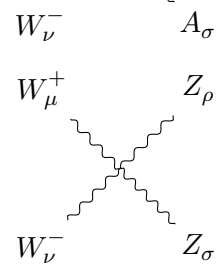
Quartic $W^+W^-W^+W^-$, $W^+W^-\gamma\gamma$, W^+W^-ZZ, and $W^+W^-\gamma Z$ couplings are also described by $\mathcal{L}_{\text{EW}}$ [32], with the Feynman rules
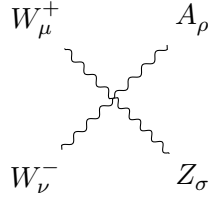
$$= ig^2\Big[2\eta_{\sigma\mu}\eta_{\rho\nu} - \eta_{\sigma\rho}\eta_{\mu\nu} - \eta_{\sigma\nu}\eta_{\rho\mu}\Big];\qquad(2.75)$$

$$= -ie^2\Big[2\eta_{\sigma\rho}\eta_{\mu\nu} - \eta_{\sigma\mu}\eta_{\rho\nu} - \eta_{\sigma\nu}\eta_{\rho\mu}\Big];\qquad(2.76)$$

$$= -ig^2\cos^2\theta_w\Big[2\eta_{\sigma\rho}\eta_{\mu\nu} - \eta_{\sigma\mu}\eta_{\rho\nu} - \eta_{\sigma\nu}\eta_{\rho\mu}\Big];\qquad(2.77)$$

$$= -ieg\cos\theta_w\left[2\eta_{\sigma\rho}\eta_{\mu\nu} - \eta_{\sigma\mu}\eta_{\rho\nu} - \eta_{\sigma\nu}\eta_{\rho\mu}\right]. \tag{2.78}$$

2.2.2 Higgs sector

Expanding around the vacuum expectation value, in the unitary gauge [45] the φ field takes the form

$$\varphi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.79}$$

The contribution to the Lagrangian from the field H, using the potential given by Equation 2.61 and the covariant derivative by Equation 2.67, is

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\varphi)^\dagger D^\mu\varphi - V(\varphi) \tag{2.80}$$

$$= \frac{1}{2}\partial_\mu H\,\partial^\mu H - \mu^2 H^2 + \left(m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu\right)\left(1 + \frac{H}{v}\right)^2 - \lambda v H^3 - \lambda H^4. \tag{2.81}$$

This describes a real scalar field with mass $m_H = \sqrt{2}\mu$, leading to the propagator

$$\xrightarrow{\quad k \quad}\quad\text{- - - - - - - -}\quad = \frac{i}{k^2 - m_H^2 + i\epsilon}. \tag{2.82}$$

The middle term in Equation 2.81 is equal to $\mathcal{L}_{K0}$ plus cubic and quartic interactions between the H field and the massive electroweak gauge fields. The associated Feynman rules are [6, 32]

$$= 2i\frac{m_W^2}{v}g_{\mu\nu}; \tag{2.83}$$

$$= 2i\frac{m_Z^2}{v}g_{\mu\nu}; \tag{2.84}$$

$$= 2i\frac{m_W^2}{v^2}g_{\mu\nu}; \tag{2.85}$$

$$= 2i\frac{m_Z^2}{v^2}g_{\mu\nu}. \tag{2.86}$$

The trailing two terms in Equation 2.81 give cubic and quartic self-couplings for the H field with Feynman rules

$$= -3i\frac{m_H^2}{v}; \tag{2.87}$$

$$= -3i\frac{m_H^2}{v^2}. \tag{2.88}$$

	Representation		Charge
Field	$SU(3)$	$SU(2)$	Y
l_L	**0**	$\frac{1}{2}$	-1
q_L	**3**	$\frac{1}{2}$	$+\frac{1}{3}$
e_R	**0**	**0**	-2
u_R	**3**	**0**	$+\frac{4}{3}$
d_R	**3**	**0**	$-\frac{2}{3}$

Table 2.1 Fermion fields of the Standard Model and their gauge group representations.

2.2.3 Fermion sector

The fermion fields of the Standard Model can be expressed in the chiral basis as l_L, q_L, e_R, u_R, and d_R. The left-handed fields are written as isospin doublets,

$$l_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \qquad q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \tag{2.89}$$

whereas the right-handed fields are $SU(2)$ singlets, with $\tau^3 \psi_R = 0$. This means that the weak interaction only couples to left-handed quarks and leptons. There are no right-handed neutrinos in the Standard Model.

The kinetic contribution to the Lagrangian density due to a single fermion field ψ is given by

$$\mathcal{L}_\psi = i\bar{\psi}\slashed{D}\psi \tag{2.90}$$

$$= i\bar{\psi}\gamma^\mu \left(\partial_\mu - ig\tau_a W_\mu^a - ig'\frac{Y}{2}B_\mu - ig_S T_a G_\mu^a \right)\psi, \tag{2.91}$$

where γ^μ are the Dirac matrices, and the generators (τ_a, Y, T_a) are evaluated according to the charge of ψ under the corresponding group. These representations are shown in Table 2.1 for the Standard Model fermion fields. The equation of motion for ψ corresponding to this form of the free-field Lagrangian density, $\mathcal{L} = i\bar{\psi}\slashed{\partial}\psi$, is the Dirac equation [46]. Following quantisation,

the Dirac propagator for a fermion is [6]

$$\xrightarrow{\quad k \quad} \;=\; \frac{i(\not{k} + m)}{k^2 - m^2 + i\epsilon}. \tag{2.92}$$

Only the quark fields q_L, u_R, d_R are charged under $SU(3)$ and the theory of quarks and gluons is known as *quantum chromodynamics* (QCD). The term in Equation 2.91 corresponding to quark–gluon interactions is

$$g_S \bar{\psi} \gamma^\mu T_a G^a_\mu \psi, \tag{2.93}$$

and the associated Feynman rule is given by [6]

$$= ig\gamma_\mu T^a_{ij}, \tag{2.94}$$

where the fermion fields have indices i and j in the $SU(3)$ group space.

After spontaneous symmetry breaking, the electroweak interactions in Equation 2.91 with all the fermion fields can be written [6]

$$\sum_\psi \bar{\psi} \gamma^\mu \left(g\tau_a W^a_\mu + g' \frac{Y}{2} B_\mu \right) \psi = g \left(W^+_\mu J^{+\mu}_W + W^-_\mu J^{-\mu}_W + Z_\mu J^\mu_Z \right) + e A_\mu J^\mu_{\text{EM}}, \tag{2.95}$$

where the currents take the forms

$$J^{+\mu}_W = \frac{1}{\sqrt{2}} \left(\bar{\nu}_L \gamma^\mu e_L + \bar{u}_L \gamma^\mu d_L \right), \tag{2.96}$$

$$J^{-\mu}_W = \frac{1}{\sqrt{2}} \left(\bar{e}_L \gamma^\mu \nu_L + \bar{d}_L \gamma^\mu u_L \right), \tag{2.97}$$

$$J^\mu_Z = \frac{1}{\cos\theta_w} \sum_\psi \bar{\psi} \gamma^\mu \left(\tau^3 - \sin^2\theta_w \, Q \right) \psi, \tag{2.98}$$

$$J^\mu_{\text{EM}} = \sum_\psi \bar{\psi} \gamma^\mu \psi. \tag{2.99}$$

This gives rise to interactions between fermions and the electroweak gauge bosons, with momentum-space Feynman rules:

$$W_\mu^- \qquad = \frac{ig}{\sqrt{2}}\gamma_\mu \tag{2.100}$$

$$Z_\mu \qquad = \frac{ig}{\cos\theta_w}\gamma_\mu\left(\tau_3 - \sin^2\theta_w\, Q\right); \tag{2.101}$$

$$A_\mu \qquad = ieQ\gamma_\mu. \tag{2.102}$$

Massive fermions

Explicit mass terms for the fermion fields (e.g. $-m_e[\bar{e}_L e_R + \bar{e}_R e_L]$) are prohibited from appearing in $\mathcal{L}_\psi$ by the gauge symmetry. Instead, there is a Yukawa coupling between each fermion field and the φ field [6],

$$\mathcal{L}_{\text{Yukawa}} = -\frac{\sqrt{2}}{v}m_e\,\bar{l}_L^a\,\varphi_a\,e_R - \frac{\sqrt{2}}{v}m_d\,\bar{q}_L^a\,\varphi_a\,d_R - \frac{\sqrt{2}}{v}m_u\,\bar{q}_L^a\epsilon_a{}^b\varphi_b^\dagger\,u_R + \text{h.c.}, \tag{2.103}$$

where a and b are $SU(2)$ indices, and h.c. stands for the Hermitian conjugate of the previous terms (i.e. $\bar{e}_L e_R + \text{h.c.} = \bar{e}_L e_R + \bar{e}_R e_L$). Expanding around the vacuum expectation value by substituting φ from Equation 2.79, this becomes

$$\begin{aligned}
\mathcal{L}_{\text{Yukawa}} = &- m_e\,\bar{e}_L e_R - m_d\,\bar{d}_L d_R - m_u\,\bar{u}_L u_R + \text{h.c.} \\
&- \frac{m_e}{v}\,\bar{e}_L\,h\,e_R - \frac{m_d}{v}\,\bar{d}_L\,h\,d_R - \frac{m_u}{v}\,\bar{u}_L\,h\,u_R + \text{h.c..}
\end{aligned} \tag{2.104}$$

	Generation, i		
Field	1	2	3
ν^i	ν_e	ν_μ	ν_τ
e^i	e	μ	τ
u^i	u	c	t
d^i	d	s	b

Table 2.2 Fermion generations in the Standard Model.

The first line of Equation 2.104 gives masses to the fermion fields. The second line describes interactions between the fermion and H fields, with the Feynman rule

$$\vcenter{\hbox{}} = -i\frac{m_f}{v} \tag{2.105}$$

with $f \in \{e, u, d\}$.

    The Standard Model contains three generations of fermions, each containing a copy of the fields discussed above. The fermion content for each generation is shown in Table 2.2. The left-handed quark fields can be written

$$u_L^i = (u_L, c_L, t_L) \tag{2.106}$$

$$d_L^i = (d_L, s_L, b_L) \tag{2.107}$$

in the basis where the qqW couplings are diagonal, but the qqH couplings are mixed. The modified representations $u_L'^i$ and $d_L'^i$ denote the left-handed quark fields in the basis of definite mass, where Equation 2.105 expresses diagonal couplings. These representations are given by

the transformations

$$u_L^i = U_u^{ij} u_L'^j, \tag{2.108}$$

$$d_L^i = U_d^{ij} d_L'^j. \tag{2.109}$$

Substituting Equations 2.108 and 2.109 into Equation 2.96, the W^+ current is given by

$$J_W^{+\mu} = \frac{1}{\sqrt{2}} \left(\bar{\nu}_L \gamma^\mu e_L + \bar{u}_L'^i \gamma^\mu \left[U_u^\dagger U_d \right]_{ij} d_L'^j \right) \tag{2.110}$$

$$= \frac{1}{\sqrt{2}} \left(\bar{\nu}_L \gamma^\mu e_L + \bar{u}_L'^i \gamma^\mu V_{ij} d_L'^j \right), \tag{2.111}$$

where in the last line $V \equiv U_u^\dagger U_d$ is used. The 3×3 Cabibbo–Kobayashi–Maskawa (CKM) [47, 48] matrix V is required to be unitary. This leaves four degrees of freedom, interpretable as three rotation angles, θ_{12}, θ_{23}, θ_{13}, and a CP-violating phase angle δ [49]:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{2.112}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.113}$$

$$= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{2.114}$$

where $s_{ij} \equiv \sin\theta_{ij}$ and $c_{ij} \equiv \cos\theta_{ij}$. The *Wolfenstein parameterisation* [50] expresses V in terms of $\mathcal{O}(1)$ parameters, λ, A, ρ, η, with $\lambda = \sin\theta_{12}$, $A\lambda^2 = \sin\theta_{23}$, and $A\lambda^3(\rho - i\eta) = \sin\theta_{13}e^{-i\delta}$. Expanding in terms of λ,

$$V = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\lambda^4). \tag{2.115}$$

This parameterisation emphasizes that the third generation of quarks is relatively decoupled from the first two. Consequently, the t quark decays almost exclusively to the lighter b quark, which is long-lived.

2.2.4 Summary and status

The Standard Model is a powerful quantum field theory, able to describe many phenomena in particle physics. It has a complex history of development and study, and has fuelled generations of particle physics experiments. The phenomena described by the Standard Model are encapsulated in the full Lagrangian density

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_G + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \sum_{\psi} \mathcal{L}_{\psi} + \mathcal{L}_{\text{Yukawa}} \tag{2.116}$$

along with the mechanical principles set out in Section 2.1. The construction and form of $\mathcal{L}_{\text{SM}}$ relies on a number of symmetries under discrete and continuous transformations.

The theory involves a number of subtle intricacies which allow it to give a description of particle physics at the energy scales of experiments. $\mathcal{L}_{\text{SM}}$ is constructed from fundamental quantum fields, but interacting particles precipitate out by perturbative expansion at such energies. These energy scales (and the particle masses) are much lower than the *ultraviolet cut-off*, the scale at which the Standard Model is expected to break down. The inclusion of this cut-off allows for the cancellation of many divergent quantities that appear in the calculations of amplitudes for processes, leading to physically verifiable predictions.

Another consequential behaviour of the Standard Model is that the interactions of quarks and gluons, governed by QCD, exhibit *asymptotic freedom*. That is, at higher energy scales the

$SU(3)$ gauge couplings of particles becomes weaker. Conversely, the force due to QCD dominates at low energy scales, leading to *confinement*. In the phenomenological *Lund string model* [51, 52], two separating quarks have a gluon *string* between them, a force field concentrated in a flux tube by the gluon self-interaction. As the distance between them increases, so does the energy in the string, until the field yields a pair of new quarks or gluons. These quarks are then subject to the same process, leading to the creation of many new particles. This process is known as *hadronisation*. The result is that no bare quark or gluon is observable, only a spray of stable hadrons – a *jet*.

A successful theory is able to agree with, predict, and explain observations. Under these criteria, the Standard Model has proved to be very successful. To date, all particles corresponding to the fields mentioned in this section have been observed by experiments. This includes quarks in various bound states and resonances, except for the t quark for which observation of $t\bar{t} \rightarrow W^+ W^- b\bar{b}$ in the leptonic channel at the Tevatron was announced in 1995 [53–55]. In 2012, the ATLAS [56] and CMS [57] collaborations announced the observation of a scalar boson at the Large Hadron Collider. When this is interpreted as the H boson, the particle content of the Standard Model is complete. Recently, measurements of processes involving these particles, for example $t\bar{t}H$ associated production, have been found to agree with Standard Model predictions [58, 59].

However, at the time of writing, there still exist some observations which are unexplained by the Standard Model. The theory requires modification or replacement if it is to become a complete description of Nature.

2.2.5 Beyond the Standard Model

A few of the remaining issues with the current description of particle physics are discussed in this section. This list is not exhaustive, and the topics here remain areas of active research.

Grand Unified Theory

The Standard Model contains 19 free parameters to be determined by experiment. In Section 2.2.1, it is shown that the electromagnetic $U(1)$ and weak $SU(2)$ interactions can be combined into one $SU(2) \times U(1)$ *electroweak* interaction. Then symmetry breaking 'freezes out' the low-energy interactions that are observed, and provides a mechanism to generate and relate the masses of the associated vector bosons.

In a similar way, it might be imagined that the QCD $SU(3)$ interaction is also frozen out of some *Grand Unified Theory* (GUT) at a higher energy scale. Indeed, it is found that the coupling constants for the three observed forces take on similar values at a unification scale $\Lambda \sim 10^{16}$ GeV, due to the running of the coupling strengths [60]. This indicates that there exists a theory of unified gauge interactions, from which the separate $SU(3) \times SU(2) \times U(1)$ structure of the Standard Model is the result of a symmetry breaking pattern.

The simplest such theory [61] contains the unified gauge group $SU(5) \supset SU(3) \times SU(2) \times U(1)$. In this theory, the Standard Model particles fit into three generations of $SU(5)$ representations. Another GUT that contains the Standard Model is based on the spin group $SO(10)$ [62]. Despite the neat arrangement of Standard Model particles in these theories, they also predict further phenomena such as the decay of protons. To date, however, the proton lifetime has been experimentally measured to be greater than 2.1×10^{29} yr at 90% confidence level [22, 63].

Gravity

The Standard Model provides no description or prediction for gravitation. The current best explanation for this phenomenon is given by Einstein's theory of general relativity (GR) [64]. The theory provides vital corrections to the classical Newtonian theory that agree with experimental observations to a remarkable degree. Such observations include the moving perihelion of Mercury's orbit around the Sun, gravitational lensing of light, and black holes [65, 66]. Recently, a key prediction of GR was verified by observations of gravitational waves emerging from a black hole merger, an extreme astrophysical event [67, 68]. The peak gravitational wave strain is

10^{-21}, necessitating extremely precise measurements using laser interferometry. In comparison to the Standard Model gauge forces, the gravitational interaction is very weak (e.g. the QCD interaction is 10^{41} times stronger) [69], and therefore its effects are negligible at the scale of particle physics experiments.

GR is a theory of classical fields, in contrast to the quantum fields that make up the Standard Model. There have been many attempts to construct a quantum theory of gravity that recovers GR in the classical limit. Many such theories predict the existence of a spin-2 *graviton* [70], with a mass constrained to less than 7.7×10^{-23} eV at a 95% confidence level by gravitational wave measurements [22, 71]. So far, the production and decay of these exotic particles at collider experiments have not been detected, and the existence of a Kaluza–Klein graviton [72, 73] with mass below 2.3 TeV has been excluded at a 95% confidence level [74].

Some modern approaches evangelise a reconsideration of axioms and assumptions that play roles in the construction of the classical and quantum theories in an attempt to unify them. For example, in one approach [75–77] spacetime and its geometry emerge from quantum information theory, where distance is a property of the entanglement of quantum states. Owing to the finiteness of Hilbert space, theories of this category also predict that Lorentz invariance should be violated [78]. Experiments have not produced any unexplainable result showing a significant degree of violation [79–82].

The hierarchy problem and fine-tuning

The gravitational force is significantly weaker than the Standard Model gauge interactions – the weak coupling is of order 10^{24} times larger, for example [83]. This discrepancy is manifested in the mass of the H field being disproportionately smaller than the GUT energy scale. In the Standard Model, quadratic divergences arise when summing over diagrams for scalar particles with free momenta, such as in the H self-coupling [84],

$$\text{\raisebox{-1ex}{\textit{k}}} \quad \sim \quad -\frac{3}{4\pi^2}\frac{m_t^2}{v^2}\Lambda^2. \tag{2.117}$$

Here Λ is a large ultraviolet cut-off at the GUT scale, and other loop diagrams cancel out the divergence to leave the relatively small bare mass for H. Such precise fine-tuning is considered unnatural in a physical theory.

A solution is proposed by *supersymmetric* theories, whereby every fermion and boson field in the Standard Model has a boson or fermion *superpartner*, respectively [85, 86]. This additional symmetry would precisely cancel out quadratic divergences in perturbative calculations. Additionally, supersymmetry permits a modification to the running of the coupling constants, such that the unification of forces has better agreement at Λ [60]. Extensive searches at collider experiments have found no evidence for any supersymmetric extension to the Standard Model [87].

Massive neutrinos

The Sun burns hydrogen into helium in two fusion processes: a proton–proton (pp) chain reaction, and the CNO cycle [88]. The overall reaction in the pp chain is summarised by

$$4p \rightarrow {}^4\text{He} + 2e^+ + 2\nu_e. \tag{2.118}$$

This reaction occurs through four possible branches, each producing neutrinos with different energy distributions. The rates of each of these can be predicted [89], and consequently the total neutrino flux emitted from the Sun due to the pp chain is expected to be $5.98 \pm 0.04 \times 10^{10} \, \text{cm}^{-2}\text{s}^{-1}$ [90]. Importantly, these neutrinos are expected to be only electron-type (ν_e), as indicated by Equation 2.118, assuming their state is fixed from production to detection. However, measurements have found the solar ν_e flux, detected via charged current interactions, to have a large deficit which cannot be explained by the Standard Model. For example, the Sudbury Neutrino Observatory (SNO) reported [91] a charged current ${}^8\text{B}$ ν_e flux 3.3σ lower than that from electron scattering, which is sensitive to all neutrinos.

The solution is a theory of *neutrino oscillation* [92, 93], whereby neutrinos change type (between ν_e, ν_μ, and ν_τ) as they propagate. Evidence for this phenomenon is given by subsequent

results from SNO [94], for example, where the total flux from neutral current interactions (sensitive to all neutrinos equally) was found to be consistent with the Standard Solar Model. Such a change of state while travelling requires that there exists a rest frame for the particle and therefore neutrinos are required to have non-zero mass, in contradiction to the Standard Model. Analogously to the quark mixing mechanism discussed in Section 2.2.3, the physical mass eigenstates $(\nu_1, \nu_2, \nu_3)_L$ and weak interaction eigenstates $(\nu_e, \nu_\mu, \nu_\tau)_L$ for the neutrinos in this model are distinct and related by a transformation

$$|\nu_l\rangle = \sum_{j=1}^{3} U_{lj}^* |\nu_j\rangle, \quad l = e, \mu, \tau, \tag{2.119}$$

where U is the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) [92, 95] matrix. If it is constrained to be unitary, the matrix can be parameterised in terms of three mixing angles θ_{12}, θ_{23}, θ_{13} and a CP-violating phase angle δ. The oscillation probabilities can be shown [22] to depend on the difference of the square masses, $\Delta m_{ij}^2 = m_i^2 - m_j^2$, $i > j$. In the full three-neutrino mixing model, the 'survival' probability for a ν_e to be detected as ν_e is given by [22, 96]

$$P(\nu_e \to \nu_e) = \cos^4 \theta_{13} \left[1 - \sin^2(2\theta_{12}) \sin^2 \left(\frac{\Delta m_{21}^2 L}{4E} \right) \right] + \sin^4 \theta_{13}, \tag{2.120}$$

where L is the distance from emission to detection, and E is the energy of the ν_e.

A modification to the Standard Model is required to describe the mechanism by which neutrinos acquire mass. Its form depends on whether the ν_i, $i = 1, 2, 3$ are *Dirac* fermions [97], like the electron, or *Majorana* fermions, where the field is a solution to the Majorana wave equation [98]. In the Dirac case, *lepton number* is conserved in interactions. For Majorana neutrinos, however, it is not a conserved quantity. In the later case, neutrinoless double-beta decay is a predicted process, since the neutrino would be its own antiparticle. There are some experiments looking for this phenomenon, but it has so far not been observed [99–102]. If neutrinos are Majorana fermions, this adds another two degrees of freedom to the theory, in the form of CP-violating phase angles.

The *seesaw mechanism* [103–107] provides a convenient explanation for the small mass of the left-handed neutrinos, relative to the massive Standard Model particles. It introduces right-handed chargeless Majorana neutrino fields, N_{iR}, with large masses M_i, $i = 1, 2, 3$. Following the Standard Model electroweak symmetry breaking process, in the basis where the mass matrix is diagonal, the neutrinos acquire a Dirac mass term (mixing the right- and left-handed fields), of the form $m_{il}^D \, \overline{N_{iR}} \, \nu_{lL} + \text{h.c.}$ For $|m_{il}^D| \ll M_j$, $i, j = 1, 2, 3$, $l = e, \mu, \tau$, the left-handed neutrinos acquire an effective Majorana mass term $m_{l'l}^{LL} \cong -[m^D]_{l'j}^{\mathsf{T}} M_j^{-1} m_{jl}^D$ [22]. Empirically, and in many GUTs, $m^{LL} \sim 0.1\,\text{eV}$, $m^D \sim 100\,\text{GeV}$, $M \sim 10^{14}\,\text{GeV}$.

Baryon asymmetry

The amount of baryonic matter versus antimatter observed in the present-day universe is highly unbalanced [108, 109]. The Standard Model predicts the conservation of charge, so the baryogenesis mechanisms produce equal proportions of particles and antiparticles, to leading order. *CP*-violation in the Standard Model does permit a matter–antimatter asymmetry, but the measured strength of this violation is incompatible with large-scale observations [110–112]. The seesaw mechanism, discussed above, could provide a solution [113, 114]. *CP*-violation in decays of heavy right-handed neutrinos in the primordial universe could lead to the asymmetry observed today.

Dark matter

Measurements of nearby galaxies' velocity curves by redshift do not match density predictions based on the light emitted from them [115–117]. Modifications to gravitational theories have so far been unsuccessful in describing this while preserving the excellent agreement between GR and observations [118]. Therefore, these observations suggest the existence of massive *dark matter*, distributed in a 'halo' around the galaxy [119]. The universal energy density contributed by dark matter is estimated to be around 27% by satellite measurements of the *cosmic microwave background* [120–122].

It is expected that there exists some modification to the Standard Model to account for the presence of dark matter. Candidate cold dark matter particles can be classed as axions and *weakly interacting massive particles* (WIMPs). Axionic bosons are presented as a possible solution to the *strong CP problem* [123]. They have a small predicted mass – 10^{-6} to $1\,\mathrm{eV}$, depending on the model [124] – and are expected to interact minimally with Standard Model fields. WIMPs, conversely, are theoretical particles of moderate to high mass (1 to $10^4\,\mathrm{GeV}$) also with low interaction cross sections. Candidates for WIMPs can be provided by GUTs, which often predict the existence of heavy bosons [125]. Finally (but not exhaustively), sterile neutrinos of an $SU(2)_L \times U(1)_Y$ singlet field with mass $\sim 1\,\mathrm{keV}$ could fit the profile for cold dark matter.

Indirect detection experiments aim to detect the products of interactions between dark matter particles, often in the form of an excess astronomical photon signal [126]. Despite many searches and fluctuations seen by experiments, no significant signal has so far been reported [127]. Direct detection experiments, on the other hand, aim to observe the scattering of dark matter with Standard Model particles, usually large amounts of material nuclei shielded in underground chambers. Currently, the nucleon–WIMP interaction cross section has been excluded to approximately $10^{-46}\,\mathrm{cm}^2$ for a WIMP mass of $100\,\mathrm{GeV}$ at a 90% confidence level [128]. Modern collider experiments also have extensive programmes searching for dark matter particles, but no signal has been detected [129].

2.3 $t\bar{t}$ production cross sections

The t quark is the heaviest particle in the Standard Model, with a directly measured mass of $m_t = 172.69 \pm 0.25\,\mathrm{(stat)} \pm 0.41\,\mathrm{(syst)}\,\mathrm{GeV}$ [130]. The next most massive particle is the H boson at $125.10 \pm 0.14\,\mathrm{GeV}$ [22]. There is currently no mechanism that precisely predicts the masses of the Standard Model particles (that is, they are free parameters of the model), and it is interesting to note that m_t is significantly larger than the masses of the other five quarks. It is similar in magnitude to the electroweak energy scale at $v \approx 246\,\mathrm{GeV}$ [131]. Interactions involving t quarks are especially sensitive to effects described beyond the Standard Model.

The large mass of the t quark means that there is extensive phase space available for its decay. The measured total decay width is $\Gamma_t = 1.42^{+0.19}_{-0.15}\,\text{GeV}$ [22]. This is related to the lifetime – the time for the survival probability to reach e^{-1} – by $\tau = \hbar/\Gamma$, giving the t quark lifetime as $\tau_t \approx 4.6 \times 10^{-25}\,\text{s}$. Because of this extremely short lifetime, it decays effectively instantly ($c\tau_t \approx 1.4 \times 10^{-16}\,\text{fm}$) and hence the t quark is the only quark to never hadronise. Consequently, the 4-momentum of the t quark is conserved in its direct decay products which can be mathematically reconstructed. This permits the study of a 'bare' quark, a unique opportunity in experimental particle physics.

2.3.1 Motivation for measurements

A precise understanding of t quarks and the mechanisms of their production tests the limits of the accuracy of the Standard Model. t quarks are mostly produced via $t\bar{t}$ pair production in collider experiments [22]. The rate of $t\bar{t}$ pair production via the cross section is sensitive to modifications to the Standard Model [132–134].

The total, or *inclusive*, cross section provides a measurement of the rate of production of $t\bar{t}$ in collider experiments, and it is different for pp collisions versus $p\bar{p}$. Total cross sections have been measured at the a variety of collision energies, spanning almost an order of magnitude [135]. A summary of these measurements is shown in Figure 2.1. As can be seen from the figure, the measured total cross sections are in good agreement with theoretical predictions from the Standard Model, calculated up through $\mathcal{O}(\alpha_S^4)$ [134].

Additionally, the kinematic properties of the $t\bar{t}$ system are subject to modifications from extensions to the theory [137]. It is therefore interesting to study the distributions of $t\bar{t}$ production cross section as functions of measurable kinematic variables. Such distributions are reported as *differential* cross sections, $\frac{\mathrm{d}\sigma}{\mathrm{d}x}$, with the relation

$$\sigma_{\text{total}} = \int_{\mathcal{X}} \frac{\mathrm{d}\sigma}{\mathrm{d}x}\,\mathrm{d}x, \tag{2.121}$$
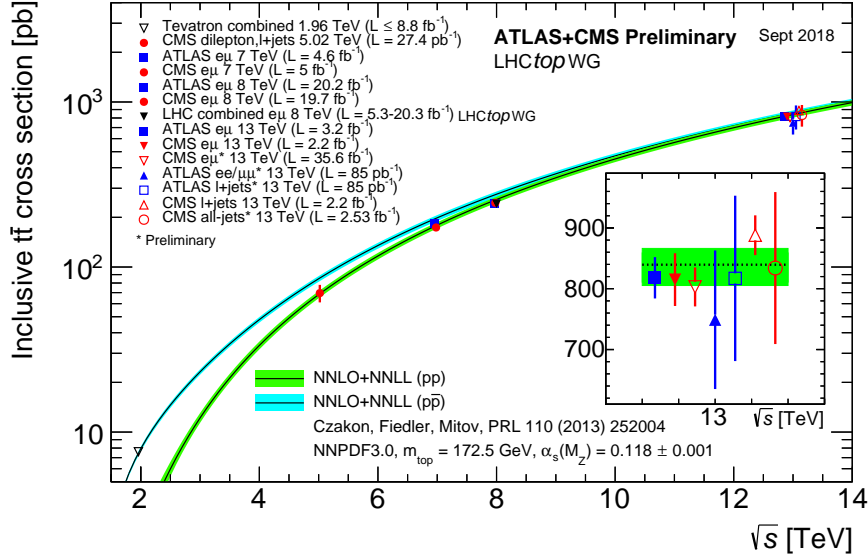
Figure 2.1 Predictions and measurements for total $t\bar{t}$ cross sections at collider experiments, as a function of the centre-of-mass energy $\sqrt{s}$. From [136] with theoretical predictions from [134].

where $x \in \mathcal{X}$ stands for the kinematic variable of interest. Measurements of $t\bar{t}$ differential cross sections can be used to perform intricate tests of Standard Model predictions at the TeV scale.

$t\bar{t}$ pair production is also a dominant background process in many searches for physics beyond the Standard Model [138, 139] and studies of the Standard Model H boson [140] at collider experiments. For example, an analysis searching for the production of $t\bar{t}H$ (with $H \to b\bar{b}$) [141] found that approximately 85–95% of background events in the signal and control regions are from $t\bar{t}$ production. Diagrams for the $t\bar{t}H$ signal process and one such possible background process are shown in Figure 2.2. This background contribution was suppressed using stacked multivariate models to discriminate events containing a H boson. However, the uncertainty in modelling the large $t\bar{t} + \geq 1b$ background is the dominant systematic uncertainty in the analysis.

(a) $t\bar{t}H(H \to b\bar{b})$ (b) $t\bar{t}b\bar{b}$
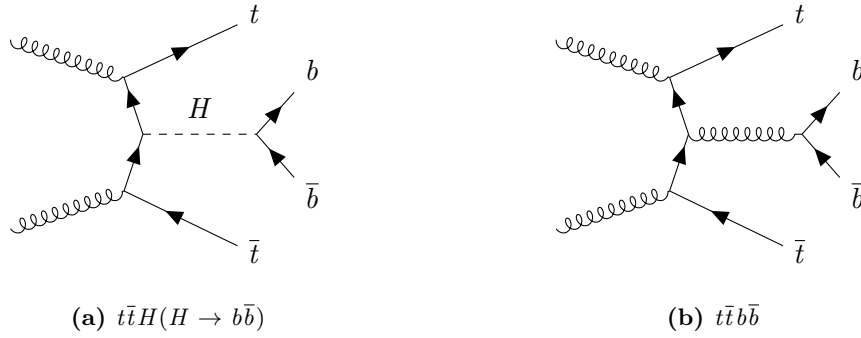
Figure 2.2 Signal process (a) for one channel of a $t\bar{t}H$ search. Production of $t\bar{t}b\bar{b}$ (b) contributes an important background to the $t\bar{t}H$ signal.

2.3.2 Decay channels

As discussed in Section 2.2.3, the third quark generation is almost independent of the lighter two generations in the electroweak interaction eigenbasis. This means that, for experimental considerations, the t quark decays exclusively to b in the process $t \to W^+b$ or $\bar{t} \to W^-\bar{b}$:
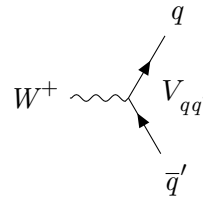
$$ \tag{2.122} $$

t quark–antiquark pair ($t\bar{t}$) decay processes therefore contain at least two b quarks, observed via the decay products of B hadrons. The decay modes are stratified according to the products of the two W bosons into separate experimental analyses. Approximately two thirds of W boson decays are to quarks, resulting in a hadronic final state. The other third are decays to

Decay mode	BR (%)	Unc. (%)
$W^+ \to e^+ \nu_e$	10.71	± 0.16
$W^+ \to \mu^+ \nu_\mu$	10.63	± 0.15
$W^+ \to \tau^+ \nu_\tau$	11.38	± 0.21
$W^+ \to$ hadrons	67.41	± 0.27
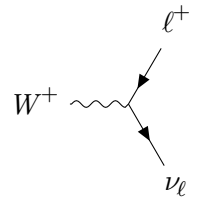
Table 2.3 Branching ratios $\mathrm{BR} = \Gamma_i/\Gamma_{\mathrm{total}}$ for the decay modes i of the W^+ boson. The rightmost column indicates the total uncertainty. Values from [22].

leptons. The corresponding vertices are given by

$$q \in \{u, c\}, \quad q' \in \{d, s\}, \tag{2.123}$$

$$l \in \{e, \mu, \tau\}. \tag{2.124}$$

The measured branching ratios for an on-shell W^+ boson are shown in Table 2.3.

All $t\bar{t}$ decay channels contain at least two *b*-tagged jets, as described in Section 3.3.4. In the *dilepton* channel both W bosons decay to leptons, with a final state containing $e^- e^+$, $e^- \mu^+$, $e^+ \mu^-$, or $\mu^- \mu^+$, plus two neutrinos. The leptonic signature of this process allows for a relatively clean measurement [142], although the smaller branching ratio for $t\bar{t} \to$ leptons means that fewer events pass selection, compared to the channels containing hadrons. The neutrinos are not detected in current analyses and since there are two of them, the $t\bar{t}$ system cannot be fully reconstructed. In the *lepton+jets* decay channel, one of the W bosons decays hadronically and the other to leptons. In this case, there is only one invisible neutrino, so its 4-momentum can be determined by conservation considerations. The initial longitudinal momenta of the partons in the hard scattering are unknown, so a *pseudo-top* algorithm [143, 144] is used to
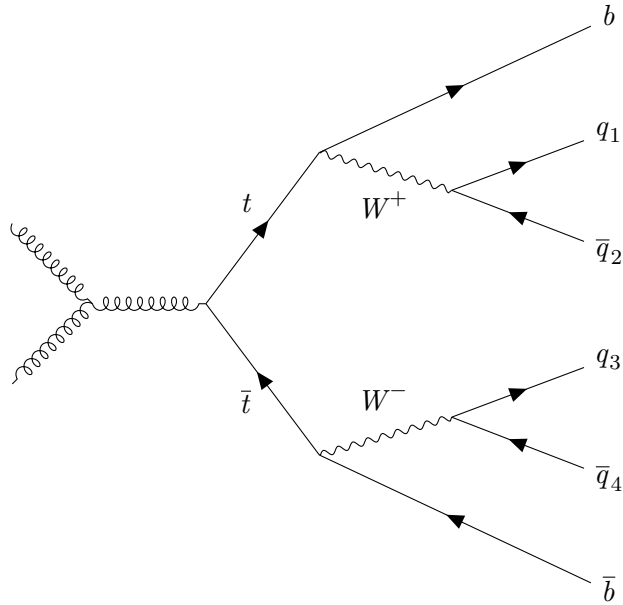
Figure 2.3 Tree-level diagram for one mode of the production of $t\bar{t}$ and decay in the fully hadronic channel.

infer the properties of the $t\bar{t}$ system in this channel. The *fully hadronic* decay mode gives the only channel in which the $t\bar{t}$ system can be completely reconstructed from the observed final state, since there are no invisible decay products. A diagram for this decay (in the *gg* fusion production mode) is shown in Figure 2.3. The data analysis in Chapter 4 is performed in the fully hadronic channel, and motivations for these cross section measurements are discussed in Section 2.3.1 below.

In the t quark rest frame, the b and W decay products travel back-to-back by conservation of zero total 3-momentum. In the laboratory frame, however, the t quarks decay with some momentum which is carried forward. Higher momenta particles decay into more closely collimated products. Therefore for a t quark with large momentum, the B hadron and the (leptonic or hadronic) products of the W boson are close in the detector. It is said that the t quarks are *boosted*. For example, an ATLAS analysis [145] targeted boosted t quarks in the fully hadronic channel, requiring the two reconstructed t quarks to have transverse momenta of at least 500 GeV and 350 GeV. In these cases, separate selection criteria are needed to filter

relevant collision events. Therefore, separate experimental channels containing boosted t quarks exist for the lepton+jets and fully hadronic decay modes.

2.3.3 Previous measurements

The t quark was discovered via $t\bar{t}$ pair production 1995 by the CDF and D0 collaborations at the Tevatron [54, 55]. These experiments have since performed measurements of differential $t\bar{t}$ production cross sections for $p\bar{p}$ collisions with centre-of-mass energy $\sqrt{s} = 1.96$ TeV [146–150]. At the LHC (described in Section 3.1), the ATLAS and CMS collaborations have published multiple $t\bar{t}$ differential production cross section measurements for $\sqrt{s} = 7$, 8, and 13 TeV [142, 145, 151–174]. Brief summaries of measurements with ATLAS at $\sqrt{s} = 13$ TeV are given here.

In the dilepton channel, $3.2\,\text{fb}^{-1}$ of LHC pp collision data recorded in 2015 were analysed [142]. Collision events with one e and one μ of opposite charge, plus two b-tagged jets (detailed in Sections 3.3.3 and 3.3.4) were selected. Differential $t\bar{t}$ production cross sections are reported as a function of the transverse momentum p_T and absolute rapidity $|y|$ for both the t quark and $t\bar{t}$ system, in addition to the invariant mass $m_{t\bar{t}}$ of the $t\bar{t}$ system. The resulting normalised differential cross sections as functions of t quark p_T and $m_{t\bar{t}}$ are shown in Figure 2.4.

The same dataset was used to perform measurements of differential $t\bar{t}$ production cross sections in the lepton+jets decay channel [161]. Events containing exactly one e or μ, plus least two jets of hadrons were selected. In this analysis, selection criteria are defined such that the momenta of the t quarks were categorised as resolved or boosted, and the cross sections were then measured separately in each topology. In the resolved regime, absolute and normalised differential cross sections are reported as functions of the hadronically-decaying t quark p_T and absolute rapidity, as well as invariant mass, p_T, and absolute rapidity of the $t\bar{t}$ system. The normalised differential cross section as a function of the hadronically-decaying t quark is shown in Figure 2.5a, and as a function of $m_{t\bar{t}}$ in Figure 2.5b. In the boosted regime, differential cross sections are reported as functions of the p_T and absolute rapidity of the hadronically-decaying t quark, shown in Figures 2.5c and 2.5d.
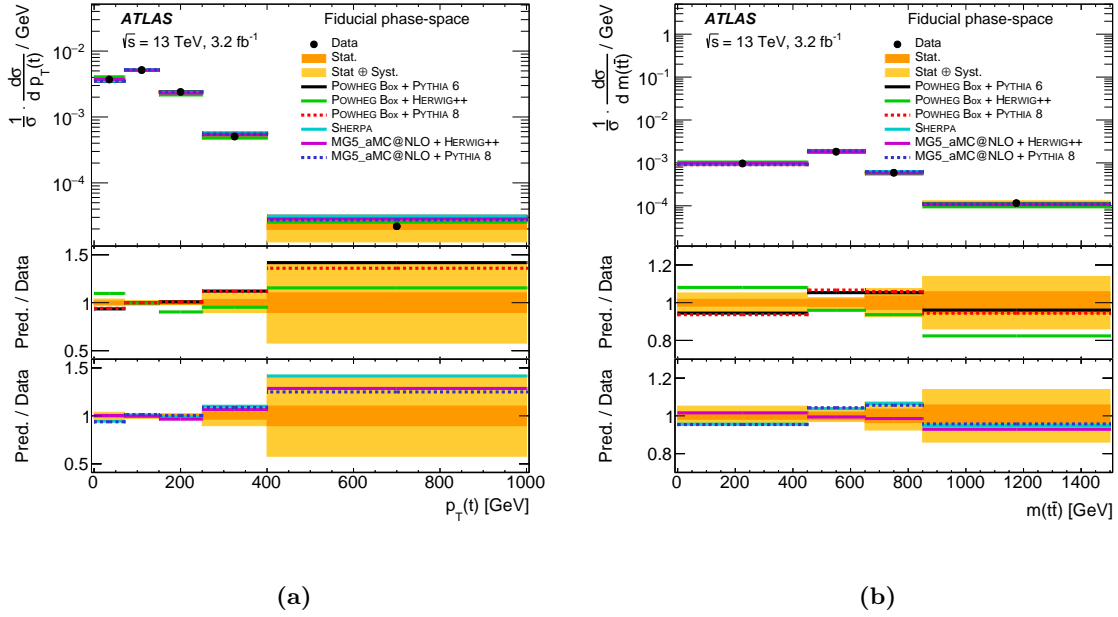
Figure 2.4 Normalised $t\bar{t}$ fiducial differential cross sections as functions of (a) t quark p_{T} and (b) $m_{t\bar{t}}$ in the $e\mu$ channel. These are a subset of the results from [142].

A larger dataset of $36.1\,\mathrm{fb}^{-1}$ LHC pp collisions, collected by ATLAS in 2015 and 2016, was analysed in the all-hadronic channel, where both t quarks decay to hadrons [145]. This analysis used events where the t quarks were highly boosted, by selecting those containing at least two large-R jets with $p_{\mathrm{T}}(t_1) > 500\,\mathrm{GeV}$ and $p_{\mathrm{T}}(t_2) > 350\,\mathrm{GeV}$ that both fall inside a mass window, $|m_{\mathrm{jet}} - m_t| < 50\,\mathrm{GeV}$. Any events with an e or μ were vetoed, and further cuts were made to reduce the mostly QCD-produced multi-jet background, increasing the purity of the sample. The cross sections were *unfolded* to both a fiducial phase space at *particle level* and a reduced phase space at *parton level*. (These terms are explained in Section 4.6 below.) The differential cross sections are measured as functions of the leading (higher p_{T}) and subleading (lower p_{T}) t quark p_{T} and absolute rapidity $|y^t|$, as well as the $t\bar{t}$ system p_{T}, invariant mass, and other kinematic quantities. Other variables of interest are defined in the $t\bar{t}$ system rest frame, for example $\chi^{t\bar{t}} = \exp(|y^t - y^{\bar{t}}|)$ and the cosine of the relative polar angle, $\cos\theta^\star$. These variables, among others, are defined and explained in Section 4.2. The particle-level normalised differential $t\bar{t}$ production cross sections are shown as functions of the leading t quark p_{T} and
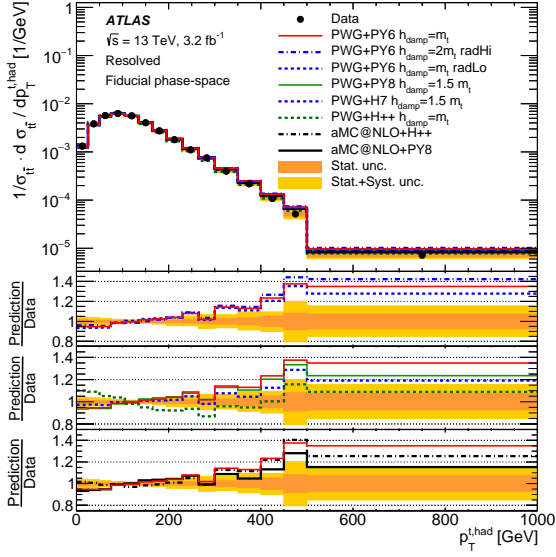
(a) Resolved, hadronic t quark p_T
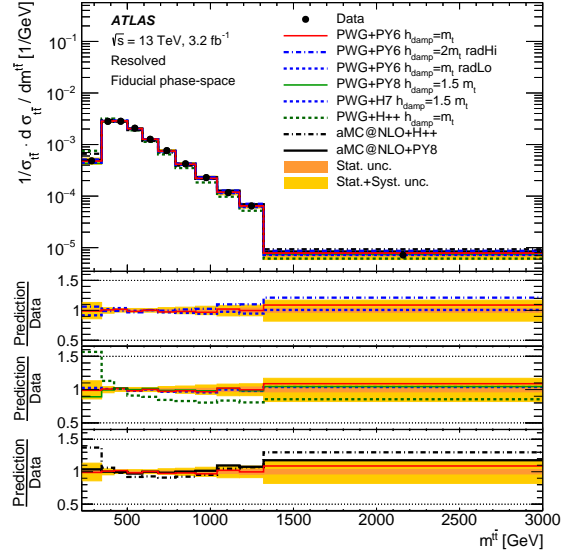
(b) Resolved, $m_{t\bar{t}}$
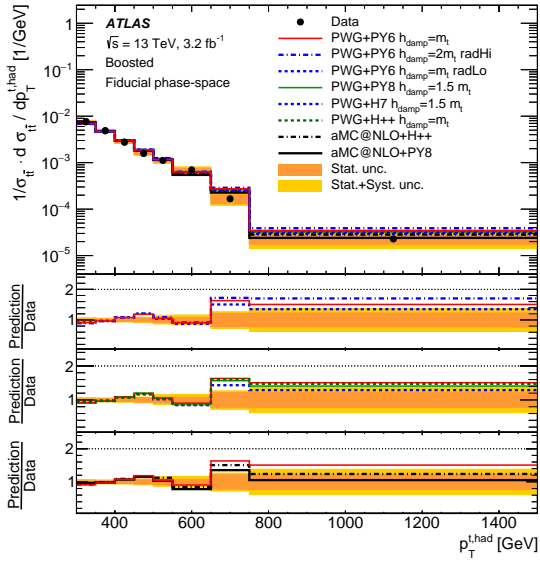
(c) Boosted, hadronic t quark p_T
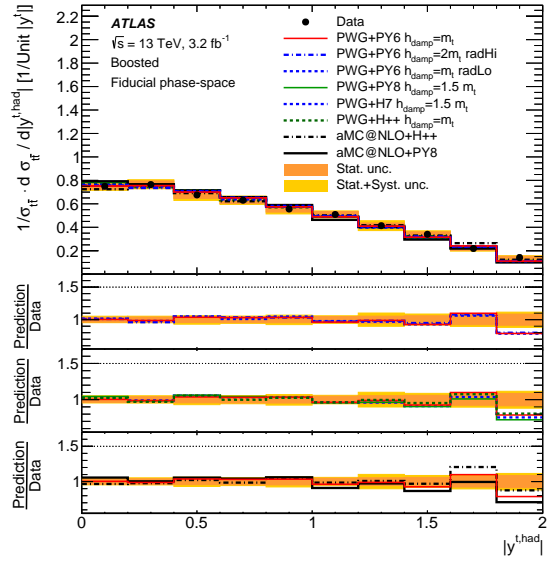
(d) Boosted, hadronic t quark absolute rapidity

Figure 2.5 Normalised $t\bar{t}$ fiducial differential cross sections measured in the lepton+jets channel. These are a subset of the results from [161].

absolute rapidity, in addition to $|\cos\theta^{\star}|$ and $\chi_{t\bar{t}}$ at particle level in Figure 2.6 and at parton level in Figure 2.7. These results serve as a baseline for the measurements in Chapter 4.

2.3.4 The fully resolved all-hadronic channel

The analysis described in Chapter 4 is performed in the all-hadronic channel, where both W bosons from the $t\bar{t}$ decay into a final state of hadrons. In contrast to the all-hadronic boosted $t\bar{t}$ analysis summarised above, events are selected where all jets can be resolved with $R = 0.4$. This targets collision events producing relatively low-p_{T} t quarks, such that their decay products are separated. In this channel, the $t\bar{t}$ system can be fully reconstructed from the observed final state. There are no invisible decay products, such as neutrinos, and the measurement achieves good resolution in η.

The potentially high precision of this measurement technique permits further phenomenological studies, for example the extraction of a t quark pole mass [175–177]. By reference to Equation 2.92, the t quark propagator with momentum k is given by

$$\xrightarrow[t]{k} \quad = \frac{i(\not{k} + m_t)}{k^2 - m_t^2 + i\epsilon}, \tag{2.125}$$

with the pole mass defined as the positive square root of the real part of the complex pole. The strong dependence of absolute and differential $t\bar{t}$ production cross sections on m_t can be calculated [134], and a maximum likelihood fit to data performed. Measurements with ATLAS at pp collision energies of 7 and 8 TeV report t quark pole mass measurements in agreement with predictions within experimental uncertainties of approximately 1 GeV [178, 179]. No such analysis has been performed with ATLAS data at $\sqrt{s} = 13$ TeV to date, although a recent CMS result from data collected in 2016 at this collision energy reported $m_t = 170.5 \pm 0.8$ GeV [180]. This analysis was performed in the dileptonic decay channel, whereas the high resolution afforded by the fully hadronic analysis presented in Chapter 4 could lead to a more precise measurement.

(a)

(b)

(c)

(d)
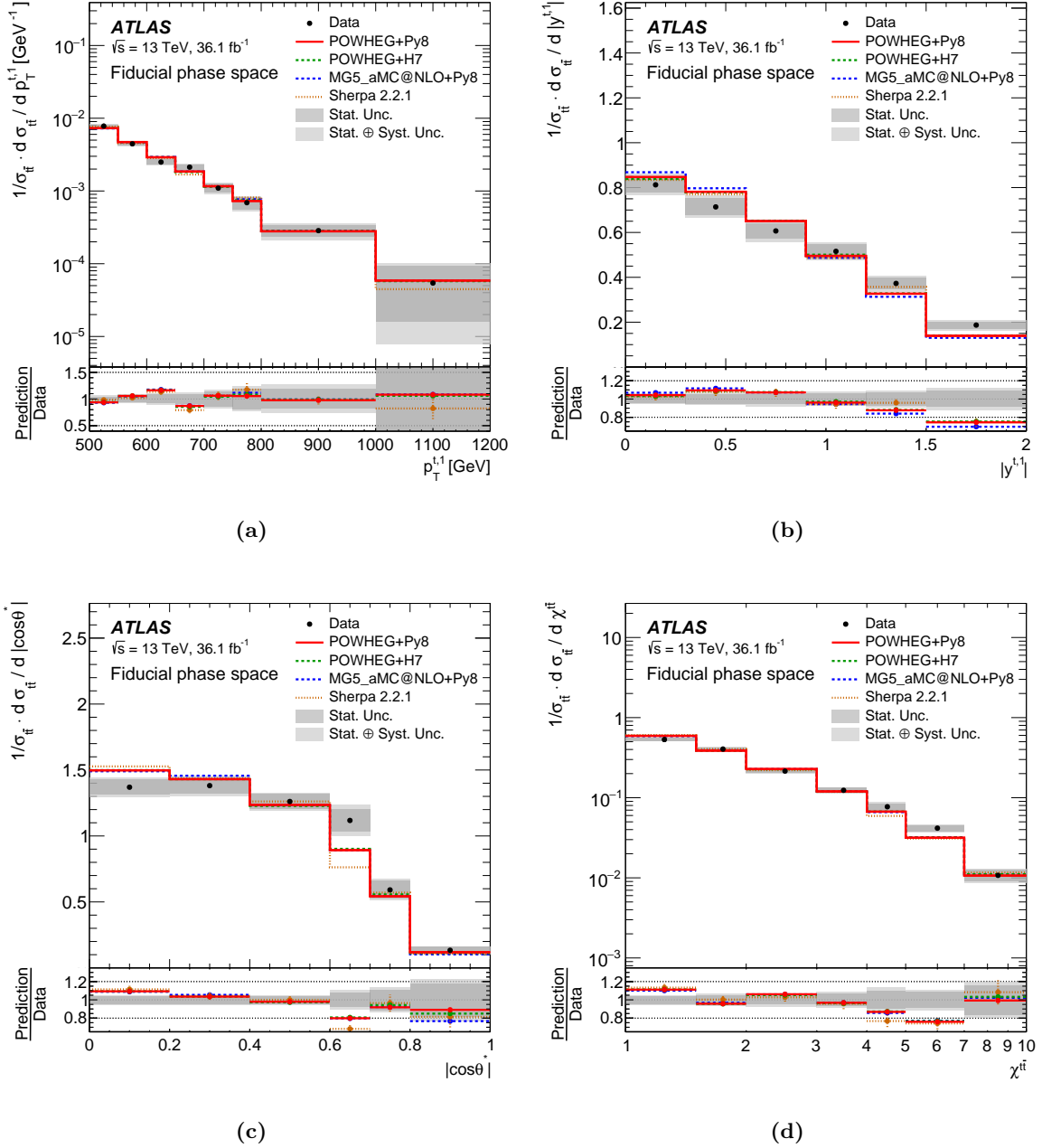
Figure 2.6 Normalised $t\bar{t}$ fiducial differential cross sections in the fully boosted all-hadronic channel, unfolded to particle level. Cross sections are reported as functions of (a) leading t quark p_{T}, (b) leading t quark absolute rapidity, (c) $|\cos\theta^{\star}|$, and (d) $\chi_{t\bar{t}}$, defined in the text. These are a subset of the results from [145].
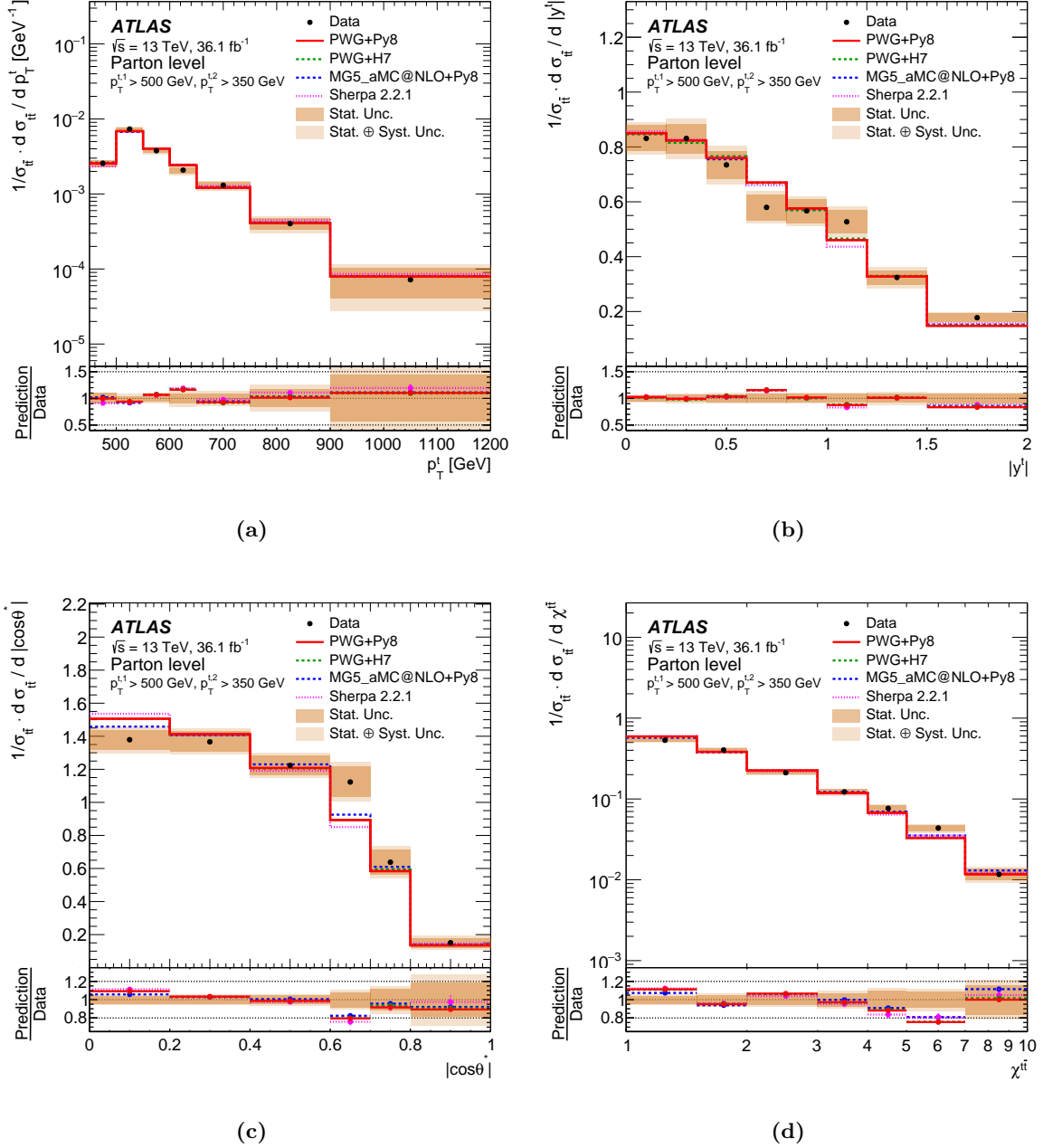
Figure 2.7 Normalised $t\bar{t}$ differential cross sections in the fully boosted all-hadronic channel, unfolded to parton level. Cross sections are reported as functions of (a) t quark p_{T}, (b) t quark absolute rapidity, (c) $|\cos\theta^{\star}|$, and (d) $\chi_{t\bar{t}}$, defined in the text. These are a subset of the results from [145].

Another motivation for performing this measurement is given by the fact that it is the first such analysis in this channel performed by ATLAS at the LHC [181]. The resulting $t\bar{t}$ production cross section measurements complement those performed in other $t\bar{t}$ decay channels on the $36.1\,\mathrm{fb}^{-1}$ 2015+2016 ATLAS dataset. Uniquely in this channel, low-p_T t quarks can be fully reconstructed and their kinematic distributions compared with theoretical predictions. The parton-level differential cross sections from one set of predictions [182] are shown in Figure 2.8 for t quark p_T, $t\bar{t}$ invariant mass, and $t\bar{t}$ absolute rapidity. It can be seen that the predicted distributions are highly sensitive to perturbative corrections in the Standard Model, and may also exhibit similar sensitivity to effects of new physics beyond the established theory. The low-p_T region contains the bulk of this sensitivity, peaking at approximately $75\,\mathrm{GeV}$.

The good angular resolution enabled by the fully resolved topology allows precise measurements of QCD radiation emitted almost collinearly to the t quark. Correlations between the $t\bar{t}$ decay final product kinematics and additional hadronic radiation can be studied in detail. NNLO QCD predictions with NLO electroweak corrections [183] predict a high sensitivity of differential cross sections as functions of kinematic properties of this additional radiation to perturbative effects. Therefore a precision measurement of such distributions could constrain the parameter space for theoretical modelling.

The CMS Collaboration performed measurements in the fully resolved all-hadronic $t\bar{t}$ decay channel [184]. The analysis was done on $2.53\,\mathrm{fb}^{-1}$ of $\sqrt{s} = 13\,\mathrm{TeV}$ LHC pp collision data, collected in 2015 with the CMS detector [185]. The selected events were combined with a separate sample containing boosted t quarks to perform a combined analysis. The $t\bar{t}$ differential cross section as a function of the leading t quark transverse momentum was unfolded to parton level and extrapolated to the full phase space. The absolute cross section distribution and a comparison to Monte Carlo predictions are shown in Figure 2.9. This analysis observed that the t quark p_T is distributed *softer* (lower average p_T) than predicted by Standard Model calculations at next-to-leading order (NLO).

(a)

(b)

(c)
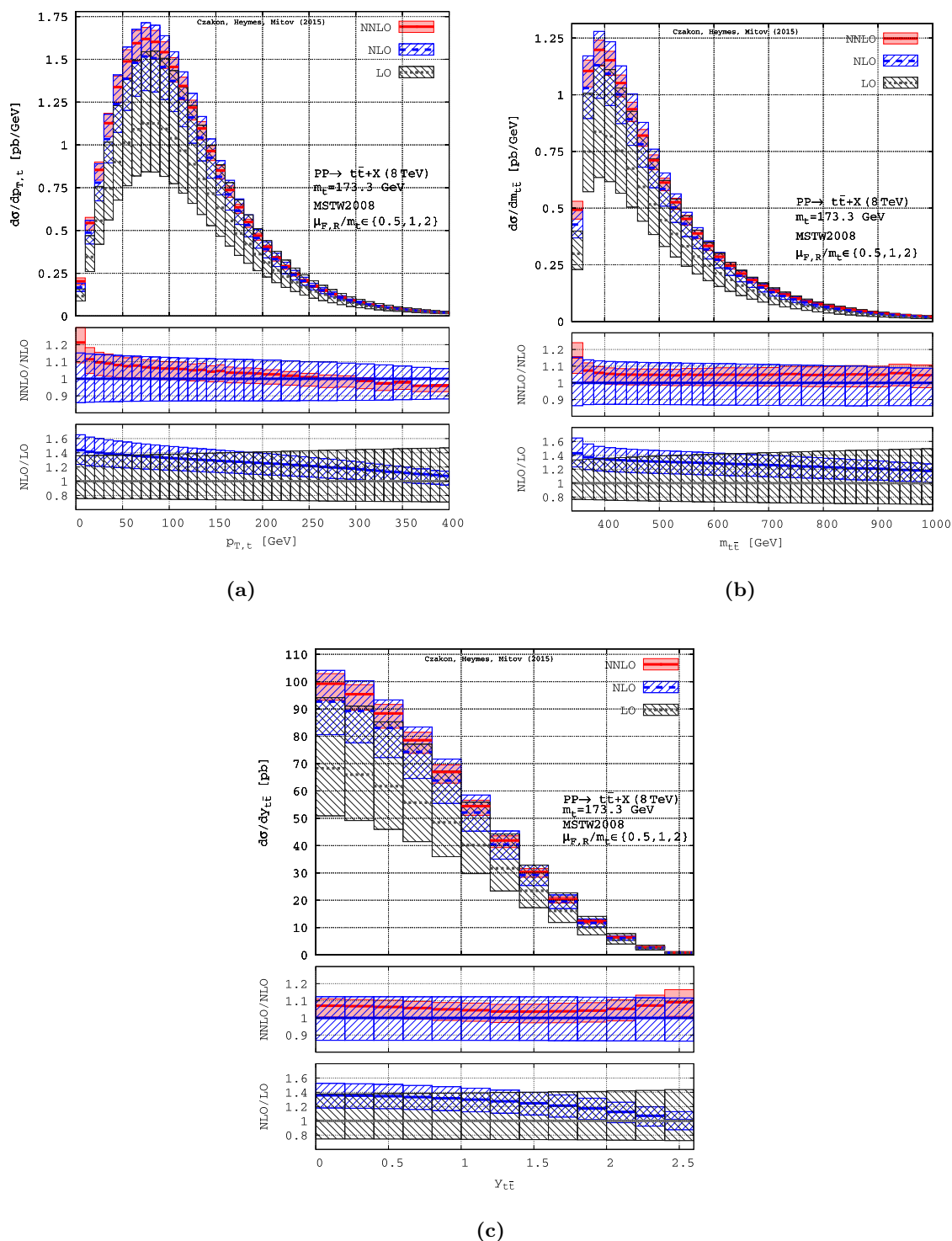
Figure 2.8 Predicted differential $t\bar{t} + X$ cross sections as functions of (a) t quark p_{T}, (b) $t\bar{t}$ invariant mass, and (c) $t\bar{t}$ absolute rapidity. The calculations are shown in perturbative expansions up to leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO), in the full phase space. From [182].
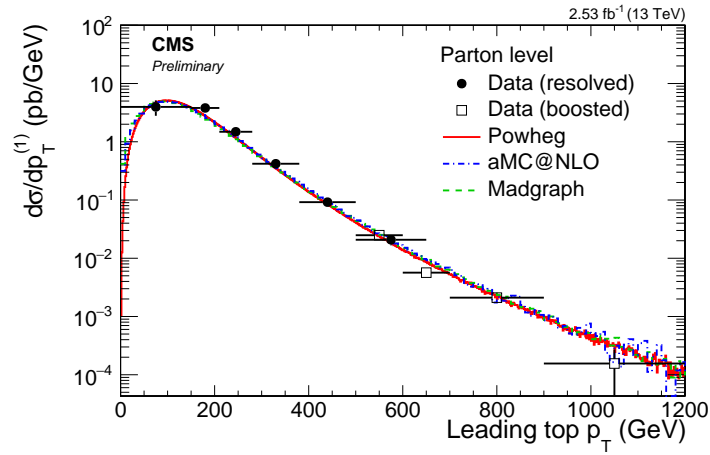
(a)
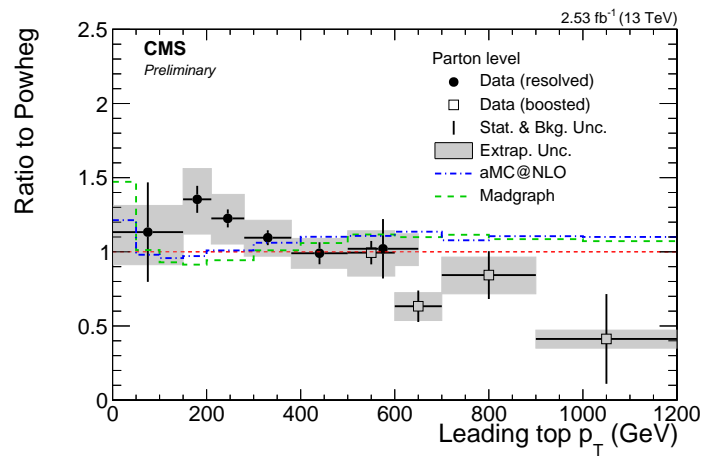
(b)

Figure 2.9 (a) Absolute differential $t\bar{t}$ production cross section as a function of the leading t quark p_{T}, measured with CMS. Collision events were selected using all-hadronic $t\bar{t}$ decays in the resolved and boosted channels, extrapolated to the full phase space, and unfolded to parton level. (b) Ratio Data/MC where MC is the prediction from POWHEG+PYTHIA8. From [184].

Chapter 3

Experimental setup and method

Phenomenological predictions made with theories such as the Standard Model and its extensions can be tested by experiment. In order to carry out the analysis in Chapter 4, data are collected from experimental conditions provided by a complex arrangement of machinery, described in Section 3.1. A bespoke detector system acquires, filters, and processes the data, as detailed in Section 3.2.

3.1 The Large Hadron Collider

The European Organisation for Nuclear Research (CERN) hosts a complex of machinery, infrastructure, and experiments for fundamental high energy physics research. At the heart of this is a chain of particle accelerators culminating in the Large Hadron Collider (LHC) [186], a pair of intersecting near-circular rings that provide high energy particle collisions to detectors. While the LHC supports heavy ionic beams (e.g. Pb and Xe), the proton–proton (pp) programme is the main concern of this thesis.

First, molecular hydrogen is released into a linear accelerator, *Linac 2*, which ionises $H_2 \rightarrow 2p + 2e^-$ and accelerates the protons to an energy of $50\,\text{MeV}$. The protons are fed into the *Proton Synchrotron* (PS) at $1.4\,\text{GeV}$ via a *booster* (PSB), an arrangement of four vertically stacked synchrotron rings of radius $25\,\text{m}$. The $628\,\text{m}$-circumference PS accelerates the
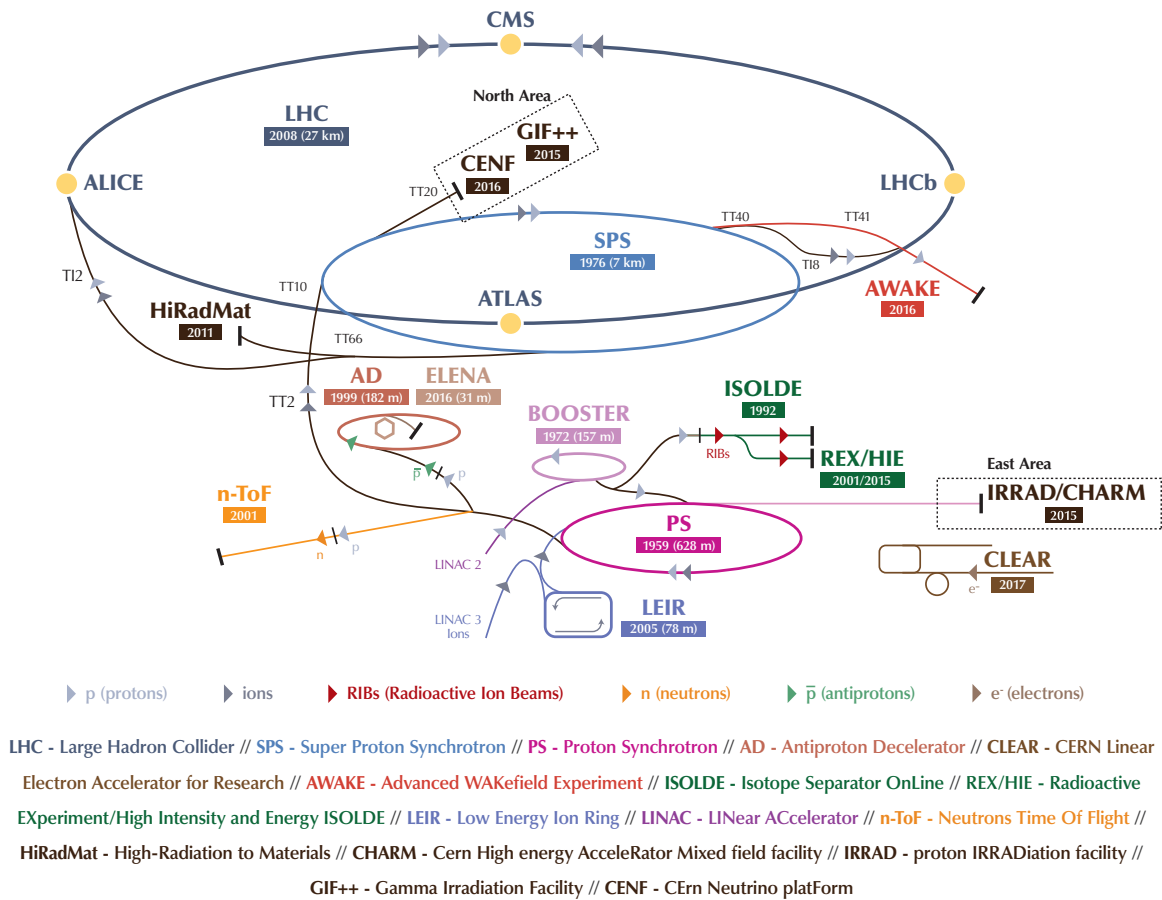
Figure 3.1 Schematic of the CERN accelerator complex. Two proton beams are injected into the LHC rings from the SPS in opposite directions via transfer lines (TI2 and TI8). From [190].

proton beam to an energy of 25 GeV before injecting them into the *Super Proton Synchrotron* (SPS), a larger synchrotron of circumference 7 km. The SPS accelerates the proton beam, arranged into 240 bunches, to an energy of 450 GeV [187, 188]. The bunches are injected into the counter-circulating LHC rings, eventually containing up to a total of 2808 bunches of approximately 1.2×10^{11} protons, spaced 25 ns apart [189]. The LHC rings are 27 km in circumference, located approximately 100 m underground near Geneva. A schematic diagram of the LHC injector chain, along with other experimental facilities at CERN, is shown in Figure 3.1.

An extensive system of electromagnets controls the beam parameters throughout the LHC machine. The most abundant create dipole fields, which constantly deflect the beams to steer

ALIGNMENT TARGET
MAIN QUADRIPOLE BUS-BARS
HEAT EXCHANGER PIPE
SUPERINSULATION
SUPERCONDUCTING COILS
BEAM PIPE
VACUUM VESSEL
BEAM SCREEN
AUXILIARY BUS-BARS
SHRINKING CYLINDER / HE I-VESSEL
THERMAL SHIELD (55 to 75K)
NON-MAGNETIC COLLARS
IRON YOKE (COLD MASS, 1.9K)
DIPOLE BUS-BARS
SUPPORT POST
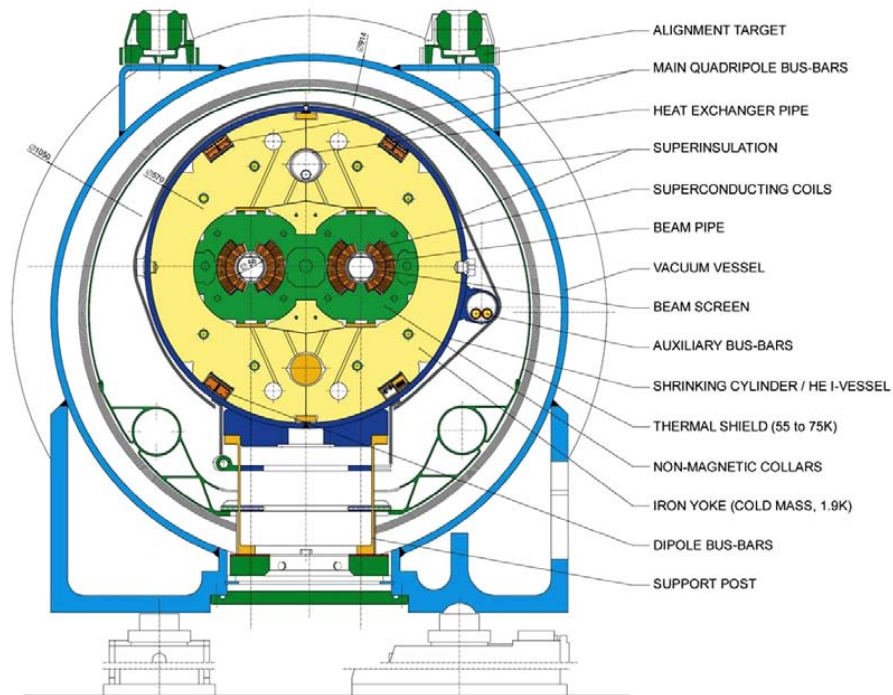
Figure 3.2 Cross section of an LHC dipole. Two magnetic dipole fields are generated by the superconducting solenoids within the cold mass. Lengths are in mm. From [186].

them around each ring. Since the beams have the same charge but are counter-circulating, the field intersecting one beam must oppose the other. This is achieved with coils placed around the beampipes, as shown in Figure 3.2. They produce a dipole field of strength 8.3 T, generated by a current of 11 850 A. This current flows through copper-clad superconducting Nb–Ti wires embedded in a 1.9 K liquid helium circuit. 1 232 units of dipole magnets are placed around the LHC ring, each 15 m in length, and a cryogenics system handles 120 t of helium [188, 191, 192].

Magnetic fields with more than two poles do not change the beam momentum. Instead, they are required in order to control the shape and dispersion of the beams. 392 quadrupole magnets lie around the LHC, arranged in pairs of focussing and defocussing units along the transverse beam axes. Sextupole, octopole, and decapole magnet systems adorn the ends of the dipole units to correct for imperfections in the steering fields fields and provide high-order corrections to the beam [191, 193].

The LHC accelerates the beams with two independent banks of linear accelerating radiofrequency cavities operating at 400.8 MHz. Each system contains eight single-cell superconducting cavities, arranged into two cryomodules cooled by liquid helium at 4.5 K. The accelerating field produced is $5.5\,\mathrm{MV\,m^{-1}}$ [194]. Although the design energy of the LHC is 7 TeV per beam, the data analysed in this thesis (during LHC Run 2) was collected using beams of energy 6.5 TeV, giving a centre-of-mass collision energy of $\sqrt{s} = 13\,\mathrm{TeV}$. This collision energy was chosen to maximise the energy frontier and the potential for experiments to observe new phenomena both within and beyond the Standard Model, while fitting within engineering constraints.

The two counter-circulating beams in the LHC are made to intersect at four points around the ring. These interaction points are surrounded by detector equipment, comprising separate particle physics experiments. ATLAS [195] and CMS [185] are general-purpose particle detectors, whereas LHCb [196] and ALICE's [197] designs are informed by their particular physics programmes. The data analysed in Chapter 4 were collected with ATLAS, and its design is discussed in detail in Section 3.2.

As the proton beams approach the interaction points, they are 'squeezed' by specialised electromagnets to greatly increase the instantaneous collision luminosity, given by [198]

$$L = \frac{n_b f N^2 \gamma}{4\pi\varepsilon\beta^*}F,\tag{3.1}$$

where n_b is the number of colliding bunches containing N protons, f is the revolution frequency, γ is the beams' relativistic γ-factor, ε is the normalised emittance determined by the injection system, and F is a geometric factor which accounts for effects from the bunch length and crossing angle. Here the amplitude β^* is a measure of the size of the beam at the interaction point, as determined by the LHC magnets nearby. Using $\beta^* = 30\,\mathrm{cm}$, the instantaneous collision luminosity delivered to ATLAS by the LHC is approximately $2 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ [199].

A centralised 25 ns-cycle clock system synchronises the LHC machine and detector readout and trigger systems [200], detailed in Section 3.2.5 for ATLAS. The LHC facility also contains infrastructure for vacuum and beam instrumentation systems. The complex of accelerators is centrally managed, independently of the experimental detectors, with the aim of creating
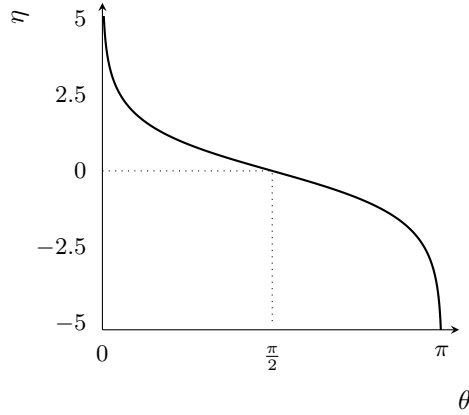
Figure 3.3 Pseudorapidity as a function of the polar angle, $\eta = -\log\tan\frac{\theta}{2}$.

the conditions desired for measurement. The first pp collisions (with $\sqrt{s} = 0.9\,\mathrm{TeV}$) were successfully delivered to ATLAS in November 2009 [201, 202].

3.2 ATLAS

The ATLAS experiment [195] is situated at one of the interaction points around the LHC ring, where the beams are steered to produce pp collisions. A local right-handed coordinate system is defined with its origin at the interaction point. The x axis points towards the centre of the LHC, the y axis upwards towards the surface, and the z axis along the beamline. An angular system is also used, with the radial coordinate r being the perpendicular distance from the beamline and ϕ the azimuthal angle in the *x–y transverse plane*, with $\phi = 0$ pointing along the x axis. The polar angle θ is measured with $\theta = 0$ aligned with the positive z axis. Often the *pseudorapidity* is used, defined by $\eta = -\log\tan\frac{\theta}{2}$, shown in Figure 3.3. A Euclidean distance metric $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is used in the η–ϕ plane, mapped onto the surfaces of cylindrical detectors.

ATLAS consists of integrated systems of magnets, detectors, triggers, and data acquisition. The main hardware is arranged in an approximately cylindrical barrel, measuring $25\,\mathrm{m}$ in diameter and $46\,\mathrm{m}$ long [203]. Four electromagnet systems provide strong magnetic field coverage across the entire detector body. Near to the interaction point, the *inner detector*

provides position and momentum information for charged particles emerging from collisions. The *electromagnetic* and *hadronic calorimeters* are placed at a larger radius, serving to measure the energy of particles. The *muon spectrometer* takes up most of the volume of ATLAS, forming the outermost layer of the detector equipment. In total, the detector and magnet hardware has a mass of approximately 7000 t. An illustrated overview of the whole ATLAS detector is shown in Figure 3.4. Each of these systems is discussed in detail in the following subsections.

3.2.1 Magnets

Four electromagnetic systems provide strong magnetic fields throughout the detector body. These fields exert a Lorentz force, $\mathbf{f} = q\mathbf{v} \times \mathbf{B}$, on particles with electric $U(1)$ charge q. By definition, $\mathbf{f} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}$ and the particle momentum is related to its velocity by $\mathbf{p} = \gamma(v)m_0\mathbf{v}$. The solution to this equation of motion determines the trajectory followed by the particle. In a linear magnetic field, for example, charged particles follow a helical path. The momentum per unit charge of the particle can be determined from the radius of curvature of the trajectory, and the sign of its charge can be found from the direction of the deflection due to the magnetic field. Therefore, the magnets in ATLAS provide vital conditions for the precise measurement of the momentum of charged particles in the inner detector and muon spectrometer. The geometry of the magnet systems is illustrated in Figure 3.5.

A solenoid coil [205] is wrapped around the outside of the inner detector, aligned along the longitudinal z axis, with a radius of 1.235 m. Superconducting NbTi/Cu wire carries a current of 7600 A, producing an approximately uniform axial field with flux density 2 T. The solenoid was designed to be thin (coil thickness 45 mm, 174 mm including infrastructure) and light (cold mass 5.7 t), so as to be almost transparent to particles travelling through the detector and to reduce the required mechanical support. This reduces degradation in the performance of the calorimeters due to the material at smaller radius.

A system of large electromagnets produce a toroidal magnetic field in the outer regions of ATLAS. In the central region ($-1.35 < \eta < 1.35$), the *barrel toroid* [206] contains eight NbTi/Cu superconducting coils carrying 20.5 kA to produce a 0.6 T average magnetic field,

Muon spectrometer   Tile calorimeter   LAr calorimeter

Toroid magnets   Solenoid magnet   SCT   Pixel detector   TRT
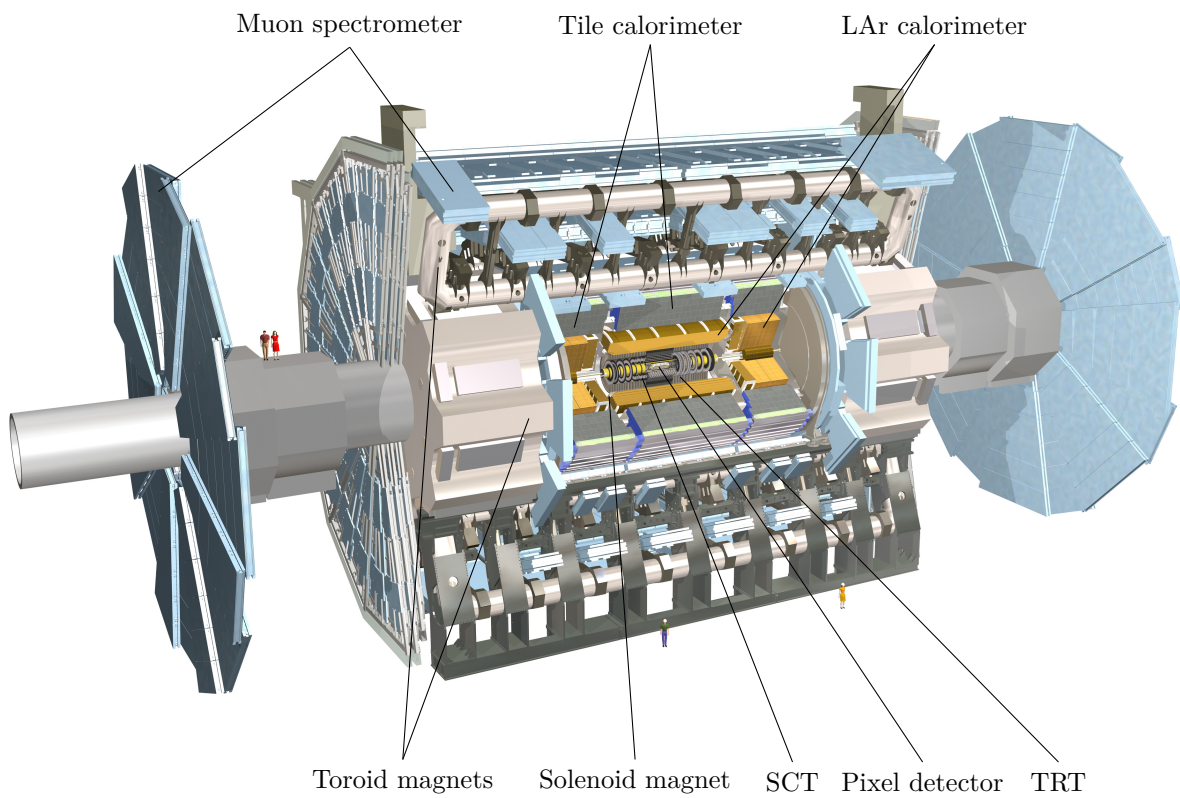
Figure 3.4 Illustrated overview of the ATLAS detector with cutaway. Colours are false and people are added for scale. From [204] with added labels.
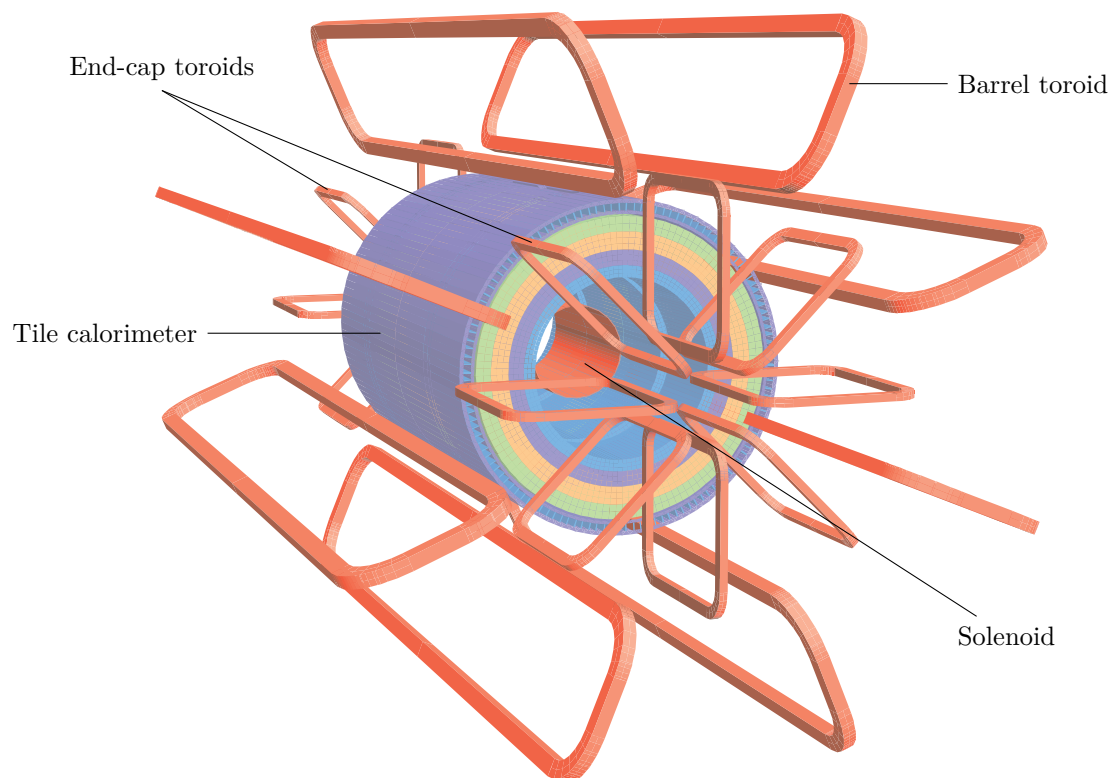
Figure 3.5 Geometry of the magnet systems in ATLAS. Magnet windings are shown in red. Also shown are the layers of the hadronic calorimeter and return yoke steel, in other colours. The eight barrel toroid coils and the end-cap toroid coils lie outside the calorimeter. The cylindrical solenoid coil is inside the calorimeter. From [195] with added labels.

peaking at 3.9 T. *End-cap toroid* magnets [207] cover the forward regions ($1.55 < |\eta| < 2.70$), producing a similar field. The toroid magnets in ATLAS produce a large magnetic field in the muon spectrometer, which makes up most of the volume of the detector. This size allows for a state-of-the-art transverse momentum resolution for charged particles travelling through the outer detector region, as detailed in Section 3.2.4.

3.2.2 Inner detector

The inner detector [208] enables the tracking of charged particles close to the interaction point. It is composed of a series of three independent but complementary subdetectors: the *pixel detector*, *semiconductor tracker*, and *transition radiation tracker*. Each is described below. It covers the central region of ATLAS, in the pseudorapidity range $-2.5 < \eta < 2.5$. The positional resolution of the subdetectors increases with radius. The design of the inner detector includes considerations to make it as transparent as possible, so as to reduce multiple scattering and to allow particles to travel to the outer detectors unimpeded. It also must withstand the high dose of radiation it will be exposed to over its lifetime. The layout of the ATLAS inner detector is shown in Figure 3.6

Pixel detector

The pixel detector [210, 211] is made up of four concentric cylindrical layers, the closest having a radius of 33.25 mm. Each layer contains a grid of reverse-biased diodes. When a charged particle travels through the doped silicon, it forms electron–hole pairs and a current flows. This is detected by the electronic front-end readout system. For the outer three layers, 90% of pixels measure $50\,\mu\text{m} \times 400\,\mu\text{m}$, and the intrinsic position measurement resolution is $10\,\mu\text{m} \times 115\,\mu\text{m}$. The innermost layer, the *insertable B-layer* [211], was added in 2014 to improve the resolution of primary and secondary decay vertices to approximately $10\,\mu\text{m}$ [212]. This layer has an intrinsic resolution of $8\,\mu\text{m} \times 40\,\mu\text{m}$. In addition, there are five disks of sensors on each end of the detector. In all layers, the pixel detector contains a total of 1.4×10^8 readout channels and covers an active area of $2.2\,\text{m}^2$.

End-cap SCT

Barrel SCT

Pixel detector

Barrel TRT

End-cap TRT
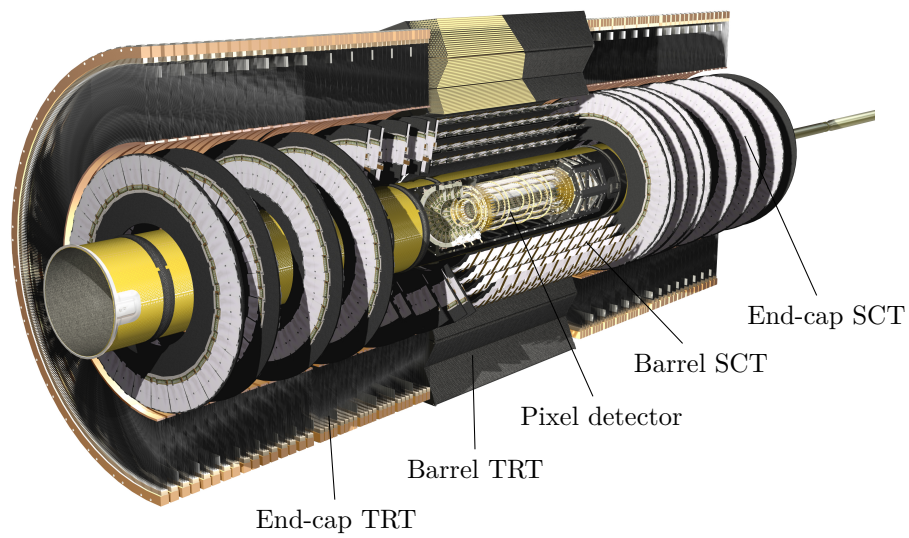
Figure 3.6 Illustration of the ATLAS inner detector, containing the pixel detector, semiconductor tracker (SCT), and transition radiation tracker (TRT). The *insertable B-layer*, installed in 2014, is not shown. From [209] with added labels.

Semiconductor tracker

The semiconductor tracker (SCT) [208] is also based on silicon sensors. In contrast to the pixel detector, however, it uses 128 mm long strips of sensors rather than grids. This is because four the SCT barrel detector are placed at a larger radii (300, 373, 447 and 520 mm) and therefore must over a much larger active area of $61\,\text{m}^2$. Nine wheels split into three rings in the forward regions extend the coverage of the SCT to $-2.5 < \eta < 2.5$. SCT modules are arranged in pairs with a 40 mrad relative angle, allowing two-dimensional position measurements. The intrinsic resolution of the SCT is 17 µm.

Transition radiation tracker

The outermost component of the inner detector is the transition radiation tracker (TRT) [208]. It consists of approximately 370 000 straw tubes filled with a $Xe/CO_2/O_2$ gas mixture. Each tube acts as a capacitor, with a central gold-plated tungsten wire anode and the outer tube acting as a cathode. When a charged particle crosses the tube, it ionises the gas leading to an electron avalanche and a detectable current in the wire. The straws are arranged into a barrel section containing about 50 000 straw tubes, spaced horizontally at distances from 560 to 1070 mm from the interaction point, and two end-caps each of 18 wheels containing straws aligned in the radial direction. Each readout channel provides a drift time measurement, giving a spatial resolution of 170 µm per straw. About 30 to 40 hits per particle are detected by the TRT. Combining the measurements from many straws, a resolution of 50 µm is achieved. Additionally, the straw tubes contain material to stimulate transition radiation (keV photons). The probability of radiation emission is proportional to the particle's relativistic γ factor, so the signal is different for particles of different rest mass. In this way, the TRT can be used to provide information for limited particle identification.

3.2.3 Calorimeters

The purpose of the ATLAS calorimetry system is to measure the energy, position, and direction of particles emerging from the primary collision point. ATLAS's system is divided into an *electromagnetic calorimeter* and *hadronic calorimeter*, which absorb and measure the energies of γ and e, and hadrons, respectively. In contrast to the detector components at smaller radii, the calorimeters are designed to have a high probability of interaction with particles emerging from collisions. As such, nearly every Standard Model is absorbed in the calorimeters, except for the minimally ionising μ and weakly interacting neutrinos. Neutral particles that do not induce any signal in the inner detector, such as photons and neutral hadrons, appear in the calorimeters and can therefore be identified.

A high-energy incident particle undergoes multiple scattering, with mean free path λ between inelastic collisions with nuclei in the calorimeter material. This produces a *shower* of lower energy particles travelling through the calorimeter, which share the total energy of the incident particle. Particles are also subject to electromagnetic interactions with the surrounding material. Over one radiation length X_0, a particle's energy becomes e^{-1} times the initial, on average. Summing the total radiation produced by a shower constitutes a measurement of the incident particle energy.

ATLAS uses *sampling calorimeters*, made up of alternating layers of dense and scintillating material. Showers are produced in the dense material and the secondary particles induce measurable signals in the scintillator.

Electromagnetic calorimeter

The electromagnetic calorimeter system uses liquid argon as the active scintillating medium. This was chosen for its stable, linear response to long-term radiation exposure [213]. The liquid argon components of the ATLAS calorimeters are highlighted in Figure 3.7. Electrodes and grounded lead absorber plates are placed placed in an 'accordion' shape, providing complete

Hadronic end-cap
calorimeter

Barrel electromagnetic
calorimeter

Electromagnetic end-cap
calorimeter

Forward calorimeter
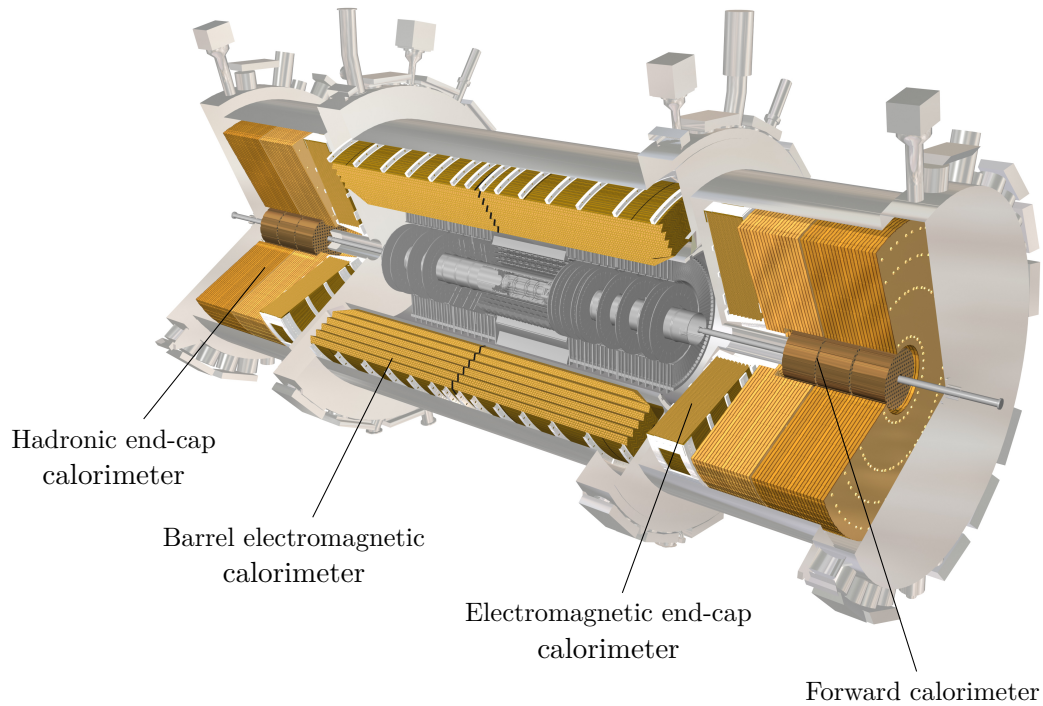
Figure 3.7 Illustration highlighting the liquid argon components of the ATLAS calorimeters. From [214] with added labels.

coverage in ϕ with no azimuthal cracks between calorimeter modules. Incident showers ionise the argon, inducing a current pulse in the electrodes.

In central barrel region of the detector, at $|\eta| < 1.475$, the electromagnetic calorimeter consists of three layers of different granularities. A segment of this layout is illustrated in Figure 3.8, where the dimensions and cell granularities are also given. The first layer has fine granularity in η. Combined with the square-prism second layer, which makes up the majority of the electromagnetic calorimeter, the detector is able to locate the origin of neutral particles, which do not leave any trace in the inner detector. The outermost layer has the largest granularity, and mainly serves to discriminate between electromagnetic and hadronic showers and to measure the energy loss after the electromagnetic calorimeter. The region with $|\eta| < 1.8$ is preceded by a thin liquid argon pre-sampler with granularity $(\Delta\eta \times \Delta\phi) = (0.025 \times 0.1)$. This serves to estimate the energy lost before the calorimeter.
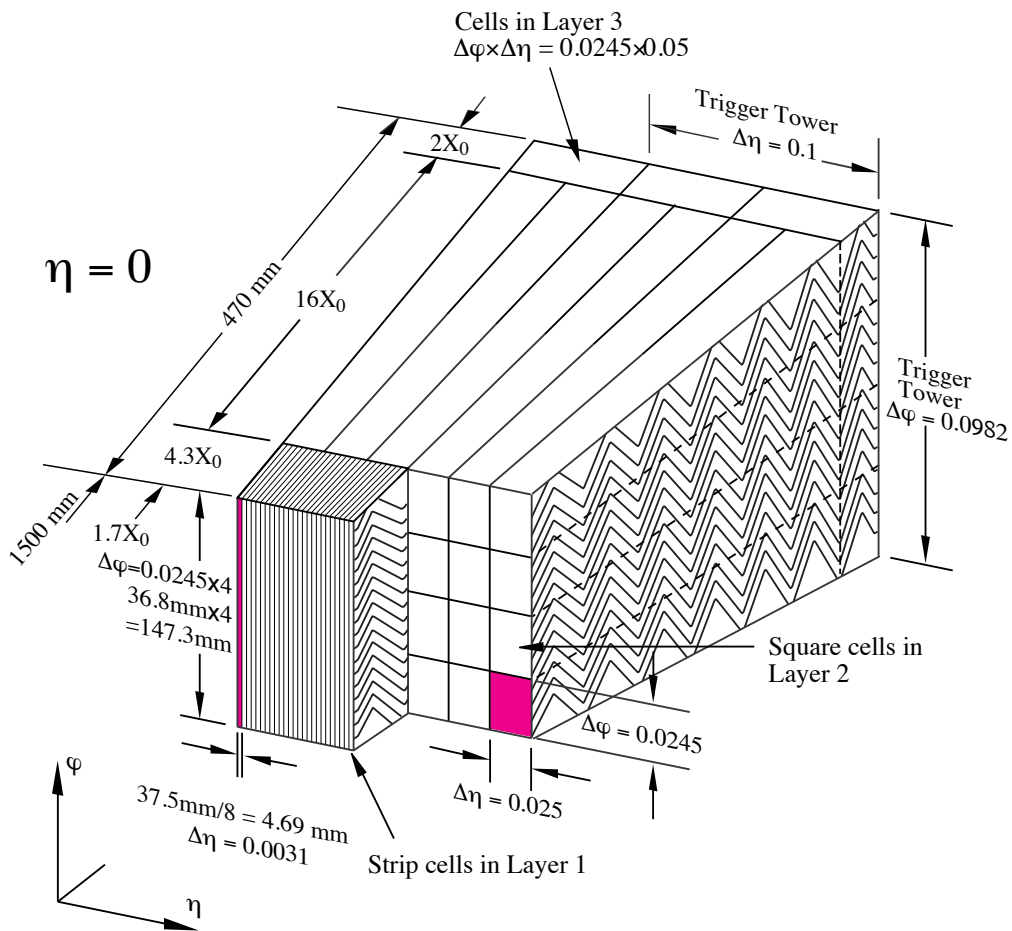
Figure 3.8 Drawing showing a section of the barrel electromagnetic calorimeter. The liquid argon cells are arranged into three distinct layers, with granularity decreasing for larger radius. From [195].

Two endcap electromagnetic calorimeters are placed in the forward regions [213]. Each consists of two coaxial *wheels*. The inner wheel covers $1.375 < |\eta| < 2.5$, and is constructed similarly to the barrel electromagnetic calorimeter, shown in Figure 3.8. The outer wheel covers $2.5 < |\eta| < 3.2$ and has a larger granularity.

Hadronic calorimeter

Heavier hadrons penetrate the electromagnetic calorimeter and enter the hadronic calorimeter system. In the barrel region sits the *tile calorimeter* at $|\eta| < 1.7$, which uses steel absorbing plates and plastic scintillating tiles [215]. Light is produced by hadronic interactions with the steel and diffused in the tiles then directed along wavelength-shifting fibres to photomultiplier tubes. The tile calorimeter is divided into a central barrel covering $|\eta| < 1.0$, and two extended barrels covering $0.8 < |\eta| < 1.7$. Each barrel is made up of 64 trapezoidal modules covering $\Delta\phi \approx 0.1$. A drawing of one such module of the tile calorimeter is shown in Figure 3.9.

Hadronic end-cap calorimeters cover the forward regions at $1.5 < |\eta| < 3.2$, and are liquid argon sampling calorimeters [213]. They use copper plates as the shower-inducing material, with width 25 mm in the inner wheels, and 50 mm in the outer wheels. These plates are aligned perpendicular to the beam axis. The hadronic end-cap calorimeters are placed directly behind the electromagnetic end-cap calorimeters, as shown in Figure 3.7.

Forward calorimeter

Sampling calorimeters cover the extreme forward regions at $3.1 < |\eta| < 4.9$, designed to serve as both electromagnetic and hadronic calorimeters [213]. Copper rods lie parallel to the beam axis in the innermost layer, producing showers for the electromagnetic portion of the calorimeter. The showers travel through 250 mm of liquid argon in a cylindrical tube, with granularity $(\Delta\eta \times \Delta\phi) = (0.1 \times 0.1)$. This is surrounded by the hadronic part of the forward calorimeter, which uses tungsten as the absorber. Tungsten rods are placed in a matrix with two sampling gaps of 275 and 500 mm, filled with liquid argon. The hadronic part is segmented by $(\Delta\eta \times \Delta\phi) = (0.2 \times 0.2)$.
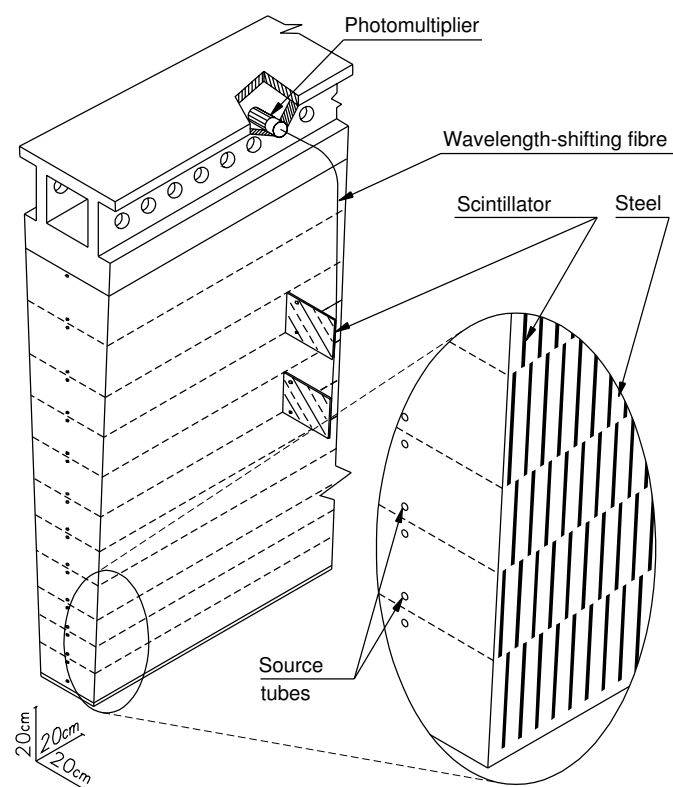
Figure 3.9 Drawing showing a section of the tile calorimeter. The steel and scintillator are layered together, with a wavelength-shifting fibre carrying signals from the scintillator to a photomultiplier located at the end of each unit. From [195].

