Jets, Substructure, and Searching for Dark Matter at the Large Hadron Collider

by

Siddharth Madhavan Narayanan

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2018

Author . . . .
Department of Physics
August 32, 2018

Certified by. . . . .
Christoph M. E. Paus
Professor of Physics
Thesis Supervisor

Accepted by . . . .
Somebody
Chairman, Department Committee on Graduate Theses

Jets, Substructure, and Searching for Dark Matter at the Large Hadron Collider

by

Siddharth Madhavan Narayanan

Submitted to the Department of Physics
on August 32, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Astrophysical observations of gravitational interactions provide strong evidence for the existence of dark matter (DM). Many theories propose and experiments test the hypothesis that DM may have a particle physics origin, but this remains unproven. One such experiment is the Compact Muon Solenoid (CMS) at the Large Hadron Collider (LHC). If DM couples to particles present in protons, it is possible that DM is produced in collisions at the LHC. Because DM, by its very nature, is effectively invisible to CMS, we must look for collisions in which DM is produced in association with one or more Standard Model (SM) particles. This thesis describes three different scenarios for the SM particle hypothesis: a single top quark, a single Higgs boson, or two light quarks. All three cases result in complicated detector signatures due to the hadronization of final-state quarks. Improved jet substructure techniques to identify these unique signatures are presented. Since the observed data is consistent with SM backgrounds in all three searches, we translate this result into the most stringent constraints to date on the relevant beyond-SM models.

Thesis Supervisor: Christoph M. E. Paus
Title: Professor of Physics

Acknowledgments

Thank you to Professor Markus Klute for nice barbecues.

Contents

List of Figures

List of Tables

Chapter 1

The Search for $t + p_{\mathrm{T}}^{\mathrm{miss}}$

In this chapter, we discuss the search for dark matter produced in association with a single top quark ("mono-top"). Since the initial state of pp collisions do not contain any appreciable contribution from top quarks, any process that produces a single top quark must involve some flavor violation. In the Standard Model, any such process is heavily suppressed by off-diagonal elements of the CKM matrix. The SM production mechanism for the mono-top signature (Figure 1.1) involves a b quark in the final state, and thus does not couple the third generation with the first or second. True production of mono-top must introduce some such coupling as an extension to the SM, in addition to one (or more) invisible particle to serve as a DM candidate.

To illustrate how beyond-SM physics can produce this final state, we introduce two DM models: a flavor-changing neutral current V and a charged, colored scalar ϕ. These models will also be used to benchmark the sensitivity of the analysis. However, it should be emphasized that the search is motivitated and designed agnostically, without reliance on any specific model; the assumption is that the mono-top final state alone is indicative of new physics, regardless of the specific production mechanism.

The FCNC V is assumed to couple to a fermionic DM candidate χ. A partial Lagrangian of the interaction terms is given by:

$$\mathcal{L}_{\mathrm{int}} = V_\mu \bar{\chi} \gamma^\mu (g_\chi^V + g_\chi^A \gamma_5)\chi + \bar{q}_u \gamma^\mu (g_u^V + g_u^A \gamma_5)q_u V_\mu + \bar{q}_d \gamma^\mu (g_d^V + g_d^A \gamma_5)q_d V_\mu + \mathrm{h.c.}, \quad (1.1)$$
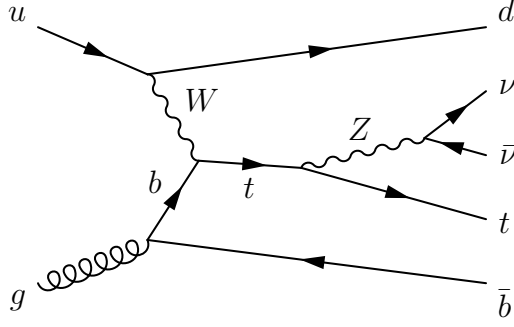
Figure 1.1: Production of mono-top in the SM, in which a top quark is produced in addition to a Z boson and bottom quark. The Z decays to neutrinos, providing large $p_{\mathrm{T}}^{\mathrm{miss}}$.

The model comes with 22 free parameters, broadly organized in three sets:

- The masses m_V and m_χ. (2)

- The couplings g_χ^V and g_χ^A. These, respectively, control the strength of the vector and axial interactions between V and χ. (2)

- The four coupling matrices g_q^X, where $q = u, d$ and $X = V, A$. As before, X determines the type of spin-1 interaction. In principle, different coupling strengths can be permitted for up- and down-type quarks, so this indexed by q. Each g_q^X is a 3×3 matrix, cross-coupling the three quark generations. To preserve $\mathrm{SU}(2)_{\mathrm{L}}$ symmetry, we require $g_u^V - g_u^A = g_d^V - g_d^A$. ($3 \times 6 = 18$)

It is the $g_{u,d}^{V,A}$ matrices that determine whether the model can produce mono-top, or mono-bottom, or mono-up, etc. If $g_{u,d}$ is strongly diagonal (i.e. strongest couplings are within generations), then mono-light quark production will dominate, resulting in the mono-jet final state (Figure 1.2a). On the other hand, if we assume the only non-zero elements are those that couple the first and third generations, then mono-top production at the LHC is the best way to probe this model (Figure 1.2b). It is this latter choice that will be made in the rest of this chapter; other choices are best probed using a combination of multiple DM channels, which is left as future work. Furthermore, to respect $\mathrm{SU}(2)_{\mathrm{L}}$ symmetry, we make the assumption that $g_u^V = g_d^V$ and $g_u^A = g_d^A$.

(a) $(g_u^{V,A})_{ij} \approx \delta_{ij}$ (b) $(g_u^{V,A})_{ij} \approx \delta_{i1}\delta_{j3} + \delta_{i3}\delta_{j1}$