(a)   **(b)**
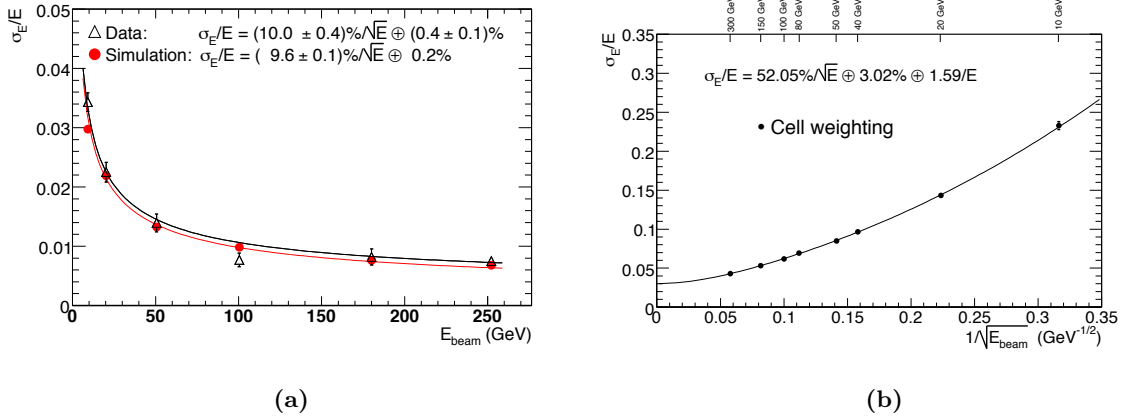
Figure 3.10 (a) Combined electron and pion test beam performance for a barrel electromagnetic calorimeter module. (b) Combined liquid argon and tile calorimeter response in a pion test beam. From [195].

Calorimeter performance

The analysis in Chapter 4 uses measurements from the ATLAS calorimetry system to measure the energy of hadronic jets. In order to accurately reconstruct the properties of the $t\bar{t}$ system, the jet energy must be precisely known. Also, the precise detection of the position and energy of b jets is essential for tagging, and thus identifying the $t\bar{t}$ production event.

Test beam experiments were used to measure the performance of the calorimeter. For the barrel electromagnetic calorimeter, combined electron and pion test beam measurements are shown in Figure 3.10a. After noise subtraction, the energy response is fit to [216]

$$\frac{\sigma(E)}{E} = \frac{(10.1 \pm 0.4)\% \times \sqrt{\text{GeV}}}{\sqrt{E}} \oplus (0.4 \pm 0.1)\%. \tag{3.2}$$

The energy response to hadrons in the barrel calorimeter was assessed using pion test beams incident on a combined module of liquid argon electromagnetic and hadronic tile calorimeters. The fractional energy resolution is shown in Figure 3.10b, and fit to [217]

$$\frac{\sigma(E)}{E} = \frac{1.6\,\text{GeV}}{E} \oplus \frac{(52.0 \pm 1.0)\% \times \sqrt{\text{GeV}}}{\sqrt{E}} \oplus (3.0 \pm 0.1)\%. \tag{3.3}$$

Thin gap
chambers (TGC)

Resistive plate
chambers (RPC)

Cathode strip
chambers (CSC)

Monitored drift
tubes (MDT)
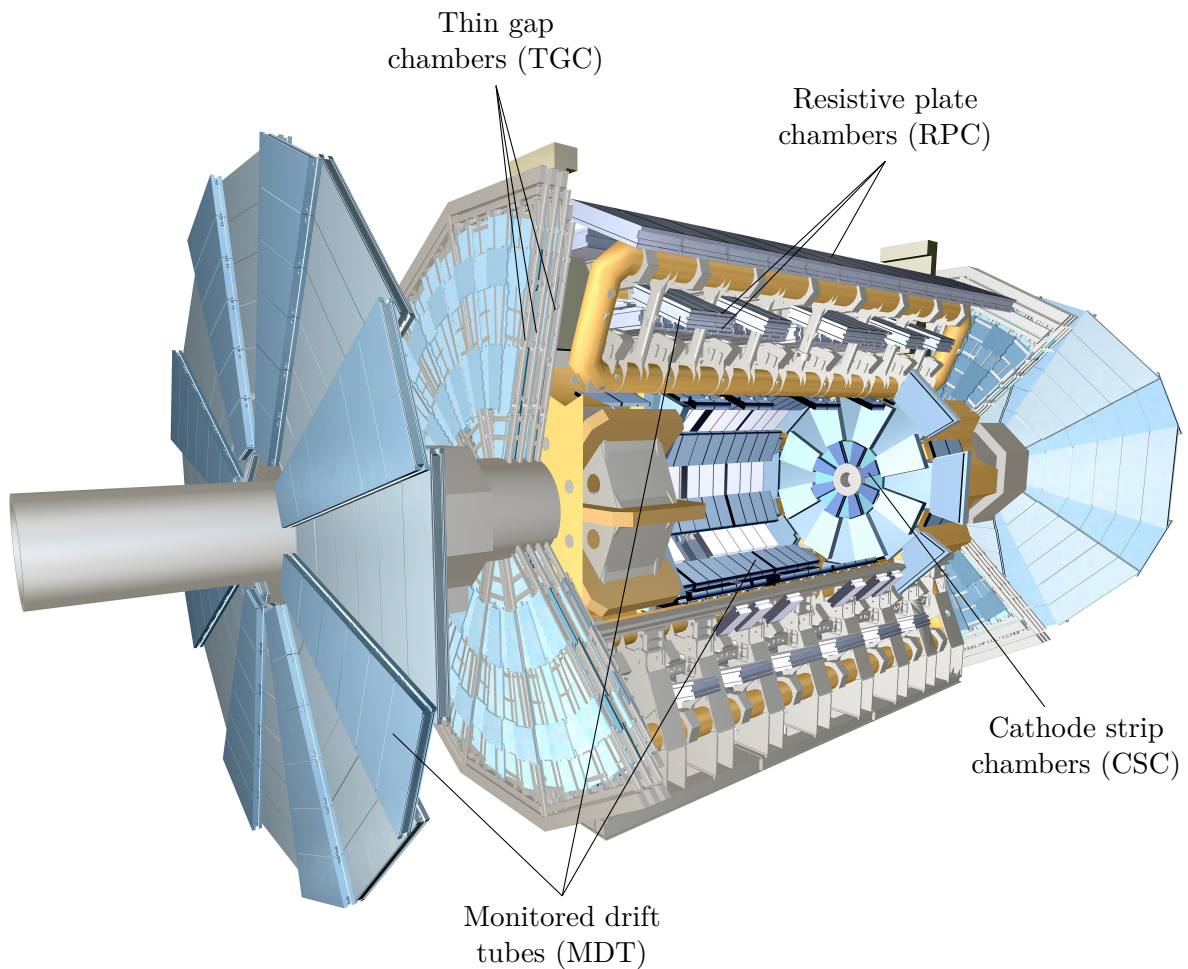
Figure 3.11 Illustration of the ATLAS muon spectrometer components. The toroid magnet coils are also shown in yellow. From [219] with added labels.

3.2.4 Muon spectrometer

The outermost portion, and most of the volume, of the ATLAS experiment contains the muon spectrometer [218]. The dense material in the calorimeters absorbs most of the particles emerging from LHC collisions at the centre of the detector, so only minimally ionising μ leptons and invisible particles reach the muon spectrometer. This detector's role is to detect the presence of μ particles and measure their position and momentum. The layout and components of the muon spectrometer are shown in Figure 3.11.

μ leptons leave at least three hits in the detector, and are matched to tracks left in the ATLAS inner detector. The muon spectrometer is immersed in the magnetic field supplied by ATLAS's toroid magnet system. This allows the μ transverse momentum to be measured via the *sagitta s* [220] of the curved trajectory, by

$$\frac{p_{\mathrm{T}}}{q} = \frac{L^2 B}{8s},$$ (3.4)

where L is the length of the path in an approximately constant magnetic field B, and q is the electric charge of the particle.

270 000 aluminium *monitored drift tubes* (MDTs) with 30 mm diameter form the bulk of the p_{T} measurement apparatus in the muon spectrometer. The tubes are filled with a $97 : 3$ $\mathrm{Ar/CO_2}$ gas mix held at 3 bar, with a W-Re anode wire through the centre. The mode of operation is similar to the TRT, described above. The MDT intrinsic spatial resolution is 80 μm, but this is improved to an average of 35 μm in the z direction by multiple layers of tubes per chamber providing multiple hits. MDTs cover 99.5% of the active area of the detector, equal to 5500 m². The remainder contains *cathode strip chambers* (CSCs), placed in the high-flux region at $2.0 < |\eta| < 2.7$. These are multi-wire proportional chambers, with radial anode wires and cathode strips placed in orthogonal planes in order to provide radial and transverse measurements. The four CSC layers achieve a combined resolution of 40 μm in the η direction and 4 mm in the ϕ direction. Together, MDTs and and CSCs provide state-of-the-art precision measurements of μ lepton p_{T}.

The event selection for the analysis in Chapter 4 rejects any collision events containing μ leptons. The muon spectrometer is used to identify those events containing such particles. In the barrel region, where $|\eta| < 1.05$, 596 resistive plate chambers (RPC) provide 10 mm resolution in both the z and ϕ directions. Importantly, the RPCs have a short response time of < 25 ns. They are made up of two parallel resistive plates held at a 9.8 kV potential difference provides a $4.9\,\mathrm{kV\,mm^{-1}}$ constant electric field. The gas between the plates is ionised by passing μ leptons, inducing an avalanche which cause current spikes in matrices of aluminium strips on the back of the resistive plates. In the endcap at $1.05 < |\eta| < 2.4$, *thin gap chambers* (TGCs) are

able to withstand the exposure to a higher flux, while maintaining a $< 25\,\text{ns}$ temporal resolution. They are multi-wire proportional chambers, similar to the CSCs albeit with more distance between the anode wires and cathode cases. The wires are arranged parallel to the MDTs and orthogonal to the readout strips, providing a resolution of 2 to $6\,\text{mm}$ in the r direction, and 3 to $7\,\text{mm}$ in the ϕ direction. The RPC and TGC trigger chambers provide measurements for timing bunch crossings and providing information to the trigger for rapid decision making.

3.2.5 Trigger and data acquisition

The LHC provided *pp* collisions to ATLAS in 2015 and 2016 at a centre-of-mass energy of $13\,\text{TeV}$ and a bunch spacing of $25\,\text{ns}$. This means there is an unfiltered event rate of up to $40\,\text{MHz}$. This is too large to fully read out the complete detector state for every event for processing and storage, since the upstream processing system does not provide adequate bandwidth or capacity. The ATLAS *trigger* system makes a decision on whether to full read out an event. It targets a readout rate of approximately $1\,\text{kHz}$ to storage, equivalent to an overall rejection factor of 40×10^3 against minimum bias events, while maintaining high efficiency for events of interest for the ATLAS physics research programme. To do this, it uses a cascading chain of decision making algorithms in hardware and software. In parallel, the data acquisition system buffers the data from the various subdetector readout systems and controls the data flow when the trigger decision is received. These systems are detailed here. An overview of the components of the ATLAS trigger and data acquisition system is shown in Figure 3.12.

Level 1 trigger

The *Level 1* trigger [222] is an ASIC/FPGA hardware-based trigger system implemented in fast electronics. It takes input from reduced-granularity portions of the calorimeters and muon spectrometer. It targets an event acceptance rate of at most $100\,\text{kHz}$, to be passed to the readout system for the High Level Trigger to process.

For the calorimeters, special *trigger towers* with typical granularity $(\Delta\eta \times \Delta\phi) = (0.1 \times 0.1)$ supply data from the electromagnetic and hadronic calorimeter systems. The towers are formed
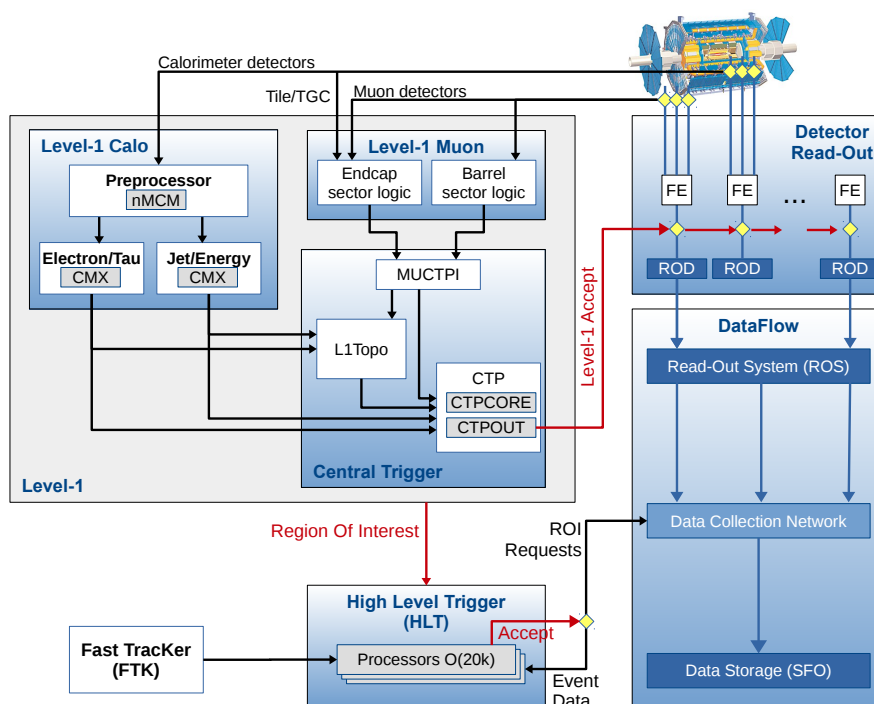
Figure 3.12 Functional overview of the ATLAS trigger and data acquisition system. Arrows indicate the direction of data flow. From [221].

by aggregating constituent cells in the calorimeter front-end analogue electronics, and provide approximately 7200 inputs to the calorimeter trigger system. Here the signals are digitised and a series of programmable selections filter the event signature for e, γ, hadronically-decaying τ, and jets. Selections can also be applied for the total calorimeter energy and missing transverse energy. The surviving object count for each threshold is passed to the *central trigger processor* (CTP), described below.

In the muon spectrometer, the RPC and TGC front-end electronics amplify, shape, and discriminate signal patterns. There are approximately 8×10^5 inputs to the muon trigger system. The time resolution of the muon spectrometer trigger chambers are designed such that the bunch crossing from which a signal originates can be determined from time-of-flight. Six programmable thresholds are applied and the resulting multiplicities for each bunch crossing are forwarded to the CTP.

A *topological trigger* system [223] receives input from the calorimeter and muon trigger systems in the form of trigger objects. It applies up to 128 algorithmic cuts based on the location of the objects in the detector, thereby discriminating event topologies and shapes. The surviving 128-bit trigger objects encoding energy and resolution are passed to the CTP. The topological trigger system operates with a latency < 200 ns and has complete coverage in η and ϕ.

The CTP combines the information provided from the systems described above and computes the Level 1 accept/reject decision. It can be programmed with up to 96 *trigger menu items* which define a series of event selection rules. If an event survives any one of these items, a 'Level 1 accept' signal is passed to the readout system. The CTP can also automatically reject events due to *prescale* or *deadtime* requirements. Each trigger menu item can be prescaled so as to dampen the acceptance rate of the events passing the cuts. Deadtime refers to the time between Level 1 accept signals in which upstream (front-end drivers) or downstream (HLT) components of the trigger and data acquisition system are saturated, blocking the flow of data. The CTP contains systems to automatically monitor and regulate acceptance rates and deadtime.

High level trigger

The *high level trigger* (HLT) system [221, 224] is implemented in software run on approximately 80 000 x86-64 CPU cores. Following a Level 1 accept signal, the detector readout systems flush the complete detector state from buffers for each subdetector system. This is passed to the HLT, where the full granularity of the event data is available.

The event input rate from the detector to the HLT is substantially reduced by the Level 1 trigger system and the architecture of the HLT permits large-scale parallel processing. Therefore the HLT algorithms can be dramatically more complex than those implemented at Level 1, also enhanced by the flexibility of the software implementation. On average, the decision time is approximately 75 ms per event, peaking at approximately 1 s for accepted events.

Algorithms are combined into *chains* that can be applied to the event data sequentially. Early algorithms may request region-of-interest information from the Level 1 system, bypassing

the need to process the complete event data. Later algorithms can use the objects constructed by earlier ones in combination to make complex decisions based on the entire detector state. If an event survives any HLT chain, it is accepted and read out to permanent storage.

The HLT targets an event output rate of 1 kHz. The main constraints on the output rate are from downstream data storage capacity and network bandwidth of the readout and storage systems.

Trigger configuration

At all states of operation, the trigger system must be correctly configured to record collision events of interest, while limit processing load to allow adequate throughput. Furthermore, although the Level 1 and HLT systems are configured separately, they must use compatible modes of operation. The ATLAS trigger system is dynamically configured through a relational ORACLE SQL database [225], hosted on the ATLAS machine local area network.

The database contains multiple tables, related by many-to-many link tables. No records are duplicated, for efficiency, and every possible trigger configuration used to collect data with ATLAS is stored in the database. Primitive records corresponding to Level 1 items and HLT algorithms are stored in tables, which are parents to tables containing records for chains and menus and sets of prescales for the items. Ultimately, a particular trigger configuration is defined by two ID numbers: one for Level 1 and one for the HLT. During data collection, an appropriate configuration is chosen for the instantaneous luminosity supplied by the LHC. The data flow rates at various stages of the trigger and data acquisition system are monitored during operation in the ATLAS control room.

Readout system and data processing

The ATLAS readout system [226] serves to buffer data from the detector front-end electronics for events passing a Level 1 accept while the HLT performs a more detailed decision-making process. During the HLT operation, the readout system handles requests from the HLT for data

fragments, and ultimately either passes the data downstream for permanent storage or deletes it, depending on the HLT decision. This is implemented with approximately 100 PC server units each fit with four custom *RobinNP* cards [227]. The input from the detector readout drivers is passed along one of around 1850 optical links at a rate of 100 kHz, with an average fragment size of 1.5 kB. Each card contains 8 GB of DDR3 random access memory to buffer the event data.

The target event acceptance rate of 1 kHz corresponds to an output data transfer rate of approximately 1.5 GB/s from the ATLAS data acquisition system [227]. This is sent over a network to permanent digital storage at CERN. There is it distributed across the *Worldwide LHC Computing Grid* (WLCG) [228], a globally-distributed network of storage and compute clusters which is shared with the other LHC experiments. The clusters are hosted at over 130 ATLAS Collaboration member universities and scientific institutions [229].

Raw ATLAS data are processed on the WLCG to extract information useful for physics analyses and reduce the size of files used. The full information of the detector readout summarised and compressed into representations of physics objects, such as the jets used in the analysis in Chapter 4. This is done with ATLAS offline software [230], which applies a framework of reduction and skimming algorithms to store the transformed data in xAOD format, where it can then be used for analysis. The WLCG also provides massively-distributed computing resources for batched custom data processing. These facilities were used to produce the intermediate and final results in Chapter 4.

3.3 Physics object reconstruction

Signals in each part of the ATLAS detector are combined to form *physics objects*, corresponding to data structures which may be used in downstream data analyses. Offline software performs this reconstruction on saved data which passes trigger selections. Objects can be either photons, electrons, muons, τ jets, hadronic jets, or missing transverse energy. The reconstruction of

leptonic objects and hadronic jets are detailed in this section. Since the analysis in Chapter 4 does not use photons or missing transverse energy, their reconstruction is omitted.

3.3.1 Tracks and vertices

Tracks are used to build the physics objects in this section and also to identify vertices. They describe the trajectory of a charged particle moving in the solenoidal magnetic field through the ATLAS inner detector. Each track is described by five parameters: the transverse impact parameter d_0, the longitudinal impact parameter z_0, azimuthal angle ϕ, polar angle θ, and the ratio of charge to momentum q/p. Tracks are reconstructed from the inner detector readout using a series of sequential algorithms [231].

The data from the pixel and strip detectors are processed into *hits*, defined by three-dimensional space points. *Drift circles* in the TRT straw detectors are also defined. Clusters of three hits in the pixel detector and innermost layer of the SCT then define *track seeds*. Extending these into the outer layers of the SCT gives *track candidates*. The track candidates are fit to hits using consecutive local pattern recognition and global pattern recognition algorithms based on a Kalman filter [231, 232], making use of accurate geometry and material information. Cuts on the fit χ^2 value and subdetector track scores removes poor-quality tracks. Ambiguities are removed by using the track corresponding to the maximum-χ^2 fit. Surviving tracks are extended into the TRT and matched with compatible drift circles. Finally, the track is refit to the information from all inner detector systems simultaneously. An illustration of tracks found by this algorithm for a simulated event containing a $t\bar{t}$ decay is shown in Figures 3.13 and 3.14.

A second, outside–in algorithm is used to find track segments which are not found by the inside–out sequence described above. These may exist because no seeds are found in the silicon detectors, which is the case when the charged particle loses a large amount of energy through material interactions, or is produced by photon conversion or other in-flight decay processes within the inner detector volume. Here the track segments are formed by drift circles from the TRT not associated with tracks from the first reconstruction algorithm. They are indicated by black circles in Figure 3.14. These seeds are extrapolated inwards to hits in the inner silicon
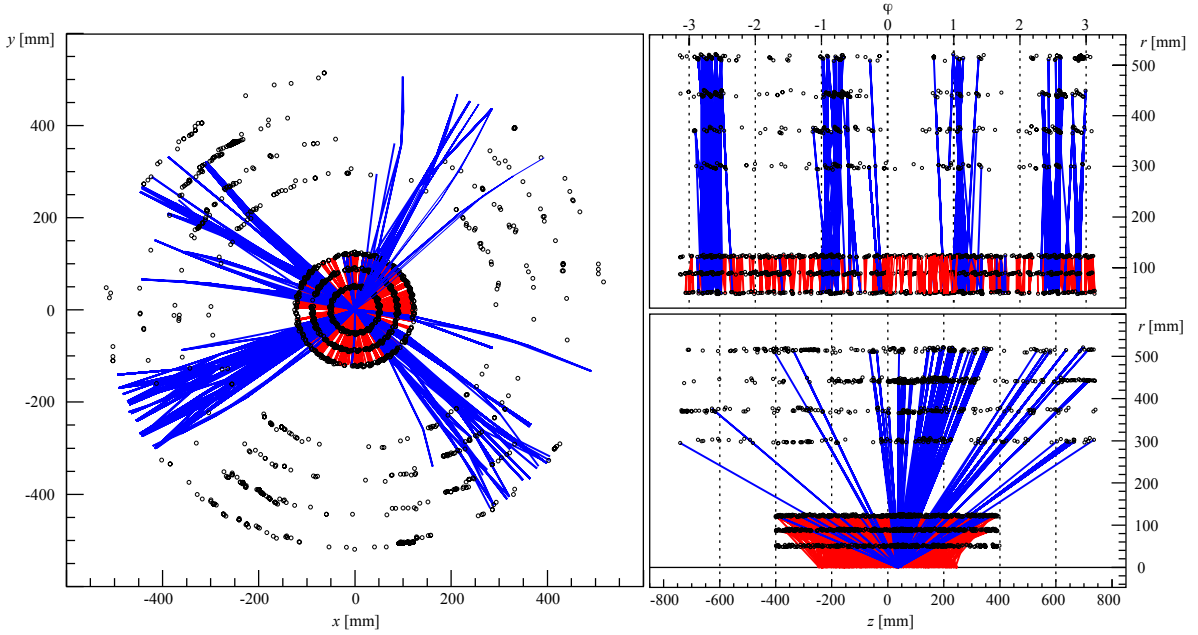
Figure 3.13 Inside–out track reconstruction for a simulated $t\bar{t}$ decay event. Hits in the inner detectors are indicated by black circles. Track seeds are shown in red, and fully reconstructed tracks are in blue. From [231].

detectors to form tracks. Again, ambiguities are removed with fit quality criteria and a complete fit is performed to extract the track parameters.

After complete tracks have been identified, they are used to determine primary and secondary vertices. Vertices are points from which particles emerge, indicating the occurrence of a physical process at that location. Since the LHC collides bunches containing 1.15×10^{11} protons [189] at high instantaneous luminosity, there are typically multiple *pp* interactions per bunch crossing. The distribution of the number of interactions per crossing is shown for the 2015+2016 data-taking period in Figure 3.15. For the dataset considered in the analysis in Chapter 4, the mean number of interactions per bunch crossing is 23.7.

Tracks are assigned to a vertex by an adaptive fitting algorithm [233], restricted to the beam spot region. The *primary vertex* is the *pp* interaction vertex emitting the most transverse momentum. The other *pp* interaction vertices in the collision event are referred to as 'pileup'. *Secondary vertices* are those where a decay has taken place subsequent to the primary *pp*
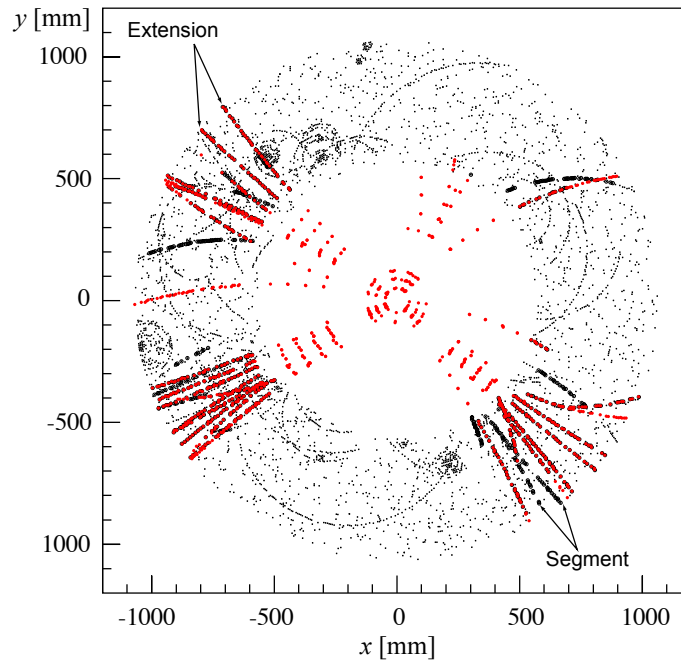
Figure 3.14 Track finding for the same simulated $t\bar{t}$ event as in Figure 3.13. Hits are indicated by small black dots in the transverse plane. Red dots show hits in associated with tracks built using the inside–out algorithm, seeded in the silicon detectors and extended into the TRT. Black circles indicate hits that form track segments in the TRT which are not found by the inside–out approach. These segments are used by the outside–in track finding algorithm. From [231].
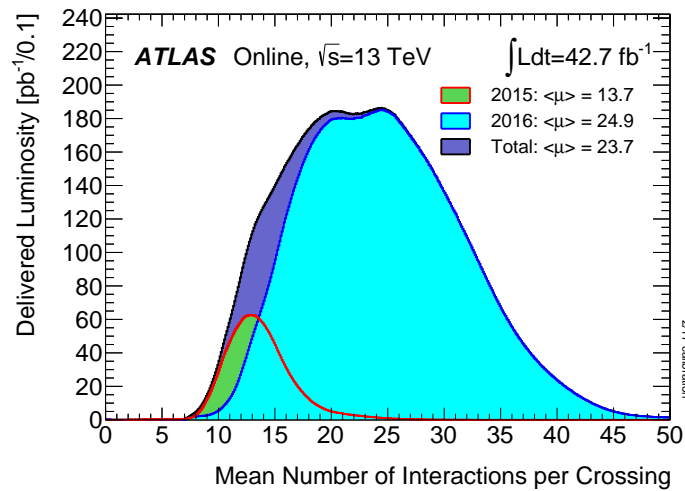
Figure 3.15 Distribution of number of interactions per pp bunch crossing in ATLAS for the 2015+2016 data-taking period.

interaction. They are characterised by large transverse impact parameters, indicating the propagation of some long-lived particle. Secondary vertices are used in *b*-tagging algorithms, detailed in Section 3.3.4 below.

3.3.2 Leptons

Electrons

Electron objects are defined by charged tracks which deposit all their energy in the electromagnetic calorimeter. Their reconstruction begins with seed clusters identified in the calorimeter, using a sliding window algorithm in η–ϕ space to identify clusters with energy above $2.5\,\text{GeV}$. These are combined with a clustering algorithm to give a total cluster energy measurement. The calorimeter clusters are matched with tracks from the inner detector. The tracks are refit with an electron hypothesis [234] to account for the 20–50% energy loss in the inner detector, due to emission of bremsstrahlung photons, and they are extended into the middle calorimeter layer in order to match to a cluster. Ambiguities in matching tracks to clusters are resolved with track quality criteria. Finally, the combined electron object quality is assessed using a multivariate likelihood to distinguish objects derived from signal e leptons from possible backgrounds such as hadronic jets or converted γ particles. The electron identification efficiency is shown as a function of the transverse energy for three likelihood working points in Figure 3.16.

Muons

Construction of muon objects uses data from the ATLAS inner detector and muon spectrometer systems. Tracks are first reconstructed independently in each detector system. In the muon spectrometer, seed track segments are found using a Hough transform [236]. Track candidates are formed using a combinatorial search over the track segment seeds. Candidates are accepted or rejected using fit and track quality criteria. Combined muon objects are formed by aligning tracks found in both the inner detector and muon spectrometer, mostly using an outside–in method. After alignment, the track is refit throughout the detector, where hits in the
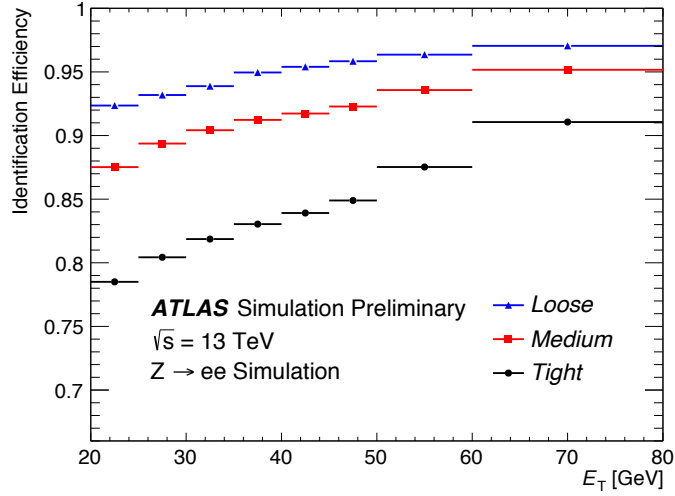
Figure 3.16 Electron identification efficiency as a function of the transverse energy in simulated $Z \to ee$ events. The analysis in Chapter 4 uses the 'Tight' likelihood threshold. From [235].

muon spectrometer can be added to or removed from the muon object. The quality of the reconstructed object is assessed using variables such as the number of hits in the subdetectors, the quality of the global track fit, and the degree of imbalance in the charge and momentum in the inner detector and muon spectrometer. The muon reconstruction efficiency is shown as a function of transverse momentum for simulated and measured $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events in Figure 3.17.

τ leptons

τ leptons decay into lighter leptons or hadrons before they reach the detector equipment, with a mean lifetime of $c\tau_\tau = 8.7\,\mu m$ [22]. Leptonic decay modes, with a branching fraction of 35.2% [22], have observable final states containing e and μ leptons. These leptonically-decaying τ leptons may be reconstructed as electron or muon objects. τ leptons decaying to hadrons are identified using clusters at the cores of jets reconstructed as described in Section 3.3.3 below. Tracking information and calorimeter shower shapes are used by a boosted decision tree to identify jets originating from τ decays. Some results from a study of the performance of this
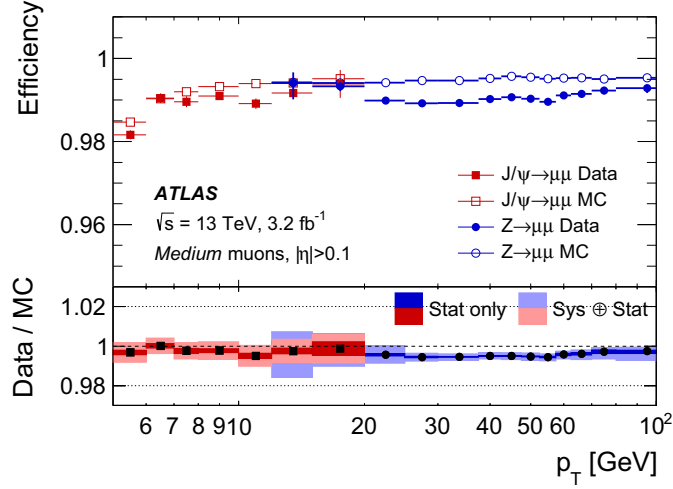
Figure 3.17 Muon reconstruction efficiency for as a function of the transverse momentum in $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events. Muon objects are defined using the 'Medium' likelihood threshold, used in the analysis in Chapter 4. From [237].

(a) **(b)**
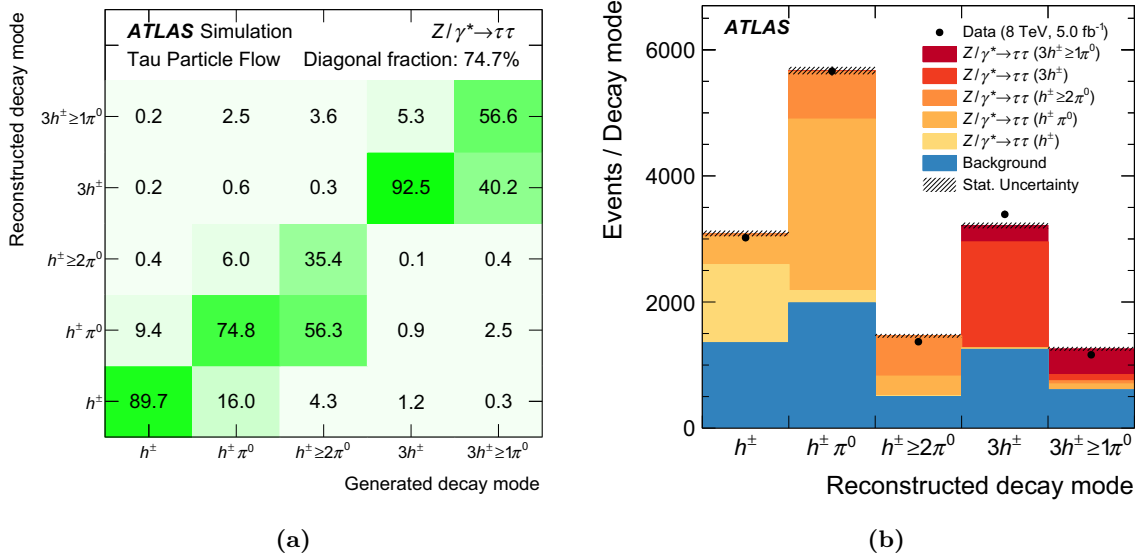
Figure 3.18 (a) Generated versus reconstructed decay modes for hadronically-decaying τ leptons; (b) Abundances of reconstructed decay modes for hadronically-decaying τ leptons in simulation and data. From [238].

reconstruction method [238] are shown in Figure 3.18. The 'Medium' working point for this discriminator is used to define hadronically-decaying τ leptons in the analysis in Chapter 4.
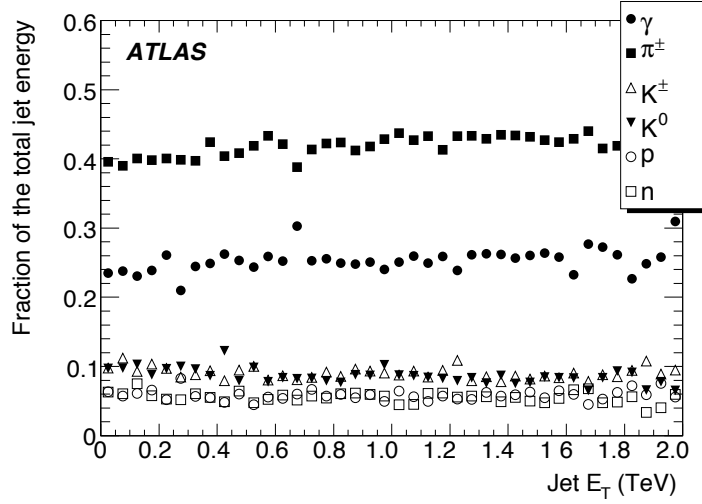
Figure 3.19 Fractional energy carried by different particle types as a function of the transverse jet energy. The jets are produced by simulation of QCD processes. From [239].

3.3.3 Jets

Sprays of hadronic activity from QCD processes result in many tracks and calorimeter deposits in the ATLAS detector. The total energy deposited relates to the energy of the original gluon or quark. Tracks and calorimeter clusters are combined or separated to define *jet* objects for analyses. Jets are the objects of interest for the analysis in Chapter 4, since the final state is fully hadronic.

The energy in a jet is distributed among many different particle types, as shown in Figure 3.19. The largest proportion of energy (approximately 40%) is carried by light charged hadrons, which leave a track in the inner detector and deposit energy in the calorimeters. Some of the energy is carried by neutral γ particles, which do not leave any track in the inner detector, but do deposit energy in the electromagnetic calorimeter. Therefore a jet is recognised by a signature of multiple tracks aligned with energy deposits in both the electromagnetic and hadronic calorimeters. Jets are constructed using a sequence of clustering algorithms, detailed in this section.

The fine segmentation of the ATLAS calorimeters allows for good jet resolution, but also admits a large amount of cell-to-cell electrical noise, in addition to noise from soft

radiation from pileup vertices. For robustness against this noise, *topological cell clusters* are first constructed [240, 241]. These start with seed cells with signal-to-noise ratio $\Gamma = E_{\text{cell}}/\sigma_{\text{noise,cell}} > S = 4$. Directly neighbouring cells are then included in the cluster. Secondary neighbours are included if $\Gamma > N = 2$. Finally, a ring of 'guard cells' with $\Gamma > P = 0$ are added. A splitting algorithm prohibits very large merged clusters by identifying local maxima [242].

These topological cell clusters form the input to the jet clustering algorithm. They are defined as massless pseudo-particles with four-momentum constructed from the calorimeter energy and direction as the energy-weighted barycentre of the cluster in (η, ϕ) space. The *anti-k_t* algorithm [243] is used to group the inputs into jet objects. It is a sequential combination algorithm which calculates the 'distance' between every pair of inputs i and j as

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta_{ij}^2}{R^2} \tag{3.5}$$

and also the distance to the beam axis,

$$d_{iB} = k_{ti}^{2p}. \tag{3.6}$$

Here k_t is the transverse momentum and $p = -1$ gives anti-k_t its name. $\Delta_{ij} = (y_i - y_j)^2 - (\phi_i - \phi_j)^2$ where y_i and ϕ_i are the rapidity and azimuthal angle of the inputs, respectively. The *jet radius parameter* R controls the characteristic size of the final jets, and is set to $R = 0.4$ for all the resolved jets in this thesis. The sequential algorithm merges entities by summing their four-momenta if $d_{ij} < d_{iB}$ for any j. If, however, $d_{ij} > d_{iB}$ for all j then entity i is declared a final jet and removed from the list of entities. The algorithm runs until this list is empty. An example illustration of the resultant jets using $R = 1.0$ is shown in Figure 3.20. The final jet object has a four vector $(E, \mathbf{p})$ that is the sum of its massless constituents. The jet mass is then defined by $m_{\text{jet}} = \sqrt{E^2 - |\mathbf{p}|^2}$.

The anti-k_t algorithm produces conical jets when they are isolated, and shares constituents between overlapping jets according to their momenta. Crucially, the clustering algorithm is *infrared safe*, meaning it is insensitive to soft QCD radiation emitted by particles in the jet, and also *collinear safe*, meaning a constituent particle may split into two particles travelling
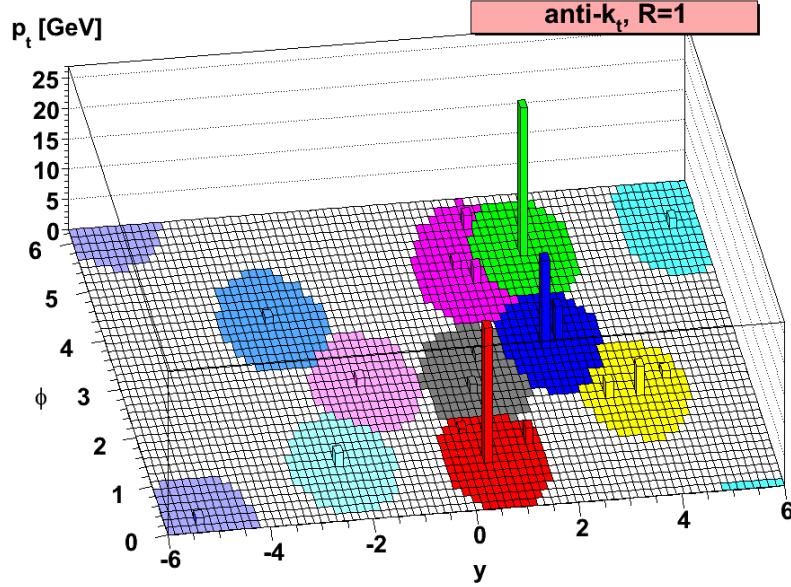
Figure 3.20 Example clustering of jets using the anti-k_t algorithm. From [243].

collinearly and this leaves the resulting jet unchanged. These properties are required for the definition of the jet to be able to be used in measurements that compare to theory. They ensure that the jet definition is not highly sensitive to increasing orders in the perturbative expansion of QCD matrix elements.

A series of corrections is applied to the resultant jet objects to calibrate them [244]:

1. Origin correction – moves the origin of the jet momentum from the centre of the detector to the primary vertex of the event. This correction improves the resolution in η by approximately 80% for jets with $p_T > 20$ GeV.

2. Pileup corrections – the jet momentum is adjusted so that the corrected p_T is given by

$$p_T^{\text{corrected}} = p_T - \rho A - \alpha(p_T, \eta)(N_{PV} - 1) - \beta(p_T, \eta)\mu. \qquad (3.7)$$

The first subtraction removes contributions from pileup with median energy density ρ, calculated in the central detector region. The correction is proportional to the jet area A [245]. Then a residual correction is applied with the functions α and β determined
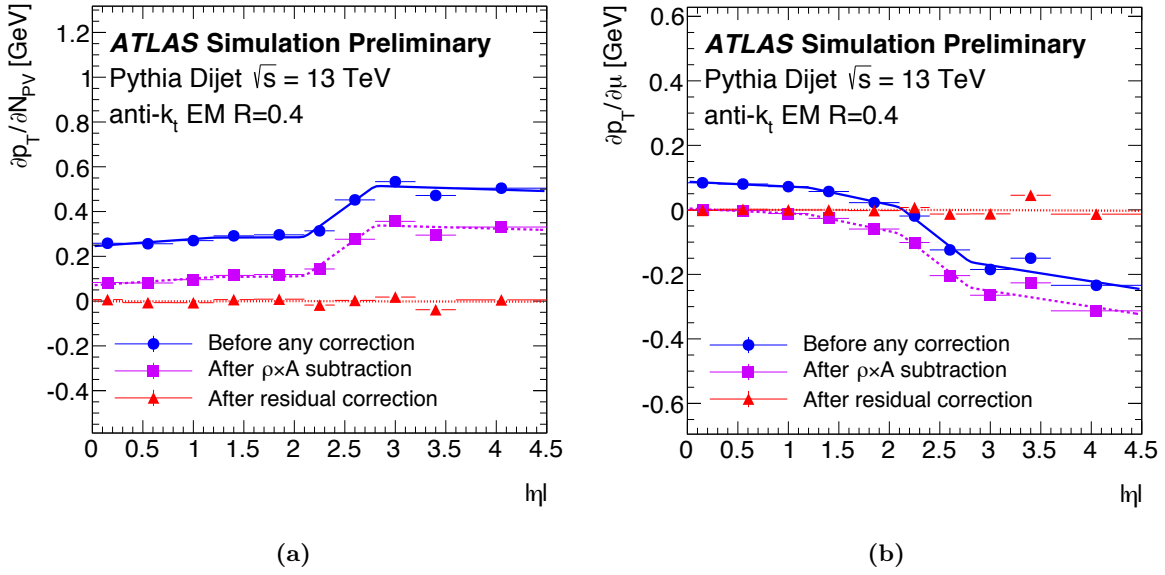
Figure 3.21 Dependence of reconstructed jet p_{T} on (a) in-time pileup and (b) out-of-time pileup in bins of jet $|\eta|$. The red curve shows the effect of of the residual pileup correction functions (a) α and (b) β in Equation 3.7. From [246].

empirically from a piecewise linear fit to simulation [246]. Here N_{PV} is the number of pp interaction vertices in that event, correcting for 'in-time' pileup collisions that occur during the same bunch crossing. For the last term, μ is the average number of pp vertices per event, providing a correction for 'out-of-time' pileup. The effects of these corrections are shown in Figure 3.21.

3. Absolute Monte Carlo-based calibration – the reconstructed jet energy and pseudorapidity are calibrated to agree with simulation by a correction in bins of simulated and reconstructed pseudorapidity. In each bin, a multiplicative correction factor is determined from the mean of a Gaussian fit to the distribution of ratio of the reconstructed to simulated jet energy. Biases in the reconstructed jet pseudorapidity due to detector layout are also corrected.

4. Global sequential calibration – biases are corrected in five variables for each jet object:

   a) Jet energy fraction deposited in the first layer of the tile calorimeter;

b) Jet energy fraction deposited in the third layer of the liquid argon electromagnetic calorimeter;

c) Number of tracks with $p_\mathrm{T} > 1\,\mathrm{GeV}$;

d) Jet radius, weighted by the p_T of constituent tracks;

e) Number of muon track segments.

5. Residual in-situ calibration – a correction to biases from Monte Carlo simulations is applied by empirical measurements to well-known real collision events. This is done by balancing dijet events in the forward and central detector regions up to jet p_T of $1.2\,\mathrm{TeV}$. Reconstructed jet momenta are also balanced with well-measured photons or reconstructed Z bosons (with $Z \to ee$ or $Z \to \mu\mu$) in the central region, up to jet p_T of $944\,\mathrm{GeV}$. Finally a correction is applied based on measurements of multi-jet events where one high-p_T jet ($300 < p_\mathrm{T} < 2000\,\mathrm{GeV}$) recoils from several low-p_T jets in the central detector region.

Despite these corrections, the *jet energy scale* contributes a significant systematic uncertainty in many analyses, including the one presented in Chapter 4. This is due to the different responses of the ATLAS calorimeters for hadronic and leptonic particles of the same energy. For hadronic jets, a per-cell weighting is applied to account for the smaller response, resulting in two possible energy scales being defined for jet reconstruction. Since hadronic objects generally penetrate further into the calorimeter material, this provides an estimate for which energy scale should be used for reconstruction.

Since jets are abundant in hadron collisions, it is advantageous to discriminate those originating from the primary event of interest from those resulting from pileup interactions. This is done in ATLAS with a *k*-nearest neighbour likelihood discriminant formed from track-based variables for each jet, resulting in a *jet vertex tagger* (JVT) variable [247]. Cutting on the JVT results in an event selection with a jet efficiency that is approximately independent of the variable number of pileup vertices in the event.

3.3.4 *b* tagging

The analysis presented in Chapter 4 is concerned with identifying the decay products of t quarks in order to measure the rate of $t\bar{t}$ production. As discussed in Sections 2.2.3 and 2.3.2, t quark decays result in a b quark and a $W^{\pm}$ boson. The b quark undergoes the hadronisation process to form a B hadron, which is relatively long-lived since the CKM coupling of the third quark generation to the lighter two is weak. The average lifetime of the $B^{\pm}$ meson, for example, is 1.64×10^{-12} s [22], giving a typical decay length of $\gamma c\tau \approx 5$ mm for relativistic factor $\gamma = 10$. This means experimental techniques can be used to *tag* jets containing B hadrons [248, 249], which is useful for many analyses of particle collision data, especially those containing t quarks.

Due to its long lifetime, it is expected that the decay of a B hadron will occur at some resolvable distance from the primary vertex. The point of B hadron decay in space is called the *secondary vertex*. For b quarks produced with significant energy, the B decay products will be produced at an acute relative angle in the laboratory frame, so are expected to remain within the same clustered jet. One indication of whether a jet contains a B hadron is given by the *impact parameters* of the constituent tracks. The transverse and longitudinal impact parameters are denoted d_0 and $z_0 \sin\theta$. They equal to the distance of closest approach of the track to the primary vertex in the transverse and longitudinal planes, respectively. The signed significance of the impact parameter measurements is used for each track to determine the likelihood under different *jet flavour hypotheses*: b, c, and *light* (u, d, s, g). These are constructed from simulated events where the quark is matched to each jet within $\Delta R < 0.3$. The transverse impact parameter and likelihood ratio for tracks in simulated $t\bar{t}$ events are shown in Figure 3.22b. The impact parameter significance is given a negative sign if the point of closest approach of the track to the primary vertex lies in the opposite direction to the jet. It is positive otherwise [250].

To further improve the discriminating power of b from c flavour jets, properties of the reconstructed secondary vertex are considered [250]. A list of possible secondary vertices is made from pairs of well-measured tracks in the jet, each with at least seven hits and a good global track fit. These two-track vertex candidates are required to have a significant displacement
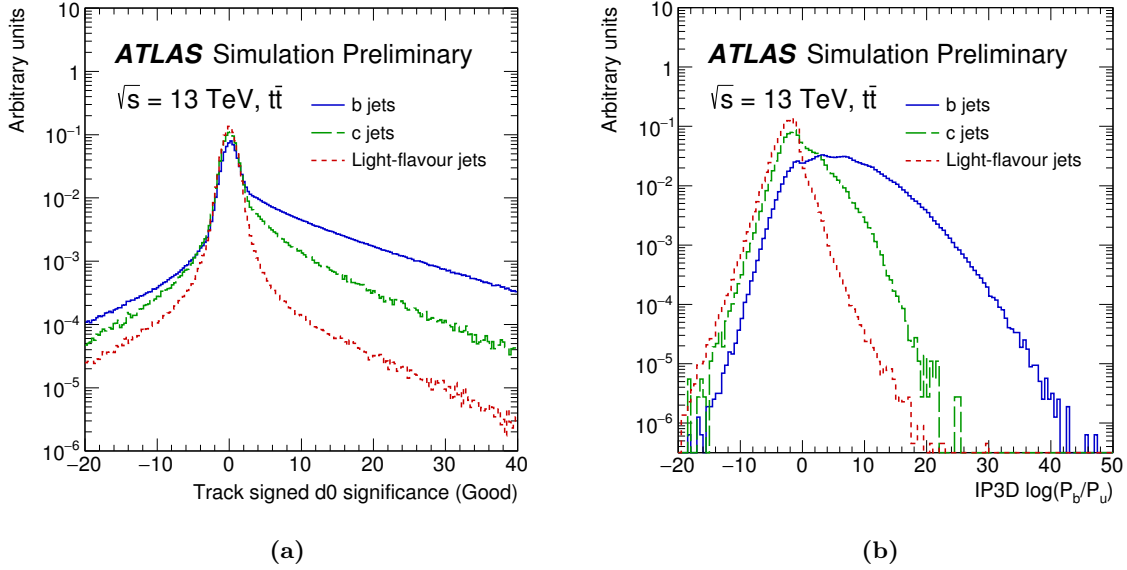
Figure 3.22 (a) Signed transverse impact parameter significance and (b) *b*-to-light log likelihood ratio for tracks belonging to *b*, *c*, and light flavoured jets in simulated $t\bar{t}$ decays. From [249].

from the primary vertex. Pairs of tracks which are likely to have originated from decays of other long-lived particles such as K_S^0 and Λ are removed, as well as those likely to be from interactions with detector material and those with combined invariant mass above $6\,\text{GeV}$. The surviving tracks are then all used to determine a final secondary vertex, with an iterative pruning procedure used to ensure a good fit.

The outputs from the impact parameter and secondary vertex calculations are combined with normalised jet four-momentum and decay chain multi-vertex information [251] to form the feature set for a boosted decision tree discriminator, *MV2c10* [248]. The model is trained using simulated $t\bar{t}$ events, with the signal sample containing *b* flavour jets and the background made up of 7% *c* and 93% light flavour jets. The distribution of the discriminator scalar output is shown in Figure 3.23.

Jets with an associated discriminant value above a chosen threshold are referred to as *b*-jets. For the analysis described in Chapter 4, the threshold MV2c10 > 0.8244273 is used. This corresponds to a 70% *b*-jet selection efficiency. At this working point, the rejection factor is 12.17 for *c* flavour jets, and 381.3 for light flavour jets [252]. The preceeding ATLAS measurement in
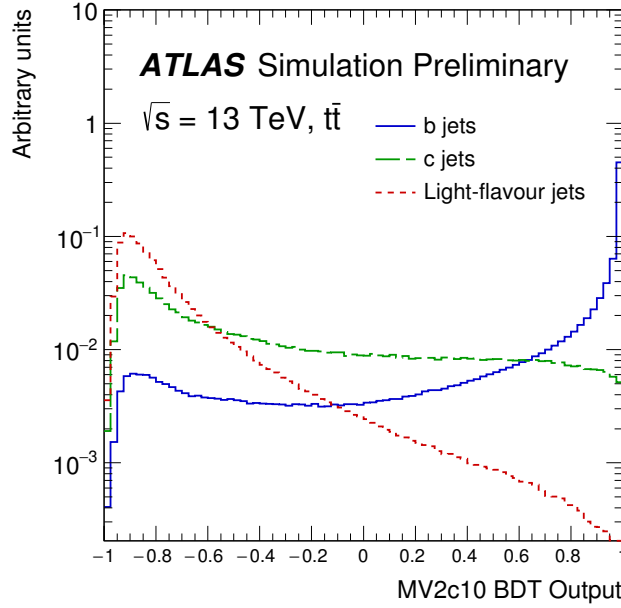
Figure 3.23 Distribution of the MV2c10 discriminant for simulated b, c, and light flavoured jets in $t\bar{t}$ events. From [249].

the resolved channel [161] used a working point with efficiency of 77%, with rejection factor 6.21 for c flavour jets and 134.3 for light flavour jets. For the anaylsis described in Chapter 4, a greater than tenfold larger data sample is analysed. This presents the opportunity to use a b tagging working point with lower efficiency. This results in a higher purity, albeit smaller, sample. Accordingly, the accepted sample will incur reduced associated systematic errors, but increased relative statistical uncertainty. Studies of the effects due to different choices of b tagging working point were performed and found that optimal total uncertainty is reached at the 70% efficiency level [253]. This choice reduces the QCD multi-jet background estimation systematic uncertainties to a level comparable with the overall statistical uncertainty. The use of a working point with higher efficiency – and therefore lower purity – does improve the statistical precision of the analysis, despite a lower relative statistical uncertainty. Conversely, working points with efficiency lower than 70% were determined to increase the relative statistical uncertainty to undesirable levels.

3.3.5 Overlap removal

The reconstruction algorithms of physics objects described in this section do not use detector signals exclusively. That is, a signal used in the construction of one object can also be used in another. To remove these ambiguities, spatially overlapping pairs of objects are considered and one element is removed from the event. The distance between two objects is measured using the metric $\Delta R_y = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$, where Δy and $\Delta\phi$ are the differences in object rapidity and azimuthal angle, respectively. The overlap removal procedure used for the analysis described in Chapter 4 is given by:

1. If an electron and muon share a track in the inner detector, the electron is removed.

2. If a jet and electron have $\Delta R_y < 0.2$, the jet is removed.

3. If a jet and electron have $\Delta R_y < 0.4$, the electron is removed.

4. If a muon and jet have $\Delta R_y < 0.2$, the jet is removed if its p_{T}, total track p_{T}, and number of tracks are consistent with radiation emitted by or energy lost from the muon.

5. If a muon and jet have $\Delta R_y < 0.4$, the muon is removed.

6. If a jet from a hadronically-decaying τ lepton and another jet have $\Delta R_y < 0.2$, the non-τ jet is removed.

The precedence of the objects and order of the overlap removals are chosen to favour preservation of the reconstruction efficiency of the $t\bar{t}$ signal.

3.4 Event simulation

Many high energy physics experiments aim to make comparisons of experimental data and theoretical predictions. To do this, simulations of $\sqrt{s} = 13\,\mathrm{TeV}$ pp collisions are made to predict the signal detected by ATLAS under the Standard Model hypothesis [254]. This process is divided into a linked chain of separate simulations, each employing Monte Carlo sampling methods.

A fixed-order transition matrix element is calculated and integrated over the final state phase space to derive a predicted cross section. For each point sampled, interacting quarks and gluons are assigned initial momenta according to the *parton distribution functions* which describe the fraction of the proton momentum carried by the parton. The outgoing scattered particles and their momenta are given by application of the transition matrix to the incoming state.

Particles which are charged under $SU(3)_C$ are modelled to emit *QCD radiation* as the energy scale of the process evolves. For high-energy outgoing quarks and gluons, the 'parton splitting' process is modelled by the *DGLAP equations* [255–257], until the partonic energy scale reaches about 1 GeV. Outgoing low-energy quarks and gluons undergo hadronisation, as described in Section 2.2.4. This is simulated with non-perturbative phenomenological models [51, 258]. Decays of short-lived particles with lifetime $c\tau < 10$ mm are simulated at this stage.

The resulting simulation of the outgoing particles from the 'hard scattering' process are overlaid with simulations of the underlying event. This is the propagation of the remnants of the colliding protons, simulated with phenomenological models tuned with data. Additionally, *pileup* collisions between other protons in the beam are overlaid.

The description of decayed particles produced by the event generation procedures are input into a detector simulation, a detailed computer model of the ATLAS detector materials and geometry, including misalignments and electrical malfunctions. The GEANT4 simulation software [259–261] is continually calibrated, tuned, and validated throughout the lifetime of the experiment. The output from the full event simulation is treated by the same reconstruction pipeline as real data from the experiment. This preserves compatibility and consistency between the simulated and real datasets.

The full simulation of a single *pp* collision event producing $t\bar{t}$ takes on average 33 minutes to run [254], normalised for comparisons of different CPUs [262]. This includes a factor of 160 for simulating the pileup effects for an instantaneous collision luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. This represents a significant use of resources for the experiment. For the analysis presented in Chapter 4, multiple such samples are required to quantify the systematic uncertainty in

the result due to the choice in procedure for the event generator and parton shower. It is not viable to produce full simulations for these variation samples in addition to the nominal sample. Therefore an approximate detector simulation, ATLFAST-II [263], is used. This model applies parameterised functions to the energies of particles incident on the calorimeter system, rather than simulating the scattering and propagation processes through it. This fast simulation also parameterises the particle showers in the calorimeters, taking the active and passive material geometry into account. These approximations result in a speed-up factor of 20, with a $t\bar{t}$ event taking $101\,\mathrm{s}$ on average [254].

Chapter 4

Measurements of $t\bar{t}$ differential production cross sections

In this chapter, an analysis resulting in multiple measurements of $t\bar{t}$ differential production cross sections as functions of kinematic variables in the resolved all-hadronic decay channel is presented. It is performed on $36.1\,\text{fb}^{-1}$ of $\sqrt{s} = 13\,\text{TeV}$ pp collision data collected by ATLAS at the LHC in 2015 and 2016. Much of the contents of this chapter are also included in documentation by the ATLAS collaboration [253, 264].

4.1 Selection and reconstruction

The dataset is filtered to select collision events which match the desired signature of $t\bar{t}$ decaying hadronically into jets that can be individually resolved by the detector. The cutoff values are chosen to ensure good purity of this signal process in the selected sample, to reduce contamination from background processes, and also to maintain efficiency. The first selection takes place at the trigger level. Then a series of object and event selections are made after reconstruction of the t quarks and $t\bar{t}$ system in each event. These are detailed in this section.

# jets	jet p_{T} [GeV]	# b-jets	MC yield
≥ 6	> 45	≥ 0	$329\,317 \pm 482$
≥ 5	> 65	≥ 0	$240\,166 \pm 412$
≥ 2	> 75	≥ 2	$303\,399 \pm 464$

Table 4.1 Comparison of trigger strategies. The minimum number of jets, p_{T} for those jets, and number of b-tagged jets are shown with the associated yields after applying to a Monte Carlo simulation, normalised to the data luminosity. The top row corresponds to the strategy used for the analysis in this chapter.

4.1.1 Trigger

The analysis in this chapter uses collision events which pass a single algorithm running in the ATLAS high level trigger system. The chosen algorithm selects events with at least six hadronic jets, all of which have $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ and $|\eta| < 2.4$. This trigger was not prescaled for the runs when the dataset used in this analysis was recorded. In subsequent data-taking periods in 2017 and 2018, this trigger algorithm was associated with a large prescale factor to account for the increased instantaneous luminosity delivered by the LHC. This means that adding this additional dataset has a negligible effect on the yield while adding complications and bifurcations to the data analysis. For these reasons, the dataset considered in this analysis is restricted to the $36.1\,\mathrm{fb}^{-1}$ collected by ATLAS in 2015 and 2016.

A comparison with two other possible trigger strategies is shown in Table 4.1, where the top row corresponds to the chosen strategy. The yield is shown for various trigger selections applied to a simulated $t\bar{t}$ all-hadronic sample, scaled to the data luminosity. The numbers shown are given after applying the event selections described below.

The chosen trigger algorithm results in the largest yield when applied to Monte Carlo samples (detailed in Section 4.4), due to the lower p_{T} requirement. A trigger requiring b-jets is not used so that a '0b' region can be used in the background estimation described in Section 4.5. A 'turn on curve' for this trigger is shown in Figure 4.1. It shows the efficiency of selecting the sixth jet as a function of its transverse momentum after the five leading have been selected
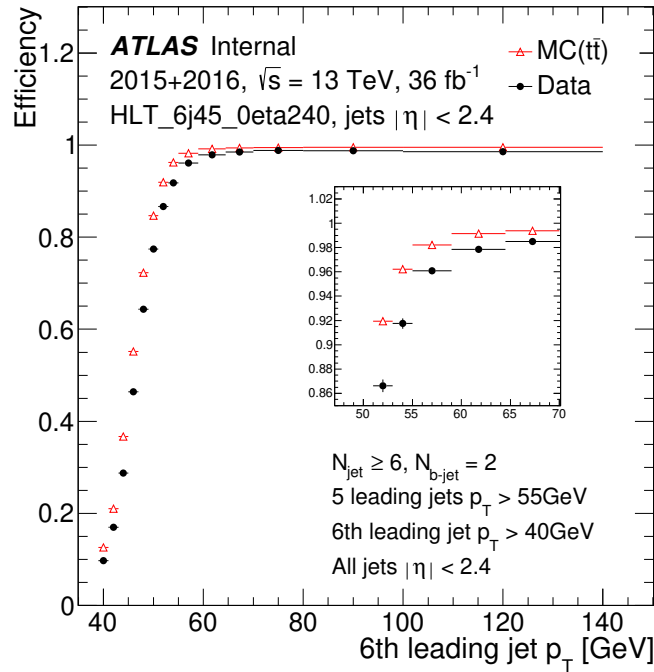
Figure 4.1 Turn-on curve for the 6-jet trigger used in the analysis presented in this chapter. The efficiency of selecting the sixth jet is shown as a function of its transverse momentum. The five leading jets are required to have $p_{\mathrm{T}} > 55\,\mathrm{GeV}$. From [253].

with $p_{\mathrm{T}} > 55\,\mathrm{GeV}$. This is performed only on events with exactly 2 b-tagged jets. The trigger efficiency in data when the sixth jet has $p_{\mathrm{T}} = 55\,\mathrm{GeV}$ is 92.0%. The efficiency of the trigger selection on Monte Carlo data is higher than for data throughout the curve. The difference is 4% at $55\,\mathrm{GeV}$. This is found to have less than 1% impact on the final cross section calculations. This threshold is chosen for all six leading jets to balance signal efficiency with trigger efficiency.

4.1.2 Event selection

Collision events containing at least six jets with $p_{\mathrm{T}} > 55\,\mathrm{GeV}$ are selected. Extra jets are included in the event if they have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$. Any events containing e, μ, or leptonically-decaying τ leptons – reconstructed as per Section 3.3.2 – with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ are rejected. Events containing any reconstructed hadronically-decaying τ leptons with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are also rejected. This is found to slightly improve the background rejection in Monte Carlo simulation, while maintaining signal purity and efficiency.

The surviving background processes are dominated by the production of multi-jet events from QCD scattering. This process typically produces showers of light quarks, whereas the signal $t\bar{t}$ decay process produces heavy b quarks. Therefore requiring events to contain b-jets strongly suppresses the multi-jet background while maintaining efficiency for $t\bar{t}$ decay events. Events containing exactly two b-tagged jets, as described in Section 3.3.4, are selected. While allowing events with more than two b-tagged jets increases the signal yield by 17%, it is also found to increase the multi-jet background contribution by 61% and the other background process contributions by 23%, thereby lowering the signal purity. Additionally, the inclusion of extra b-tagged jets would introduce further combinatorial complications to the downstream reconstruction steps in this analysis. Therefore only events containing exactly two b-tagged jets are selected.

Following these event selections, an attempt is made to reconstruct the t and $\bar{t}$ quark four-momenta and the $t\bar{t}$ system of the hypothesised signal process. This allows more selections to be made, further reducing background contributions, as well as defining kinematic variables of the t quarks to be studied. The same selections are used at both detector and particle level in this analysis.

4.1.3 t quark reconstruction

Simulated $t\bar{t}$ events are used to determine the experimental resolutions of various invariant mass quantities derived from combinations of jets. To do this, the parton-level $t\bar{t}$ decay products are matched to reconstructed jets within $\Delta R < 0.3$. The widths of Gaussian fits to the invariant mass distributions for reconstructed $t \rightarrow bqq'$ and $W \rightarrow qq'$ are denoted σ_t and σ_W, respectively. The detector-level invariant mass distributions and fits for the reconstructed $t \rightarrow Wb$ and W bosons are shown in Figure 4.2. Fits are performed in the cores of the distributions (150 to 200 GeV for m^t and 65 to 100 GeV for m^W) to avoid interference from effects in the tails. This results in extracted resolutions of $\sigma_t = 10.7$ GeV and $\sigma_W = 5.90$ GeV, assuming perfect matching.

(a)   **(b)**
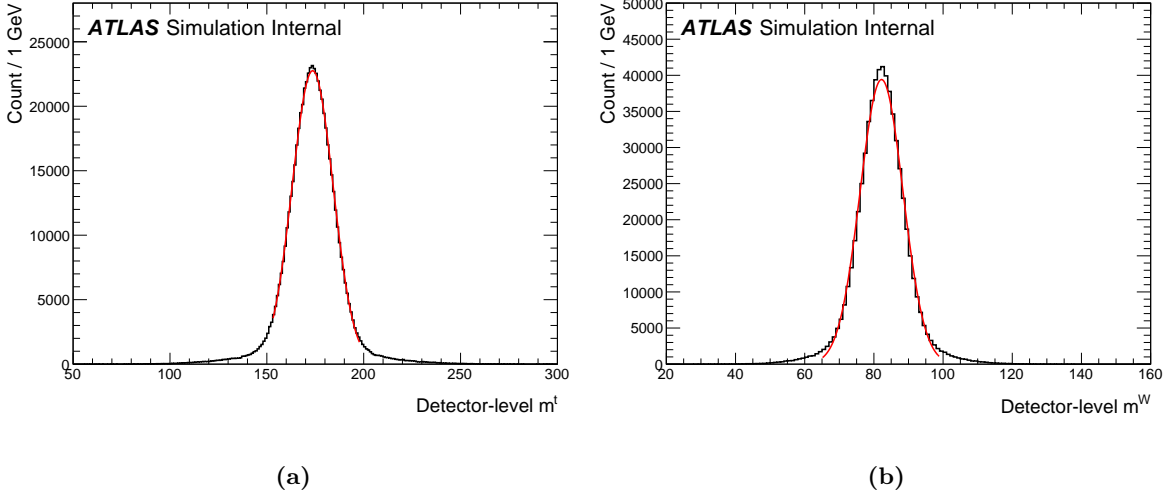
Figure 4.2 Distributions (black histograms) and fits (red curves) for (a) the invariant mass of bqq' jets from t decays and (b) pairs of jets matched to quarks from W decays in $t\bar{t}$ simulations.

In what follows, subscript b denotes a jet which has been assigned a b-tag and j denotes a non-b-tagged jet. A χ^2 discriminant is defined

$$\chi^2 = \frac{\left(m_{b_1 j_1 j_2} - m_{b_2 j_3 j_4}\right)^2}{\sigma_t^2} + \frac{\left(m_{j_1 j_2} - m_W\right)^2}{\sigma_W^2} + \frac{\left(m_{j_3 j_4} - m_W\right)^2}{\sigma_W^2}, \tag{4.1}$$

where $m_{b_1 j_1 j_2}$ is the invariant mass of the combined system of one b-tagged jet and two non-b-tagged jets. Similarly, $m_{j_1 j_2}$ is the invariant mass of the combination of two non-b-tagged jets. This quantity follows a chi-squared distribution with four degrees of freedom for background events, reduced to two degrees of freedom for signal events due to the extra constraints of the originating t quarks. The accepted mass of the W boson is used as $m_W = 80.385\,\text{GeV}$ [22].

The form of Equation 4.1 is chosen to test the $t\bar{t}$ hypothesis, matching pairs of non-b-tagged jets to decaying W bosons and balancing the invariant masses of the two Wb systems in the first term. The t quark mass does not appear explicitly to avoid sculpting background distributions to resemble those from the signal. Of at least 48 possible assignments of jets in each event, the permutation associated with the minimum value of χ^2 is chosen to define the reconstructed leading and subleading W bosons (W_1 and W_2) and the leading and subleading t quarks (t_1 and t_2). The reconstructed $t\bar{t}$ system has four-momentum defined by the sum of

the four-momenta of t_1 and t_2. This assignment of jets is found match the true configuration in approximately 60% of cases for events containing six jets, decreasing to 40–50% for higher jet multiplicities [265].

4.1.4 Background suppression

Events with $\chi^2_{\min} > 10$ are rejected, since these fit the $t\bar{t}$ hypothesis poorly. The cut value is chosen to balance the optimal points for the particle and parton level analyses, obtained from Monte Carlo simulations. Studies demonstrate that the efficiency is largely insensitive to the exact choice of cut, and the impact of using this middle value is at most 1%. The distributions of $\chi^2_{\min}$ in simulated $t\bar{t}$ and non-all-hadronic events and data are shown in Figure 4.3a, after the preselection detailed in Section 4.1.2 is applied but before any others. They are shown inclusive of other cuts described here, in Figure 4.4a.

In the dominant $g \to b\bar{b}$ background process, the b-jets are typically produced at an acute angle in the laboratory frame. In contrast, the signal $t\bar{t}$ process produces the two b-jets at a larger relative angle. To further suppress the background contribution, a selection is used where events are required to have $\Delta R_{b_1 b_2} > 2.0$ to pass. The threshold value is chosen by hand to balance signal purity against selection efficiency, using simulated data. The distribution for this variable in simulated $t\bar{t}$ and non-all-hadronic events and data, after the selection in Section 4.1.2 but before any others, is shown in Figure 4.3b. They are shown inclusive of other cuts described here, in Figure 4.4b.

Similarly, a cut is placed on the relative angle of the momenta of the b jets and their associated W boson. For the energy of t quarks produced in $t\bar{t}$ events by the LHC, it is expected that the maximum such angle, $\Delta R_{bW}^{\max}$ is rarely very large. Background events are suppressed by rejecting events with $\Delta R_{bW}^{\max} \geq 2.2$. The threshold value is chosen by hand to balance signal purity against selection efficiency, using simulated data. The distributions for this variable are shown in Figure 4.3c for all events passing the preselection detailed in Section 4.1. They are shown inclusive of other cuts described here, in Figure 4.4c. The distribution for the signal sample before other selections is double-peaked. The peak in the rejected region

(a) $\chi^2_{\min}$

(b) ΔR_{bb}

(c) $\Delta R_{bW}^{\max}$

(d) m_t
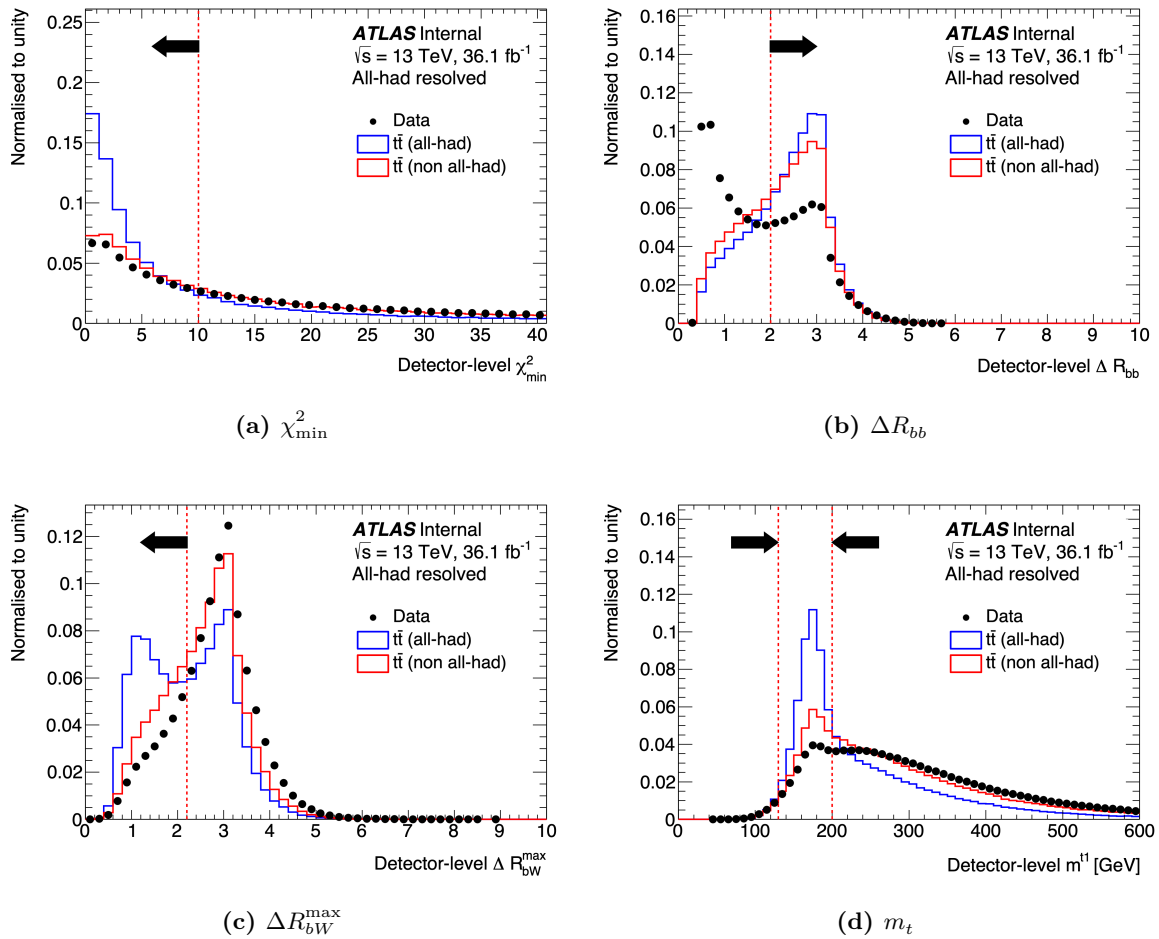
Figure 4.3 Distributions of variables used in selections after event reconstruction, shown for simulated $t\bar{t}$ and non-all-hadronic events as well as data. All samples are shown after the primary selection described in Section 4.1.2, but before any other cuts. Each is normalised to unity. Cut values are indicated by dotted red lines, and the accepted regions by arrows.

(a) $\chi^2_{\min}$

(b) ΔR_{bb}

(c) $\Delta R^{\max}_{bW}$

(d) m_t
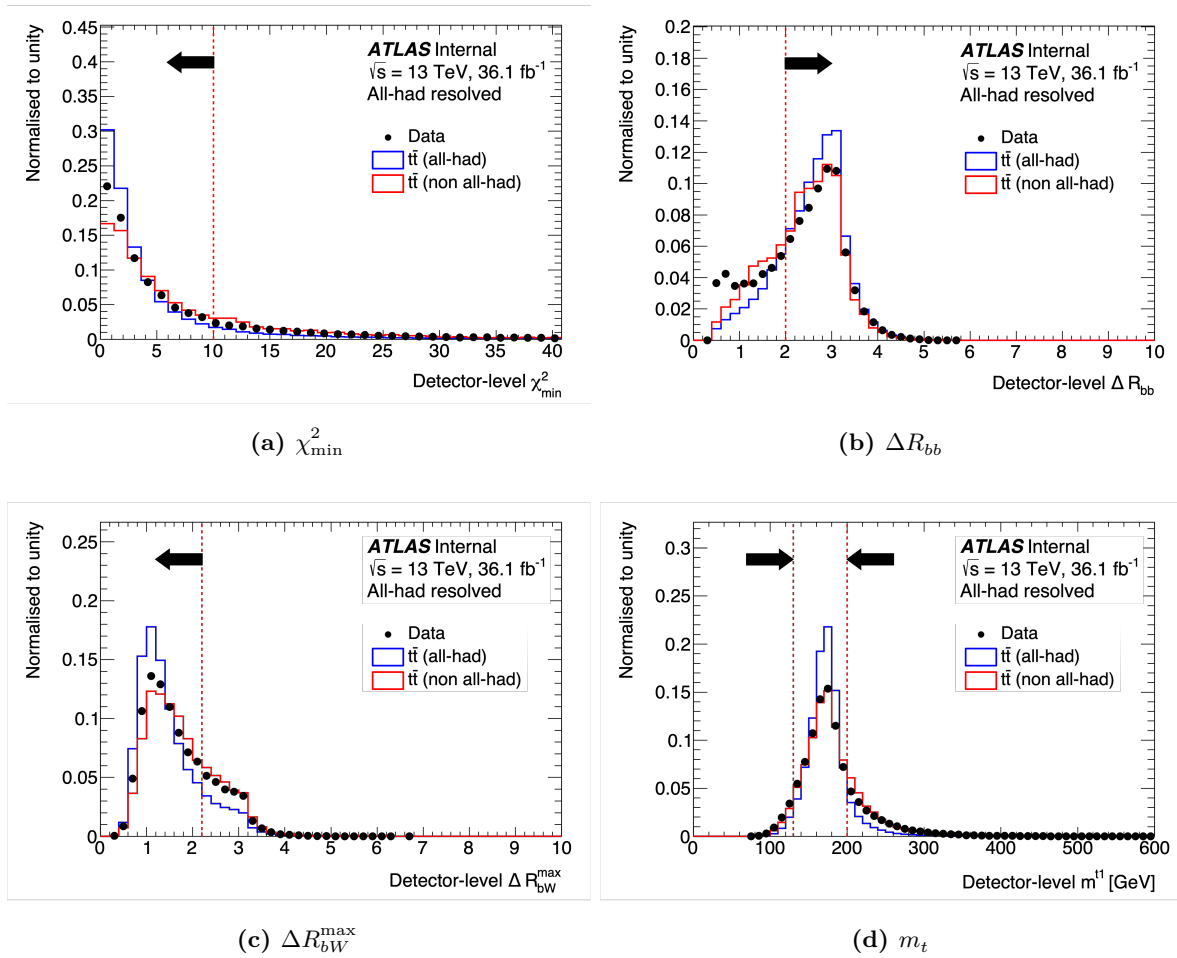
Figure 4.4 Distributions of variables used in selections, after other cuts. Shown for simulated $t\bar{t}$ and non-all-hadronic events as well as data. Each is normalised to unity. All samples are shown inclusive of the selections in other panels. Cut values are indicated by dotted red lines, and the accepted regions by arrows.

corresponds to events where the minimum χ^2 reconstruction assigns jets incorrectly. This peak is strongly suppressed in signal events after applying the $\chi^2_{\min}$, ΔR_{bb}, and m_t cuts described in this section, as demonstrated in Figures 4.4c and 4.5c.

Finally, a selection window is defined for the invariant mass of the reconstructed t quarks in the event. Both are required to have $130\,\mathrm{GeV} < m_t < 200\,\mathrm{GeV}$ to pass selection. The window is determined by the signal region definition for the multi-jet background estimation detailed in Section 4.5. It is chosen to allow reasonable signal and background purities, and to suppress the combined statistical and systematic uncertainties from the data-driven background estimation. The distributions for m_t are shown in Figure 4.3d for events with only the preselection described in Section 4.1.2 applied. They are also shown inclusive of the other cuts described in this subsection in Figure 4.4d.

After this full selection is made, simulations show that the surviving sample is dominated by the $t\bar{t}$ all-hadronic signal. Detector-level distributions, scaled to the data luminosity, for the cut variables $\chi^2_{\min}$, ΔR_{bb}, and $\Delta R_{bW}^{\max}$ for events passing all selections, except for in the independent variable, are shown in Figure 4.5. Other $t\bar{t}$ decay channels are strongly suppressed by the lepton veto selection and subsequent cuts. Yields and efficiencies for the sequence of selections are shown in Table 4.3 for the data and simulated all-hadronic $t\bar{t}$ signal and non-all-hadronic $t\bar{t}$ background samples. The initial sample from data discards events that fail to meet criteria for experimental conditions, such as the status of the ATLAS calorimeters or the quality of the primary vertex reconstruction. This corresponds to an initial selection efficiency of 97.6% for the $36.1\,\mathrm{fb}^{-1}$ dataset. The initial sample for simulations consists of all events from the nominal all-hadronic and non-all-hadronic background $pp \to t\bar{t}$ samples, described in Section 4.4. The smallest efficiency after the trigger decision is due to the $N_{b\text{-jets}} = 2$ selection in all samples, at 6.88% in data. Since the corresponding efficiencies for the simulated signal and non-all-hadronic $t\bar{t}$ background samples are 39.7% and 41.2%, respectively, this cut strongly suppresses the multi-jet background. For subsequent cuts, the background efficiencies are always smaller than those for signal, since the selections are chosen to maximise signal purity while suppressing the contamination from background. The final selection efficiencies are 9.84×10^{-5} for data and 2.15×10^{-3} for the simulated signal.

(a) $\chi^2_{\min}$

(b) ΔR_{bb}

(c) $\Delta R_{bW}^{\max}$
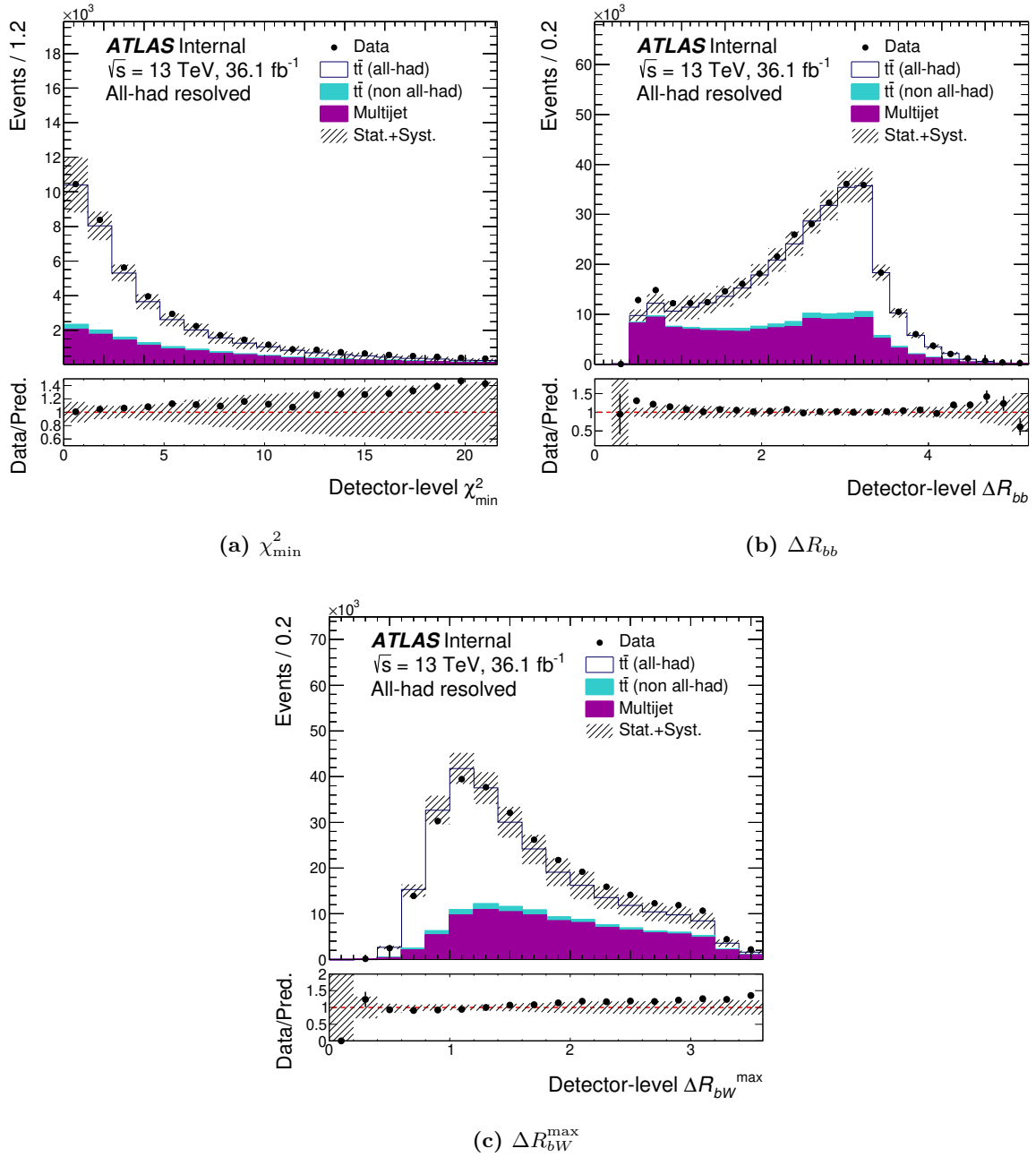
Figure 4.5 Distributions of variables used for event selection, inclusive of other cuts. Each distribution includes only the events passing the initial selection and cuts on variables except for the one being shown.

Sample	Yield	Proportion
$t\bar{t}$ all-hadronic	$29459\,^{+2044}_{-2454}$	67.70%
$t\bar{t}$ non-all-hadronic	$1494\,^{+140.4}_{-122.6}$	3.43%
Multi-jet	$12564\,^{+1896}_{-1894}$	28.87%
Total MC	$43517\,^{+2747}_{-3018}$	100%
Data	44621 ± 211	100%

Table 4.2 Event yields after the full event selection is applied. The Monte Carlo samples are normalised to the data luminosity. The uncertainties shown for simulation are the sum in quadrature of the statistical and detector systematics. The uncertainty on the data is entirely statistical.

Selection	Data			Signal MC			Non-all-hadronic background MC		
	Yield	Eff. [%]	Cum. eff. [%]	Yield	Eff. [%]	Cum. eff. [%]	Yield	Eff. [%]	Cum. eff. [%]
Initial	4.4254×10^8	97.6	97.6	1.3699×10^7	100	100	1.6312×10^7	100	100
Trigger	5.1478×10^7	11.6	11.4	9.5296×10^5	6.96	6.96	3.9369×10^5	2.41	2.41
$N_{e,\mu} = 0$	5.0382×10^7	97.9	11.1	9.2150×10^5	96.7	6.73	1.6928×10^5	43.0	1.04
$N_{\text{jets}} \geq 6$	1.6034×10^7	31.8	3.54	3.7324×10^5	40.5	2.72	6.4639×10^4	38.2	0.396
$N_{b\text{-jets}} = 2$	1.1031×10^6	6.88	0.243	1.4801×10^5	39.7	1.08	2.6611×10^4	41.2	0.163
$\chi^2_{\min} \leq 10$	3.3172×10^5	30.1	0.0732	8.0904×10^4	54.7	0.590	8.7963×10^3	33.1	0.0539
$\Delta R_{bb} > 2.0$	1.5599×10^5	49.2	0.0344	5.6377×10^4	69.7	0.412	5.4205×10^3	61.6	0.0332
$\Delta R_{bW}^{\max} < 2.2$	7.2657×10^4	46.6	0.0160	3.7222×10^4	66.0	0.272	2.5394×10^3	46.8	0.0156
m_t window	4.4621×10^4	61.4	0.00984	2.9459×10^4	79.1	0.215	1.4919×10^3	58.7	0.00915

Table 4.3 Selection yields and efficiencies for the data and nominal simulated all-hadronic signal and non-all-hadronic background $t\bar{t}$ samples, normalised to the data luminosity.

The final yields after selection, with combined statistical and systematic uncertainties, are shown in Table 4.2, stratified by sample. To start, the $t\bar{t}$ all-hadronic sample contains $39\,885\,000$ events, corresponding to an integrated luminosity of $54.654\,\text{fb}^{-1}$. The $t\bar{t}$ non-all-hadronic sample contains $119\,432\,000$ events, corresponding to an integrated luminosity of $163.66\,\text{fb}^{-1}$. These are normalised to the luminosity of the data sample, $36.1\,\text{fb}^{-1}$. The remaining background is almost entirely dominated by events containing jets seeded by u, d, s, c, or b quarks, or by the scattering of gluons. This irreducible multi-jet background is estimated using a data-driven method, described in Section 4.5.

4.2 Observables

The analysis presented in this chapter measures differential cross sections as functions of many observables. They are unfolded at particle level or parton level, as described in Section 4.6. Many of these variables rely on the reconstruction of the t quarks and $t\bar{t}$ system, described in Section 4.1.3. Here t_1 and t_2 refer to the leading- and subleading-p_T reconstructed t or $\bar{t}$ quark, respectively. A complete list of variables unfolded to particle level and parton level and their descriptions are shown in Tables 4.4 and 4.5, respectively.

The kinematics of the $t\bar{t}$ decay can be observed through the reconstructed t quark transverse momenta, $p_\text{T}^{t_1}$ and $p_\text{T}^{t_2}$ and the rapidities, $|y^{t_1}|$ and $|y^{t_2}|$. For the combined $t\bar{t}$ system, the transverse momentum $p_\text{T}^{t\bar{t}}$, rapidity $|y^{t\bar{t}}|$, and invariant mass $m^{t\bar{t}}$ are measured. The cross sections as functions of these variables are unfolded to both particle and parton levels. Combined, these measurements capture the total and constituent four-momenta of the $t\bar{t}$ system. This enables comparisons with fixed-order calculations, which test predictions from theoretical models [132–134, 137], characterise the $t\bar{t}$ production process as a background for other analyses [138–140], and enable general searches and determinations of limits for new physics [266, 267].

Unfolding to particle level, cross sections as functions of particular variables may be used to constrain possible forms for the transition matrix element and parton shower process used in

simulations. The rapidity of the t and $\bar{t}$ quarks in the $t\bar{t}$ rest frame are given by $y^{\star} = (y^{t_1} - y^{t_2})/2$ and $-y^{\star}$. The observable $\chi^{t\bar{t}} = \exp(2|y^{\star}|)$ is highly sensitive to differences in t quark rapidities, peaking at low values for many processes beyond the Standard Model [268, 269]. The longitudinal motion of the $t\bar{t}$ system is indicated by the observable $|y_{\text{boost}}| = |y^{t_1} + y^{t_2}|/2$, which is sensitive to the form of the parton distribution functions used to sample the initial state momenta [157, 269]. The *out-of-plane momentum* is the projection of the three-momentum of one of the t quarks along the normal to the plane formed between the other t quark and the beam axis:

$$P_{\text{out}} = \mathbf{p}^{t_1} \cdot \frac{\mathbf{p}^{t_2} \times \hat{\mathbf{z}}}{|\mathbf{p}^{t_2} \times \hat{\mathbf{z}}|} \ . \tag{4.2}$$

Observable	Description
N_{jets}	Number of reconstructed jets in the event
$p_{\text{T}}^{t_1}$	Leading t quark transverse momentum
$|y^{t_1}|$	Leading t quark absolute rapidity
$p_{\text{T}}^{t_2}$	Subleading t quark transverse momentum
$|y^{t_2}|$	Subleading t quark absolute rapidity
$p_{\text{T}}^{t\bar{t}}$	$t\bar{t}$ system transverse momentum
$|y^{t\bar{t}}|$	$t\bar{t}$ system absolute rapidity
$m^{t\bar{t}}$	$t\bar{t}$ system invariant mass
$H_{\text{T}}^{t\bar{t}}$	$p_{\text{T}}^{t_1} + p_{\text{T}}^{t_2}$; scalar sum of t quark transverse momenta
$\chi^{t\bar{t}}$	$\exp(|y_{t_1} - y_{t_2}|)$; sensitive to small differences in rapidity
$Z^{t\bar{t}}$	$p_{\text{T}}^{t_2}/p_{\text{T}}^{t_1}$; ratio of subleading to leading t quark transverse momenta
$\cos\theta^{\star}$	Cosine of the relative polar angle of the t quarks in the $t\bar{t}$ rest frame
$\Delta\phi$	Relative azimuthal between the t quarks
$|y_{\text{boost}}|$	Absolute average rapidity of the t quarks
$|P_{\text{out}}|$	Magnitude of the out-of-plane momentum
$|P_{\text{cross}}|$	Magnitude of the cross product of jet directions
R_{Wb}^{leading}	$p_{\text{T}}^{W}/p_{\text{T}}^{b}$; ratio of W boson to b quark p_{T} for the leading t quark
$R_{Wb}^{\text{subleading}}$	$p_{\text{T}}^{W}/p_{\text{T}}^{b}$; ratio of W boson to b quark p_{T} for the subleading t quark
R_{Wt}^{leading}	$p_{\text{T}}^{W}/p_{\text{T}}^{t}$; ratio of W boson to t quark p_{T} for the leading t quark
$R_{Wt}^{\text{subleading}}$	$p_{\text{T}}^{W}/p_{\text{T}}^{t}$; ratio of W boson to t quark p_{T} for the subleading t quark

Table 4.4 Observables for one-dimensional $t\bar{t}$ differential cross sections unfolded to particle level.

Observable	Description
$p_{\mathrm{T}}^{t_1}$	Leading t quark transverse momentum
$|y^{t_1}|$	Leading t quark absolute rapidity
$p_{\mathrm{T}}^{t_2}$	Subleading t quark transverse momentum
$|y^{t_2}|$	Subleading t quark absolute rapidity
$p_{\mathrm{T}}^{t\bar{t}}$	$t\bar{t}$ system transverse momentum
$|y^{t\bar{t}}|$	$t\bar{t}$ system absolute rapidity
$m^{t\bar{t}}$	$t\bar{t}$ system invariant mass
$H_{\mathrm{T}}^{t\bar{t}}$	$p_{\mathrm{T}}^{t_1} + p_{\mathrm{T}}^{t_2}$; scalar sum of t quark transverse momenta
$\chi^{t\bar{t}}$	$\exp(|y_{t_1} - y_{t_2}|)$; sensitive to small differences in rapidity
$\Delta\phi$	Relative azimuthal between the t quarks
$|y_{\mathrm{boost}}|$	Absolute average rapidity of the t quarks

Table 4.5 Observables for one-dimensional $t\bar{t}$ differential cross sections unfolded to parton level.

This observable and $\Delta\phi$ are particularly sensitive to additional radiation in the main scattering process, and therefore they are strongly effected by effects beyond leading order in the perturbative expansion of the transition matrix element [270]. Hence measurements of these allow the validation of matrix element calculations.

Many of these were previously measured with $\sqrt{s} = 8\,\mathrm{TeV}$ ATLAS data in the $t\bar{t}$ lepton+jets decay channel [157], and in the fully hadronic, highly boosted channel at $\sqrt{s} = 13\,\mathrm{TeV}$ [145]. These analyses showed only modest agreement between theory and measurement. The analysis presented here provides complementary measurements to these. In the fully hadronic decay mode, the reconstruction of the $t\bar{t}$ system does not depend on missing energy in the detector, carried away by invisible neutrinos. Since determination of missing energy has a large experimental uncertainty, the fully hadronic analysis presented in this chapter is able to measure angular quantities with smaller resolutions than those with leptonic decay modes. For example, the resolution of $|\eta^{t_1}|$ is 0.4 in the fully hadronic channel, compared to 0.8 in the lepton+jets channel [271]. Therefore, good resolutions can be achieved for the directional observables, $|P_{\mathrm{out}}|$, $\chi^{t\bar{t}}$, and $\Delta\phi$. Cross sections as functions of another angular variable, $\cos\theta^{\star}$, characterise the production angle of the t quarks in the $t\bar{t}$ rest frame. The variable is generally measured

well, and some models beyond the Standard Model predict changes to its distribution [272]. Measurements in the fully hadronic, highly boosted decay channel show modest differences with predictions [145]. A measurement in the resolved case complements this.

Distributions of combinations of the transverse momenta of reconstructed particles in $t\bar{t}$ events can demonstrate particular sensitivities to differences between theory and observations [157, 273, 274]. In previous differential cross section measurements performed in channels containing leptons, fewer events are observed than predicted at high values of $H_{\mathrm{T}}^{t\bar{t}}$. For cross sections as functions of the ratio variable R_{Wt}, the data show an excess of events at larger values. In addition, information on the sharing of transverse momenta between the t quark decay products can be used to constrain matrix element and parton shower models [145, 157]. This motivates the measurements of differential cross sections as functions of $H_{\mathrm{T}}^{t\bar{t}}$, $Z^{t\bar{t}}$, and R_{Wt} for both reconstructed t quarks (denoted 'leading' and 'subleading'), as well as R_{Wb}.

The fully hadronic resolved channel studied in this analysis is the only $t\bar{t}$ decay channel where the four-vectors for all decay products can be determined. Therefore it is uniquely sensitive to a combination of jet directions never measured before,

$$P_{\mathrm{cross}} = \left[\hat{\mathbf{p}}_{b_1} \times \left(\hat{\mathbf{p}}_{j_1} \times \hat{\mathbf{p}}_{j_2} \right) \right] \times \left[\hat{\mathbf{p}}_{b_2} \times \left(\hat{\mathbf{p}}_{j_3} \times \hat{\mathbf{p}}_{j_4} \right) \right] \;, \tag{4.3}$$

where $\hat{\mathbf{p}}$ denotes a unit vector in the direction of the jet momentum. The differential $t\bar{t}$ production cross section as a function of the absolute magnitude of P_{cross} is measured in this analysis, unfolded to particle level. Since the jet direction is well-measured, compared to energy, $|P_{\mathrm{cross}}|$ can be determined with good resolution. The variable encapsulates information on the directions of the six jets from the $t\bar{t}$ system decay. It therefore enables fine discrimination between the predictions of angular variables from different theoretical models.

Some of the observables defined above are used in two-dimensional differential cross sections, unfolded to particle and parton levels. The particular combinations used are listed in Tables 4.6 and 4.7. The differential $t\bar{t}$ production cross section is reported as a function of the *internal variable* in bins of the *external variable*.

These measurements can provide an improved understanding of correlations between the kinematic properties of the $t\bar{t}$ system, such as in distributions of $p_{\mathrm{T}}^{t_1}$ in bins of $p_{\mathrm{T}}^{t_2}$ and $|y^{t_1}|$ in bins of $|y^{t_2}|$. Such results may be used to validate and tune theoretical models at increasingly higher orders of α_S [169, 275]. Previous measurements have shown increasing discrepancies between predictions and observations with increasing $m^{t\bar{t}}$ [275]. This motivates determinations of differential cross sections in bins of $m^{t\bar{t}}$, in order to probe the modelled characteristics fo the $t\bar{t}$ system at high invariant mass. Furthermore, the distributions of $p_{\mathrm{T}}^{t\bar{t}}$ and $|y^{t\bar{t}}|$ are especially sensitive to the effects of QCD radiation, and can therefore offer insights into the validation of perturbative calculations [169].

Double differential cross sections are useful for a number of downstream analyses, offering insights into many questions at the horizon of particle physics. At particle level, the cross sections as functions of variables stratified by event jet multiplicity characterise the $t\bar{t}$ system associated with different levels of additional radiation. These can be used to tune and constrain parameters of Monte Carlo generator models [162]. Double differential cross sections may also be used to determine parton distribution functions, in particular information about the gluon content of the proton at small Bjorken x [169, 276]. A value of the t quark pole mass can be extracted from template fits to cross sections [175, 277]. Since kinematic distributions of the $t\bar{t}$ system and its decay products are sensitive to the value of the pole mass, new measurements of double (or more) differential distributions can be used to constrain its value to increasingly tighter limits.

4.3 Binning optimisation

The results from this analysis are reported as cross sections in bins of variables of interest. Therefore the definition of the bins, in the form of the positions of their edges, are parameters of the data analysis. Traditionally in particle physics, bins for distributions are defined by hand, based on *a priori* reasoning and adjusted with empirical observations. The method described in this section is based on the detector resolution and tuned for downstream steps.

External	Internal
$p_{\mathrm{T}}^{t_2}$	$p_{\mathrm{T}}^{t_1}$
$m^{t\bar{t}}$	$p_{\mathrm{T}}^{t_1}$
	$p_{\mathrm{T}}^{t_2}$
	$p_{\mathrm{T}}^{t\bar{t}}$
	$\lvert y^{t\bar{t}} \rvert$
N_{jets}	$p_{\mathrm{T}}^{t_1}$
	$p_{\mathrm{T}}^{t_2}$
	$p_{\mathrm{T}}^{t\bar{t}}$
	$\Delta\phi$
	$\lvert P_{\mathrm{out}} \rvert$
	$\lvert P_{\mathrm{cross}} \rvert$

Table 4.6 Observables for two-dimensional $t\bar{t}$ differential cross sections unfolded to particle level. Cross sections are reported as functions of the internal variable in bins of the external variable.

External	Internal
$p_{\mathrm{T}}^{t_2}$	$p_{\mathrm{T}}^{t_1}$
$\lvert y^{t_1} \rvert$	$\lvert y^{t_2} \rvert$
$m^{t\bar{t}}$	$p_{\mathrm{T}}^{t_1}$
	$\lvert y^{t_1} \rvert$
	$p_{\mathrm{T}}^{t_2}$
	$\lvert y^{t_2} \rvert$
	$p_{\mathrm{T}}^{t\bar{t}}$
	$\lvert y^{t\bar{t}} \rvert$

Table 4.7 Observables for two-dimensional $t\bar{t}$ differential cross sections unfolded to parton level. Cross sections are reported as functions of the internal variable in bins of the external variable.

Initial binnings for observables are obtained according to an algorithm based on resolution. Resolutions for some observables at particle level, determined as the root-mean-square (RMS) standard deviation across 200 fine bins, are shown in Figures 4.6 to 4.8. Resolutions for t quark and $t\bar{t}$ kinematic variables at parton level are shown in Figure 4.9.

For each observable, the mean and RMS standard deviation of the difference between the simulated true and measured values, is determined across 200 fine bins. Then the fine bins are merged, starting from the leftmost, until two criteria are simultaneously satisfied:

- The merged bin width is greater than $\delta \times RMS$, where RMS is the total standard deviation of the merged bin. Here δ is a regularisation parameter to be chosen by hand.

- The merged bin error is lower than $5\% \times N$, where N is the number of fine bins within the merged bin. For 200 fine bins and using the Poisson error $(\sqrt{N})$, this constraint is equivalent to requiring at least 400 entries in each merged bin.

When both of these criteria are met, the merged bin is accepted and a new bin is defined by merging from the next fine bin. If an accepted bin is smaller than the previous one, those bins are merged. The parameter δ is scanned between 1.0 and 1.9 in intervals of 0.1. A suitable value for δ is chosen by considering the resulting response matrix to be used in the unfolding procedure, described below.

The binning schemes are reassessed for observables used in two-dimensional cross section measurements. First, the procedure is run for external variable, with stricter requirements of $\delta = 2.0$ and a maximum relative statistical error of 1.5%. Then for each bin of the external variable, the resolution of the internal variable is estimated as above, and the binning algorithm run with $\delta = 2.0$ and a maximum relative statistical error of 3.5%.

For some observables, such as $\cos\theta^\star$, the left-to-right binning procedure may be unsuitable, since it creates bins closer to the start of the variable range before the end. To check sensitivity to this effect, the procedure is also run for these variables in right-to-left mode. Bin edges which contrast strongly between these results are adjusted by hand to minimise the effect of the direction of the merging algorithm.

(a) $p_{\mathrm{T}}^{t_1}$

(b) $|y^{t_1}|$

(c) $p_{\mathrm{T}}^{t_2}$
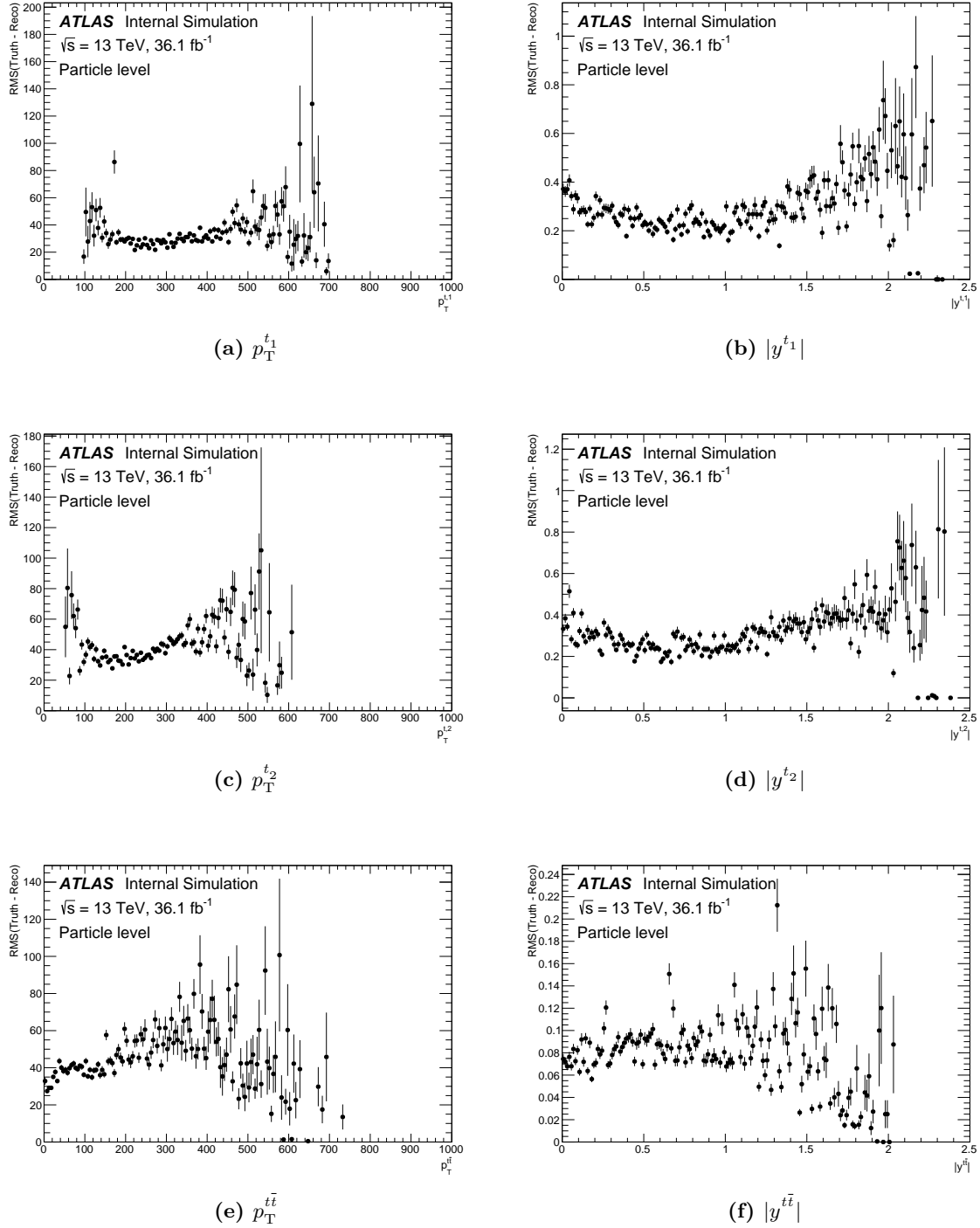
(d) $|y^{t_2}|$

(e) $p_{\mathrm{T}}^{t\bar{t}}$

(f) $|y^{t\bar{t}}|$

Figure 4.6 Resolutions for particle level observables (a) $p_{\mathrm{T}}^{t_1}$, (b) $|y^{t_1}|$, (c) $p_{\mathrm{T}}^{t_2}$, (d) $|y^{t_2}|$, (e) $p_{\mathrm{T}}^{t\bar{t}}$, and (f) $|y^{t\bar{t}}|$.

(a) $m^{t\bar{t}}$

(b) $H_{\mathrm{T}}^{t\bar{t}}$

(c) $\chi^{t\bar{t}}$

(d) $Z^{t\bar{t}}$

(e) $\cos\theta^{\star}$

(f) $\Delta\phi$
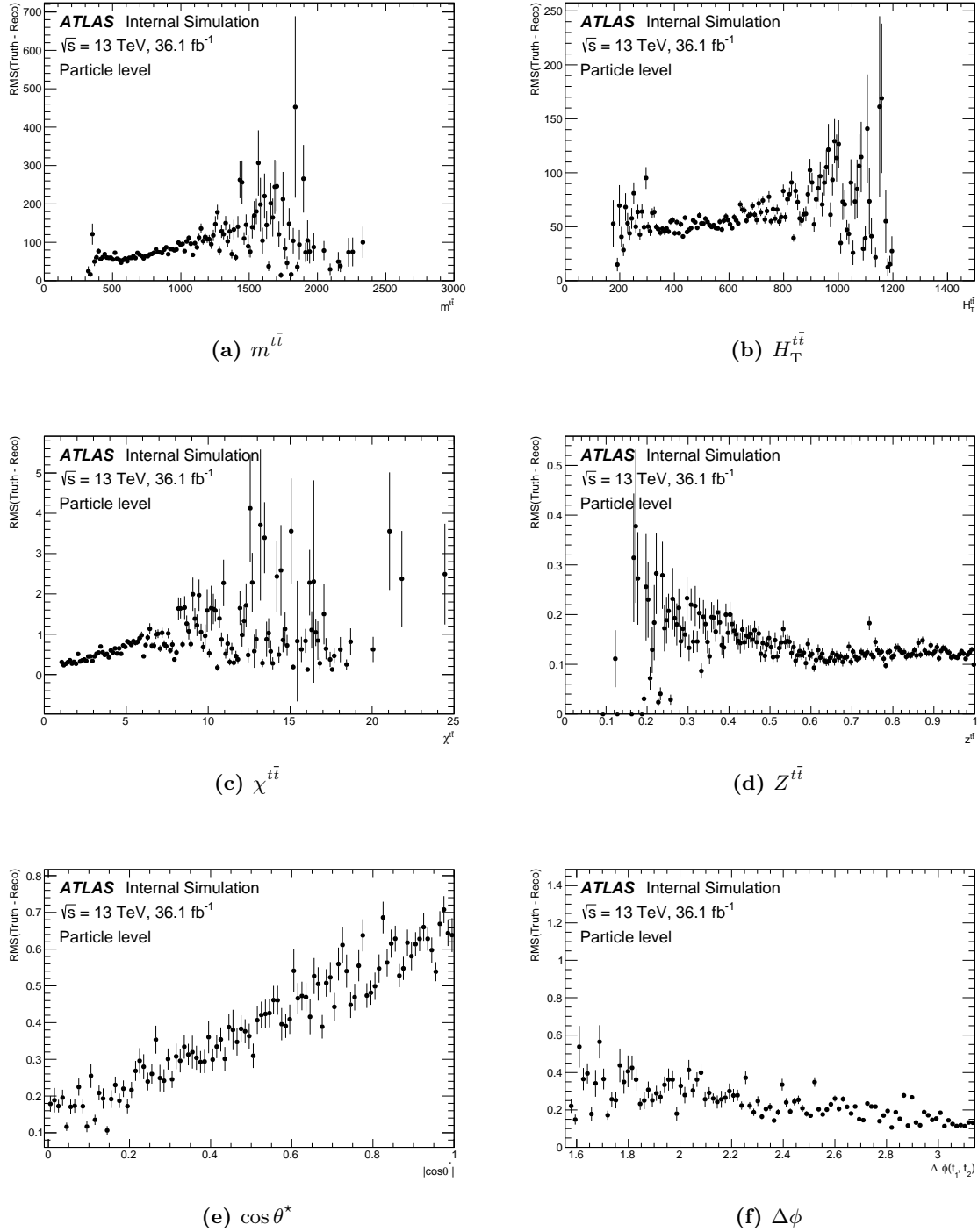
Figure 4.7 Resolutions for particle level observables (a) $m^{t\bar{t}}$, (b) $H_{\mathrm{T}}^{t\bar{t}}$, (c) $\chi^{t\bar{t}}$, (d) $Z^{t\bar{t}}$, (e) $|\cos\theta^{\star}|$, and (f) $\Delta\phi$.
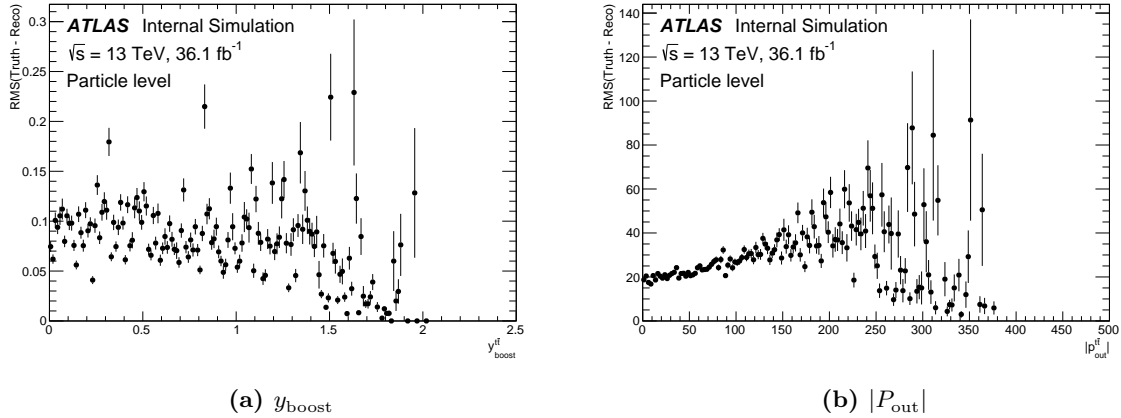
(a) y_{boost}

(b) $|P_{\mathrm{out}}|$

Figure 4.8 Resolutions for particle level observables (a) y_{boost} and (b) $|P_{\mathrm{out}}|$.

The definition of the bin edges for a distribution is a source of implicit regularisation in the unfolding procedure described in Section 4.6. Therefore the level of agreement in the closure and stress tests for the unfolding procedure depends on the output of the binning algorithm. In turn, the covariance matrix of the resulting unfolded distribution is dependent on the binning scheme. Therefore an iterative step is performed, whereby the bin edges are manually altered to reach a satisfactory distribution. This alteration most commonly takes the form of merging small neighbouring bins in order to reduce the migration of events between them. Finally, the merged bin edges may be adjusted slightly by hand to give more regularly spaced intervals between them.

It is notable that compared to an ATLAS analysis on the same dataset using the lepton+jets channel [271], this channel achieves finer binning in the directional distributions, such as $\Delta\phi$ and $\chi^{t\bar{t}}$, under the same procedure. This is despite a much larger signal region sample for the lepton+jets channel, in data after selection. With more data from the LHC and an optimised trigger strategy, the fully hadronic $t\bar{t}$ decay channel offers potentially greater prospects for precision measurements using angular observables than leptonic channels.

(a) $p_{\mathrm{T}}^{t_1}$

(b) $|y^{t_1}|$

(c) $p_{\mathrm{T}}^{t_2}$
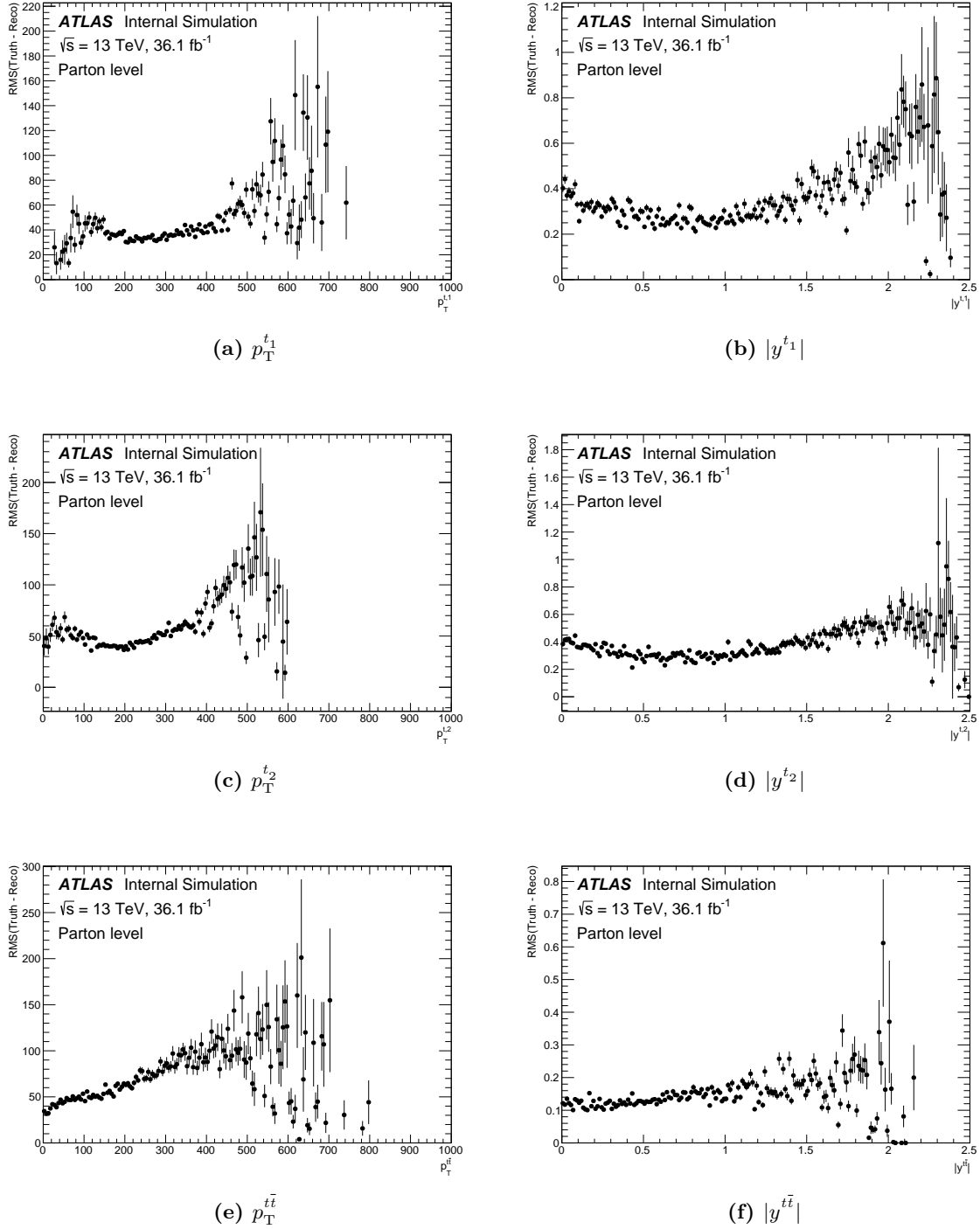
(d) $|y^{t_2}|$

(e) $p_{\mathrm{T}}^{t\bar{t}}$

(f) $|y^{t\bar{t}}|$

Figure 4.9 Resolutions for parton level observables (a) $p_{\mathrm{T}}^{t_1}$, (b) $|y^{t_1}|$, (c) $p_{\mathrm{T}}^{t_2}$, (d) $|y^{t_2}|$, (e) $p_{\mathrm{T}}^{t\bar{t}}$, and (f) $|y^{t\bar{t}}|$.

Sample	Matrix element	Parton shower	Settings	Detector	Events	$\int \mathcal{L}\,\mathrm{d}t$ [fb^{-1}]	k factor
PWG+PY8 (nominal)	POWHEG–BOX v2	PYTHIA8	A14 tune $h_{\mathrm{damp}} = 1.5\,m_t$ $\mu_{R,F} = 1.0$	GEANT4	39 885 000	54.654	1.1397
PWG+PY8 Var. Up	POWHEG–BOX v2	PYTHIA8	Var3cUp A14 tune $h_{\mathrm{damp}} = 3\,m_t$ $\mu_{R,F} = 0.5$	ATLFASTII	39 976 000	54.781	1.1397
PWG+PY8 Var. Down	POWHEG–BOX v2	PYTHIA8	Var3cDown A14 tune $h_{\mathrm{damp}} = 1.5\,m_t$ $\mu_{R,F} = 2.0$	ATLFASTII	19 995 000	27.399	1.1397
aMC@NLO+PY8	MADGRAPH5 _aMC@NLO 2.6.0	PYTHIA8	A14 tune $\mu_q = 0.5\,H_{\mathrm{T}}$	ATLFASTII	20 000 000	28.114	1.1692
SHERPA	SHERPA 2.2.1	SHERPA ME+PS@NLO	–	ATLFASTII	9 993 000	30.233	1.1484
PWG+H7	POWHEG–BOX v2	HERWIG7	H7-UE MMHT tune $h_{\mathrm{damp}} = 1.5\,m_t$	ATLFASTII	19 997 000	27.388	1.1392

Table 4.8 Generators and settings used to produce $t\bar{t}$ signal samples for the analysis presented in this chapter.

4.4 Simulation samples

$pp \to t\bar{t}$ collision events are simulated using Monte Carlo computer simulations, as described in Section 3.4. The packages and settings used for the various signal samples in this analysis are summarised in Table 4.8.

For the generation of $t\bar{t}$ events [278, 279], matrix elements are generated at next-to-leading order (NLO) using the Powheg–Box v2 (r3026) generator [280–282] with the NNPDF3.0NLO PDF set [283]. The parton shower and underlying event are generated by Pythia8 [284] using the NNPDF2.3LO PDF set [283] and the ATLAS A14 tune [285]. For comparison, a sample is also generated with the combined simulation from Sherpa 2.2.1 [286, 287].

The impact of the matrix element modelling choice is estimated with a variational sample generated with MadGraph_aMC@NLO+Pythia8 at next-to-leading order. The parameter for the shower starting scale function is set to $\mu_q = H_{\mathrm{T}}/2$ for the MadGraph_aMC@NLO [288] matrix element calculation.

To estimate the effect of the choice of parton shower algorithm, a Powheg+Herwig7 sample is generated using the same Powheg settings as for the nominal sample above. The parton showering process is simulated with Herwig7 [289] using the H7-UE tune [290] and the MMHT2014lo38cl PDF set [291].

The parameter m_t, for the t quark mass, was set to $172.5\,\mathrm{GeV}$ for all generators. The resummation damping parameter h_{damp}, which controls the p_{T} of the first additional emission beyond first order, was set to $1.5\,m_t$ for most samples. This regulates the high-p_{T} emission that the $t\bar{t}$ system recoils against in Powheg. That is, a larger value for h_{damp} leads to higher-p_{T} additional QCD radiation [281, 282]. The signal samples are filtered for only those where both W bosons from the $t\bar{t}$ decay hadronically. Additionally, $t\bar{t}$ events with at least one leptonically-decaying W boson are produced with the same settings to estimate the background contamination from misidentified leptons.

The effects of different levels of QCD radiation in the initial and final states are evaluated using variations of the renormalisation scale μ_R and the factorisation scale μ_F, in addition to h_{damp}. The renormalisation scale refers to the energy at which ultraviolet divergences are subtracted. The QCD coupling obeys the renormalisation group equation [292]

$$\mu_R^2 \frac{\mathrm{d}\alpha_S}{\mathrm{d}\mu_R^2} = \beta(\alpha_S) = -\left(b_0\alpha_S^2 + b_1\alpha_S^3 + b_2\alpha_S^4 + \dots\right), \tag{4.4}$$

where the negative sign indicates asymptotic freedom. Hence α_S decreases with increasing μ_R, so smaller values for the renormalisation scale lead to an increase in QCD radiation p_{T}. The factorisation scale μ_F parameterises the approximation that the pp collision can be factorised into partonic hard scattering process and non-perturbative intra-hadronic interactions [293]. μ_F refers to the collinear cutoff, used to avoid infrared divergences in the cross section calculation due to near-collinear gluon emissions, for example. The proton parton density functions and splitting functions are sensitive to μ_R [292, 294–296], and it also obeys a renormalisation group equation [297]. Similarly to the renormalisation scale, increasing the factorisation scale leads to a decrease in measured QCD radiation p_{T}. For the 'PWG+PY8 Var. Up' sample, both energy scales are halved, h_{damp} is increased to $3\,m_t$, and the 'Var3cUp' ATLAS A14 tune is used. For the 'PWG+PY8 Var. Down' sample, the energy scales are doubled, $h_{\text{damp}} = 1.5\,m_t$, and the 'Var3cDown' ATLAS A14 tune is used. These settings are chosen since combined variations have been shown to cover the envelope of all individual variations [298], hence they provide suitable coverage for the radiation uncertainty.

All samples are generated at NLO and normalised by a k factor to scale the cross section to NNLO+NNLL precision. This scales the simulation to the best known cross sections from theoretical predictions [299], including resummed calculations [300]. The sample size, integrated luminosity, and k factor are given for each $t\bar{t}$ sample in Table 4.8.

4.5 Background estimation

The nominal simulated signal sample indicates that it consists of approximately 5% non-hadronically decaying $t\bar{t}$ events, after the selection described in Section 4.1. Less than 2% of the total data yield originates from single t production events, but since this contamination is well within the statistical uncertainty for both data and Monte Carlo samples, it is not considered any further in the analysis.

A significant irreducible background is present in the selected data sample, originating from multi-jet production from QCD scattering events. It constitutes approximately one third of the total yield, and therefore must be subtracted from the data before unfolding to the final cross section results. The contribution from the background is estimated using a data-driven technique derived from the *ABCD method.*

4.5.1 The ABCD method

The ABCD method [301] is commonly used in high energy particle physics to determine the background contribution in a *signal region* of phase space. It is generally applicable when events can be divided into regions by two uncorrelated variables which can each provide good discrimination between signal and background events.

Data are divided into four independent regions of phase space according to values of the discriminating test variables, denoted i and j:

	j fail	j pass
i fail	A	B
i pass	C	D

The region D is the signal region, where events pass both i and j selections.

The method operates under the criterion that i and j can independently separate signal from background. This ensures that the event counts in the control regions, A, B, and C, are
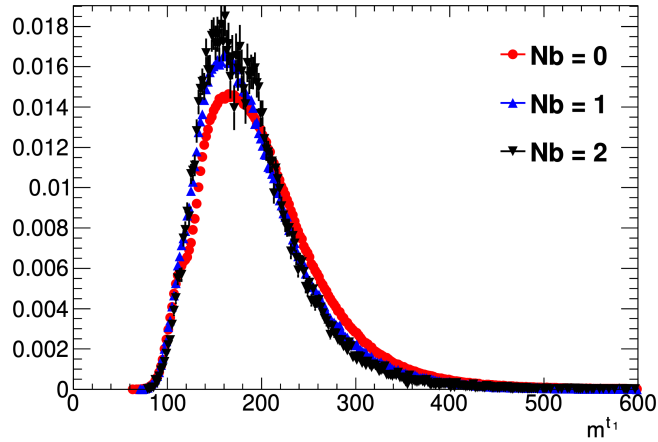
Figure 4.10 Distribution of m^{t_1} at particle level in data minus non-all-hadronic $t\bar{t}$ simulated background events, for different numbers of b-jets after all other selection criteria are applied.

independent from the count in region D. Further to this, it is also required that the i and j parameters are independently distributed across the background process. This means that the ratios of background events passing the i and j selections are independent. In other words, whether an event passes the i selection is not affected by its selection by j.

This gives the relation

$$\frac{D}{B} = \frac{C}{A}, \tag{4.5}$$

where A, B, C, and D stand for the number of events in that region of phase space after selection. Therefore the number of events in the signal region is $D = BC/A$, which can be calculated directly from data. This procedure is applied differentially for every bin in each distribution of interest.

4.5.2 Multi-jet background estimation

For the analysis presented in this chapter, the best performing pair of discriminating variables was found to be the *b-jet multiplicity* N_b, and the *t quark mass window*, (m^{t_1}, m^{t_2}). The distribution of m^{t_1} for $N_b = 0, 1, 2$ is shown in Figure 4.10. The distributions of m^{t_1} (and m^{t_2})

are very similar for $N_b = 1$ and $N_b = 2$. However for $N_b = 0$, larger differences are observed, introducing some residual correlations into the analysis of the background. Therefore a selection based on $N_b = 1, 2$ is expected to perform well, with the $N_b = 0$ region used as validation to assess the uncertainty in the estimation.

The selection on (m^{t_1}, m^{t_2}) results in three regions:

- **Tail**: If at least one of m^{t_1} or m^{t_2} is lower than $120\,\mathrm{GeV}$ or higher than $250\,\mathrm{GeV}$;

- **Peak**: If both m^{t_1} and m^{t_2} fall within the range 130 to $200\,\mathrm{GeV}$;

- **Gap**: Otherwise.

Then when combined with the selections $N_b = 0, 1, 2$, nine regions are defined:

		(m^{t_1}, m^{t_2})		
		Tail	Gap	Peak
	0	A	B	C
N_b	1	D	E	F
	2	G	H	S

This indicates that an estimate of the number of events from background contributions in region S is given by

$$S = \frac{F \times G}{D}, \tag{4.6}$$

where the background yields F, G, and D are found by subtracting the simulated non-all-hadronic $t\bar{t}$ estimate from the data in those regions. The *gap* region serves to strongly reduce the contamination by the signal in the control regions, improving the robustness of the background estimation. The yields in the gap region are small and it is not otherwise used for the background estimate. Doing this increases the statistical uncertainty of the measurement by approximately 1%, since events are removed from the control regions.

Region	N_b	(m^{t_1}, m^{t_2})	Sig. purity [%]	MC bkg. purity [%]
A	0	Tail	1.87	0.19
B	0	Gap	0.45	0.06
C	0	Peak	0.96	0.08
D	1	Tail	3.35	0.69
E	1	Gap	6.86	0.96
F	1	Peak	16.1	1.16
G	2	Tail	16.1	2.90
H	2	Gap	33.9	4.21
S	2	Peak	66.1	3.35

Table 4.9 Purities of simulated signal and non-all-hadronic $t\bar{t}$ background events in each region used for the background estimation. The remainder is estimated using the data-driven method presented in the text.

All other selection criteria described in Section 4.1 are applied and the estimate of the background yield in the signal region S is given by Equation 4.6. A parallel estimate uses regions A and C in order to assess the systematic uncertainty in the background estimate, which could arise due to correlations between the t quark kinematics and heavy flavour composition of the event. This alternative estimate is given by

$$S = \frac{C \times G}{A}. \tag{4.7}$$

The calculation and effects of the systematic uncertainty from background modelling are discussed in Section 4.7.3.

The fraction of signal and non all-hadronic $t\bar{t}$ background in each region is given in Table 4.9. As expected, the largest degree of contamination from the signal in the control regions is for $N_b > 0$.

The signal region distributions of interest for the one-dimensional particle level differential cross sections, listed in Table 4.4, are shown in Figures 4.11 to 4.16. For each observable, the distribution of events after selection of reconstruction is shown for $t\bar{t}$ all-hadronic signal and non-all-hadronic background events simulated with Powheg+Pythia8, plus the estimation of
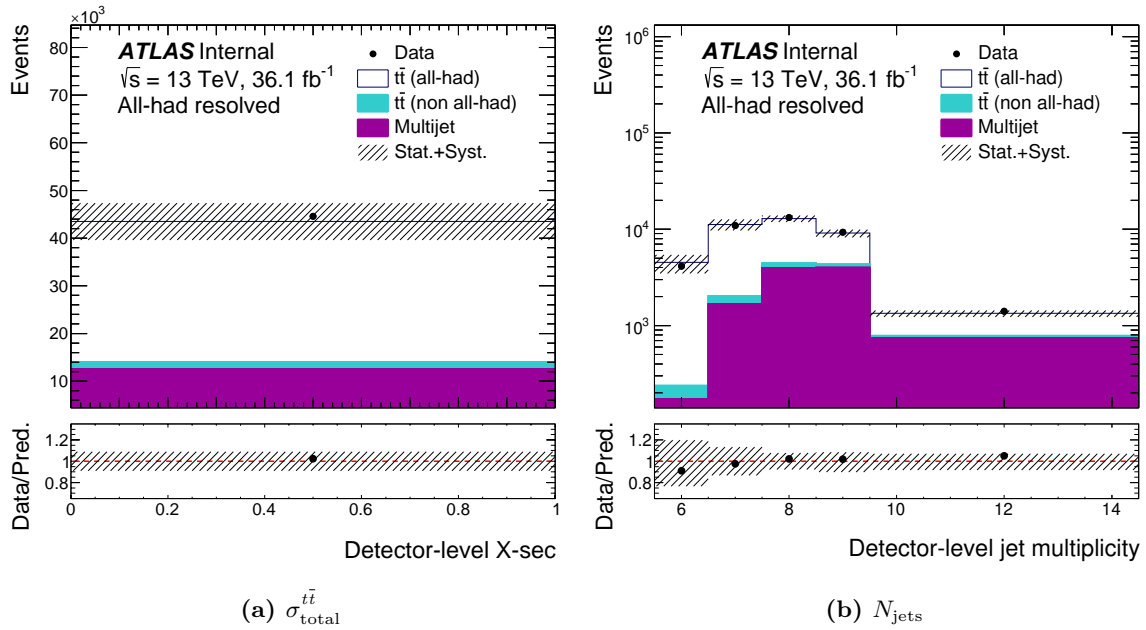
(a) $\sigma_{\mathrm{total}}^{t\bar{t}}$ **(b)** N_{jets}

Figure 4.11 Distributions in the signal region for (a) $\sigma_{\mathrm{total}}^{t\bar{t}}$ and (b) N_{jets}. The shaded area indicates the total statistical and systematic uncertainty.

the multi-jet background from the method presented in this section. These distributions are overlaid by the data yields, and good agreement within uncertainties is seen throughout. The shaded band in each plot represents the total combined statistical and systematic uncertainties, including the experimental, signal modelling, and background modelling uncertainties discussed in Section 4.7. The corresponding distributions for parton level observables, and the two-dimensional distributions at both particle and parton levels, are shown in Appendix A.

(a) $p_\mathrm{T}^{t_1}$
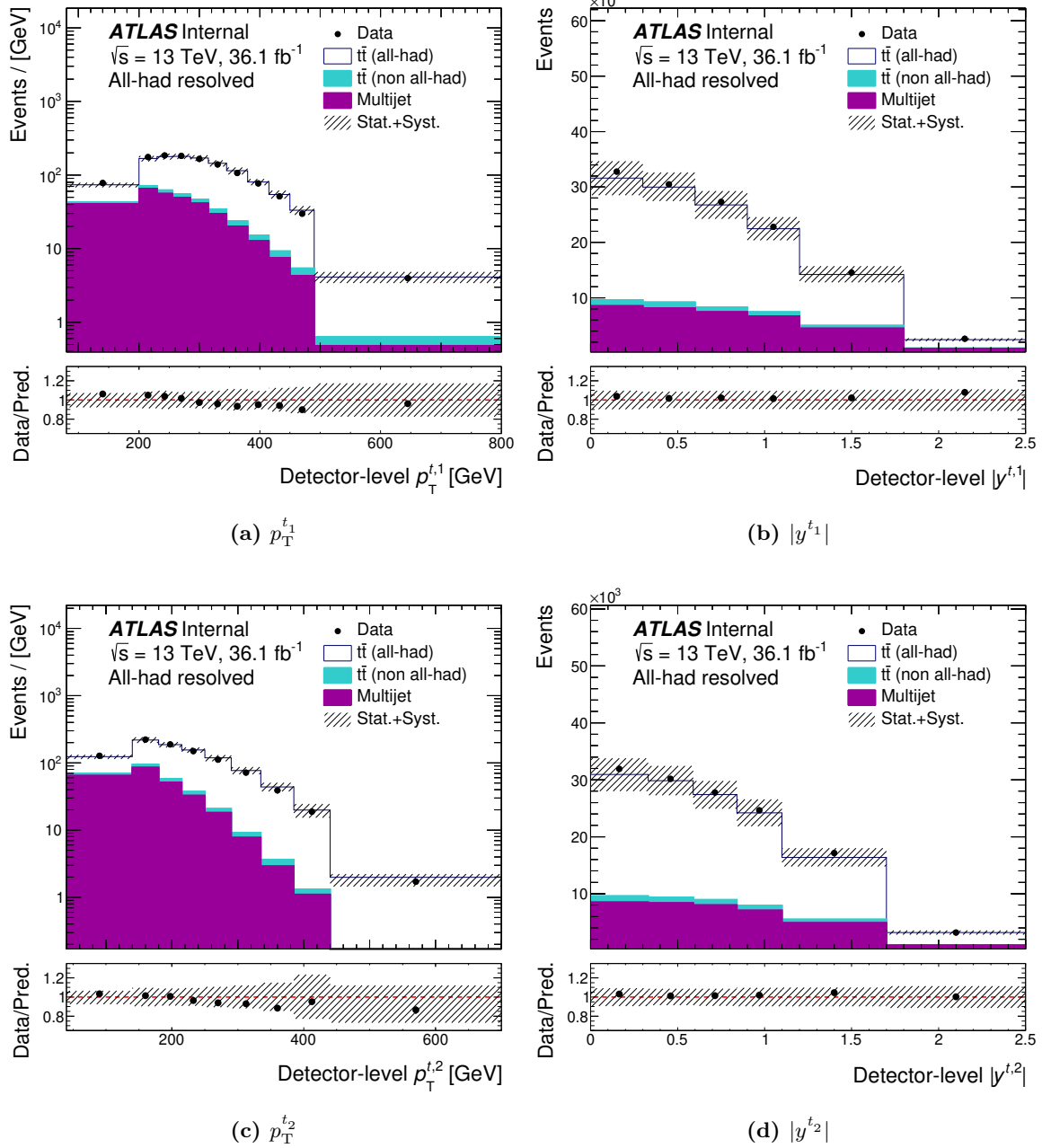
(b) $|y^{t_1}|$

(c) $p_\mathrm{T}^{t_2}$

(d) $|y^{t_2}|$

Figure 4.12 Distributions in the signal region for (a) $p_\mathrm{T}^{t_1}$, (b) $|y^{t_1}|$, (c) $p_\mathrm{T}^{t_2}$, (d) $|y^{t_2}|$. The shaded area indicates the total statistical and systematic uncertainty.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
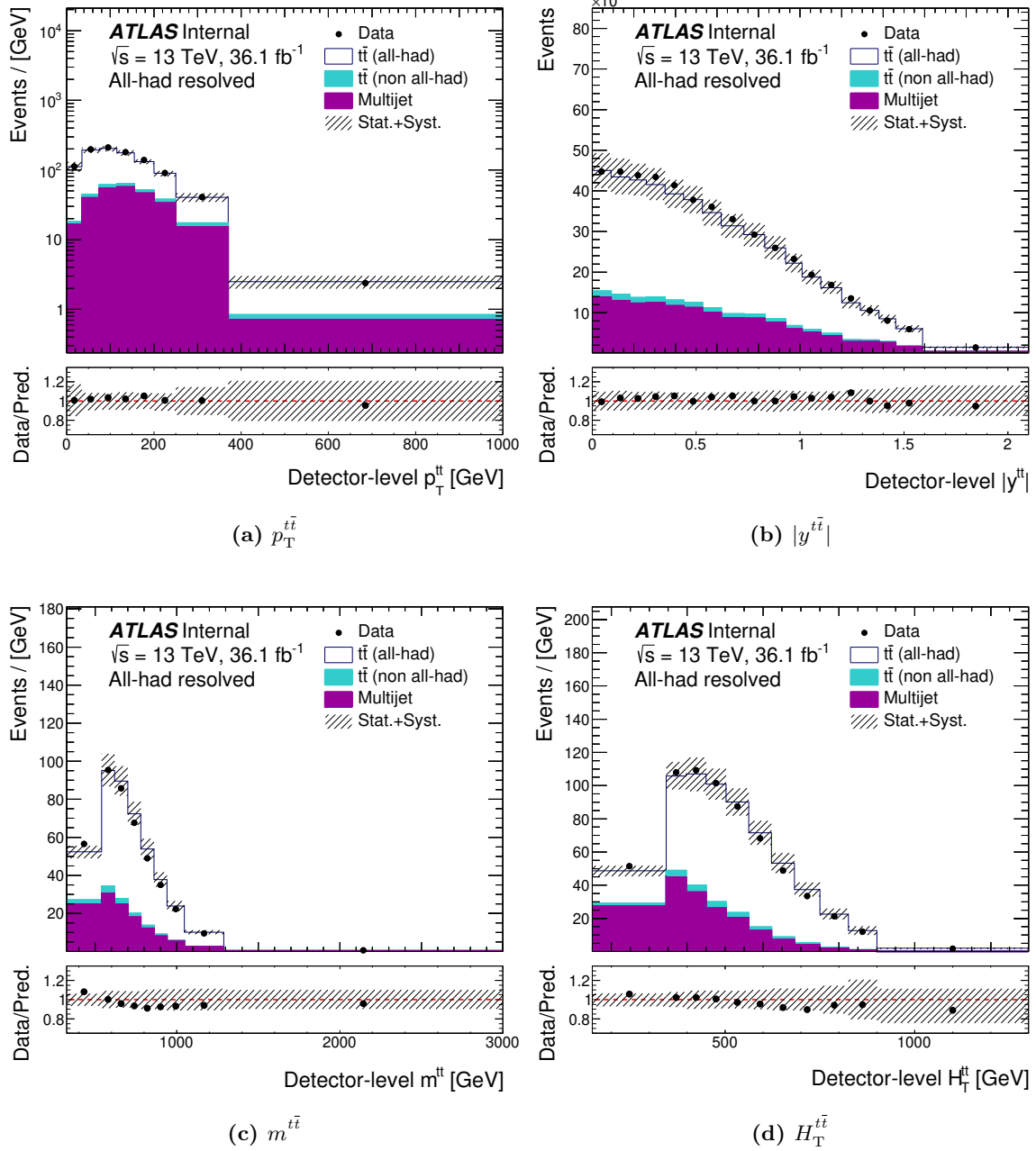
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure 4.13 Distributions in the signal region for (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, (d) $H_{\mathrm{T}}^{t\bar{t}}$. The shaded area indicates the total statistical and systematic uncertainty.

(a) $\chi^{t\bar{t}}$ **(b)** $Z^{t\bar{t}}$

(c) $\cos\theta^{\star}$ **(d)** $\Delta\phi$
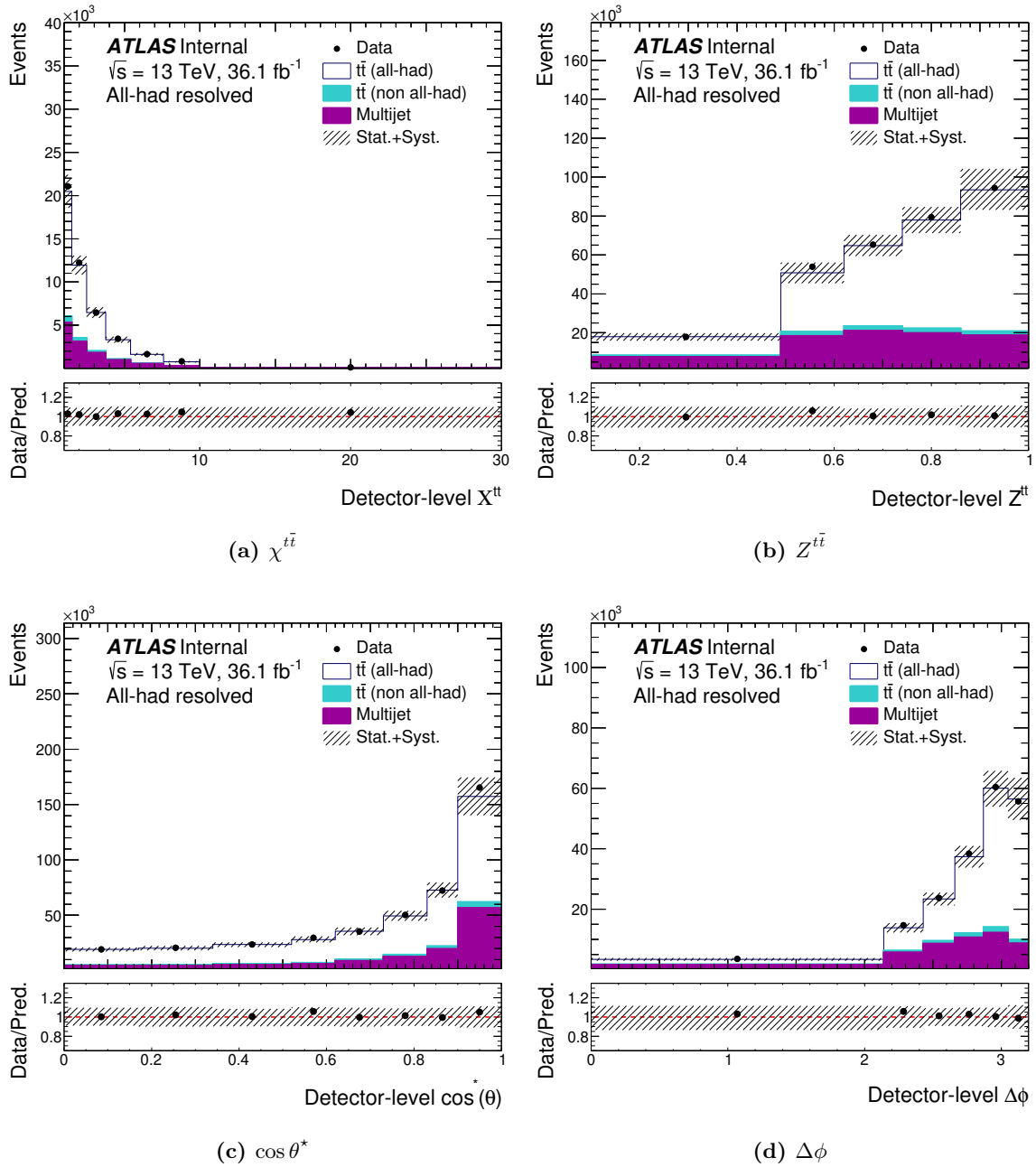
Figure 4.14 Distributions in the signal region for (a) $\chi^{t\bar{t}}$, (b) $Z^{t\bar{t}}$, (c) $\cos\theta^{\star}$, (d) $\Delta\phi$. The shaded area indicates the total statistical and systematic uncertainty.

(a) $|y_{\text{boost}}|$
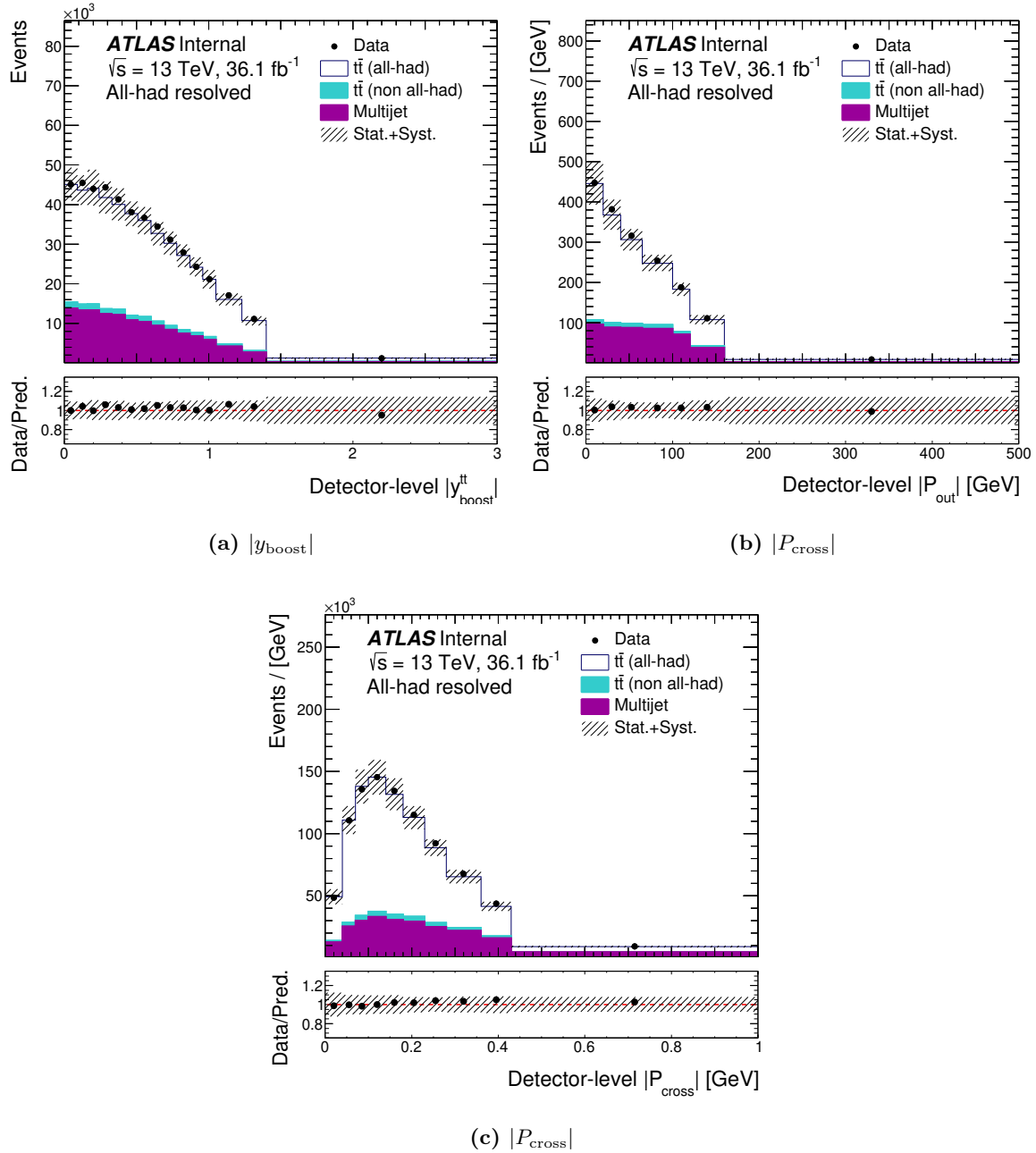
(b) $|P_{\text{cross}}|$

(c) $|P_{\text{cross}}|$

Figure 4.15 Distributions in the signal region for (a) $|y_{\text{boost}}|$, (b) $|P_{\text{cross}}|$, and (c) $|P_{\text{cross}}|$. The shaded area indicates the total statistical and systematic uncertainty.

(a) R_{Wb}^{leading}

(b) $R_{Wb}^{\text{subleading}}$

(c) R_{Wt}^{leading}
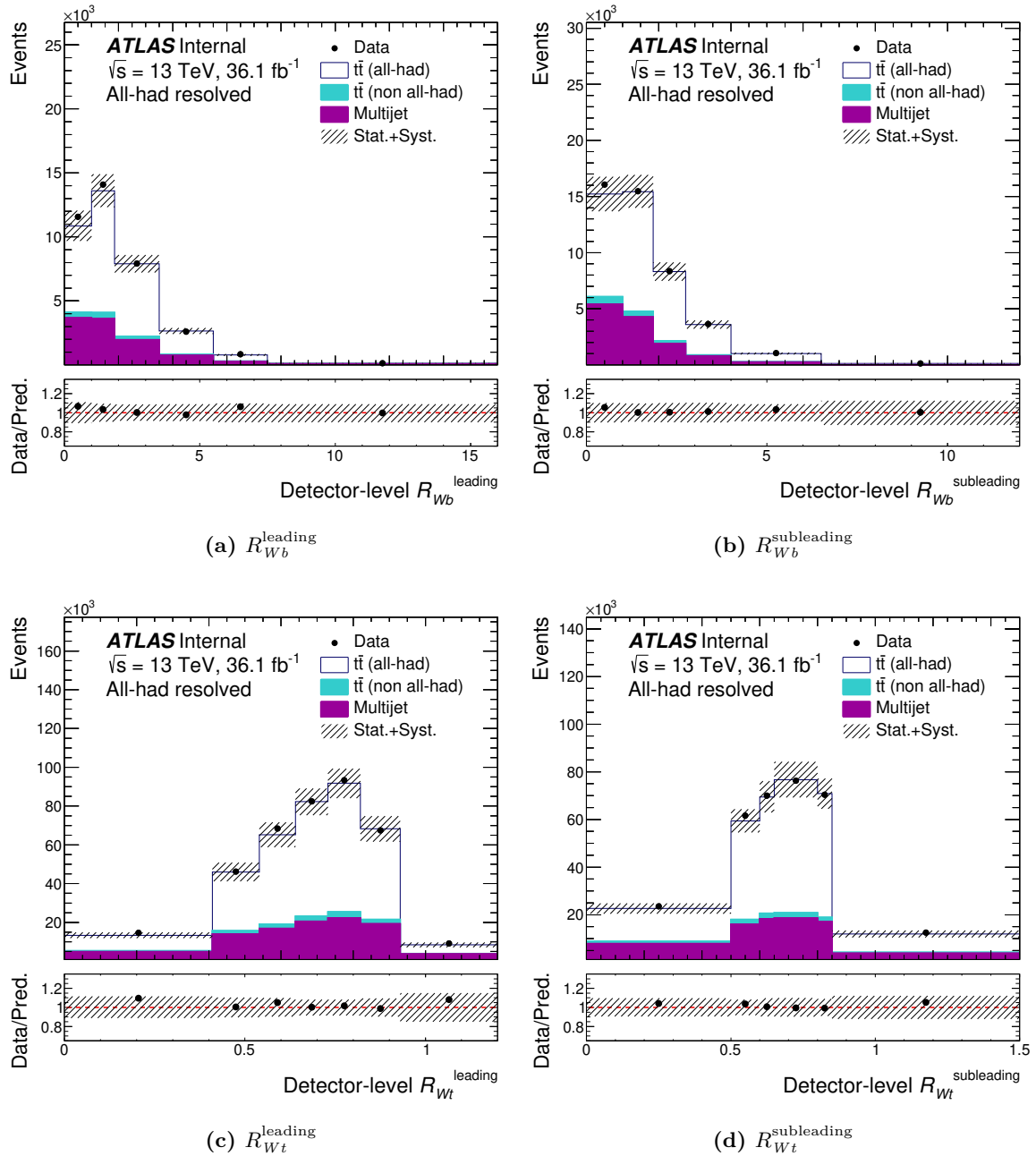
(d) $R_{Wt}^{\text{subleading}}$

Figure 4.16 Distributions in the signal region for (a) R_{Wb}^{leading}, (b) $R_{Wb}^{\text{subleading}}$, (c) R_{Wt}^{leading}, (d) $R_{Wt}^{\text{subleading}}$. The shaded area indicates the total statistical and systematic uncertainty.

4.6 Unfolding

A full mathematical treatment of unfolding is presented in detail in Chapter 6. This section gives an overview of the method used for the analysis presented in this chapter.

To enable comparisons of the measured cross sections to theoretical models and other experiments, the background-subtracted data are *unfolded* to two definitions of the t quark. At *particle level*, observables are determined from simulated values of the stable products of in-flight decays and the parton shower and hadronisation processes, before the simulation of the interaction of the particles with the detector. At *parton level*, quantities are given by their simulated values in the final state output of the transition matrix element calculation.

Due to differences in these definitions, some variables are only defined at particle level. For example, N_{jets} is not used at parton level since the number of outgoing particles is fixed, and also jets are not defined at this level. Radiative QCD emissions in the parton shower can lead to multiple jets after hadronisation at particle level, so the variable is well-defined here.

Per-bin corrections for the effects of limited acceptance and efficiency allow the unfolded spectra to be extrapolated to regions of phase space not directly accessible to the detector. At particle level, the differential cross sections are unfolded to a *fiducial phase space*, defined by cuts on the particle level objects so that their selection closely resembles that of the data at reconstruction level. This is done to reduce the impact of uncertainties introduced by extrapolating to poorly measured regions of phase space. At parton level, these selections are not made and the differential cross section is unfolded to the full phase space. This preserves generality in the parton level result, permitting direct comparisons to predictions from theory. The distributions have different numbers and locations of bins at particle and parton levels, due to the different resolutions of the observables, as shown in Section 4.3.

The unfolding procedure can be summarised in a single expression. For an observable X, the cross section in the jth bin is given by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}X_j} = \frac{1}{\mathcal{L}}\,\frac{1}{\Delta X_j}\,\frac{1}{\varepsilon_j}\sum_i \left[\mathcal{M}_{ij}^{-1}\right]^{\mathsf{T}} f_i^{\text{acc}}\left(N_i^{\text{reco}} - N_i^{\text{bkg}}\right). \tag{4.8}$$

Here $i = 1, \ldots, N$ indexes the reconstructed bin and $j = 1, \ldots, M$ indexes the unfolded (particle or parton level) bin. N_i^{reco} is the number of reconstructed events given by data and N_i^{bkg} is the corresponding estimate of the background contribution in the ith bin. $\mathcal{L}$ is the is the integrated luminosity of the dataset, and ΔX_j is the bin width.

The effects of limited detector acceptance and efficiency are accounted for by correction factors, f_i^{acc} and $1/\varepsilon_j$, respectively. The acceptance correction is applied to the background-subtracted data in each bin before unfolding to particle level only. It is determined by the ratio of the number of events passing both the particle-level and reconstruction-level selections to the number of events passing the reconstruction-level selection:

$$f_i^{\text{acc}} = \frac{N_i^{\text{reco}\wedge\text{particle}}}{N_i^{\text{reco}}}. \tag{4.9}$$

This corrects for events that are generated outside the fiducial phase space but pass the reconstruction level selection. The efficiency correction is applied after the unfolding procedure, to correct for the inefficiency of reconstruction to the fiducial and full phase spaces. It is given by the ratio of the number of events passing both the particle- or parton-level ('part.') selection and the reconstruction-level selection criteria to the number of events passing the particle- or parton-level selection:

$$\varepsilon_j = \frac{N_j^{\text{reco}\wedge\text{part.}}}{N_j^{\text{part.}}}. \tag{4.10}$$

These correction factors are evaluated using the nominal signal Monte Carlo simulation sample.

The term $\mathcal{M}_{ij}^{-1}$ is a proxy for the application of an unfolding procedure to the background-subtracted data. In lieu of a naive matrix inversion, a regularised approach is preferred in order to control the variance of the unfolded result. In the analysis presented in this chapter, the iterative Richardson–Lucy method [302–305] – a form of the *expectation maximisation* algorithm [306] – is performed by the RooUnfold [307] software package. In the kth iteration, the estimation of the ith bin population in background-subtracted data, $\hat{\nu}_i^{(k)}$, is given by the

folding equation,

$$\hat{\nu}_i^{(k)} = \sum_j R_{ij}\,\hat{\mu}_j^{(k)}, \tag{4.11}$$

where $R_{ij} = (\varepsilon_j/f_i^{\mathrm{acc}}) \times \mathcal{M}_{ij}$ is an element of the *response matrix* formed by the *migration matrix* $\mathcal{M}$ and efficiency and acceptance factors. The element $\mathcal{M}_{ij}$ gives the probability for the value of the observable of interest to be measured in the ith bin of the reconstructed histogram, given that its true value lies in the jth bin of the simulated histogram. In the unfolding step, the value of the estimated true bin population, $\hat{\mu}_j^{(k)}$, is updated by scaling by $(N_i^{\mathrm{reco}} - N_i^{\mathrm{bkg}})$:

$$\hat{\mu}_j^{(k+1)} = \sum_i \frac{R_{ij}\,\hat{\mu}_j^{(k)}}{\hat{\nu}_i^{(k)}} \left(N_i^{\mathrm{reco}} - N_i^{\mathrm{bkg}}\right). \tag{4.12}$$

For an increasing number of iterations, the observed data are better reconstructed by application of the folding through Equation 4.11. However, the corresponding unfolded distributions have very large variances and bin-to-bin correlations for a large number of iterations. This is controlled by limiting the number of iterative steps in the procedure. The total number of iterations is therefore a *regularisation parameter*. It was found that using four iterations results in unfolded differential cross sections with satisfactory properties, and the procedure is robust to statistical fluctuations. This is demonstrated in the remainder of this section. Additionally, using three or five total iterations does not significantly affect the conclusions of the analysis, indicating the unfolding is stable with respect to the regularisation parameter.

For the example of the $p_{\mathrm{T}}^{t_1}$ observable, the acceptance correction factor and efficiency when unfolding to particle level are shown in Figures 4.17a and 4.17b, respectively. The migration matrix is shown in Figure 4.17c. The same plots are shown for the $|P_{\mathrm{out}}|$ variable, unfolding to particle level, in Figure 4.18. For the two-dimensional differential cross sections, the histograms are flattened along the external axis to give a vector of values. The acceptance correction factor, efficiency, and migration matrix are shown for the unfolding to particle level of the $p_{\mathrm{T}}^{t_1}$ cross section in bins of $m^{t\bar{t}}$ in Figure 4.19, and for the $p_{\mathrm{T}}^{t_2}$ cross section in bins of N_{jets} in Figure 4.20.

(a)

(b)

(c)

Figure 4.17 (a) Acceptance correction, (b) efficiency, and (c) reconstruction-to-particle-level migration matrix for the $p_{\mathrm{T}}^{t_1}$ observable.

For unfolding to parton level, these plots are shown for the cross section in $p_{\mathrm{T}}^{t_1}$ in Figure 4.21 and for $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$ in Figure 4.22.

(a)

(b)

(c)

Figure 4.18 (a) Acceptance correction, (b) efficiency, and (c) reconstruction-to-particle-level migration matrix for the P_{out} observable.

(a)

(b)

(c)

Figure 4.19 (a) Acceptance correction, (b) efficiency, and (c) reconstruction-to-particle-level migration matrix for the $p_{\mathrm{T}}^{t_1}$ observable in bins of $m^{t\bar{t}}$.

(a)

(b)

(c)

Figure 4.20 (a) Acceptance correction, (b) efficiency, and (c) reconstruction-to-particle-level migration matrix for the $p_T^{t_2}$ observable in bins of N_{jets}.

(a)

(b)

(c)

Figure 4.21 (a) Acceptance correction, (b) efficiency, and (c) reconstruction-to-parton-level migration matrix for the $p_{\mathrm{T}}^{t_1}$ observable.
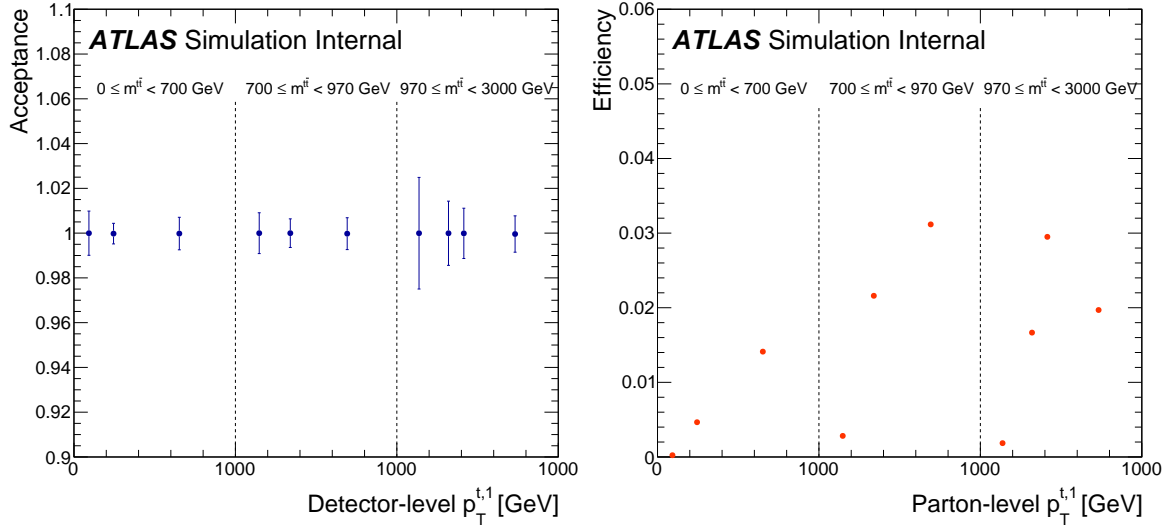
(a)

(b)
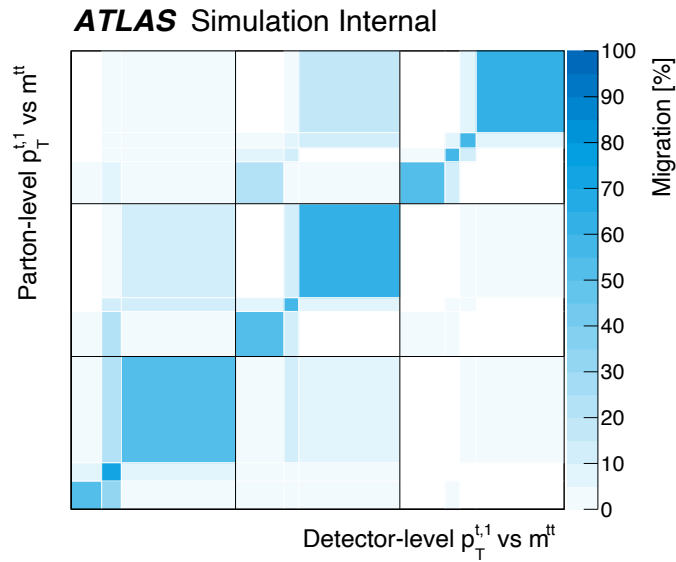
(c)

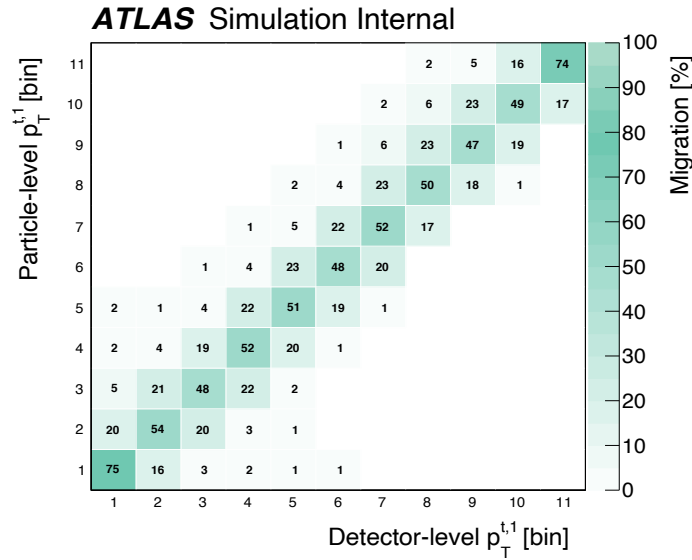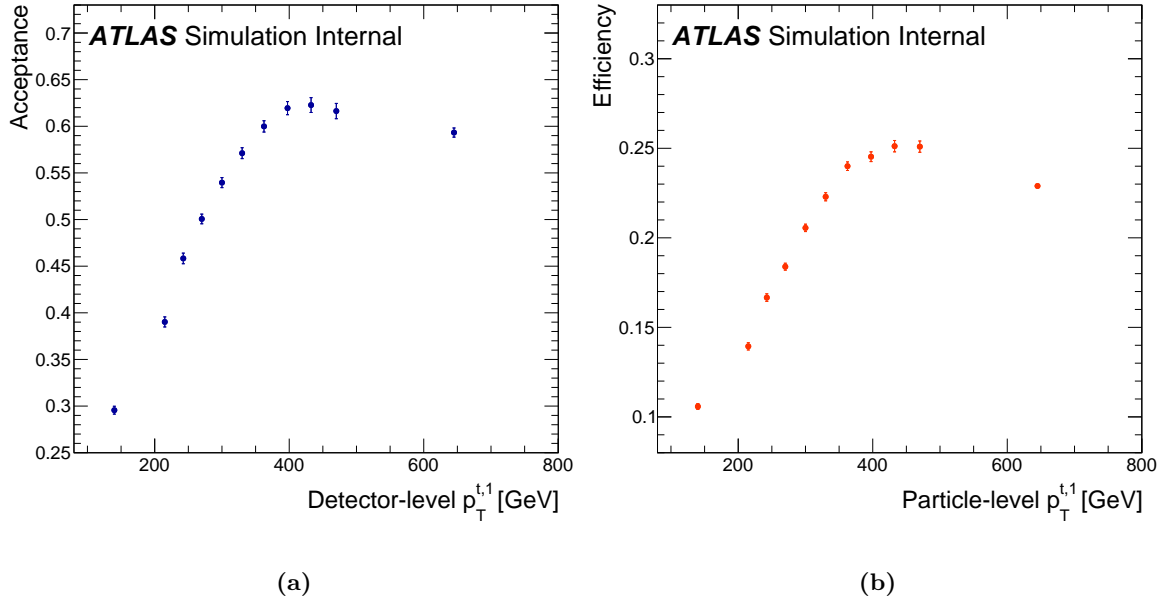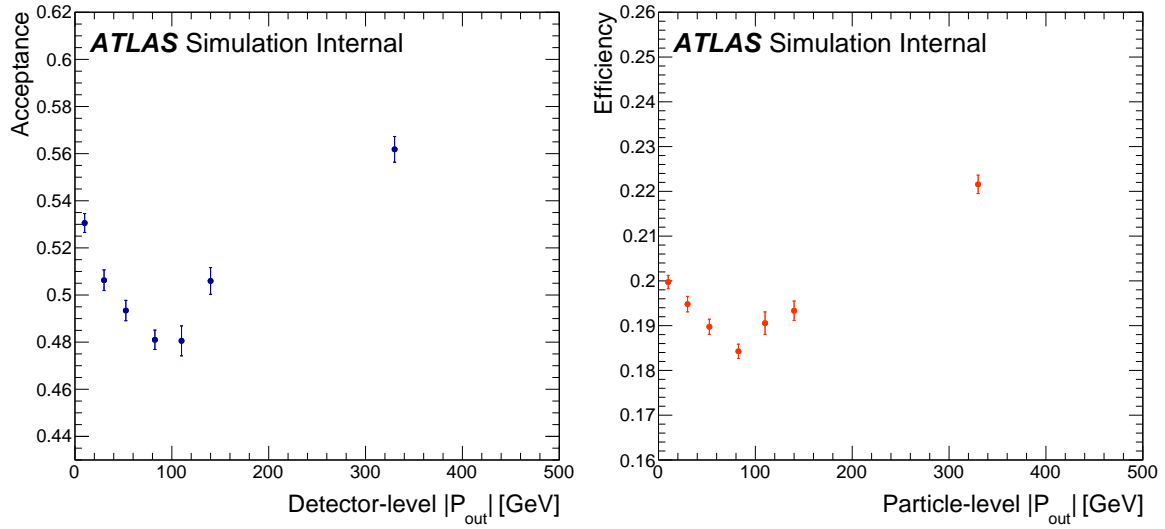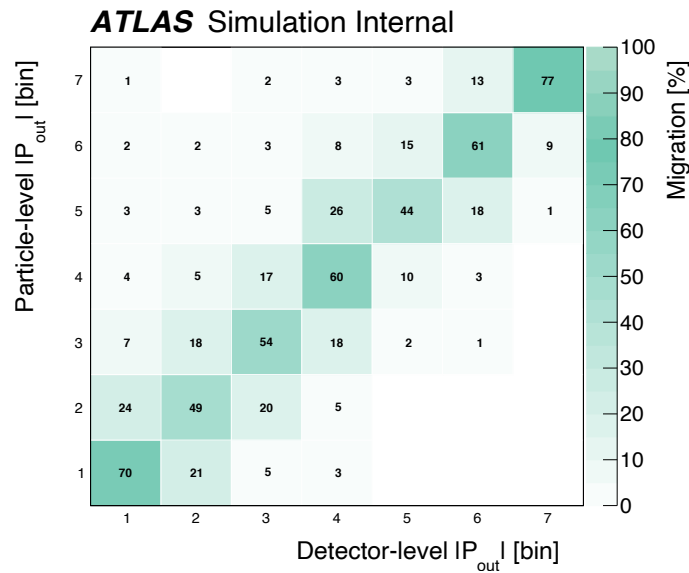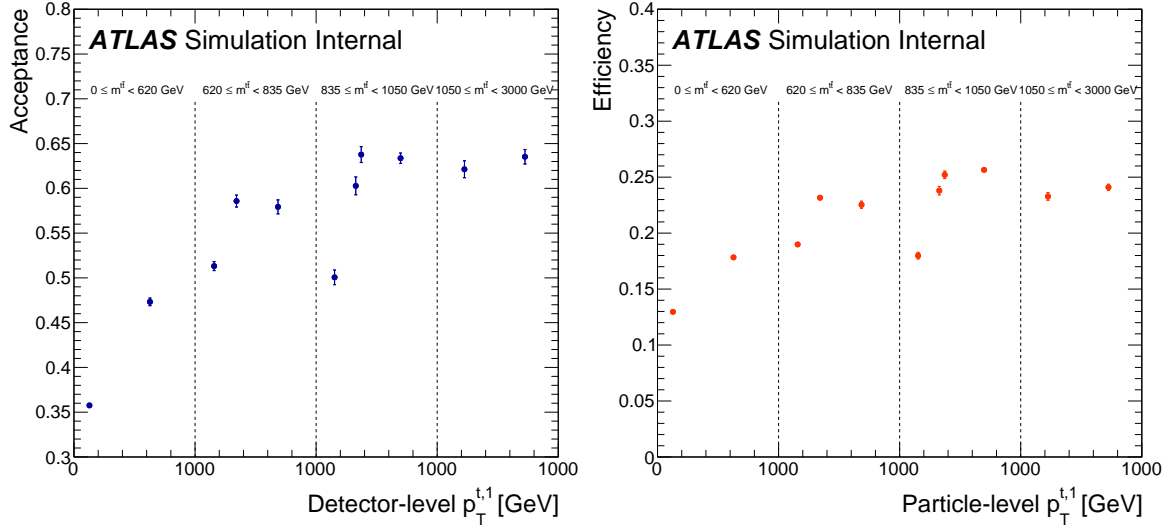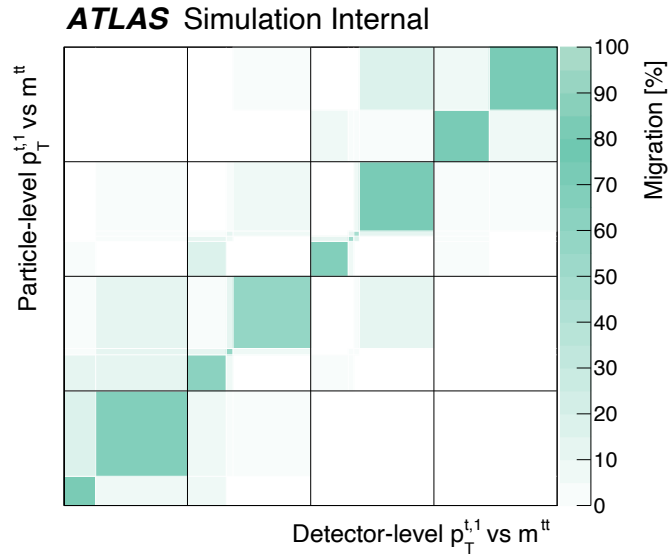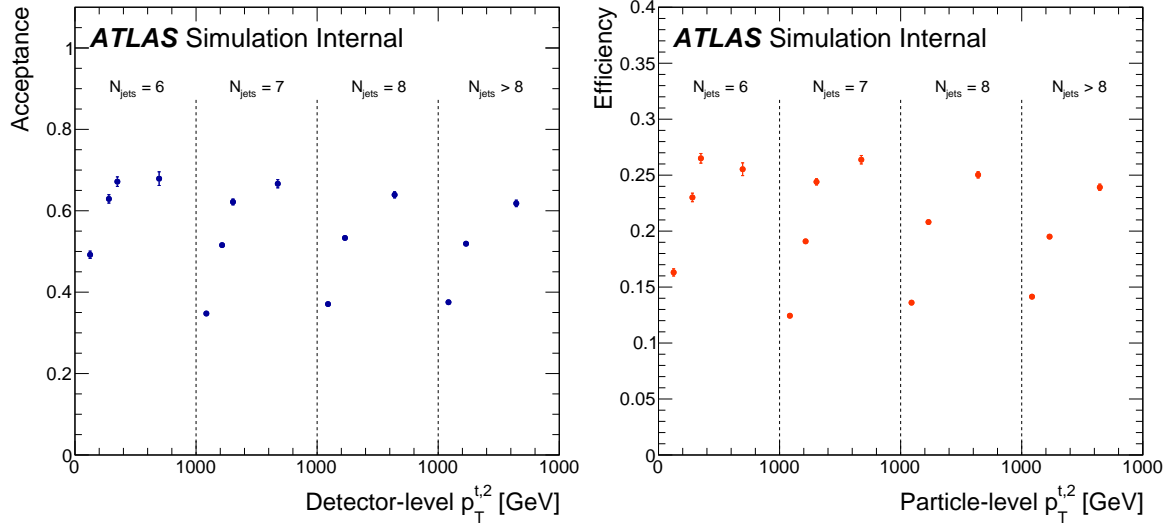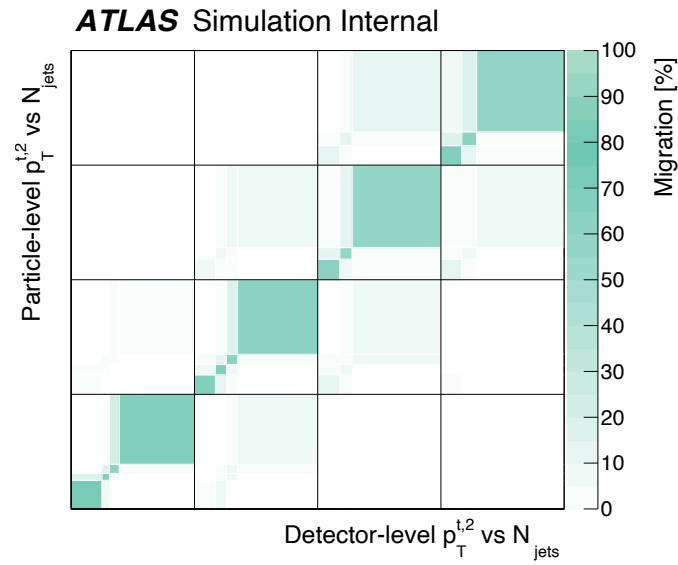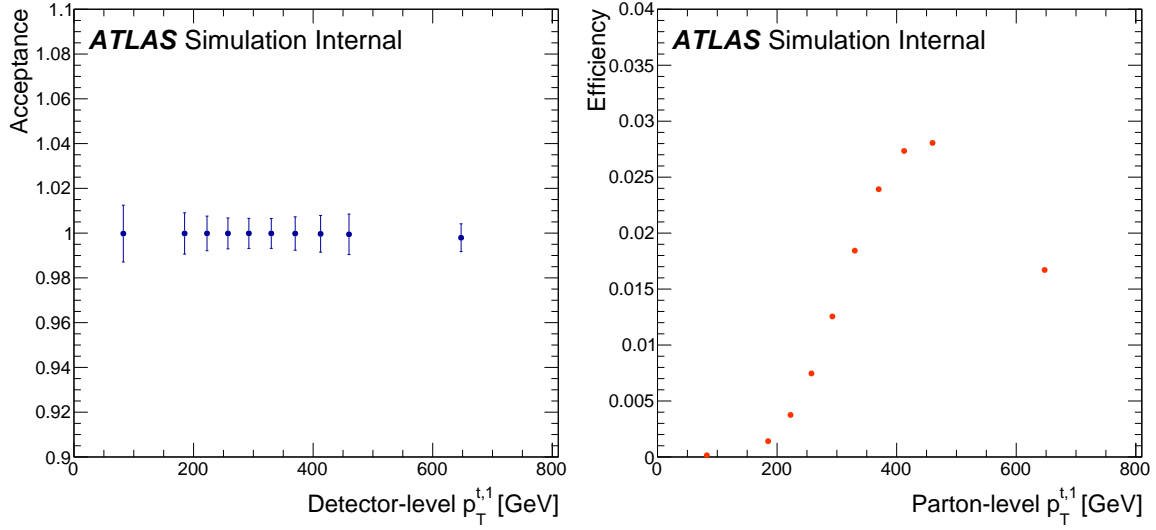**Figure 4.22** (a) Acceptance correction, (b) efficiency, and (c) reconstruction-to-parton-level migration matrix for the $p_{\mathrm{T}}^{t_1}$ observable in bins of $m^{t\bar{t}}$.
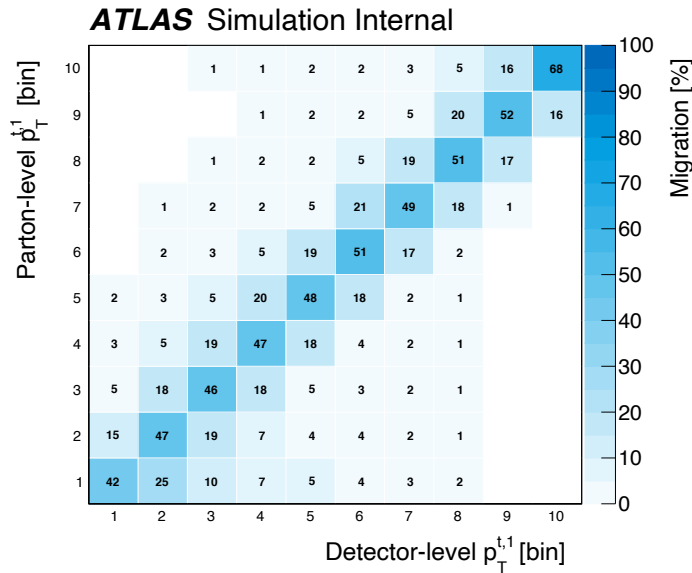
4.6.1 Closure tests

The unfolding procedure is validated and assessed for stability using *closure* and *stress* tests. For closure, it is required to be able to sufficiently recover the particle level or parton level distribution, given the reconstructed distribution, from the simulated data with which the response is formed.

First, the nominal simulated sample is split evenly into two subsamples, 'half-0' and 'half-1'. The response, formed of the acceptance correction factors, efficiencies, and migrations, is determined from half-1 by comparing the simulated detector-level distribution to the distribution at particle or parton level. The half-0 subsample is then regarded as pseudodata and passed through the unfolding mechanism described above, applying the corrections from half-1. A χ^2 test statistic is calculated using the statistical uncertainties of the samples. If the corresponding *p*-value is not less than 0.05, the procedure is declared robust. For $p < 0.05$, the bin edges are manually adjusted (moved or merged) where there is most disagreement and the procedure is repeated.

The unfolded pseudodata and particle level distribution is shown for $p_{\mathrm{T}}^{t_1}$ in Figure 4.23a and for $|P_{\mathrm{out}}|$ in Figure 4.23b. They are also shown for $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$ in Figure 4.23c, and for $p_{\mathrm{T}}^{t_2}$ in bins of N_{jets} in Figure 4.23d. The same plots for the parton level unfolding are shown for $p_{\mathrm{T}}^{t_1}$ and $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$ in Figure 4.24. After appropriate adjustments to the bins, the closure for all 1D distributions unfolded to particle level give to a mean *p*-value of 0.57 for absolute cross sections. For 1D distributions unfolded to parton level, the mean *p*-value is 0.66. Correspondingly, the mean *p*-values are 0.59 and 0.67 for for absolute 2D cross sections unfolded to particle and parton level, respectively. In the case of normalised cross sections, the level of per-bin agreement is very similar, but the reduced number of degrees of freedom lowers the *p*-values by approximately 4% to 7%. No final *p*-values are below 0.06. The unfolding procedure, including responses and regularisation, is therefore considered robust in terms of closure. This means it can appropriately recover particle level and parton level distributions from data.

(a) $p_{\mathrm{T}}^{t_1}$

(b) $|P_{\mathrm{out}}|$

(c) $p_{\mathrm{T}}^{t_1}$ vs. $m^{t\bar{t}}$
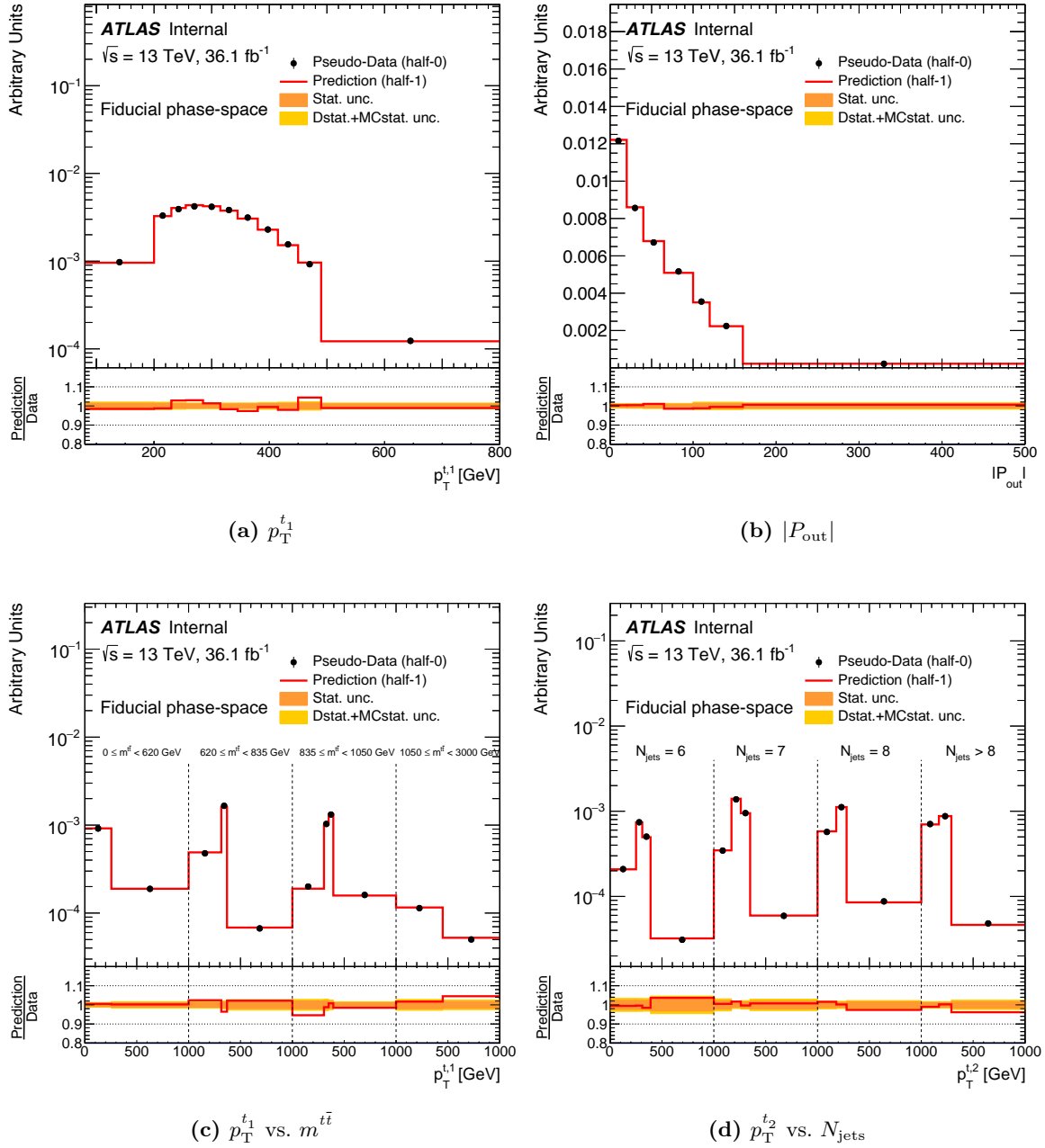
(d) $p_{\mathrm{T}}^{t_2}$ vs. N_{jets}

Figure 4.23 Closure tests for (a) $p_{\mathrm{T}}^{t_1}$, (b) $|P_{\mathrm{out}}|$, (c) $p_{\mathrm{T}}^{t_1}$ vs. $m^{t\bar{t}}$, and (d) p_{T}^{t} vs. N_{jets}, unfolded to particle level.
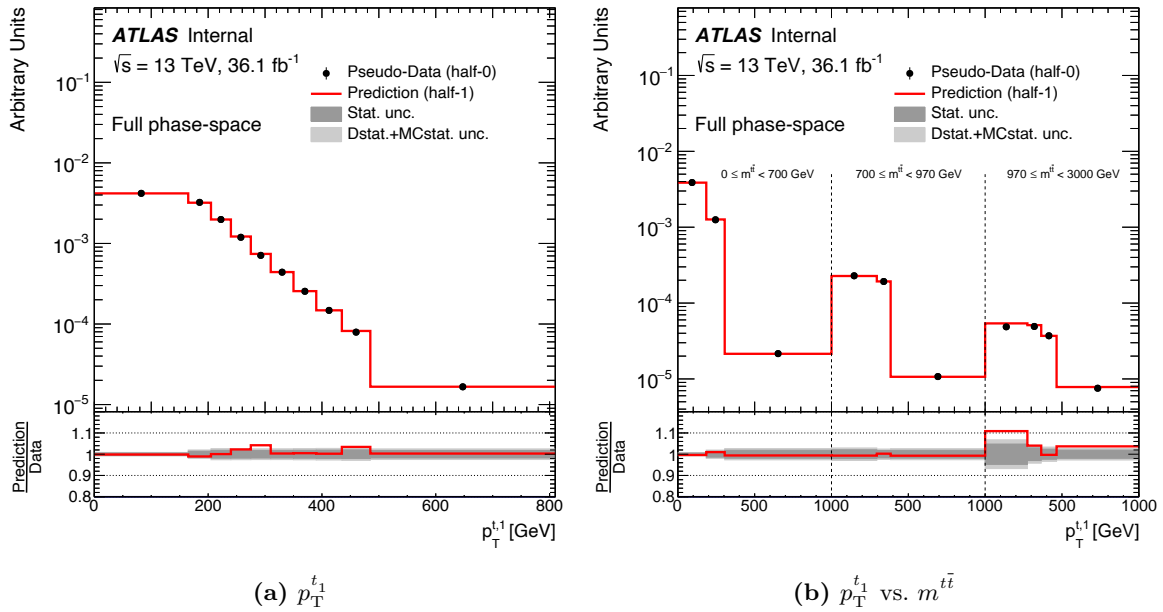
(a) $p_{\mathrm{T}}^{t_1}$ **(b)** $p_{\mathrm{T}}^{t_1}$ vs. $m^{t\bar{t}}$

Figure 4.24 Closure tests for (a) $p_{\mathrm{T}}^{t_1}$ and (b) $p_{\mathrm{T}}^{t_1}$ vs. $m^{t\bar{t}}$, unfolded to parton level.

4.6.2 Stress tests

Stress tests are performed to assess the degree of bias introduced by the unfolding regularisation. Similarly to the closure tests described above, half of the nominal simulation sample is used to calculate the ingredients for the unfolding procedure, and the other half is considered as pseudodata. Only the nominal simulation sample is used, so that the effects of bias in the unfolding procedure are isolated. The pseudodata are reweighted to change the shape of the simulated observed distribution. This distribution is then unfolded and the result is compared to the reweighted simulated particle or parton level simulated distribution. The disagreement between these distributions provides an indication of the bias in the unfolding.

The reweighting factor for each bin in a distribution is a multiple k of the ratio of the reconstructed bin populations from pseudodata to simulation,

$$f_i = k \times \frac{N_i^{\mathrm{pseudodata}}}{N_i^{\mathrm{MC\ truth}}}. \tag{4.13}$$

This is done with $k = -5, -3, -1, 1, 3, 5$. It is expected that the closure becomes worse at larger values of $|k|$, since the bias is amplified.

The simulated stressed distributions and pseudodata unfolded to particle level are shown for $p_{\mathrm{T}}^{t_1}$ in Figure 4.25, $|P_{\mathrm{cross}}|$ in Figure 4.26, $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$ in Figure 4.27, and for $p_{\mathrm{T}}^{t_2}$ in bins of N_{jets} in Figure 4.28. They are also shown for the unfolding to parton level for $p_{\mathrm{T}}^{t_1}$ in Figure 4.29 and for $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$ in Figure 4.30. This procedure confirms that the unfolding is able to successfully recover distributions that are different from those used to form the response.

4.6.3 Bump insertion

The ability of the unfolding procedure to reconstruct an unexpected deviation from the simulation is also tested. The regularisation used in the unfolding could smooth out these deviations, which should be able to be resolved by the analysis. To do this, the $m^{t\bar{t}}$ distribution is reweighted by a factor

$$f_i = 1 + k \, \exp\left[-\frac{(m^{t\bar{t}} - m_0)^2}{2\sigma^2} \right] \tag{4.14}$$

for the ith bin, where $m_0 = 900 \, \mathrm{GeV}$, $\sigma = 60 \, \mathrm{GeV}$ and $k = -0.35$ are chosen for this test. This inserts a negative bump into the pseudodata around $m^{t\bar{t}}$, which should be reconstructed in the unfolded distribution.

The unfolded pseudodata, stressed according to Equation 4.14, and the original and stressed simulated $m^{t\bar{t}}$ spectra are shown at particle level in Figure 4.31a and at parton level in Figure 4.31b. In both cases, the inserted bump is preserved through the unfolding.

4.7 Systematic uncertainties

The measured differential $t\bar{t}$ production cross section results shown in Section 4.8 are affected by several sources of systematic uncertainties. The methods used to estimate these are described in this section. The dominant sources of uncertainty are related to the modelling of the detector

(a) $k = 1$

(b) $k = 3$

(c) $k = 5$
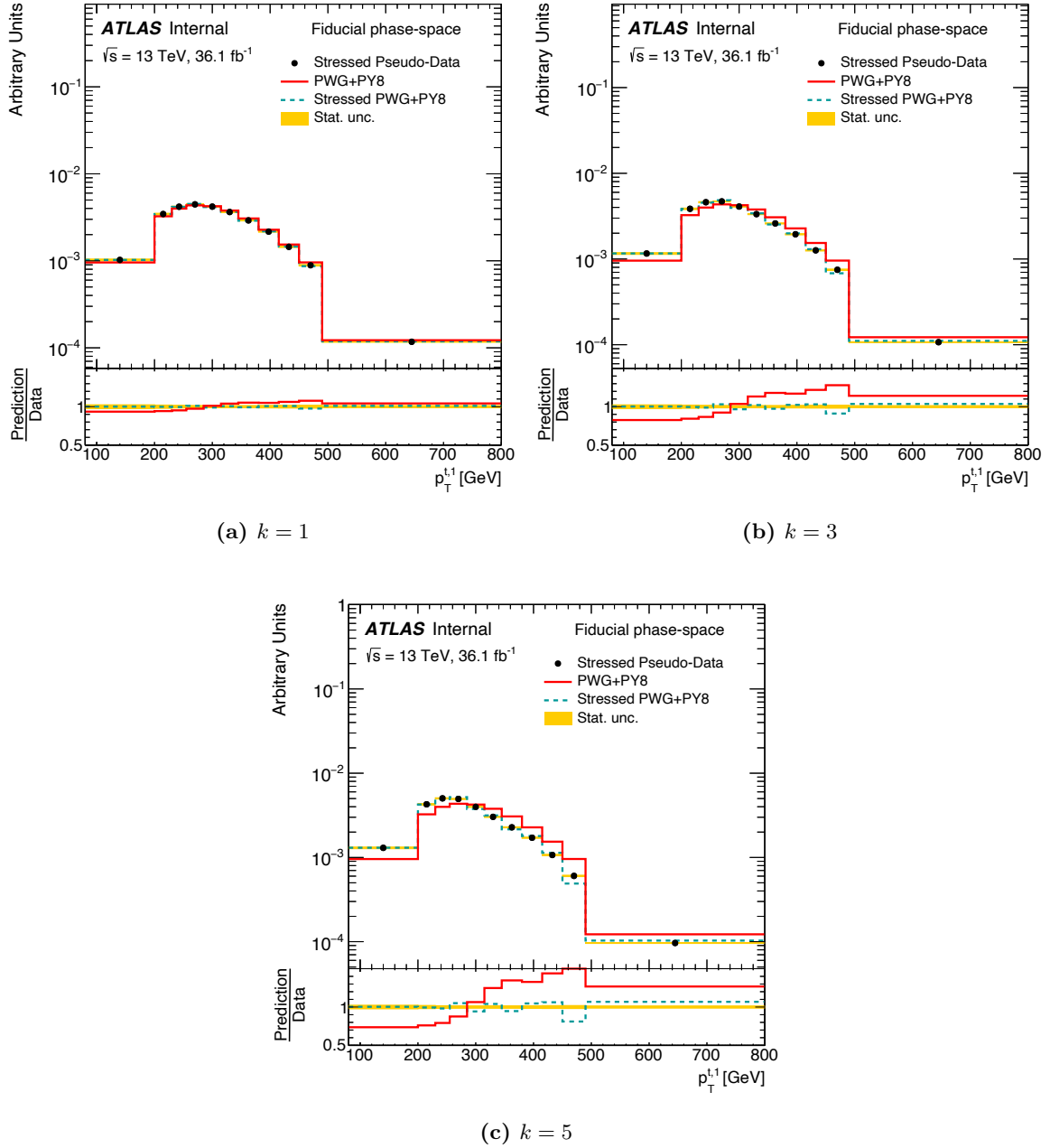
Figure 4.25 Stress tests for the unfolding of $p_{\mathrm{T}}^{t_1}$ to particle level with stress factors (a) 1, (b) 3, (c) 5.

(a) $k = 1$

(b) $k = 3$

(c) $k = 5$
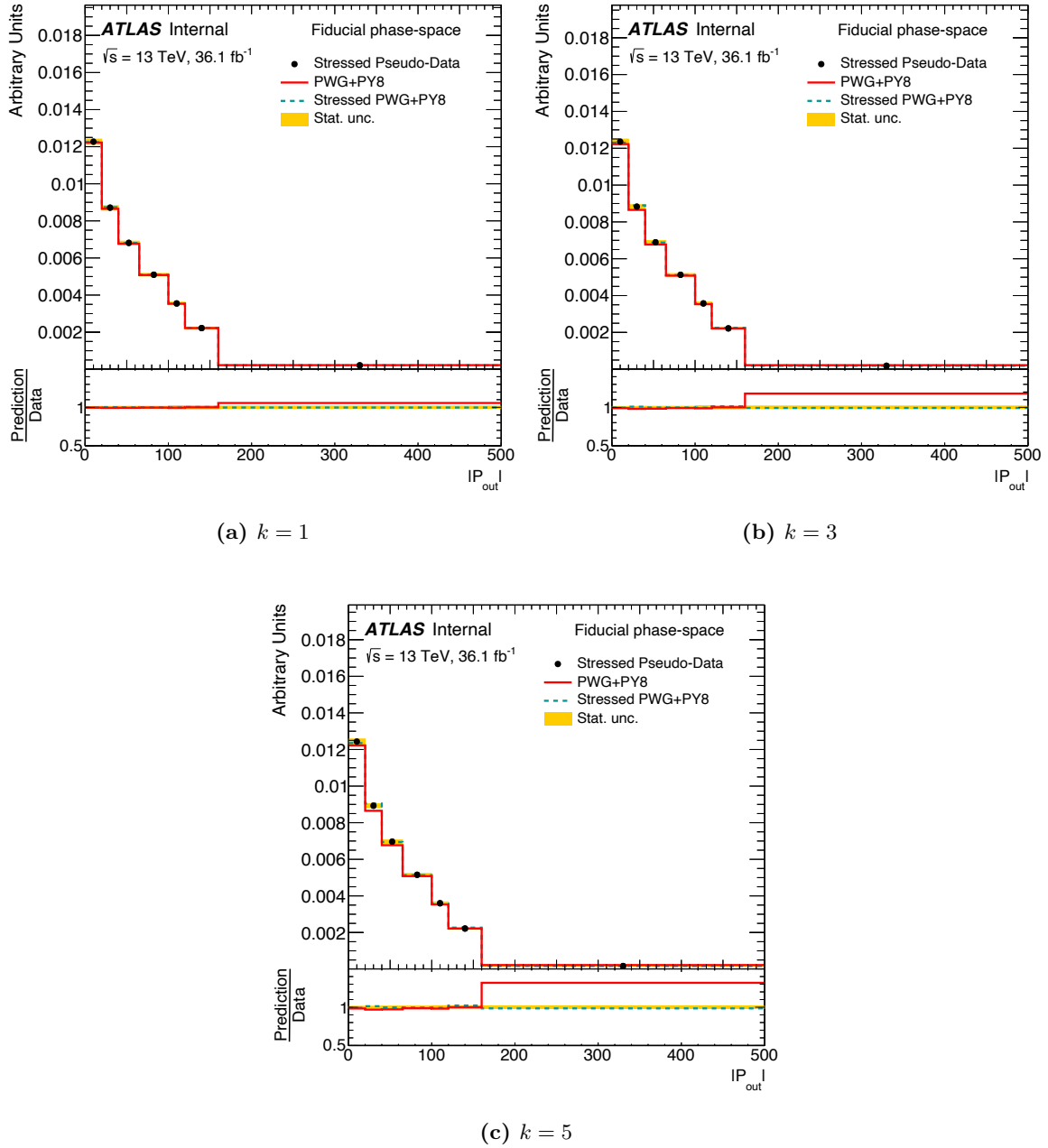
Figure 4.26 Stress tests for the unfolding of $|P_{\text{out}}|$ to particle level with stress factors (a) 1, (b) 3, (c) 5.

(a) $k = 1$

(b) $k = 3$

(c) $k = 5$
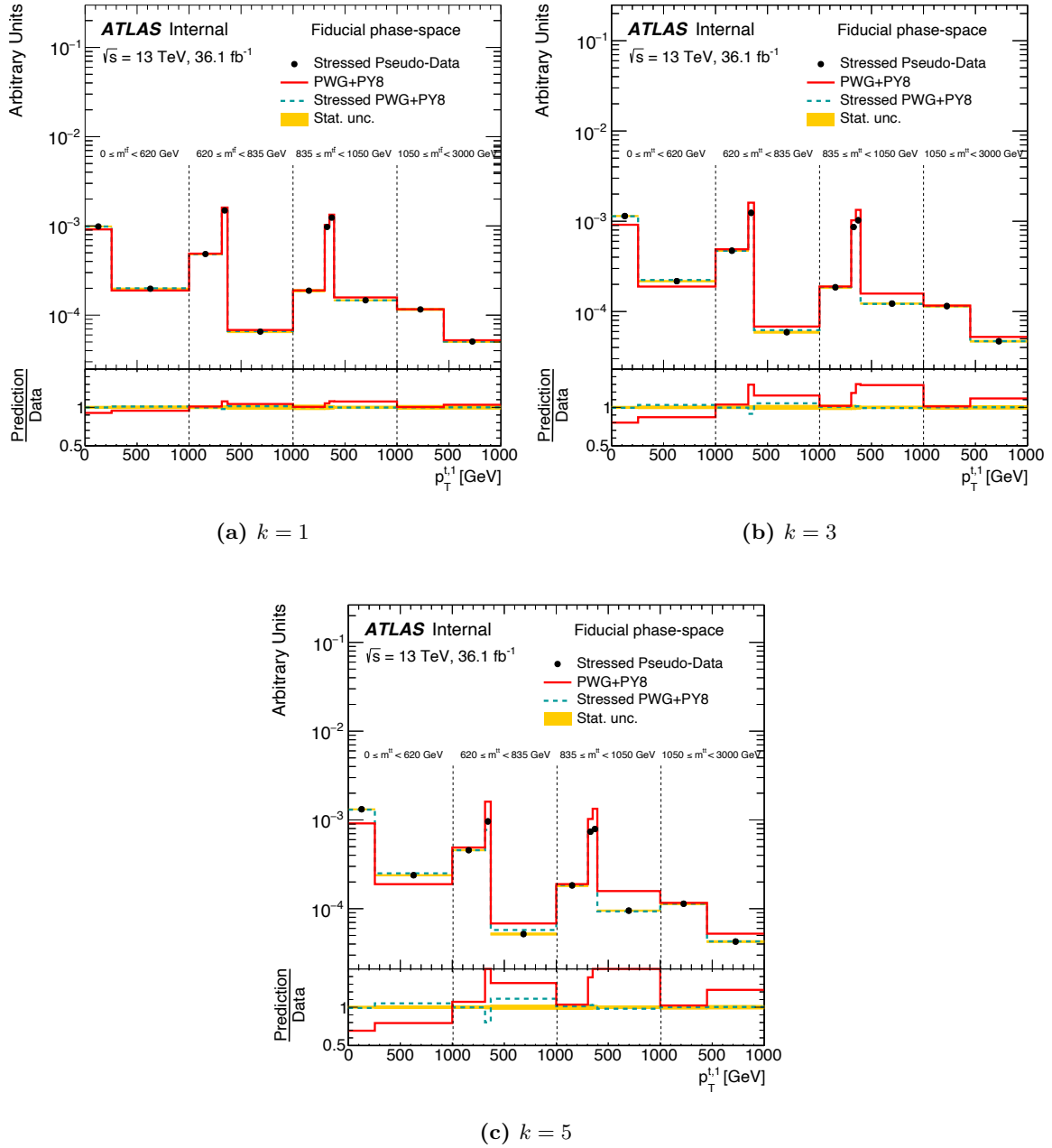
Figure 4.27 Stress tests for the unfolding of $p_{\mathrm{T}}^{t_1}$ vs. $m^{t\bar{t}}$ to particle level with stress factors (a) 1, (b) 3, (c) 5.

(a) $k = 1$

(b) $k = 3$

(c) $k = 5$
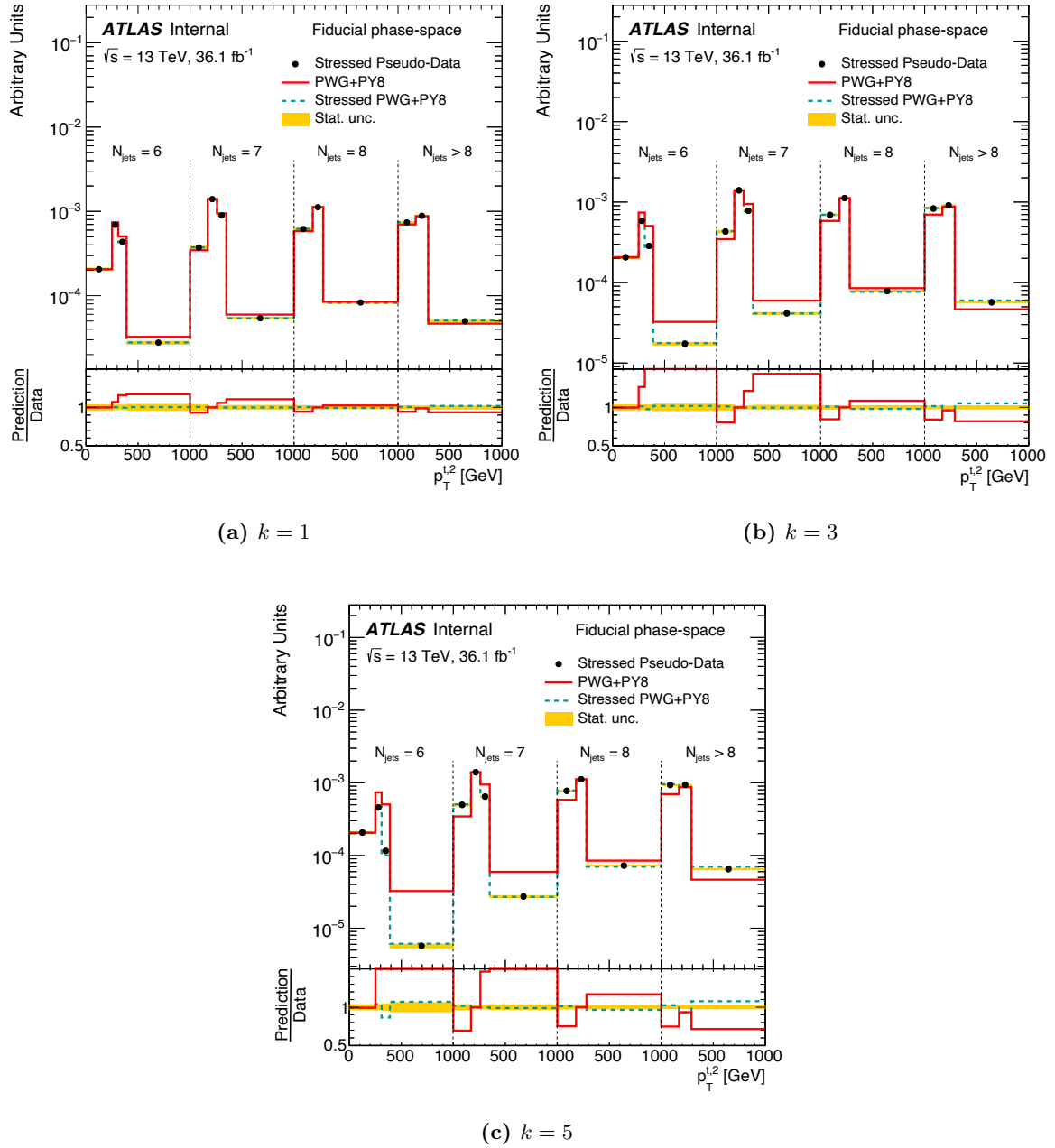
Figure 4.28 Stress tests for the unfolding of $p_T^{t_2}$ vs. N_{jets} to particle level with stress factors (a) 1, (b) 3, (c) 5.

(a) $k = 1$

(b) $k = 3$

(c) $k = 5$
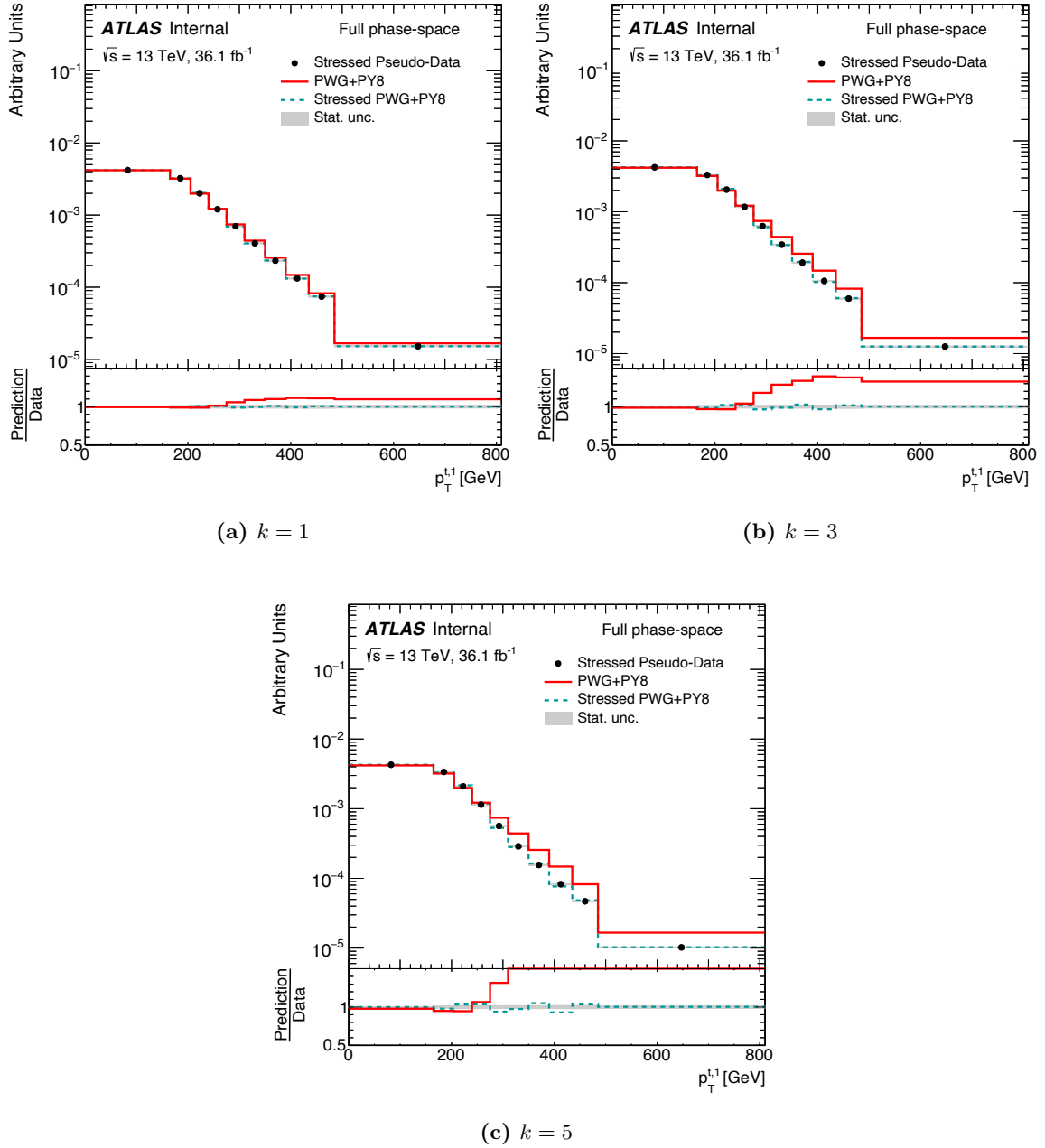
Figure 4.29 Stress tests for the unfolding of $p_{\mathrm{T}}^{t_1}$ to parton level with stress factors (a) 1, (b) 3, (c) 5.

(a) $k = 1$

(b) $k = 3$

(c) $k = 5$
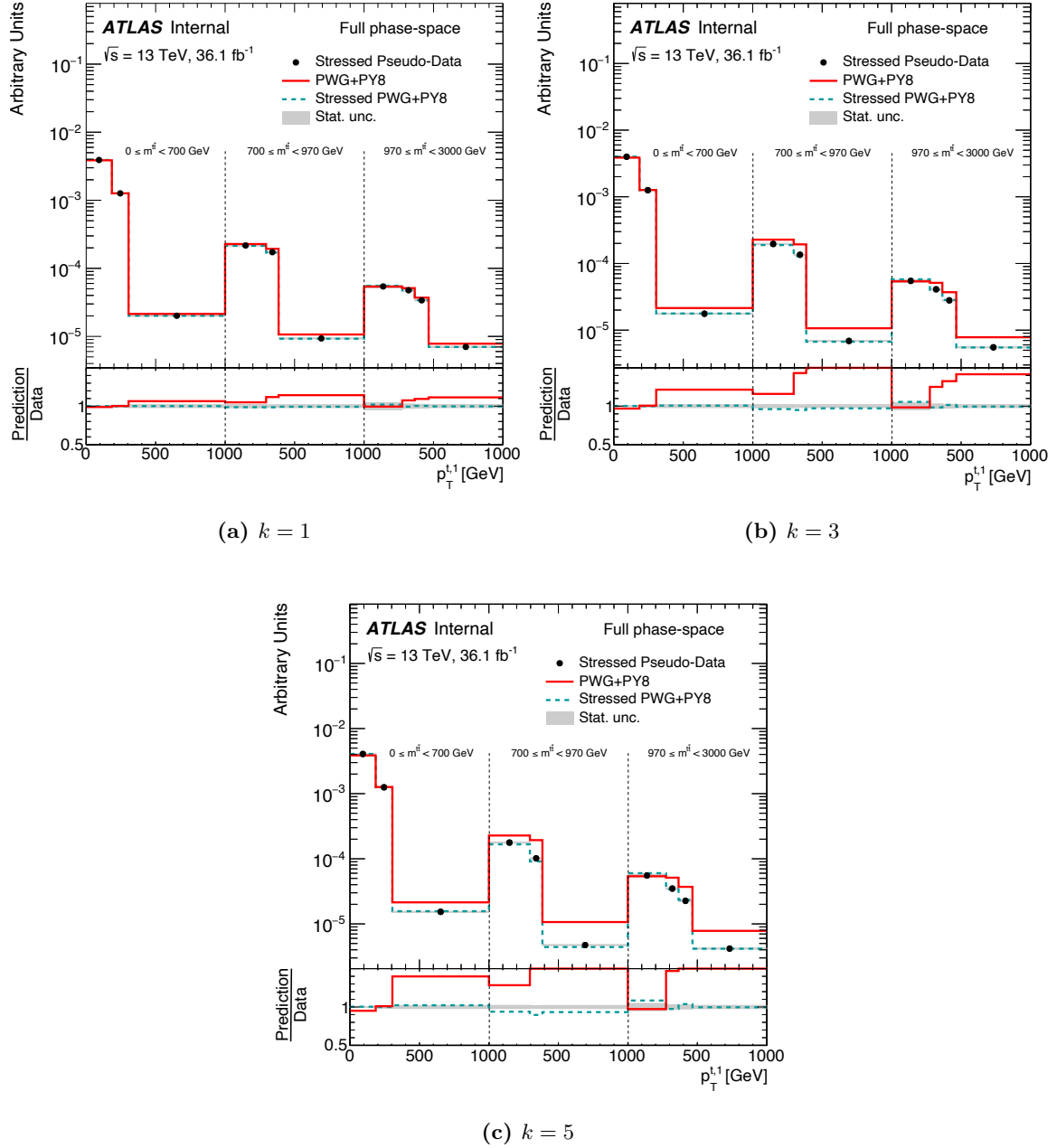
Figure 4.30 Stress tests for the unfolding of $p_T^{t_1}$ vs. $m^{t\bar{t}}$ to parton level with stress factors (a) 1, (b) 3, (c) 5.

(a) Particle level   (b) Parton level
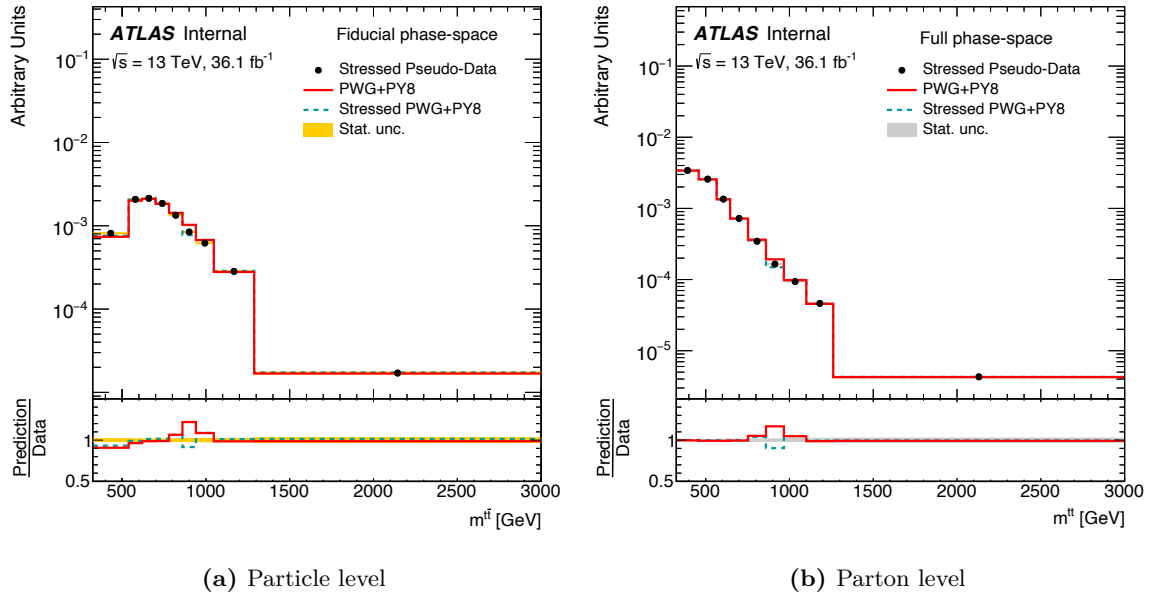
Figure 4.31 Bump tests for the unfolding in $m^{t\bar{t}}$ to (a) particle level and (b) parton level.

and the simulation of signal and background processes. Systematic uncertainties from unfolding are found to have less impact.

For each source of systematic uncertainty, one or two alternative Monte Carlo samples are generated using the fast detector simulation described in Section 3.4. An overview of the samples used is given in Section 4.4. The deviation from the nominal value is recorded for each bin in a distribution. The absolute systematic variation is given by the difference between the number of events in the bin for the systematic variation sample and the nominal sample, $\Delta N = N_{\text{syst.}} - N_{\text{nominal}}$. Where two alternative systematic samples are available, the deviations are generally asymmetric and recorded as 'up' and 'down' variations. When only a single alternative sample is used, the systematic uncertainty is taken as the symmetric deviation from the nominal value.

Each source of systematic uncertainty is evaluated before and after performing the unfolding procedure described in Section 4.6, using the acceptance correction factors, efficiencies, and migrations derived from the nominal signal sample with fast detector simulation. The unfolded distributions are compared to the 'true' distributions (at particle or parton level) for the

alternative sample. The relative systematic uncertainty is then given by the symmetrised relative difference between the unfolded and true distributions.

4.7.1 Experimental uncertainties

Systematic uncertainties arising from the experimental method and modelling of the detector in the simulation are assessed. For detector modelling, they are evaluated in each case by swapping the signal and non-all-hadronic $t\bar{t}$ background samples with the systematic variation samples. The data-driven estimate of the multi-jet background is then recalculated for each variation using the method described in Section 4.5.

Luminosity

The relative uncertainty of the integrated luminosity of the dataset collected with ATLAS in 2015 and 2016 is 2.9%, taken from data collected with van der Meer scans [308, 309]. In this procedure, each beam position is scanned across the transverse plane and the distribution of observed luminosity versus beam position recorded.

Trigger

No uncertainties are typically assigned for the trigger in analyses that work in the plateau region of the turn-on curve, shown in Figure 4.1 for the analysis presented in this chapter. For this analysis, the multi-jet trigger efficiency ranges from approximately 94% at the threshold $p_{\mathrm{T}} > 55\,\mathrm{GeV}$, to approximately 98% in the high p_{T} region. Since this difference in efficiency is far smaller than the other effects mentioned in this section, no uncertainty is evaluated. It is expected that the impact of including such an uncertainty is negligible.

Jet reconstruction

The systematic contributions from the uncertainty of the jet energy scale are estimated by varying the energies of jets according to uncertainties derived from simulation, test beam data,

(a) (b)
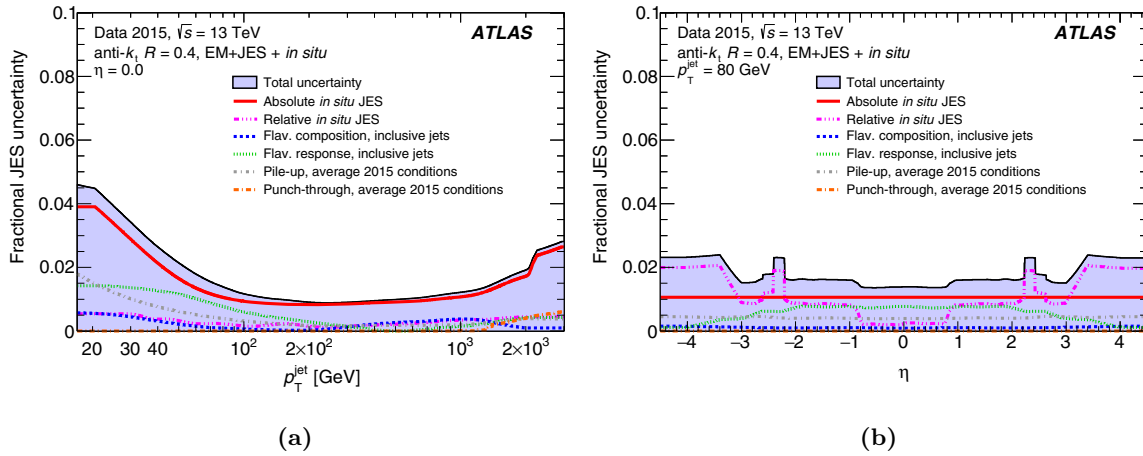
Figure 4.32 Combined systematic uncertainty from the jet energy scale of calibrated jets in ATLAS. The contributions from categories of components are shown as a function of (a) jet transverse momentum p_T at $\eta = 0$, and (b) pseudorapidity η at $p_T = 80\,\text{GeV}$. Data from pp collisions at $\sqrt{s} = 13\,\text{TeV}$ taken in 2015 were used. From [310].

and *in-situ* calibration measurements [244, 310]. Contributions from jet flavour composition, η-intercalibration, punch-through, single-particle response, calorimeter response to different jet flyovers, and pile-up are taken into account, giving more than 80 separate correlated systematic uncertainty terms [310]. The contributions to the fractional jet energy scale uncertainty are shown for the categories as functions of jet p_T and pseudorapidity in Figure 4.32. A reduced set of 29 near-independent components is used in this analysis, derived from the principal components of eigenbasis decompositions of the set of nuisance parameters within each category [311–313]. Since the analysis presented in this chapter is performed in the all-hadronic final state containing many jets, the resulting differential cross section measurements are dominated by the systematic uncertainty arising from the determination of the jet energy scale, at approximately 5–10% combined, except for in statistically-limited bins.

The uncertainty due to the difference in jet energy resolution between the data and simulated events is calculated by applying a smearing to the simulated jet p_T according to the resolution for p_T and η [314]. The relative uncertainty from the jet energy resolution is found to be approximately 1%.

The systematic uncertainty in the track-based tagging of jets from pileup events is evaluated by randomly discarding events according to the difference between the data and simulation [315].

b-jet tagging

The systematic uncertainties associated with tagging b-jets are separated by the b-jet tagging efficiency, the c-jet tagging efficiency, and the misidentified light flavour jet tagging efficiency. The efficiencies are all estimated from data and parameterised as functions of p_{T} and η [316]. A systematic uncertainty is assigned for each correction applied to correct the efficiency differences between data and the simulation. The uncertainties in the simulation modelling of the b-jet tagging performance are assessed by studying b-jets in dileptonic $t\bar{t}$ events. While the systematic uncertainty from the light flavour jet tagging efficiency correction is generally $< 1\%$, the b-jet tagging efficiency systematic uncertainty can be as large as 5%.

Lepton reconstruction

Uncertainties related to the modelling of the lepton energy and momentum scales and resolution in simulation are estimated from $Z \rightarrow ee/\mu\mu$, $J/\psi \rightarrow ee/\mu\mu$, and $W \rightarrow e\mu$ processes [235, 237, 317]. Since the efficiency of the zero lepton selection is very high, the e, μ, and leptonically decaying τ reconstruction, trigger, and identification are found to contribute negligibly to the total systematic uncertainties in the results.

The uncertainties relating to the reconstruction of hadronically decaying τ leptons is estimated from simulations. Only the τ-jet energy scale uncertainties are calculated and are found to be negligibly small.

4.7.2 Signal modelling

The choice of simulation used to model the all-hadronic $t\bar{t}$ signal process affects the kinematic properties of the $t\bar{t}$ system and its decay products. It also affects the object reconstruction efficiencies and indirectly the estimate of the non-all-hadronic $t\bar{t}$ background.

Matrix element

The uncertainty due to the choice of the matrix element generator is evaluated by unfolding the MadGraph5_aMC@NLO+Pythia8 sample using the acceptance correction factors, efficiencies, and migrations from the Powheg+Pythia8 nominal sample. The unfolded spectra are compared to the particle and parton level spectra from the systematic variation sample and the difference gives the systematic uncertainty.

Parton shower

The uncertainty due to the choice of the parton shower model is determined by unfolding the Powheg+Herwig7 sample, again using the ingredients from the nominal sample. The deviation between the unfolded and true varied spectra is taken as the systematic uncertainty contribution from the choice of parton shower model. The resulting systematic uncertainties are found to exhibit strong dependence on the spectrum and bin. In extreme cases, the relative uncertainties are as large as 30%, although they are 1–5% in most bins.

Initial- and final-state radiation

The level of initial- and final state radiation (IFSR) from QCD effects affects the distribution of N_{jets} in addition to the kinematics of the $t\bar{t}$ system. In order to evaluate the uncertainty due to the choice of IFSR modelling, simulated $t\bar{t}$ samples with modified settings are used. The renormalisation, factorisation, and h_{damp} parameters for these 'Var. Up' and 'Var. Down' samples are shown in Table 4.8. In each case, the unfolded spectrum from the nominal sample is compared to the unfolded spectrum from the variation sample and the difference is taken as the corresponding side of the systematic uncertainty. For most bins, the total IFSR uncertainty is at the level of a few percent, comparable in magnitude to the parton shower uncertainty.

Parton distribution functions

The systematic uncertainty due to the choice of PDF set is assessed using the thirty PDF sets provided by PDF4LHC15 [318]. The choice of PDF impacts the acceptance correction factors, efficiencies, and migrations used in the unfolding. Therefore the uncertainty from the PDF choice is assessed by unfolding the nominal $t\bar{t}$ signal sample using these ingredients reweighted according to the difference in spectra. The resulting relative systematic uncertainty is found to be $< 1\%$, with some excesses at 1–2% in statistically limited bins.

Monte Carlo sample size

The limited size of the nominal simulated signal sample contributes a sampling uncertainty to the unfolded differential cross section results. To estimate the size of this uncertainty, 10 000 alternative samples are generated by sampling from a joint Gaussian with mean given by the nominal estimated bin counts with their corresponding independent variances. The smeared spectra are unfolded and the sample standard deviation of their unfolded distributions is used as the systematic uncertainty due to limited Monte Carlo sample size. The resulting uncertainty is found to be $< 0.5\%$ in most cases, peaking at 1–2% in bins with low population.

4.7.3 Background modelling

The data-driven estimation of the multi-jet background, detailed in Section 4.5, contributes two sources of uncertainty. The first is a statistical effect, stemming from the limited number of events used to perform the estimation. This uncertainty is accounted for in the same way as the Monte Carlo sample size uncertainty above.

The second component of uncertainty is from the assumption of independence used in the ABCD-inspired method. To calculate this, the same background estimation procedure is repeated using the alternative ratio, given by $S = CG/A$ from Equation 4.7. This makes use of the regions with $N_b = 0$. The systematic error is then taken to be the difference in the unfolded distributions for the usual and alternative estimates of the multi-jet background.
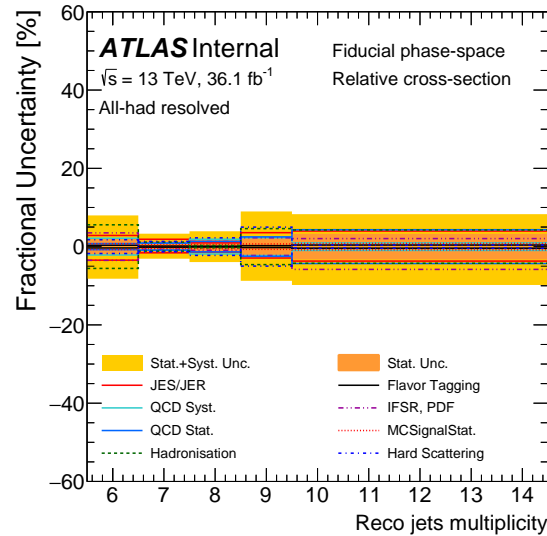
Figure 4.33 Uncertainty composition for the normalised cross section as a function of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

The relative systematic uncertainty from background modelling is calculated by this method to be usually $< 5\%$, and very often smaller than the jet energy scale uncertainty. In some rare cases, it is found to be the largest uncertainty. This is in regions where the signal purity is very low, amplifying the effect of the background modelling in the sources of systematic uncertainty.

4.7.4 Total uncertainty composition

For the event yields passing selection, the total uncertainties are given in Table 4.2. The symmetrised total relative uncertainty is approximately 7.6% for the $t\bar{t}$ all-hadronic signal sample, 8.8% for the $t\bar{t}$ non-all-hadronic background, and 15% for the multi-jet background. The symmetrised total relative uncertainty on the total Monte Carlo yield is 6.6%.

The compositions of the total uncertainty for the one-dimensional relative differential cross sections unfolded to particle level are shown in Figures 4.33 to 4.38. The uncertainty breakdowns for all other cross sections (relative 2D particle level, relative 1D parton level, relative 2D particle level, absolute 1D particle level, absolute 2D particle level, absolute 1D parton level, absolute 2D particle level) are shown in Appendix B.

(a) $p_T^{t_1}$
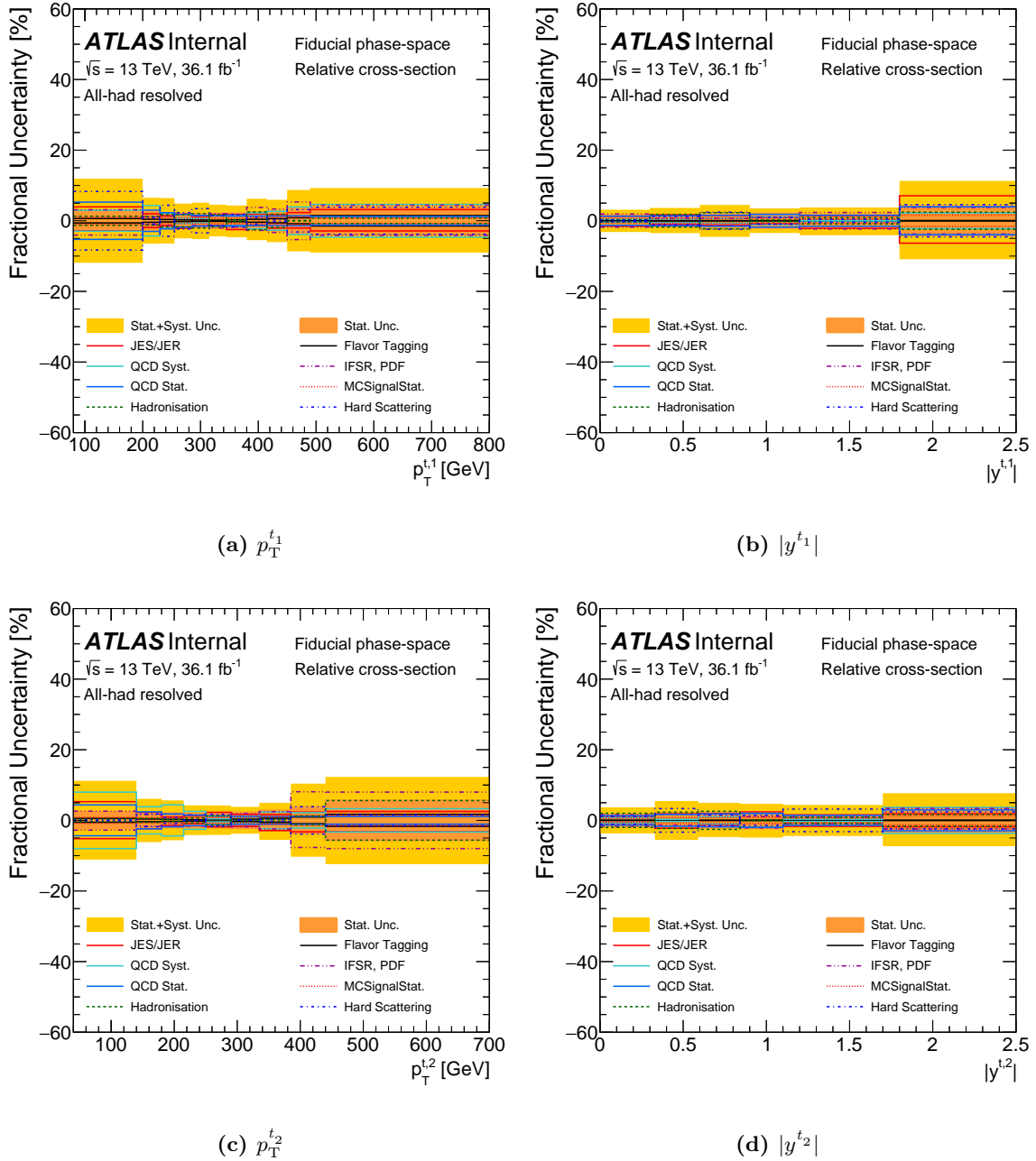
(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure 4.34 Uncertainty composition for the normalised cross section as a function of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
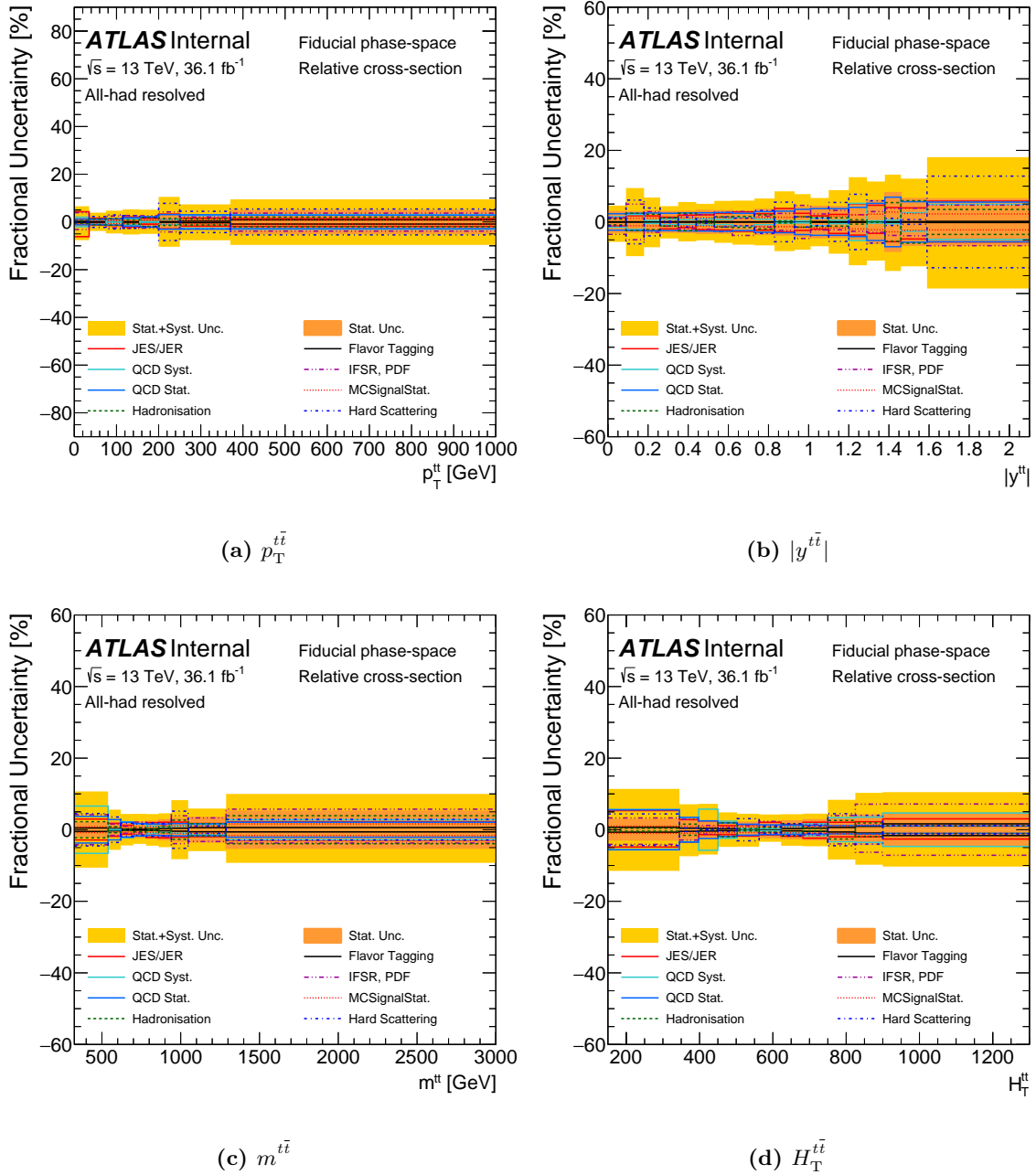
Figure 4.35 Uncertainty composition for the normalised cross section as a function of (a) $p_T^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_T^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $\chi^{t\bar{t}}$

(b) $Z^{t\bar{t}}$

(c) $\cos\theta^{\star}$

(d) $\Delta\phi$
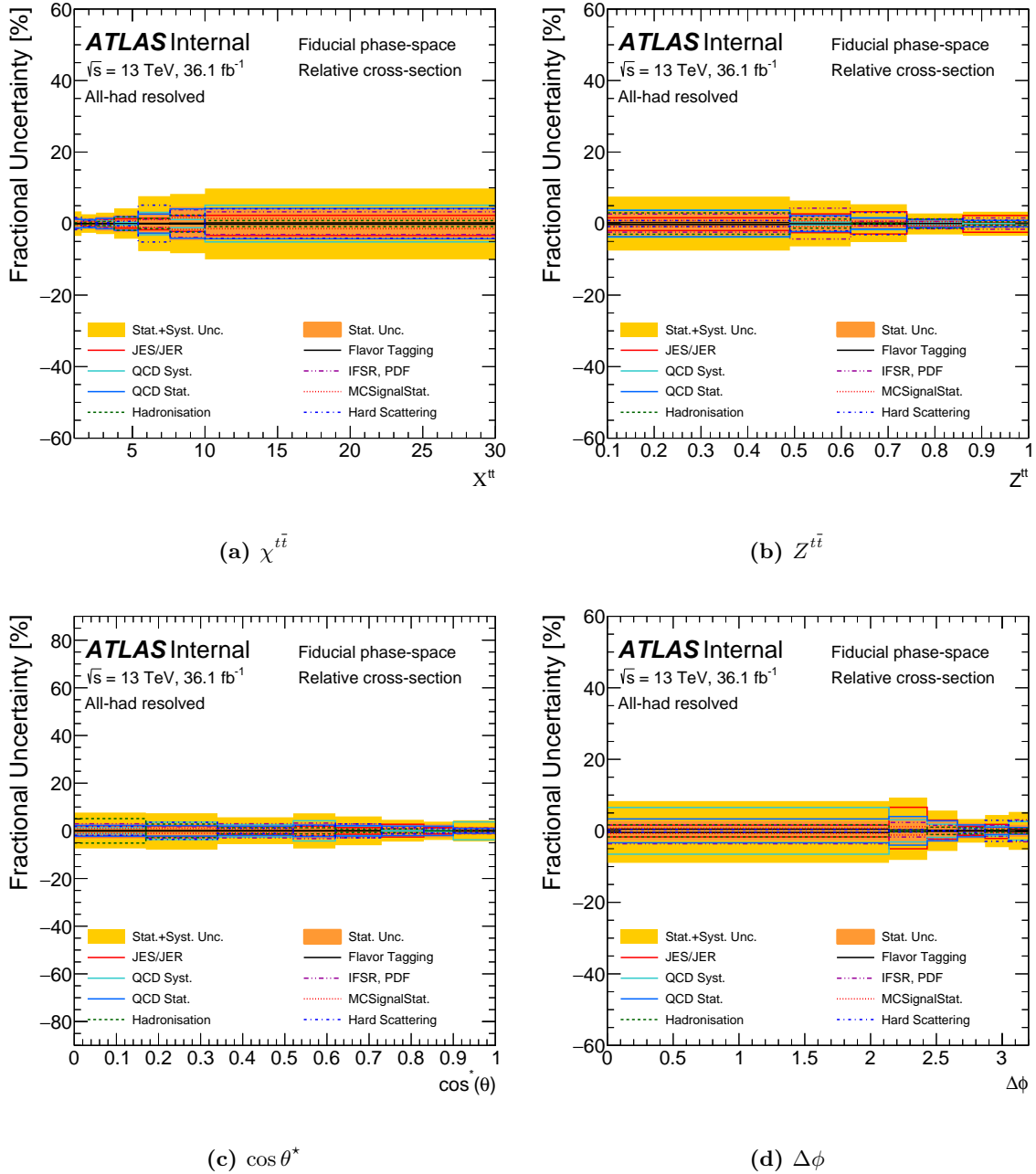
Figure 4.36 Uncertainty composition for the normalised cross section as a function of (a) $\chi^{t\bar{t}}$, (b) $Z^{t\bar{t}}$, (c) $\cos\theta^{\star}$, and (d) $\Delta\phi$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $|y_{\mathrm{boost}}|$
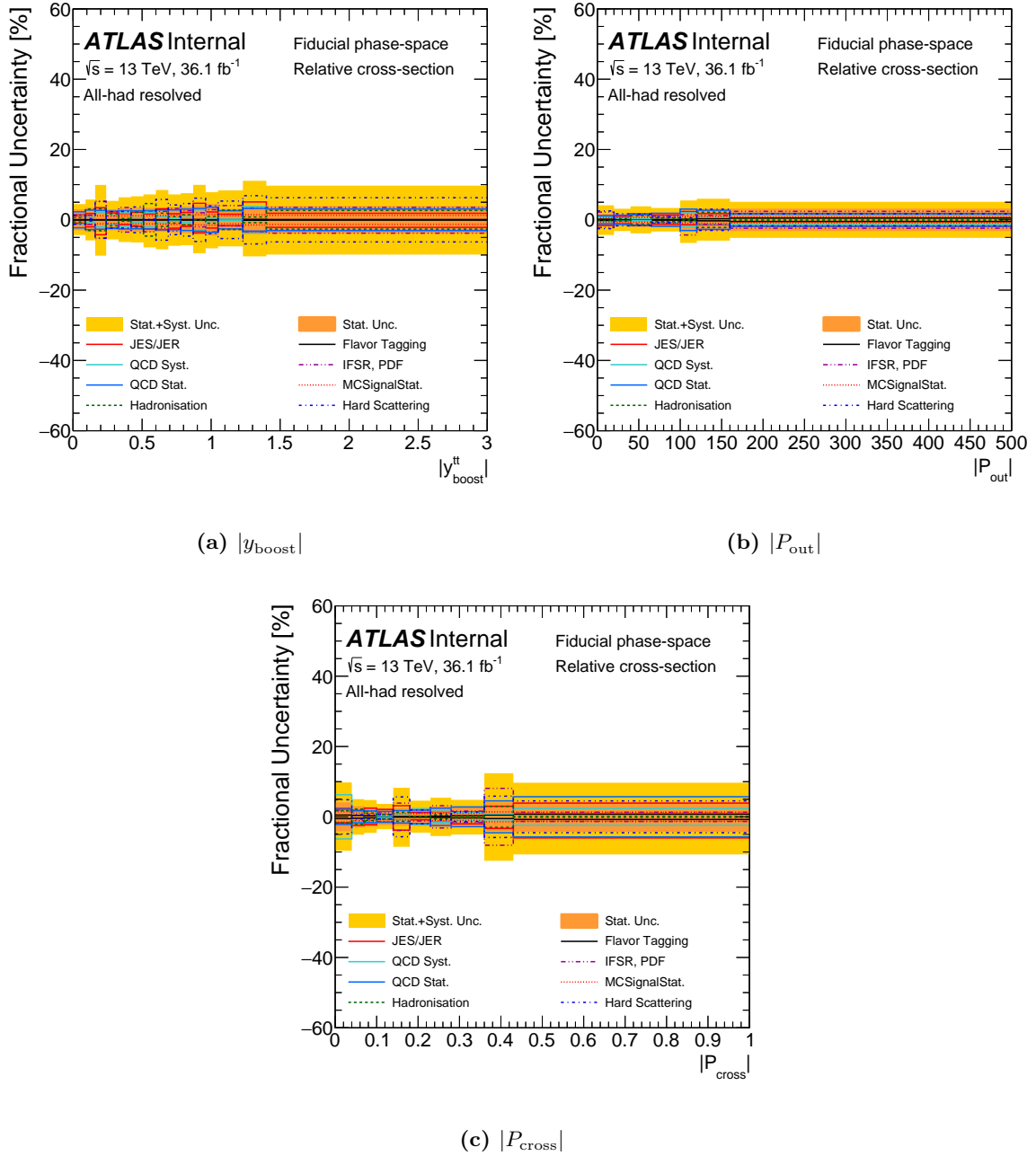
(b) $|P_{\mathrm{out}}|$

(c) $|P_{\mathrm{cross}}|$

Figure 4.37 Uncertainty composition for the normalised cross section as a function of (a) $|y_{\mathrm{boost}}|$, (b) $|P_{\mathrm{out}}|$, and (c) $|P_{\mathrm{cross}}|$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) R_{Wb}^{leading}

(b) $R_{Wb}^{\text{subleading}}$

(c) R_{Wt}^{leading}
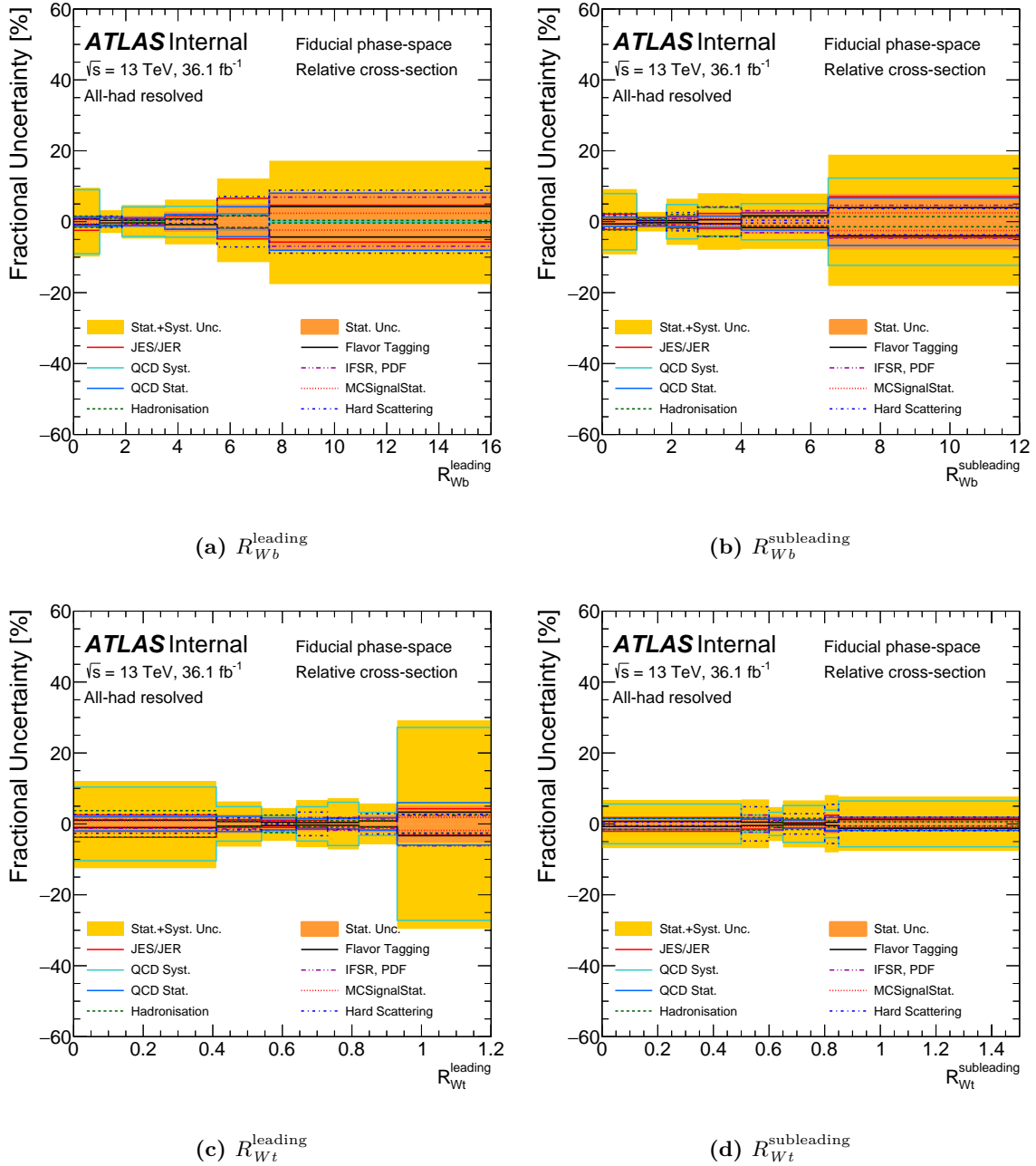
(d) $R_{Wt}^{\text{subleading}}$

Figure 4.38 Uncertainty composition for the normalised cross section as a function of (a) R_{Wb}^{leading}, (b) $R_{Wb}^{\text{subleading}}$, (c) R_{Wt}^{leading}, and (d) $R_{Wt}^{\text{subleading}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

4.8 Results

Unfolded differential $t\bar{t}$ production cross sections are shown in this section. Selected one- and two-dimensional distributions are reported, unfolded to particle and parton levels. Each figure displays the unfolded data with statistical uncertainty and combined statistical and systematic uncertainties. The data are also compared with multiple predictions from simulations, described in Section 4.4. The cross sections reported in this section are normalised, relative to the inclusive $t\bar{t}$ production cross section. The corresponding absolute differential cross sections are provided in Appendix C, displayed in the same order as this section. The results are discussed in Section 4.9.

4.8.1 Particle level 1D differential cross sections

The $t\bar{t}$ production cross section is reported as a function of the number of jets per event, N_{jets}, in Figure 4.39. Cross sections are reported as functions of kinematic variables for the reconstructed t quarks in Figure 4.40, and for the $t\bar{t}$ system in Figure 4.41. Distributions for the remaining observables are given in Figures 4.42 to 4.44.
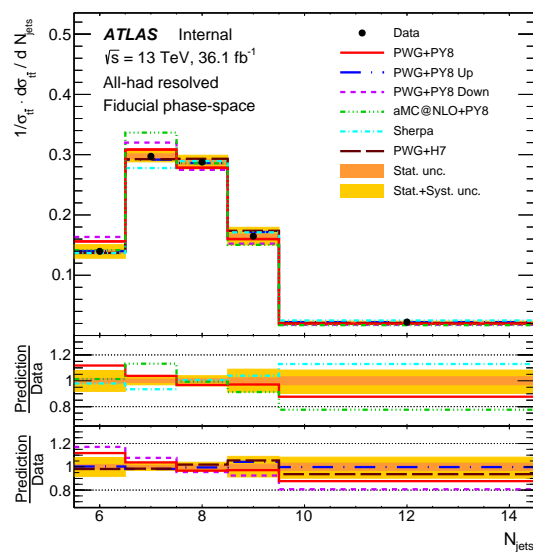
Figure 4.39 Normalised differential cross section as a function of N_{jets}, unfolded to particle level. The bottom panels show the ratios of predictions from simulations to the data.

(a) $p_{\mathrm{T}}^{t_1}$
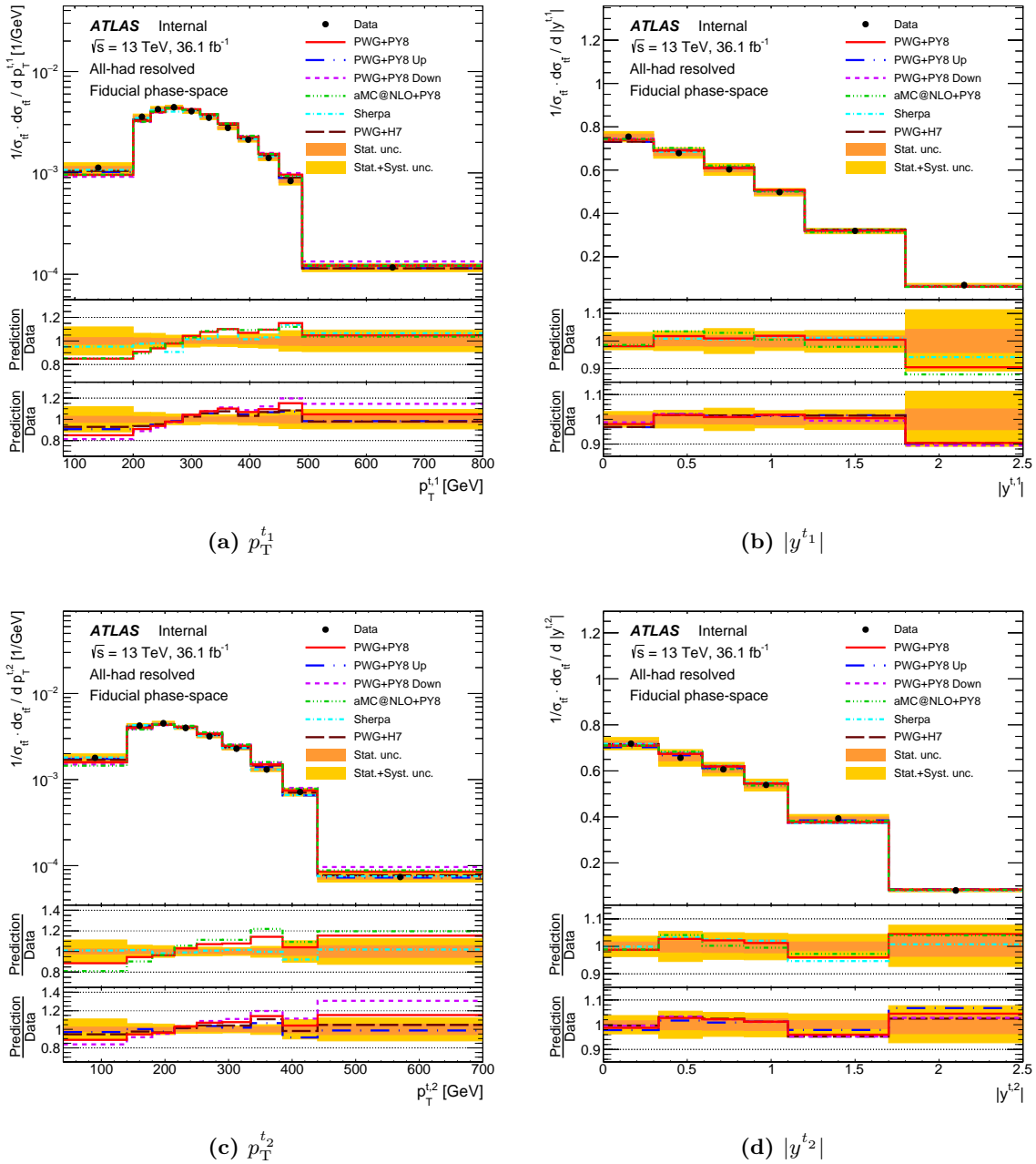
(b) $|y^{t_1}|$

(c) $p_{\mathrm{T}}^{t_2}$

(d) $|y^{t_2}|$

Figure 4.40 Normalised differential cross sections as functions of (a) $p_{\mathrm{T}}^{t_1}$, (b) $|y^{t_1}|$, (c) $p_{\mathrm{T}}^{t_2}$, and (d) $|y^{t_2}|$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
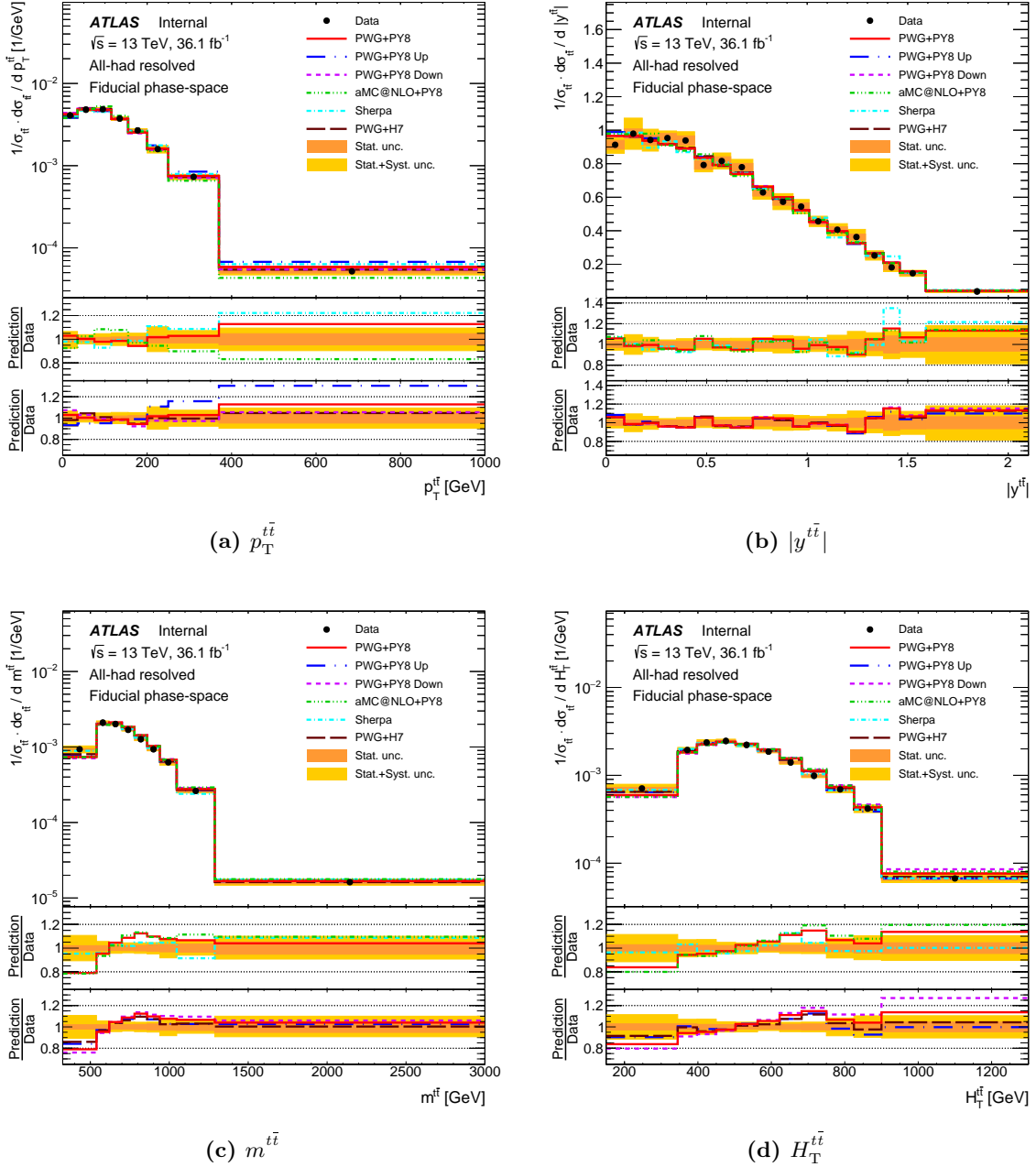
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure 4.41 Normalised differential cross sections as functions of (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $\chi^{t\bar{t}}$

(b) $Z^{t\bar{t}}$

(c) $\cos\theta^{\star}$

(d) $\Delta\phi$
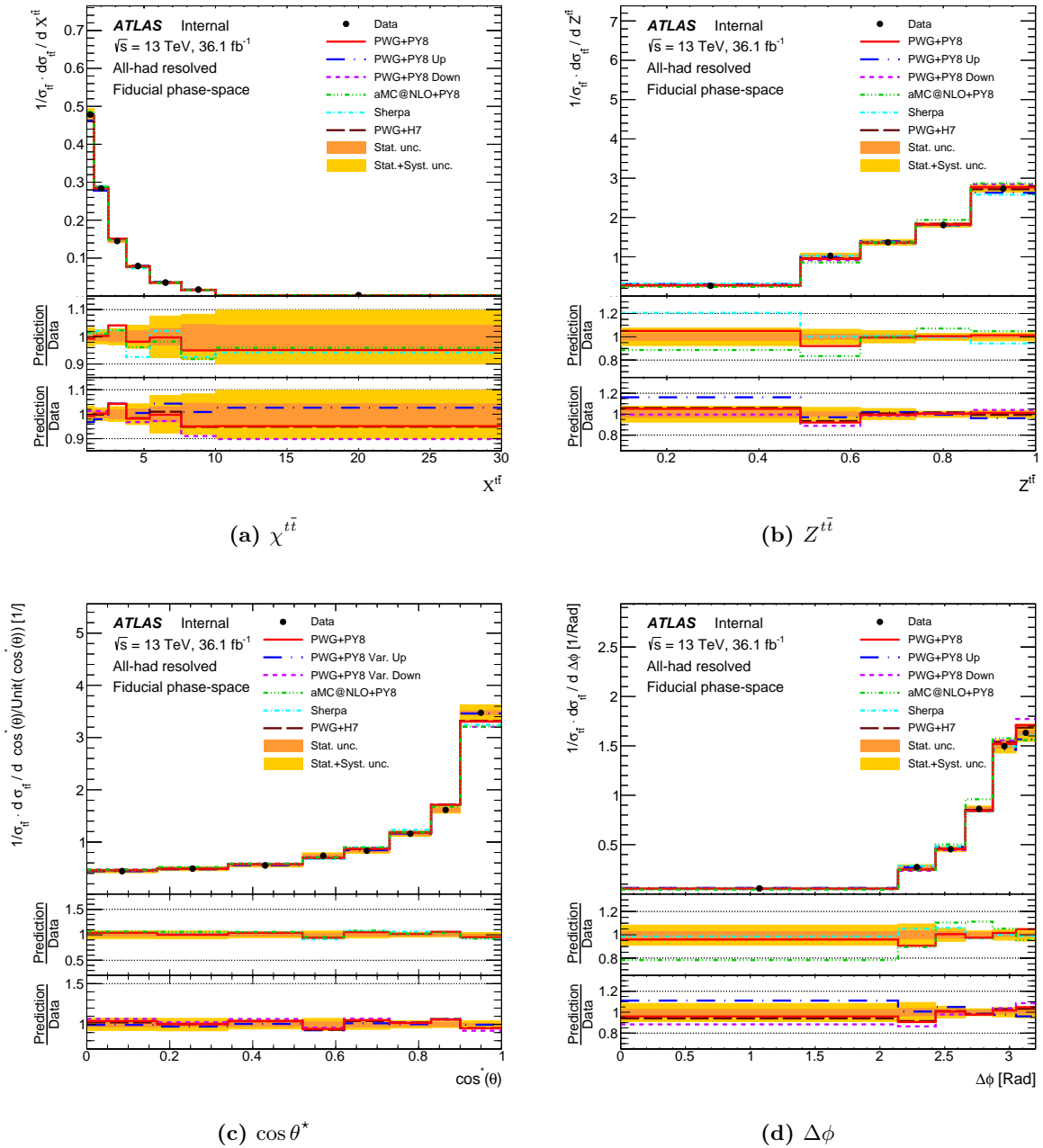
Figure 4.42 Normalised differential cross sections as functions of (a) $\chi^{t\bar{t}}$, (b) $Z^{t\bar{t}}$, (c) $\cos\theta^{\star}$, and (d) $\Delta\phi$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $|y_{\text{boost}}|$
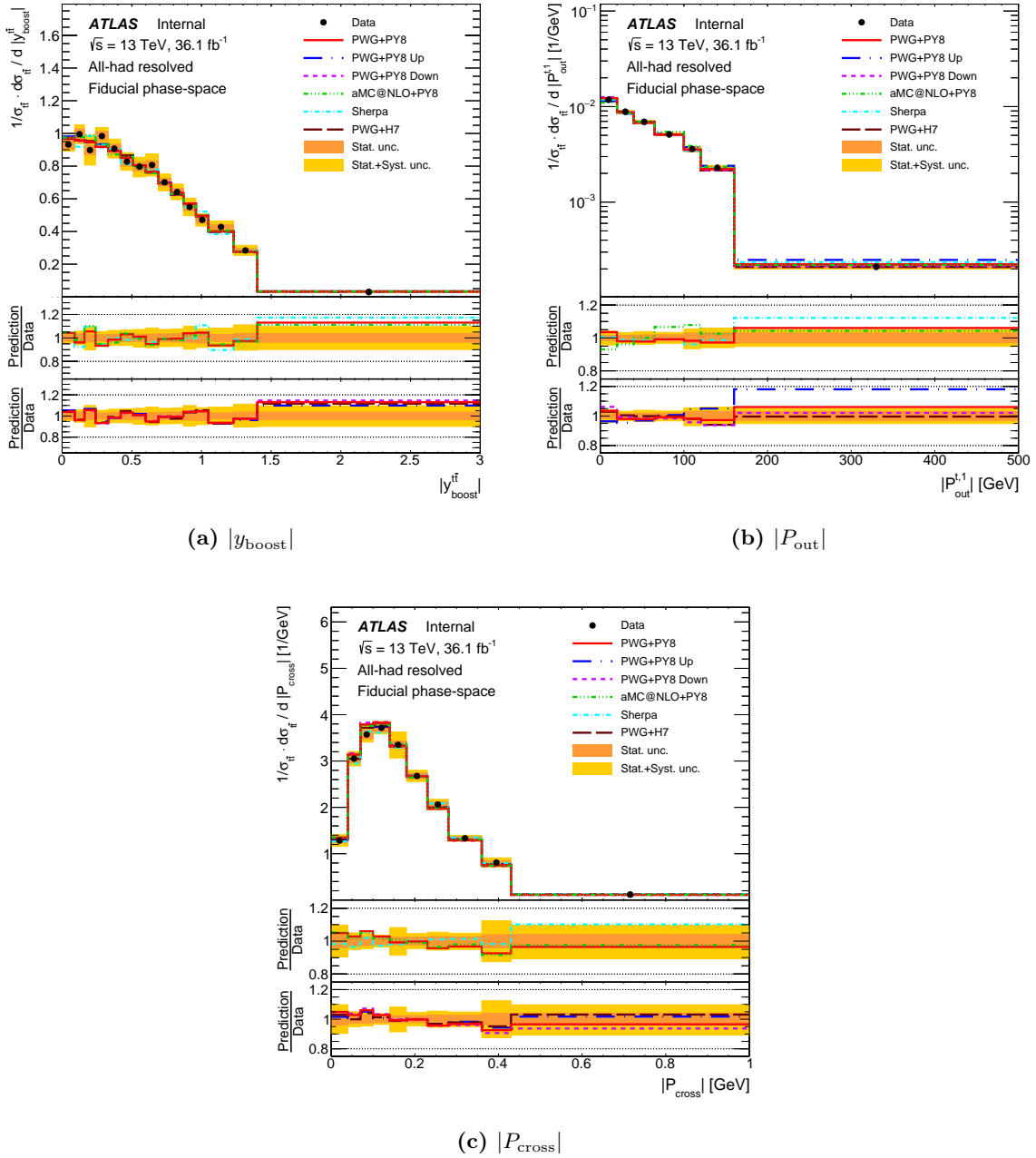
(b) $|P_{\text{out}}|$

(c) $|P_{\text{cross}}|$

Figure 4.43 Normalised differential cross sections as functions of (a) $|y_{\text{boost}}|$, (b) $|P_{\text{out}}|$, and (c) $|P_{\text{cross}}|$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulations to the data.

179

(a) R_{Wb}^{leading}

(b) $R_{Wb}^{\text{subleading}}$

(c) R_{Wt}^{leading}
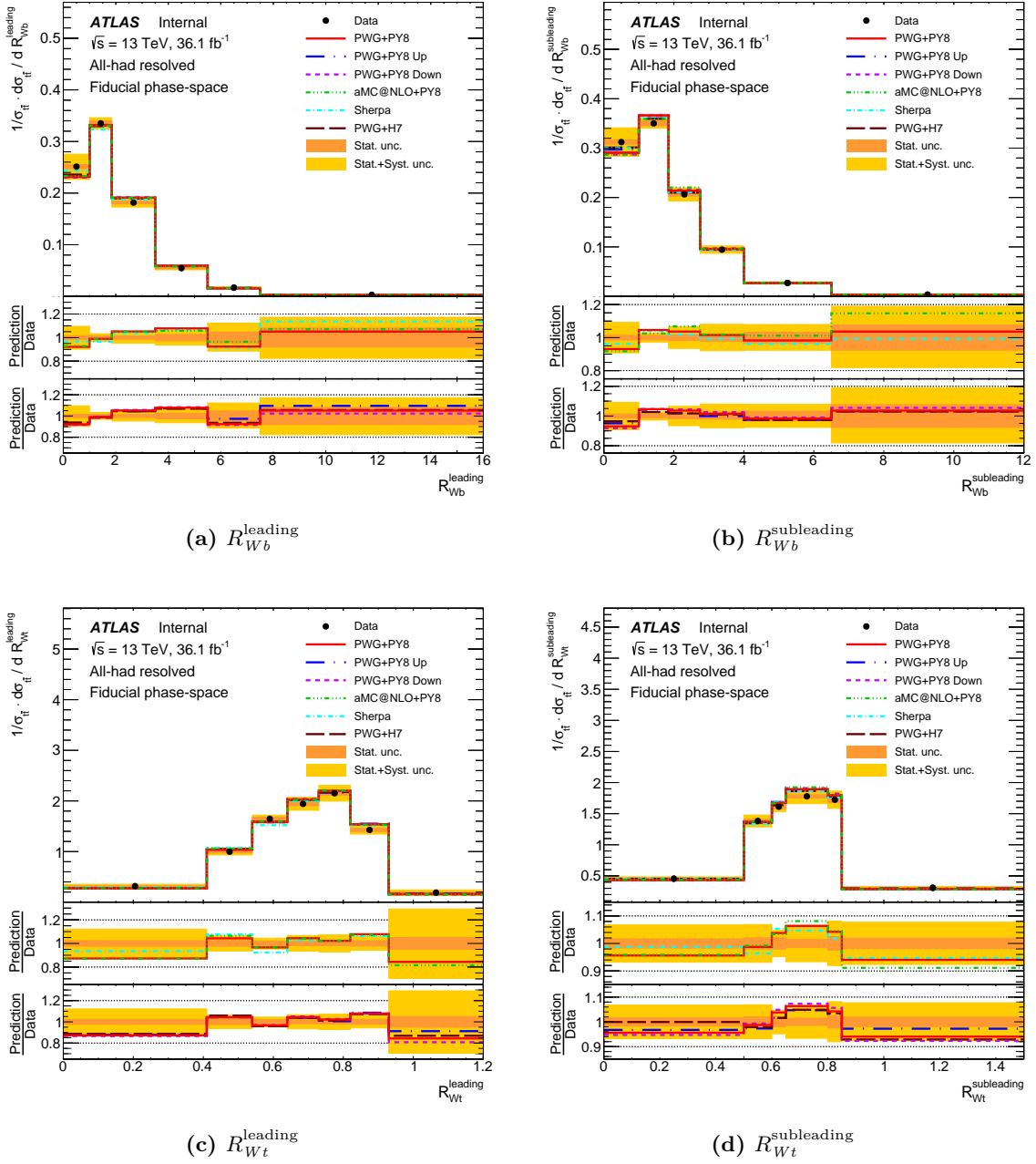
(d) $R_{Wt}^{\text{subleading}}$

Figure 4.44 Normalised differential cross sections as functions of (a) R_{Wb}^{leading}, (b) $R_{Wb}^{\text{subleading}}$, (c) R_{Wt}^{leading}, and (d) $R_{Wt}^{\text{subleading}}$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

4.8.2 Particle level 2D differential cross sections

Normalised differential $t\bar{t}$ production cross sections are reported in two-dimensional distributions with external variables $p_{\mathrm{T}}^{t_2}$ and $m^{t\bar{t}}$ in Figures 4.45 to 4.49. They are given in distributions in bins of N_{jets} in Figures 4.50 to 4.55. In each figure, panel (a) shows the value of the normalised cross section as a function of the internal variable, in bins of the external variable. Panel (b) compares different predictions by taking the ratio to the unfolded data. This panel also shows the statistical and combined statistical and systematic uncertainties on the data.
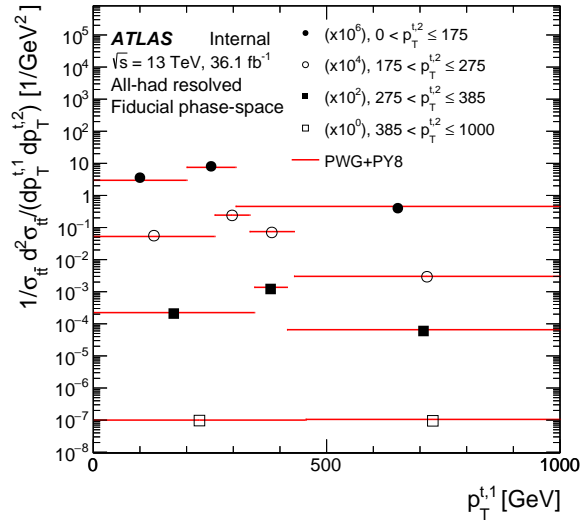
(a)
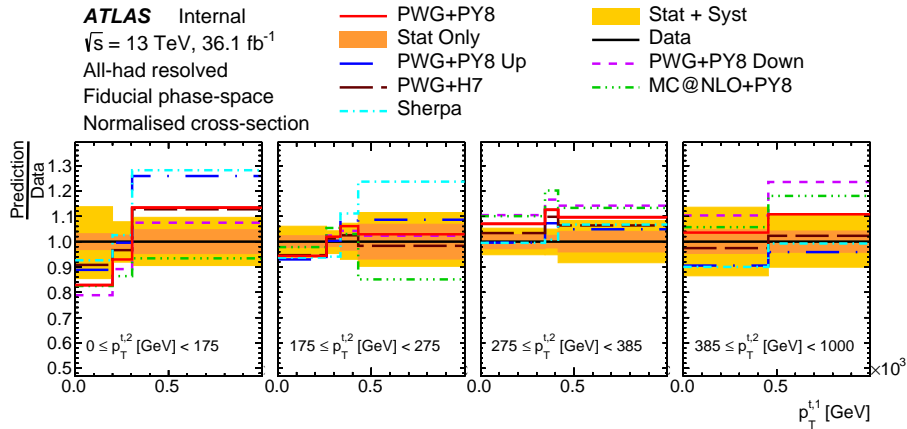
(b)

Figure 4.45 (a) Normalised differential cross section as a function of $p_T^{t_1}$ in bins of $p_T^{t_2}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.
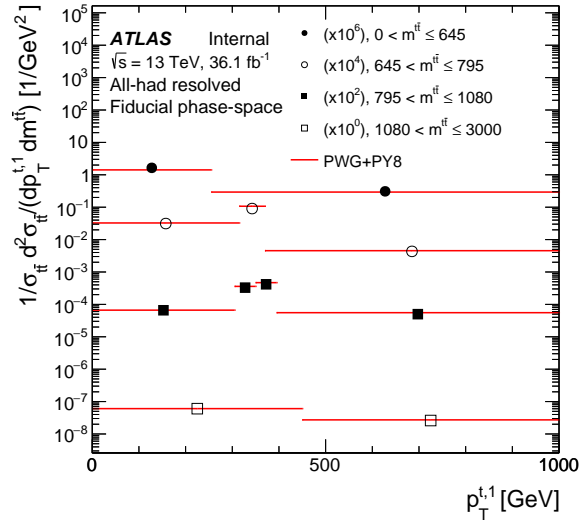
(a)
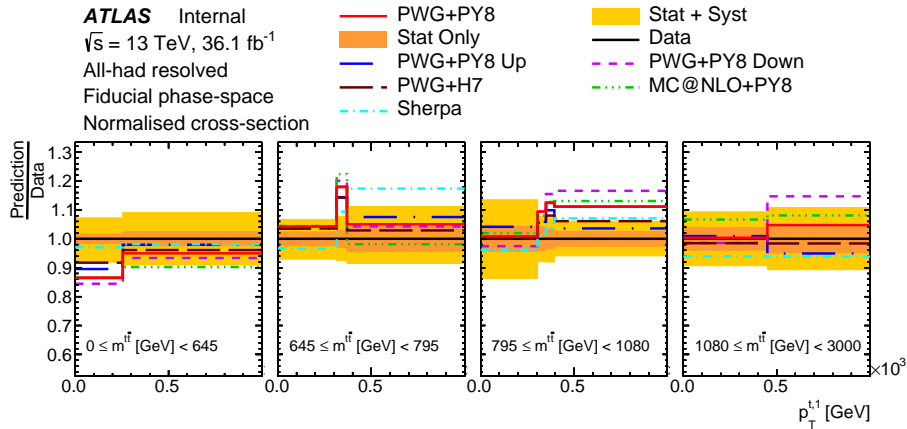
(b)

Figure 4.46 (a) Normalised differential cross section as a function of $p_{\mathrm{T}}^{t,1}$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.
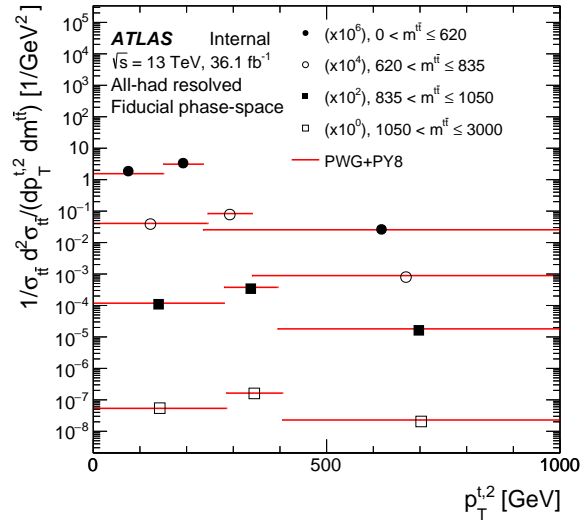
(a)
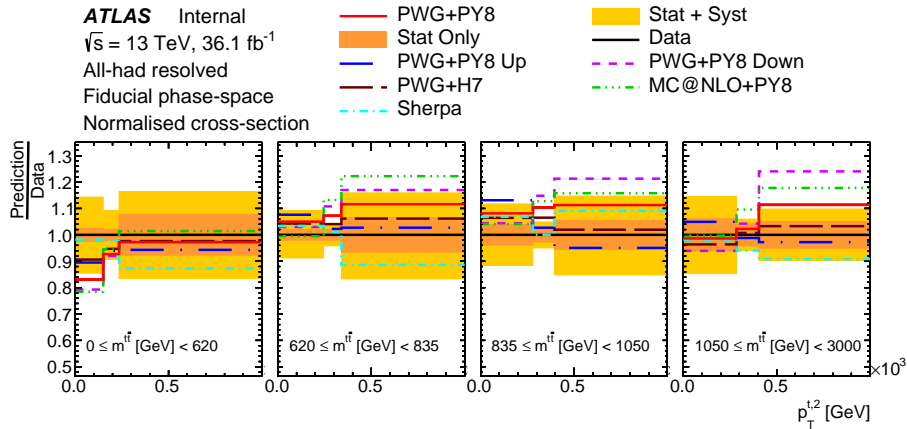
(b)

Figure 4.47 (a) Normalised differential cross section as a function of $p_{\mathrm{T}}^{t_2}$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.
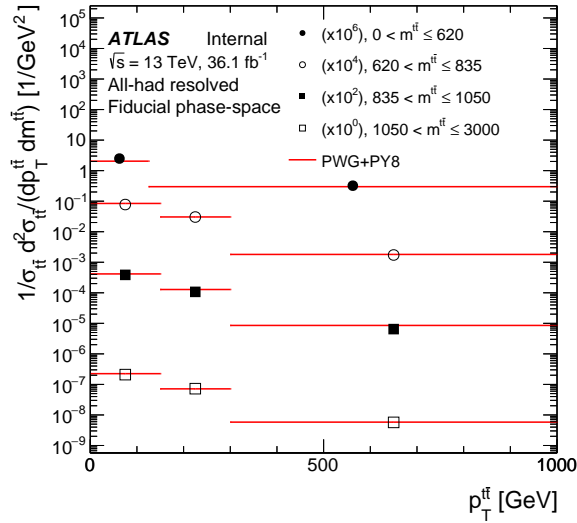
(a)

(b)

Figure 4.48 (a) Normalised differential cross section as a function of $p_{\mathrm{T}}^{t\bar{t}}$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)

Figure 4.49 (a) Normalised differential cross section as a function of $|y^{t\bar{t}}|$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)

Figure 4.50 (a) Normalised differential cross section as a function of $p_T^{t_1}$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)

Figure 4.51 (a) Normalised differential cross section as a function of $p_{\mathrm{T}}^{t_2}$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)

Figure 4.52 (a) Normalised differential cross section as a function of $p_T^{t\bar{t}}$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)

Figure 4.53 (a) Normalised differential cross section as a function of $\Delta\phi$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.
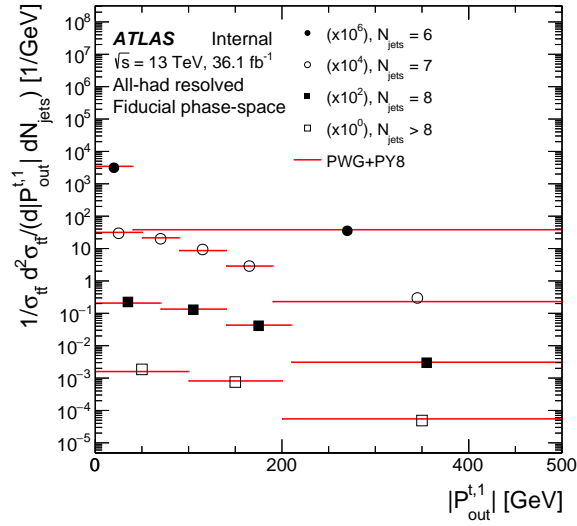
(a)
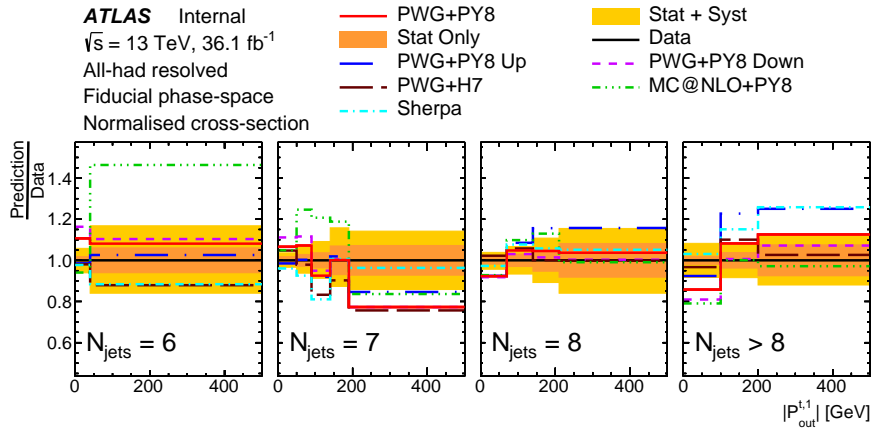
(b)

Figure 4.54 (a) Normalised differential cross section as a function of $|P_{\text{out}}|$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.
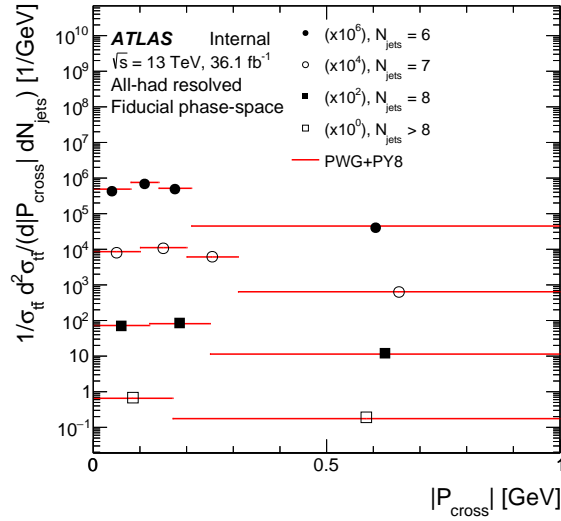
(a)
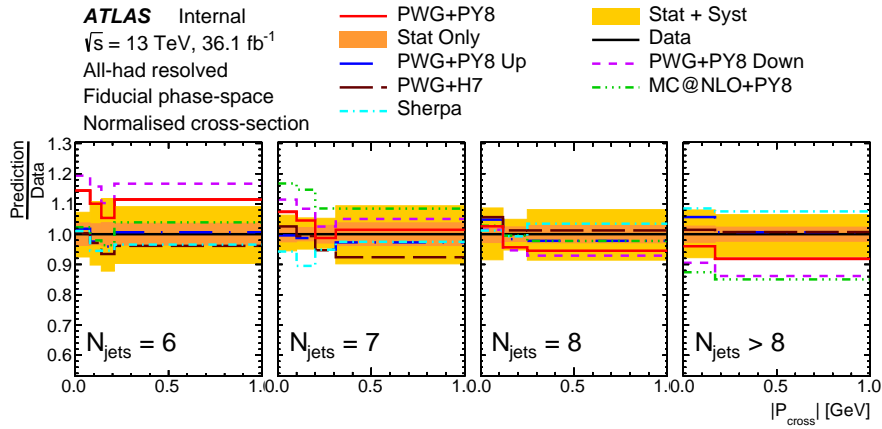
(b)

Figure 4.55 (a) Normalised differential cross section as a function of $|P_{\text{cross}}|$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

4.8.3 Parton level 1D differential cross sections

Cross sections are reported as functions of kinematic variables for the reconstructed t quarks in Figure 4.56, and for the $t\bar{t}$ system in Figure 4.57. Distributions for the remaining observables are given in Figure 4.58.

(a) $p_T^{t_1}$
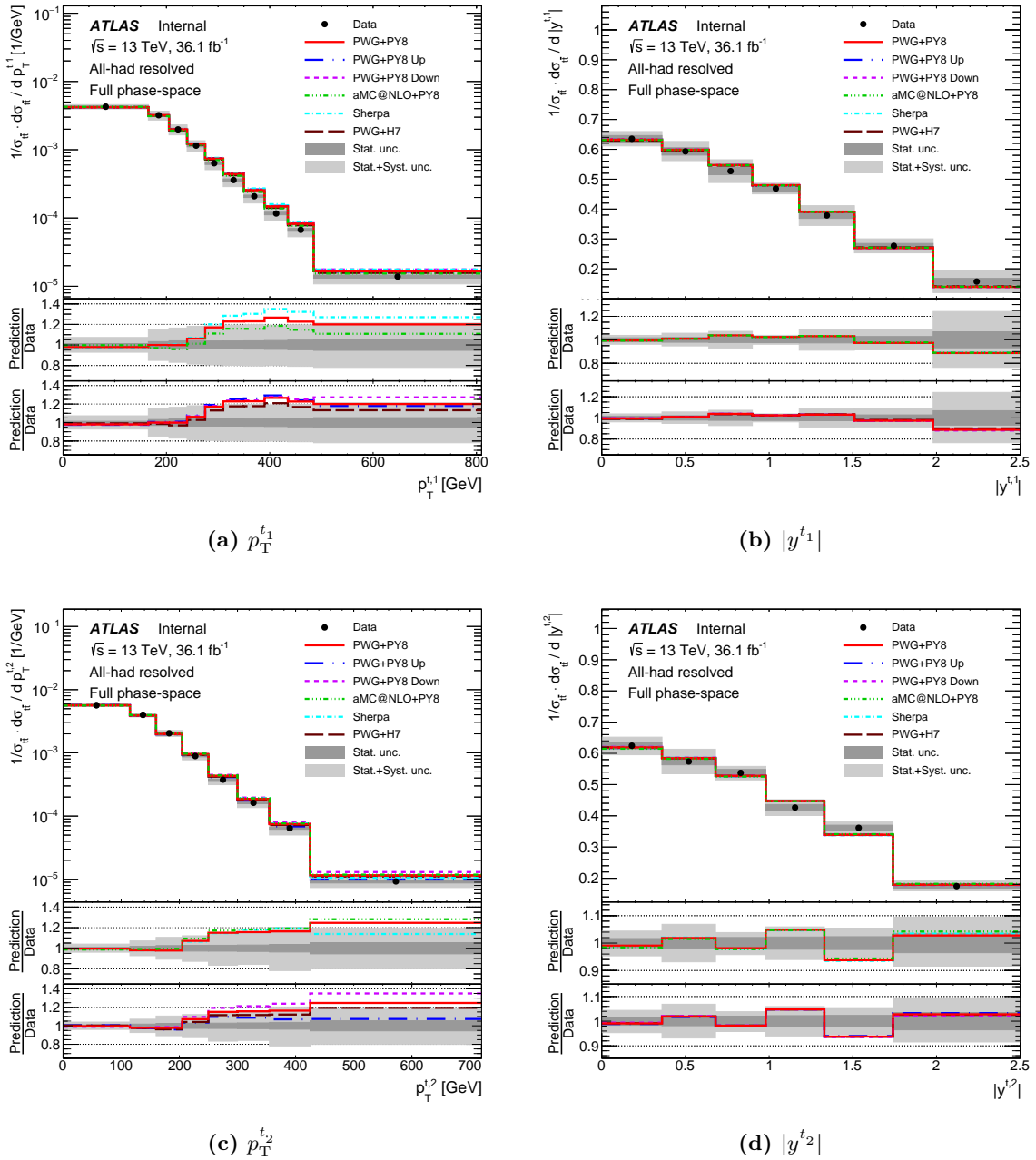
(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure 4.56 Normalised differential cross sections as functions of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, all unfolded to parton level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
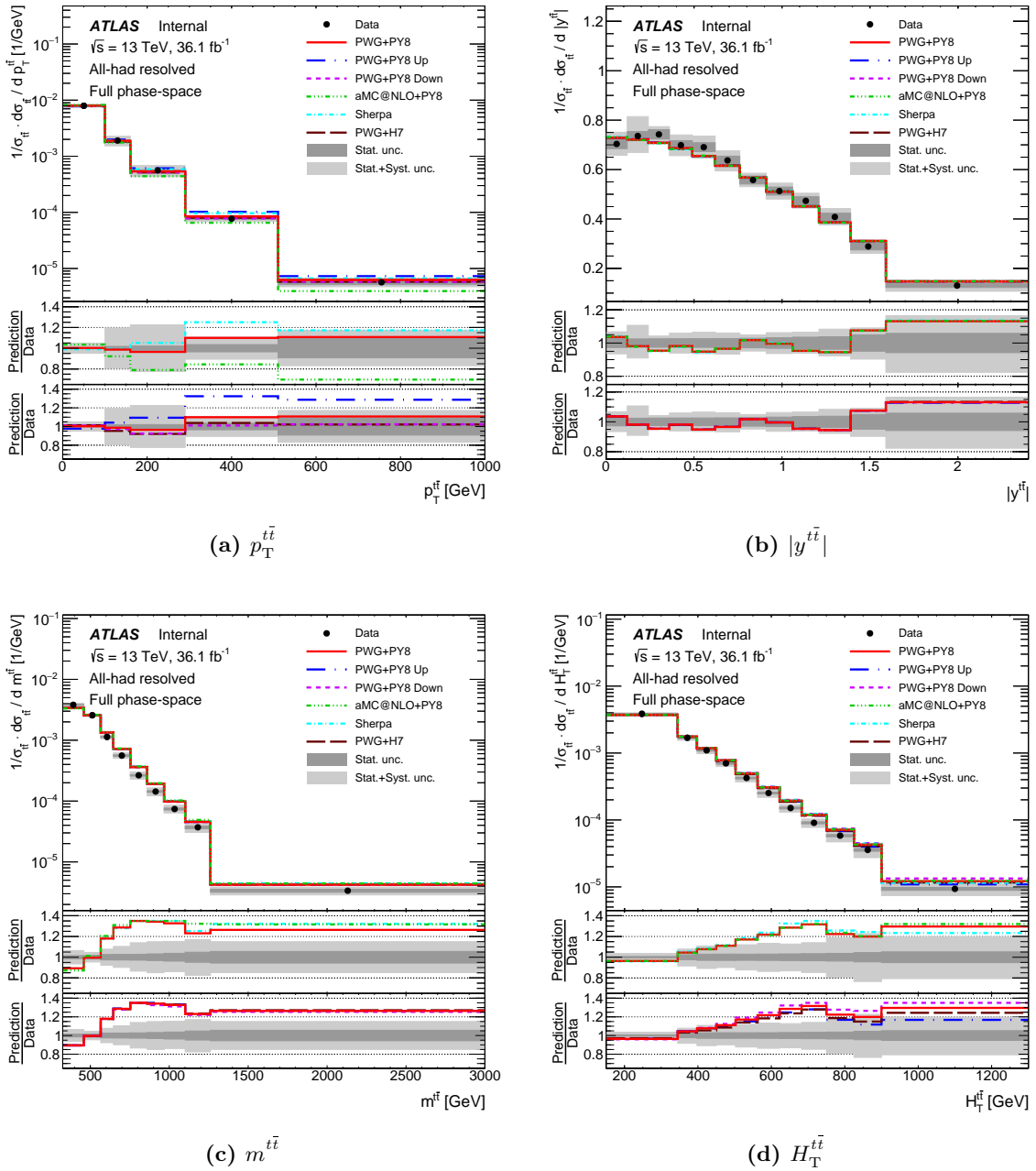
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure 4.57 Normalised differential cross sections as functions of (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$, all unfolded to parton level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $\chi^{t\bar{t}}$

(b) $\Delta\phi$
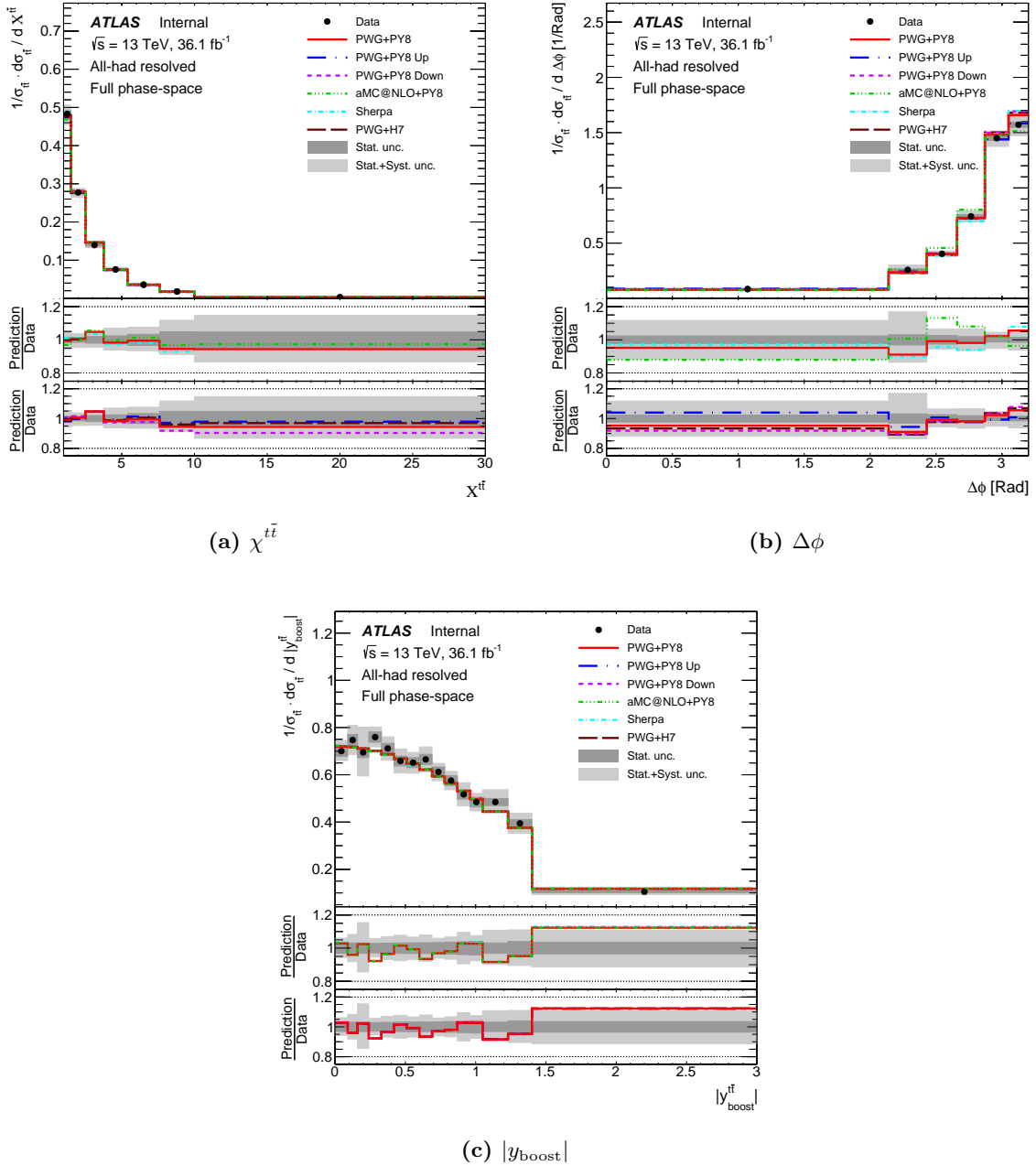
(c) $|y_{\text{boost}}|$

Figure 4.58 Normalised differential cross sections as functions of (a) $\chi^{t\bar{t}}$, (b) $\Delta\phi$, and (c) $|y_{\text{boost}}|$, all unfolded to parton level. The bottom panels show the ratios of predictions from simulation to the data.

4.8.4 Parton level 2D differential cross sections

Normalised differential $t\bar{t}$ production cross sections are reported in two-dimensional distributions with external variables $p_{\mathrm{T}}^{t_2}$ and $|y^{t_1}|$ in Figures 4.59 and 4.60. They are reported in bins of $m^{t\bar{t}}$ in Figures 4.61 to 4.66. In each figure, panel (a) shows the value of the normalised cross section as a function of the internal variable, in bins of the external variable. Panel (b) compares different predictions by taking the ratio to the unfolded data. This panel also shows the statistical and combined statistical and systematic uncertainties on the data.
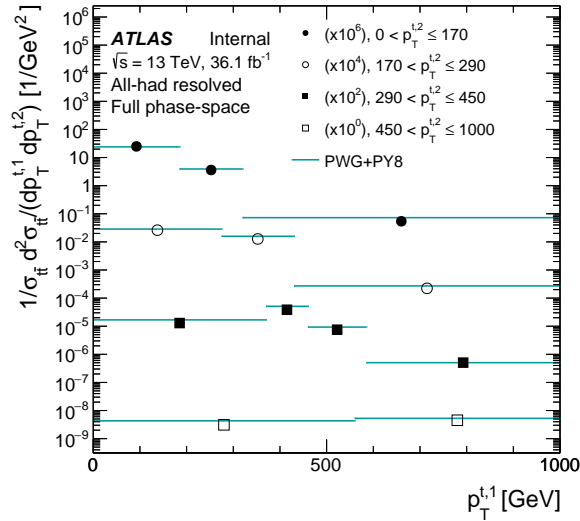
(a)
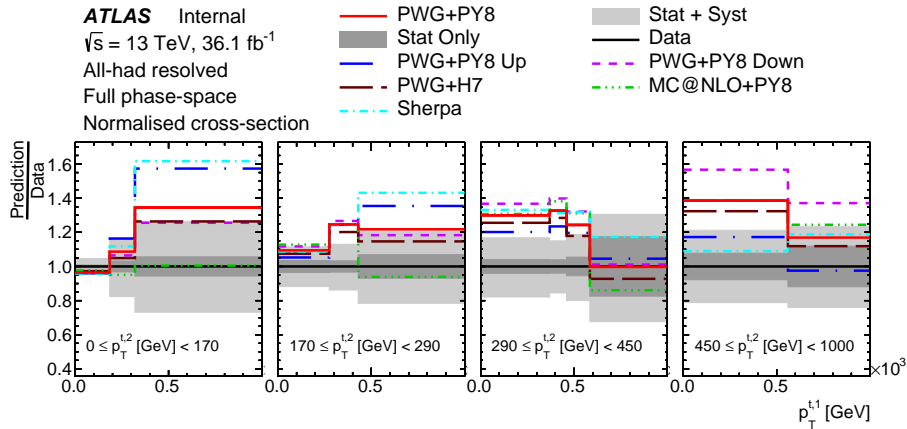
(b)

Figure 4.59 (a) Normalised differential cross section as a function of $p_T^{t_1}$ in bins of $p_T^{t_2}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
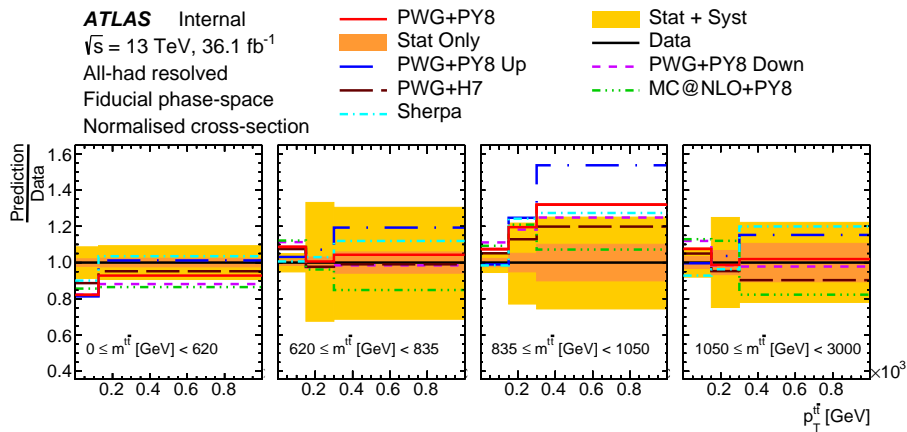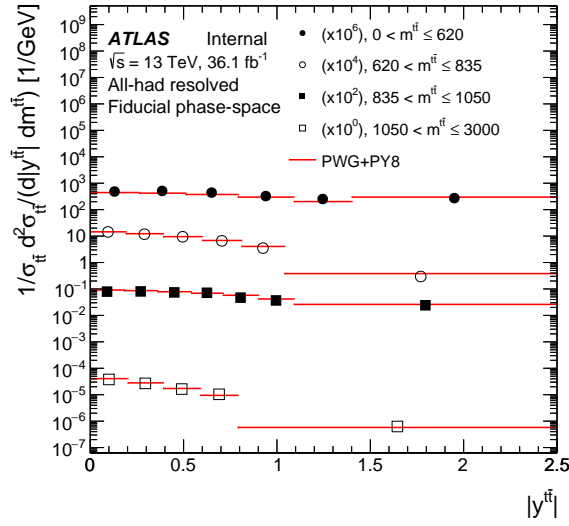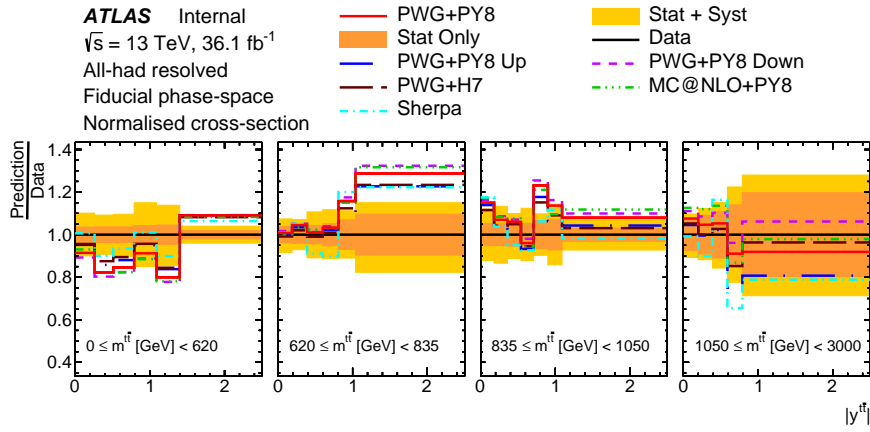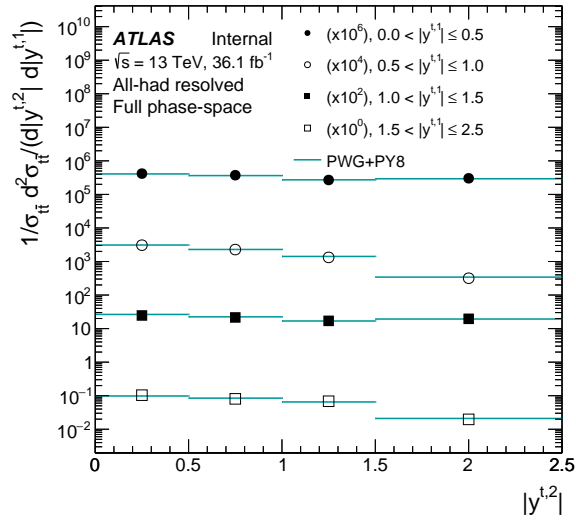
(a)
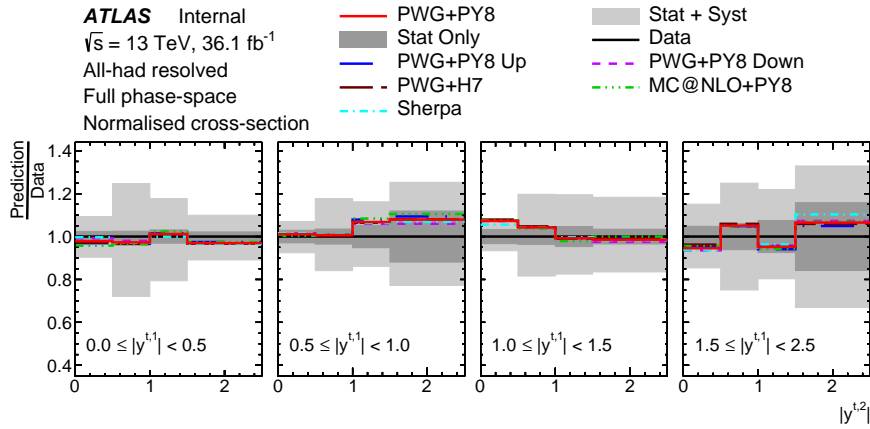
(b)

Figure 4.60 (a) Normalised differential cross section as a function of $|y^{t_2}|$ in bins of $|y^{t_1}|$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
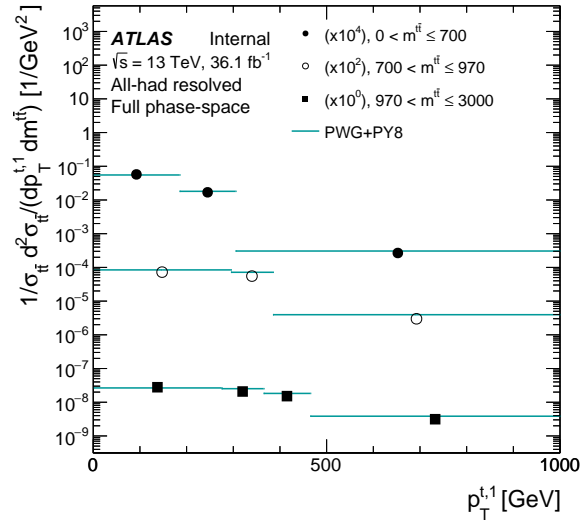
(a)
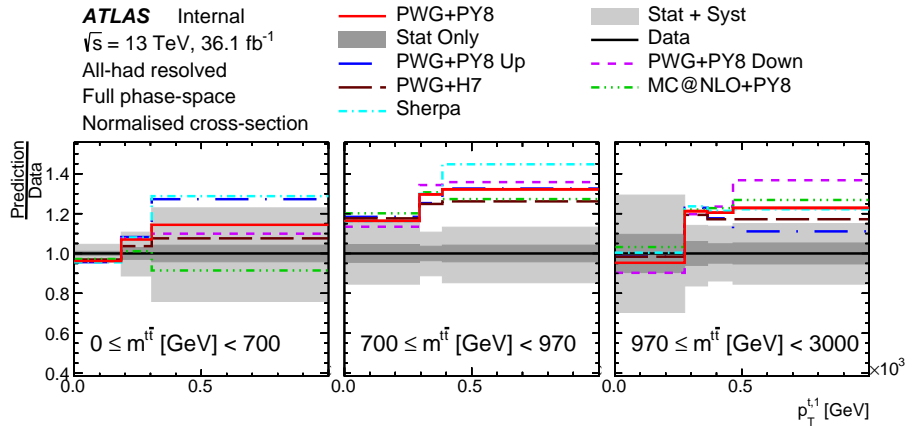
(b)

Figure 4.61 (a) Normalised differential cross section as a function of $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
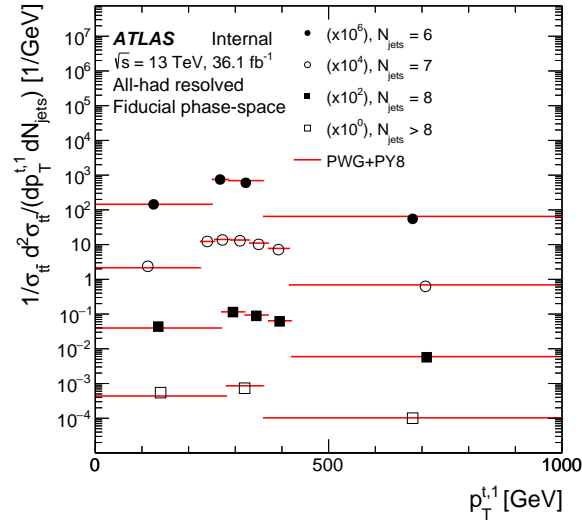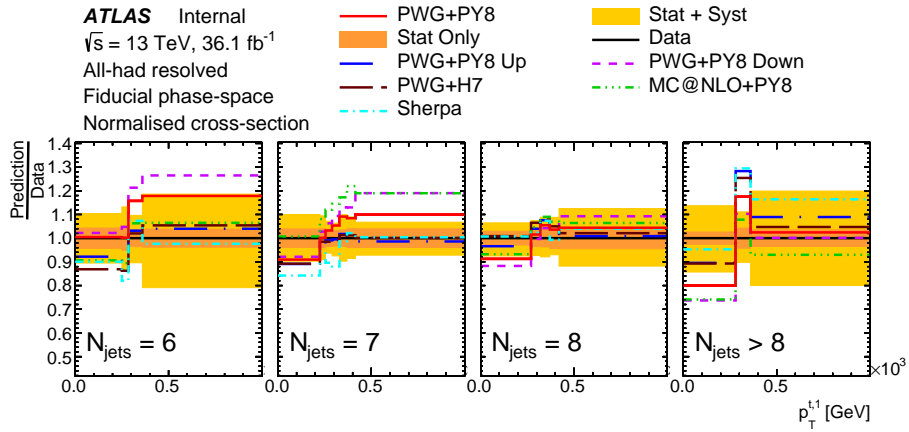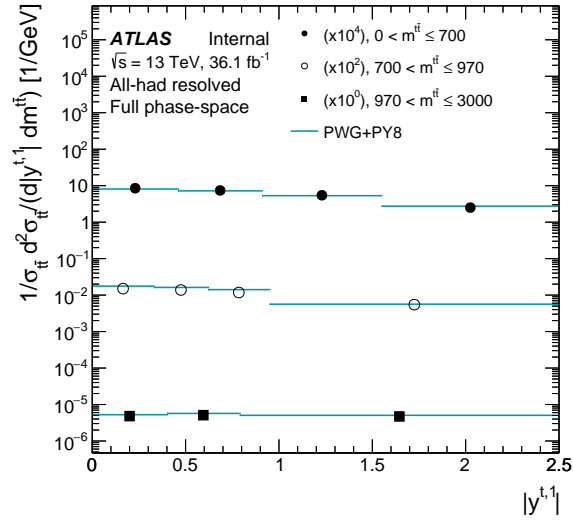
(a)
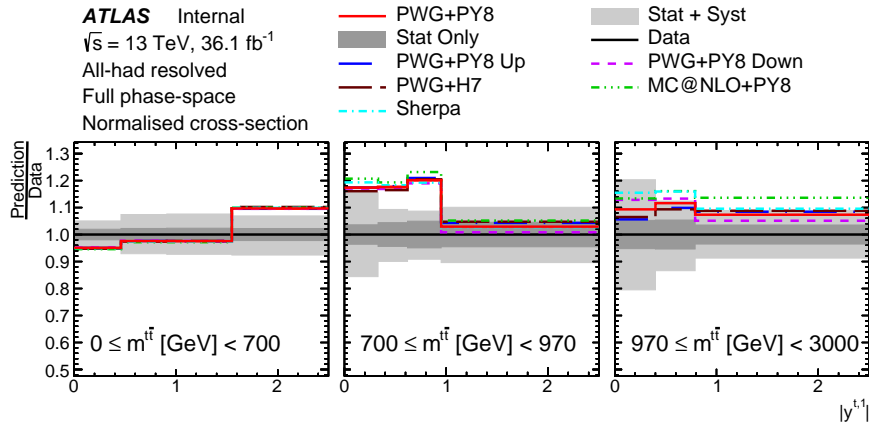
(b)

Figure 4.62 (a) Normalised differential cross section as a function of $|y^{t_1}|$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
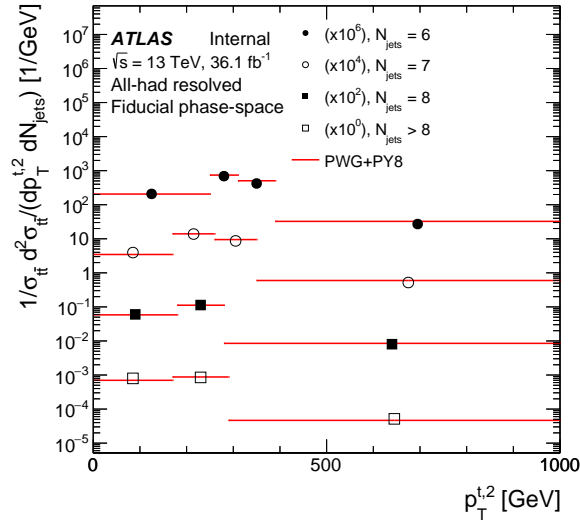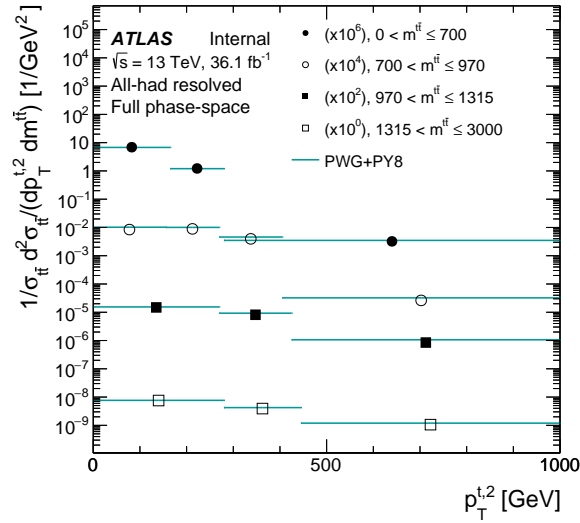
(a)
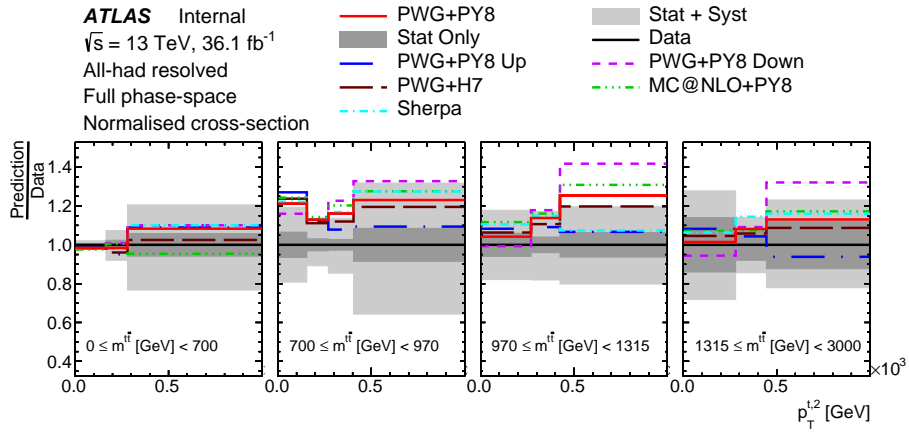
(b)

Figure 4.63 (a) Normalised differential cross section as a function of $p_T^{t_2}$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
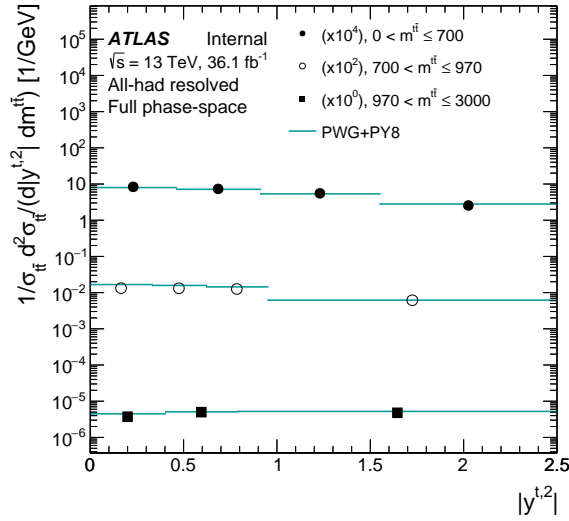
(a)
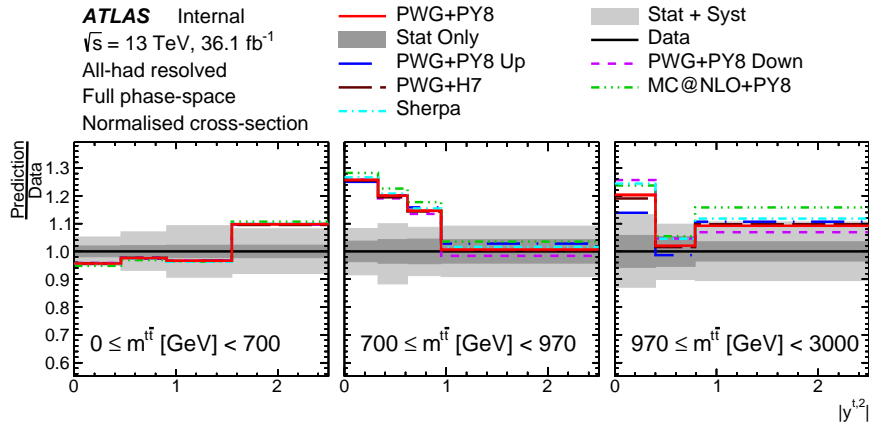
(b)

Figure 4.64 (a) Normalised differential cross section as a function of $|y^{t_2}|$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
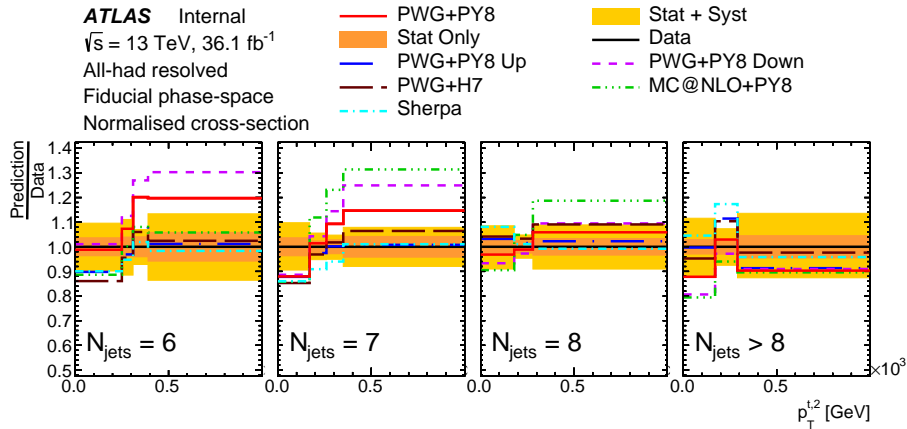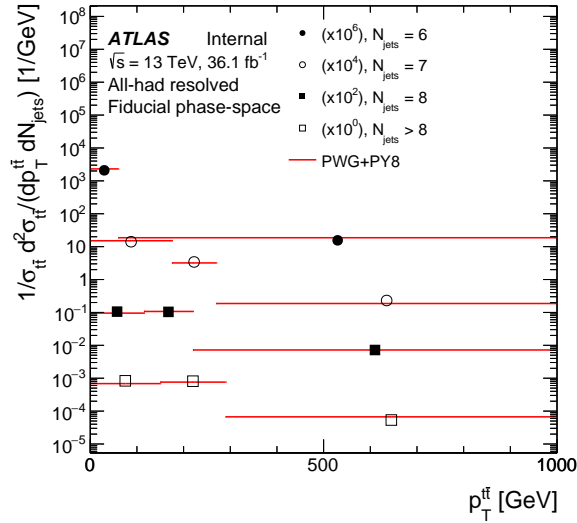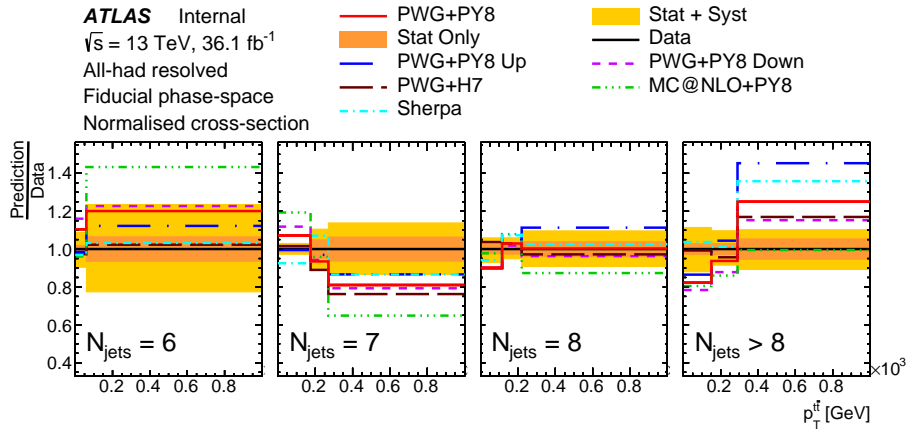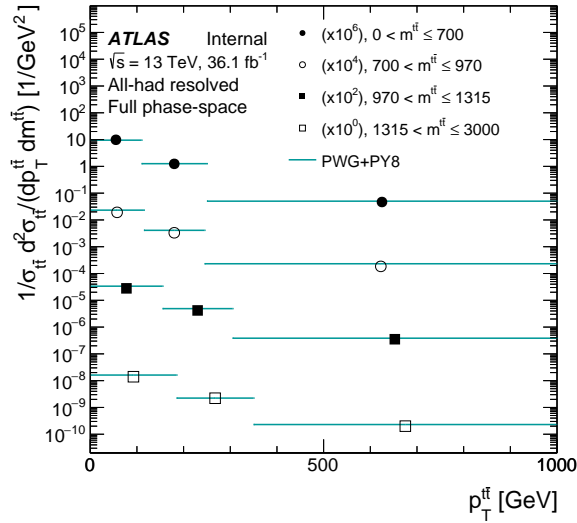
(a)
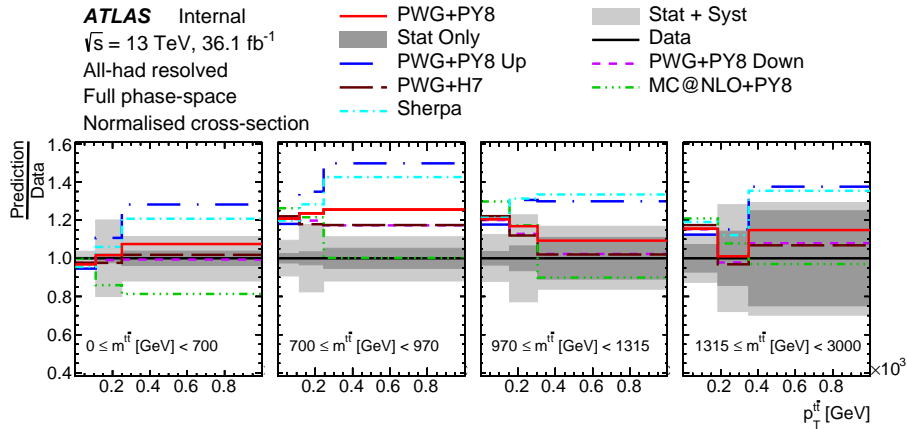
(b)

Figure 4.65 (a) Normalised differential cross section as a function of $p_{\mathrm{T}}^{t\bar{t}}$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
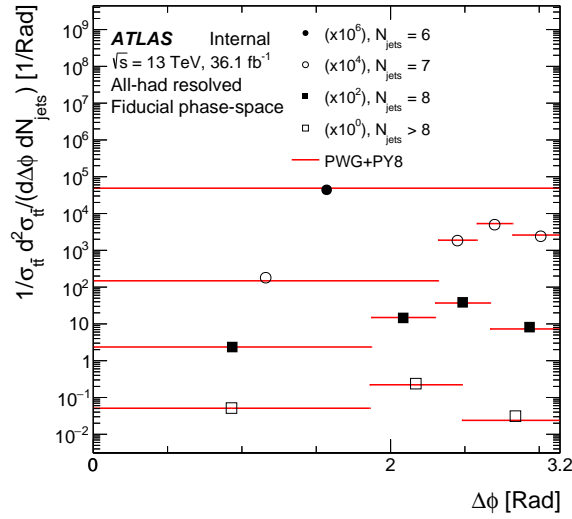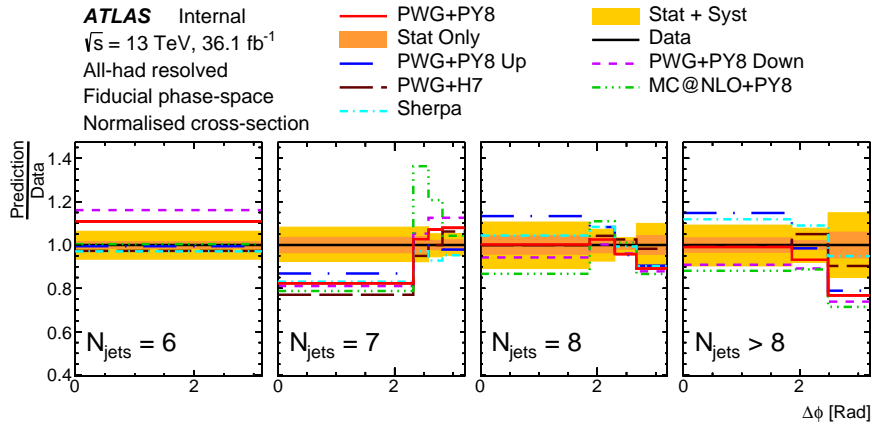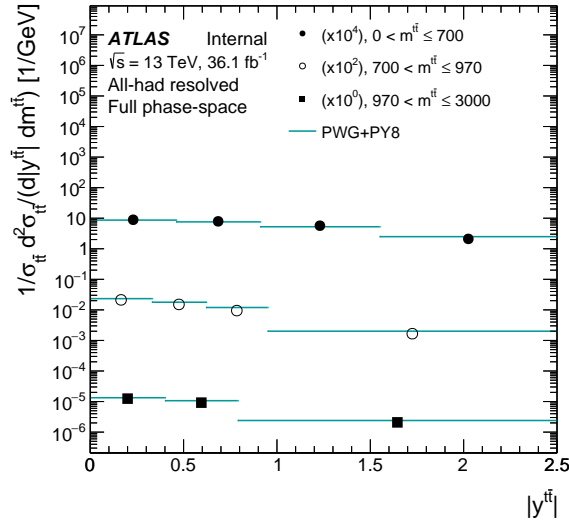
(a)
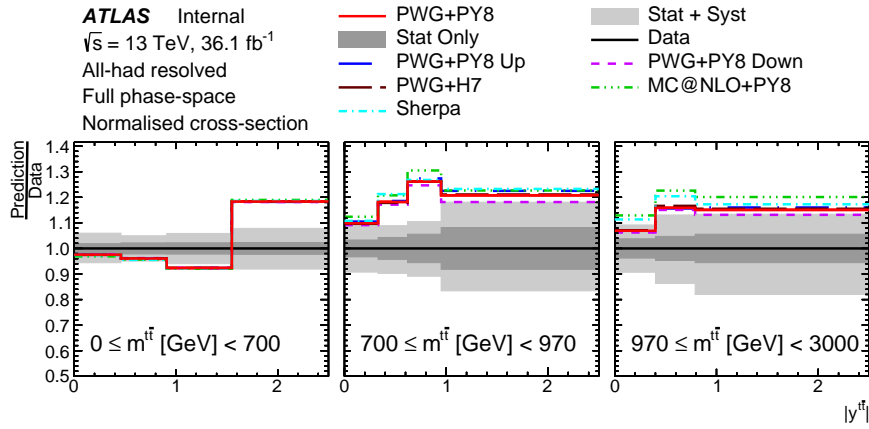
(b)

Figure 4.66 (a) Normalised differential cross section as a function of $|y^{t\bar{t}}|$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.

4.8.5 Comparison to simulations

The measured differential cross sections are compared to theoretical predictions by means of a χ^2 test. Values are calculated using the total experimental covariance matrix for each cross section. The χ^2 test statistic is given by the inner product

$$\chi^2 = \mathbf{v}^\mathsf{T}\Sigma^{-1}\mathbf{v}, \tag{4.15}$$

where $\mathbf{v}$ is the column vector of differences between the measured and predicted cross section values. The experimental covariance matrix Σ is calculated from the sum of contributions,

$$\Sigma = \Sigma_1 + \Sigma_{\mathrm{ME}} + \Sigma_{\mathrm{PS}} + \Sigma_{\mathrm{IFSR}} + \Sigma_{\mathrm{PDF}}, \tag{4.16}$$

and its rank gives the number of degrees of freedom (NDF) in the χ^2 calculation. The covariance matrix Σ_1 includes covariances from statistical effects, as well as uncertainties in modelling the detector and background contribution. The statistical uncertainty is calculated by performing pseudo-experiments, where in each the distribution of data is given by sampling from a Poisson distribution with expectation values given by the nominal prediction. The bin counts are modified by adding a contribution from systematic uncertainties discussed in Section 4.7, sampled from Gaussian distributions. The varied bin counts are limited to be positive and passed through the unfolding procedure, and the resulting distribution of cross sections is used to determine Σ_1.

The other contributions from signal modelling uncertainties in Equation 4.16, cannot currently be represented as smooth variations to pass through the unfolding procedure. Therefore their contribution is added to Σ_1 separately. The matrices are calculated independently to account for systematic modelling uncertainties in the $t\bar{t}$ matrix element (Σ_{ME}), parton shower (Σ_{PS}), initial- and final-state radiation (Σ_{IFSR}), and parton distribution functions (Σ_{PDF}). Each is computed by multiplying the measured cross section in each bin by the relative systematic uncertainty and assuming bin-to-bin correlations of 100%.

The χ^2/NDF values for one-dimensional absolute and normalised differential cross sections unfolded to particle level are shown in Tables 4.10 and 4.11, respectively. The χ^2/NDF values for the two-dimensional distributions unfolded to particle level are given in Table 4.12. For differential cross sections unfolded to parton level, the χ^2/NDF values are given in Tables 4.13 and 4.14. The tables also show the corresponding p-values, the probabilities that the χ^2 are larger than or equal to their observed values under the Standard Model hypothesis given by the simulation. The level of compatibility with the various simulations is discussed in Section 4.9.

Observable	PWG+PY8		PWG+PY8 Var. Up		PWG+PY8 Var. Down		aMC@NLO+PY8		SHERPA		PWG+H7	
	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value
N_{jets}	7.4/5	0.19	1.0/5	0.96	19.9/5	**< 0.01**	26.2/5	**< 0.01**	9.5/5	0.09	6.3/5	0.28
$p_{\text{T}}^{t_1}$	21.7/11	**0.03**	18.8/11	0.07	26.5/11	**< 0.01**	14.5/11	0.21	29.3/11	**< 0.01**	8.2/11	0.69
$|y^{t_1}|$	2.8/6	0.83	4.1/6	0.67	3.0/6	0.81	4.6/6	0.59	1.7/6	0.95	4.4/6	0.62
$p_{\text{T}}^{t_2}$	19.9/9	**0.02**	10.7/9	0.29	34.7/9	**< 0.01**	34.4/9	**< 0.01**	2.7/9	0.98	8.4/9	0.49
$|y^{t_2}|$	3.4/6	0.76	2.2/6	0.90	4.0/6	0.68	3.1/6	0.80	5.5/6	0.48	7.0/6	0.32
$p_{\text{T}}^{t\bar{t}}$	5.9/8	0.66	58.7/8	**< 0.01**	7.2/8	0.52	20.1/8	**< 0.01**	27.8/8	**< 0.01**	4.6/8	0.80
$|y^{t\bar{t}}|$	11.7/18	0.86	12.8/18	0.81	12.9/18	0.80	15.2/18	0.65	23.9/18	0.16	11.9/18	0.85
$m^{t\bar{t}}$	17.6/9	**0.04**	12.8/9	0.17	23.1/9	**< 0.01**	22.5/9	**< 0.01**	10.8/9	0.29	10.9/9	0.29
$H_{\text{T}}^{t\bar{t}}$	24.5/11	**0.01**	24.3/11	**0.01**	34.2/11	**< 0.01**	27.3/11	**< 0.01**	16.4/11	0.13	9.8/11	0.55
$\chi^{t\bar{t}}$	3.6/7	0.83	4.8/7	0.69	7.3/7	0.39	4.2/7	0.76	7.7/7	0.36	5.8/7	0.57
$Z^{t\bar{t}}$	4.1/5	0.53	11.8/5	**0.04**	5.7/5	0.33	11.9/5	**0.04**	13.4/5	**0.02**	5.4/5	0.37
$\cos\theta^{\star}$	8.3/8	0.41	4.8/8	0.77	13.9/8	0.09	10.0/8	0.27	15.2/8	0.06	8.2/8	0.42
$\Delta\phi$	4.2/6	0.64	3.9/6	0.69	10.5/6	0.10	31.6/6	**< 0.01**	3.5/6	0.74	3.9/6	0.69
$|y_{\text{boost}}|$	11.6/15	0.71	12.2/15	0.67	12.8/15	0.61	11.1/15	0.75	16.7/15	0.34	11.5/15	0.71
$|P_{\text{out}}|$	2.7/7	0.91	28.3/7	**< 0.01**	6.3/7	0.51	15.1/7	**0.03**	6.8/7	0.45	3.2/7	0.86
$|P_{\text{cross}}|$	4.9/10	0.90	2.7/10	0.99	7.3/10	0.69	2.9/10	0.98	2.9/10	0.98	3.2/10	0.98
$R_{\text{Wb}}^{\text{leading}}$	5.0/6	0.54	3.9/6	0.69	5.5/6	0.48	3.5/6	0.75	8.9/6	0.18	6.5/6	0.37
$R_{\text{Wb}}^{\text{subleading}}$	4.5/6	0.61	3.8/6	0.71	5.4/6	0.49	2.4/6	0.88	3.2/6	0.78	4.0/6	0.68
$R_{\text{Wt}}^{\text{leading}}$	12.8/7	0.08	15.1/7	**0.04**	14.1/7	0.05	12.5/7	0.09	16.8/7	**0.02**	12.2/7	0.09
$R_{\text{Wt}}^{\text{subleading}}$	2.5/6	0.86	1.7/6	0.94	3.7/6	0.71	3.4/6	0.76	6.6/6	0.36	5.6/6	0.47

Table 4.10 Comparison of the measured particle level single differential absolute cross sections with the predictions from several simulations. For each prediction a χ^2 and a p-value is calculated using the covariance matrix of the measured spectrum. The number of degrees of freedom (NDF) is equal to the number of bins in the distribution.

Observable	PWG+PY8 χ^2/NDF	p-value	PWG+PY8 Var. Up χ^2/NDF	p-value	PWG+PY8 Var. Down χ^2/NDF	p-value	aMC@NLO+PY8 χ^2/NDF	p-value	SHERPA χ^2/NDF	p-value	PWG+H7 χ^2/NDF	p-value
N_{jets}	5.0/4	0.29	0.8/4	0.94	13.7/4	**< 0.01**	28.2/4	**< 0.01**	6.5/4	0.17	1.5/4	0.82
$p_{\text{T}}^{t_1}$	15.3/10	0.12	11.6/10	0.31	21.2/10	**0.02**	12.7/10	0.24	31.2/10	**< 0.01**	8.7/10	0.56
$|y^{t_1}|$	1.1/5	0.96	2.1/5	0.83	0.9/5	0.97	2.0/5	0.85	0.9/5	0.97	2.0/5	0.85
$p_{\text{T}}^{t_2}$	18.9/8	**0.02**	9.1/8	0.34	37.7/8	**< 0.01**	44.9/8	**< 0.01**	3.4/8	0.91	10.1/8	0.26
$|y^{t_2}|$	3.8/5	0.57	1.6/5	0.90	4.9/5	0.42	3.1/5	0.68	5.8/5	0.32	4.9/5	0.43
$p_{\text{T}}^{t\bar{t}}$	4.5/7	0.72	18.2/7	**0.01**	10.2/7	0.18	15.9/7	**0.03**	13.2/7	0.07	6.4/7	0.49
$|y^{t\bar{t}}|$	12.0/17	0.80	13.2/17	0.72	12.7/17	0.76	16.2/17	0.51	22.1/17	0.18	11.8/17	0.81
$m^{t\bar{t}}$	19.9/8	**0.01**	12.8/8	0.12	26.9/8	**< 0.01**	26.5/8	**< 0.01**	9.9/8	0.27	10.3/8	0.25
$H_{\text{T}}^{t\bar{t}}$	22.7/10	**0.01**	19.3/10	**0.04**	34.8/10	**< 0.01**	34.6/10	**< 0.01**	18.2/10	0.05	12.9/10	0.23
$\chi^{t\bar{t}}$	3.4/6	0.76	4.0/6	0.68	6.7/6	0.35	4.5/6	0.61	8.5/6	0.20	3.3/6	0.77
$Z^{t\bar{t}}$	3.8/4	0.43	11.6/4	**0.02**	4.9/4	0.30	23.2/4	**< 0.01**	12.7/4	**0.01**	4.4/4	0.35
$\cos\theta^{\star}$	7.6/7	0.37	5.8/7	0.57	9.2/7	0.24	8.1/7	0.32	14.9/7	**0.04**	8.7/7	0.27
$\Delta\phi$	3.8/5	0.57	2.8/5	0.73	8.3/5	0.14	30.7/5	**< 0.01**	3.2/5	0.67	3.6/5	0.60
$|y_{\text{boost}}|$	10.1/14	0.76	11.6/14	0.64	9.9/14	0.77	12.8/14	0.55	15.3/14	0.36	11.4/14	0.66
$|P_{\text{out}}|$	3.4/6	0.76	6.1/6	0.41	7.6/6	0.27	12.4/6	0.05	1.1/6	0.98	2.6/6	0.85
$|P_{\text{cross}}|$	6.0/9	0.74	3.0/9	0.96	8.2/9	0.52	4.8/9	0.85	1.7/9	0.99	2.5/9	0.98
$R_{\text{Wb}}^{\text{leading}}$	4.8/5	0.45	2.8/5	0.73	5.4/5	0.37	2.8/5	0.73	6.7/5	0.25	4.1/5	0.54
$R_{\text{Wb}}^{\text{subleading}}$	4.9/5	0.43	4.2/5	0.52	5.2/5	0.40	2.8/5	0.72	4.1/5	0.54	2.0/5	0.84
$R_{\text{Wt}}^{\text{leading}}$	14.4/6	**0.03**	16.2/6	**0.01**	14.9/6	**0.02**	16.7/6	**0.01**	18.8/6	**< 0.01**	15.8/6	**0.01**
$R_{\text{Wt}}^{\text{subleading}}$	3.2/5	0.67	2.7/5	0.75	4.0/5	0.56	4.2/5	0.52	7.5/5	0.18	3.5/5	0.63

Table 4.11 Comparison of the measured particle level single differential normalised cross-sections with the predictions from several simulations. For each prediction a χ^2 and a p-value is calculated using the covariance matrix of the measured spectrum. The number of degrees of freedom (NDF) is equal to the number of bins in the distribution minus one.

Observables	PWG+PY8 χ^2/NDF	PWG+PY8 p-value	PWG+PY8 Var. Up χ^2/NDF	PWG+PY8 Var. Up p-value	PWG+PY8 Var. Down χ^2/NDF	PWG+PY8 Var. Down p-value	aMC@NLO+PY8 χ^2/NDF	aMC@NLO+PY8 p-value	SHERPA χ^2/NDF	SHERPA p-value	PWG+H7 χ^2/NDF	PWG+H7 p-value
$p_T^{t_1}$ vs. $p_T^{t_2}$	21.7/12	**0.04**	25.1/12	**0.01**	32.0/12	**< 0.01**	33.2/12	**< 0.01**	29.8/12	**< 0.01**	7.3/12	0.83
$p_T^{t_1}$ vs. $m^{t\bar{t}}$	26.9/11	**< 0.01**	18.8/11	0.07	37.9/11	**< 0.01**	36.9/11	**< 0.01**	13.7/11	0.25	13.8/11	0.24
$p_T^{t_2}$ vs. $m^{t\bar{t}}$	15.3/12	0.23	7.3/12	0.83	31.2/12	**< 0.01**	26.3/12	**< 0.01**	10.9/12	0.53	8.4/12	0.76
$p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$	35.9/11	**< 0.01**	96.3/11	**< 0.01**	35.5/11	**< 0.01**	33.5/11	**< 0.01**	48.4/11	**< 0.01**	17.3/11	0.10
$|y^{t\bar{t}}|$ vs. $m^{t\bar{t}}$	35.3/24	0.06	25.8/24	0.36	46.5/24	**< 0.01**	38.0/24	**0.03**	36.5/24	0.05	18.4/24	0.78
$p_T^{t_1}$ vs. N_{jets}	28.4/19	0.08	20.0/19	0.40	47.0/19	**< 0.01**	60.6/19	**< 0.01**	38.2/19	**< 0.01**	22.9/19	0.24
$p_T^{t_2}$ vs. N_{jets}	26.7/14	**0.02**	22.2/14	0.08	45.7/14	**< 0.01**	88.2/14	**< 0.01**	30.1/14	**< 0.01**	31.5/14	**< 0.01**
$p_T^{t\bar{t}}$ vs. N_{jets}	34.6/11	**< 0.01**	52.7/11	**< 0.01**	60.0/11	**< 0.01**	135.0/11	**< 0.01**	33.6/11	**< 0.01**	17.8/11	0.09
$\Delta\phi$ vs. N_{jets}	42.1/12	**< 0.01**	23.6/12	**0.02**	77.5/12	**< 0.01**	120.0/12	**< 0.01**	20.6/12	0.06	24.7/12	**0.02**
$|P_{\text{out}}|$ vs. N_{jets}	49.6/14	**< 0.01**	50.2/14	**< 0.01**	78.8/14	**< 0.01**	90.0/14	**< 0.01**	38.1/14	**< 0.01**	17.4/14	0.23
$|P_{\text{cross}}|$ vs. N_{jets}	20.5/13	0.08	6.0/13	0.95	43.9/13	**< 0.01**	44.4/13	**< 0.01**	23.5/13	**0.04**	9.4/13	0.74

(a) Absolute

Observables	PWG+PY8 χ^2/NDF	PWG+PY8 p-value	PWG+PY8 Var. Up χ^2/NDF	PWG+PY8 Var. Up p-value	PWG+PY8 Var. Down χ^2/NDF	PWG+PY8 Var. Down p-value	aMC@NLO+PY8 χ^2/NDF	aMC@NLO+PY8 p-value	SHERPA χ^2/NDF	SHERPA p-value	PWG+H7 χ^2/NDF	PWG+H7 p-value
$p_T^{t_1}$ vs. $p_T^{t_2}$	16.6/11	0.12	21.7/11	**0.03**	25.5/11	**< 0.01**	36.7/11	**< 0.01**	29.5/11	**< 0.01**	10.2/11	0.52
$p_T^{t_1}$ vs. $m^{t\bar{t}}$	31.6/10	**< 0.01**	16.8/10	0.08	47.6/10	**< 0.01**	46.6/10	**< 0.01**	11.2/10	0.34	15.5/10	0.11
$p_T^{t_2}$ vs. $m^{t\bar{t}}$	21.1/11	**0.03**	6.3/11	0.85	39.8/11	**< 0.01**	36.0/11	**< 0.01**	8.9/11	0.64	8.3/11	0.69
$p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$	50.6/10	**< 0.01**	104.0/10	**< 0.01**	52.8/10	**< 0.01**	42.1/10	**< 0.01**	62.3/10	**< 0.01**	28.1/10	**< 0.01**
$|y^{t\bar{t}}|$ vs. $m^{t\bar{t}}$	34.0/23	0.07	24.4/23	0.38	43.5/23	**< 0.01**	40.8/23	**0.01**	39.6/23	**0.02**	20.8/23	0.60
$p_T^{t_1}$ vs. N_{jets}	22.5/18	0.21	22.0/18	0.23	34.1/18	**0.01**	65.5/18	**< 0.01**	40.1/18	**< 0.01**	24.9/18	0.13
$p_T^{t_2}$ vs. N_{jets}	25.8/13	**0.02**	14.6/13	0.34	48.9/13	**< 0.01**	113.0/13	**< 0.01**	18.6/13	0.14	23.4/13	**0.04**
$p_T^{t\bar{t}}$ vs. N_{jets}	28.8/10	**< 0.01**	16.7/10	0.08	55.8/10	**< 0.01**	144.0/10	**< 0.01**	21.7/10	**0.02**	11.9/10	0.29
$\Delta\phi$ vs. N_{jets}	32.3/11	**< 0.01**	22.6/11	**0.02**	59.6/11	**< 0.01**	140.0/11	**< 0.01**	25.7/11	**< 0.01**	20.9/11	**0.03**
$|P_{\text{out}}|$ vs. N_{jets}	53.2/13	**< 0.01**	35.3/13	**< 0.01**	85.0/13	**< 0.01**	111.0/13	**< 0.01**	28.4/13	**< 0.01**	26.3/13	**0.02**
$|P_{\text{cross}}|$ vs. N_{jets}	14.3/12	0.28	6.0/12	0.92	30.4/12	**< 0.01**	50.4/12	**< 0.01**	16.9/12	0.15	8.4/12	0.76

(b) Normalised

Table 4.12 Comparison of the measured particle level double differential (a) absolute and (b) normalised cross sections with the predictions from several simulations. For each prediction a χ^2 and a p-value is calculated using the covariance matrix of the measured spectrum. The number of degrees of freedom (NDF) is equal to the number of bins in the distribution.

Observable	PWG+PY8 χ^2/NDF	p-value	PWG+PY8 Var. Up χ^2/NDF	p-value	PWG+PY8 Var. Down χ^2/NDF	p-value	aMC@NLO+PY8 χ^2/NDF	p-value	SHERPA χ^2/NDF	p-value	PWG+H7 χ^2/NDF	p-value
$p_\mathrm{T}^{t_1}$	31.2/10	**< 0.01**	38.0/10	**< 0.01**	35.4/10	**< 0.01**	16.7/10	0.08	60.4/10	**< 0.01**	20.3/10	**0.03**
$|y^{t_1}|$	2.8/7	0.91	2.8/7	0.91	2.8/7	0.90	2.8/7	0.90	2.6/7	0.92	2.7/7	0.91
$p_\mathrm{T}^{t_2}$	27.7/8	**< 0.01**	9.9/8	0.27	56.1/8	**< 0.01**	32.9/8	**< 0.01**	27.8/8	**< 0.01**	17.0/8	**0.03**
$|y^{t_2}|$	5.3/6	0.51	5.3/6	0.50	5.1/6	0.53	5.7/6	0.46	5.8/6	0.45	5.2/6	0.51
$p_\mathrm{T}^{t\bar{t}}$	6.4/5	0.27	44.4/5	**< 0.01**	5.8/5	0.32	52.4/5	**< 0.01**	24.2/5	**< 0.01**	8.3/5	0.14
$|y^{t\bar{t}}|$	8.2/12	0.77	8.0/12	0.79	8.3/12	0.76	8.2/12	0.77	9.1/12	0.69	8.2/12	0.77
$m^{t\bar{t}}$	24.5/9	**< 0.01**	26.7/9	**< 0.01**	22.4/9	**< 0.01**	26.5/9	**< 0.01**	29.1/9	**< 0.01**	24.6/9	**< 0.01**
$H_\mathrm{T}^{t\bar{t}}$	35.5/11	**< 0.01**	25.9/11	**< 0.01**	53.3/11	**< 0.01**	33.0/11	**< 0.01**	49.0/11	**< 0.01**	24.3/11	**0.01**
$\chi^{t\bar{t}}$	3.5/7	0.84	3.1/7	0.88	5.7/7	0.58	4.7/7	0.69	3.2/7	0.87	3.0/7	0.88
$\Delta\phi$	5.6/6	0.47	3.0/6	0.81	9.2/6	0.16	21.1/6	**< 0.01**	14.1/6	**0.03**	8.4/6	0.21
$|y_\mathrm{boost}|$	13.7/15	0.55	13.2/15	0.58	13.7/15	0.55	13.5/15	0.56	15.1/15	0.45	13.5/15	0.56

(a) Absolute

Observable	PWG+PY8 χ^2/NDF	p-value	PWG+PY8 Var. Up χ^2/NDF	p-value	PWG+PY8 Var. Down χ^2/NDF	p-value	aMC@NLO+PY8 χ^2/NDF	p-value	SHERPA χ^2/NDF	p-value	PWG+H7 χ^2/NDF	p-value
$p_\mathrm{T}^{t_1}$	26.5/9	**< 0.01**	31.4/9	**< 0.01**	28.8/9	**< 0.01**	18.3/9	**0.03**	47.6/9	**< 0.01**	20.3/9	**0.02**
$|y^{t_1}|$	2.6/6	0.86	2.6/6	0.86	2.7/6	0.85	2.5/6	0.87	2.4/6	0.88	2.4/6	0.88
$p_\mathrm{T}^{t_2}$	18.9/7	**< 0.01**	8.0/7	0.33	38.2/7	**< 0.01**	22.9/7	**< 0.01**	18.2/7	**0.01**	12.8/7	0.08
$|y^{t_2}|$	5.3/5	0.38	5.3/5	0.38	5.1/5	0.40	5.6/5	0.34	5.9/5	0.32	5.2/5	0.39
$p_\mathrm{T}^{t\bar{t}}$	7.1/4	0.13	49.8/4	**< 0.01**	6.5/4	0.17	59.1/4	**< 0.01**	27.2/4	**< 0.01**	9.4/4	0.05
$|y^{t\bar{t}}|$	9.0/11	0.63	8.7/11	0.65	9.0/11	0.63	9.0/11	0.62	9.9/11	0.54	8.9/11	0.63
$m^{t\bar{t}}$	34.8/8	**< 0.01**	38.1/8	**< 0.01**	31.6/8	**< 0.01**	38.2/8	**< 0.01**	41.4/8	**< 0.01**	35.1/8	**< 0.01**
$H_\mathrm{T}^{t\bar{t}}$	34.0/10	**< 0.01**	27.0/10	**< 0.01**	48.1/10	**< 0.01**	35.4/10	**< 0.01**	43.4/10	**< 0.01**	26.4/10	**< 0.01**
$\chi^{t\bar{t}}$	3.9/6	0.69	2.9/6	0.82	6.0/6	0.43	4.6/6	0.60	3.6/6	0.73	3.1/6	0.80
$\Delta\phi$	5.2/5	0.40	3.2/5	0.67	7.4/5	0.19	22.8/5	**< 0.01**	14.0/5	**0.02**	7.4/5	0.19
$|y_\mathrm{boost}|$	14.7/14	0.40	14.3/14	0.43	14.6/14	0.40	14.7/14	0.40	16.2/14	0.30	14.6/14	0.41

(b) Normalised

Table 4.13 Comparison of the measured parton level (a) absolute and (b) normalised single differential cross sections with the predictions from several simulations. For each prediction a χ^2 and a p-value is calculated using the covariance matrix of the measured spectrum. The number of degrees of freedom (NDF) is equal to the number of bins in the distribution.

Observables	PWG+PY8		PWG+PY8 Var. Up		PWG+PY8 Var. Down		aMC@NLO+PY8		Sherpa		PWG+H7	
	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value
$p_T^{t_1}$ vs. $p_T^{t_2}$	34.9/12	**< 0.01**	41.5/12	**< 0.01**	55.9/12	**< 0.01**	66.3/12	**< 0.01**	54.5/12	**< 0.01**	25.6/12	**0.01**
$\|y^{t_2}\|$ vs. $\|y^{t_1}\|$	7.6/16	0.96	8.2/16	0.94	7.4/16	0.96	9.5/16	0.89	6.2/16	0.99	8.4/16	0.94
$p_T^{t_1}$ vs. $m^{t\bar{t}}$	40.8/10	**< 0.01**	36.4/10	**< 0.01**	61.7/10	**< 0.01**	66.5/10	**< 0.01**	62.4/10	**< 0.01**	31.4/10	**< 0.01**
$\|y^{t_1}\|$ vs. $m^{t\bar{t}}$	23.0/11	**0.02**	21.5/11	**0.03**	24.8/11	**< 0.01**	29.6/11	**< 0.01**	27.9/11	**< 0.01**	20.2/11	**0.04**
$p_T^{t_2}$ vs. $m^{t\bar{t}}$	34.4/13	**< 0.01**	21.1/13	0.07	61.3/13	**< 0.01**	44.5/13	**< 0.01**	34.6/13	**< 0.01**	26.2/13	**0.02**
$\|y^{t_2}\|$ vs. $m^{t\bar{t}}$	34.9/11	**< 0.01**	30.6/11	**< 0.01**	39.5/11	**< 0.01**	40.7/11	**< 0.01**	38.2/11	**< 0.01**	32.9/11	**< 0.01**
$p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$	33.2/12	**< 0.01**	60.9/12	**< 0.01**	33.4/12	**< 0.01**	89.4/12	**< 0.01**	44.5/12	**< 0.01**	36.3/12	**< 0.01**
$\|y^{t\bar{t}}\|$ vs. $m^{t\bar{t}}$	39.5/11	**< 0.01**	41.1/11	**< 0.01**	36.6/11	**< 0.01**	47.4/11	**< 0.01**	43.0/11	**< 0.01**	39.5/11	**< 0.01**

(a) Absolute

Observables	PWG+PY8		PWG+PY8 Var. Up		PWG+PY8 Var. Down		aMC@NLO+PY8		Sherpa		PWG+H7	
	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value	χ^2/NDF	p-value
$p_T^{t_1}$ vs. $p_T^{t_2}$	34.6/11	**< 0.01**	55.1/11	**< 0.01**	52.6/11	**< 0.01**	78.2/11	**< 0.01**	62.6/11	**< 0.01**	28.8/11	**< 0.01**
$\|y^{t_2}\|$ vs. $\|y^{t_1}\|$	6.6/15	0.97	7.0/15	0.96	6.6/15	0.97	8.2/15	0.92	5.5/15	0.99	7.0/15	0.96
$p_T^{t_1}$ vs. $m^{t\bar{t}}$	46.4/9	**< 0.01**	37.4/9	**< 0.01**	72.0/9	**< 0.01**	91.9/9	**< 0.01**	61.6/9	**< 0.01**	40.0/9	**< 0.01**
$\|y^{t_1}\|$ vs. $m^{t\bar{t}}$	24.6/10	**< 0.01**	23.6/10	**< 0.01**	25.6/10	**< 0.01**	33.3/10	**< 0.01**	30.2/10	**< 0.01**	22.2/10	**0.01**
$p_T^{t_2}$ vs. $m^{t\bar{t}}$	34.2/12	**< 0.01**	24.3/12	**0.02**	54.7/12	**< 0.01**	46.3/12	**< 0.01**	34.3/12	**< 0.01**	28.9/12	**< 0.01**
$\|y^{t_2}\|$ vs. $m^{t\bar{t}}$	39.9/10	**< 0.01**	35.3/10	**< 0.01**	44.5/10	**< 0.01**	47.5/10	**< 0.01**	43.9/10	**< 0.01**	37.9/10	**< 0.01**
$p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$	40.5/11	**< 0.01**	74.8/11	**< 0.01**	40.6/11	**< 0.01**	114.0/11	**< 0.01**	55.4/11	**< 0.01**	45.0/11	**< 0.01**
$\|y^{t\bar{t}}\|$ vs. $m^{t\bar{t}}$	44.8/10	**< 0.01**	46.8/10	**< 0.01**	41.0/10	**< 0.01**	55.0/10	**< 0.01**	49.4/10	**< 0.01**	45.1/10	**< 0.01**

(b) Normalised

Table 4.14 Comparison of the measured parton level (a) absolute and (b) normalised double differential cross-sections with the predictions from several simulations. For each prediction a χ^2 and a p-value is calculated using the covariance matrix of the measured spectrum. The number of degrees of freedom (NDF) is equal to the number of bins in the distribution minus one.

4.9 Discussion

A simulation is said to agree with the observed unfolded data if the p-value is larger than the conventional cut-off of 0.05. The χ^2 test statistics indicate that the absolute and normalised differential cross section measurements are sometimes contradictory. That is, there are several cases where the absolute cross section predicted by a simulation agrees with data, but the corresponding normalised cross section does not, or *vice versa*.

Particle level

Universally, the POWHEG+HERWIG7 simulation best agrees with data, with only the normalised cross section in R_{Wt}^{leading} showing incompatibility at particle level. However, this cross section cannot be well-described by any of the simulations tested. $H_{\text{T}}^{t\bar{t}}$ and $m^{t\bar{t}}$ are the next-worst modelled observables, whereas rapidity-based variables such as $|y^{t_1}|$, $|y^{t_2}|$, $|y^{t\bar{t}}|$, $\chi^{t\bar{t}}$, and $|y_{\text{boost}}|$ show good modelling performance. The best described observable is $|P_{\text{cross}}|$, which is a new variable only accessible in this analysis channel.

The nominal POWHEG+PYTHIA8 simulation also shows mostly good agreement with the data, although it describes the transverse momenta of the t quarks weakly. The predictions from MADGRAPH_aMC@NLO+PYTHIA8 and POWHEG+PYTHIA8 Var3cDown do not agree with the data in many distributions.

For the double differential cross sections summarised in Table 4.12, there is a larger discrepancy between the predictions from simulation and the measured data. The MADGRAPH_aMC@NLO+PYTHIA8 and POWHEG+PYTHIA8 Var3cDown simulations fail to sufficiently model any of the measured double differential cross sections. POWHEG+HERWIG7 again gives the best performance, but it displays discrepancies with both absolute and normalised cross sections in $p_{\text{T}}^{t_2}$ and $\Delta\phi$ in bins of N_{jets}, plus with normalised cross sections in $p_{\text{T}}^{t\bar{t}}$ in bins of $m^{t\bar{t}}$ and, interestingly, $|P_{\text{out}}|$ in bins of N_{jets}. None of the simulations model the cross sections in $\Delta\phi$ vs. N_{jets} or $p_{\text{T}}^{t\bar{t}}$ vs. $m^{t\bar{t}}$ to good agreement.

Parton level

The values for the χ^2 test statistic in Table 4.13 show that there is a similar level of (dis)agreement between each of the simulations and the data for one-dimensional differential cross sections. In general, the normalised cross sections show worse agreement across all distributions, since the measurement uncertainty is reduced by the normalisation. Therefore the higher precision of measurements for normalised differential cross sections means that the data can better discriminate between the simulation models, preferring POWHEG+HERWIG7 and POWHEG+PYTHIA8 in this case.

Universally, $m^{t\bar{t}}$ and $H_{\mathrm{T}}^{t\bar{t}}$ are modelled poorly at parton level. Also the transverse momenta of the t quarks are poorly modelled at parton level by all generators except for the subleading t quark p_{T} by POWHEG+PYTHIA8 Var3cUp.

For the two-dimensional differential cross sections at parton level, poor agreement is seen in all distributions for all simulations except for the cross section in y^{t_2} in bins of y^{t_1}, which demonstrates excellent agreement with all simulations. This result involves two well-measured angular observables and agrees with the findings from the results at particle level, that rapidity is generally modelled well. The simultaneous lack of agreement in the other cross sections means that the data does not discriminate between the simulation models.

4.9.1 Comparison to measurements in the lepton+jets channel

Many of the differential cross sections presented in this chapter are also measured by a complementary ATLAS analysis in the lepton+jets $t\bar{t}$ decay channel [271]. There are several differences between the analysis strategies. The phase space of the lepton+jets analysis is significantly larger, since leptons and jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ pass selection, compared to $55\,\mathrm{GeV}$ for the analysis in this chapter. Therefore even with the larger all-hadronic branching ratio for $t\bar{t}$, the lepton+jets channel dataset contains significantly more events after selection. On the other hand, the analysis presented in this chapter achieves better resolution in many variables, particularly those involving jet directions due to the lack of missing transverse energy. This

means that the differential cross sections are often presented with much finer binnings in the bulk of the distributions. These various effects combine non-trivially, making it difficult to conclude which analysis provides the better discrimination between theoretical predictions, without a complex combination of the results.

A preliminary comparison of the results allows a few conclusions to be made. Similar levels of agreement between data and predictions are seen across both analyses for the and POWHEG+HERWIG7 and POWHEG+PYTHIA8 simulations, except for in $p_T^{t\bar{t}}$ and N_{jets} for POWHEG+HERWIG7. That is, where one analysis demonstrates a good agreement between a particular model and the data for a distribution, so does the other.

For measurements of two-dimensional differential cross sections, a similar level of compatibility is again observed between the two analyses. For example, the cross section as a function of $\Delta\phi$ in bins of N_{jets} cannot be described well by any simulation models. Similar results are obtained for $|P_{\text{out}}|$ and $p_T^{t\bar{t}}$ in bins of N_{jets}, while the hadronic channel presented here gives better discrimination between the models for $|y^{t\bar{t}}|$ vs. $m^{t\bar{t}}$. Both analyses indicate that $p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$ is described poorly by all models.

Consistent mismodelling is observable in both analyses for the cross sections in $p_T^{t_1}$, $p_T^{t_2}$, and $p_T^{t\bar{t}}$. In bins where the hadronic analysis is statistically limited by the number of events passing selection, the lepton+jets analysis shows more disagreement between the data and predictions.

While both analyses observe the distribution of $m^{t\bar{t}}$ to disagree with predictions, they do so in opposite fashions. In the fully hadronic channel the measured distribution has smaller values on average than the data, whereas for the lepton+jets analysis $m^{t\bar{t}}$ is found to be generally larger than predicted.

At parton level, the analyses are also compatible in suggesting that few observables can be modelled accurately. For example, the cross section as a function of $m^{t\bar{t}}$ shows significant disagreement between data and all the simulated predictions. When considering the double differential cross section results, both analyses indicate that none of the models is able to predict the data distribution better than the others.

Overall, both the hadronic and lepton+jets analyses suggest that the POWHEG+HERWIG7 and POWHEG+PYTHIA8 simulation models provide the best predictions for $t\bar{t}$ differential production cross sections. In some cases the analyses are complementary, with each providing better power for discrimination between the models in different observable distributions.

4.9.2 Outlook

This chapter presents the first comprehensive measurements of differential $t\bar{t}$ production cross sections in the fully hadronic channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$. The cross sections are reported in one- and two-dimensional distributions of multiple observables, unfolded to both particle level and parton level. Some of the results can provide the power to discriminate between theoretical predictions from simulated models. These results may also be used to improve the modelling of t quarks, useful for many other future experiments.

Some of the results presented show poor agreement between predictions and observations, so can provide an indication of where the most impactful improvements to modelling could be made. In particular, the two-dimensional differential cross sections unfolded to particle level will be very useful for improving the modelling in regions of phase space containing many additional jets from initial- and final-state QCD radiation.

The results at parton level are compared to the most accurate theoretical calculations currently available. They could be used in the future to perform measurements of parton distribution functions, and extract a value for the t quark pole mass.

Despite non-negligible irreducible background contributions, effective event selection and background estimation strategies permit state-of-the-art precision measurements, able to constrain theoretical predictions in events with many hadronic jets. However the analysis presented here is severely limited by the efficiency of the hadronic trigger used to select events. Future upgrades to the detector, such as the ability to use tracking information early in the trigger chain, will improve the efficiency for this trigger strategy. This will also improve the state of jet energy calibration, pileup suppression, and b-tagging. A multi-jet plus heavy-flavour

tag trigger strategy is particularly motivated here, which in turn will allow more precise measurements of $t\bar{t}$ production cross sections.

While the analysis presented here motivates the definition of the bin edges for the reported distributions using detector resolution, the procedure still requires a 'human in the loop' to check and modify them at many stages in the analysis. In future versions of this analysis, an end-to-end binning solution could be developed to adjust the number and locations of the bins according to the final post-unfolding experimental covariance matrix. This would allow correlations between bin populations, introduced by the unfolding procedure, to be taken into account.

Two-dimensional differential cross sections are reported in this analysis, often with N_{jets} as the external variable at particle level. With the efficiency gains and dataset size increase from future runs of ATLAS, it may be possible to report two-dimensional kinematic cross sections (e.g. $p_{\text{T}}^{t_1}$ vs. $p_{\text{T}}^{t_2}$) in bins of N_{jets}. These three-dimensional distributions may be used to more powerfully discriminate between approaches to $t\bar{t}$ modelling, especially when including extra jets.

Similarly, future iterations of the analysis presented in this chapter could make use of identified radiative 'extra' jets in the event. The ability to better determine correlations between the kinematics of the t quarks, $t\bar{t}$ system, and extra jets could be used to further improve future theoretical models. Measurements of some such variables are only possible with the fully hadronic $t\bar{t}$ decay mode, since the leading fermion from the decay could be an invisible neutrino in other channels. Since the initial state radiation scales linearly with the partonic centre-of-momentum energy, it is possible that the leading extra jet is also the highest-p_{T} object in the event. This makes it a good candidate for providing a reference for the energy scale of the event. Additionally, correlations between the extra jets and t quark kinematics are particularly useful for modelling since the radiative emission is expected to be approximately collinear. The p_{T} ratios between extra jets in $t\bar{t}$ events is particularly sensitive to soft gluon emissions that may not be resolved as jets by the detector. This allows tests of resummation effects.

Finally, the results presented in this chapter can provide greater power to discriminate between and improve modelling approaches in combination with the differential cross sections measured in complementary $t\bar{t}$ decay channels. While current approaches to combine distributions from different analyses require the binning definitions to be carefully harmonised, the novel unfolding method presented in Chapter 6 allows the distributions to be rebinned after unfolding, thereby providing flexibility in combining differential cross sections from different analysis or experiments.

Chapter 5

Gaussian processes

Gaussian processes (GPs) are a general class of descriptions of functions. They have a simple and compact definition, yet may be used to describe non-parametrically a broad range of functions. They can be conveniently interpreted as distributions over functions.

As is the case with many machine learning methods, GPs can be used for classification and regression. For the generally multi-class classification task, the aim is to label data points according to the category they belong to. An example from high energy physics is the sorting of particle collision events according to the physical process from which they originate. Another example is the classification of objects (e.g. tracks, calorimeter hits) in a detector according to the type of particle that produced them, known as *tagging*.

The aim of regression is essentially to estimate a function. The task is to predict the function's output at values of the input generally not measured before. After defining commonly used notation in Section 5.1 and a formal definition of a Gaussian process in Section 5.2, the use of GPs in regression is discussed in Section 5.4. An application of using GP regression for the optimisation of computer simulations is presented in Section 5.4.5.

5.1 Notation

Scalar quantities are written as lowercase letters in normal weight, e.g. x. A column vector is represented by boldface $\mathbf{x}$ with its ith element denoted x_i. A capital letter is used to represent a matrix, e.g. X, whose elements are written X_{ij}. The jth column of the matrix Y is the vector $\mathbf{y}_j$, with ith element Y_{ij} by definition. Terms and expressions used commonly in this chapter are listed in Table 5.1.

5.2 Definition of a Gaussian process

A *random process* is a collection of indexed random variables. In the context of physical random processes the index is often interpretable as position or time, although in general it may be any real-valued control variable. A *Gaussian process* (GP) is therefore a collection of indexed random variables, any finite subset of which have a joint Gaussian distribution [319].

A GP may be considered as a distribution on an infinite-dimensional space of functions; a generalisation of a Gaussian distribution over a finite vector space [320]. Consider the set of d-dimensional indices $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Then a GP is a distribution over functions f such that $f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n)$ are distributed according to a joint Gaussian [321]. In physical stochastic processes, for example Brownian motion, the index is often time or space, but may in general be any d-dimensional real variable.

Just as a multivariate Gaussian distribution is entirely defined by a mean vector and covariance matrix, a GP over f is completely described by a mean function and covariance function. These are defined as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \tag{5.1}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathrm{cov}[f(\mathbf{x}), f(\mathbf{x}')] \tag{5.2}$$

$$= \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \tag{5.3}$$

Term	Description
$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$	Multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ
A^{T}	Transpose of the matrix A
$|A|$	Determinant of the matrix A
$\mathbf{v}^{\mathsf{T}}$	Row vector. Transpose of the column vector $\mathbf{v}$
$\mathbb{E}[x]$	Expectation value of the random variable x
$\mathrm{cov}[x, y]$	Covariance of random variables x and y
$\mathrm{cov}[\mathbf{x}]$	Shorthand for the covariance matrix with elements $V_{ij} = \mathrm{cov}[x_i, x_j]$
$\hat{x}$	Estimator for the random variable x
$\mathbf{x}$	Column vector representing a single d-dimensional index
$\mathbf{y}$	Vector of n observations
X	$d \times n$ matrix of n indices. The jth column represents a single index $\mathbf{x}_j$
$\mathbf{f}(X)$	Vector of function values with $f_i = f(\mathbf{x}_i)$
$\overline{\mathbf{f}}$	Mean values of $\mathbf{f}$
$\mathbf{x}_*$	A single prediction point
X_*	Matrix of prediction points
$k(\mathbf{x}, \mathbf{x}')$	Kernel function evaluated at $\mathbf{x}$ and $\mathbf{x}'$
$K(X, X')$	Matrix of kernel values with elements $K_{ij} = k(\mathbf{x}_i, \mathbf{x}'_j)$
K	Shorthand for the square matrix $K(X, X)$
K_*	Shorthand for $K(X, X_*)$
K_{**}	Shorthand for the square matrix $K(X_*, X_*)$

Table 5.1 Definitions of commonly used symbols in this chapter.

Then the GP is written $f \sim \mathrm{GP}(m, k)$, which is understood to mean that the function f is sampled from the GP defined by m and k. The covariance function is often called the *kernel function* (or simply *kernel*) in the literature, and these terms will be used interchangeably in this thesis.

5.3 Probabilistic model

Because GPs can be considered as probability distributions, it is useful to derive some common properties from probability theory. To do this, some notation is first introduced.

Consider the vector of n function values $\mathbf{f}(X) = (f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_n))^{\mathsf{T}}$, where X is the $d \times n$ *design matrix*

$$
X = \begin{pmatrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{d1} & X_{d2} & \ldots & X_{dn} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{pmatrix}, \tag{5.4}
$$

and similarly the mean vector $\mathbf{m}(X) = (m(\mathbf{x}_1), m(\mathbf{x}_2), \ldots, m(\mathbf{x}_n))^{\mathsf{T}}$. The matrix K is defined with respect to the kernel function as

$$
K(X, X') = \begin{pmatrix} k(\mathbf{x}_1, \mathbf{x}_1') & k(\mathbf{x}_1, \mathbf{x}_2') & \ldots & k(\mathbf{x}_1, \mathbf{x}_m') \\ k(\mathbf{x}_2, \mathbf{x}_2') & k(\mathbf{x}_2, \mathbf{x}_2') & \ldots & k(\mathbf{x}_2, \mathbf{x}_m') \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1') & k(\mathbf{x}_n, \mathbf{x}_2') & \ldots & k(\mathbf{x}_n, \mathbf{x}_m') \end{pmatrix} \tag{5.5}
$$

for X containing n indices and X' containing m indices.

The function values $\mathbf{f}(X)$ are distributed according to a joint Gaussian with mean vector $\mathbf{m}(X)$ and $n \times n$ symmetric covariance matrix $K(X, X)$. This is written

$$\mathbf{f}(X) \sim \mathcal{N}\big(\mathbf{m}(X),\, K(X, X)\big) \tag{5.6}$$

and the associated probability density is given by

$$P\big(\mathbf{f}(X) \,|\, \mathbf{m}(X), K(X, X)\big) = \frac{1}{\sqrt{(2\pi)^n\, |K(X, X)|}}$$
$$\times \exp\left\{ -\frac{1}{2}\left[\mathbf{f}(X) - \mathbf{m}(X)\right]^{\mathsf{T}} K(X, X)^{-1} \left[\mathbf{f}(X) - \mathbf{m}(X)\right] \right\}. \tag{5.7}$$

Since the only design matrix used here is X, this symbol may be dropped to give a more compact notation: $\mathbf{f} \sim \mathcal{N}(\mathbf{m}, K)$. Taking logarithms of Equation 5.7, the *log prior probability density* is

$$\log P(\mathbf{f} \,|\, \mathbf{m}, K) = -\frac{1}{2}(\mathbf{f} - \mathbf{m})^{\mathsf{T}} K^{-1}(\mathbf{f} - \mathbf{m}) - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi. \tag{5.8}$$

The term *prior probability* is used here in the sense that no observations have been included in the model. In regression, the prior is updated using data to yield a posterior probability, as discussed in Section 5.4.

5.3.1 Sampling from a Gaussian process

It is sometimes useful to visualise a subset of the functions described by a GP by sampling from it. A simple and efficient algorithm achieves this [319]. Consider a GP given by $\mathbf{m}$ and K (or equivalently $m(\mathbf{x})$, $k(\mathbf{x}, \mathbf{x}')$, and X), where a single sampled function is evaluated at n indices. Firstly, a sample of n random numbers is drawn from the standard Gaussian distribution: $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, I)$. This then undergoes a linear transformation into $\mathbf{f} = \mathbf{m} + L\mathbf{u}$, where $LL^{\mathsf{T}} = K$. The values $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ are now distributed according to a joint Gaussian with mean $\mathbf{m}$ and covariance matrix K, as desired. The matrix L is usually taken to be the Cholesky decomposition [322] of the positive semi-definite covariance matrix K, and can informally be
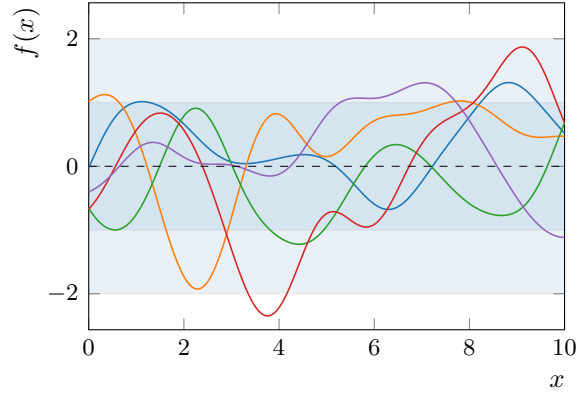
Figure 5.1 Five functions (solid) randomly sampled from a GP with constant zero mean (dashed) and the squared exponential covariance function given by Equation 5.10. The functions are evaluated at 200 points distributed linearly along the horizontal axis. The darker shaded area represents the 1σ uncertainty band (68% confidence region), while the lighter shaded area represents the 2σ uncertainty band (95% confidence region).

considered the square root of K. As a Cholesky factor, L is a lower triangular matrix and provides an efficient method to determine $\frac{1}{2}\log|K|$,

$$\frac{1}{2}\log|K| = \sum_{i=1}^{n} \log L_{ii}. \tag{5.9}$$

This is required in the evaluation of $\log P(\mathbf{f} \mid \mathbf{m}, K)$ in Equation 5.8.

As a simple example, this sampling procedure was followed for five draws from a one-dimensional GP with constant zero mean and kernel function given by a squared exponential,

$$k(x, x') = \exp\left[-\frac{1}{2}(x - x')^2\right]. \tag{5.10}$$

This kernel gives a variance of 1 for $x' = x$, and a smoothly falling covariance as x and x' move further apart. For $|x - x'| = 1$, the covariance is $1/\sqrt{e} \approx 0.607$. A more general version of the squared exponential, along with other kernels, is given in Section 5.6. The functions sampled from this GP are evaluated at 200 points spaced evenly in the range $[0, 10]$. This was done using the NumPy Python package [323, 324], and the sampled functions are shown in Figure 5.1.

5.4 Gaussian process regression

The regression task can be stated as follows: given observations of a dependent variable y at values of the independent variable $\mathbf{x}$, predict the value of the underlying function $f(\mathbf{x}_*)$ at some arbitrary new value $\mathbf{x}_*$ [325].

Section 5.4.1 introduces regression with GPs for the case of noiseless observations, where the function to be estimated is known exactly at the measured points. This situation is extended to the case of noisy observations, which is widely discussed in the literature [319, 320, 325], in Section 5.4.2. Finally, a further generalisation is made to consider the case of correlated data in Section 5.4.3. This important result is used in Chapter 6.

5.4.1 Noiseless observations

In the case where the n observations are noiseless, the underlying function is known exactly at those points: $\mathbf{f}(X) = \mathbf{y}$. The predictive values of the underlying function at previously unseen indices are given by $\mathbf{f}(X_*)$. Therefore the joint distribution of the observed and predictive points according to the GP prior is

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}(X_*) \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mathbf{m}(X) \\ \mathbf{m}(X_*) \end{pmatrix}, \begin{pmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{pmatrix} \right], \tag{5.11}$$

where the K matrices are linked to the kernel function k as defined by Equation 5.5. Using a compact notation (see Table 5.1), this is written

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \mathbf{m} \\ \mathbf{m}_* \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_*^\mathsf{T} & K_{**} \end{pmatrix} \right]. \tag{5.12}$$

The predictive distribution is the conditional distribution for $\mathbf{f}_*$ given the observations $\mathbf{y}$. It is said that the prior GP is *conditioned* on the observations to give the posterior GP. It can be shown [319, 326] that this posterior distribution is a GP given by

$$\mathbf{f}_* \mid \mathbf{y} \sim \mathcal{N}\left(\overline{\mathbf{f}}_*,\ \mathrm{cov}[\mathbf{f}_*]\right), \tag{5.13}$$

$$\text{where} \quad \overline{\mathbf{f}}_* = K_*^\mathsf{T}\, K^{-1}\,(\mathbf{y} - \mathbf{m}) + \mathbf{m}_*, \tag{5.14}$$

$$\mathrm{cov}[\mathbf{f}_*] = K_{**} - K_*^\mathsf{T}\, K^{-1} K_*. \tag{5.15}$$

As an example, a zero-mean GP with squared exponential kernel given by Equation 5.10 was used as a prior. The posterior process, conditioned on an increasing number of noiseless data points, is shown in Figures 5.2a to 5.2e. The uncertainty of the predictive distribution at $\mathbf{x}_*$ vanishes approaching an observed point $\mathbf{x}$, reflecting the exact knowledge of the underlying function at that point. Far away from observations, the uncertainty remains large. This corresponds to the lack of knowledge of the underlying function in those regions.

5.4.2 Noisy observations

The result from Section 5.4.1 is often extended to the case where the observations y_i are subject to some independent identically-distributed Gaussian noise ϵ with variance σ_ϵ^2, such that

$$y_i = f(\mathbf{x}_i) + \epsilon(\mathbf{x}_i) \tag{5.16}$$

$$\mathrm{cov}[y_i, y_j] = k(\mathbf{x}_i, \mathbf{x}_j) + \sigma_\epsilon^2 \delta_{ij}, \tag{5.17}$$

where δ_{ij} is the Kronecker delta. Since the underlying function $f(\mathbf{x})$ is now hidden from direct measurement, it is sometimes referred to as the *latent function*.

Equation 5.12 is modified to reflect this contribution from the noise,

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{f}_* \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{m} \\ \mathbf{m}_* \end{pmatrix}, \begin{pmatrix} K + \sigma_\epsilon^2 I & K_* \\ K_*^\mathsf{T} & K_{**} \end{pmatrix} \right], \tag{5.18}$$
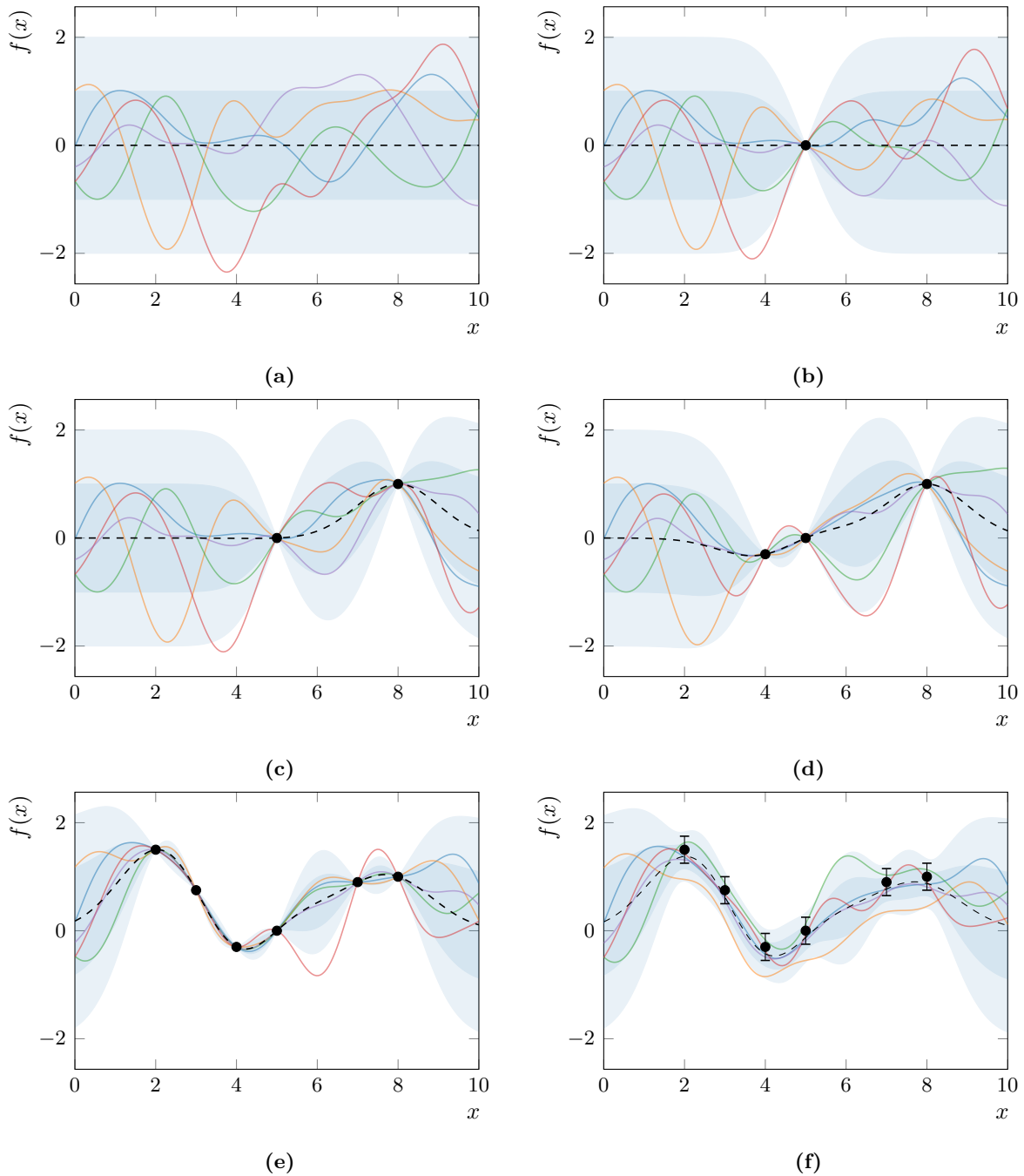
Figure 5.2 Predictive mean $\mathbf{f}_* \mid \mathbf{y}$ (dashed) and functions sampled from the GP (solid). The darker and lighter shaded areas indicate the 1σ and 2σ uncertainty bands, respectively. The GP prior with constant zero mean and unit variance is shown in Panel (a). Panels (b) to (e) show the posterior distribution for the GP conditioned on an increasing number of noiseless data points. The same observations with Gaussian noise term $\sigma_\epsilon = 0.25$ are used to derive the posterior distribution shown in Panel (f).

and the predictive distribution is therefore given by

$$\mathbf{f}_* \mid \mathbf{y} \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, \operatorname{cov}[\mathbf{f}_*]\right), \tag{5.19}$$

$$\text{where} \quad \bar{\mathbf{f}}_* = K_*^{\mathsf{T}} \left[K + \sigma_\epsilon^2 I\right]^{-1} (\mathbf{y} - \mathbf{m}) + \mathbf{m}_*, \tag{5.20}$$

$$\operatorname{cov}[\mathbf{f}_*] = K_{**} - K_*^{\mathsf{T}} \left[K + \sigma_\epsilon^2 I\right]^{-1} K_*. \tag{5.21}$$

The regression example in Section 5.4.1 was repeated with Gaussian noise with standard deviation $\sigma_\epsilon = 0.25$. The predictive distribution given by Equations 5.19 to 5.21 is shown in Figure 5.2f. Owing to the increased uncertainty on the observations, the predictive mean function no longer passes exactly through the data points.

5.4.3 Correlated observations

Finally, GP regression may be further generalised to include the case where the observations are correlated [319, 327, 328]. In this situation, the covariance of two observations can be written

$$\operatorname{cov}[y_i, y_j] = k(\mathbf{x}_i, \mathbf{x}_j) + V_{ij}, \tag{5.22}$$

where V_{ij} are elements of the positive semi-definite covariance matrix V. The predictive distribution is then given by

$$\mathbf{f}_* \mid \mathbf{y} \sim \mathcal{N}\left(\bar{\mathbf{f}}_*, \operatorname{cov}[\mathbf{f}_*]\right), \tag{5.23}$$

$$\text{where} \quad \bar{\mathbf{f}}_* = K_*^{\mathsf{T}} \left[K + V\right]^{-1} (\mathbf{y} - \mathbf{m}) + \mathbf{m}_*, \tag{5.24}$$

$$\operatorname{cov}[\mathbf{f}_*] = K_{**} - K_*^{\mathsf{T}} \left[K + V\right]^{-1} K_*. \tag{5.25}$$

This important result is used in Chapter 6.

5.4.4 Smoothing properties

A feature of GP regression is that it smooths out fluctuations in the data, i.e. it is a *regulariser*. This can be understood by considering the simple case of observations with Gaussian independent identically-distributed noise, presented in Section 5.4.2. The same argument applies in the general case of observations with a general covariance matrix V.

From Equation 5.20, a zero-mean GP regression for noisy observations has the predictive mean function

$$\overline{f}(\mathbf{x}_*) = K_*^{\mathsf{T}} \left[K + \sigma_\epsilon^2 I \right]^{-1} \mathbf{y} \tag{5.26}$$

$$= \mathbf{h}(\mathbf{x}_*)^{\mathsf{T}} \mathbf{y}, \tag{5.27}$$

where $\mathbf{h}(\mathbf{x}_*) = \left[K + \sigma_\epsilon^2 I \right]^{-1} K_*$ is a column vector for a single test point $\mathbf{x}_*$. The vector of functions $\mathbf{h}$ is independent of the data points and is known as the *weight function* [329]. Equation 5.27 shows that $\overline{f}(\mathbf{x}_*)$ is a linear weighted sum of the data, therefore GP regression is considered a *linear smoother* [330].

The smoothing properties of GP regression can be understood through the eigendecomposition of the matrix K with eigenvalues and eigenvectors $\{(\lambda_i, \mathbf{v}_i)\}$:

$$K = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^{\mathsf{T}}. \tag{5.28}$$

When trained on noisy data, a zero-mean GP regression gives the predictive mean values

$$\overline{\mathbf{f}} = K \left[K + \sigma_\epsilon^2 I \right]^{-1} \mathbf{y} \tag{5.29}$$

at the training points, where all functions are evaluated at X. The data can be represented in the eigenbasis of K as $\mathbf{y} = \sum_{i=1}^{n} \gamma_i \mathbf{v}_i$ with coefficients $\gamma_i = \mathbf{v}_i^{\mathsf{T}} \mathbf{y}$. Substituting this and Equation 5.28 into Equation 5.29, the predictive mean values can be written as

$$\overline{\mathbf{f}} = \sum_{i=1}^{n} \frac{\gamma_i \lambda_i}{\lambda_i + \sigma_\epsilon^2} \mathbf{v}_i. \tag{5.30}$$

Since K is a positive semi-definite covariance matrix, its eigenvalues λ_i are real and non-negative. Therefore the term $\lambda_i/(\lambda_i + \sigma_\epsilon^2) < 1$ for non-zero noise. This means that $\bar{\mathbf{f}}$ is a damped representation of the data along the principal components of the covariance matrix K. Furthermore, most practical kernel functions have larger eigenvalues for more slowly varying eigenfunctions (i.e. eigenfunctions with fewer roots) [319], so high-frequency components of the data are damped more strongly. This has the effect of smoothing out fine structure in the observations. In this sense, GP regression may be considered as a low-pass filter.

5.4.5 Application: Efficient Bayesian optimisation of computer simulations

An example application of the result given by Equations 5.13 to 5.15 is in the optimisation of a computer simulation [331, 332]. Since the return value of a program is the same each time it is run, the observations in this case are noiseless. (This assumes identical initial conditions and the absence of stochastic effects in the simulation, but noisy simulation results can also be considered and treated using the augmented regression procedure for noisy observations discussed in Section 5.4.2.)

The problem definition is as follows: find the minimum point of the *objective function* $f(\mathbf{x})$, given by

$$\mathbf{x}^\star = \arg\min_{\mathbf{x}} f(\mathbf{x}). \tag{5.31}$$

The objective function is considered a black-box function, with no closed form or gradient information available. This prohibits the use of gradient-based optimisation algorithms. It is also very expensive to calculate, in that it uses a large amount of time, energy, or money, and $\mathbf{x}$ often spans a large number of dimensions. For these reasons, standard approaches such as grid search or random search can be wasteful and time-consuming.

This situation often arises in machine learning research and engineering, with the design of a neural network, for example. In this case the objective function would be the cross-validation loss [333], and $\mathbf{x}$ is the set of hyperparameters (such as the number of hidden layers) of the

model. The training time can often be a number of hours or days, so it is desirable to minimise the number of times the objective function has to be evaluated. Another example from high energy physics is the tuning of a Monte Carlo event generator [334]. Here the objective function is the chi-square, which summarises the level of agreement between the simulation's predictions and experimental data. The number of tuning parameters is large (there are 20 continuous parameters for PYTHIA 8.1 [335] used in the Monash tune [334, 336], for example) so $\mathbf{x}$ is in a high-dimensional space, which discourages the use of exhaustive search methods. Again, this program is expensive to run and the number of iterations used to determine the optimal parameters should be kept to a minimum.

At the nth iteration, the objective function is estimated by a probabilistic *surrogate model* formed of a GP conditioned on the observations $\{(\mathbf{x}_i, y_i)\}$ for $i = 1, \ldots, n$. This model is calculated from Equations 5.13 to 5.15. For each step, an inexpensive *acquisition function* is evaluated to determine the value of $\mathbf{x}_{n+1}$, the next point to be evaluated. One such acquisition function is the *expected improvement* [337–340] over the minimum value evaluated so far, $y_{\min}$,

$$a_{\mathrm{EI}}(\mathbf{x}; \xi) = \mathbb{E}\big[y_{\min} - f(\mathbf{x}) - \xi\big] \tag{5.32}$$

$$= \sigma(\mathbf{x}) \left[\gamma(\mathbf{x}; \xi)\, \Phi\big(\gamma(\mathbf{x}; \xi)\big) + \phi\big(\gamma(\mathbf{x}; \xi)\big) \right], \tag{5.33}$$

$$\text{where} \quad \gamma(\mathbf{x}; \xi) = \frac{y_{\min} - \overline{f}(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}. \tag{5.34}$$

Here $\overline{f}(\mathbf{x})$ and $\sigma(\mathbf{x}) = \sqrt{\mathrm{cov}[f(\mathbf{x}), f(\mathbf{x})]}$ are the mean and standard deviation of the surrogate GP model evaluated at $\mathbf{x}$. $\phi(\mathbf{x})$ and $\Phi(\mathbf{x})$ are the probability density function and cumulative distribution function of the standard Gaussian distribution, respectively. The next sample point is then chosen as the value of $\mathbf{x}$ that maximises the acquisition function,

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}} a_{\mathrm{EI}}(\mathbf{x}; \xi). \tag{5.35}$$

The parameter ξ determines the trade-off between *exploration* and *exploitation*: larger values assign more importance to areas of high uncertainty (exploration), whereas smaller values increase the relative importance of areas of potentially closer to the true minimum of the objective function (exploitation). It is typically set to approximately 0.01 [339].

The procedure performs a Bayesian update step at each iteration. To start, the unconditioned prior probability distribution is chosen to be a sufficiently uninformative GP, with mean and covariance functions covering the expected range of the objective function. At the nth iteration, Bayes' theorem is applied to derive the posterior probability distribution, given a Gaussian likelihood for the observed point $(\mathbf{x}_n, y_n)$, to obtain the posterior GP given by Equations 5.13 to 5.15. At each iteration the hyperparameters of the surrogate model are also adjusted to the point that maximises the marginal likelihood, as defined in Equation 5.47. This GP model is then used as the prior for the $n + 1$th step, where it is updated with the measured point $(\mathbf{x}_{n+1}, y_{n+1})$.

In addition to finding a good estimate for $\mathbf{x}^\star$ in the parameter space, it is often desirable to find a solution that is efficient to compute in terms of CPU time, energy consumption, monetary cost, or some other quantity under budget. For the case of reducing CPU walltime, this is done by penalising points in parameter space that yield a long evaluation time for $f(\mathbf{x})$. Along with the objective function $f(\mathbf{x})$, a *duration function* $c(\mathbf{x})$ can be estimated by a GP model at each iteration of the optimisation algorithm [338]. Assuming that $f(\mathbf{x})$ and $c(\mathbf{x})$ are independent, it is simple to compute the *expected improvement per second*, $a_{\mathrm{EI}}(\mathbf{x}; \xi)/c(\mathbf{x})$. The maximum of this quantity then gives the next point to evaluate.

As an illustrative example, a simple analytic function was used as a proxy for a black-box objective function, shown in Figure 5.3. GP-based optimisation was performed using the Scikit-Optimize [337] Python implementation with $a_{\mathrm{EI}}(x, 0.01)$, defined in Equation 5.32, as the acquisition function. This algorithm found an acceptable global minimum in 15 iterations.

5.5 Mean function

In the context of regression, it is common to use a GP with constant mean function $m(\mathbf{x}) = 0$ as the prior over functions [319, 341]. This is because in practical applications the kernel function (described in Section 5.6) provides enough flexibility to adequately update the prior given the data, resulting in a posterior process with non-zero mean function.
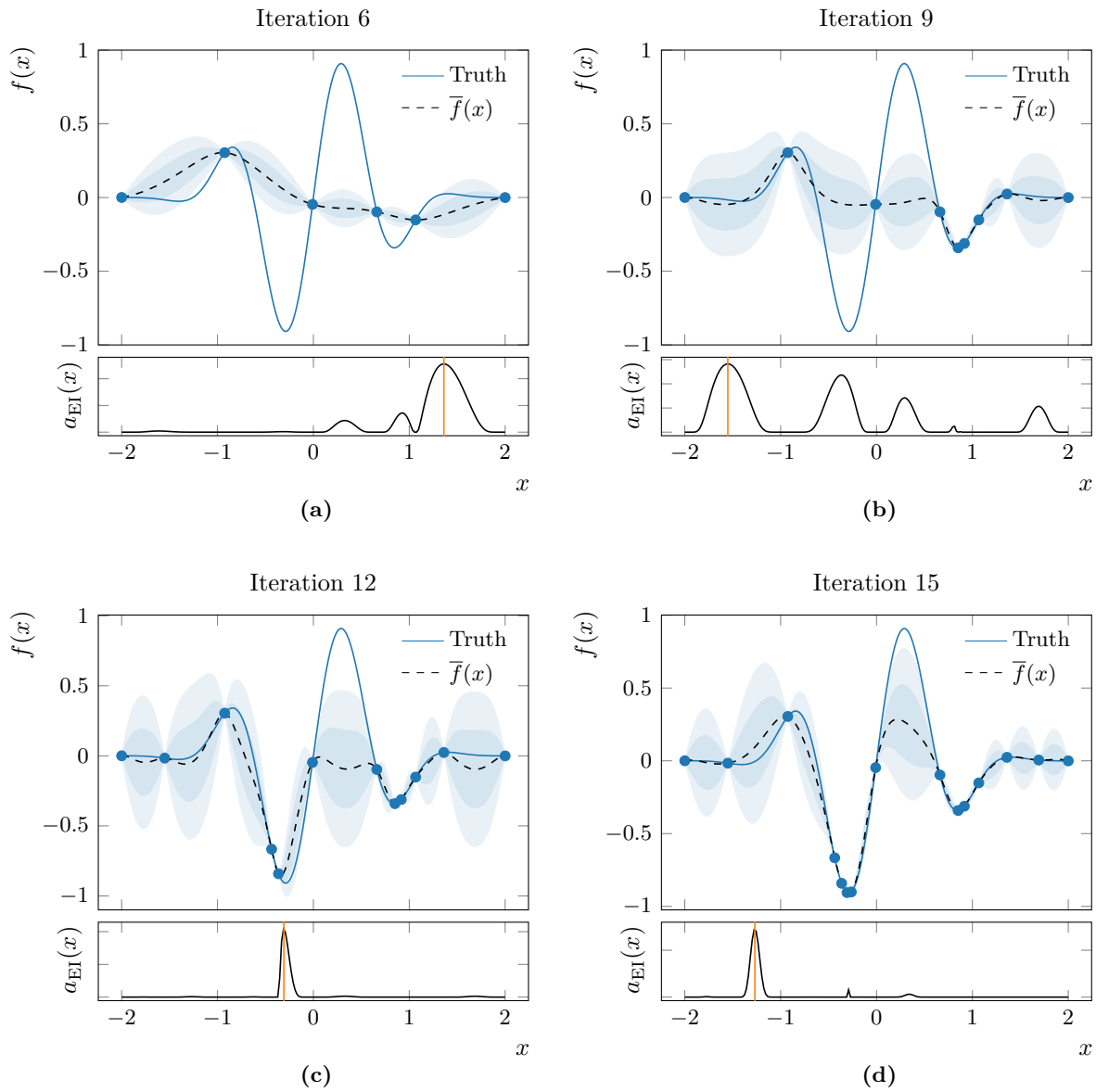
Figure 5.3 One-dimensional toy example of optimisation of a black box function with GP regression. The goal is to find the global minimum of the hidden truth function (dashed line), while minimising the number of function evaluations. A GP (solid line and shaded regions) conditioned on the observations so far (circles) is used as the surrogate model, updated at each step. At each iteration, the acquisition function (bottom panels) is evaluated and its maximum (vertical line) is chosen as the next function evaluation point.

Additionally, any desired structure or behaviours in the functions described by the GP can be specified by the kernel function. Therefore the vast majority of the interesting behaviour of a GP for regression is described by its covariance, and the mean function has little impact in most situations.

It is illuminating to note that the results given by Equations 5.13 to 5.15, 5.19 to 5.21, and 5.23 to 5.25 are identical to those obtained by modelling the residuals $\mathbf{y} - \mathbf{m}$ with a zero-mean GP, then adding $\mathbf{m}_*$ to the predictive mean. This means that for regression, a global mean function can be subtracted from the data before modelling the residual process with a zero-mean GP.

5.6 Kernel function

The kernel function is crucial to the definition of a GP as it encodes prior knowledge and assumptions of the underlying objective function [319]. It does this by describing the similarity of two points $\mathbf{x}$ and $\mathbf{x}'$ in the index space. For example, a reasonable assumption may be that points that are close in $\mathbf{x}$ will also be close in y, and therefore will be highly correlated. Conversely, points that are distant might be assumed to be independent and will therefore have a small correlation. The distance metric and characteristic length scales are provided in the kernel function, so it wholly encapsulates the meaning of similarity for the model. In the context of Bayesian regression, the kernel function can be interpreted as the covariance of the prior process.

A kernel function $k(\mathbf{x}, \mathbf{x}')$ is said to be *stationary* if it is a function of only $\mathbf{x} - \mathbf{x}'$; that is, it is invariant to translations in the continuous index space. Kernel functions of only $\|\mathbf{x} - \mathbf{x}'\|$ are *isotropic* and are classed as *radial basis functions*. It is possible that the kernel function is parameterised by some set of parameters $\boldsymbol{\theta}$. This is denoted by $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$. The contents of $\boldsymbol{\theta}$ are referred to as the *hyperparameters* of the GP model.

In the definition of a GP, the kernel function provides the covariance of the joint Gaussian distribution for $\mathbf{f}$. Therefore there are constraints on its form such that it corresponds to a valid

covariance. Since for any covariance function $\text{cov}[\mathbf{x}, \mathbf{x}'] = \text{cov}[\mathbf{x}', \mathbf{x}]$, k must also be symmetric. The *Gram matrix*, with elements $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, is a valid covariance matrix if it is positive semi-definite, i.e. $\mathbf{v}^\mathsf{T} K \mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^n$. This constrains k to functions which give a Gram matrix with non-negative eigenvalues.

5.6.1 Examples of kernels

In this section, some examples of commonly encountered kernel functions are presented. Functions drawn from a zero-mean GP corresponding to each kernel, using the sampling procedure described in Section 5.3.1, are shown in Table 5.2.

Constant

The constant kernel has a single parameter and is given by

$$k(\mathbf{x}, \mathbf{x}') = \sigma_0^2. \tag{5.36}$$

Processes with this kernel describe a family of constant functions with a fixed covariance between them. On its own, the constant kernel is not particularly interesting or useful. However, it may be combined with other kernels to give an overall amplitude (via multiplication) or bias (via addition) [321].

Linear

Linear kernels are defined by functions that only depend on their arguments through the inner product, $\mathbf{x} \cdot \mathbf{x}'$. The *homogeneous* linear kernel is given by $k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$. Allowing for a general covariance matrix Σ and combining this with a constant kernel results in the general inhomogeneous linear kernel,

$$k(\mathbf{x}, \mathbf{x}') = \sigma_0^2 + \mathbf{x}^\mathsf{T} \Sigma \mathbf{x}'. \tag{5.37}$$

A zero-mean GP regression with this kernel function is identical to the result obtained with Bayesian linear regression [341]. That is, this is the covariance relation for estimators at $\mathbf{x}$ and $\mathbf{x}'$ given by the linear regression $\hat{y}(\mathbf{x}) = \mathbf{w}^\mathsf{T}\mathbf{x} + \epsilon$ with priors $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and $\epsilon \sim \mathcal{N}(0, \sigma_0^2)$. Polynomial kernels, $k(\mathbf{x}, \mathbf{x}') = (\sigma_0^2 + \mathbf{x}^\mathsf{T}\Sigma\mathbf{x}')^p$ for integer p, are also valid covariance functions [319].

Matérn

The *Matérn* kernels are a class of stationary kernel functions, expressed in terms of $r = \|\mathbf{x} - \mathbf{x}'\|$. A general expression is given by [319, 342]

$$k_\nu(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}r}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}r}{l}\right), \tag{5.38}$$

where Γ is the gamma function [343], K_ν is a modified Bessel function [344], and ν and l are positive real parameters. These covariance functions relate to processes with spectral density $S_\nu(\omega) = (1 + \omega^2)^{\frac{1}{2}-\nu}$ [345], which are the solutions to general Laplacian stochastic partial differential equations [346]. The Matérn family of kernel functions are therefore widely applicable to many natural random processes.

The parameter l may be interpreted naturally as the characteristic length scale of the random process described by the corresponding GP. Special cases are encountered for $\nu = p + \frac{1}{2}$, $p = 0, 1, 2, \ldots$, when Equation 5.38 becomes [319, 347]

$$k_{\nu=p+\frac{1}{2}}(r) = \exp\left(-\frac{\sqrt{2\nu}r}{l}\right) \frac{p!}{(2p)!} \sum_{q=0}^{p} \frac{(p+q)!}{i!\,(p+q)!} \left(\frac{\sqrt{8\nu}r}{l}\right)^{p-q}. \tag{5.39}$$

The cases where $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$ are expanded upon below. For values of $\nu > \frac{5}{2}$, the resulting GP is often indistinguishable [319] from that which uses the squared exponential kernel obtained in the limit where $\nu \to \infty$, also included here.

- $(\nu = \frac{1}{2})$ *Ornstein–Uhlenbeck*

$$k_{\frac{1}{2}}(r) = \exp\left(-\frac{r}{l}\right) \tag{5.40}$$

This kernel is named after the work of Ornstein and Uhlenbeck [348] in the mathematical modelling of Brownian motion. Specifically, it models mean-reverting Gauss–Markov processes [349], and has applications in the modelling of financial markets [350].

- $(\nu = \frac{3}{2}, \frac{5}{2})$

$$k_{\frac{3}{2}}(r) = \left(1 + \frac{\sqrt{3}r}{l}\right) \exp\left(-\frac{\sqrt{3}r}{l}\right) \tag{5.41}$$

$$k_{\frac{5}{2}}(r) = \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right) \tag{5.42}$$

The Matérn-$\frac{3}{2}$ and Matérn-$\frac{5}{2}$ kernels are commonly used in applications of machine learning [319, 327] and spatial statistics [351], owing to their properties under differentiation [342]. They do not make as strong smoothness assumptions as the squared exponential kernel (below), but provide ample flexibility for most practical datasets.

- $(\nu \to \infty)$ *Squared exponential*

$$\lim_{\nu \to \infty} k_{\nu}(r) = \exp\left(-\frac{r^2}{2l^2}\right) \tag{5.43}$$

The squared exponential kernel is obtained from the Matérn class in the limit as $\nu \to \infty$. Since the kernel function has infinitely many derivatives, functions sampled from a GP with this kernel are smooth. The squared exponential covariance function is used expansively in the literature.

Gibbs

An extension to the squared exponential kernel function can be made to account for a varying length scale $l_i(\mathbf{x})$. This is given by the *Gibbs kernel* [352], with positive length functions

$l_i(\mathbf{x})$, $i = 1, \ldots, d$ (one for each axis of $\mathbf{x}$),

$$k(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{d} \sqrt{\frac{2\, l_i(\mathbf{x})\, l_i(\mathbf{x}')}{l_i^2(\mathbf{x}) + l_i^2(\mathbf{x}')}} \exp\left(-\frac{|x_i - x_i'|^2}{l_i^2(\mathbf{x}) + l_i^2(\mathbf{x}')} \right) \tag{5.44}$$

Here the first multiplicative term in the product is for normalisation, so that $k(\mathbf{x}, \mathbf{x}) = 1$ and the kernel function is positive semi-definite.

5.6.2 Combining kernels

Kernel functions may be combined under certain normalisation rules to describe various behaviours in the resulting GP [321]. Most commonly, the constant kernel is used as a multiplicative factor to give an overall amplitude to the covariance function. For example, combining the constant and squared exponential kernels results in

$$k(\mathbf{x}, \mathbf{x}') = \sigma_0^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2} \right). \tag{5.45}$$

This kernel function describes a GP with characteristic length scale l, and variance $\mathrm{cov}[\mathbf{x}, \mathbf{x}] = \sigma_0^2$. The set of hyperparameters is the union of the hyperparameters of the constituent kernels, in this case $\boldsymbol{\theta} = \{\sigma_0^2, l\}$.

5.6.3 Kernels for applications in high energy physics

In limited applications to date, Gaussian processes have been used in high energy physics to perform efficient and robust background modelling [353] and unfolding [354, 355], which is the focus of Chapter 6. Combinations of kernel functions are used to specify an adequate prior probability distribution for the data. In the stationary case, the kernel function in Equation 5.45 is mostly used. For non-stationary problems, the Gibbs kernel in Equation 5.44 (including an amplitude factor) is used to allow a varying length scale, usually with a simple linear parameterisation $l(x) = bx + c$.

Kernel name	$k(x, x')$	Covariance	Samples from GP	Stationary
Linear	$\sigma_0^2 + x \cdot x'$			No
Ornstein–Uhlenbeck	$\exp\left(-\dfrac{\|x - x'\|}{l}\right)$			Yes
Squared exponential	$\exp\left(-\dfrac{\|x - x'\|^2}{2l^2}\right)$			Yes
Gibbs	$\sqrt{\dfrac{2\,l(x)\,l(x')}{l^2(x) + l^2(x')}}\,\exp\left(-\dfrac{\|x - x'\|^2}{l^2(x) + l^2(x')}\right)$			No

Table 5.2 Properties of common kernel functions in one dimension and GPs defined by them.

Empirically, the choice of such standard kernels with sensible hyperparameters often leads to a sufficiently noninformative prior probability distribution for the problem at hand. Additionally, there exist methods to inform the selection of a suitable kernel function, such as nested Bayesian search [319] or automatic model search [321].

In certain cases, it may be appropriate to assign a systematic uncertainty for the kernel choice. Here the difference in marginal likelihood from Equation 5.47 can be used to quantify the uncertainty.

5.7 Hyperparameter optimisation

In machine learning, the aim of model selection is to minimise the *generalisation error*. This in turn maximises the ability of a particular model to make accurate predictions that agree with previously unseen data. There is therefore a natural desire to simultaneously minimise both bias and variance of the resulting estimator. For many families of models, including GPs, it is not possible to simultaneously minimise these two properties. Instead, there is a *bias–variance trade-off* whereby they are antagonistic properties and a penalty is paid for highly prioritising one over the other. The choice of the optimal trade-off is an open problem, and there are many approaches taken in the literature [319, 356–358]. There are many choices pertaining to model selection; the decision to use a GP is the first, if the setting is appropriate. Then the mean and kernel functions must be chosen. If the kernel is parameterised, as is usually the case, these hyperparameters must be chosen to optimise the bias–variance trade-off with respect to some predefined criteria. In this section, focus is placed on an approximate Bayesian method to determine the optimal hyperparameters for a particular choice of kernel.

Bayesian statistics, combined with the simple algebra of GPs, provides a convenient and efficient method for model selection. This is based on the *marginal likelihood*, obtained by integrating the likelihood for Gaussian-distributed data, $\mathbf{y} \,|\, \mathbf{f} \sim \mathcal{N}(\mathbf{f}, V)$, times the prior

(Equation 5.8), over the latent function $f(\mathbf{x})$ of the GP,

$$P(\mathbf{y} \,|\, \mathbf{m}, K) = \int P(\mathbf{y} \,|\, \mathbf{f}) \, P(\mathbf{f} \,|\, \mathbf{m}, K) \, \mathrm{d}\mathbf{f}. \tag{5.46}$$

Here all the indices used are the training set, X. After taking logarithms and writing the covariance matrix's dependence on the hyperparameters as $K_{\boldsymbol{\theta}} = k_{\boldsymbol{\theta}}(X, X)$, the log marginal likelihood is given by

$$\log P(\mathbf{y} \,|\, \mathbf{m}, K_{\boldsymbol{\theta}}) = -\frac{1}{2}(\mathbf{y} - \mathbf{m})^{\mathsf{T}} \left[K_{\boldsymbol{\theta}} + V\right]^{-1} (\mathbf{y} - \mathbf{m}) - \frac{1}{2} \log |K_{\boldsymbol{\theta}} + V| - \frac{n}{2} \log 2\pi. \tag{5.47}$$

The method of maximum marginal likelihood [319] is an approximate approach for finding an optimal set of hyperparameters $\boldsymbol{\theta}$. The proper Bayesian treatment of hyperparameters would integrate $\boldsymbol{\theta}$ out of any expressions used for prediction [359], but this integral is usually intractable. The *evidence approximation* [360, 361] uses Laplace's method [362–364] to replace the integral with the value of the marginal likelihood at its maximum, which is valid for suitably peaked $P(\boldsymbol{\theta} \,|\, \mathbf{y})$. Then the optimal hyperparameters are those which maximise the marginal likelihood, or $\log P(\mathbf{y} \,|\, \mathbf{m}, K_{\boldsymbol{\theta}})$ given by Equation 5.47. This method is also known as *type-II maximum likelihood* [365] or *empirical Bayes* [366], and has a long-standing history of applications in statistical inference and machine learning.

It is possible to interpret separately each of the terms in Equation 5.47. The first term $-\frac{1}{2}(\mathbf{y} - \mathbf{m})^{\mathsf{T}} \left[K_{\boldsymbol{\theta}} + V\right]^{-1} (\mathbf{y} - \mathbf{m})$ is the least-squares log likelihood, and it quantifies the fit to data of the GP model. That is, models with greater flexibility to describe the observed data give larger values for this term. The second term is related to the simplicity of the model. Models with higher variance give larger values for $|K_{\boldsymbol{\theta}}|$, so $-\frac{1}{2} \log |K_{\boldsymbol{\theta}} + V|$ penalises over complex models. In this sense, the balance between the first two terms of Equation 5.47 represents the bias–variance trade-off involved in model selection. The $-\frac{n}{2} \log 2\pi$ term is for normalisation.

An example of performing this procedure for one hyperparameter is shown in Figure 5.4. Data points were generated by sampling from a one-dimensional GP with the squared exponential kernel (Equation 5.10) with $l = 1.0$ and Gaussian noise $V = \sigma_{\epsilon}^2 I$ with $\sigma_{\epsilon} = \sqrt{0.25}$. GP regression

was performed for these data points, following Section 5.4.2. In Figure 5.4a, a short length scale of $l = 0.25$ was used. Here the fit to data is good, but the variance at points away from the observations is large and the model is overly flexible. In contrast, GP regression with a long length scale of $l = 3.0$ is shown in Figure 5.4b. While the complexity of the model, and accordingly the variance, is greatly reduced, the fit to data is now poor. A numerical maximisation of the marginal likelihood given by Equation 5.47 was performed using the SciPy [367] Python implementation of the Brent–Dekker method [368, 369]. This found that a value of $l = 0.90$ gives the maximum marginal likelihood, close to the original generating value of $l = 1.0$. GP regression with this value is shown in Figure 5.4c, and displays an agreeable trade-off between bias and variance. In Figure 5.4d, the marginal likelihood and its constituent terms are shown as functions of l.

(a) $l = 0.25$

(b) $l = 3.0$

(c) $l = 0.90$

(d)
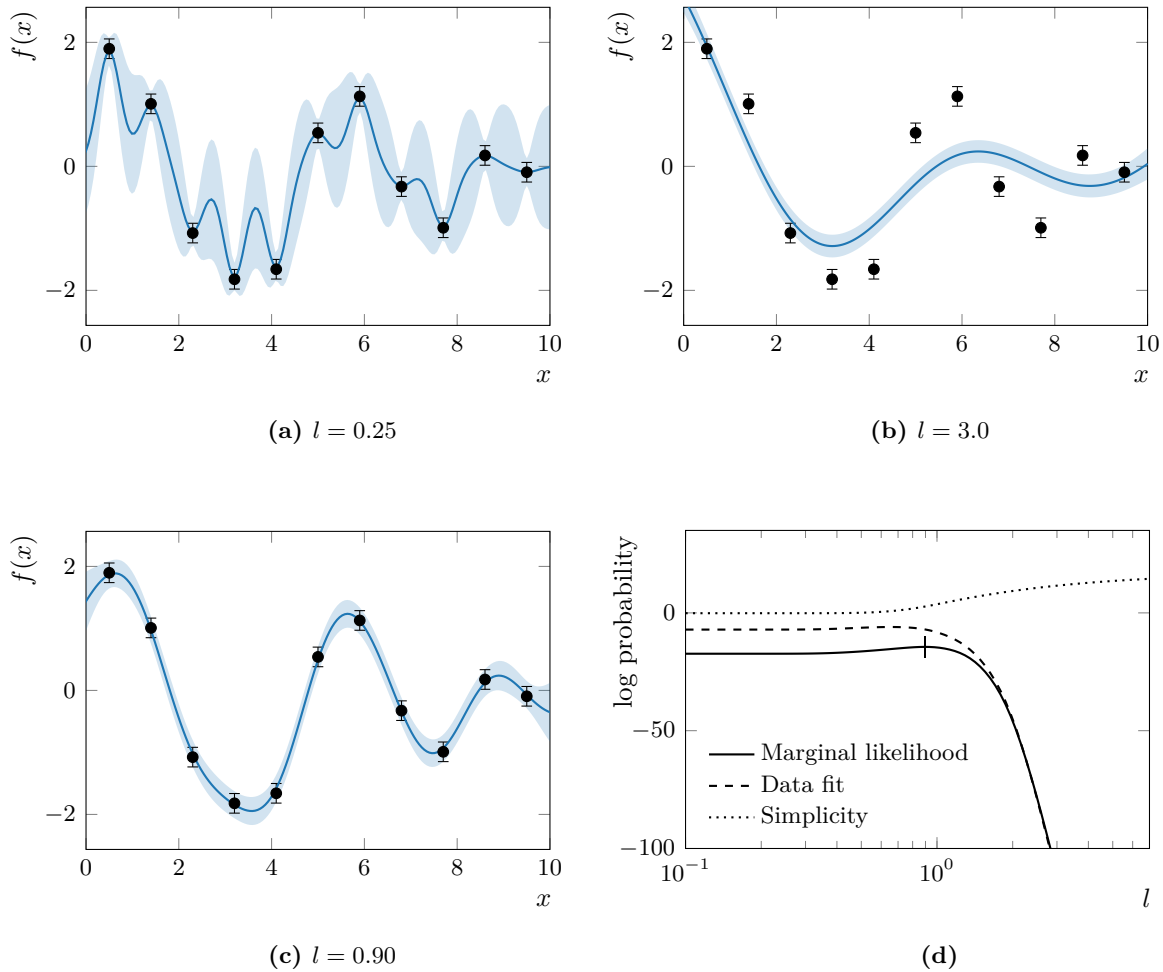
Figure 5.4 Effect of the length scale hyperparameter. Panels (a)–(c): Noisy data ($\sigma_\epsilon = \sqrt{0.25}$) are generated by sampling from a one-dimensional GP using the squared exponential kernel (Equation 5.10) with $l = 1.0$. A GP regression is fit to the data points using short and long length scales in Panels (a) and (b), respectively. In Panel (c), the length scale hyperparameter is set to the value which maximises the marginal likelihood (Equation 5.47). The log marginal likelihood is shown as a function of l in Panel (d) (solid). Its components, the data fit term $-\frac{1}{2}\mathbf{y}^{\mathsf{T}}[K + \sigma_\epsilon^2 I]^{-1}\mathbf{y}$ (dashes) and the simplicity or negative complexity term $-\frac{1}{2}\log|K + \sigma_\epsilon^2 I|$ (dots), are also shown. The point of maximum marginal likelihood is indicated with a vertical mark (|).

Chapter 6

Unfolding with Gaussian processes

The contents of this chapter have also been written in a paper [354] and submitted for peer-reviewed publication.

6.1 Introduction

The distribution of an experimentally measured random variable is distorted by the effects of limited detector resolution, acceptance, and efficiency. This distortion can be in the form of random noise and bias in the distribution. In order to compare measurements of the same underlying physical distribution with different detectors, or to compare to theoretical predictions, it is sometimes desirable to *unfold* the measured distribution.

In high energy particle physics, there is an established literature of a variety of approaches and implementations of unfolding [307, 370–372]. In other fields, the statistical techniques used for unfolding are often called *deconvolution* or *restoration* [373–376]. Universally they aim to solve an *inverse problem*, loosely defined as inference of an *a priori* unknown function which is related to the observed data through a convolutional model.

In this chapter, a newly proposed method for unfolding in particle physics is described. The unfolding problem is defined in in Section 6.2. Section 6.3 presents the maximum likelihood solution, and the need for regularisation. In Section 6.4, a connection is made between the

maximum a posteriori estimator for unfolding and the solution of a regression problem which conditions prior knowledge encoded in a Gaussian process on the maximum likelihood solution. Two example applications are shown in Section 6.5.

6.2 Problem definition

In particle physics, distributions of measured random variables are often reported as binned histograms, rather than continuous functions. Therefore the method presented here will deal with discretised bin populations, where the probability for a measurement to land in bin i is given by

$$P_i = \int\limits_{\text{bin } i} g(x)\,\mathrm{d}x. \tag{6.1}$$

Here the continuous random variable x is distributed according to the probability density function $g(x)$. It is important to note that the process of discretisation introduces a source of implicit regularisation. The method presented in this chapter can be generally applied to continuous, unbinned distributions, but this form of data representation is less common in particle physics.

Four histograms are defined:

- **n**: Data counts, i.e. n_i is the number of events in bin i for $i = 1, \ldots, N$;

- $\boldsymbol{\nu}$: The *reconstruction* histogram, $\boldsymbol{\nu} = E[\mathbf{n}]$, the expected histogram of measured data;

- $\boldsymbol{\mu}$: The *truth* histogram with discretisation $\mu_j = \int_{\text{bin } j} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$, $j = 1, \ldots, M$;

- $\hat{\boldsymbol{\mu}}$: The *unfolded* histogram, an estimator for $\boldsymbol{\mu}$.

The truth and observed histograms are related through the effects of limited detector response, acceptance, and contributions from background processes. For simplicity, the background is taken to be zero here, although this assumption is relaxed in Section 6.6. In the

unfolding problem, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are modelled as being linearly related by

$$\boldsymbol{\nu} = R\boldsymbol{\mu}, , \tag{6.2}$$

where R is an $N \times M$ *response matrix.*, defined to have elements

$$R_{ij} = P(\text{measured value in bin } i \,|\, \text{true value in bin } j). \tag{6.3}$$

The response matrix can be constructed from simulation. This is done by using Monte Carlo sampling methods, as described in Section 3.4, to generate 'true' values for the variable of interest. The same simulated event is also passed through a model of the detector, giving a simulated 'measured' value. The events fill a two-dimensional histogram of true versus measured values of the variable of interest, in N and M bins, respectively.

The elements of the response matrix are taken from the counts of the bin populations after scaling such that

$$\sum_{i=1}^{N} R_{ij} = \varepsilon_j, \tag{6.4}$$

where ε_j is the *efficiency* of bin j, $j = 1, \ldots, M$. Limited efficiency causes events to not be detected. The efficiency of bin j, ε_j, is equal to the proportion of events with true values in that bin which also appear in the simulated measurement. Written another way, $1 - \varepsilon_j$ is the proportion of events that are lost by the detector model. The effects of limited acceptance can be handled in a similar manner.

The goal is to construct a vector of estimators $\hat{\boldsymbol{\mu}}(\mathbf{n})$ for the unfolded histogram, given the observed data. This is an ill-posed inverse problem with no unique solution [377]. There exists a bias–variance trade-in the family of acceptable solutions, and this is typically handled through *regularisation.*

In a counting experiment, commonly encountered in particle physics, the population of the ith bin n_i is an integer random variable which is distributed according to the Poisson

distribution with expectation value ν_i. The probability mass function is given by

$$f(n_i; \nu_i) = \frac{\nu_i^{n_i} e^{-\nu_i}}{n_i!}. \tag{6.5}$$

The n_i are mutually independent for all i. For large values, n_i can be approximated as a real random variable distributed according to a Gaussian distribution with probability density function

$$f(n_i; \nu_i) = \frac{1}{\sqrt{2\pi n_i}} \exp\left[-\frac{(n_i - \nu_i)^2}{2n_i}\right]. \tag{6.6}$$

, where the variance has been estimated by the observed bin count. For the entire histogram, the joint probability density for all N bin populations is given by

$$f(\mathbf{n}; \boldsymbol{\nu}) = \left[(2\pi)^N |V|\right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{n} - \boldsymbol{\nu})^{\mathsf{T}} V^{-1}(\mathbf{n} - \boldsymbol{\nu})\right], \tag{6.7}$$

where

$$V = \begin{pmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_N \end{pmatrix} \tag{6.8}$$

is a diagonal covariance matrix, since the bin populations are independent random variables.

The findings presented in this chapter apply when bin counts are approximately Gaussian, e.g., for large ν_i in the case of Poisson-distributed data.

6.3 Maximum likelihood estimator

Since the data **n** are modelled as being Gaussian-distributed around $\boldsymbol{\nu}$, the likelihood is given by

$$P(\mathbf{n} \,|\, \boldsymbol{\nu}) = \left[(2\pi)^N \,|V|) \right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{n} - \boldsymbol{\nu})^{\mathsf{T}} V^{-1} (\mathbf{n} - \boldsymbol{\nu}) \right] \tag{6.9}$$

and hence the log-likelihood may be written

$$\log P(\mathbf{n} \,|\, \boldsymbol{\mu}) = -\frac{1}{2} (\mathbf{n} - \boldsymbol{\nu})^{\mathsf{T}} V^{-1} (\mathbf{n} - \boldsymbol{\nu}) - \frac{1}{2} \log |V| - \frac{N}{2} \log 2\pi \tag{6.10}$$

$$= -\frac{1}{2} (\mathbf{n} - R\boldsymbol{\mu})^{\mathsf{T}} V^{-1} (\mathbf{n} - R\boldsymbol{\mu}) - \frac{1}{2} \log |V| - \frac{N}{2} \log 2\pi, \tag{6.11}$$

where in the last line $\boldsymbol{\nu} = R\boldsymbol{\mu}$ has been substituted, by Equation 6.2, and $|V|$ is the determinant of the covariance matrix V.

It can be shown that the maximum likelihood solution for $\boldsymbol{\nu}$ is given by [356]

$$\hat{\boldsymbol{\nu}}_{\mathrm{ML}} = \mathbf{n}. \tag{6.12}$$

Rearranging Equation 6.2 gives

$$\boldsymbol{\mu} = R^{-1}\boldsymbol{\nu} \tag{6.13}$$

for invertible R. Therefore the maximum likelihood estimator for the truth histogram is given by

$$\hat{\boldsymbol{\mu}}_{\mathrm{ML}} = R^{-1}\mathbf{n}. \tag{6.14}$$

$\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$ may be obtained by explicit matrix inversion for invertible R when $N = M$ or by alternative methods, such as numerically maximising Equation 6.11 or singular value decomposition [].

The covariance matrix for the maximum likelihood estimator is given by [356]

$$U_{\mathrm{ML}} = R^{-1} V \left(R^{-1} \right)^{\mathsf{T}}.$$ (6.15)

6.3.1 Regularisation

The detector response acts to smear out fine structure in the truth distribution, so statistical fluctuations in the data can lead to a large amount of fine structure in the unfolded result. This effect yields large local fluctuations in the maximum likelihood estimator when the typical bin width is not much larger than the detector resolution. The high-frequency nature of these fluctuations often lead to strong negative correlations between estimators for neighbouring bin counts.

These undesired false features are typically reduced by a technique known as *regularisation*. An explicit regularisation may be introduced beyond the initial discretisation of the distribution by minimising a *cost functional*,

$$\Phi(\boldsymbol{\mu}) = -\alpha \log P(\mathbf{n} \,|\, \boldsymbol{\mu}) + S(\boldsymbol{\mu}),$$ (6.16)

where $S(\boldsymbol{\mu})$ penalises high-variance distributions, effectively constricting the space of possible unfolded solutions. Multiple measures of smoothness may be used, such as those based on derivatives [378, 379] or entropy [380]. The ML solution has the minimum variance for an unbiased estimator, so any reduction in variance must be balanced by introducing some bias. The regularisation parameter α controls this bias–variance trade-off.

An unfolded distribution may alternatively be obtained by iterative techniques, which converge on the ML solution (*expectation maximisation*) [302–304]. Stopping after a fixed number of iterations can yield a solution with the desired properties, although the fact that the bias–variance trade-off is controlled by a discrete parameter, rather than a continuous one, limits the possibility to tune the parameter values. Modifications to the unfolding function in the iteration are available to overcome such limitations, for instance by introducing another

tunable continuous hyperparameter that reduces the convergence speed [381]. *Fully Bayesian unfolding* [382] addresses regularisation through a non-constant prior distribution, and performs the unfolding by sampling from the posterior distribution. The method presented in the next section mirrors the fully Bayesian technique when the prior and posterior distributions are Gaussian.

6.4 Gaussian process method

6.4.1 Maximum a posteriori estimator

From Bayes' theorem, the log posterior probability is given by

$$\log P(\boldsymbol{\mu}\,|\,\mathbf{n}) = \log P(\mathbf{n}\,|\,\boldsymbol{\mu}) + \log P(\boldsymbol{\mu}) - \log P(\mathbf{n}), \tag{6.17}$$

where $P(\boldsymbol{\mu})$ is the prior probability. The last term $P(\mathbf{n})$ (the *evidence*) may be ignored since it does not depend on $\boldsymbol{\mu}$.

The prior probability is taken to be given by a GP with mean vector $\mathbf{m}$ (the values are the bin contents of a *reference histogram*) and covariance matrix $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ where $\mathbf{x}$ is the vector containing the values of bin centers of the truth histogram. From Equation 5.8, the log prior probability is then given by

$$\log P(\boldsymbol{\mu}) = -\frac{1}{2}\,(\boldsymbol{\mu} - \mathbf{m})^{\mathsf{T}}\,K^{-1}\,(\boldsymbol{\mu} - \mathbf{m}) + \dots, \tag{6.18}$$

where the unwritten terms do not depend on $\boldsymbol{\mu}$. Substituting the likelihood from Equation 6.11 and prior from Equation 6.18 into Equation 6.17, the posterior is given by

$$\log P(\boldsymbol{\mu}\,|\,\mathbf{n}) = -\frac{1}{2}\,(\mathbf{n} - R\boldsymbol{\mu})^{\mathsf{T}}\,V^{-1}\,(\mathbf{n} - R\boldsymbol{\mu}) - \frac{1}{2}\,(\boldsymbol{\mu} - \mathbf{m})^{\mathsf{T}}\,K^{-1}\,(\boldsymbol{\mu} - \mathbf{m}) + \dots, \tag{6.19}$$

again dropping terms which do not contain $\boldsymbol{\mu}$.

The *maximum a posteriori* (MAP) estimator $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ is defined as the mode of the posterior probability distribution given by Equation 6.19. The derivative for the first and second terms on the right hand side are given by

$$\frac{\partial}{\partial \boldsymbol{\mu}} \left[-\frac{1}{2} \left(\mathbf{n} - R\boldsymbol{\mu} \right)^{\mathsf{T}} V^{-1} \left(\mathbf{n} - R\boldsymbol{\mu} \right) \right] = \left(\mathbf{n} - R\boldsymbol{\mu} \right)^{\mathsf{T}} V^{-1} R, \tag{6.20}$$

$$\frac{\partial}{\partial \boldsymbol{\mu}} \left[-\frac{1}{2} \left(\boldsymbol{\mu} - \mathbf{m} \right)^{\mathsf{T}} K^{-1} \left(\boldsymbol{\mu} - \mathbf{m} \right) \right] = - \left(\boldsymbol{\mu} - \mathbf{m} \right)^{\mathsf{T}} K^{-1}, \tag{6.21}$$

respectively. Combining these and taking the transpose (V^{-1} and K^{-1} are symmetric), it is therefore required that $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ satisfies

$$\mathbf{0} = R^{\mathsf{T}} V^{-1} \left(\mathbf{n} - R\hat{\boldsymbol{\mu}}_{\mathrm{MAP}} \right) - K^{-1} \left(\hat{\boldsymbol{\mu}}_{\mathrm{MAP}} - \mathbf{m} \right) \tag{6.22}$$

$$= R^{\mathsf{T}} V^{-1} \mathbf{n} - \left[R^{\mathsf{T}} V^{-1} R + K^{-1} \right] \hat{\boldsymbol{\mu}}_{\mathrm{MAP}} + K^{-1} \mathbf{m} \tag{6.23}$$

at the extremum. The covariance of the ML solution from Section 6.3 is given by $U_{\mathrm{ML}} = R^{-1} V (R^{-1})^{\mathsf{T}}$ by Equation 6.15, so $R^{\mathsf{T}} V^{-1} R = U_{\mathrm{ML}}^{-1}$. Substituting into Equation 6.23 and rearranging for $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$,

$$\hat{\boldsymbol{\mu}}_{\mathrm{MAP}} = \left[K^{-1} + U_{\mathrm{ML}}^{-1} \right]^{-1} \left(U_{\mathrm{ML}}^{-1} R^{-1} \mathbf{n} + K^{-1} \mathbf{m} \right) \tag{6.24}$$

$$= K \left[K + U_{\mathrm{ML}} \right]^{-1} R^{-1} \mathbf{n} + U_{\mathrm{ML}} \left[K + U_{\mathrm{ML}} \right]^{-1} \mathbf{m} \tag{6.25}$$

$$= K \left[K + U_{\mathrm{ML}} \right]^{-1} \left(R^{-1} \mathbf{n} - \mathbf{m} \right) + \mathbf{m}, \tag{6.26}$$

where from Equation 6.24 to Equation 6.25, $\left[A^{-1} + B^{-1} \right]^{-1} B^{-1} = A \left[A + B \right]^{-1}$ is used, valid for invertible matrices A and B.

The matrix acting on $\left(R^{-1} \mathbf{n} - \mathbf{m} \right)$ is expected to have eigenvalues that are bounded to be less than unity (from Section 5.4.4). The resulting summary statistic $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ is a *linear smoother* [329, 383] of the term $\left(R^{-1} \mathbf{n} - \mathbf{m} \right)$. The choice of the kernel function is discussed in Section Section 6.4.2 below.

By comparing the MAP estimator from Equation 6.26 to that obtained from GP regression in Equation 5.24, the important result that $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ is the posterior mean of a GP regression

whose observations are the ML solution is obtained. Since the posterior distribution is a product of Gaussians, it is also a Gaussian and therefore the mode is identical to the mean. The connection to GP regression gives that the covariance of the MAP estimator is given by

$$U = K - K \left[K + R^{-1} V (R^{-1})^{\mathsf{T}} \right]^{-1} K. \tag{6.27}$$

Furthermore, if the observation (training) indices $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots)$ are different from the prediction (testing) indices $X_* = (\mathbf{x}_{*1}, \mathbf{x}_{*2}, \ldots)$, and the reference histogram can be obtained for bins defined by X_*, then the standard results from GP regression can be used to generalise the MAP solution to

$$\hat{\boldsymbol{\mu}}_{\mathrm{MAP}} = K_*^{\mathsf{T}} \left[K + U_{\mathrm{ML}} \right]^{-1} \left(\hat{\boldsymbol{\mu}}_{\mathrm{ML}} - \mathbf{m} \right) + \mathbf{m}_*, \tag{6.28}$$

$$U = K_{**} - K_*^{\mathsf{T}} \left[K + U_{\mathrm{ML}} \right]^{-1} K_*, \tag{6.29}$$

where $[K_*]_{ij} = k(\mathbf{x}_i, \mathbf{x}_{*j})$, $[K_{**}]_{ij} = k(\mathbf{x}_{*i}, \mathbf{x}_{*j})$, and $\mathbf{m}_*$ is the mean histogram at X_* bin positions.

This is the method of unfolding with GPs. Its novelty derives from including the detector convolution in a GP regression. The generalised results in Equation 6.28 and Equation 6.29 are simple, linear algebraic expressions once the ML solution is known. Therefore the unfolded estimator and its covariance are efficient to compute and this is an advantage over other, more computationally intensive unfolding methods. Additionally, the covariance matrix for the unfolded histogram is easily calculated as part of the result.

6.4.2 Kernel choice and optimisation

In the proposed unfolding method using GPs, the explicit regularisation is introduced via the kernel function $k(\mathbf{x}, \mathbf{x}')$ which constricts the space of possible solutions to those with a particular covariance. A common choice for the kernel function is the squared-exponential,

which in one dimension has the form

$$k(\mathbf{x}, \mathbf{x}') = \tau^2 \exp\left[\frac{-(x - x')^2}{2l^2}\right].$$

(6.30)

This kernel function is stationary in the sense that it is a function of only the distance between the inputs, $|x - x'|$. It is parameterised by the amplitude τ^2 and length scale l, which form the set of hyperparameters, $\boldsymbol{\theta} = \{\tau^2, l\}$. In this treatment, the method of maximum marginal likelihood, as described in Section 5.7, can be used.

The kernel function in Equation 6.30 is smooth and has other desirable properties that lead it to be very widely used in the literature [319]. Also the hyperparameters τ^2 and l are readily interpretable. Other kernel functions, however, may be more suitable for describing the truth distribution. An attractive feature of the approach presented here is that one may encode knowledge of the underlying physical process to derive a physically-motivated kernel [353] which may better describe the truth distribution.

The mathematics of *reproducing kernel Hilbert spaces* formalises the link between the kernel and the traditional explicit regularisation approach used in some particle physics results. For example, a *thin plate covariance* [384] leads to a solution equivalent to that of spline regularisation, known as Tikhonov regularisation in particle physics [356, 378, 379, 385]. In one dimension, this stationary kernel may be written $k(r) = \tau^2(2r^3 - 3Dr^2 + D^3)$, where $r = |x - x'| \leq D$ and D is determined by boundary conditions. This kernel contains a single parameter τ^2, which controls the global strength of the regularisation, as is the case with Tikhonov regularisation in its usual implementation. In contrast, an advantage of the GP approach presented in this paper is that the explicit regularisation may be varied locally along the spectrum by using a non-stationary kernel function. An example of this, using the Gibbs kernel [352], is provided in Section 6.5.2.

6.5 Example applications

Examples are given here of a bimodal distribution in Section 6.5.1, and a falling spectrum in Section 6.5.2. They were generated using a Python framework developed for this project [386].

6.5.1 Bimodal distribution

A set of 200 000 simulated 'truth' events is obtained by sampling from two Gaussian distributions for x with mean values 0.3 and 0.7, both with standard deviation 0.1. These truth events are histogrammed in $\boldsymbol{\mu}$. They are then smeared with a zero-mean Gaussian resolution of $\sigma = 0.075$ to generate the histogram $\boldsymbol{\nu}$. Events are accepted in the region $0 < x < 1$ and both the $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ histograms use 20 constant bins of constant width. The truth and smeared events are used to determine the square response matrix R from a 2D histogram, normalised such that $\sum_{i=1}^{N} R_{ij} = 1$, as detailed in Section 6.2. A large number of events are used to reduce the impact of statistical fluctuations in R. Finally, the observed histogram $\mathbf{n}$ is generated by applying the same smearing process to an independent sample of 20 000 events drawn from the truth distribution. The histograms $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are scaled so that they contain the same number of events as $\mathbf{n}$. The three histograms are shown in Figure 6.1.

A GP with the squared-exponential kernel function given by Equation 6.30 is used as the prior. The reference histogram, $\mathbf{m}$, is taken to be zero for all bins. After setting the GP kernel and mean, the values for the two hyperparameters τ^2 and l are chosen to be those that maximise the log marginal likelihood, given by Equation 5.47. The maximum point and contours of the log marginal likelihood are shown in Figure 6.3.

The estimator for the unfolded histogram, $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ given by Equation 6.26, is shown in Figure 6.2. The covariance matrix U is defined by Equation 6.27, and the correlation matrix with elements $\rho_{ij} = U_{ij}/\sqrt{U_{ii}U_{jj}}$ is shown in Figure 6.4. The mean correlations for the first, second, and third neighbouring bins are 0.360, -0.453, and -0.566, respectively.

The *pull* is defined to be given by the normalised residual, $p_i = (\mathbb{E}[\hat{\mu}_i] - \mu_i)/\hat{\sigma}_i$, where the elements of $\mathbb{E}[\hat{\boldsymbol{\mu}}]$ are obtained by unfolding the histogram $\boldsymbol{\nu}$ and $\hat{\sigma}_i^2 = U_{ii}(\boldsymbol{\nu})$. This quantity
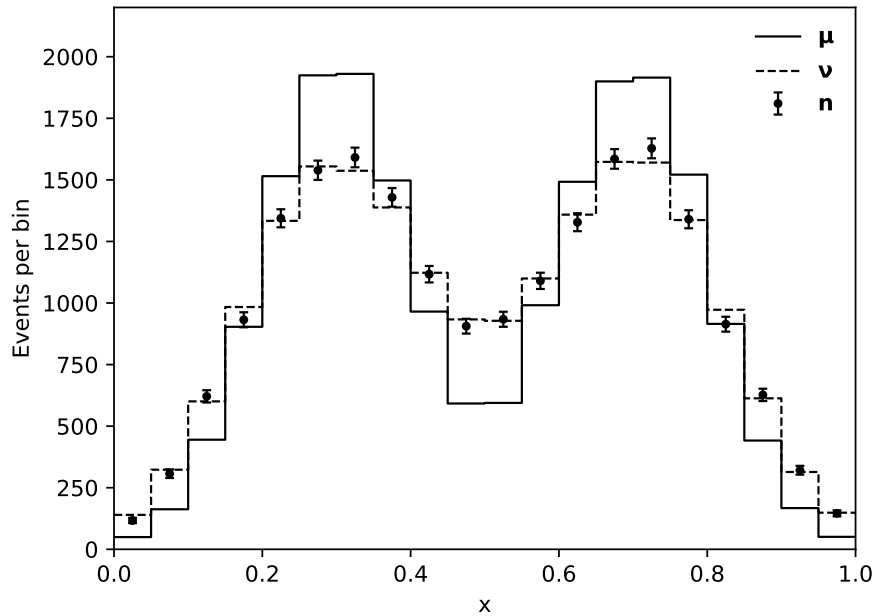
Figure 6.1 Truth ($\boldsymbol{\mu}$), expected reconstruction ($\boldsymbol{\nu}$), and observed ($\mathbf{n}$) histograms for the bimodal example. The histogram definitions are reported in the text. The error bars on $\mathbf{n}$ represent their Poisson uncertainties.
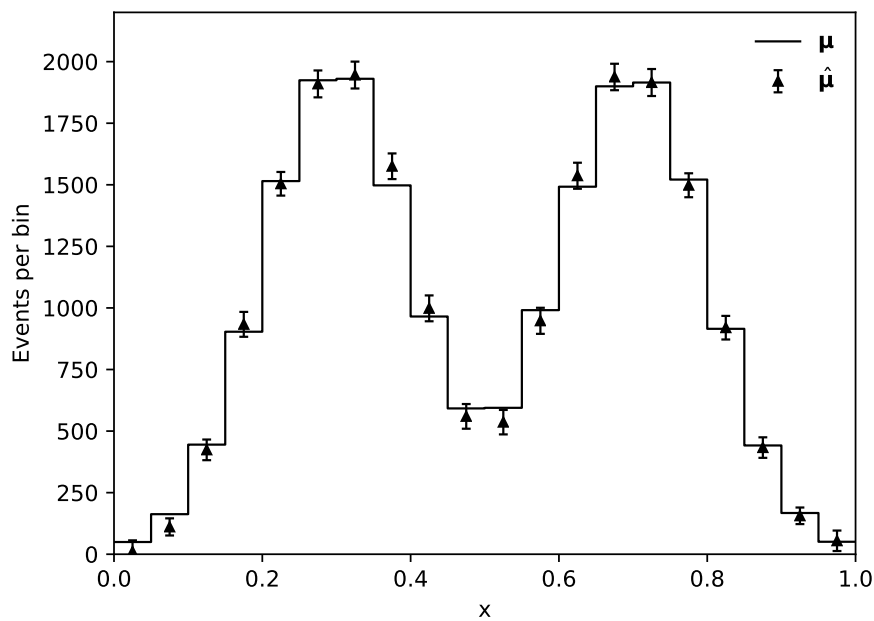
Figure 6.2 Truth ($\boldsymbol{\mu}$) and unfolded ($\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ histograms for the two-peak example. The error bars on $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ represent the standard deviations obtained from the covariance matrix as defined by Equation 6.27.
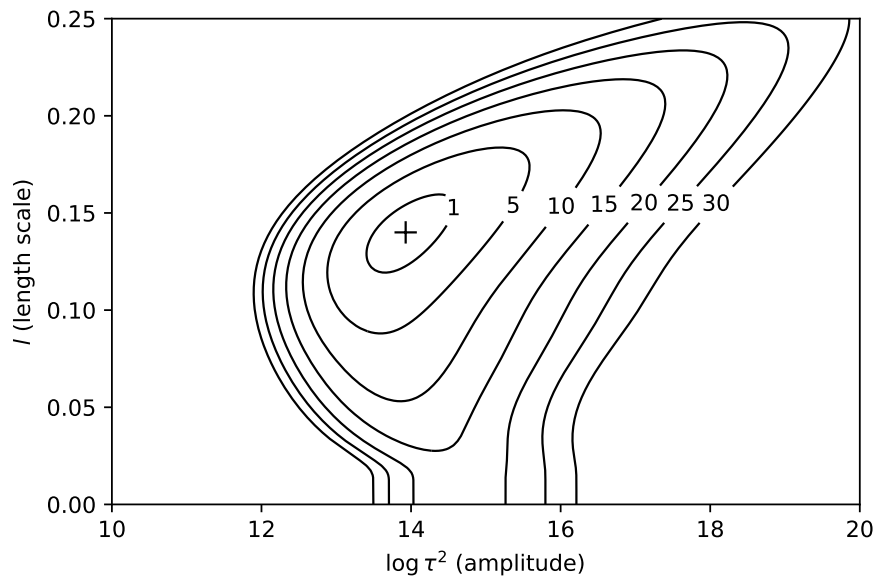
Figure 6.3 Contours of the log marginal likelihood, given by Equation 5.47, for the two-peak example as a function of the parameters for the squared-exponential kernel, τ^2 and l. The cross indicates the point of maximum marginal likelihood. The contour labels are the depth of the contour below the maximum.
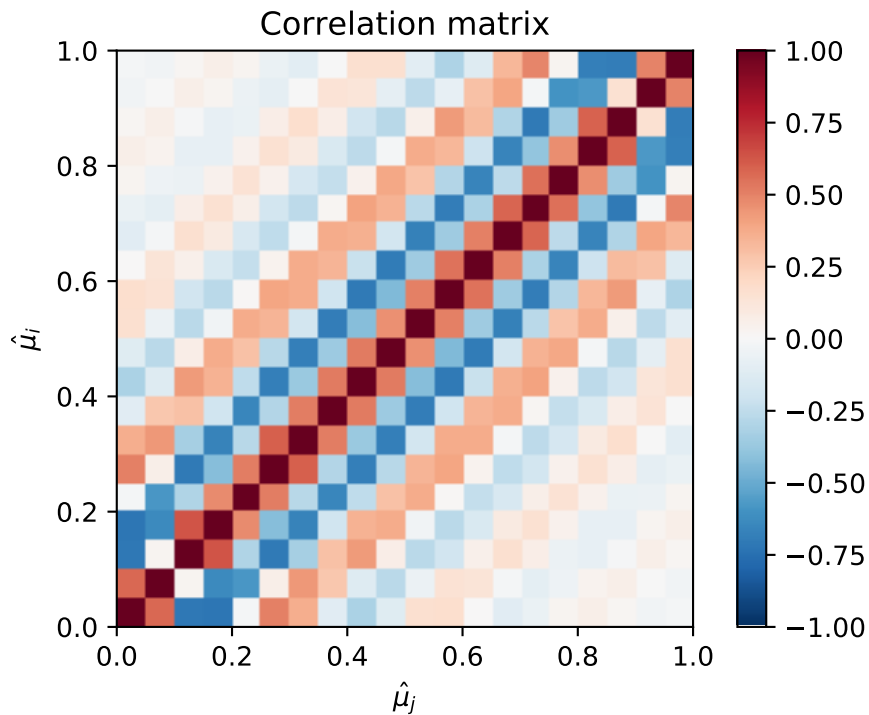
Figure 6.4 Correlation matrix for the unfolded truth estimators $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ for the bimodal example.
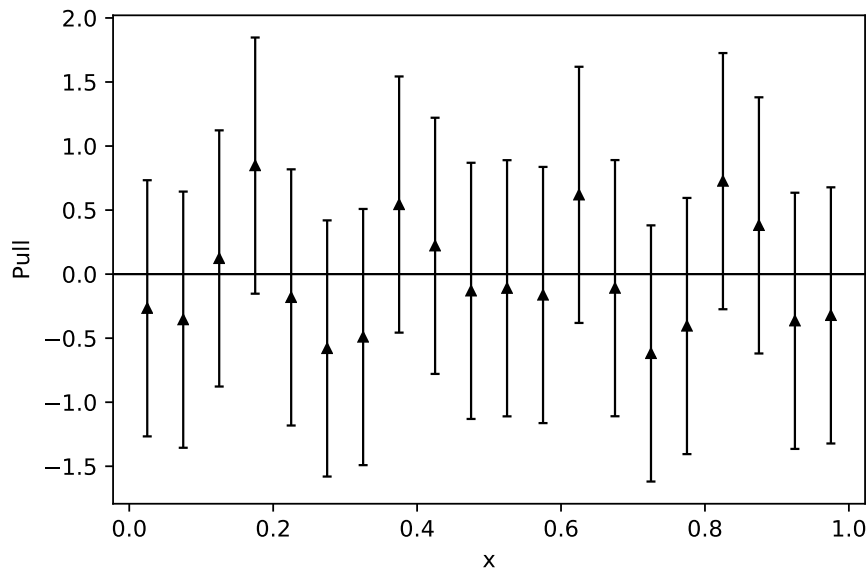
Figure 6.5 Pulls of the unfolded estimators $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ for the bimodal example.

provides an indication of the bias of the unfolded estimator, and is shown for the present example in Figure 6.5.

6.5.2 Falling spectrum

For the second example, $200\,000$ 'truth' events are sampled from an exponential distribution $f(x) = e^{-x}$ in the region $1 < x < 5$ and accumulated in 20 bins of equal width to form the histogram $\boldsymbol{\mu}$. These events are smeared according to a Gaussian distribution with variance corresponding to a resolution of $0.2\sqrt{x}$. The smeared events are placed in the histogram $\boldsymbol{\nu}$ with 30 bins of equal width in the region $0.5 < x < 5$. The observed histogram $\mathbf{n}$ is generated by applying the same smearing to 1000 independent events generated in the truth region from the same exponential distribution. The histograms $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are scaled to contain 1000 events. These three histograms are shown in Figure 6.6.

In this example, $N > M$ so while the problem is well-constrained, the $N \times M$ response matrix R is not directly invertible. To mitigate this, a form of the posterior distribution that
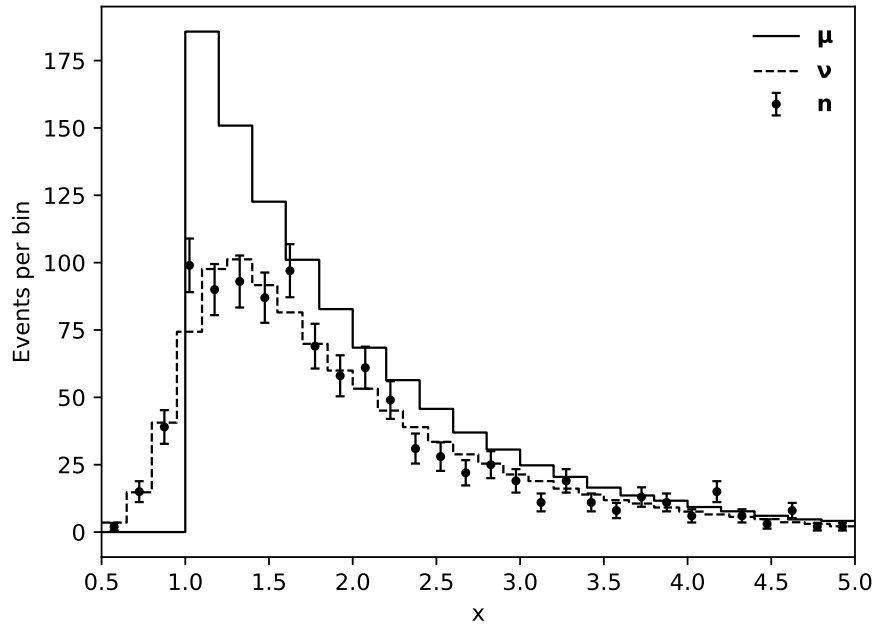
Figure 6.6 Truth ($\boldsymbol{\mu}$), expected reconstruction ($\boldsymbol{\nu}$), and observed ($\mathbf{n}$) histograms for the falling spectrum example. The error bars on $\mathbf{n}$ represent their Poisson uncertainties.
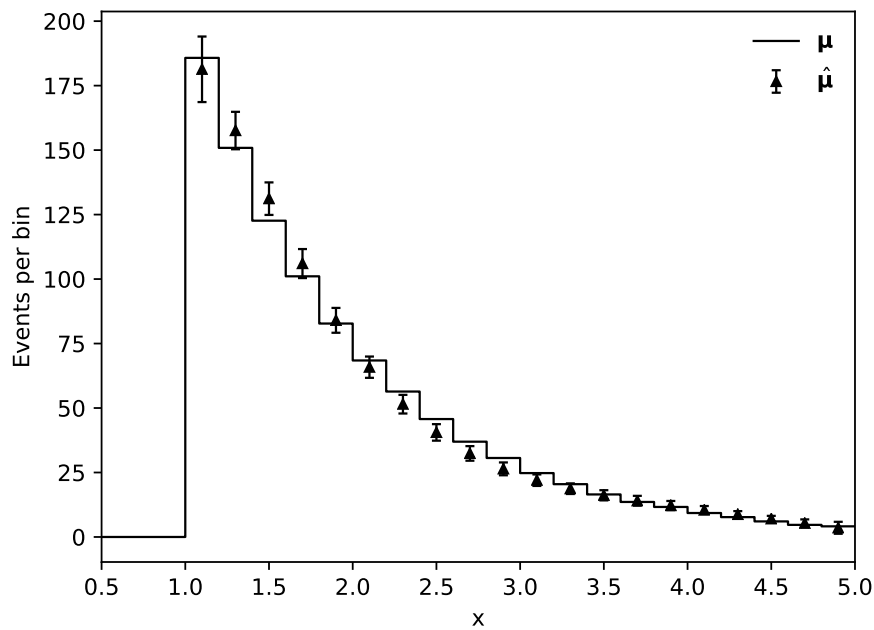
Figure 6.7 Truth ($\boldsymbol{\mu}$) and unfolded histograms ($\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$) for the falling spectrum example.
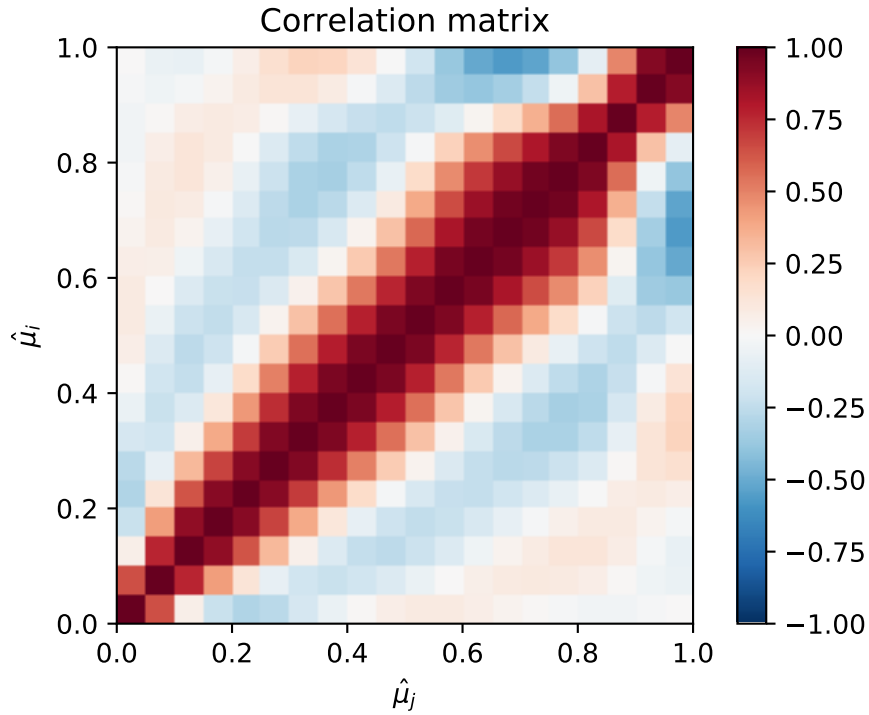
Figure 6.8 Correlation matrix for the unfolded truth estimators $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ for the falling spectrum example.
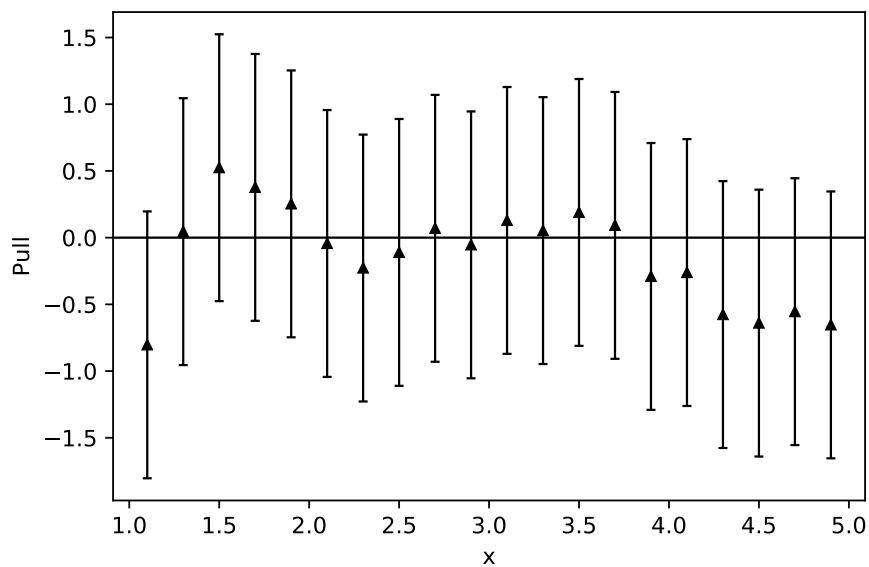
Figure 6.9 Pulls of the unfolded histogram $\hat{\boldsymbol{\mu}}$ for the falling spectrum example.

does not rely on R^{-1} [387], obtained by rearranging Equations 6.26 and 6.27, can be used:

$$\hat{\boldsymbol{\mu}}_{\text{MAP}} = KR^{\mathsf{T}} \left[RKR^{\mathsf{T}} + V \right]^{-1} (\mathbf{n} - R\mathbf{m}) + \mathbf{m} \tag{6.31}$$

$$U = K - KR^{\mathsf{T}} \left[RKR^{\mathsf{T}} + V \right]^{-1} RK. \tag{6.32}$$

In an example such as this (e.g., an invariant mass spectrum of a background process), it might be expected that there is fine structure present in the bulk of the truth distribution, but not the tail. A kernel with constant length scale, such as the squared-exponential Equation 6.30, is unsuitable in this case. A more suitable choice is given by a kernel function with variable length scale, such as the Gibbs kernel [319, 352],

$$k(x, x') = \tau^2 \sqrt{\frac{2l(x)l(x')}{l^2(x) + l^2(x')}} \, \exp\left(-\frac{(x - x')^2}{l^2(x) + l^2(x')} \right), \tag{6.33}$$

in 1D, where $l(x)$ is an arbitrary positive function of x, here chosen to be $l(x) = bx + c$. This allows for a linearly-changing length scale. The reference histogram, $\mathbf{m}$ is taken to be zero for all bins as in the previous example. The increased flexibility afforded by this kernel function is realised by introducing more regularisation parameters, $\boldsymbol{\theta} = \{\tau^2, b, c\}$. For a large number of such parameters, it can become increasingly difficult to choose the optimal point.

The resultant unfolded histogram is compared with the truth histogram in Figure 6.7. Once the GP kernel and the reference histogram are set, the parameters $\boldsymbol{\theta}$ for this example are chosen with the maximum log marginal likelihood prescription given in Section 6.4.2, corresponding to $\log \tau^2 = 9.62$, $b = 2.89$, $c = -0.858$. As expected, $b > 0$ so the length scale increases with x. The correlation matrix for the unfolded truth estimators is shown in Figure 6.8, and the pulls in Figure 6.9. The binning with $N > M$ introduces a source of implicit regularisation, and therefore bias in the estimator for the unfolded histogram.

6.6 Discussion

This chapter presents a method of unfolding using GP regression. Conditioning a GP prior on the maximum likelihood solution to the inverse problem is equivalent to constructing the MAP estimator. In this way, the use of GP regression provides regularisation to the maximum likelihood solution.

The GP is entirely described by mean and kernel functions. While the mean function has little impact on the result, the kernel function prescribes the covariance of the estimator for the unfolded solution. By choosing an appropriate kernel function, the smoothness in the unfolded estimator can be controlled both globally and locally. This approach allows the regularisation to be finely controlled and can be naturally motivated by knowledge of the underlying physics.

Traditional unfolding methods operate with fixed definitions for the binning of the histograms involved. No such restrictions exist for the method of unfolding with GPs. The posterior distribution is itself a GP, and can therefore be evaluated at arbitrary indices, X_*, by Equations 6.28 and 6.29. This means that the unfolded distribution can be re-binned at will, so long as the kernel function is known, and therefore the direct comparison of results from different detector apparatus is more easily achieved.

GPs have been introduced to a number of scientific fields to improve their statistical procedures [375, 376]. They have not, however, traditionally been used in particle physics, although recent developments in this area have shown promise [353]. This chapter demonstrates that the presented novel unfolding method is generally applicable to problems of different shapes and sizes, that the regularisation can be controlled naturally, and that the result – including the unfolded covariance matrix – can be obtained conveniently.

Non-zero background

In Section 6.2, contributions to the measured bin populations from background processes are taken to be equal to zero for simplicity. Background contributions, with expectation values given by the N-dimensional vector β, can be simply included by modifying the folding equation,

given by Equation 6.2, to

$$\boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta}. \tag{6.34}$$

Then for the estimators used in the method presented in this chapter, the data histogram is substituted for the background-subtracted data, $\mathbf{n} \to \mathbf{n} - \boldsymbol{\beta}$, throughout.

Gaussian approximation

In this chapter, it is assumed that the data may be approximated as distributed according to a Gaussian. However, this is not universally the case in particle physics. The choice of unfolding method depends on the analysis being done and should be tested against simulation. This is also the case for the traditional unfolding methods, when deciding on an acceptable regularisation scheme to control the bias–variance trade-off. In any analysis making use of the techniques presented in this chapter, it is recommended that the unfolding is tested to ensure it acceptably meets the requirements of the analysis under consideration. This is to be done with simulated *pseudo-data*, before the real data are unblinded, to avoid biasing the final result.

Future work

The statistical properties of the MAP estimator for the unfolded histogram are probed by the distributions of the pulls in Figures 6.5 and 6.9. Further studies are also possible. In particular, the frequentist properties of $\hat{\boldsymbol{\mu}}_{\mathrm{MAP}}$ could be investigated, for example its coverage. This would assess the suitability of the method for use in hybrid frequentist–Bayesian analyses often seen in particle physics [388].

The treatment of systematic uncertainties is envisaged for future work on this topic. Approximate variational approaches, as used in published particle physics analyses [389, 390], may still be employed for the unfolding method presented here. Also as an extension to this work, further research into the applications of Student-t processes [391] for unfolding in particle physics could be carried out.

Chapter 7

Conclusion

The Standard Model has a remarkable history of successful prediction and agreement with observations from measurements in the field of high energy particle physics. However, it is also known to be incomplete and extensions or alternatives to it must therefore exist.

One of the areas this could occur in is the physics of the t quark, one of the most exotic and exceptional particles in the Standard Model. As the most massive particle, it could provide a window into effects beyond the Standard Model. Additionally, measurements of the properties and production cross sections of t quarks are essential for characterising background processes relevant in explicit searches for exotic phenomena.

This thesis presents measurements of $t\bar{t}$ production cross sections at the Large Hadron Collider, using the state-of-the-art ATLAS detector. The analysis uses collisions which result in decays to a fully hadronic final state, where the jets of hadronic particles can all be resolved in the detector. This particular decay channel corresponds to t quarks with relatively low transverse momentum, and the $t\bar{t}$ system can be fully reconstructed with good resolution. The cross sections are reported differentially as functions of kinematic variables of the t quarks and of the combined $t\bar{t}$ system. They are unfolded to definitions of the t quark at particle level and parton level. Good agreement with predictions from the Standard Model, through Monte Carlo simulations, is seen across the results. Some alternative simulations are found to deviate from the data.

The use of Gaussian processes in high energy particle physics is discussed in Chapter 5. As simple and interpretable, yet flexible, non-parametric models of families of functions, GPs can be used in many areas in the field. One such applicable use case is in unfolding. A novel method of unfolding using GP regression is presented in Chapter 6. Here it is seen how the regularisation can be prescribed in a flexible manner, informed by the underlying physics. A connection is made to Bayesian statistics and the *maximum a posteriori* estimator. The performance of the method is illustrated and evaluated in two toy examples.

Acknowledgements

My foremost thanks go to Glen Cowan for his wisdom and guidance. I am grateful for his constant support and encouragement, and his infectious enthusiasm for finding things out.

My thanks are extended to the ATLAS group at Royal Holloway for their support: Tracey Berry, Veronique Boisvert, Michele Faucci Giannelli, Will Panduro Vazquez, Francesco Spanò, and Pedro Teixeira-Dias.

I am grateful to my fellow students and friends, both in London and Geneva: Deshan Abhayasinghe, Ian Connelly, Mike Fenton, Rob Harding, David Lack, Ben Sowden, Joe Taylor, and Lewis Wilkins.

This thesis would not have been possible without Jimmy Anderson, John Barnes, Carl Beard, Andrew Boothroyd, Ann Crist, Francesco Fermani, Claire Gwenlan, Alex Hickey, Barry King, Philip Saville, and Amit Visana.

Finally, I would like to thank Cristobel Soares-Smith for setting the next stone.

Addendum, April 2020

I am indebted to my examiners, Roger Barlow and Eram Rizvi.

The final revision of this thesis could not have been submitted without the kind support of Joseph Enguehard, Lauren Halliwell, Nils Hammerla, Ben Haselden, Saurabh Johri, Minsang Kim, Matt McGee, Jonathan Murray, Claudia Schulz, and Dane Sherburn, in addition to those acknowledged above.

References

[1] ATLAS Collaboration, *The Collaboration.*
    https://atlas.cern/discover/collaboration (accessed on 20/08/2019).

[2] L. D. Landau and E. M. Lifshitz, *Mechanics*, vol. 1 of *Course of Theoretical Physics.*
    Butterworth–Heinemann, third ed., 1976.

[3] E. W. Weisstein, *Euler–Lagrange Differential Equation.* From *MathWorld* – A Wolfram
    Web Resource.
    http://mathworld.wolfram.com/Euler-LagrangeDifferentialEquation.html
    (accessed on 24/03/2019).

[4] R. P. Feynman, *Space-Time Approach to Non-Relativistic Quantum Mechanics.* Reviews
    of Modern Physics **20**(2) (1948) pp. 367–387.

[5] R. P. Feynman, A. R. Hibbs, and D. F. Styer, *Quantum Mechanics and Path Integrals.*
    Dover Publications, emended ed., 2010.

[6] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory.* Westview
    Press, 1995.

[7] P. A. M. Dirac, *The Lagrangian in Quantum Mechanics.* Physikalische Zeitschrift der
    Sowjetunion **3** (1933) pp. 64–72.
    http://www.hep.anl.gov/czachos/soysoy/Dirac33.pdf.

[8] M. A. Srednicki, *Quantum Field Theory.* Cambridge University Press, 2007.

[9] E. Noether, *Invariante Variationsprobleme.* Nachrichten von der Gesellschaft der
    Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1918) pp. 235–257.
    https://eudml.org/doc/59024.

[10] H. Lehmann, K. Symanzik, and W. Zimmermann, *Zur Formulierung quantisierter
     Feldtheorien.* Il Nuovo Cimento **1**(1) (1955) pp. 205–225.

[11] G. C. Wick, *The Evaluation of the Collision Matrix.* Physical Review **80**(2) (1950)
     pp. 268–272.

[12] F. H. L. Essler, *Lecture Notes for the C6 Theory Option (Oxford MPhys Course)*, 2012. http://www-thphys.physics.ox.ac.uk/courses/C6/C6web2012/theory/lecturenotes2012.pdf (accessed on 27/03/2019).

[13] E. W. Weisstein, *Fourier Transform–Delta Function.* From *MathWorld* – A Wolfram Web Resource. http://mathworld.wolfram.com/FourierTransformDeltaFunction.html (accessed on 30/03/2019).

[14] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics.* Physical Review **76**(6) (1949) pp. 769–789.

[15] J. P. Ellis, *TikZ-Feynman: Feynman diagrams with TikZ.* Computer Physics Communications **210** (2017) pp. 103–123.

[16] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-Order QED Contribution to the Electron g − 2 and an Improved Value of the Fine Structure Constant.* Physical Review Letters **109**(111807) (2012) .

[17] P. J. Mohr, D. B. Newell, and B. N. Taylor, *CODATA Recommended Values of the Fundamental Physical Constants: 2014.* Reviews of Modern Physics **88**(035009) (2016) .

[18] nLab authors, *Poincaré group*, 2019. http://ncatlab.org/nlab/revision/Poincaré+group/27 (accessed on 2019-04-03).

[19] L. R. Moore, *Top Quark Physics in the Standard Model Effective Field Theory.* PhD thesis, University of Glasgow, 2016.

[20] C. S. Wu *et al.*, *Experimental Test of Parity Conservation in Beta Decay.* Physical Review **105**(4) (1957) pp. 1413–1415.

[21] M. Beyer, *CP Violation in Particle, Nuclear, and Astrophysics.* Springer, 2002.

[22] M. Tanabashi *et al.* (Particle Data Group), *Review of Particle Physics.* Physical Review D **98**(030001) (2018) .

[23] J. Schwinger, *The Theory of Quantized Fields.* Physical Review **82**(6) (1951) pp. 914–927.

[24] G. Lüders, *On the Equivalence of Invariance under Time Reversal and under Particle–Antiparticle Conjugation for Relativistic Field Theories.* Det Kongelige Danske Videnskabernes Selskab Matematisk-Fysiske Meddelelser **28**(5) (1954) pp. 1–17. http://cds.cern.ch/record/1071765 (accessed on 03/04/2019).

[25] W. Pauli, *Niels Bohr and the Development of Physics; essays dedicated to Niels Bohr on the occasion of his seventieth birthday.* McGraw–Hill, 1955.

[26] M. Fierz and G. Pasa, *On the relativistic theory of force-free particles with any spin.* `arXiv:1704.00662`.

[27] W. Pauli, *The Connection Between Spin and Statistics.* Physical Review **58**(8) (1940) pp. 716–722.

[28] R. P. Feynman, *Quantum Electrodynamics.* Addison–Wesley, 1998.

[29] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren.* Zeitschrift für Physik **31**(1) (1925) pp. 765–783.

[30] M. Gell-Mann, *Symmetries of Baryons and Mesons.* Physical Review **125**(3) (1962) pp. 1067–1084.

[31] W. Pauli, *Zur Quantenmechanik des magnetischen Elektrons.* Zeitschrift für Physik **43**(9-10) (1927) pp. 601–623.

[32] J. C. Romão and J. P. Silva, *A Resource for Signs and Feynman Diagrams of The Standard Model.* International Journal of Modern Physics A **27**(1230025) (2012) .

[33] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass.* Physical Review **130**(1) (1963) pp. 439–442.

[34] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons.* Physical Review Letters **13**(9) (1964) pp. 321–323.

[35] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles.* Physical Review Letters **13**(20) (1964) pp. 585–587.

[36] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories.* Physical Review **155**(5) (1967) pp. 1554–1561.

[37] P. W. Higgs, *Broken symmetries, massless particles and gauge fields.* Physics Letters **12**(2) (1964) pp. 132–133.

[38] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons.* Physical Review Letters **13**(16) (1964) pp. 508–509.

[39] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons.* Physical Review **145**(4) (1966) pp. 1156–1163.

[40] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields.* Nuclear Physics B **44**(1) (1972) pp. 189–213.

[41] S. L. Glashow, *The renormalizability of vector meson interactions.* Nuclear Physics **10** (1959) pp. 107–117.

[42] A. Salam and J. C. Ward, *Weak and electromagnetic interactions.* Il Nuovo Cimento **11**(4) (1959) pp. 568–577.

[43] S. Weinberg, *A Model of Leptons.* Physical Review Letters **19**(21) (1967) pp. 1264–1266.

[44] J. D. Jackson, *From Lorenz to Coulomb and other explicit gauge transformations.* American Journal of Physics **70**(9) (2002) pp. 917–928.

[45] S. Dawson, *Introduction to Electroweak Symmetry Breaking.* `arXiv:9901280`.

[46] P. A. M. Dirac, *The Principles of Quantum Mechanics.* Clarendon Press, fourth ed., 1981.

[47] N. Cabibbo, *Unitary Symmetry and Leptonic Decays.* Physical Review Letters **10**(12) (1963) pp. 531–533.

[48] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction.* Progress of Theoretical Physics **49**(2) (1973) pp. 652–657.

[49] L. Chau and Y. Keung, *Comments on the Parametrization of the Kobayashi–Maskawa Matrix.* Physical Review Letters **53** (1984) pp. 1802–1805.

[50] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix.* Physical Review Letters **51**(21) (1983) pp. 1945–1947.

[51] B. Andersson *et al.*, *Parton fragmentation and string dynamics.* Physics Reports **97**(2-3) (1983) pp. 31–145.

[52] T. Sjöstrand, *Old Ideas in Hadronization: The Lund String.* Speaking at *New Ideas in Hadronization*, Durham (2009). `http://home.thep.lu.se/~torbjorn/talks/durham09.pdf` (accessed on 09/06/2019).

[53] C. Quigg, *Discovery of the Top Quark.* `http://lutece.fnal.gov/Papers/PhysNews95.html` (accessed on 2019-06-09).

[54] CDF Collaboration, *Observation of Top Quark Production in pp Collisions with the Collider Detector at Fermilab.* Physical Review Letters **74**(14) (1995) pp. 2626–2631.

[55] D0 Collaboration, *Observation of the Top Quark.* Physical Review Letters **74**(14) (1995) pp. 2632–2637.

[56] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.* Physics Letters B **716**(1) (2012) pp. 1–29.

[57] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.* Physics Letters B **716**(1) (2012) pp. 30–61.

[58] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton–proton collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS experiment.* ATLAS-CONF-2018-031, 2018. https://cds.cern.ch/record/2629412.

[59] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector.* Physical Review D **97**(072016) (2018) .

[60] G. G. Ross, *Grand Unified Theories (Frontiers in Physics).* CRC Press, 1985.

[61] H. Georgi and S. L. Glashow, *Unity of All Elementary-Particle Forces.* Physical Review Letters **32**(8) (1974) pp. 438–441.

[62] H. Georgi, *The State of the Art—Gauge Theories* in AIP Conference Proceedings, vol. 23, 1975, pp. 575–582. American Institute of Physics.

[63] SNO Collaboration, *Constraints on Nucleon Decay via Invisible Modes from the Sudbury Neutrino Observatory.* Physical Review Letters **92**(102004) (2004) .

[64] A. Einstein, *Die Feldgleichungen der Gravitation.* Sitzung der physikalische-mathematischen Klasse **25** (1915) pp. 844–847. https://einsteinpapers.press.princeton.edu/vol6-doc/272. English translation at https://einsteinpapers.press.princeton.edu/vol6-trans/129 (accessed on 12/06/2019).

[65] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation.* Princeton University Press, 2017 ed., 2017.

[66] The Event Horizon Telescope Collaboration, *First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole.* The Astrophysical Journal Letters **875**(1) (2019) .

[67] LIGO Scientific Collaboration and Virgo Collaboration, *Observation of Gravitational Waves from a Binary Black Hole Merger.* Physical Review Letters **116**(061102) (2016) .

[68] D. Castelvecchi, *The black-hole collision that reshaped physics.* Nature **531**(7595) (2016) pp. 428–431.

[69] E. W. Weisstein, *Fundamental Forces.* http://scienceworld.wolfram.com/physics/FundamentalForces.html (accessed on 2019-06-12).

[70] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984.

[71] Particle Data Group, *Experimental Tests of Gravitational Theory*. http://pdg.lbl.gov/2019/reviews/rpp2018-rev-gravity-tests.pdf (accessed on 12/06/2019).

[72] T. Kaluza, *On the Unification Problem in Physics*. International Journal of Modern Physics D **27**(1870001) (2018, original 1921) .

[73] O. Klein, *The Atomicity of Electricity as a Quantum Theory Law*. Nature **118** (1926) p. 516.

[74] ATLAS Collaboration, *Combination of searches for heavy resonances decaying into bosonic and leptonic final states using 36 fb^{-1} of proton–proton collision data at $\sqrt{s} =$ 13 TeV with the ATLAS detector*. Physical Review D **98**(052008) (Sep, 2018) .

[75] C. Cao, S. M. Carroll, and S. Michalakis, *Space from Hilbert Space: Recovering Geometry from Bulk Entanglement*. arXiv:1606.08444.

[76] S. M. Carroll, *Space Emerging from Quantum Mechanics*. http://www.preposterousuniverse.com/blog/2016/07/18/space-emerging-from-quantum-mechanics/ (accessed on 12/06/2019).

[77] S. M. Carroll, *The Many Worlds of Quantum Mechanics*. CERN Colloquium, 21/03/2019. https://cds.cern.ch/record/2668839.

[78] J. Collins, A. Perez, and D. Sudarsky, *Lorentz Invariance Violation and its Role in Quantum Gravity Phenomenology*. arXiv:hep-th/0603002.

[79] D. Mattingly, *Modern Tests of Lorentz Invariance*. Living Reviews in Relativity **8** (2005) p. 5.

[80] S. Liberati, *Tests of Lorentz invariance: a 2013 update*. arXiv:1304.5795.

[81] A. Bourgoin *et al.*, *Lorentz Symmetry Violations from Matter–Gravity Couplings with Lunar Laser Ranging*. Physical Review Letters **119**(201102) (2017) .

[82] N. A. Flowers, C. Goodge, and J. D. Tasson, *Superconducting-Gravimeter Tests of Local Lorentz Invariance*. Physical Review Letters **119**(201101) (2017) .

[83] S. Hughes, *Introduction; Couloumb's Law; Superposition; Electric Energy*, 2005. Lecture notes for MIT course 8.022. http://web.mit.edu/sahughes/www/8.022/lec01.pdf (accessed on 30/06/2019).

[84] C. Csáki and P. Tanedo, *Beyond the Standard Model; Lectures at the 2013 European School of High Energy Physics.* `arXiv:1602.04228v2`.

[85] S. P. Martin, *A Supersymmetry Primer.* `arXiv:hep-ph/9709356v7`.

[86] Particle Data Group, *Supersymmetry, Part I (Theory).* `http://pdg.lbl.gov/2019/reviews/rpp2018-rev-susy-1-theory.pdf` (accessed on 02/07/2019).

[87] Particle Data Group, *Supersymmetry, Part II (experiment).* `http://pdg.lbl.gov/2018/reviews/rpp2018-rev-susy-2-experiment.pdf` (accessed on 02/07/2019).

[88] H. J. G. L. M. Lamers and E. M. Levesque, *Understanding Stellar Evolution.* IOP Publishing, 2017.

[89] E. G. Adelberger *et al.*, *Solar fusion cross sections. II. The pp chain and CNO cycles.* Reviews of Modern Physics **83**(1) (2011) pp. 195–245.

[90] N. Vinyoles *et al.*, *A New Generation of Standard Solar Models.* The Astrophysical Journal **835** (2017) p. 202.

[91] SNO Collaboration, *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory.* Physical Review Letters **87**(071301) (2001) .

[92] B. Pontecorvo, *Inverse Beta Processes and Nonconservation of Lepton Charge.* Journal of Experimental and Theoretical Physics **34** (1958) pp. 247–249. `http://www.jetp.ac.ru/files/pontecorvo1958_en.pdf` (accessed on 07/07/2019).

[93] B. Pontecorvo, *Neutrino Experiments and the Problem of Conservation of Leptonic Charge.* Journal of Experimental and Theoretical Physics **26**(5) (1968) pp. 984–988. `http://www.jetp.ac.ru/cgi-bin/dn/e_026_05_0984.pdf` (accessed on 07/07/2019).

[94] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory.* Physical Review Letters **89**(011301) (2002) .

[95] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles.* Progress of Theoretical Physics **28**(5) (1962) pp. 870–880.

[96] J. N. Bahcall and C. Peña-Garay, *Solar models and solar neutrino oscillations.* New Journal of Physics **6** (2004) p. 63.

[97] P. A. M. Dirac, *The Quantum Theory of the Electron.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **117**(778) (1928) pp. 610–624.

[98] E. Majorana, *Teoria simmetrica dell'elettrone e del positrone.* Il Nuovo Cimento **14**(4) (1937) pp. 171–184.

[99] M. J. Dolinski, A. W. P. Poon, and W. Rodejohann, *Neutrinoless Double-Beta Decay: Status and Prospects.* `arXiv:1902.04097`.

[100] SuperNEMO Collaboration, *Probing new physics models of neutrinoless double beta decay with SuperNEMO.* The European Physical Journal C **70**(4) (2010) pp. 927–943.

[101] NEMO-3 Collaboration, *Search for neutrinoless double-beta decay of ^{100}Mo with the NEMO-3 detector.* Physical Review D **89**(111101(R)) (2014) .

[102] SNO+ Collaboration, *Current Status and Future Prospects of the SNO+ Experiment.* Advances in High Energy Physics **2016** (2016) pp. 1–21.

[103] P. Minkowski, *$\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?* Physics Letters B **67**(4) (1977) pp. 421–428.

[104] T. Yanagida, *Horizontal Symmetry and Masses of Neutrinos.* Progress of Theoretical Physics **64**(3) (1980) pp. 1103–1105.

[105] R. N. Mohapatra and G. Senjanović, *Neutrino Mass and Spontaneous Parity Nonconservation.* Physical Review Letters **44**(14) (1980) pp. 912–915.

[106] J. Schechter and J. W. F. Valle, *Neutrino masses in $SU(2) \otimes U(1)$ theories.* Physical Review D **22**(9) (1980) pp. 2227–2235.

[107] R. D. Klauber, *The Seesaw Mechanism.* `http://www.quantumfieldtheory.info/TheSeesawMechanism.htm` (accessed on 26/06/2019).

[108] CERN, *The matter-antimatter asymmetry problem.* `https://home.cern/science/physics/matter-antimatter-asymmetry-problem` (accessed on 26/06/2019).

[109] L. Canetti, M. Drewes, and M. Shaposhnikov, *Matter and antimatter in the universe.* New Journal of Physics **14**(095012) (2012) .

[110] Planck Collaboration, *Planck 2015 results.* Astronomy & Astrophysics **594**(A13) (2016) .

[111] CERN, *LHCb sees a new flavour of matter–antimatter asymmetry.* `https://home.cern/news/press-release/physics/lhcb-sees-new-flavour-matter-antimatter-asymmetry` (accessed on 2019-06-26).

[112] LHCb Collaboration, *Observation of CP Violation in Charm Decays.* Physical Review Letters **122**(211803) (2019) .

[113] N. Wolchover, *Do Neutrinos Explain Matter-Antimatter Asymmetry?*
`https://www.quantamagazine.org/`
`do-neutrinos-explain-matter-antimatter-asymmetry-20160728/` (accessed on
2019-06-26).

[114] R. G. Felipe, *Neutrinos and the matter-antimatter asymmetry in the Universe.*
`arXiv:1108.2694`.

[115] V. C. Rubin and W. Ford, *Rotation of the Andromeda Nebula from a Spectroscopic
Survey of Emission Regions.* The Astrophysical Journal **159** (1970) p. 379.

[116] V. C. Rubin *et al.*, *Rotation velocities of 16 SA galaxies and a comparison of Sa, Sb, and
SC rotation properties.* The Astrophysical Journal **289** (1985) p. 81.

[117] K. Freese, *Review of Observational Evidence for Dark Matter in the Universe and in
upcoming searches for Dark Stars.* `arXiv:0812.4005`.

[118] J. A. R. Cembranos, *Modified gravity and dark matter.* Journal of Physics: Conference
Series **718**(032004) (2016) .

[119] P. Salucci, *The distribution of dark matter in galaxies.* The Astronomy and Astrophysics
Review **27**(2) (2019) .

[120] N. Jarosik *et al.*, *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP)
Observations: Sky Maps, Systematic Errors, and Basic Results.* `arXiv:1001.4744v1`.

[121] V. Salvatelli *et al.*, *New constraints on coupled dark energy from the Planck satellite
experiment.* Physical Review D **88**(023531) (2013) .

[122] University of Cambridge, *Planck captures portrait of the young Universe, revealing
earliest light.* `https://www.cam.ac.uk/research/news/`
`planck-captures-portrait-of-the-young-universe-revealing-earliest-light`
(accessed on 2019-06-28).

[123] J. E. Kim *et al.*, *A review on axions and the strong CP problem.* AIP Conference
Proceedings **1200** (2009) pp. 83–92.

[124] Particle Data Group, *Axions and Other Similar Particles*, 2018.
`http://pdg.lbl.gov/2019/reviews/rpp2018-rev-axions.pdf` (accessed on
29/06/2019).

[125] G. Arcadi, *Dark matter phenomenology of GUT inspired simplified models.* Journal of
Physics: Conference Series **718**(042003) (2016) .

[126] F. S. Queiroz, *Dark Matter Overview: Collider, Direct and Indirect Detection Searches* in
Proceedings of the 51st Recontres de Moriond on Electroweak Interactions and Unified

Theories, 2016, pp. 427–436. ARISF. http://inspirehep.net/record/1466249/.

[127] J. Conrad, *Indirect Detection of WIMP Dark Matter: a compact review.* arXiv:1411.1925v1.

[128] M. Schumann, *Direct Detection of WIMP Dark Matter: Concepts and Status.* arXiv:1903.03026v1.

[129] A. Boveia and C. Doglioni, *Dark Matter Searches at Colliders.* Annual Review of Nuclear and Particle Science **68**(1) (2018) pp. 429–459.

[130] ATLAS Collaboration, *Measurement of the top quark mass in the $t\bar{t} \to$ dilepton channel from $\sqrt{s} = 8$ TeV ATLAS data.* Physics Letters B **761** (2016) pp. 350 – 371.

[131] Particle Data Group, *Status of Higgs Boson Physics.* http://pdg.lbl.gov/2019/reviews/rpp2018-rev-higgs-boson.pdf (accessed on 11/07/2019).

[132] C. Englert, A. Freitas, M. Spira, and P. M. Zerwas, *Constraining the intrinsic structure of top-quarks.* Physics Letters B **721** (2013) pp. 261–268.

[133] *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels.* Journal of High Energy Physics **2012**(54) (2012) .

[134] *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$.* Physical Review Letters **110**(252004) (2013) .

[135] Particle Data Group, *Top Quark.* http://pdg.lbl.gov/2019/reviews/rpp2018-rev-top-quark.pdf (accessed on 07/08/2019).

[136] LHC Top Working Group, *LHC Top Working Group Summary Plots.* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots (accessed on 07/08/2019).

[137] R. Frederix and F. Maltoni, *Top pair invariant mass distribution: a window on new physics.* Journal of High Energy Physics **01**(047) (2009) .

[138] A. Bethani, *Higgs physics at the LHC: Beyond the standard model.* https://www.qmul.ac.uk/spa/media/pprc/events/Higgs_Seminar_QMUL.pdf (accessed on 07/08/2019).

[139] T. Vickey, *Searching for additional Higgs bosons at the CERN Large Hadron Collider.* https://www2.physics.ox.ac.uk/sites/default/files/page/2012/03/27/tvickey-bsm-higgs-oxford-november-21-2017-42254.pdf (accessed on 07/08/2019).

[140] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector.* Physics Letters B **784** (2018) pp. 173–191.

[141] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* Physical Review D **98**(072016) (2018) .

[142] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the $e\mu$ channel in pp collisions at $\sqrt{s} =13$ TeV using the ATLAS detector.* The European Physical Journal C **77**(5) (2017) p. 292.

[143] ATLAS Collaboration, *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.* Journal of High Energy Physics **2015**(6) (2015) p. 100.

[144] LHC Top Working Group, *Particle level objects and pseudo-top-quark definitions.* `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/` `ParticleLevelTopDefinitions?rev=17` (accessed on 21/07/2019).

[145] ATLAS Collaboration, *Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector.* Physical Review D **98**(012003) (2018) .

[146] A. W. Jung, *Top differential cross section measurements (Tevatron).* `arXiv:1201.2660`.

[147] CDF Collaboration, *First Measurement of the $t\bar{t}$ Differential Cross Section $d\sigma/dM_{t\bar{t}}$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV.* Physical Review Letters **102**(222003) (2009) .

[148] D0 Collaboration, *Dependence of the $t\bar{t}$ production cross section on the transverse momentum of the top quark.* Physics Letters B **693** (2010) pp. 515–521.

[149] R. Y. Peters, *Top Quark Pair Production Cross Section at the Tevatron.* `arXiv:1509.07629v1`.

[150] D0 Collaboration, *Measurement of differential $t\bar{t}$ production cross sections in $p\bar{p}$ collisions.* Physical Review D **90**(092006) (2014) .

[151] ATLAS Collaboration, *Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV.* The European Physics Journal C **73**(2261) (2013) .

[152] ATLAS Collaboration, *Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector.* Physical Review D **90**(072004) (2014) .

[153] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section as function of jet multiplicity and jet transverse momentum in 7 TeV proton–proton collisions with the ATLAS detector.* Journal of High Energy Physics **2015**(20) (2015) .

[154] ATLAS Collaboration, *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.* Journal of High Energy Physics **2015**(100) (2015) .

[155] ATLAS Collaboration, *Measurement of top quark pair differential cross sections in the dilepton channel in pp collisions t $\sqrt{s} = 7$ and 8 TeV with ATLAS.* Physical Review D **94**(092003) (2016) .

[156] ATLAS Collaboration, *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.* Physical Review D **93**(032009) (2016) .

[157] ATLAS Collaboration, *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.* The European Physical Journal C **2016**(538) (2016) .

[158] ATLAS Collaboration, *Measurement of jet activity in top quark events using the $e\mu$ final state with two b-tagged jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.* Journal of High Energy Physics **2016**(74) (2016) .

[159] ATLAS Collaboration, *Measurement of lepton differential distributions and the top quark mass in $t\bar{t}$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.* The European Physical Journal C **77**(804) (2017) .

[160] ATLAS Collaboration, *Measurement of jet activity produced in top-quark events with an electron, a muon and two b-tagged jets in the final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* The European Physical Journal C **77**(220) (2017) .

[161] ATLAS Collaboration, *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.* Journal of High Energy Physics **2017**(191) (2017) .

[162] ATLAS Collaboration, *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.* Journal of High Energy Physics **2018**(159) (2018) .

[163] ATLAS Collaboration, *Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* Journal of High Energy Physics **2019**(46) (2019) .

[164] CMS Collaboration, *Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV*. The European Physical Journal C **73**(2339) (2012) .

[165] CMS Collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*. The European Physical Journal C **75**(542) (2015) .

[166] CMS Collaboration, *Measurement of the integrated and differential $t\bar{t}$ production cross sections for high-p_T top quarks in pp collisions at $\sqrt{s} = 8$ TeV*. Physical Review D **94**(072002) (2016) .

[167] CMS Collaboration, *Measurement of the differential cross sections for top quark pair production as a function of kinematic event variables in pp collisions at $\sqrt{s} = 7$ and 8 TeV*. Physical Review D **94**(052006) (2016) .

[168] CMS Collaboration, *Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV*. Physical Review D **95**(092001) (2017) .

[169] CMS Collaboration, *Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions*. The European Physical Journal C **2017**(459) (2017) .

[170] CMS Collaboration, *Measurement of normalized differential $t\bar{t}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV*. Journal of High Energy Physics **2018**(60) (2018) .

[171] CMS Collaboration, *Measurements of differential cross sections of top quark pair production as a function of kinematic event variables in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Journal of High Energy Physics **2018**(2) (2018) .

[172] CMS Collaboration, *Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV*. Physical Review D **97**(112003) (2018) .

[173] CMS Collaboration, *Measurements of $t\bar{t}$ differential cross sections in proton–proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons*. Journal of High Energy Physics **2018**(149) (2018) .

[174] CMS Collaboration, *Measurement of $t\bar{t}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s} = 13$ TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions*. `arXiv:1904.05237`.

[175] S. Q. Wang *et al.*, *A precise determination of the top-quark pole mass*. The European Physical Journal C **78**(237) (2018) .

[176] M. C. Smith and S. S. Willenbrock, *Top-quark Pole Mass.* Physical Review Letters **79**(325) (1997) .

[177] G. Corcella, *The Top-Quark Mass: Challenges in Definition and Determination.* Frontiers in Physics **7** (2019) p. 54.

[178] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$ =7 and 8 TeV with the ATLAS detector.* The European Physical Journal C **74**(3109) (2014) .

[179] ATLAS Collaboration, *Measurement of the top-quark mass in $t\bar{t}$ + 1-jet events collected with the ATLAS detector in pp collisions at $\sqrt{s}$ = 8 TeV.* arXiv:1905.02302.

[180] CMS Collaboration, *Measurement of $t\bar{t}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s}$ =13 TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions.* arXiv:1904.05237v1.

[181] The New York Times, *"Because it's there." – George Mallory*, 1923. http://graphics8.nytimes.com/packages/pdf/arts/mallory1923.pdf (accessed on 21/07/2019).

[182] M. Czakon, D. Heymes, and A. Mitov, *High-Precision Differential Predictions for Top-Quark Pairs at the LHC.* Physical Review Letters **116**(082003) (2016) .

[183] M. Czakon *et al.*, *Top-pair production at the LHC through NNLO QCD and NLO EW.* Journal of High Energy Physics **2017**(186) (2017) .

[184] CMS Collaboration, *Measurement of the $t\bar{t}$ production cross section at 13 TeV in the all-jets final state.* CMS-PAS-TOP-16-013, 2016. https://cds.cern.ch/record/2161138.

[185] CMS Collaboration, *The CMS experiment at the CERN LHC.* Journal of Instrumentation **3**(S08004) (2008) .

[186] L. Evans and P. Bryant, *LHC Machine.* Journal of Instrumentation **3**(S08001) (2008) .

[187] K. Schindl, *The Injector Chain for the LHC* in Proceedings of the 9th LEP-SPS Performance Workshop, Chamonix, France, 1999, pp. 47–52. https://cds.cern.ch/record/384396.

[188] CERN, *The Large Hadron Collider.* https://home.cern/science/accelerators/large-hadron-collider (accessed on 13/07/2019).

[189] CERN, *Facts and figures about the LHC.* https://home.cern/resources/faqs/facts-and-figures-about-lhc (accessed on

13/07/2019).

[190] CERN, *The CERN Accelerator Complex.* https://cds.cern.ch/record/2636343 (accessed on 13/07/2019).

[191] CERN, *Pulling together: Superconducting electromagnets.* https://home.cern/science/engineering/pulling-together-superconducting-electromagnets (accessed on 13/07/2019).

[192] CERN, *Cryogenics: Low temperatures, high performance.* https://home.cern/science/engineering/cryogenics-low-temperatures-high-performance (accessed on 13/07/2019).

[193] D. E. Baynham, R. C. Coombs, A. Ijspeert, and R. Perin, *Design of Superconducting Corrector Magnets for LHC.* IEEE Transactions on Magnetics **30** (1994) pp. 1823–1826.

[194] D. Boussard and T. P. R. Linnecar, *The LHC Superconducting RF System* in Joint Cryogenic Engineering Conference and International Cryogenic Materials Conference, 1999. https://cdsweb.cern.ch/record/410377.

[195] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider.* Journal of Instrumentation **3**(S08003) (2008) .

[196] LHCb Collaboration, *The LHCb Detector at the LHC.* Journal of Instrumentation **3**(S08005) (2008) .

[197] ALICE Collaboration, *The ALICE experiment at the CERN LHC.* Journal of Instrumentation **3**(S08002) (2008) .

[198] W. Herr and B. Muratori, *Concept of luminosity* in CERN Accelerator School: Intermediate Course on Accelerator Physics, 2006.

[199] R. Steerenberg, *LHC Report: The LHC is full!* https://home.cern/news/news/accelerators/lhc-report-lhc-full (accessed on 14/07/2019).

[200] J. Vrerla, *Timing and Synchronization in the LHC Experiments* in 6th Workshop on Electronics for LHC Experiments, 2000.

[201] M. Cirilli and M. Barnett, *First collisions in ATLAS.* https://atlas.cern/updates/atlas-news/first-collisions-atlas (accessed on 13/07/2019).

[202] CERN, *CERN sets date for first attempt at 7 TeV collisions in the LHC.* https://home.cern/news/press-release/cern/cern-sets-date-first-attempt-7-tev-collisions-lhc (accessed on 13/07/2019).

[203] ATLAS Collaboration, *ATLAS Fact Sheet.* `https://cds.cern.ch/record/1457044` (accessed on 24/07/2019).

[204] ATLAS Experiment, CERN, *Computer generated image of the whole ATLAS detector.* `https://cds.cern.ch/record/1095924` (accessed on 14/07/2019).

[205] ATLAS Collaboration, *ATLAS central solenoid: Technical Design Report.* CERN-LHCC-97-021, ATLAS-TDR-9, 1997. `https://cdsweb.cern.ch/record/331067`.

[206] ATLAS Collaboration, *ATLAS barrel toroid: Technical Design Report.* CERN-LHCC-97-019, ATLAS-TDR-7, 1997. `https://cdsweb.cern.ch/record/331065`.

[207] ATLAS Collaboration, *ATLAS end-cap toroids: Technical Design Report.* CERN-LHCC-97-020, ATLAS-TDR-8, 1997. `https://cdsweb.cern.ch/record/331066`.

[208] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report.* CERN-LHCC-97-016, ATLAS-TDR-4, 1997. `https://cdsweb.cern.ch/record/331063`.

[209] ATLAS Experiment, CERN, *Computer generated image of the ATLAS inner detector.* `https://cds.cern.ch/record/1095926` (accessed on 28/07/2019).

[210] ATLAS Collaboration, *ATLAS pixel detector: Technical Design Report.* CERN-LHCC-98-013, ATLAS-TDR-11, 1998. `https://cdsweb.cern.ch/record/381263`.

[211] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report.* CERN-LHCC-2010-013, ATLAS-TDR-19, 2010. `https://cdsweb.cern.ch/record/1291633`.

[212] A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation.* Journal of Instrumentation **11** (2016) p. C12036.

[213] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report.* CERN-LHCC-96-041, ATLAS-TDR-2, 1996. `https://cdsweb.cern.ch/record/331061`.

[214] ATLAS Experiment, CERN, *Computer generated image of the ATLAS Liquid Argon.* `https://cds.cern.ch/record/1095928` (accessed on 31/07/2019).

[215] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report.* CERN-LHCC-96-042, ATLAS-TDR-3, 1996.

`https://cdsweb.cern.ch/record/331062`.

[216] M. Aharrouche *et al.*, *Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam.* Nuclear Instruments and Methods in Physics Research A **568** (2006) pp. 601–623.

[217] ATLAS Collaboration, *Results from a new combined test of an electromagnetic liquid argon calorimeter with a hadronic scintillating-tile calorimeter.* Nuclear Instruments and Methods in Physics Research A **449** (2000) pp. 462–477.

[218] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report.* CERN-LHCC-97-022, ATLAS-TDR-10, 1997. `https://cdsweb.cern.ch/record/331068`.

[219] ATLAS Experiment, CERN, *Computer generated image of the ATLAS Muons subsystem.* `https://cds.cern.ch/record/1095929` (accessed on 31/07/2019).

[220] E. W. Weisstein, *Sagitta.* From *MathWorld* – A Wolfram Web Resource. `http://mathworld.wolfram.com/Sagitta.html` (accessed on 2019-08-18).

[221] A. Ruiz Martínez, *The Run-2 ATLAS Trigger System.* Journal of Physics **762**(012003) (2016) .

[222] ATLAS Collaboration, *ATLAS level-1 trigger: Technical Design Report.* CERN-LHCC-98-014, ATLAS-TDR-12, 1998. `https://cdsweb.cern.ch/record/381429`.

[223] E. Simioni, *The Hardware Topological Trigger of ATLAS: Commissioning and Operations.* ATL-DAQ-SLIDE-2018-102, 2016. `https://cds.cern.ch/record/2305791`.

[224] ATLAS Collaboration, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report.* CERN-LHCC-2003-022, ATLAS-TDR-16, 2003. `https://cdsweb.cern.ch/record/616089`.

[225] A. Martyniuk *et al.*, *The Database Driven ATLAS Trigger Configuration System.* Journal of Physics: Conference Series **664**(082030) (2015) .

[226] G. Crone *et al.*, *The ATLAS ReadOut System – Performance with first data and perspective for the future.* Nuclear Instruments and Methods in Physics Research A **623** (2010) pp. 534–536.

[227] W. Panduro Vazquez, *The ATLAS Data Acquisition system in LHC Run 2.* Journal of Physics: Conference Series **898**(032017) (2017) .

[228] LHC Experiments Committee, CERN, *LHC computing Grid: Technical Design Report.* CERN-LHCC-2005-024, LCG-TDR-001, 2004. `https://cds.cern.ch/record/840543`.

[229] ATLAS Collaboration, *Software and Computing.*
https://atlas.cern/discover/detector/software-computing (accessed on
20/08/2019).

[230] R. Seuster *et al.*, *Status and Future Evolution of the ATLAS Offline Software.* Journal of
Physics: Conference Series **664**(072044) (2015) .

[231] T. Cornelissen *et al.*, *Concepts, Design and Implementation of the ATLAS New Tracking
(NEWT).* ATL-SOFT-PUB-2007-007, 2007. https://cds.cern.ch/record/1020106/.

[232] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting.* Nuclear
Instruments and Methods in Physics Research A **262** (1987) pp. 444–450.

[233] W. Waltenberger, R. Frühwirth, and P. Vanlaer, *Adaptive vertex fitting.* Journal of
Physics G: Nuclear and Particle Physics **34**(12) (2007) .

[234] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian
Sum Filter-based model for bremsstrahlung.* ATLAS-CONF-2012-047, 2012.
https://cdsweb.cern.ch/record/1449796.

[235] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using
the 2015 LHC proton–proton collision data* in 51st Rencontres de Moriond on QCD and
High Energy Interactions, 2016. http://cds.cern.ch/record/2157687.

[236] J. Illingworth and J. Kittler, *A survey of the Hough transform.* Computer Vision,
Graphics, and Image Processing **44** (1988) pp. 87–116.

[237] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in
proton–proton collision data at $\sqrt{s} =13$ TeV.* The European Physical Journal C **76**(292)
(2016) .

[238] ATLAS Collaboration, *Reconstruction of hadronic decay products of tau leptons with the
ATLAS experiment.* The European Physical Journal C **76**(295) (2016) .

[239] ATLAS Collaboration, *Expected performance of the ATLAS experiment: detector, trigger
and physics*, 2009. https://cds.cern.ch/record/1125884.

[240] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in
proton–proton collisions at $\sqrt{s} =7$ TeV.* The European Physical Journal C **73**(2304)
(2013) .

[241] ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance.*
ATL-LARG-PUB-2008-002, 2008. https://cds.cern.ch/record/1099735.

[242] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its
performance in LHC Run 1.* The European Physical Journal C **77**(490) (2017) .

[243] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm.* Journal of High Energy Physics **2008**(063) (2008) .

[244] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS.* ATLAS-CONF-2015-037, 2015. https://cds.cern.ch/record/2044941.

[245] M. Cacciari and S. G. P, *Pileup subtraction using jet areas.* Physics Letters B **659** (2008) pp. 119–126.

[246] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} =$ 13 TeV.* ATL-PHYS-PUB-2015-015, 2015. https://cds.cern.ch/record/2037613.

[247] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector.* ATLAS-CONF-2014-018, 2014. https://cds.cern.ch/record/1700870.

[248] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2* in European Physical Society Conference on High Energy Physics, 2015. https://cds.cern.ch/record/2037697.

[249] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run.* ATL-PHYS-PUB-2016-012, 2016. https://cds.cern.ch/record/2160731.

[250] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data.* ATLAS-CONF-2011-102, 2011. https://cds.cern.ch/record/1369219.

[251] G. Piacquadio and C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS.* Journal of Physics: Conference Series **119**(3) (2008) .

[252] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment.* Journal of Instrumentation **11**(P04008) (2016) .

[253] ATLAS Collaboration, *ATL-COM-PHYS-2018-507*, 2018. Unpublished ATLAS internal document in preparation (restricted). https://cds.cern.ch/record/2316994.

[254] ATLAS Collaboration, *The ATLAS Simulation Infrastructure.* The European Physical Journal C **70** (2010) pp. 823–874.

[255] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics.* Soviet Journal of Experimental and Theoretical Physics **47** (1977) pp. 641–653. http://www.jetp.ac.ru/cgi-bin/dn/e_046_04_0641.pdf.

[256] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory.* Soviet Journal of Experimental and Theoretical Physics **15** (1972) pp. 438–450.

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

[258] B. R. Webber, *A QCD model for jet fragmentation including soft gluon interference.* Nuclear Physics B **238** (1984) pp. 492–528.

[259] S. Agostinelli *et al.*, *Geant4 – a simulation toolkit.* Nuclear Instruments and Methods in Physics Research A **506** (2003) pp. 250–303.

[260] J. Allison *et al.*, *Geant4 Developments and Applications.* IEEE Transactions on Nuclear Science **53** (2006) pp. 270–278.

[261] J. Allison *et al.*, *Recent developments in Geant4.* Nuclear Instruments and Methods in Physics Research A **835** (2016) pp. 186–225.

[262] J. L. Henning, *SPEC CPU2000: measuring CPU performance in the New Millennium.* IEEE Computer **33** (2000) pp. 28–35.

[263] E. Richter-Was, D. Froidvaux, and L. Poggioli, *ATLFAST 2.0 a fast simulation package for ATLAS.* ATL-PHYS-98-131, 1998. `https://cds.cern.ch/record/683751`.

[264] ATLAS Collaboration, *Measurements of top-quark pair single- and double-differential cross-sections in the all-hadronic channel in pp collisions at $\sqrt{s} =13$ TeV using the ATLAS detector*, 2019. Unpublished ATLAS document in preparation (restricted). `https://cds.cern.ch/record/2693215`.

[265] T. Khoo and A. Sfyrla, *Studies of extra jets*, 2017. Slides from ATLAS internal meeting. `http://cern.ch/go/6XP8` (accessed on 08/02/2020).

[266] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} =7$ and 8 TeV with the ATLAS detector.* The European Physical Journal C **74**(3109) (2014) .

[267] C. Degrande, J.-M. Gérard, C. Grojean, F. Maltoni, and G. Servant, *Non-resonant new physics in top pair production at hadron colliders.* Journal of High Energy Physics **2011**(125) (2011) .

[268] ATLAS Collaboration, *Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector.* Physical Review D **96** (2017) p. 052004.

[269] ATLAS Collaboration, *Search for New Phenomena in Dijet Angular Distributions in Proton-Proton Collisions at $\sqrt{s} =8$ TeV Measured with the ATLAS Detector.* Physical

Review Letters **114**(221802) (2015) .

[270] Fermilab E706 Collaboration, *Evidence for Parton k_T Effects in High-p_T Particle Production.* Physical Review Letters **81**(2642) (1998) .

[271] ATLAS Collaboration, *ATL-COM-PHYS-2018-126.* Unpublished ATLAS internal document in preparation (restricted). `https://cds.cern.ch/record/2304451`.

[272] ATLAS Collaboration, *Search for contact interactions and large extra dimensions in the dilepton channel using proton–proton collisions at $\sqrt{s} =8$ TeV with the ATLAS detector.* The European Physical Journal C **74**(3134) (2014) .

[273] A. Denner, S. Dittmaier, S. Kallweit, and S. Pozzorini, *Next-to-Leading-Order QCD Corrections to $W^+W^-b\bar{b}$ Production at Hadron Colliders.* Physical Review Letters **106**(052001) (2011) .

[274] G. Bevilacqua, M. Czakon, A. van Hameren, C. G. Papadopoulos, and M. Worek, *Complete off-shell effects in top quark pair hadroproduction with leptonic decay at next-to-leading order.* Journal of High Energy Physics **2011**(83) (2011) .

[275] ATLAS Collaboration, *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.* The European Physical Journal C **79**(1028) (2019) .

[276] ATLAS Collaboration, *Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z/γ^* and $t\bar{t}$ cross sections.* ATL-PHYS-PUB-2018-017, 2018. `https://cds.cern.ch/record/2633819`.

[277] T. Barillari, *Top-quark mass and top-quark pole mass measurements with the ATLAS detector* in Proceedings of the Fifth Annual Large Hadron Collider Physics Conference, 2017. `arXiv:1710.06019v1`.

[278] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016* in 9th International Workshop on Top Quark Physics, 2016. `https://cds.cern.ch/record/2216168`.

[279] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO.* ATL-PHYS-PUB-2017-007, 2017. `https://cds.cern.ch/record/2261938`.

[280] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms.* Journal of High Energy Physics **2004**(040) (2004) .

[281] S. Frixione, P. Nason, and O. C, *Matching NLO QCD computations with parton shower simulations: the POWHEG method.* Journal of High Energy Physics **2007**(070) (2007) .

[282] S. Alioli *et al.*, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX.* Journal of High Energy Physics **2010**(43) (2010) .

[283] NNPDF Collaboration, *Parton distributions with LHC data.* Nuclear Physics B **867** (2013) pp. 244–289.

[284] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual.* Journal of High Energy Physics **2006**(026) (2006) .

[285] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes.* ATL-PHYS-PUB-2014-021, 2014. `https://cds.cern.ch/record/1966419`.

[286] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1.* Journal of High Energy Physics **2009**(007) (2009) .

[287] E. Bothmann *et al.*, *Event Generation with Sherpa 2.2.* `arXiv:1905.09127v2`.

[288] J. Alwall *et al.*, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.* Journal of High Energy Physics **2014**(79) (2014) .

[289] S. Bähr *et al.*, *Herwig++ physics and manual.* The European Physical Journal C **58** (2008) pp. 639–707.

[290] J. Bellm *et al.*, *Herwig 7.0/Herwig++ 3.0 release note.* The European Physical Journal C **76**(196) (2016) .

[291] L. A. Harland-Lang *et al.*, *Parton distributions in the LHC era: MMHT 2014 PDFs.* The European Physical Journal C **75**(204) (2015) .

[292] Particle Data Group, *Quantum Chromodynamics.* `http://pdg.lbl.gov/2019/reviews/rpp2019-rev-qcd.pdf` (accessed on 03/02/2020).

[293] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization of Hard Processes in QCD.* `arXiv:hep-ph/0409313`.

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

[295] Y. L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics.* Zh. Eksp. Teor. Fiz **73** (1977) p. 1216.

[296] V. N. Gribov and L. N. Lipatov, *Deep inelastic scattering in perturbation theory.* Sov. J. Nucl. Phys. **15** (1972) p. 438.

[297] G. P. Salam, *Basics of QCD, Lecture 3: PDFs and DGLAP*. Slides from ICTP–SAIFR school on QCD and LHC physics, July 2015, São Paulo, Brazil. `https://gsalam.web.cern.ch/gsalam/repository/talks/2015-SaoPaulo-lecture3.pdf` (accessed on 03/02/2020).

[298] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions to ATLAS measurements of top pair production at 7 TeV*. ATL-PHYS-PUB-2015-002, 2015. `https://cds.cern.ch/record/1981319`.

[299] M. Czakon *et al.*, *NNLO QCD predictions for fully-differential top-quark pair production at the Tevatron.* Journal of High Energy Physics **2016**(34) (2016) .

[300] G. Luisoni and S. Marzani, *QCD resummation for hadronic final states.* `arXiv:1505.04084v3`.

[301] A. Loginov, *Strategies of data-driven estimations of $t\bar{t}$ backgrounds in ATLAS.* ATL-PHYS-PROC-2010-067, 2010. `https://cds.cern.ch/record/1287126/`.

[302] L. B. Lucy, *An iterative technique for the rectification of observed distributions.* Astron. J. **79** (1974) pp. 745–754.

[303] W. H. Richardson, *Bayesian-Based Iterative Method of Image Restoration.* J. Opt. Soc. Am. **62**(1) (Jan, 1972) pp. 55–59. `http://www.osapublishing.org/abstract.cfm?URI=josa-62-1-55`.

[304] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem.* Nucl. Inst. Methods Phys. Res. A **362** (1995) pp. 487–498.

[305] G. Zech, *Iterative unfolding with the Richardson-Lucy algorithm.* Nucl. Inst. Methods Phys. Res. A **716** (2013) pp. 1–9, `arXiv:1210.5177`.

[306] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society B **39**(1) (1977) pp. 1–38.

[307] T. Adye, *Unfolding algorithms and tests using RooUnfold.* `arXiv:1105.1160v1`.

[308] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC.* The European Physics Journal C **76**(653) (2016) .

[309] S. van der Meer, *Calibration of the effective beam height in the ISR.* CERN-ISR-PO-68-31. `https://cds.cern.ch/record/296752`.

[310] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* Physical Review D **96**(072002) (2017) .

[311] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.* The European Physical Journal C **75**(17) (2015) .

[312] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV.* The European Physical Journal C **73**(2304) (2013) .

[313] ATLAS Collaboration, *Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC.* The European Physical Journal C **74**(2305) (2013) .

[314] ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8$ TeV.* ATLAS-CONF-2015-017, 2015. `https://cds.cern.ch/record/2008678`.

[315] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector.* ATLAS-CONF-2014-018, 2014. `https://cds.cern.ch/record/1700870`.

[316] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector.* ATL-PHYS-PUB-2017-011, 2017. `https://cds.cern.ch/record/2270366`.

[317] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV.* ATL-PHYS-PUB-2016-015, 2016. `https://cds.cern.ch/record/2203514`.

[318] J. Butterworth *et al.*, *PDF4LHC recommendations for LHC Run II.* Journal of Physics G: Nuclear and Particle Physics **43**(023001) (2016) .

[319] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning.* The MIT Press, 2006. `http://www.gaussianprocess.org/gpml/` (accessed on 15/05/2018).

[320] D. J. C. MacKay, *Introduction to Gaussian Processes.* NATO ASI Series F Computer and Systems Sciences **168** (1998) pp. 133–166. `https://www.ics.uci.edu/~welling/teaching/KernelsICS273B/gpB.pdf` (accessed on 23/12/2018).

[321] D. K. Duvenaud, *Automatic Model Construction with Gaussian Processes.* PhD thesis, Computational and Biological Learning Laboratory, University of Cambridge, 2014. `https://www.cs.toronto.edu/~duvenaud/thesis.pdf` (accessed on 23/12/2018).

[322] G. H. Golub and C. F. Van Loan, *Matrix Computations, Third Edition.* The Johns Hopkins University Press, 1996.

[323] T. E. Oliphant, *A guide to NumPy.* Trelgol Publishing, 2006.

[324] T. E. Oliphant, *Python for Scientific Computing.* Computing in Science & Engineering **9**(3) (2007) pp. 10–20.

[325] M. Ebden, *Gaussian Processes: A Quick Introduction.* `arXiv:1505.02965v2`.

[326] R. von Mises, *Mathematical Theory of Probability and Statistics.* Academic Press, 1964.

[327] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O'Neil, *Fast Direct Methods for Gaussian Processes.* IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(2) (2016) pp. 252–265, `arXiv:1403.6015v2`.

[328] R. Murray-Smith and A. Girard, *Gaussian Process Priors with ARMA Noise Models* in Proceedings of the Irish Signals and Systems Conference, 2001. `http://www.dcs.gla.ac.uk/~rod/publications/MurGir01.pdf` (accessed on 27/12/2018).

[329] B. W. Silverman, *Spline Smoothing: The Equivalent Variable Kernel Method.* The Annals of Statistics **12**(3) (1984) pp. 898–916.

[330] T. Hastie and R. Tibshirani, *Generalized Additive Models.* Chapman & Hall/CRC, 1990.

[331] M. U. Gutmann and J. Corander, *Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models.* Journal of Machine Learning Research **17** (2016) pp. 1–47. `http://www.jmlr.org/papers/volume17/15-017/15-017.pdf` (accessed on 20/03/2019).

[332] G. Louppe, *Bayesian Optimization with Scikit-Optimize* in PyData Amsterdam, 2017. `https://orbi.uliege.be/bitstream/2268/226433/1/PyData2017_BayesianoptimizationwithScikit-Optimize.pdf` (accessed on 27/12/2018).

[333] M. Stone, *Cross-Validatory Choice and Assessment of Statistical Predictions.* Journal of the Royal Statistical Society, Series B (Methodological) **36**(2) (1974) pp. 111–147. `http://www.jstor.org/stable/2984809` (accessed on 02/05/2019).

[334] P. Ilten, M. Williams, and Y. Yang, *Event generator tuning using Bayesian optimization.* Journal of Instrumentation **12**(4) (2017) , `arXiv:1610.08328v2`.

[335] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to* Pythia *8.1.* Computer Physics Communications **178**(11) (2008) pp. 852–867.

[336] P. Skands, S. Carrazza, and J. Rojo, *Tuning* Pythia *8.1: the Monash 2013 tune.* European Physical Journal C **74** (2014) p. 3024, `arXiv:1404.5630v1`.

[337] G. Louppe *et al.*, *Scikit-Optimize*. `https://scikit-optimize.github.io` (accessed on 27/12/2018).

[338] J. Snoek, H. Larochelle, and R. P. Adams, *Practical Bayesian Optimization of Machine Learning Algorithms*. `arXiv:1206.2944`.

[339] M. Krasser, *Bayesian Optimization*, 2018. `http://krasserm.github.io/2018/03/21/bayesian-optimization/` (accessed on 27/12/2018).

[340] D. R. Jones, M. Schonlau, and W. J. Welch, *Efficient Global Optimization of Expensive Black-Box Functions*. Journal of Global Optimization **13**(4) (1998) pp. 455–492.

[341] C. K. I. Williams and C. E. Rasmussen, *Gaussian Processes for Regression* in Advances in Neural Information Processing Systems, 1996, pp. 514–520. `http://papers.nips.cc/paper/1048-gaussian-processes-for-regression.pdf` (accessed on 11/03/2018).

[342] M. L. Stein, *Interpolation of spatial data: some theory for kriging*. Springer–Verlag, 1999.

[343] E. W. Weisstein, *Gamma function*. From *MathWorld* – A Wolfram Web Resource. `http://mathworld.wolfram.com/GammaFunction.html` (accessed on 11/03/2019).

[344] E. W. Weisstein, *Modified Bessel Function of the First Kind*. From *MathWorld* – A Wolfram Web Resource. `http://mathworld.wolfram.com/ModifiedBesselFunctionoftheFirstKind.html` (accessed on 11/03/2019).

[345] N. Durrande, *Kernel Design*. Lecture at GP Summer School, Sheffield, September 2015. `http://gpss.cc/gpss15/talks/KernelDesign.pdf` (accessed on 21/01/2020).

[346] N. Cressie and C. K. Wikle, *Statistics for Spatio-Temporal Data*. John Wiley & Sons, 2011.

[347] M. Abramowitz and I. A. Stegun (editors), *Handbook of Mathematical Functions*. Dover Publications, 1965.

[348] G. E. Uhlenbeck and L. S. Ornstein, *On the Theory of the Brownian Motion*. Physical Review **36**(5) (1930) pp. 823–841.

[349] J. L. Doob, *The Brownian Movement and Stochastic Equations*. Annals of Mathematics **43**(2) (1942) pp. 351–369.

[350] O. Vasicek, *An equilibrium characterization of the term structure*. Journal of Financial Economics **5**(2) (1977) pp. 177–188.

[351] C. Gaetan and X. Guyon, *Spatial Statistics and Modeling*. Springer, 2010.

[352] M. N. Gibbs, *Bayesian Gaussian Processes for Regression and Classification*. PhD thesis, University of Cambridge, 1997. `http://www.inference.org.uk/mng10/GP/thesis.ps` (accessed on 27/05/2018).

[353] M. Frate, K. Cranmer, S. Kalia, A. Vandenberg-Rodes, and D. Whiteson, *Modeling Smooth Backgrounds and Generic Localized Signals with Gaussian Processes*, 2017.

[354] A. J. Bozson, G. D. Cowan, and F. Spanò, *Unfolding with Gaussian Processes*. `arXiv:1811.01242`.

[355] A. J. Bozson, *Gaussian Processes for HEP*, 2018. Joint APP and HEPP Annual Conference, Institute of Physics. `https://github.com/adambozson/gp-talk/blob/master/Talk.pdf` (accessed on 22/01/2020).

[356] G. D. Cowan, *Statistical Data Analysis*. Oxford University Press, 1998.

[357] S. Geman, E. Bienenstock, and R. Doursat, *Neural Networks and the Bias/Variance Dilemma*. Neural Computation **4**(1) (1992) pp. 1–58.

[358] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[359] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[360] D. J. C. MacKay, *Bayesian Interpolation*. Neural Computation **4**(3) (1992) pp. 415–447.

[361] D. J. C. MacKay, *A Practical Bayesian Framework for Backpropagation Networks*. Neural Computation **4**(3) (1992) pp. 448–472.

[362] P. S. Laplace, *Mémoire sur la probabilité des causes par les évènemens*. Mémoires de Mathématique et de Physique, Tome Sixième (1774) pp. 621–656.

[363] S. M. Stigler, *Memoir on the Probability of the Causes of Events*. Statistical Science **1**(3) (1986) pp. 364–378. `http://www.jstor.org/stable/2245476` (accessed on 19/03/2019).

[364] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[365] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.

[366] G. Casella, *An Introduction to Empirical Bayes Data Analysis*. The American Statistician **39**(2) (1985) pp. 83–87.

[367] E. Jones, T. E. Oliphant, P. Peterson, *et al.*, *SciPy: Open source scientific tools for Python*. `http://www.scipy.org/` (accessed on 20/03/2019).

[368] R. P. Brent, *Algorithms for Minimization Without Derivatives*. Prentice–Hall, 1972.

[369] T. J. Dekker, *Finding a zero by means of successive linear interpolation* in Constructive aspects of the fundamental theorem of algebra, B. Dejon and P. Henrici (editors), 1969, pp. 37–51.

[370] G. Cowan, *A survey of unfolding methods for particle physics* in Proc. Conference on Advanced Statistical Techniques in Particle Physics, Durham, England, Mar., 2002, pp. 248–257. `http://www.ippp.dur.ac.uk/old/Workshops/02/statistics/proceedings/cowan.pdf`.

[371] V. Blobel, *Unfolding Methods in Particle Physics* in Proc. PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, 2011, pp. 240–251. CERN, Geneva, Switzerland.

[372] F. Spanò, *Unfolding in particle physics: a window on solving inverse problems*. EPJ Web Conf. **55** (2013) p. 03002.

[373] D. Foreman-Mackey, E. Agol, S. Ambikasaran, and R. Angus, *Fast and scalable Gaussian process modeling with applications to astronomical time series*. Astron. J. **154**(6) (2017) p. 220, `1703.09710`.

[374] I. Andersen, A. Szymkowiak, C. Rasmussen, L. Hanson, J. Marstrand, H. Larsson, and L. Hansen, *Perfusion quantification using Gaussian process deconvolution*. Magn. Reson. Med. **48**(2) pp. 351–361.

[375] J. A. Christen, B. Sansó, M. Santana-Cibrian, and J. X. Velasco-Hernández, *Bayesian deconvolution of oil well test data using Gaussian processes*. J. Appl. Stat. **43**(4) (2016) pp. 721–737.

[376] B. Hunt, *Bayesian Methods in Nonlinear Digital Image Restoration*. IEEE Trans. Comput. **C-26**(3) (1977) pp. 219–229.

[377] V. Blobel, *Inverse problems—unfolding*, 2003. `http://graduierten-kurse.physi.uni-heidelberg.de/WiSe2003/lectures/blobel/export/invr/master.pdf` (accessed on 24/12/2019).

[378] A. Tikhonov, *On the solution of ill-posed problems and the method of regularization*. Mat. Sb. **151**(3) (1963) pp. 501–504.

[379] D. L. Phillips, *A Technique for the Numerical Solution of Certain Integral Equations of the First Kind*. J. ACM **9**(1) (1962) pp. 84–97.

[380] M. Schmelling, *The method of reduced cross-entropy: A general approach to unfold probability distributions.* Nucl. Inst. Methods Phys. Res. A **340**(2) (1994) pp. 400–412.

[381] G. Zech, *Analysis of distorted measurements – parameter estimation and unfolding* . arXiv:1607.06910 [physics.data-an].

[382] G. Choudalakis, *Fully Bayesian Unfolding.* arXiv:1201.4612v4.

[383] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models.* Chapman and Hall, 1990.

[384] O. Williams and A. Fitzgibbon, *Gaussian process implicit surfaces* in Proc. Gaussian Processes in Practice, 2007, pp. 1–4. https://www.microsoft.com/en-us/research/publication/gaussian-process-implicit-surfaces-2/.

[385] G. Wahba, *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics, Jan, 1990.

[386] https://github.com/adambozson/gp-unfold.

[387] A. M. Stuart, *Inverse problems: A Bayesian perspective.* Acta Numerica **19** (2010) p. 451–559.

[388] K. Cranmer, *Practical Statistics for the LHC.* arXiv:1503.07622.

[389] G. Aad et al. (ATLAS Collaboration), *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** (2016) p. 538.

[390] G. Aad et al. (ATLAS Collaboration), *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** (2016) pp. 1–34.

[391] A. Shah, A. Wilson, and Z. Ghahramani, *Student-t processes as alternatives to Gaussian processes* in Proc. Artificial Intelligence and Statistics, 2014, pp. 877–885.

Appendix A

Background compositions

The background composition distributions for one-dimensional observables to be unfolded to particle level are shown in Figures 4.11 to 4.16 in the main text.

A Background compositions

Particle level 2D

(a) $p_T^{t_1}$ vs. $p_T^{t_2}$

(b) $p_T^{t_1}$ vs. $m^{t\bar{t}}$

(c) $p_T^{t_2}$ vs. $m^{t\bar{t}}$
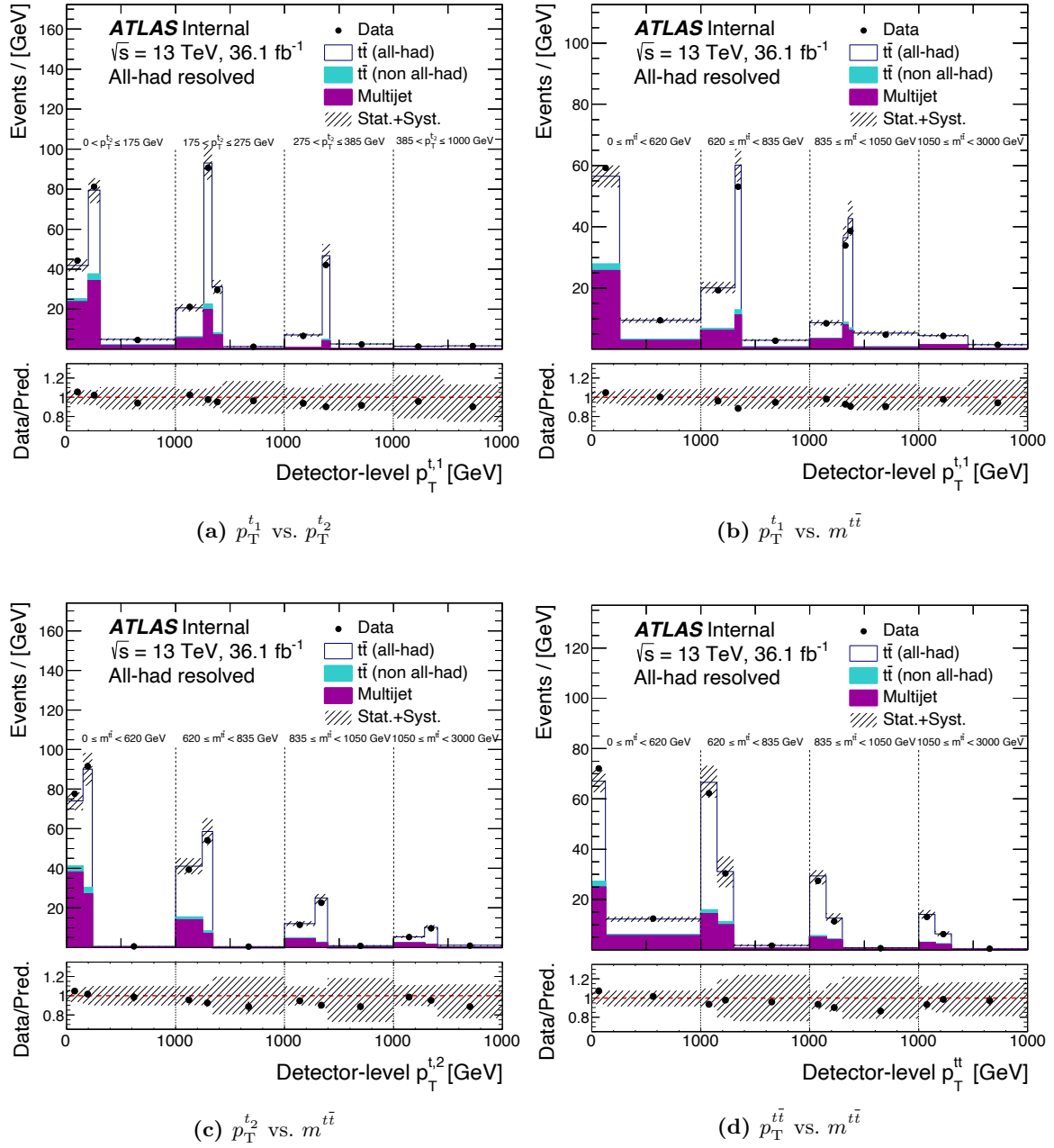
(d) $p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$

Figure A.1 Distributions in the signal region for (a) $p_T^{t_1}$ vs. $p_T^{t_2}$, (b) $p_T^{t_1}$ vs. $m^{t\bar{t}}$, (c) $p_T^{t_2}$ vs. $m^{t\bar{t}}$, and (d) $p_T^{t\bar{t}}$ vs. $m^{t\bar{t}}$. The shaded area indicated the total statistical and systematic uncertainty.

A Background compositions

(a) $|y^{t\bar{t}}|$ vs. $m^{t\bar{t}}$

(b) $p_{\mathrm{T}}^{t_1}$ vs. N_{jets}

(c) $p_{\mathrm{T}}^{t_2}$ vs. N_{jets}
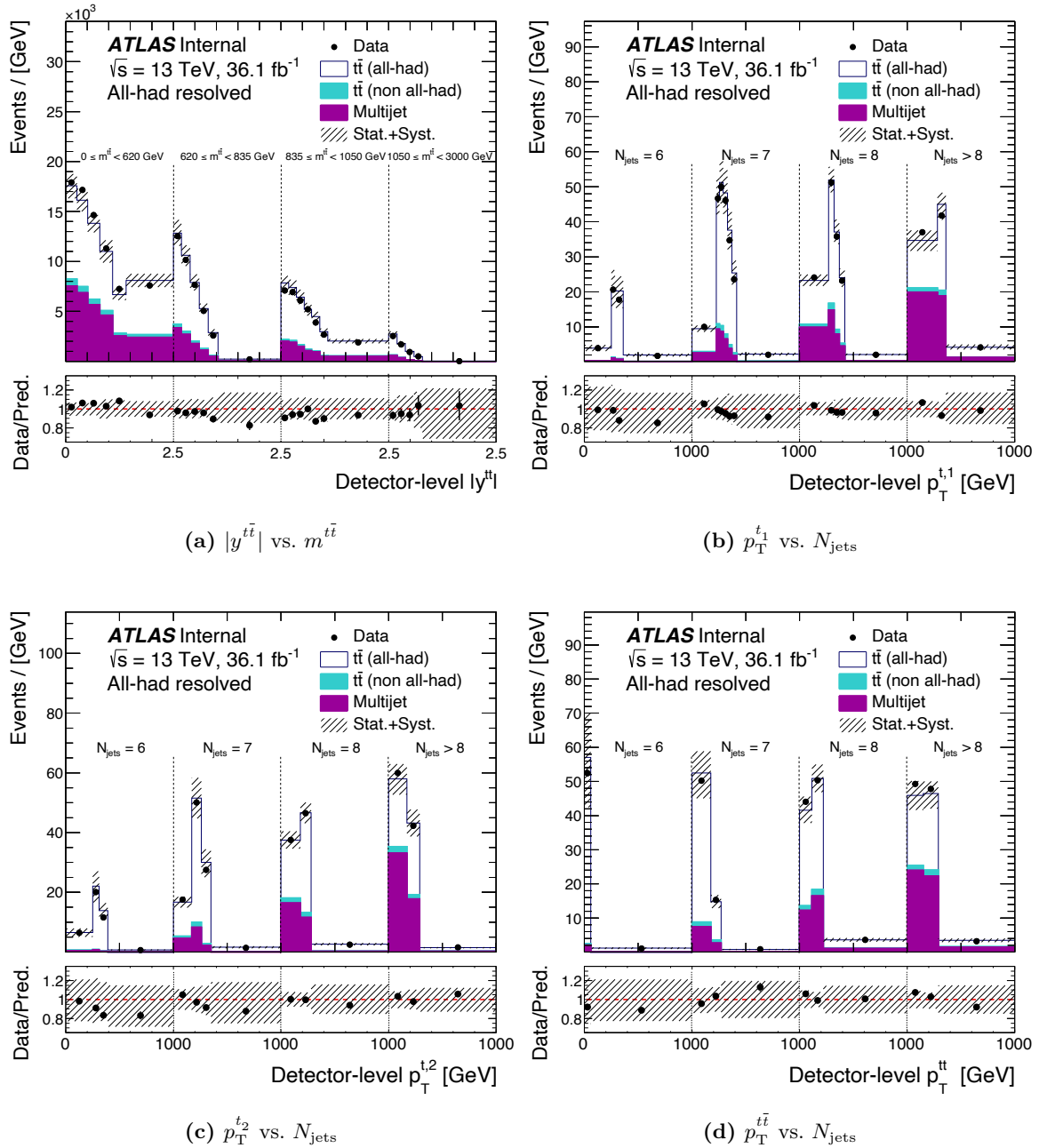
(d) $p_{\mathrm{T}}^{t\bar{t}}$ vs. N_{jets}

Figure A.2 Distributions in the signal region for (a) $|y^{t\bar{t}}|$ vs. $m^{t\bar{t}}$, (b) $p_{\mathrm{T}}^{t_1}$ vs. N_{jets}, (c) $p_{\mathrm{T}}^{t_2}$ vs. N_{jets}, and (d) $p_{\mathrm{T}}^{t\bar{t}}$ vs. N_{jets}. The shaded area indicated the total statistical and systematic uncertainty.
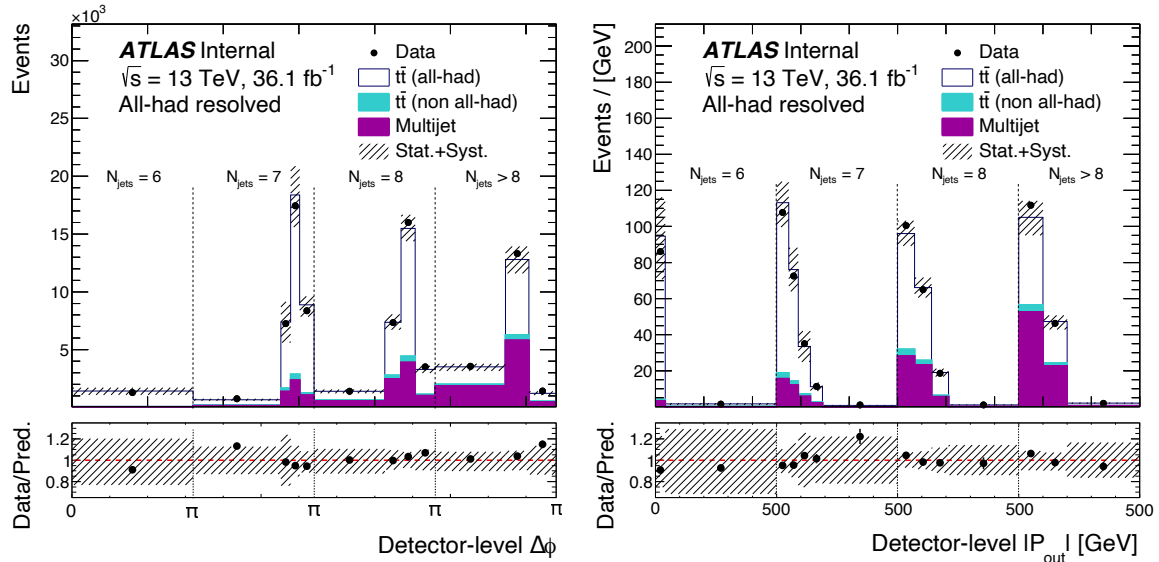
(a) $\Delta\phi$ vs. N_{jets}
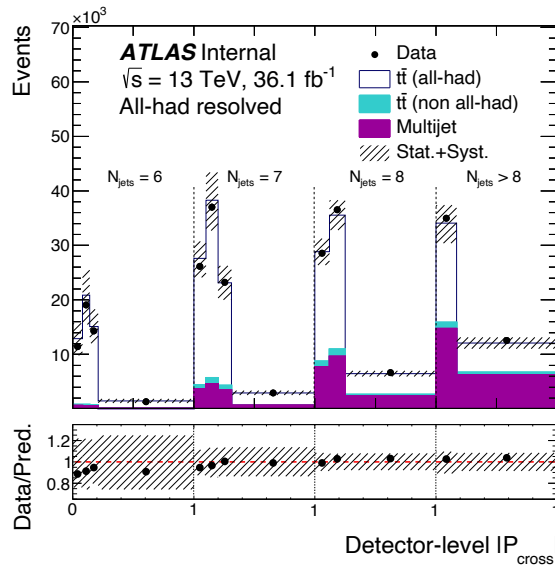
(b) $|P_{\text{out}}|$ vs. N_{jets}

(c) $|P_{\text{cross}}|$ vs. N_{jets}

Figure A.3 Distributions in the signal region for (a) $\Delta\phi$ vs. N_{jets}, (b) $|P_{\text{out}}|$ vs. N_{jets}, and (c) $|P_{\text{cross}}|$ vs. N_{jets}. The shaded area indicated the total statistical and systematic uncertainty.

Parton level 1D

(a) $p_T^{t_1}$
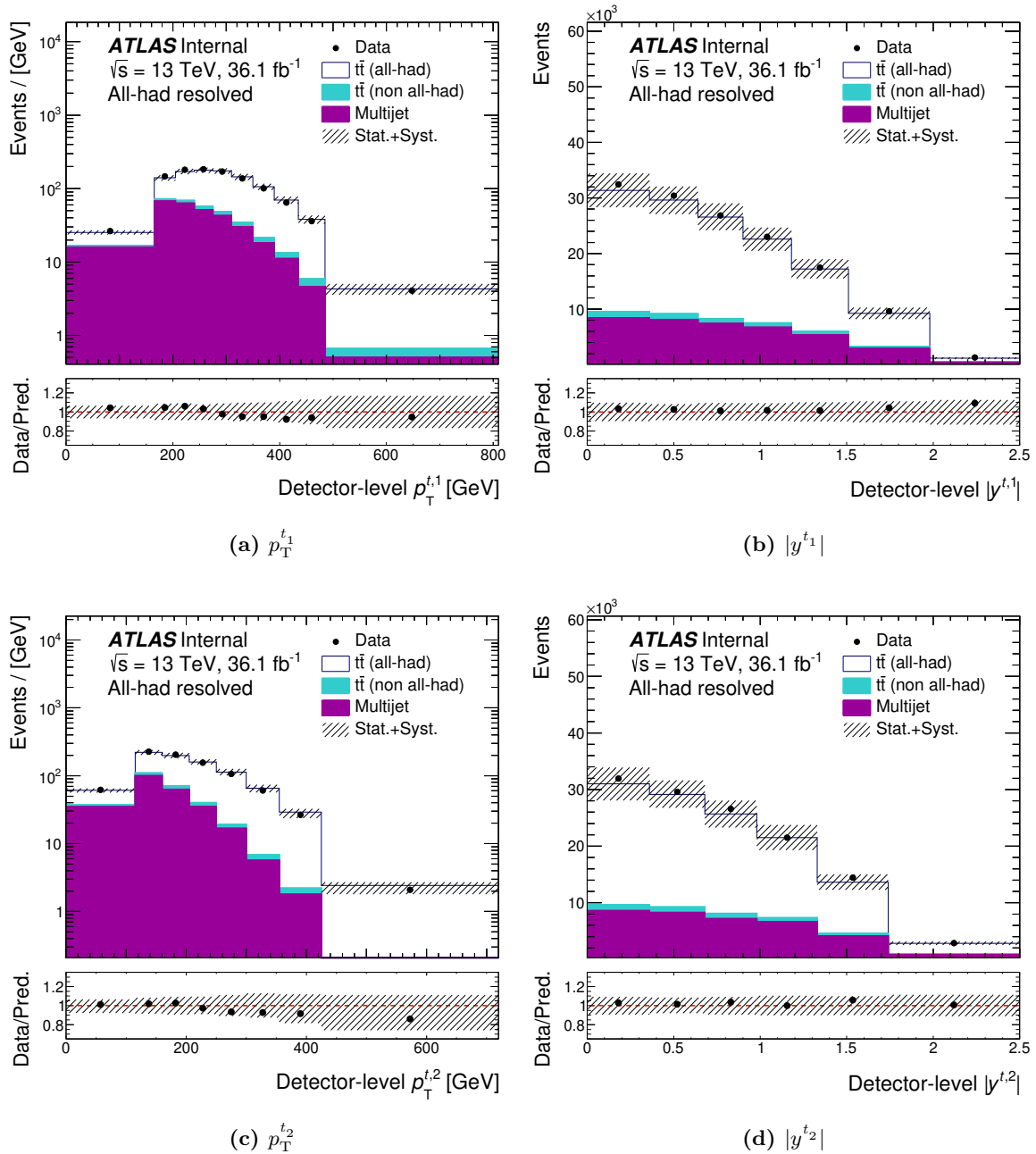
(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure A.4 Distributions in the signal region for (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$. The shaded area indicated the total statistical and systematic uncertainty.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
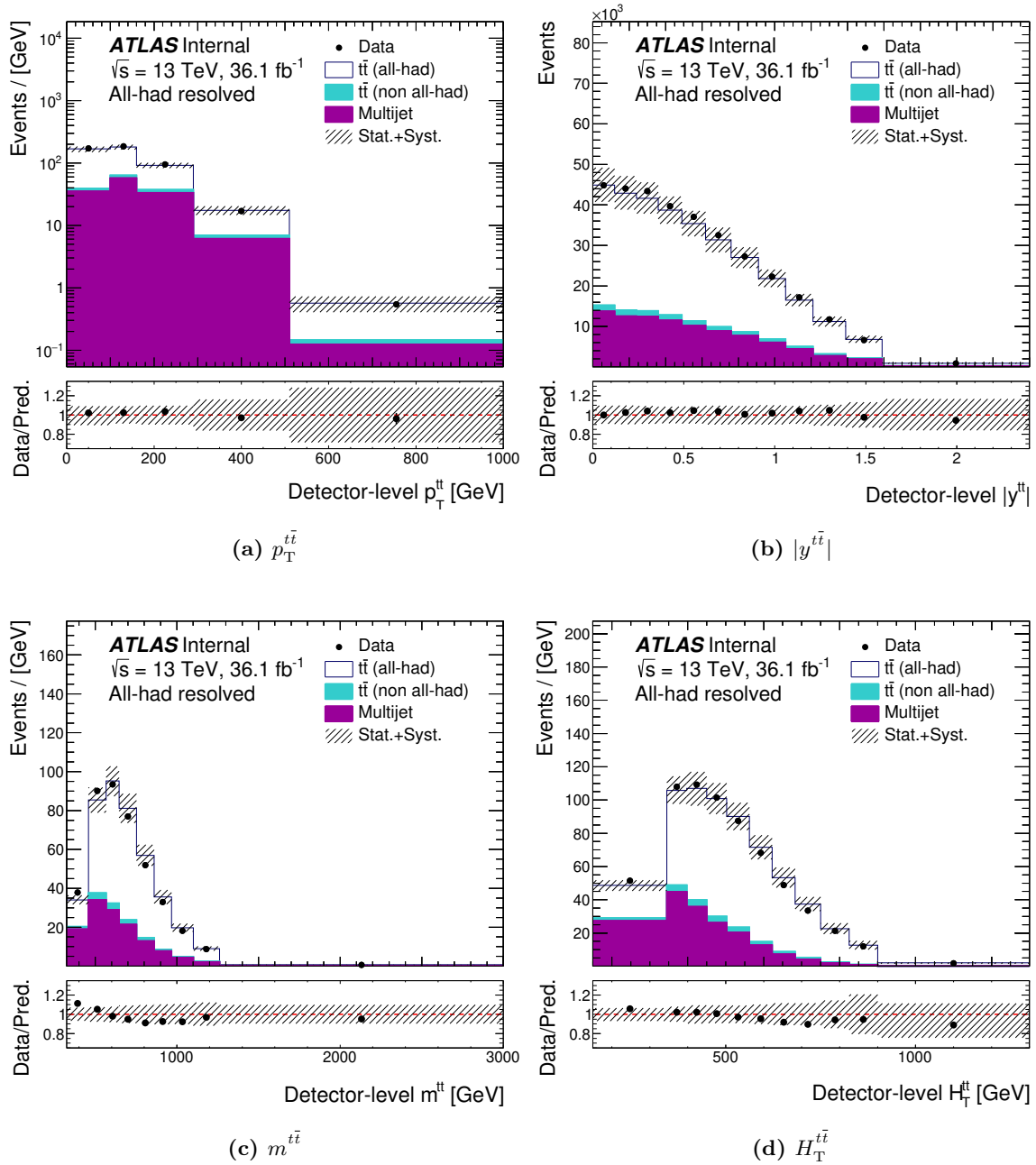
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure A.5 Distributions in the signal region for (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$. The shaded area indicated the total statistical and systematic uncertainty.

(a) $\chi^{t\bar{t}}$

(b) $\Delta\phi$
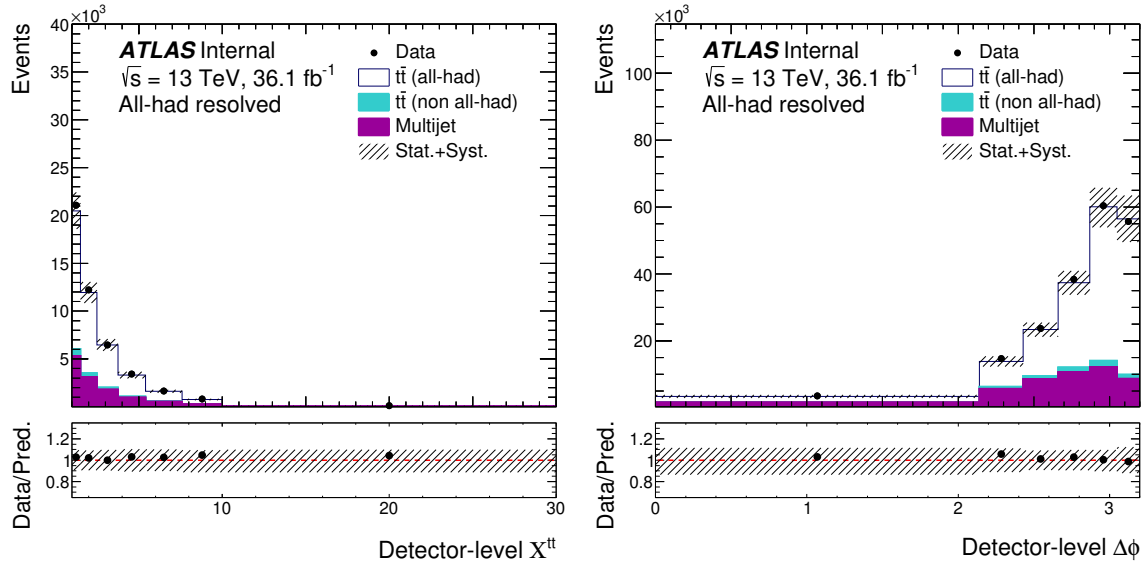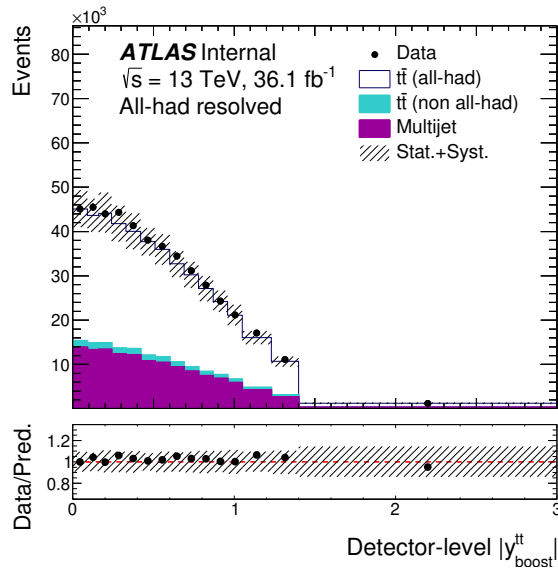
(c) $|y_{\text{boost}}|$

Figure A.6 Distributions in the signal region for (a) $\chi^{t\bar{t}}$, (b) $\Delta\phi$, and (c) $|y_{\text{boost}}|$. The shaded area indicated the total statistical and systematic uncertainty.

A Background compositions

Parton level 2D

(a) $p_\mathrm{T}^{t_1}$ vs. $p_\mathrm{T}^{t_2}$

(b) $|y^{t_1}|$ vs. $|y^{t_2}|$

(c) $p_\mathrm{T}^{t_1}$ vs. $m^{t\bar{t}}$
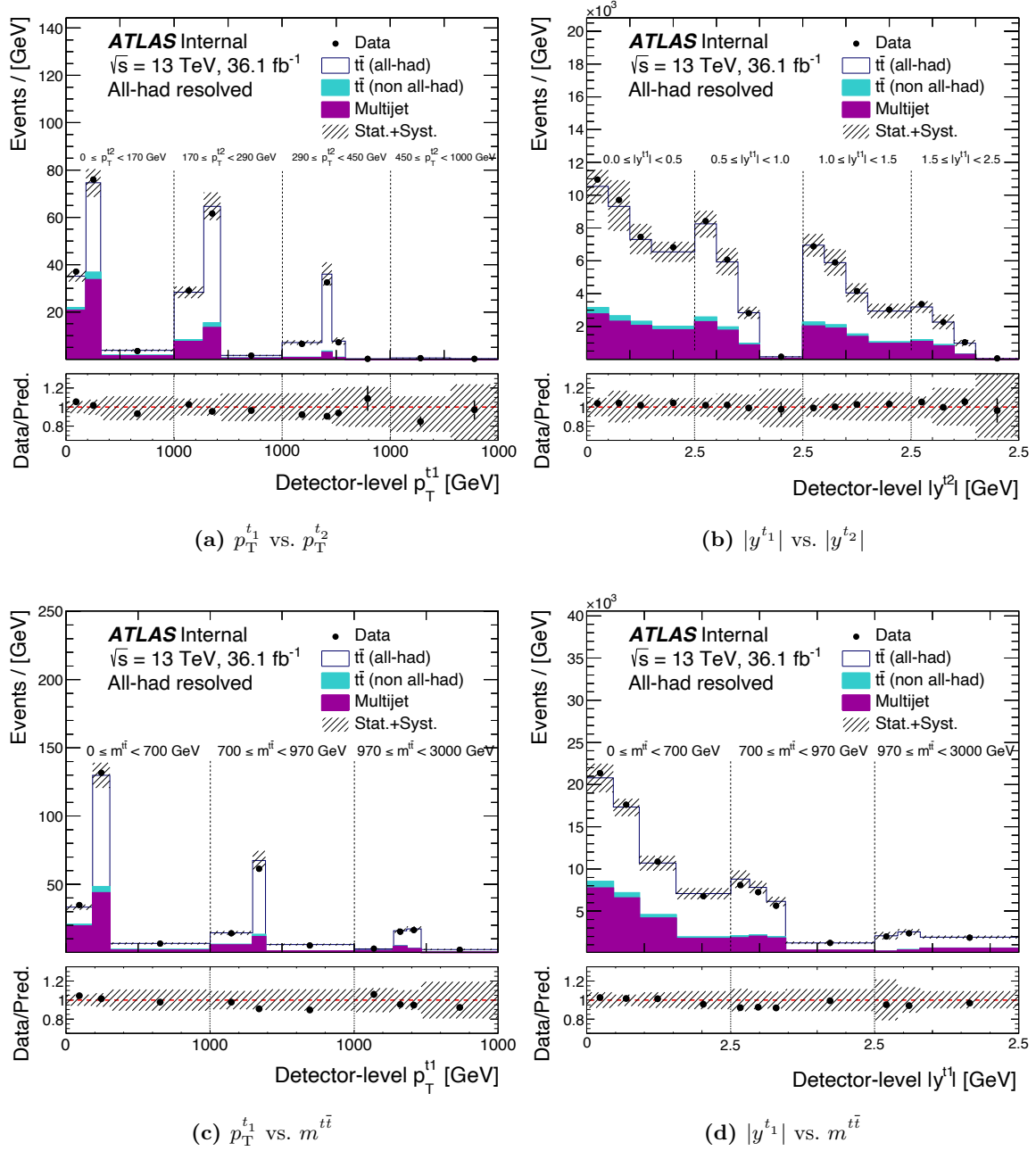
(d) $|y^{t_1}|$ vs. $m^{t\bar{t}}$

Figure A.7 Distributions in the signal region for (a) $p_\mathrm{T}^{t_1}$ vs. $p_\mathrm{T}^{t_2}$, (b) $|y^{t_1}|$ vs. $|y^{t_2}|$, (c) $p_\mathrm{T}^{t_1}$ vs. $m^{t\bar{t}}$, and (d) $|y^{t_1}|$ vs. $m^{t\bar{t}}$. The shaded area indicated the total statistical and systematic uncertainty.

302

(a) $p_{\mathrm{T}}^{t_2}$ vs. $m^{t\bar{t}}$

(b) $|y^{t_2}|$ vs. $m^{t\bar{t}}$

(c) $p_{\mathrm{T}}^{t\bar{t}}$ vs. $m^{t\bar{t}}$
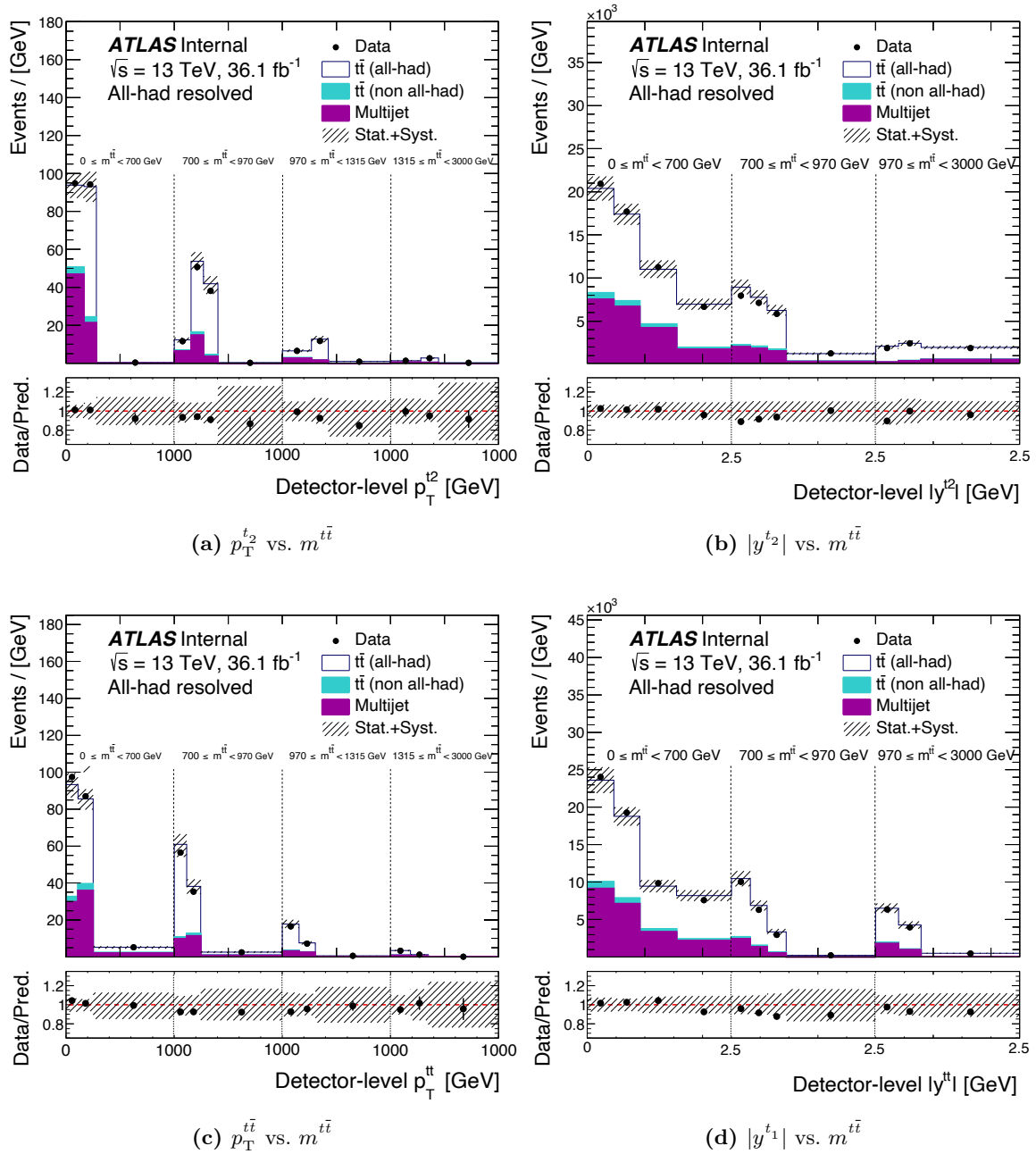
(d) $|y^{t_1}|$ vs. $m^{t\bar{t}}$

Figure A.8 Distributions in the signal region for (a) $p_{\mathrm{T}}^{t_2}$ vs. $m^{t\bar{t}}$, (b) $|y^{t_2}|$ vs. $m^{t\bar{t}}$, (c) $p_{\mathrm{T}}^{t\bar{t}}$ vs. $m^{t\bar{t}}$, and (d) $|y^{t_1}|$ vs. $m^{t\bar{t}}$. The shaded area indicated the total statistical and systematic uncertainty.

Appendix B

Uncertainty compositions

Normalised cross sections

The uncertainty compositions for one-dimensional normalised cross sections unfolded to particle level are shown in Figures 4.33 to 4.38 in the main text.

Particle level 2D
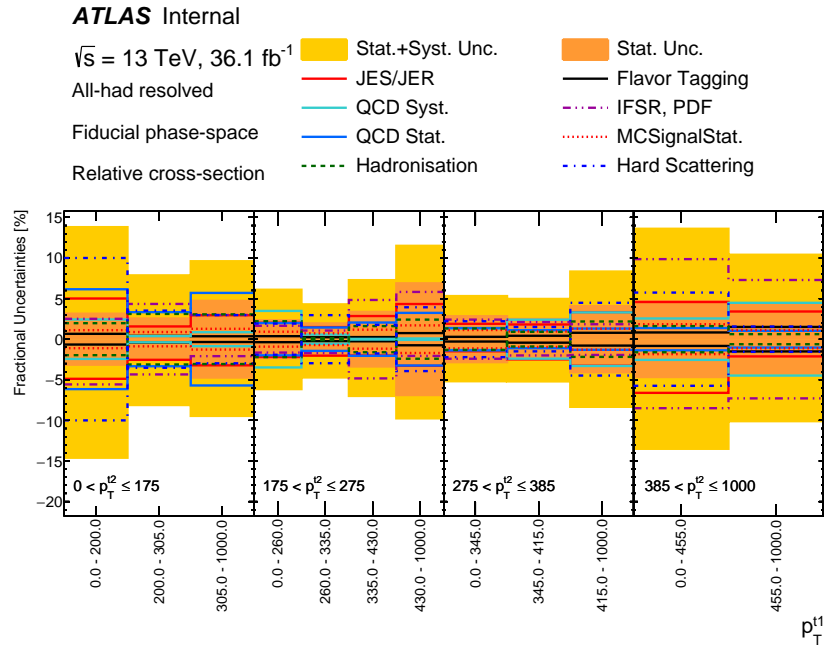
Figure B.1 Uncertainty composition for the $p_T^{t_1}$ cross section in bins of $p_T^{t_2}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
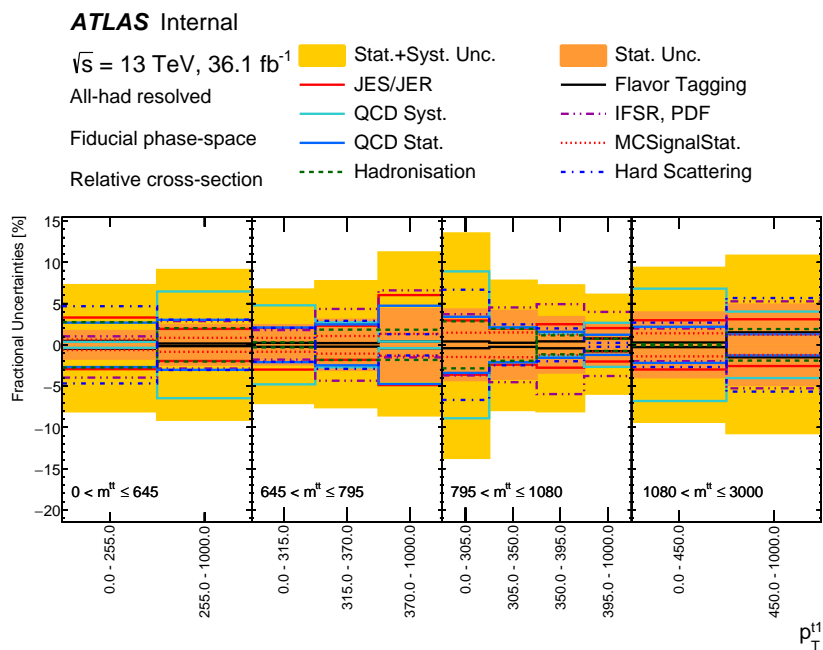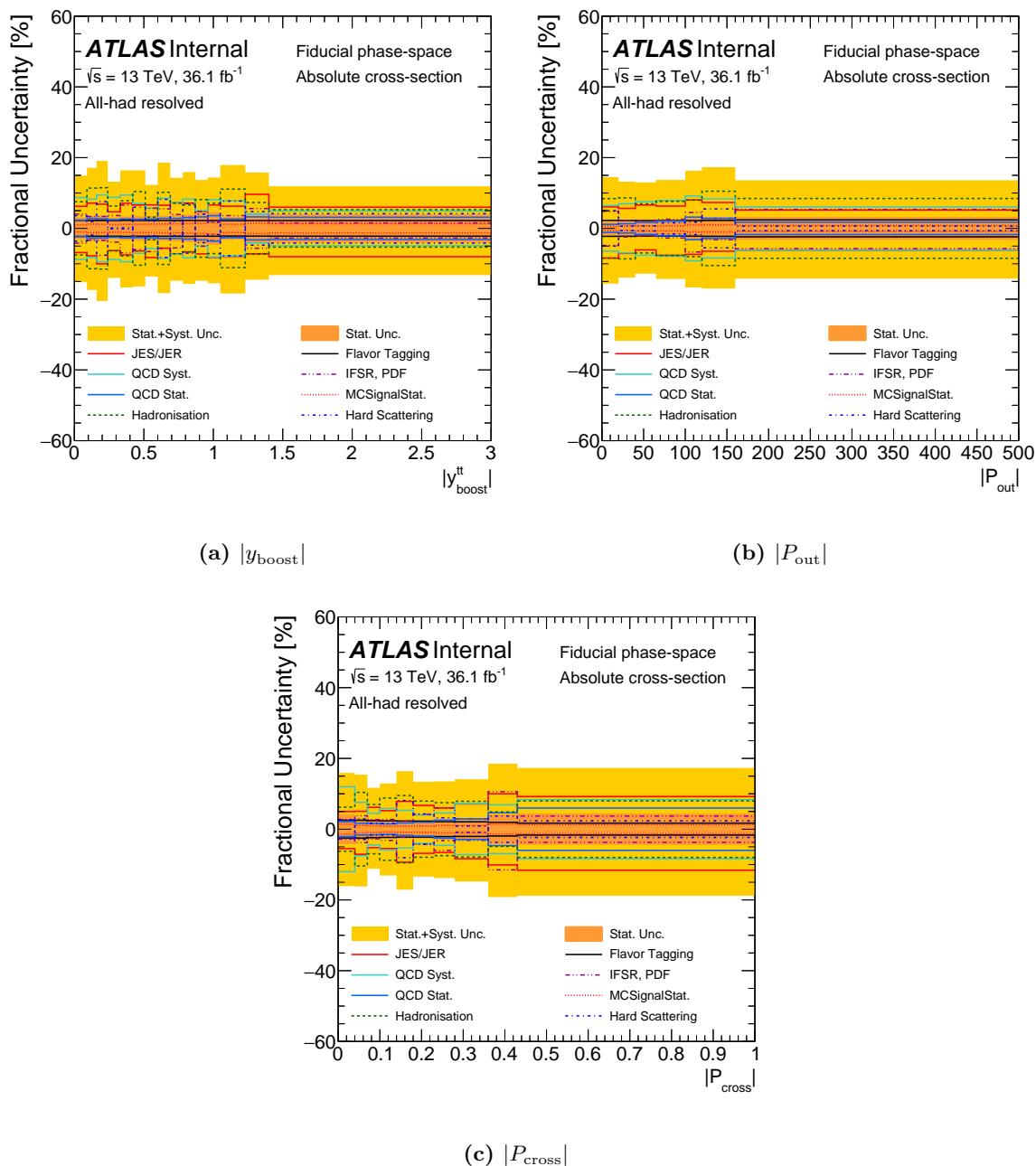
Figure B.2 Uncertainty composition for the $p_{\mathrm{T}}^{t_1}$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
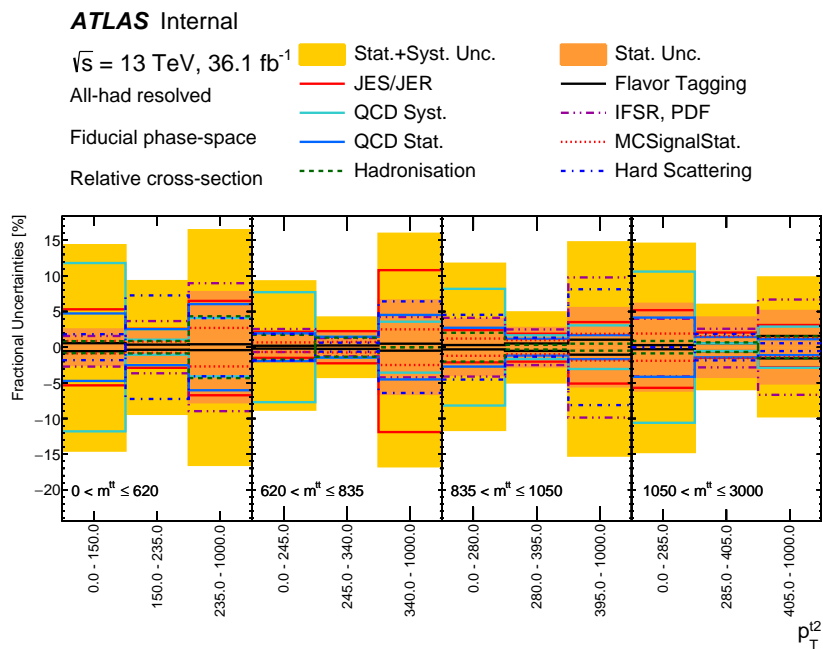
Figure B.3 Uncertainty composition for the $p_{\mathrm{T}}^{t_2}$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
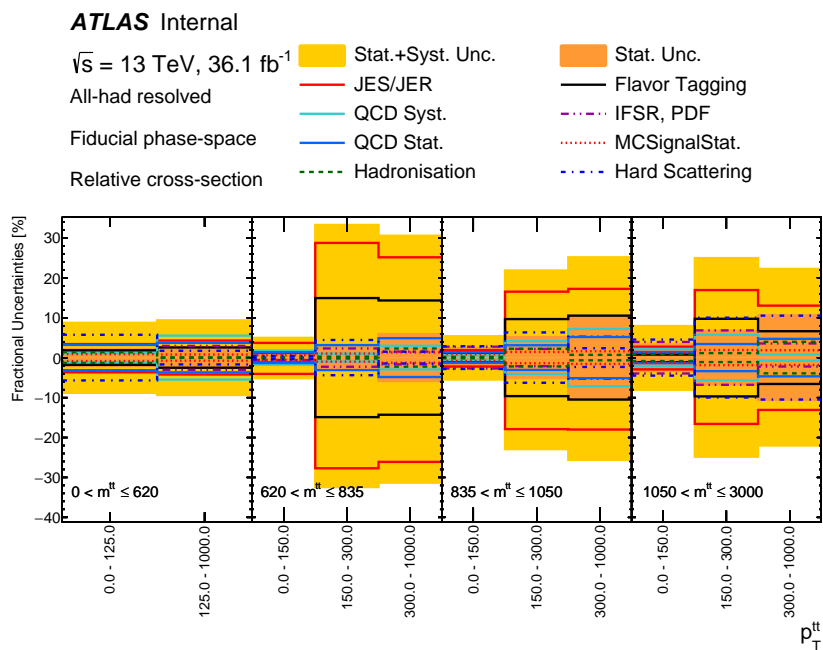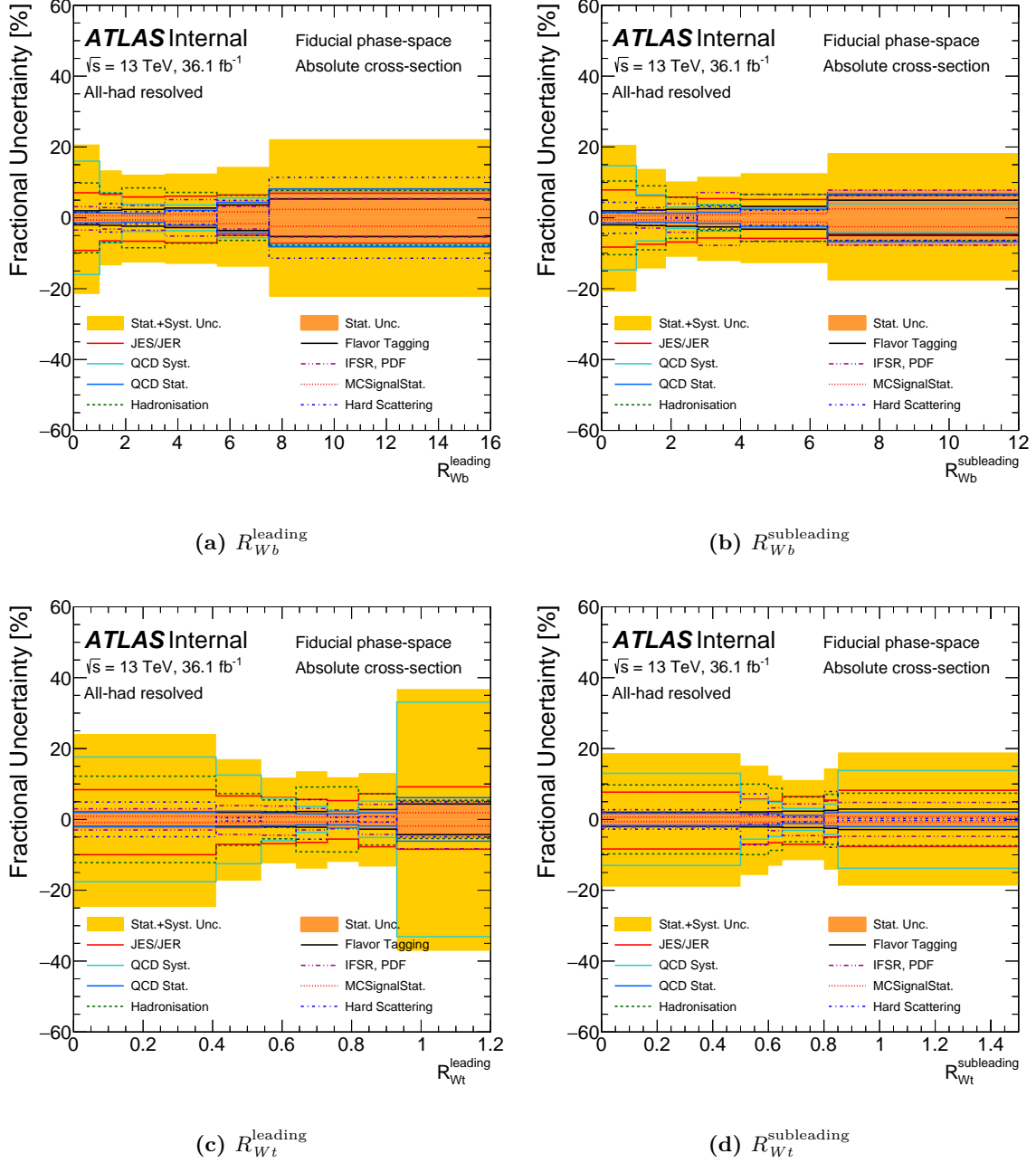
Figure B.4 Uncertainty composition for the $p_{\mathrm{T}}^{t\bar{t}}$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
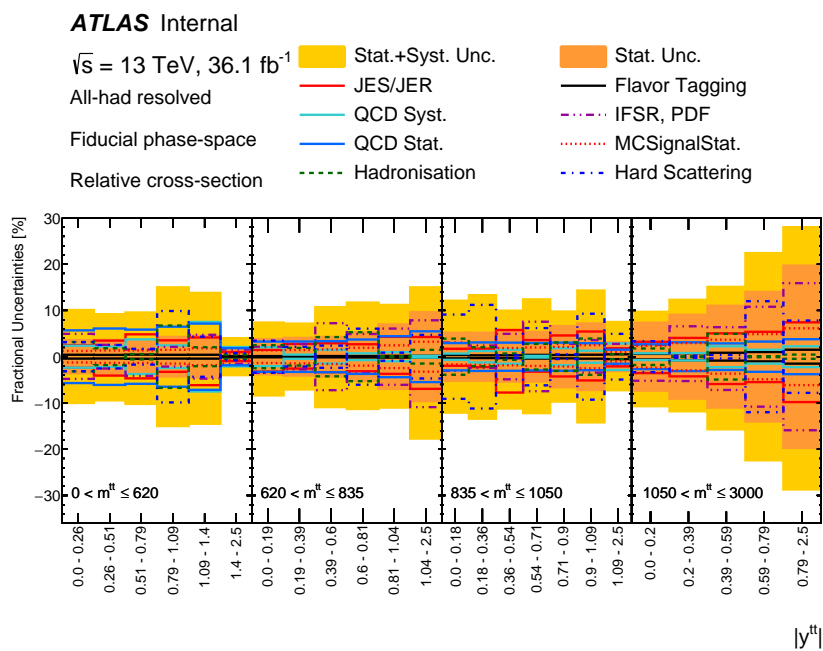
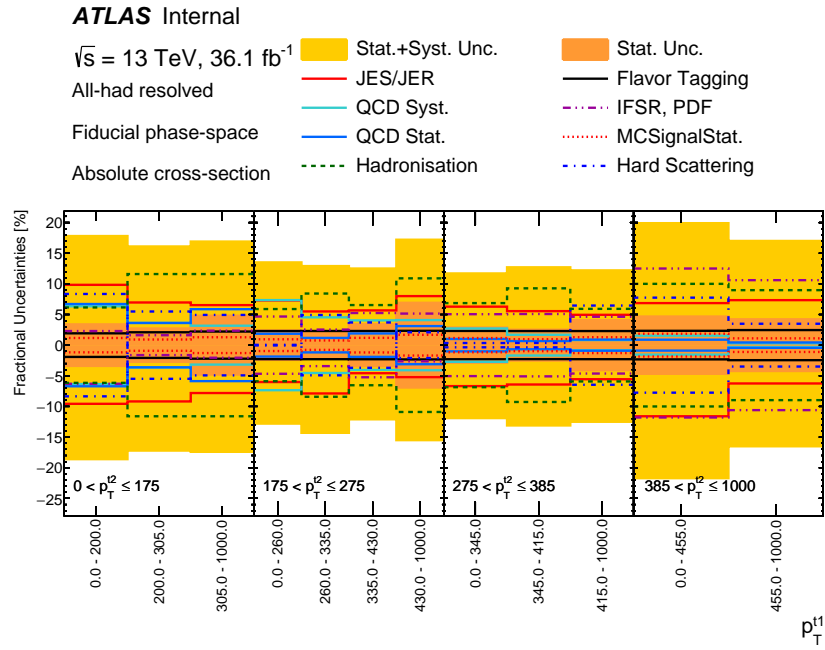**Figure B.5** Uncertainty composition for the $|y^{t\bar{t}}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
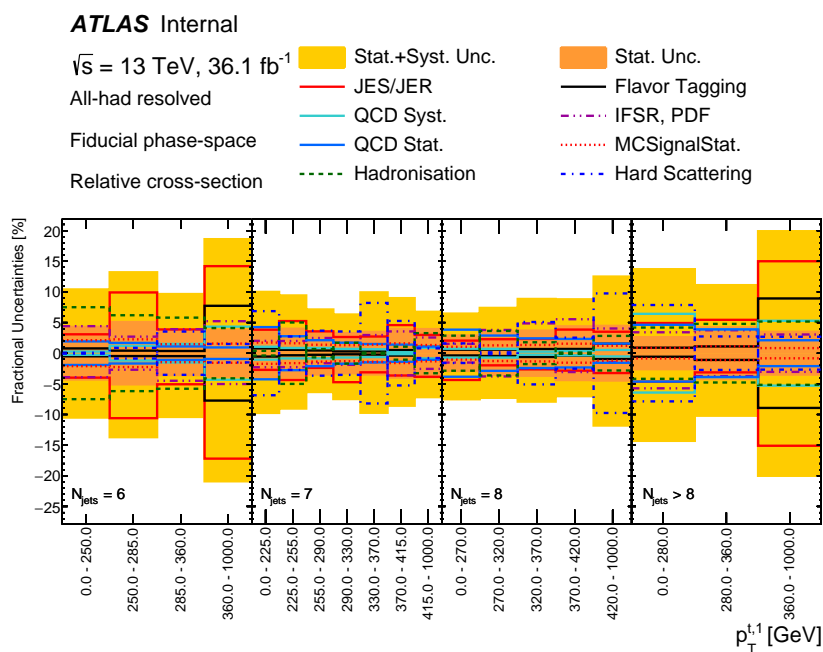
Figure B.6 Uncertainty composition for the $p_{\mathrm{T}}^{t_1}$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
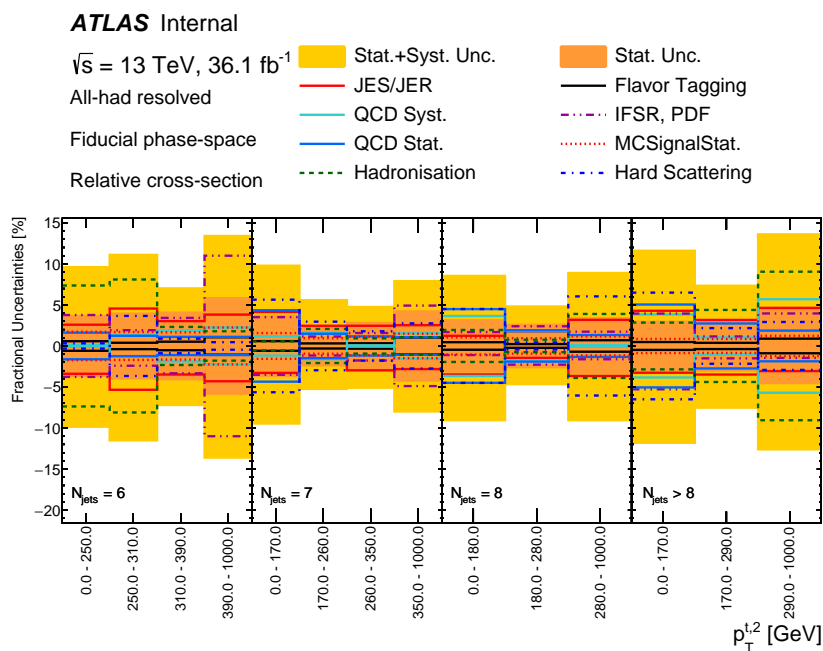
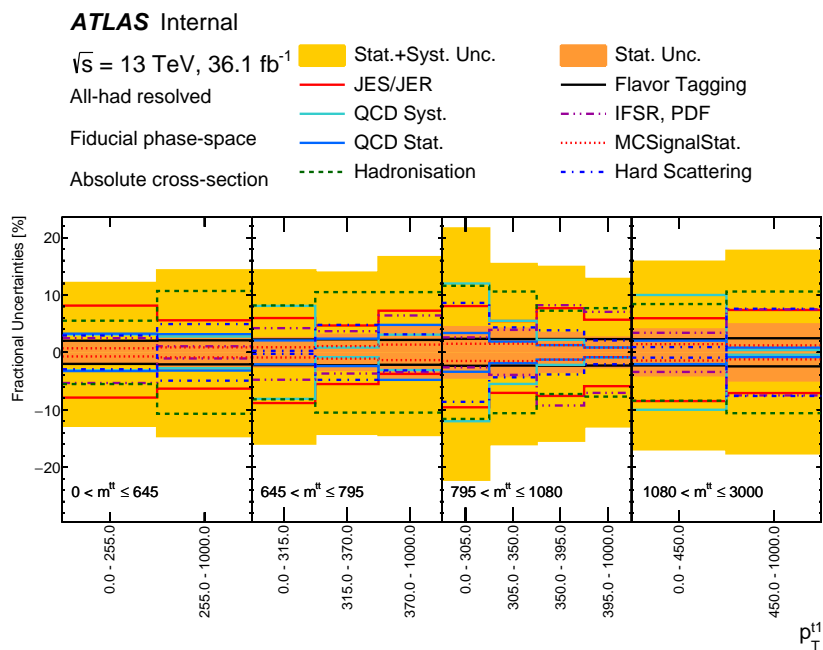**Figure B.7** Uncertainty composition for the $p_{\mathrm{T}}^{t_2}$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
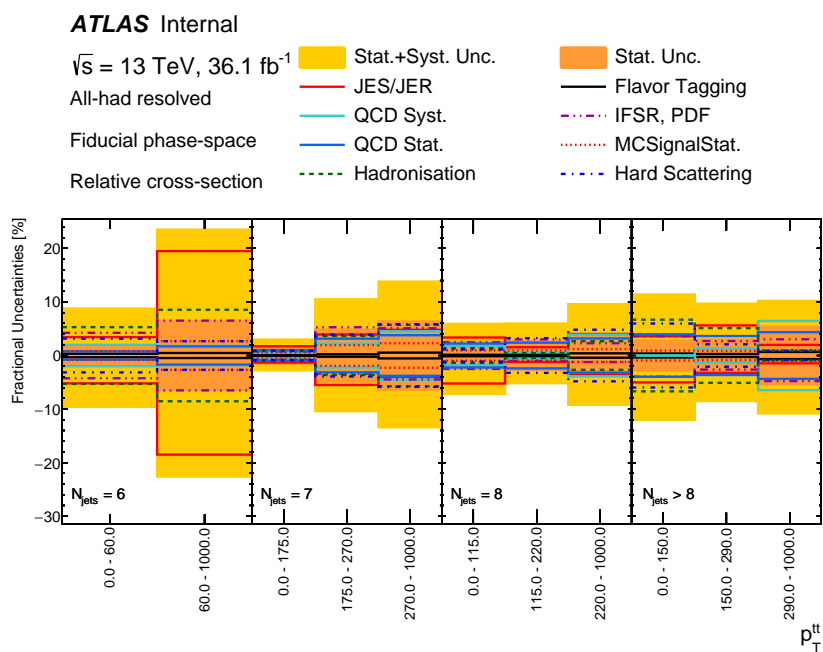
Figure B.8 Uncertainty composition for the $p_{\mathrm{T}}^{t\bar{t}}$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
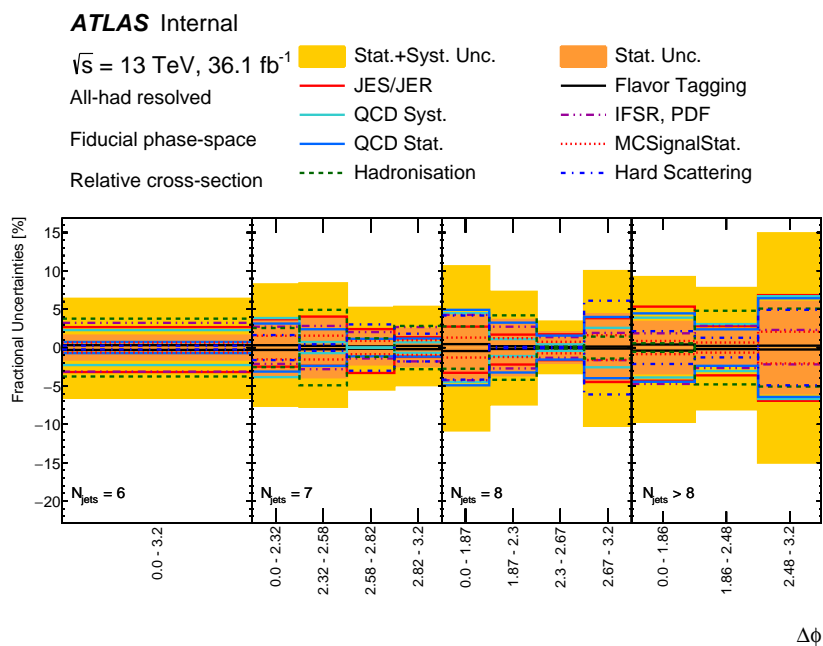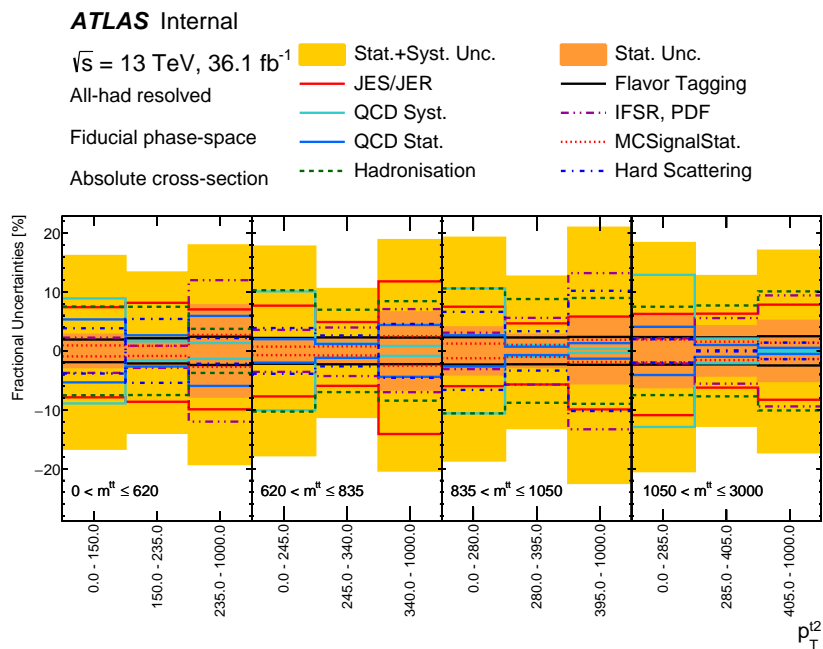

Figure B.9 Uncertainty composition for the $\Delta\phi$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
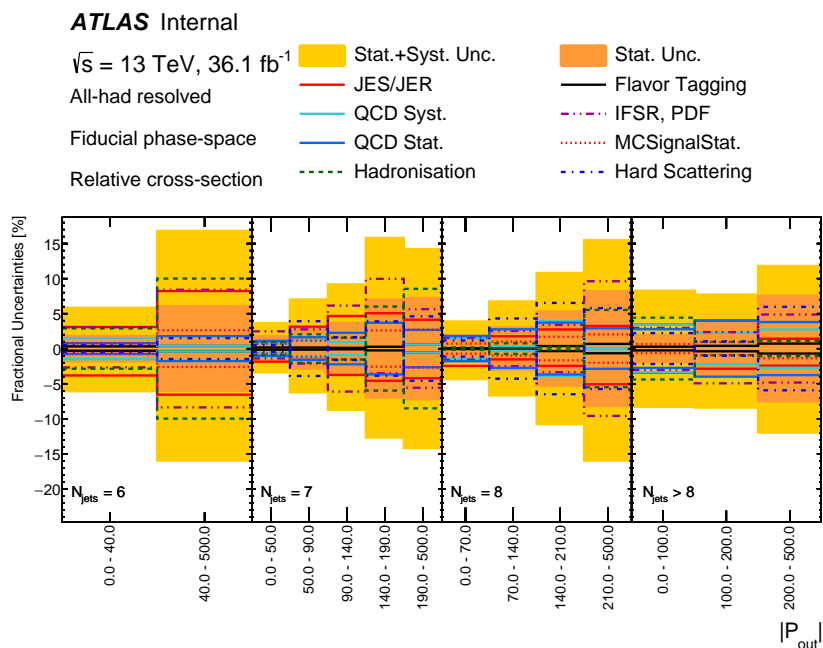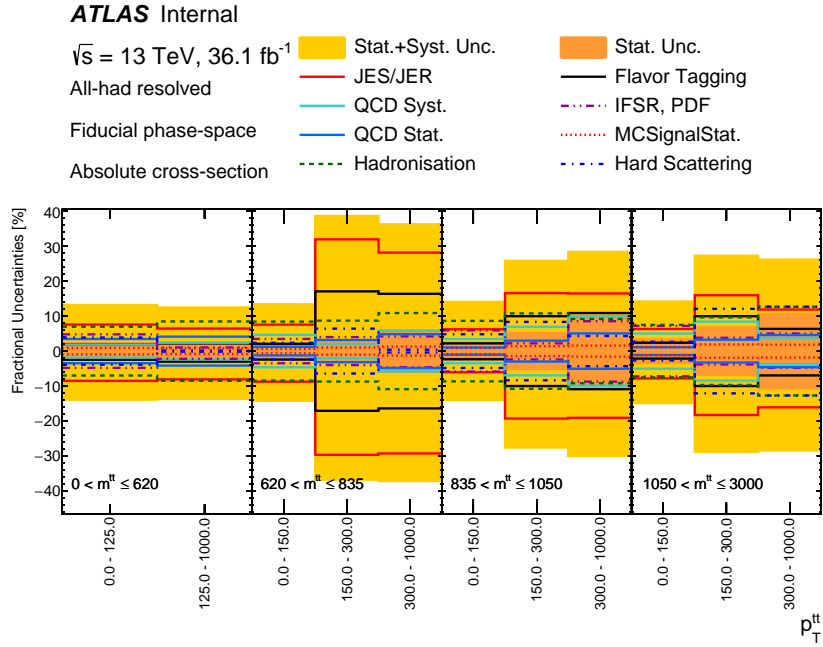
Figure B.10 Uncertainty composition for the $|P_{\text{out}}|$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
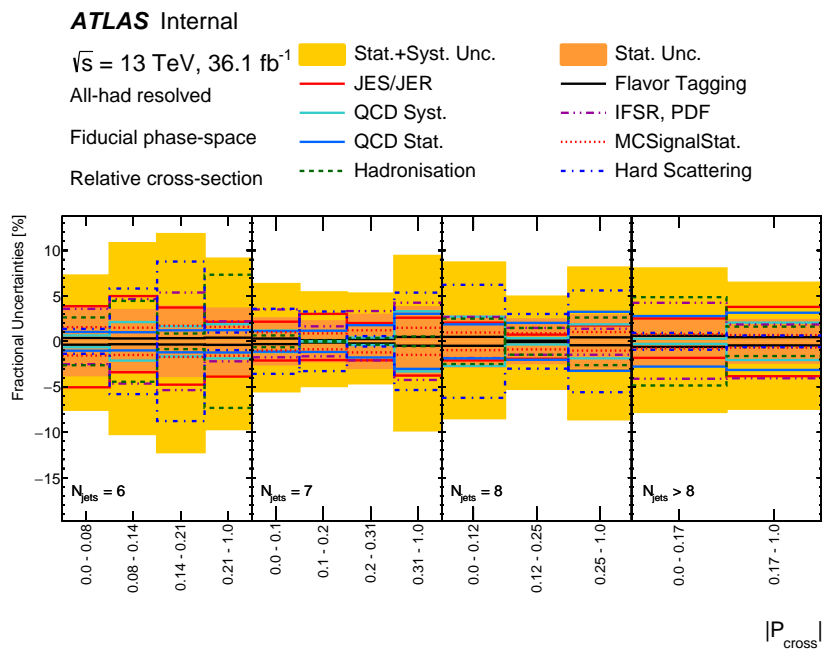
Figure B.11 Uncertainty composition for the $|P_{\text{cross}}|$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
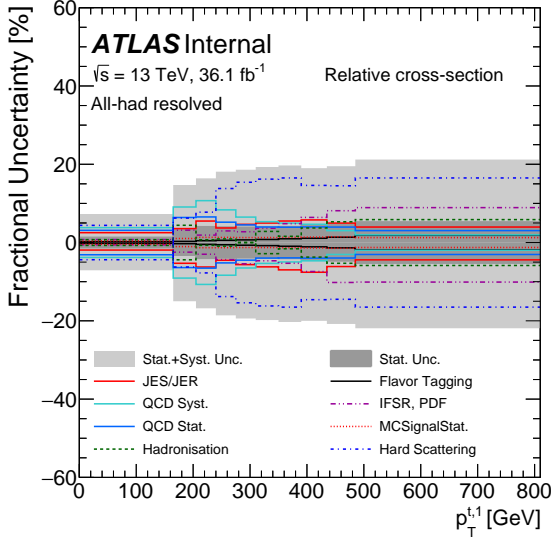
Parton level 1D
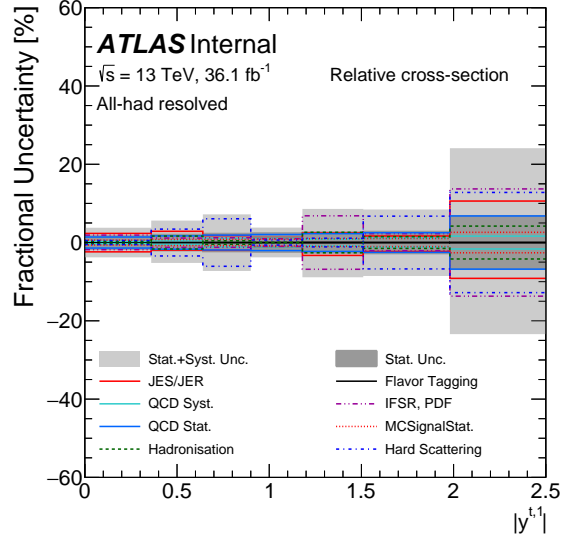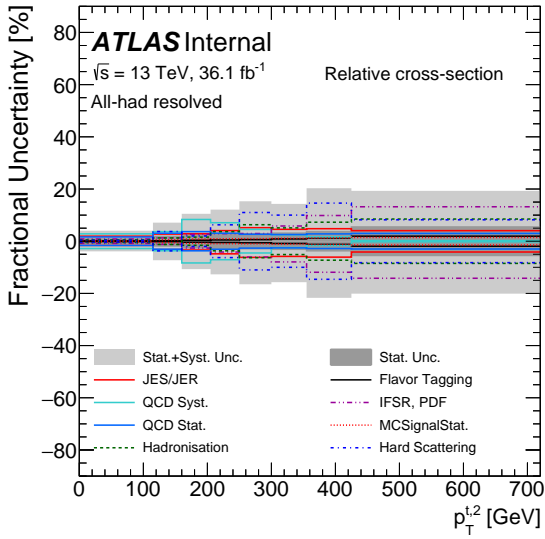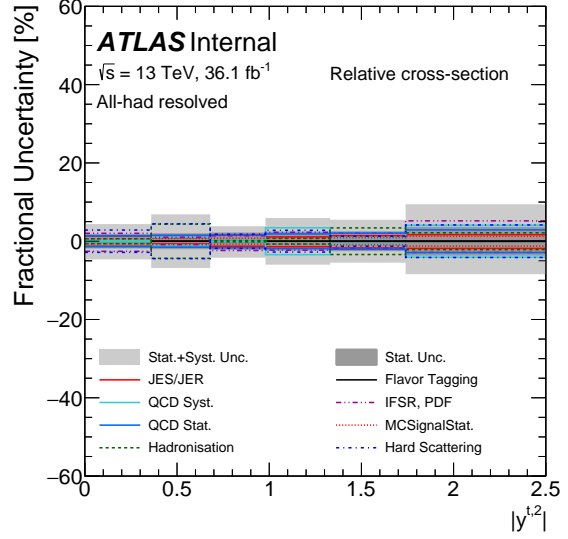
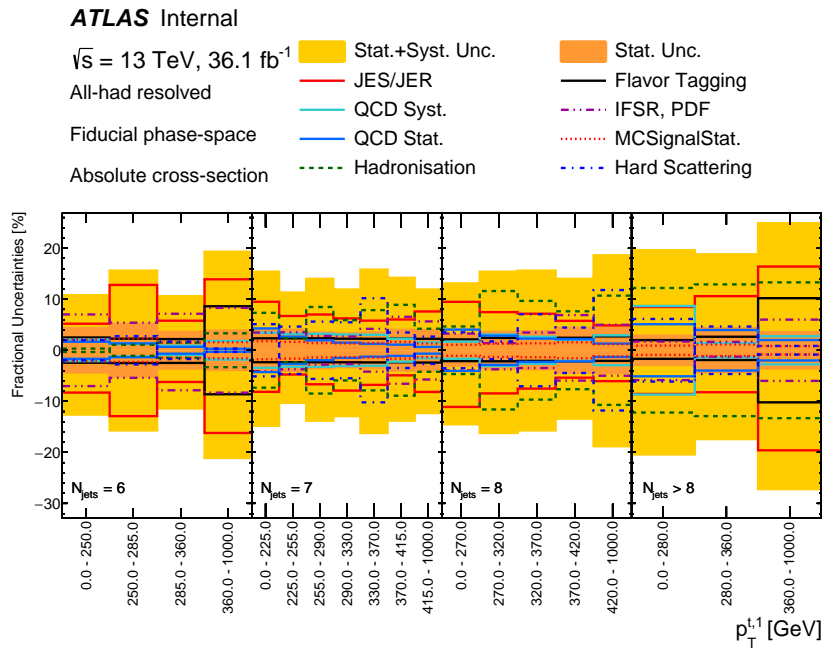**(a)** $p_T^{t_1}$

(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure B.12 Uncertainty composition for the normalised cross section as a function of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
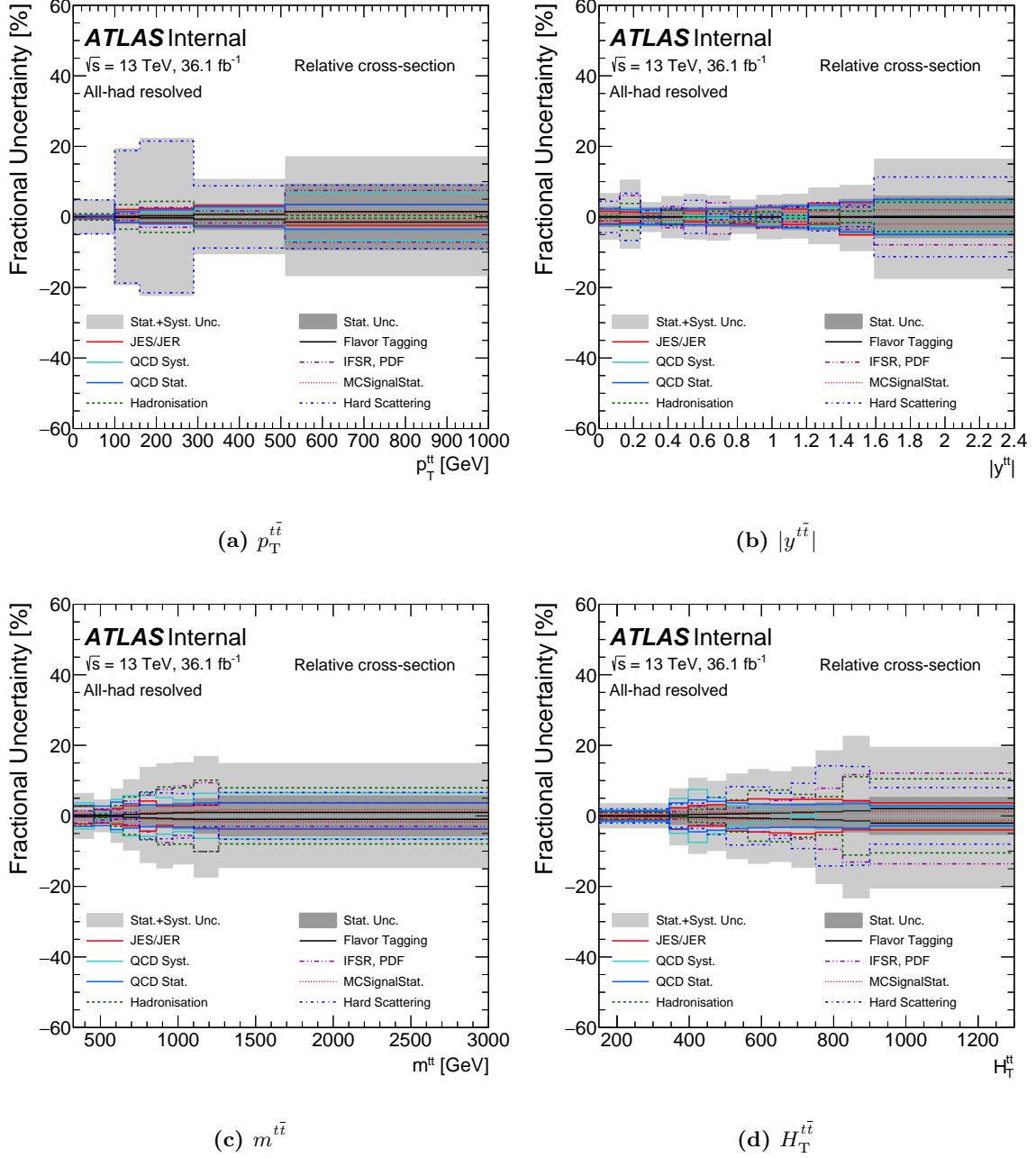
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure B.13 Uncertainty composition for the normalised cross section as a function of (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $\chi^{t\bar{t}}$

(b) $\Delta\phi$
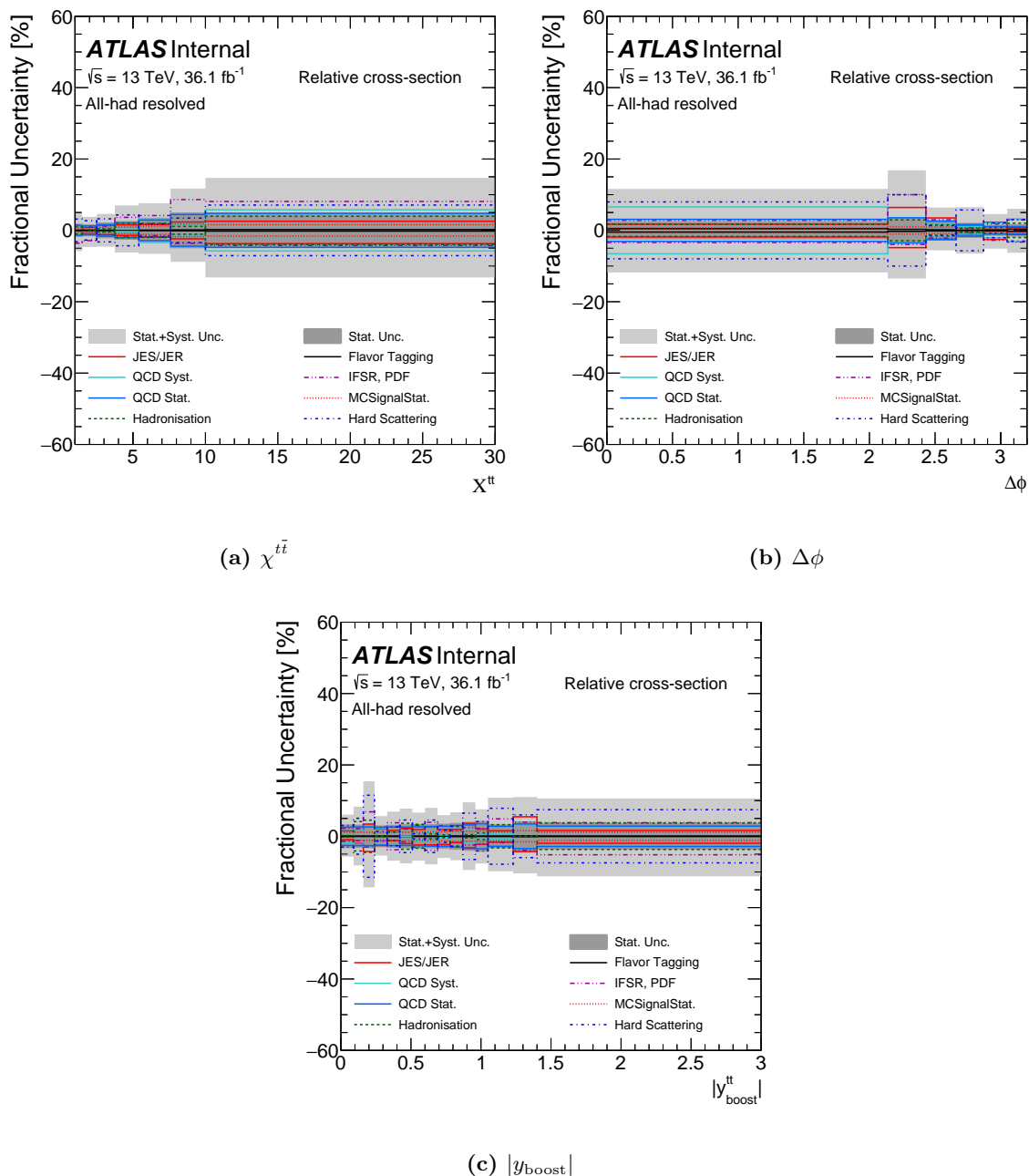
(c) $|y_{\text{boost}}|$

Figure B.14 Uncertainty composition for the normalised cross section as a function of (a) $\chi^{t\bar{t}}$, (b) $\Delta\phi$, and (c) $|y_{\text{boost}}|$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
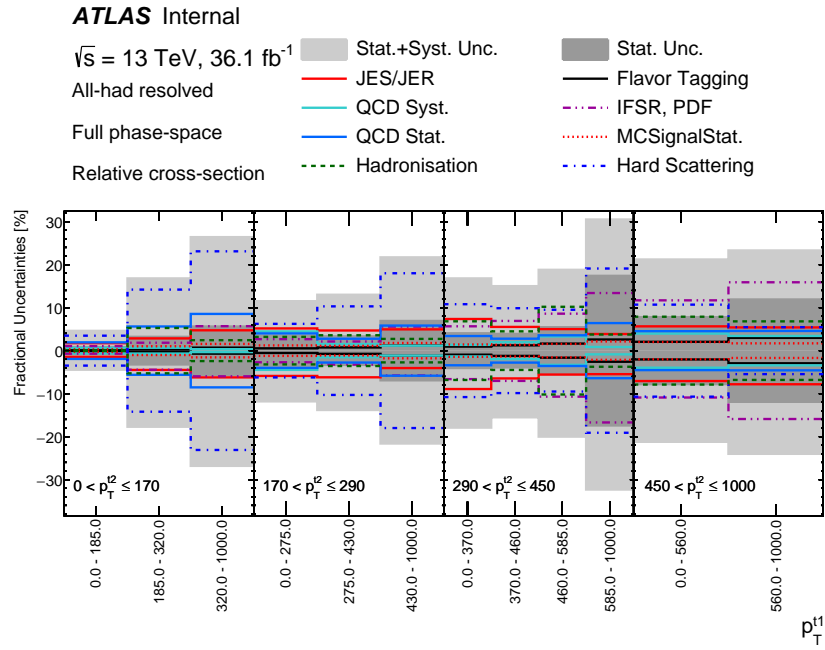
Parton level 2D

Figure B.15 Uncertainty composition for the $p_{\mathrm{T}}^{t_1}$ cross section in bins of $p_{\mathrm{T}}^{t_2}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
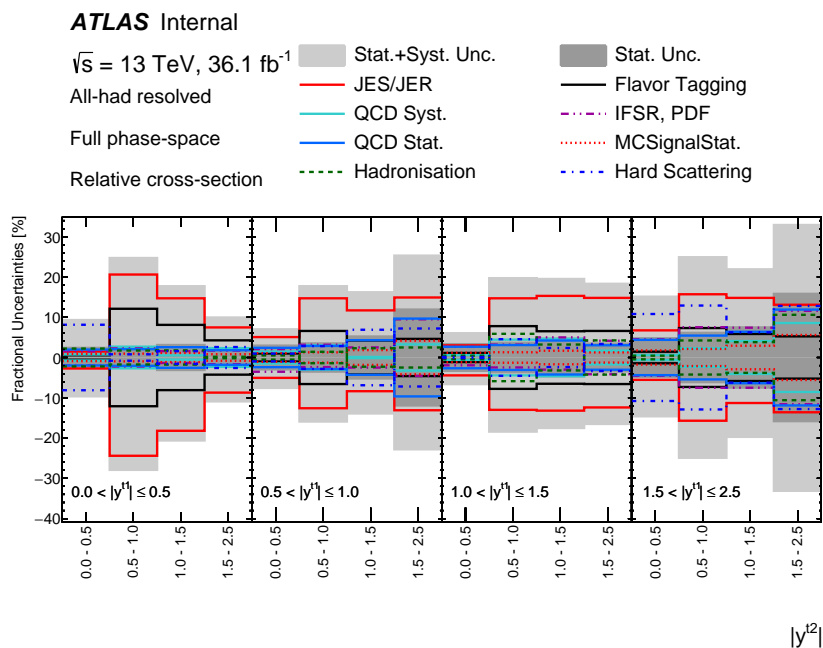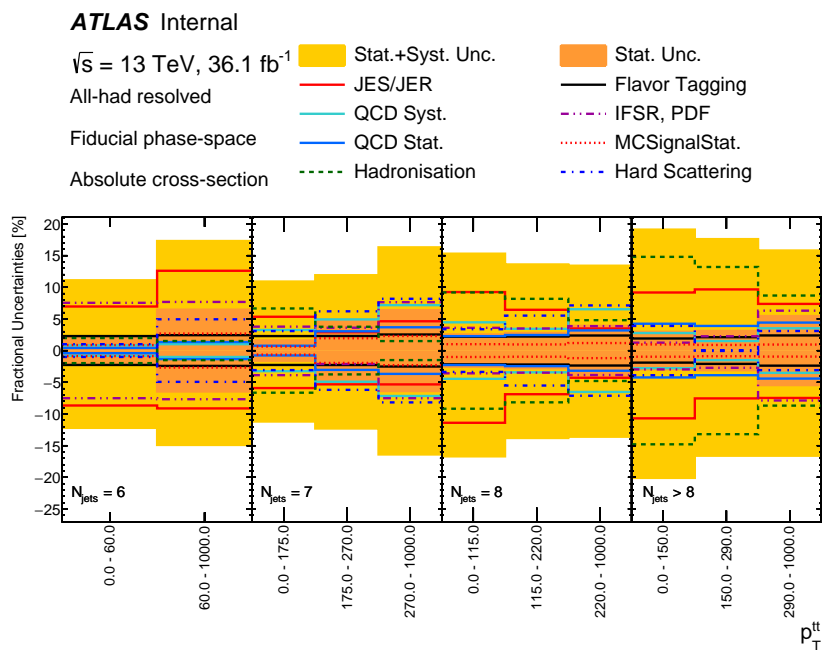
Figure B.16 Uncertainty composition for the $|y^{t_2}|$ cross section in bins of $|y^{t_1}|$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
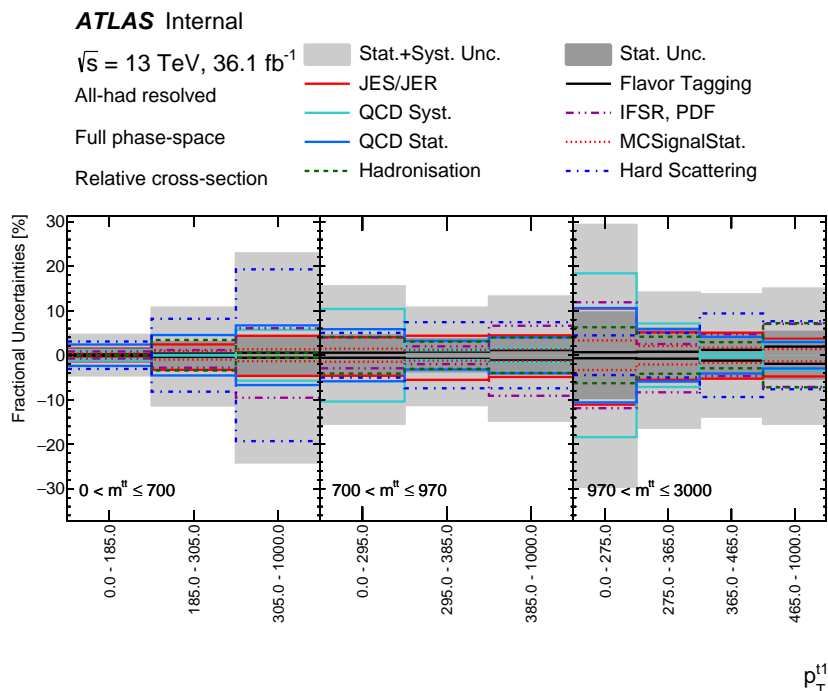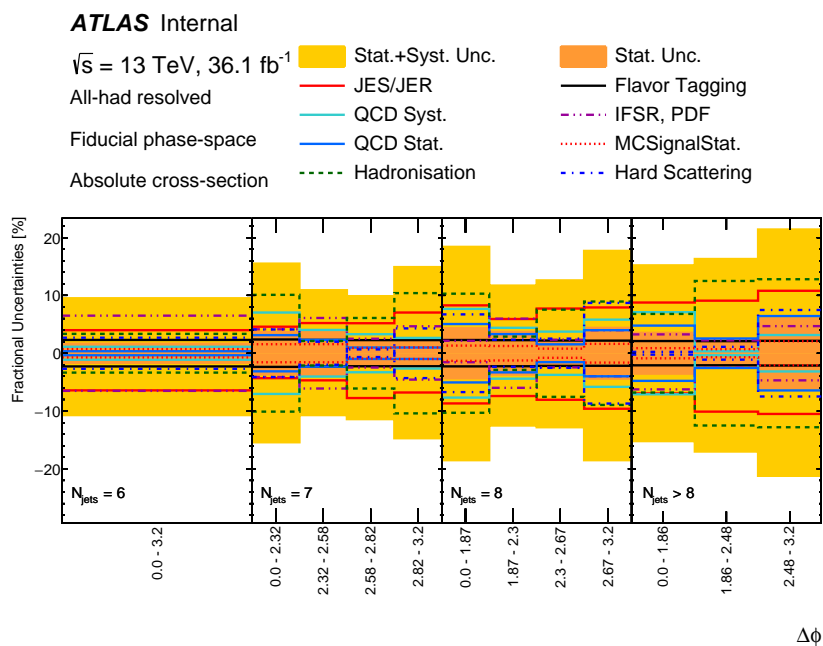

Figure B.17 Uncertainty composition for the $p_{\mathrm{T}}^{t_1}$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
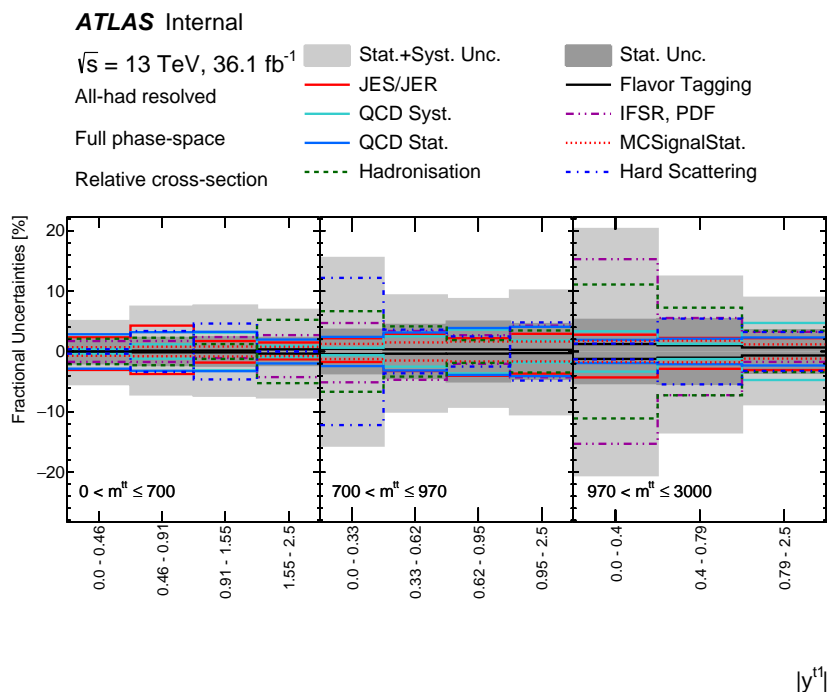
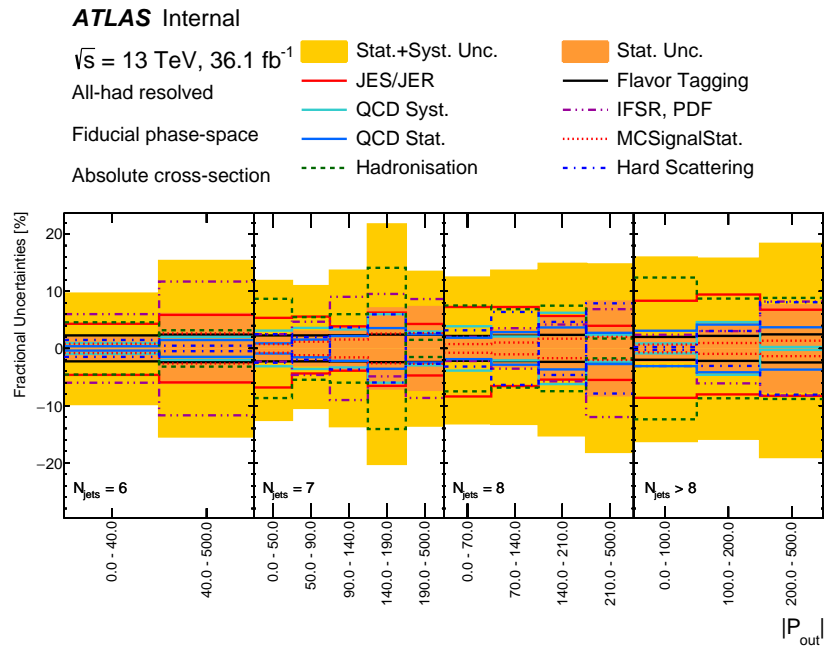**Figure B.18** Uncertainty composition for the $|y^{t_1}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
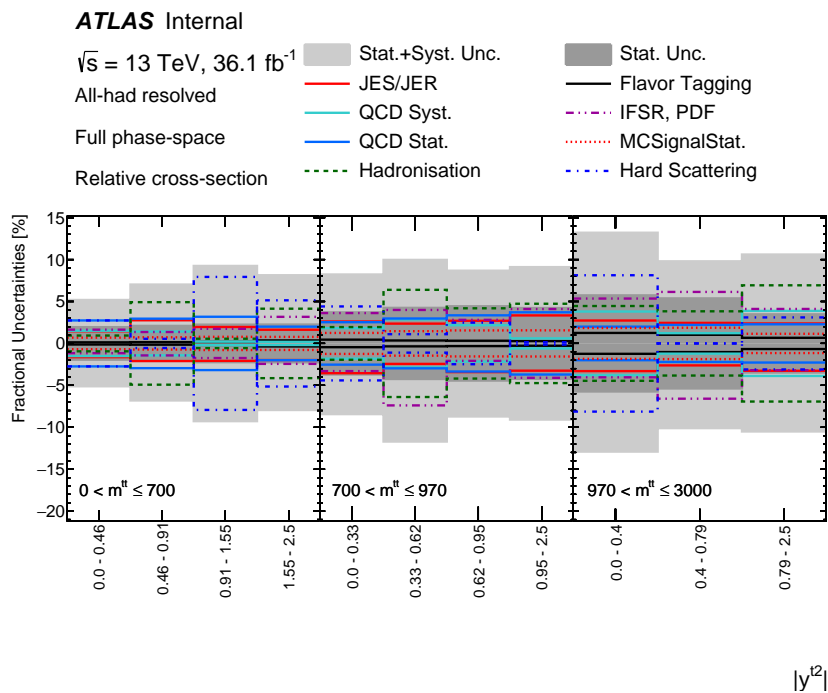
Figure B.19 Uncertainty composition for the $|y^{t_2}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
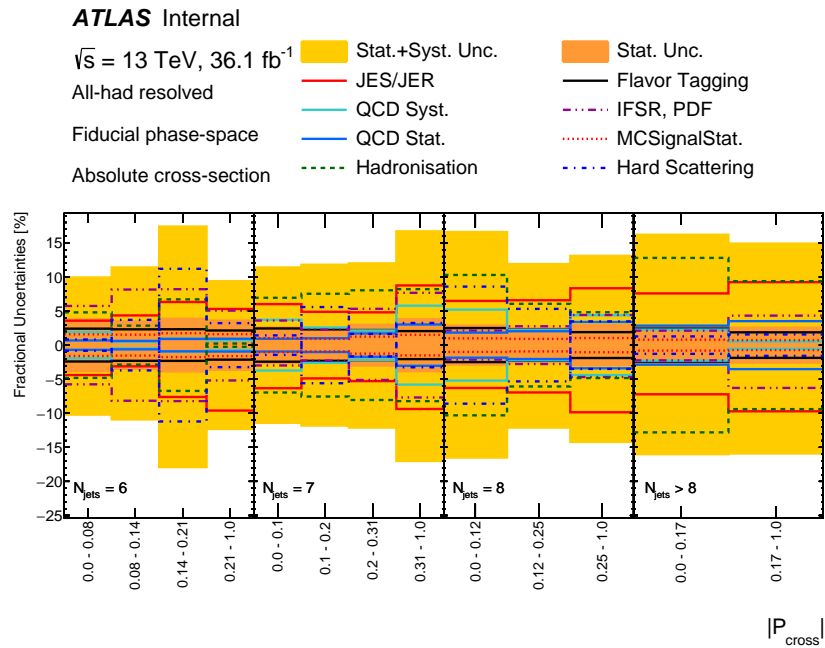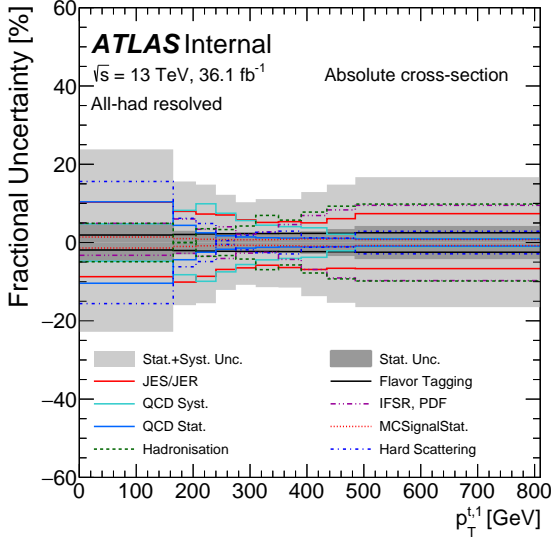
Figure B.20 Uncertainty composition for the $p_{\mathrm{T}}^{t\bar{t}}$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
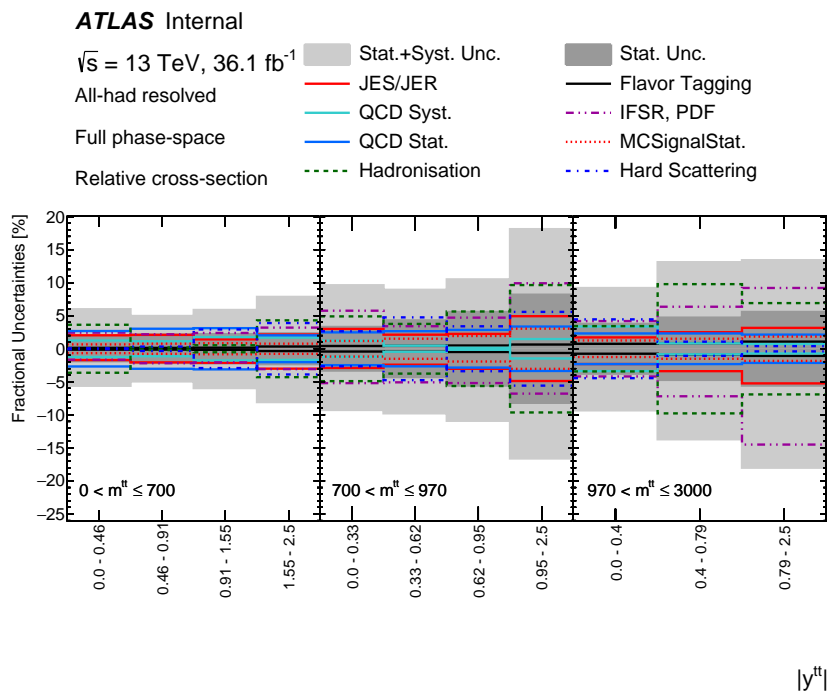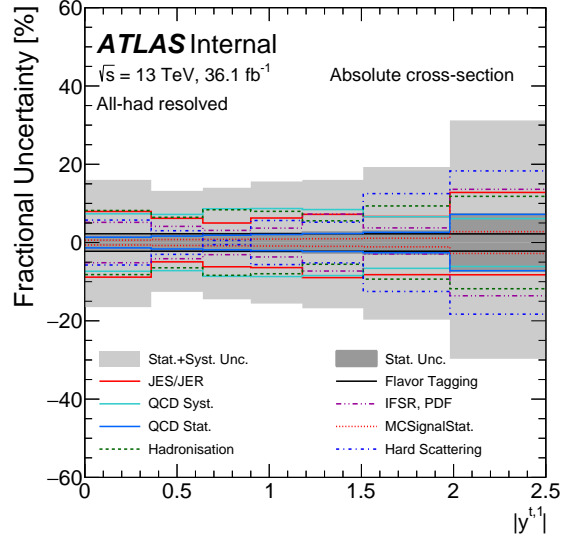

Figure B.21 Uncertainty composition for the $|y^{t\bar{t}}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
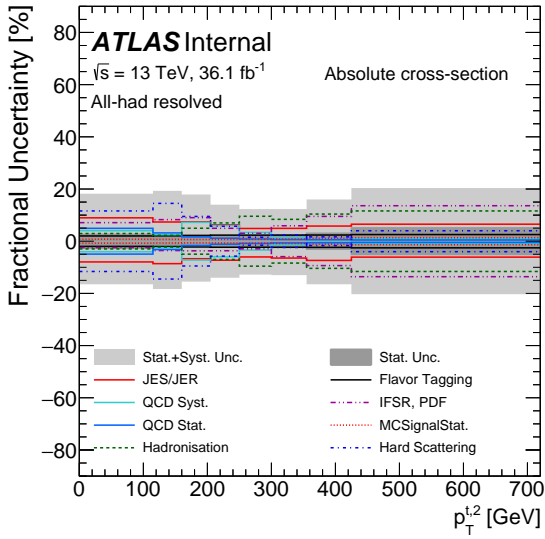
Absolute cross sections

Particle level 1D
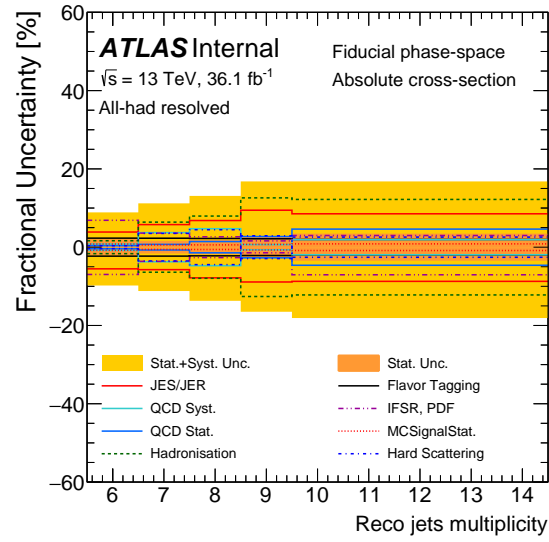
Figure B.22 Uncertainty composition for the absolute cross section as a function of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
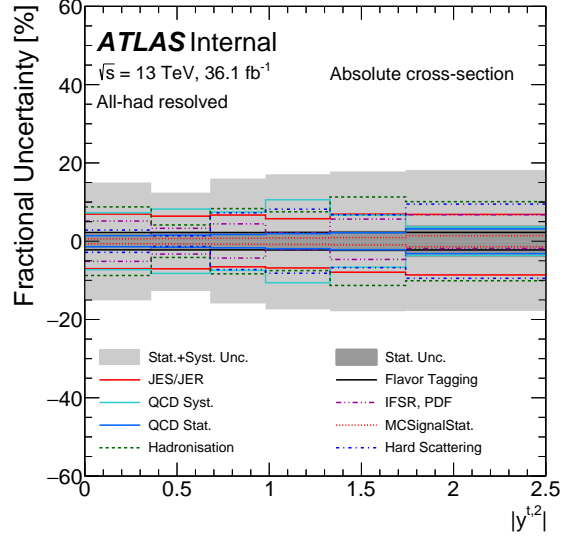
(a) $p_T^{t_1}$

(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure B.23 Uncertainty composition for the absolute cross section as a function of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $p_{\mathrm{T}}^{t\bar{t}}$

(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$
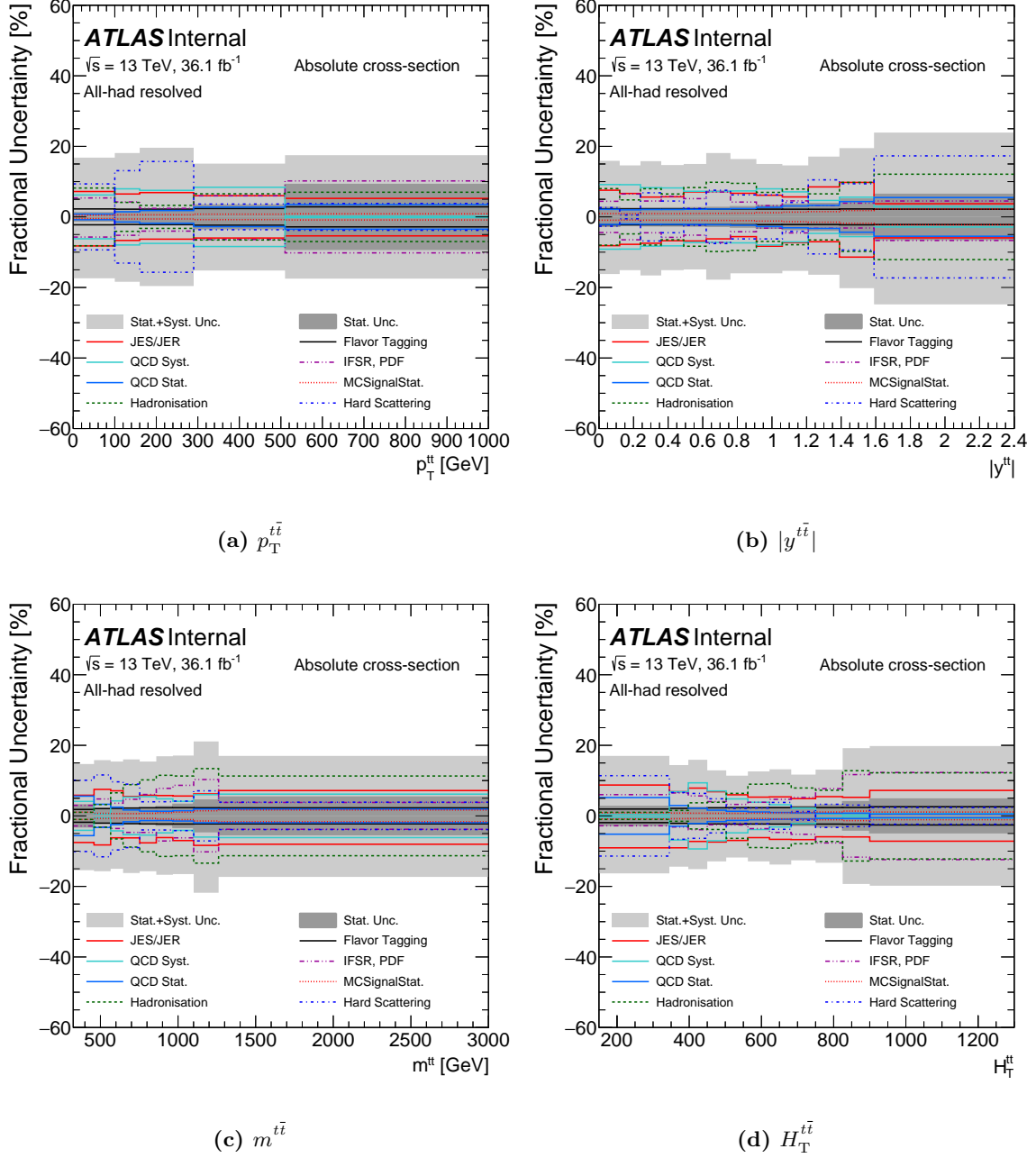
(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure B.24 Uncertainty composition for the absolute cross section as a function of (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
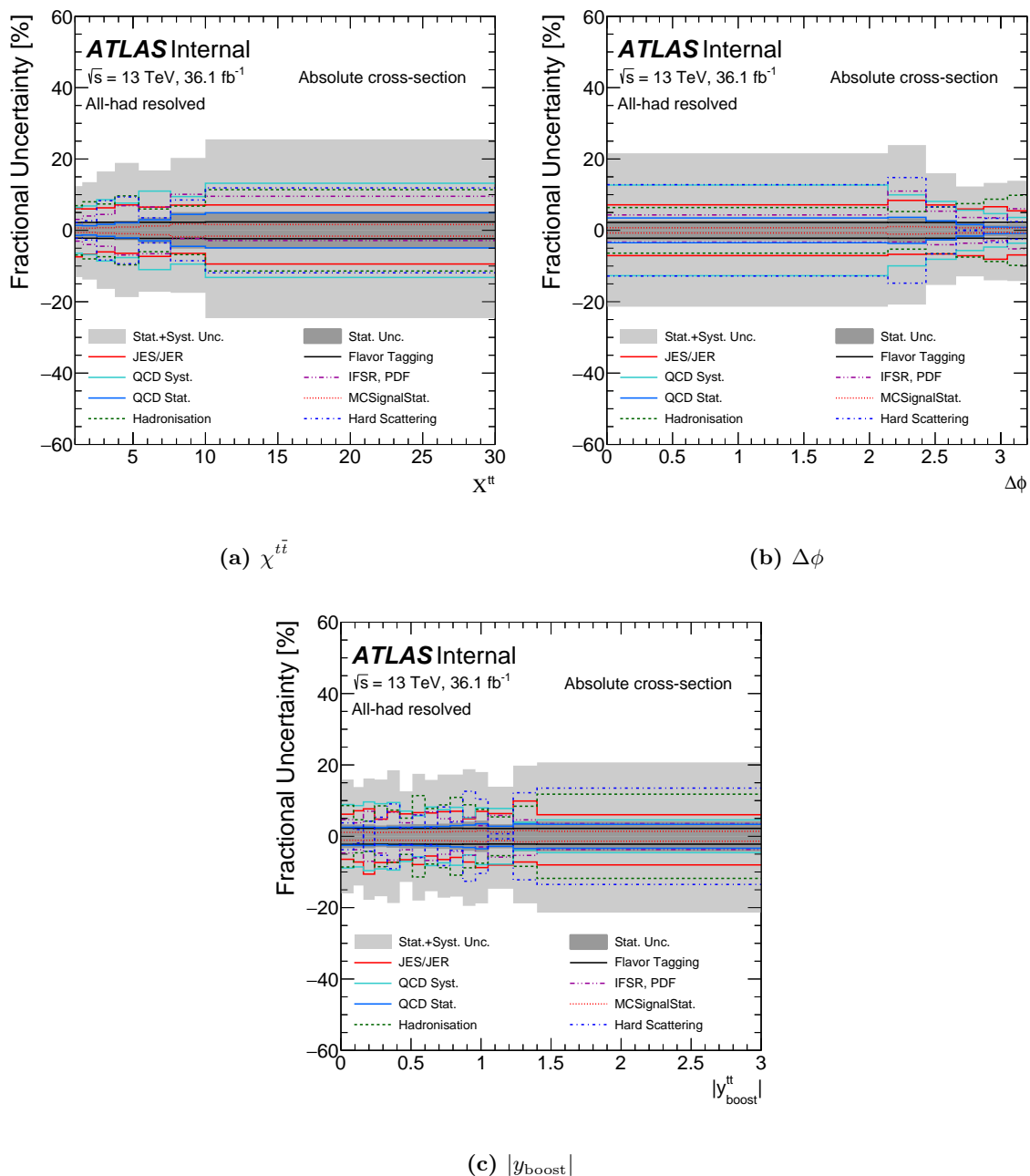
(a) $\chi^{t\bar{t}}$


(b) $Z^{t\bar{t}}$

(c) $\cos\theta^{\star}$

(d) $\Delta\phi$

Figure B.25 Uncertainty composition for the absolute cross section as a function of (a) $\chi^{t\bar{t}}$, (b) $Z^{t\bar{t}}$, (c) $\cos\theta^{\star}$, and (d) $\Delta\phi$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $|y_{\mathrm{boost}}|$

(b) $|P_{\mathrm{out}}|$

(c) $|P_{\mathrm{cross}}|$

Figure B.26 Uncertainty composition for the absolute cross section as a function of (a) $|y_{\mathrm{boost}}|$, (b) $|P_{\mathrm{out}}|$, and (c) $|P_{\mathrm{cross}}|$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

B Uncertainty compositions

**(a)** $R_{Wb}^{\mathrm{leading}}$

(b) $R_{Wb}^{\mathrm{subleading}}$

(c) $R_{Wt}^{\mathrm{leading}}$

(d) $R_{Wt}^{\mathrm{subleading}}$

Figure B.27 Uncertainty composition for the absolute cross section as a function of (a) $R_{Wb}^{\mathrm{leading}}$, (b) $R_{Wb}^{\mathrm{subleading}}$, (c) $R_{Wt}^{\mathrm{leading}}$, and (d) $R_{Wt}^{\mathrm{subleading}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Particle level 2D

Figure B.28 Uncertainty composition for the $p_{\mathrm{T}}^{t_1}$ cross section in bins of $p_{\mathrm{T}}^{t_2}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
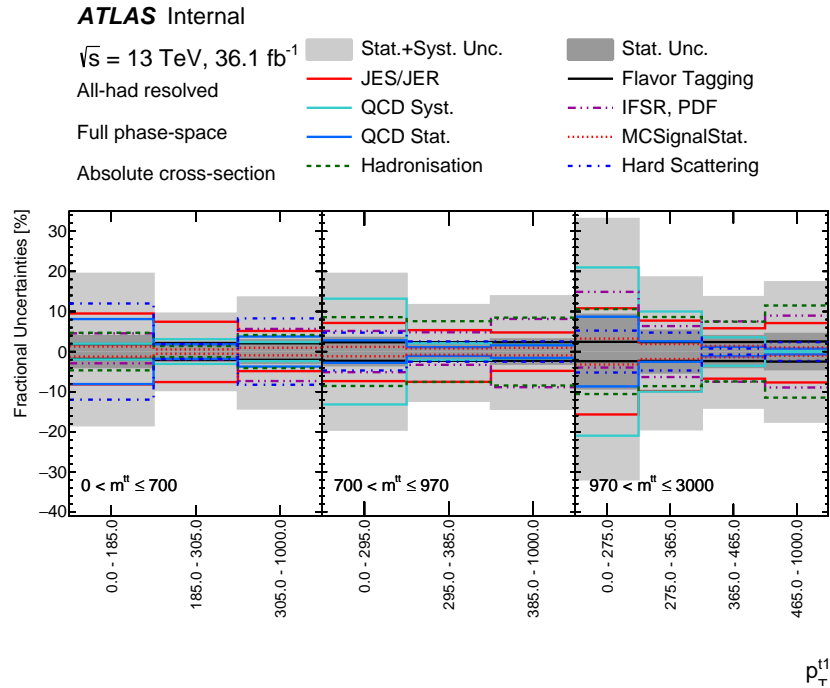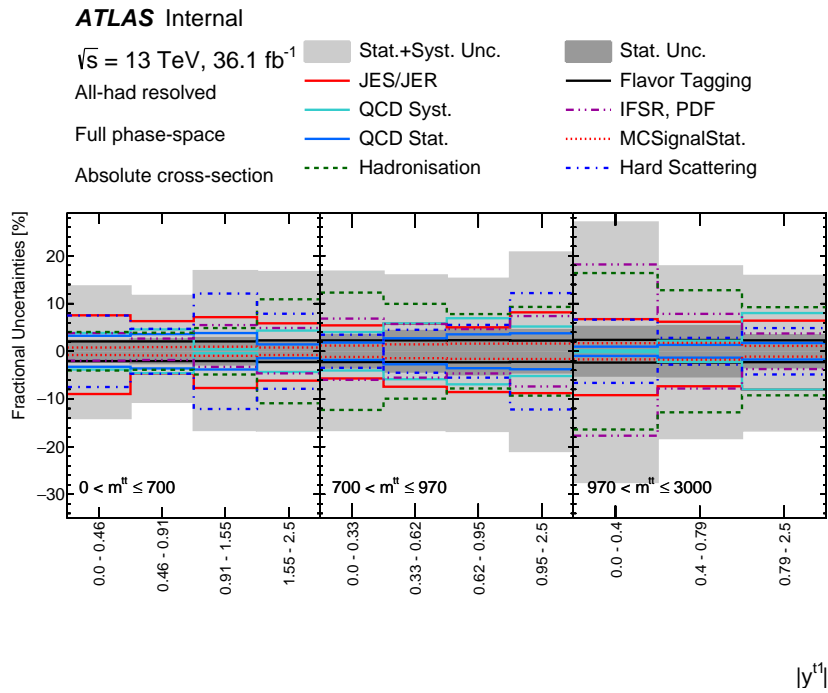
Figure B.29 Uncertainty composition for the $p_\mathrm{T}^{t_1}$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.30 Uncertainty composition for the $p_\mathrm{T}^{t_2}$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
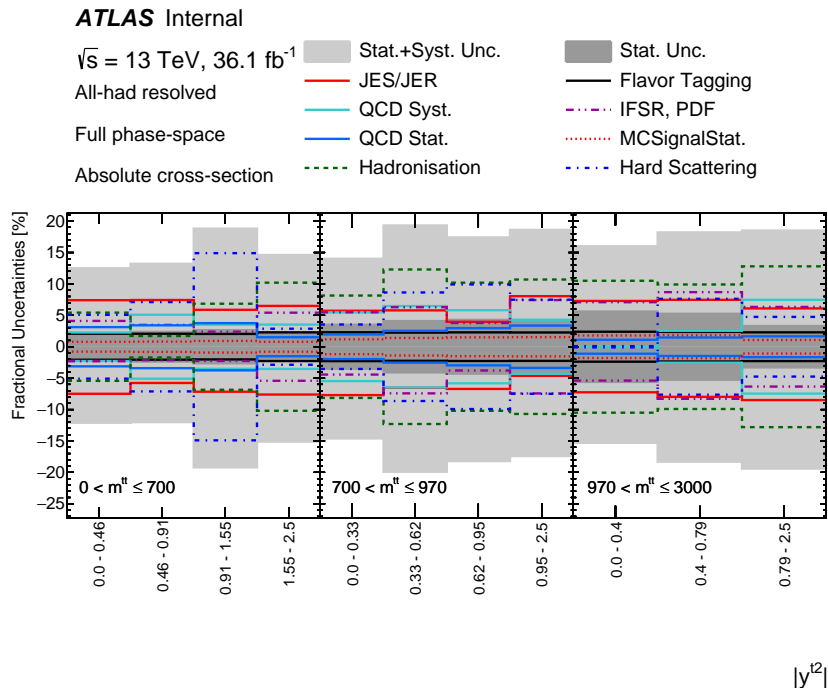
Figure B.31 Uncertainty composition for the $p_{\mathrm{T}}^{t\bar{t}}$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
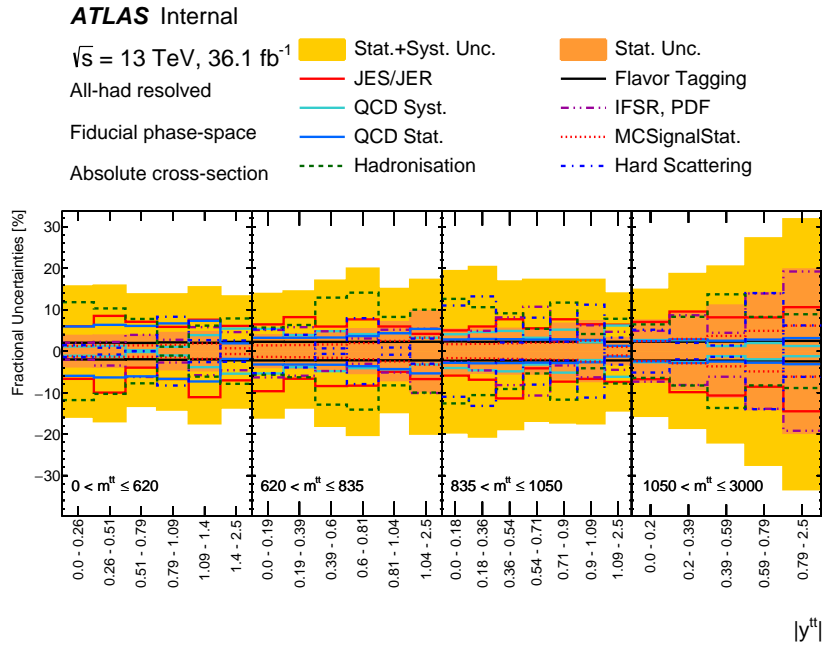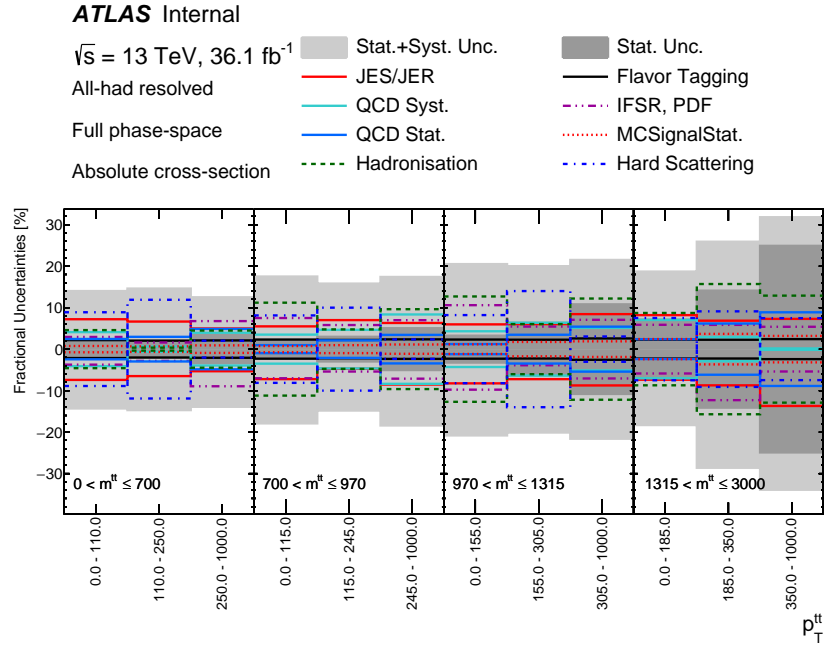

Figure B.32 Uncertainty composition for the $|y^{t\bar{t}}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
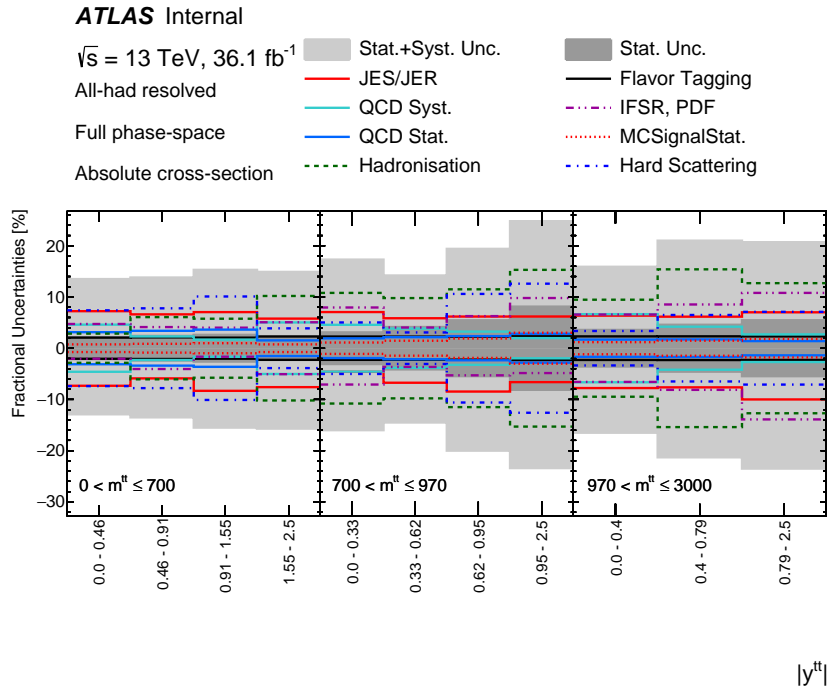
Figure B.33 Uncertainty composition for the $p_T^{t_1}$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.
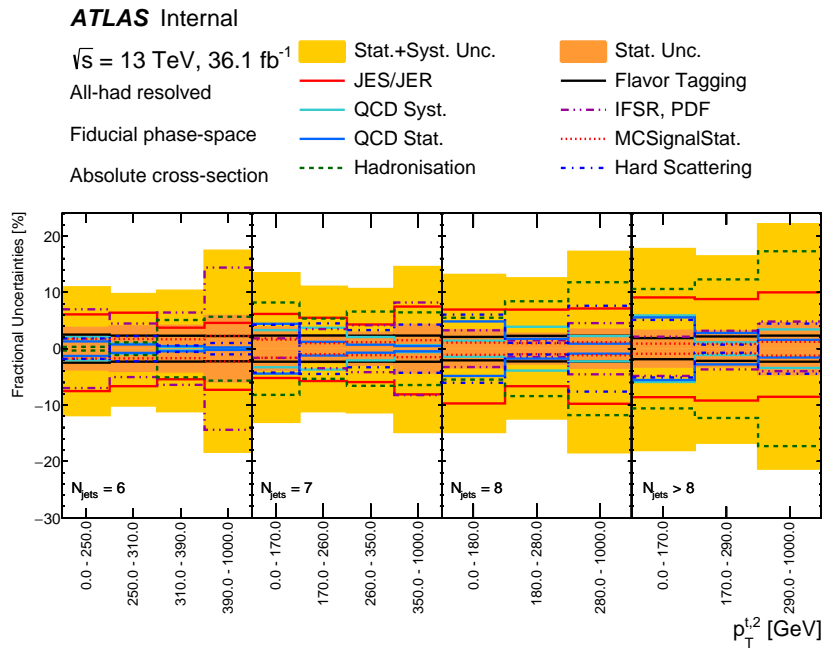
Figure B.34 Uncertainty composition for the $p_T^{t_2}$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.35 Uncertainty composition for the $p_{\mathrm{T}}^{t\bar{t}}$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.36 Uncertainty composition for the $\Delta\phi$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.37 Uncertainty composition for the $|P_{\text{out}}|$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.38 Uncertainty composition for the $|P_{\text{cross}}|$ cross section in bins of N_{jets}, unfolded to particle level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Parton level 1D

(a) $p_T^{t_1}$

(b) $|y^{t_1}|$

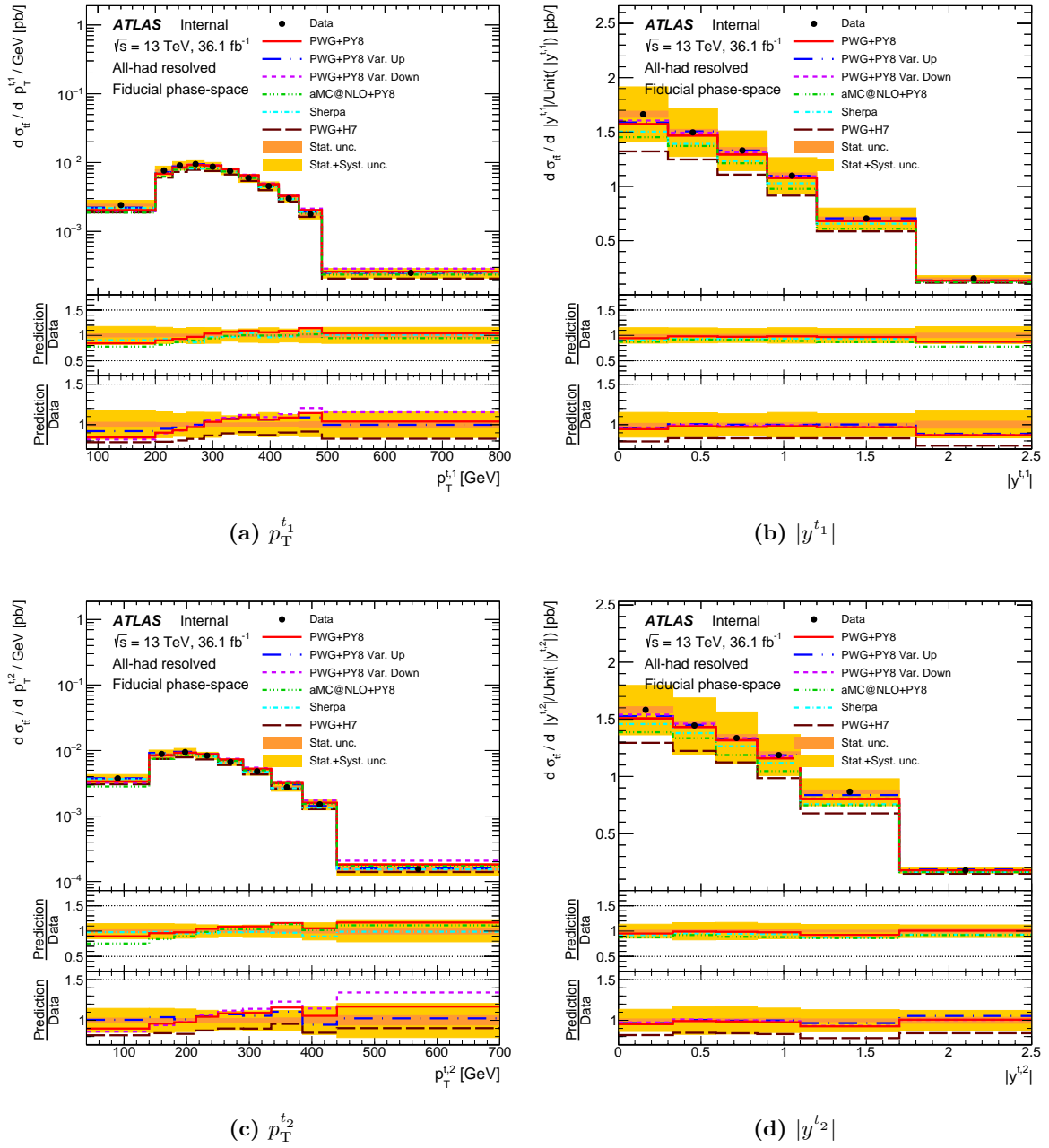**(c)** $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure B.39 Uncertainty composition for the absolute cross section as a function of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $p_\mathrm{T}^{t\bar{t}}$

(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_\mathrm{T}^{t\bar{t}}$

Figure B.40 Uncertainty composition for the absolute cross section as a function of (a) $p_\mathrm{T}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_\mathrm{T}^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

(a) $\chi^{t\bar{t}}$

(b) $\Delta\phi$

(c) $|y_{\mathrm{boost}}|$

Figure B.41 Uncertainty composition for the absolute cross section as a function of (a) $\chi^{t\bar{t}}$, (b) $\Delta\phi$, and (c) $|y_{\mathrm{boost}}|$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
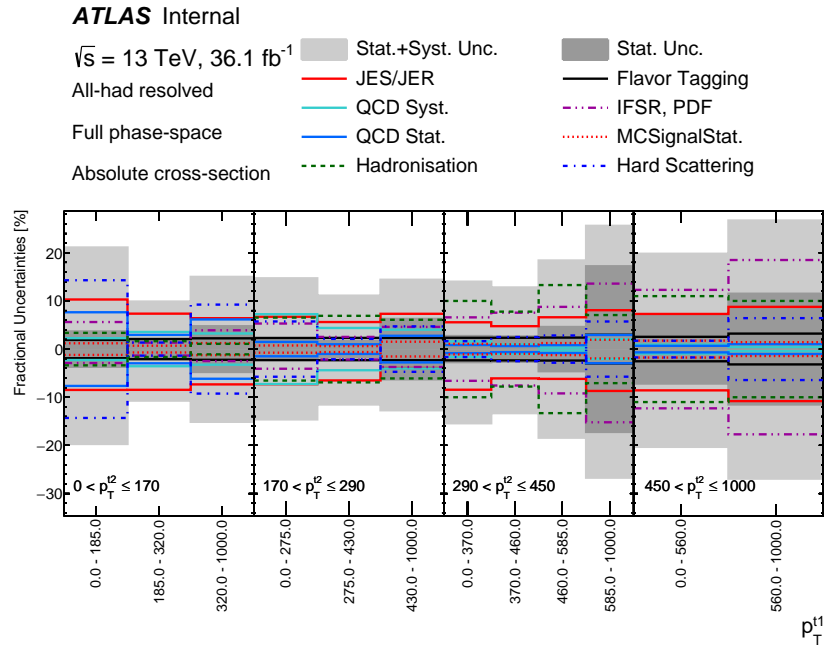
Parton level 2D

Figure B.42 Uncertainty composition for the $p_T^{t_1}$ cross section in bins of $p_T^{t_2}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.
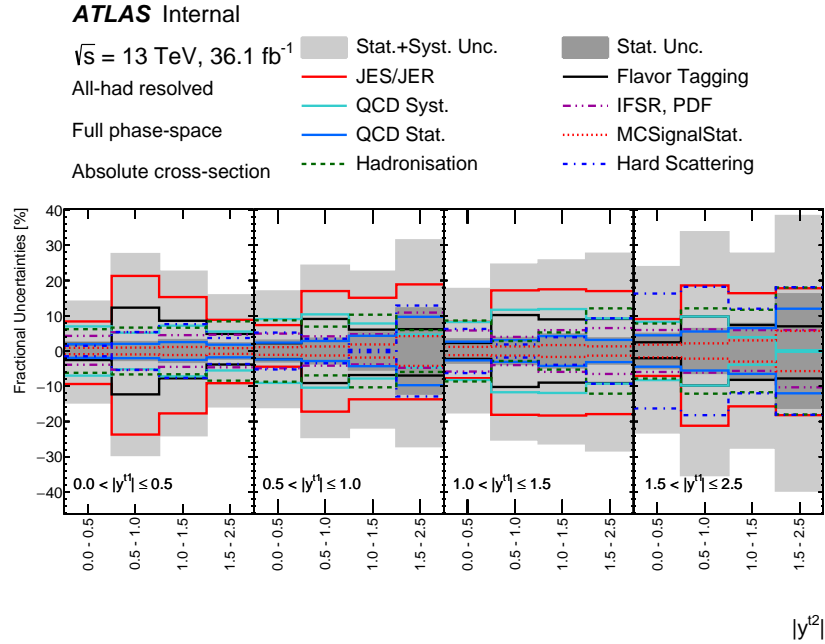
Figure B.43 Uncertainty composition for the $|y^{t_2}|$ cross section in bins of $|y^{t_1}|$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.44 Uncertainty composition for the $p_T^{t_1}$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.45 Uncertainty composition for the $|y^{t_1}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.46 Uncertainty composition for the $|y^{t_2}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.47 Uncertainty composition for the $p_{\mathrm{T}}^{t\bar{t}}$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Figure B.48 Uncertainty composition for the $|y^{t\bar{t}}|$ cross section in bins of $m^{t\bar{t}}$, unfolded to parton level. The lighter shaded area indicates the total statistical and systematic uncertainty.

Appendix C

Absolute differential cross sections

Particle level 1D differential cross sections
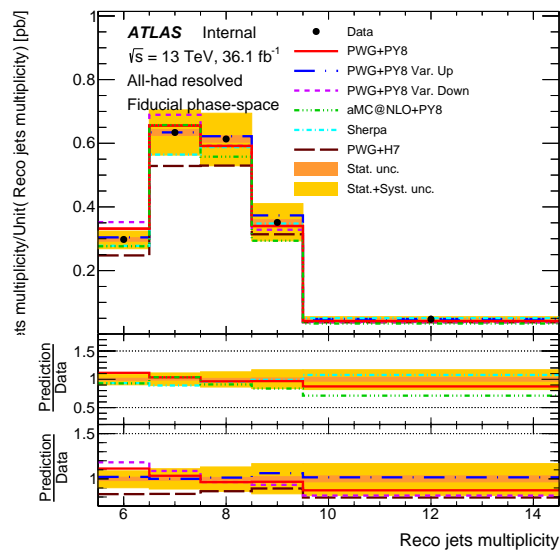
Figure C.1 Absolute differential cross section as a function of N_{jets}, unfolded to particle level. The bottom panels show the ratios of predictions from simulations to the data.

(a) $p_{\mathrm{T}}^{t_1}$
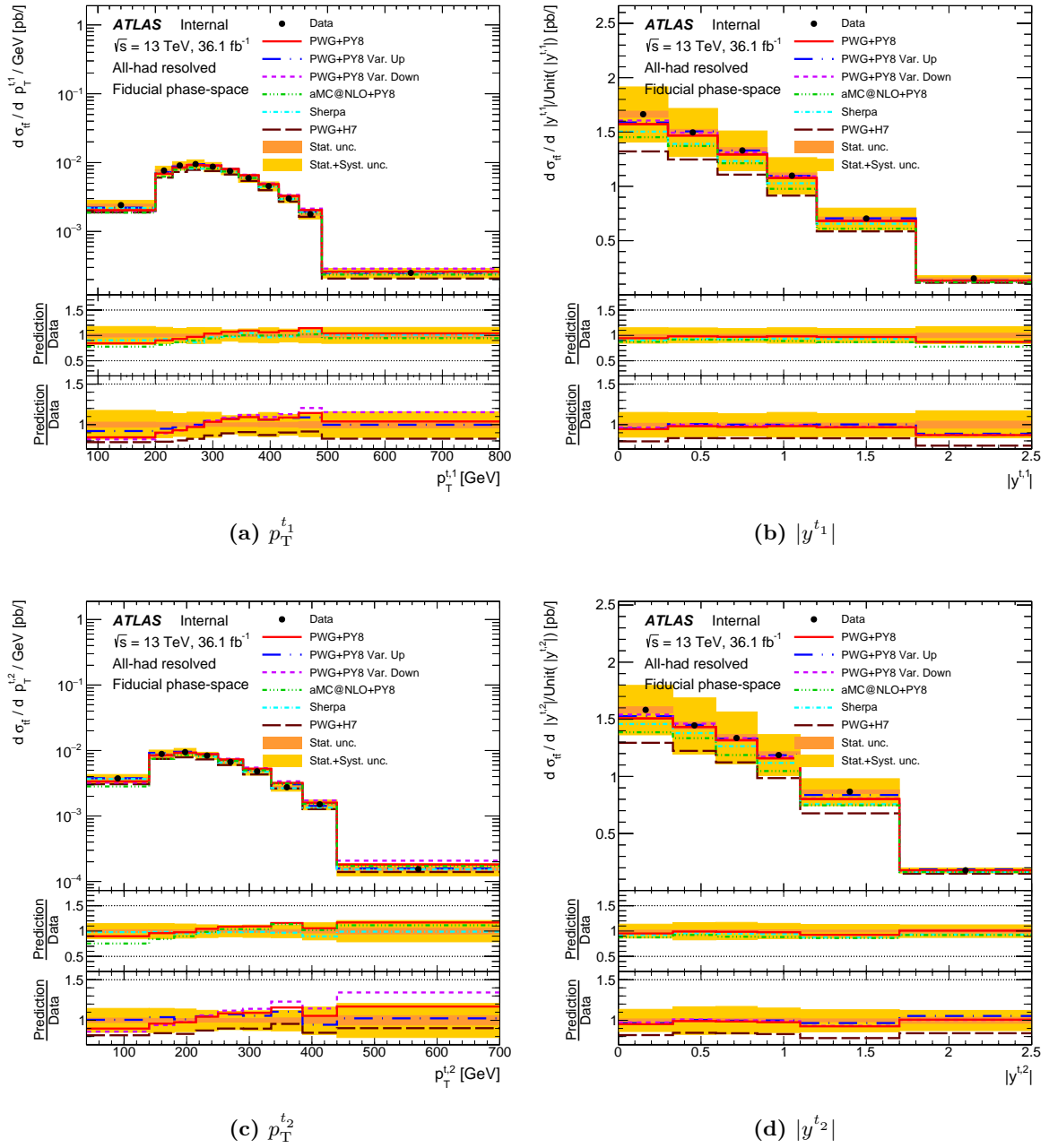
(b) $|y^{t_1}|$

(c) $p_{\mathrm{T}}^{t_2}$

(d) $|y^{t_2}|$

Figure C.2 Absolute differential cross sections as functions of (a) $p_{\mathrm{T}}^{t_1}$, (b) $|y^{t_1}|$, (c) $p_{\mathrm{T}}^{t_2}$, and (d) $|y^{t_2}|$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $p_T^{t_1}$

(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure C.3 Absolute differential cross sections as functions of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
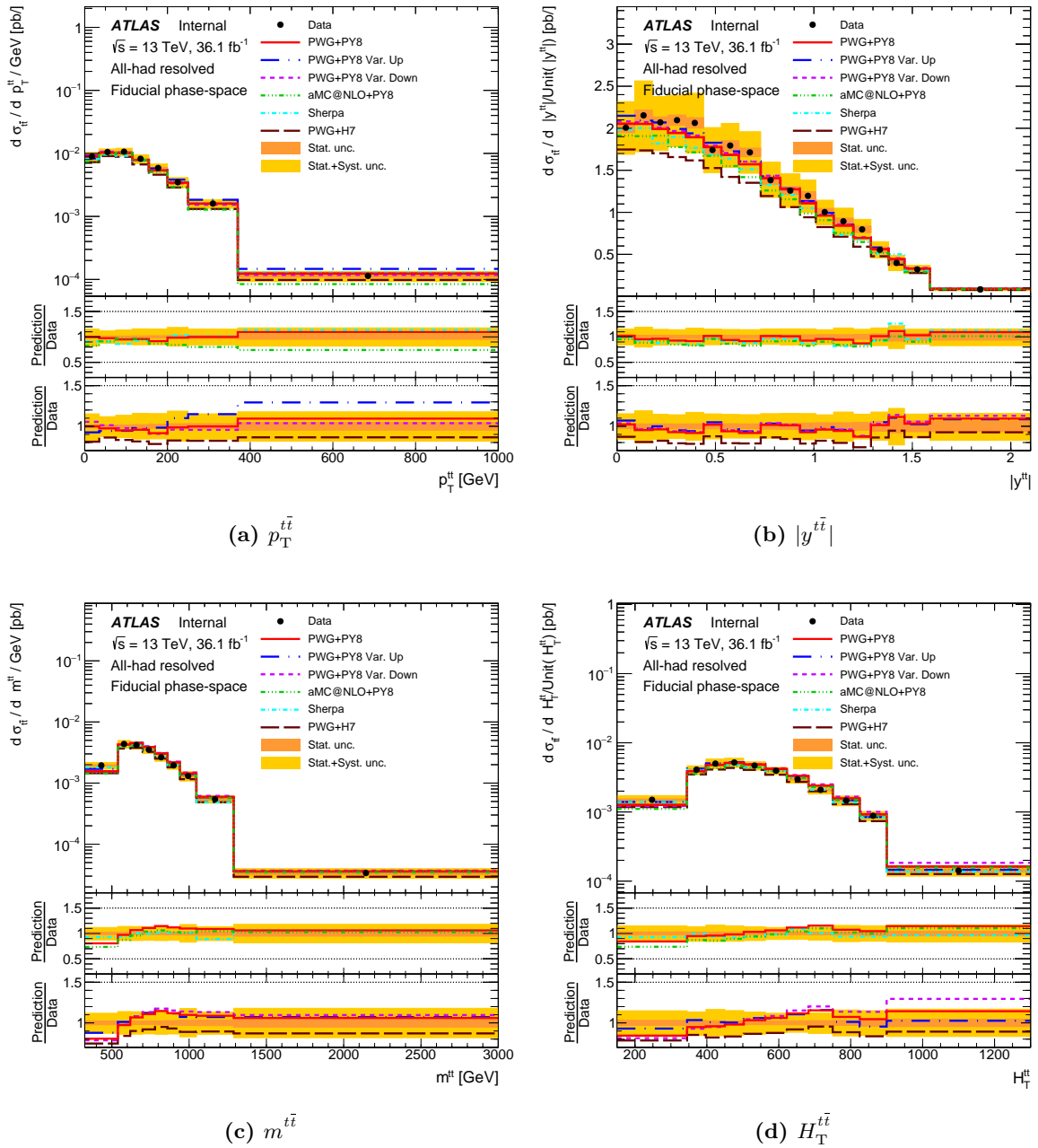
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure C.4 Absolute differential cross sections as functions of (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $\chi^{t\bar{t}}$

(b) $Z^{t\bar{t}}$

(c) $\cos\theta^*$

(d) $\Delta\phi$
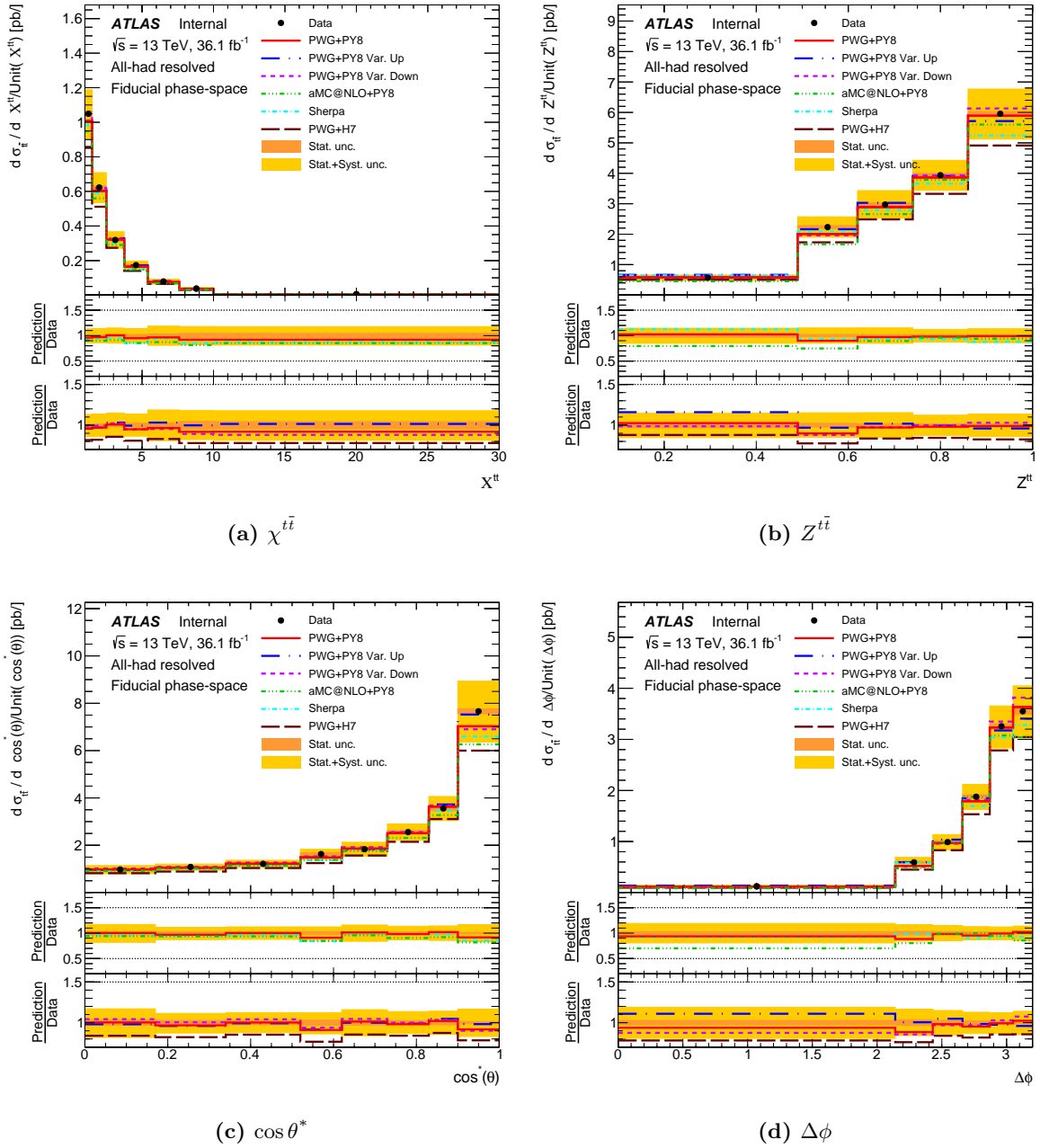
Figure C.5 Absolute differential cross sections as functions of (a) $\chi^{t\bar{t}}$, (b) $Z^{t\bar{t}}$, (c) $\cos\theta^*$, and (d) $\Delta\phi$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $|y_{\mathrm{boost}}|$
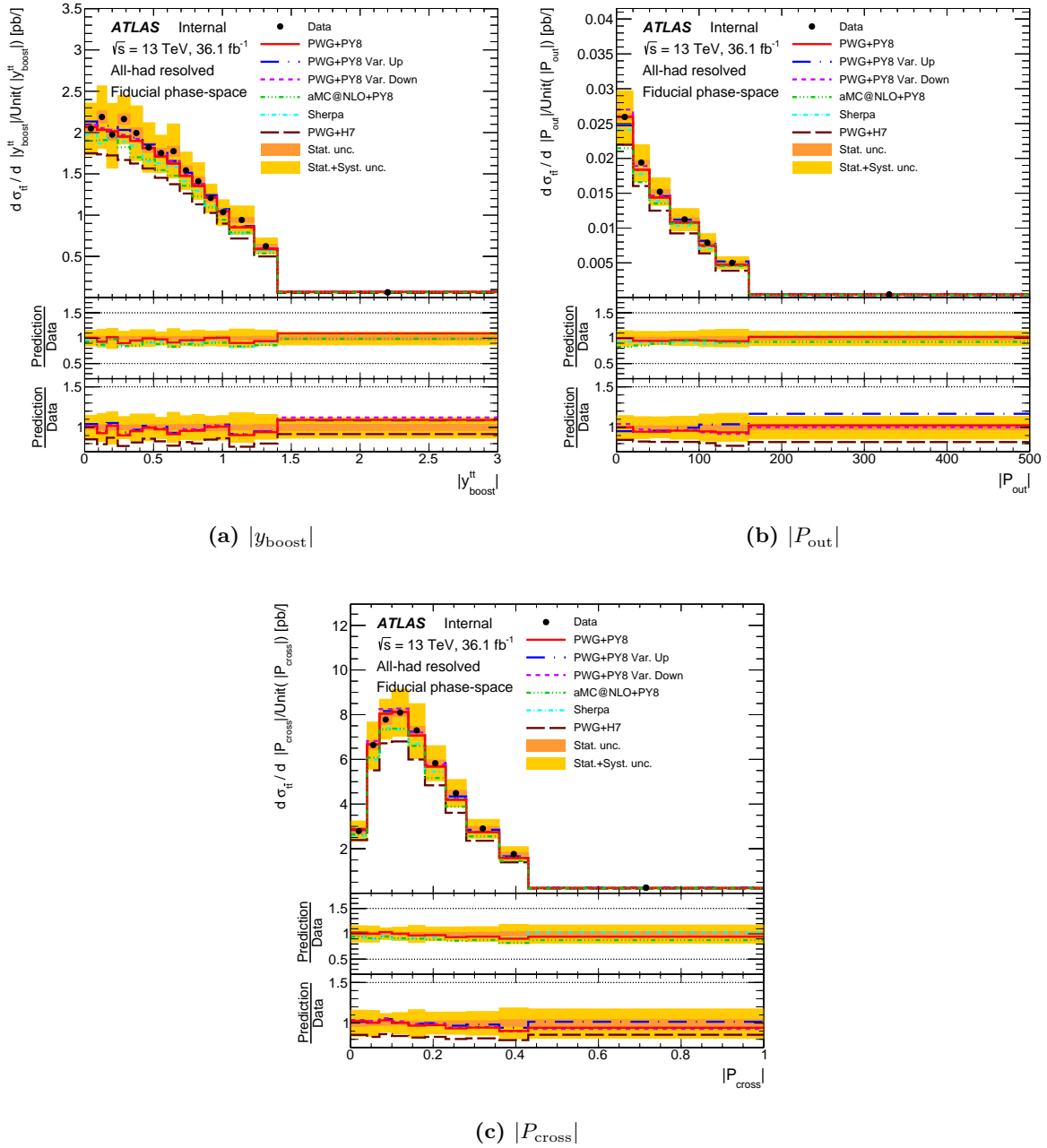
(b) $|P_{\mathrm{out}}|$

(c) $|P_{\mathrm{cross}}|$

Figure C.6 Absolute differential cross sections as functions of (a) $|y_{\mathrm{boost}}|$, (b) $|P_{\mathrm{out}}|$, and (c) $|P_{\mathrm{cross}}|$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulations to the data.

(a) R_{Wb}^{leading}

(b) $R_{Wb}^{\text{subleading}}$

(c) R_{Wt}^{leading}
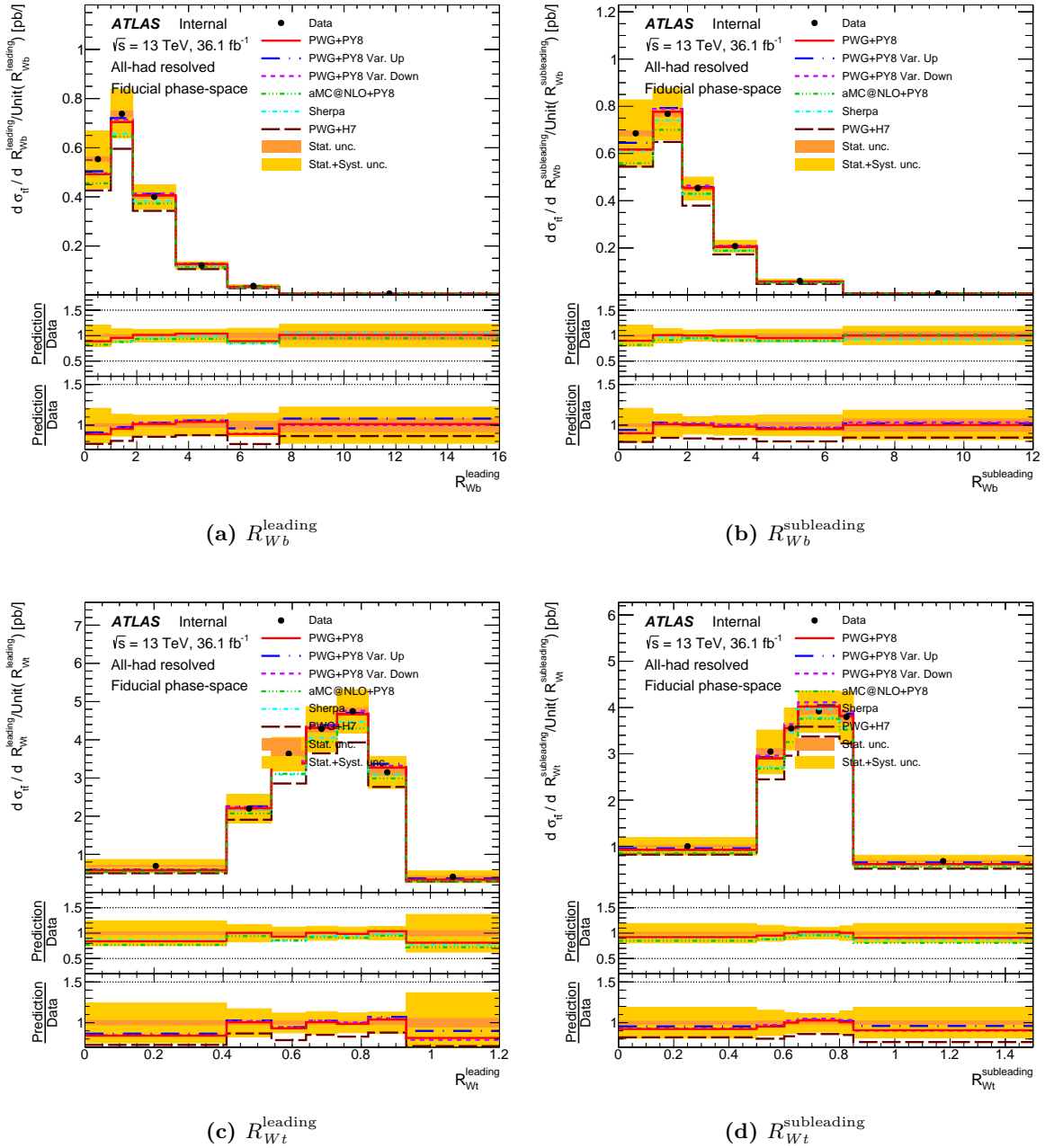
(d) $R_{Wt}^{\text{subleading}}$

Figure C.7 Absolute differential cross sections as functions of (a) R_{Wb}^{leading}, (b) $R_{Wb}^{\text{subleading}}$, (c) R_{Wt}^{leading}, and (d) $R_{Wt}^{\text{subleading}}$, all unfolded to particle level. The bottom panels show the ratios of predictions from simulation to the data.

Particle level 2D differential cross sections

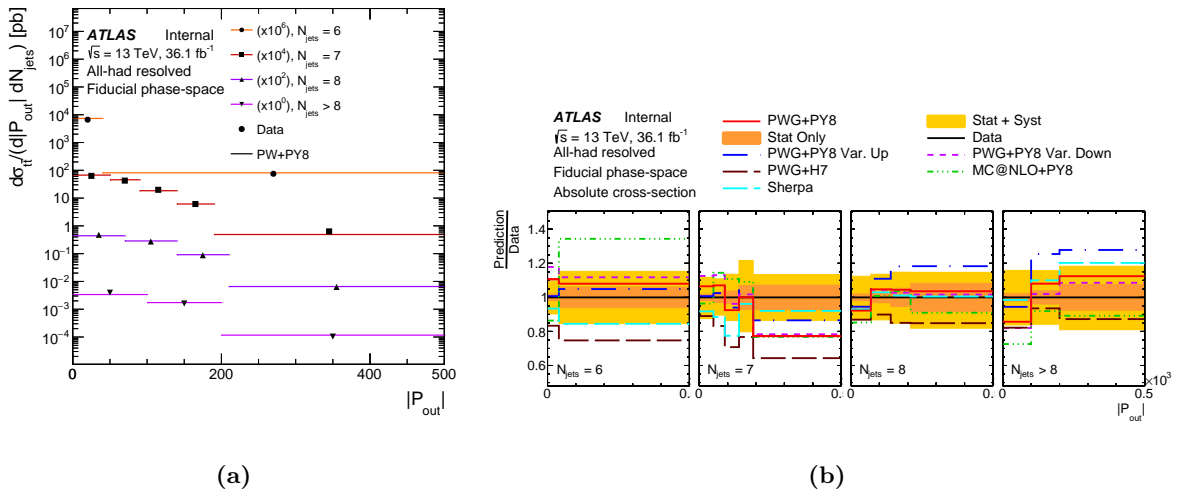(a)   (b)
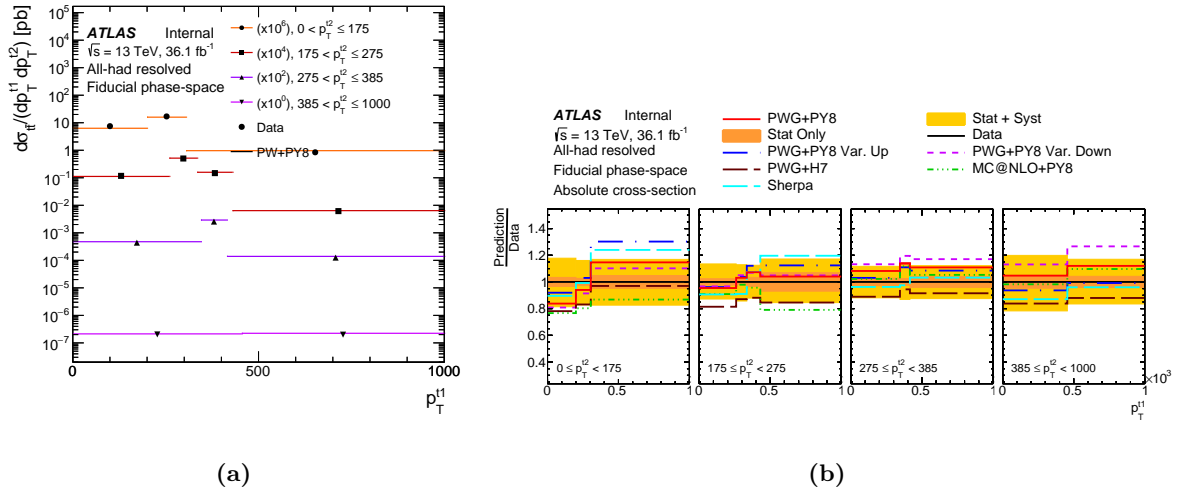
Figure C.8 Absolute differential cross section (a) as a function of $p_T^{t_1}$ in bins of $p_T^{t_2}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

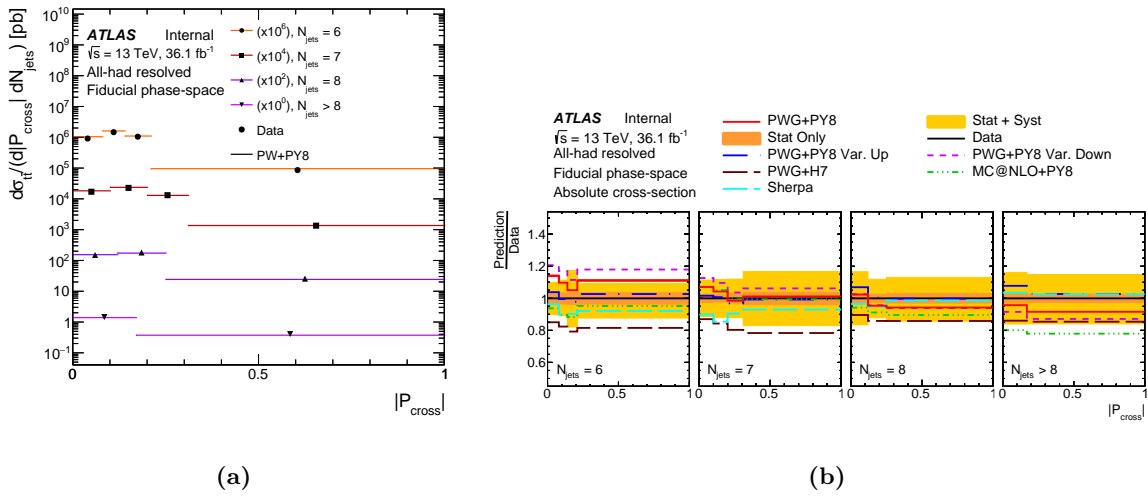
(a)   (b)
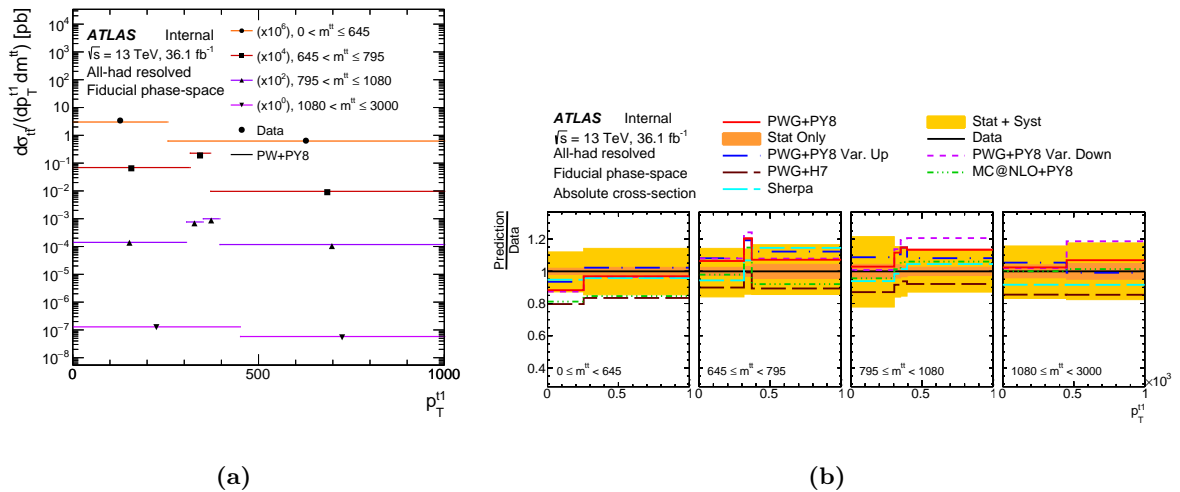
Figure C.9 Absolute differential cross section (a) as a function of $p_T^{t_1}$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)   **(b)**
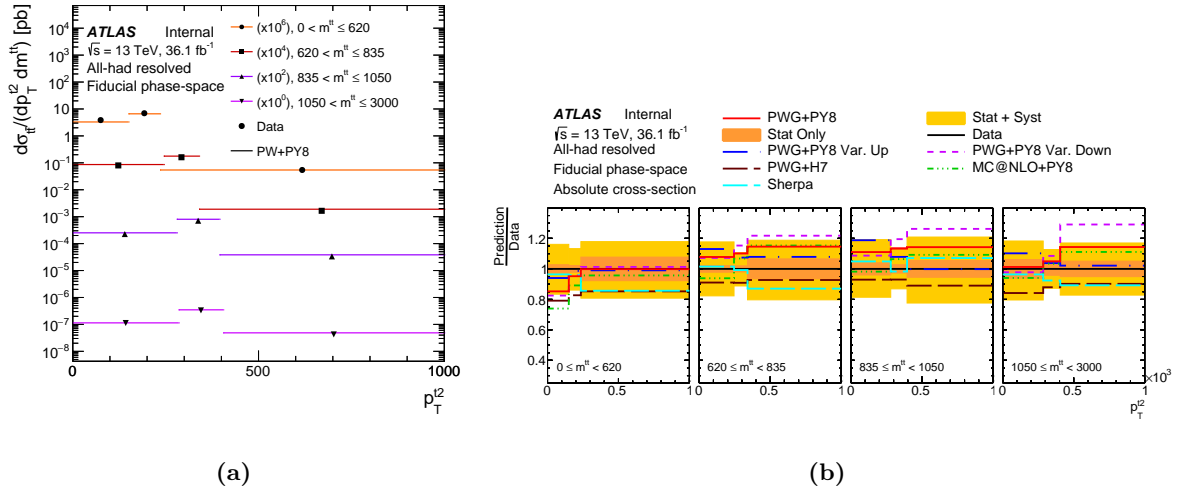
Figure C.10 Absolute differential cross section (a) as a function of $p_{\mathrm{T}}^{t_2}$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)   **(b)**
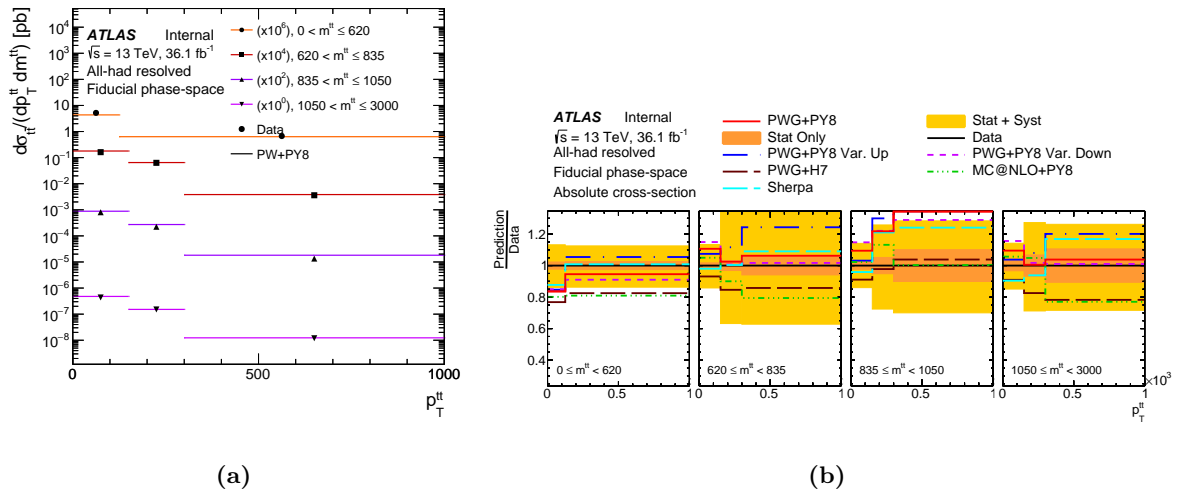
Figure C.11 Absolute differential cross section (a) as a function of $p_{\mathrm{T}}^{t\bar{t}}$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)
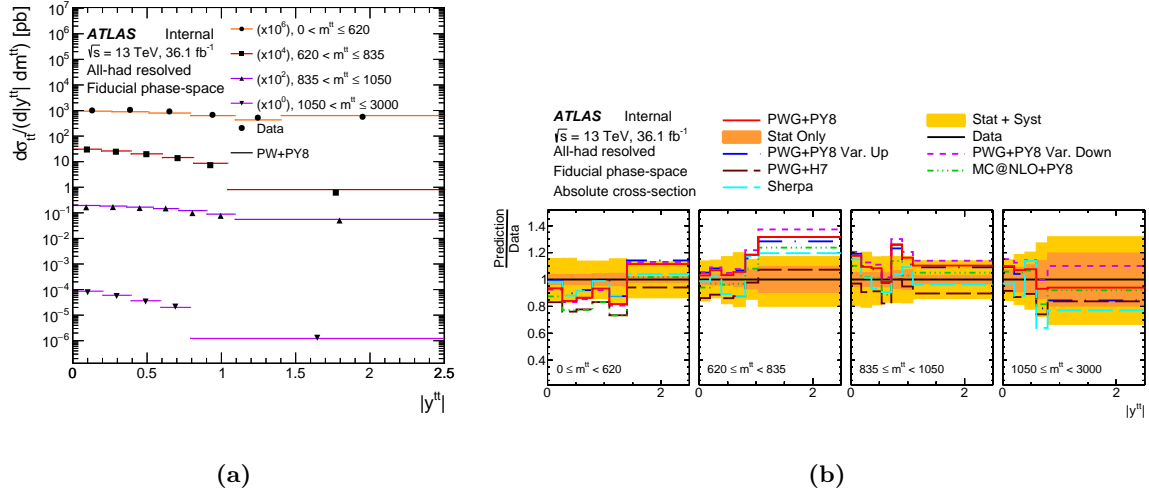
Figure C.12 Absolute differential cross section (a) as a function of $|y^{t\bar{t}}|$ in bins of $m^{t\bar{t}}$, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)

(b)
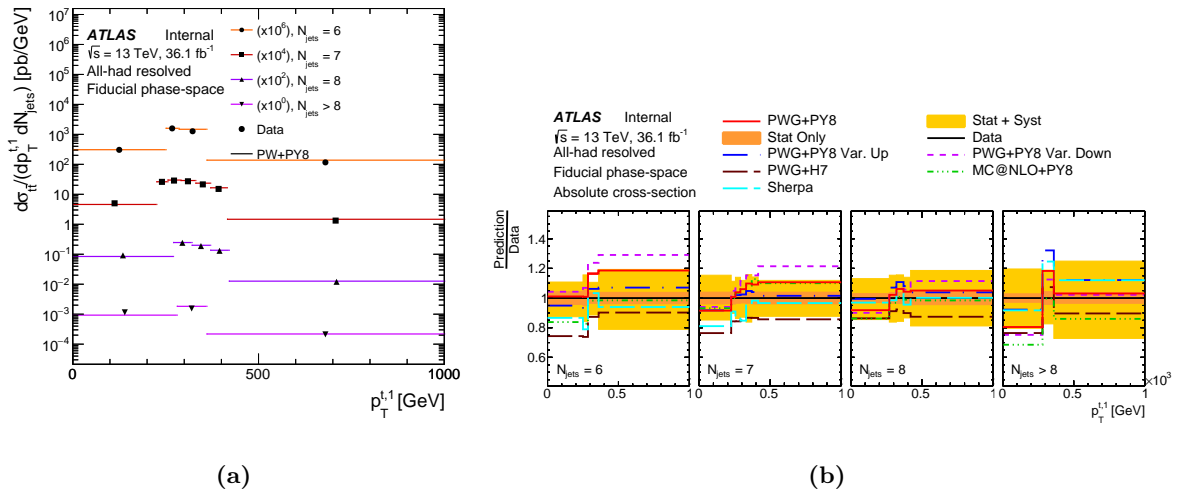
Figure C.13 Absolute differential cross section (a) as a function of $p_T^{t,1}$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.
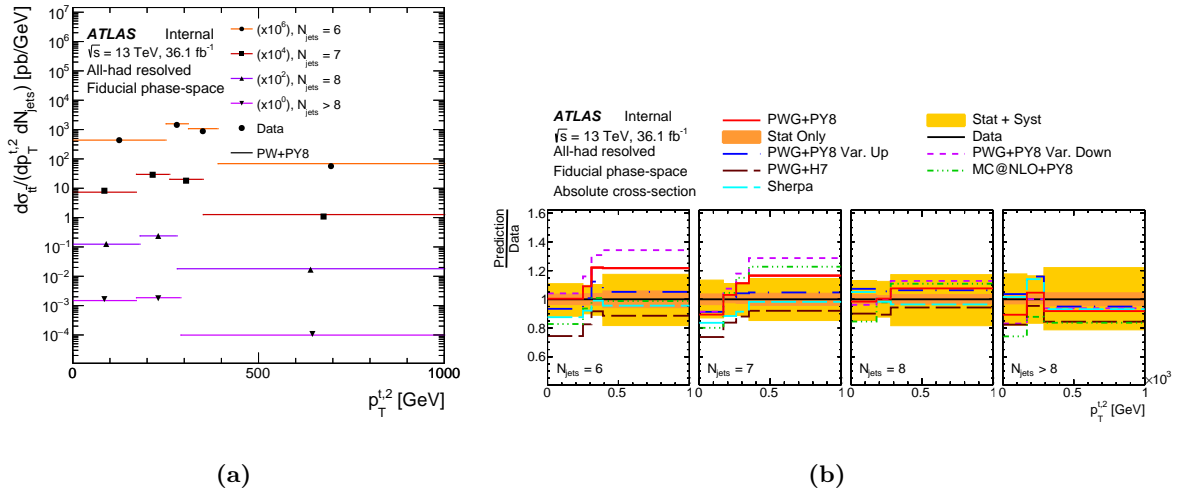
Figure C.14 Absolute differential cross section (a) as a function of $p_{\mathrm{T}}^{t_2}$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.
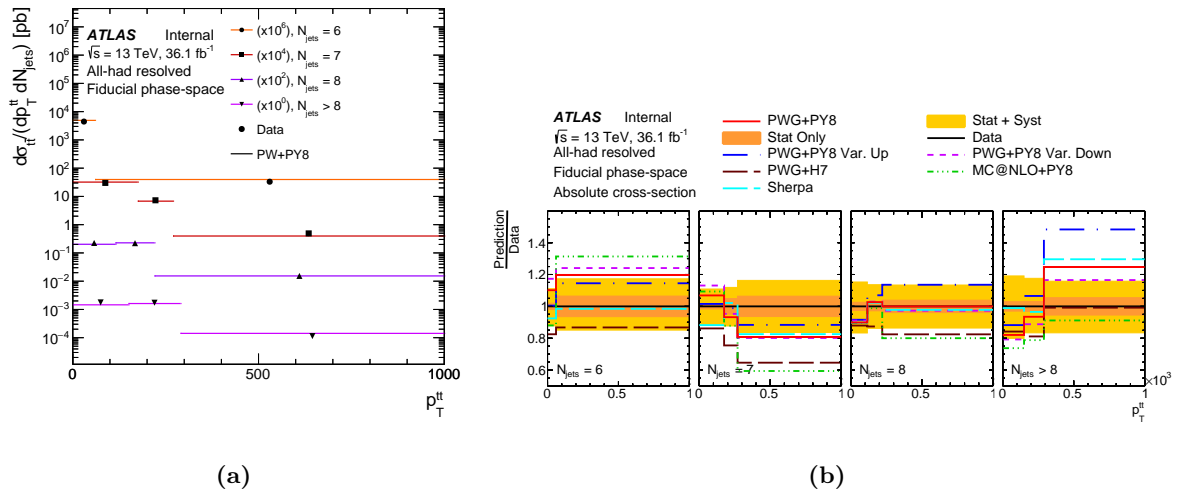
Figure C.15 Absolute differential cross section (a) as a function of $p_{\mathrm{T}}^{t\bar{t}}$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)   (b)
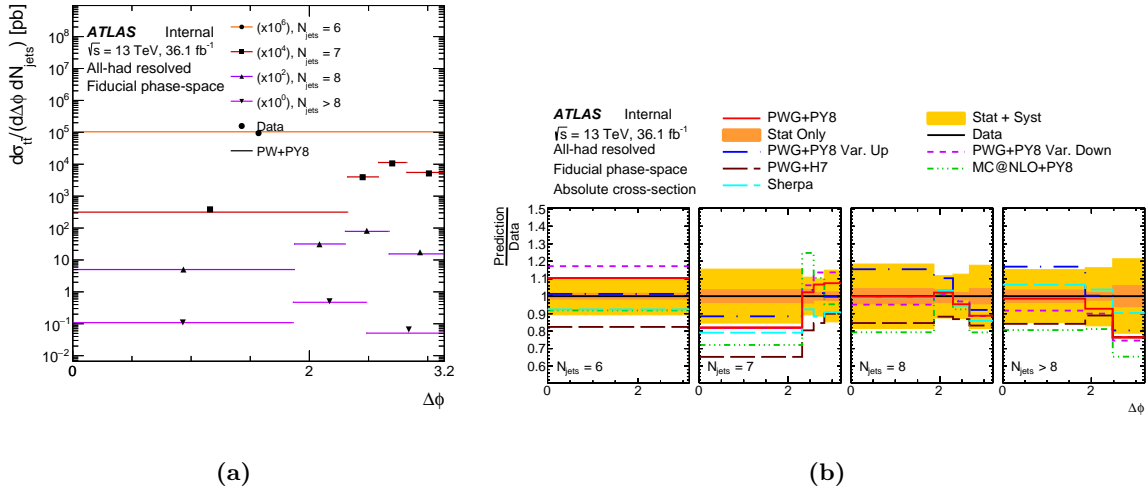
Figure C.16 Absolute differential cross section (a) as a function of $\Delta\phi$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)   (b)

Figure C.17 Absolute differential cross section (a) as a function of $|P_{\text{out}}|$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

(a)   (b)

Figure C.18 Absolute differential cross section (a) as a function of $|P_{\text{cross}}|$ in bins of N_{jets}, unfolded to particle level. (b) Ratio of predictions from simulations to data.

Parton level 1D differential cross sections

(a) $p_T^{t_1}$
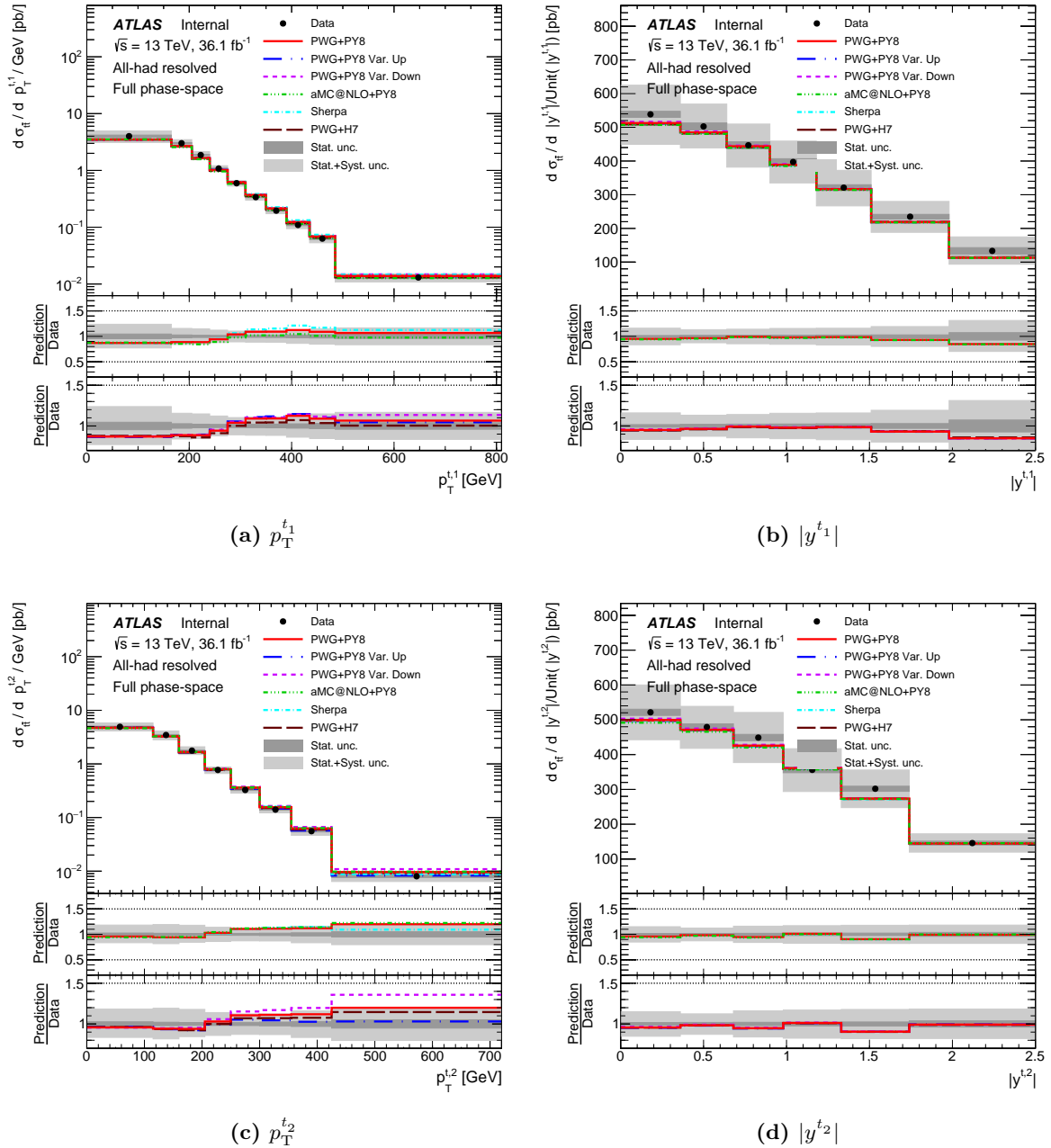
(b) $|y^{t_1}|$

(c) $p_T^{t_2}$

(d) $|y^{t_2}|$

Figure C.19 Absolute differential cross sections as functions of (a) $p_T^{t_1}$, (b) $|y^{t_1}|$, (c) $p_T^{t_2}$, and (d) $|y^{t_2}|$, all unfolded to parton level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $p_{\mathrm{T}}^{t\bar{t}}$
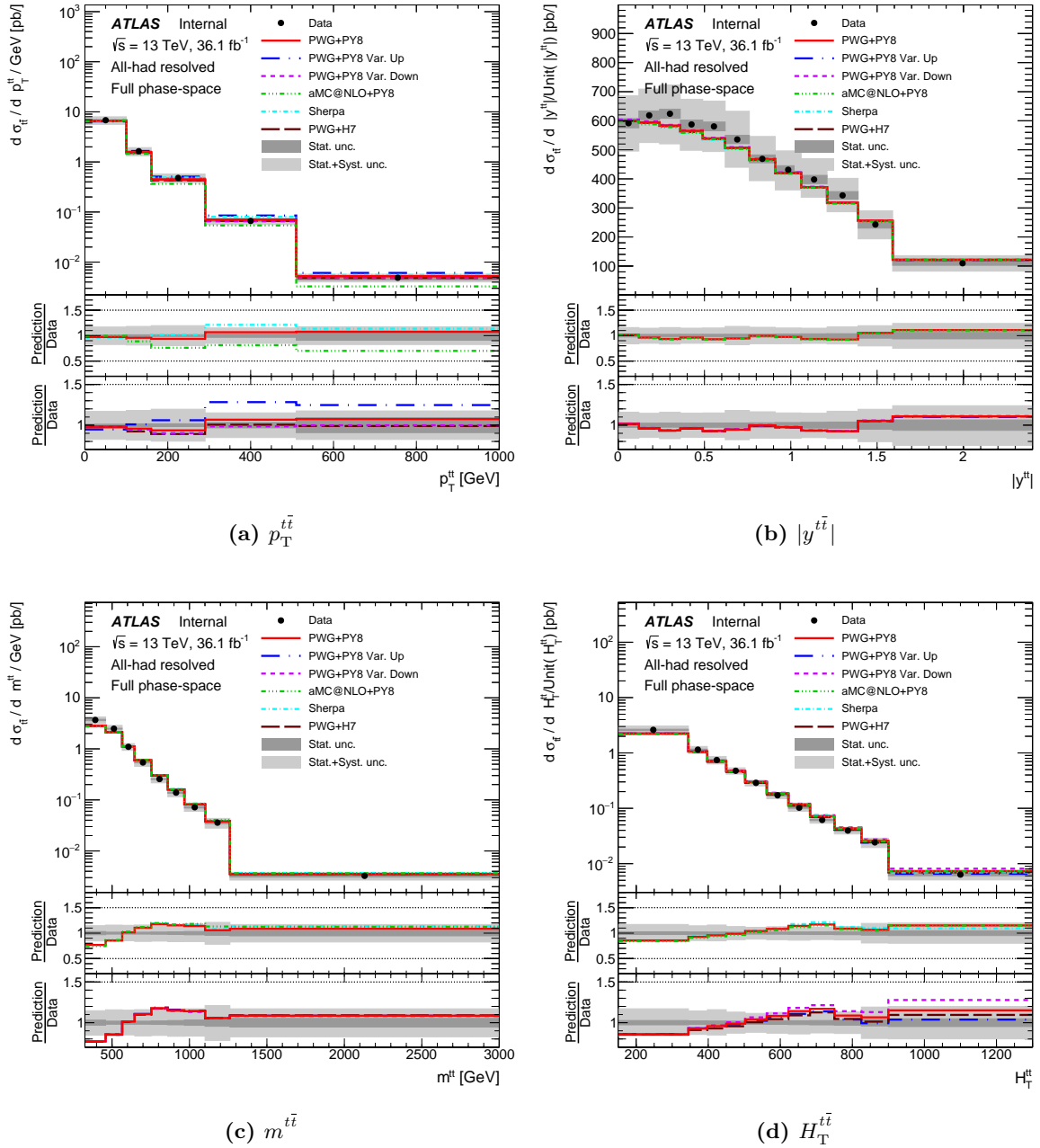
(b) $|y^{t\bar{t}}|$

(c) $m^{t\bar{t}}$

(d) $H_{\mathrm{T}}^{t\bar{t}}$

Figure C.20 Absolute differential cross sections as functions of (a) $p_{\mathrm{T}}^{t\bar{t}}$, (b) $|y^{t\bar{t}}|$, (c) $m^{t\bar{t}}$, and (d) $H_{\mathrm{T}}^{t\bar{t}}$, all unfolded to parton level. The bottom panels show the ratios of predictions from simulation to the data.

(a) $\chi^{t\bar{t}}$

(b) $\Delta\phi$
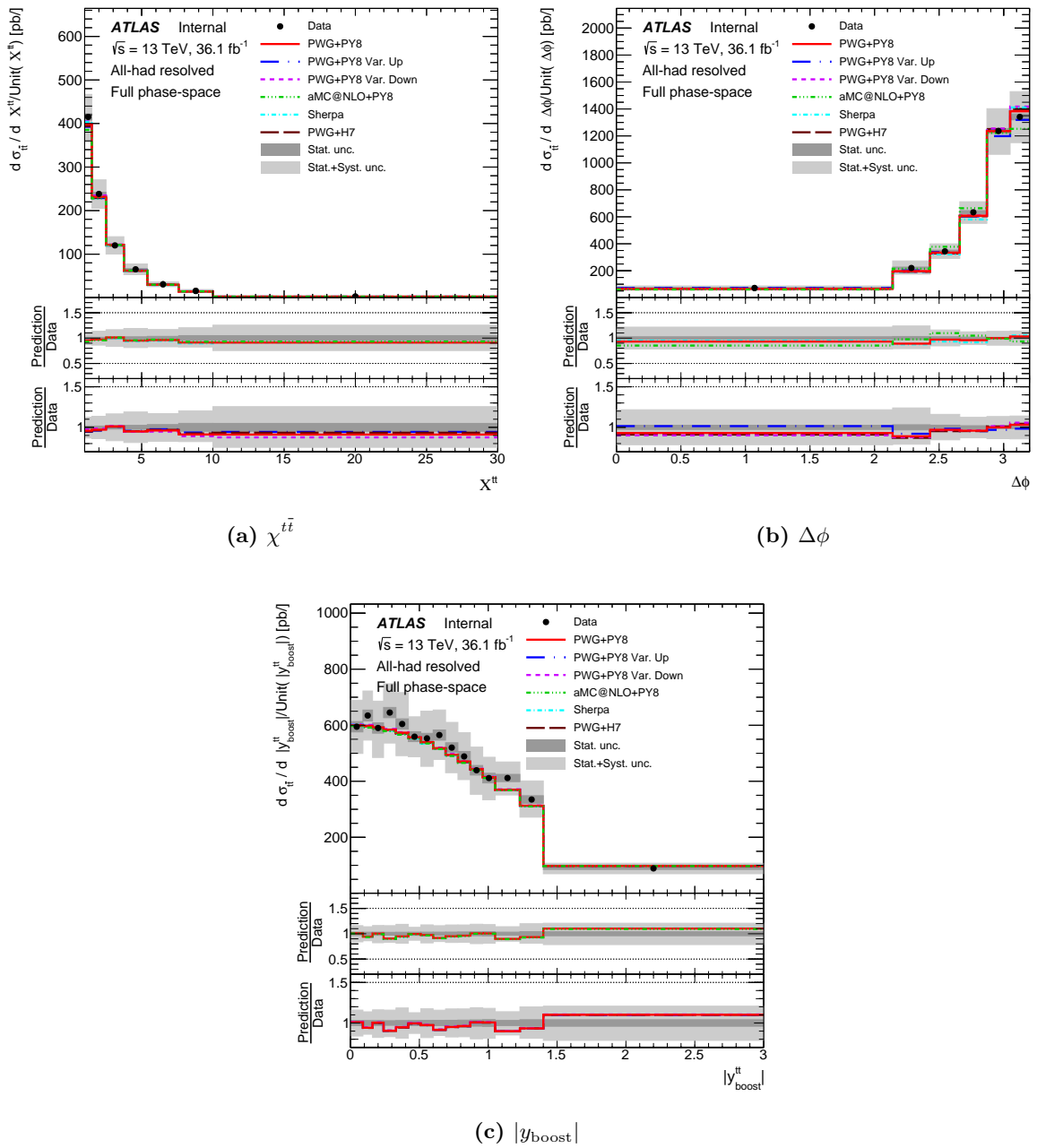
(c) $|y_{\text{boost}}|$

Figure C.21 Absolute differential cross sections as functions of (a) $\chi^{t\bar{t}}$, (b) $\Delta\phi$, and (c) $|y_{\text{boost}}|$, all unfolded to parton level. The bottom panels show the ratios of predictions from simulation to the data.

Parton level 2D differential cross sections

(a)

(b)
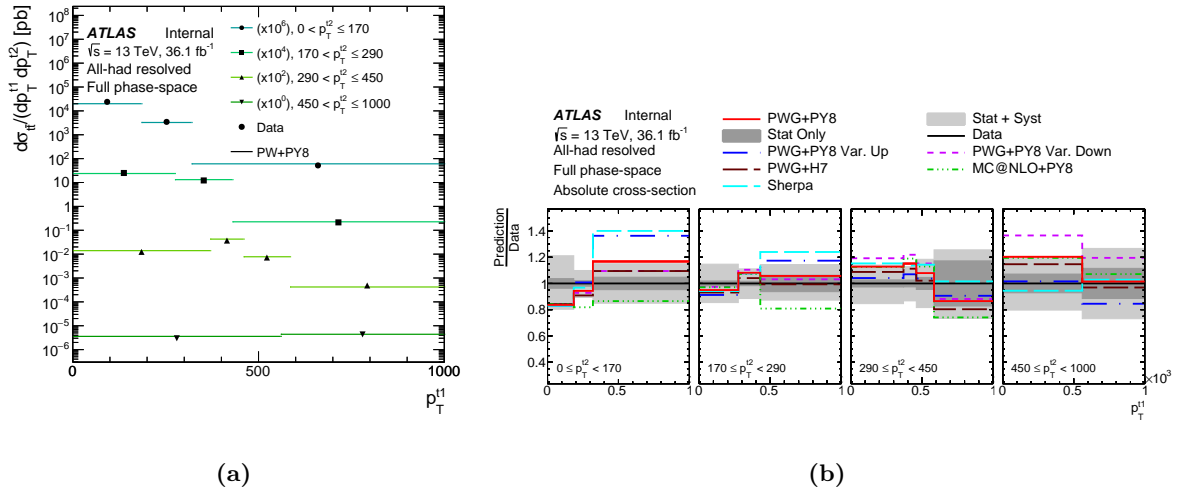
Figure C.22 (a) Absolute differential cross section as a function of $p_\mathrm{T}^{t_1}$ in bins of $p_\mathrm{T}^{t_2}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
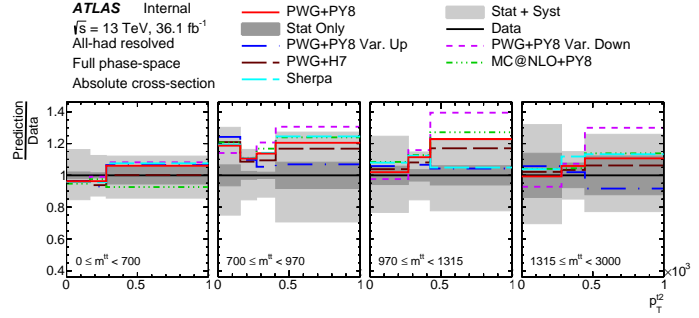

(a)

(b)
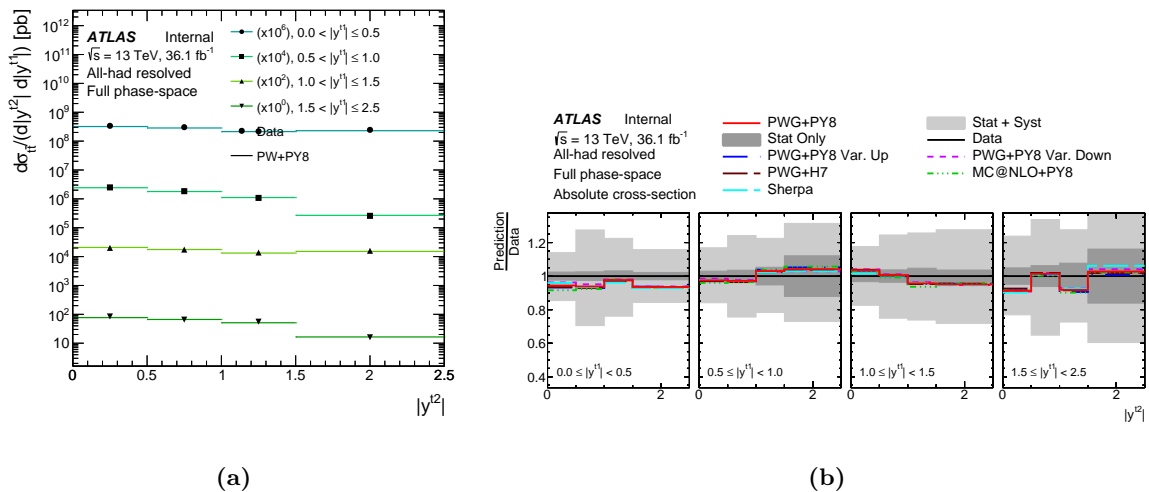
Figure C.23 (a) Absolute differential cross section as a function of $|y^{t_2}|$ in bins of $|y^{t_1}|$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
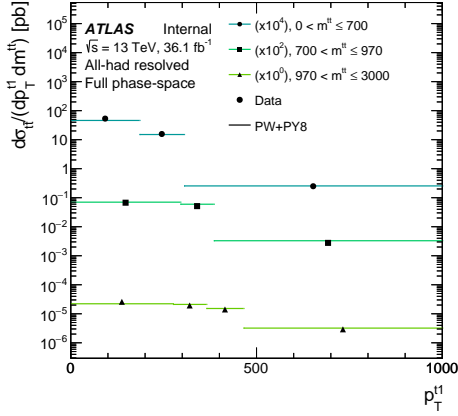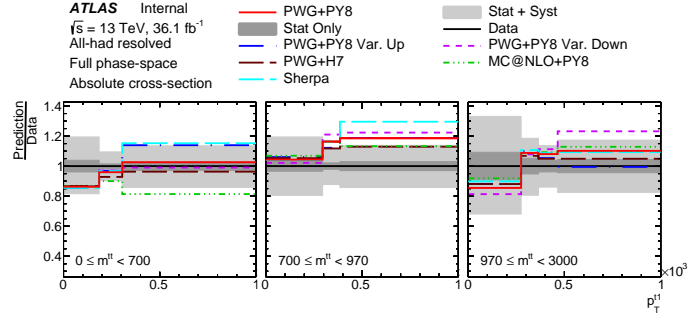
Figure C.24 (a) Absolute differential cross section as a function of $p_{\mathrm{T}}^{t_1}$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
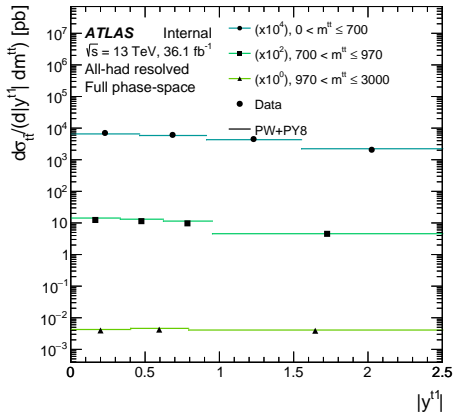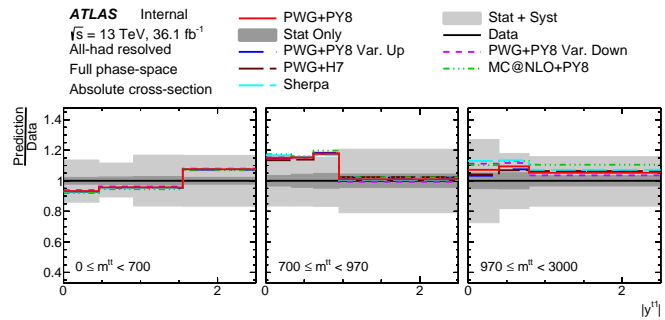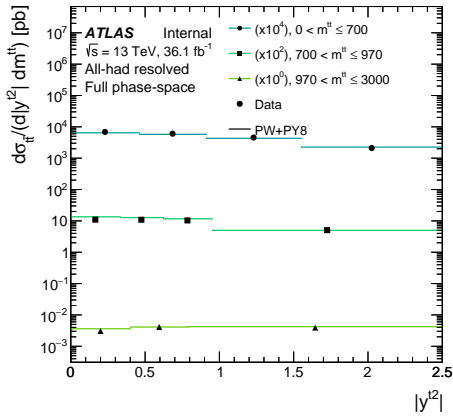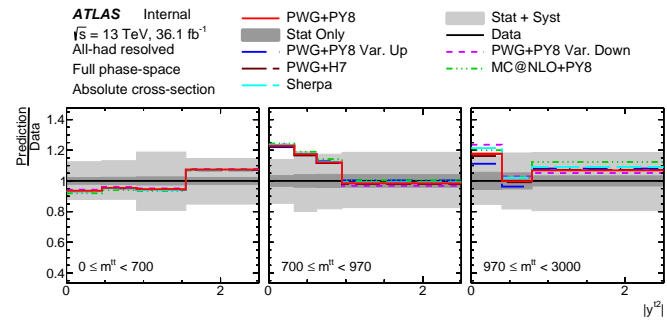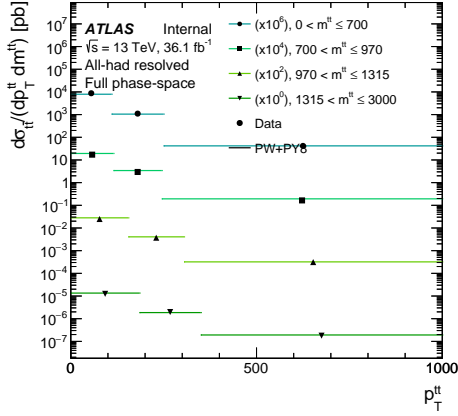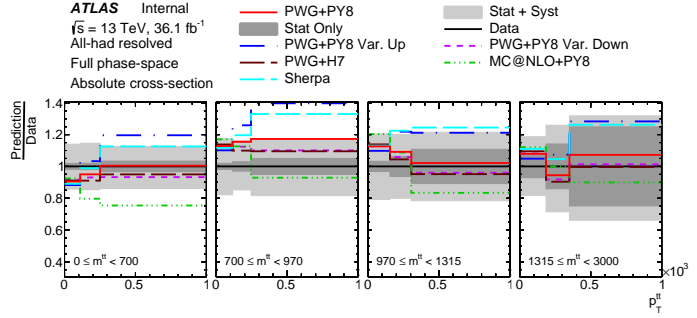

Figure C.25 (a) Absolute differential cross section as a function of $|y^{t_1}|$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
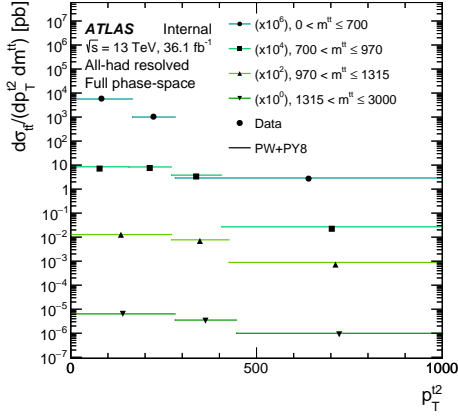
Figure C.26 (a) Absolute differential cross section as a function of $p_{\mathrm{T}}^{t_2}$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.

Figure C.27 (a) Absolute differential cross section as a function of $|y^{t_2}|$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.
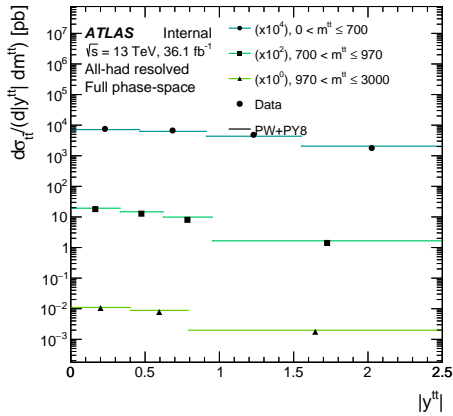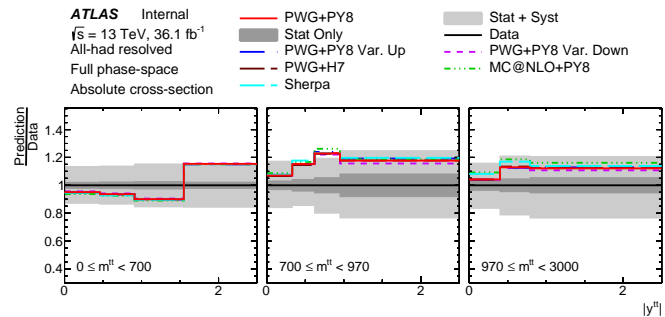
(a)

(b)

Figure C.28 (a) Absolute differential cross section as a function of $p_{\mathrm{T}}^{t\bar{t}}$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.

(a)

(b)

Figure C.29 (a) Absolute differential cross section as a function of $|y^{t\bar{t}}|$ in bins of $m^{t\bar{t}}$, unfolded to parton level. (b) Ratio of predictions from simulations to data.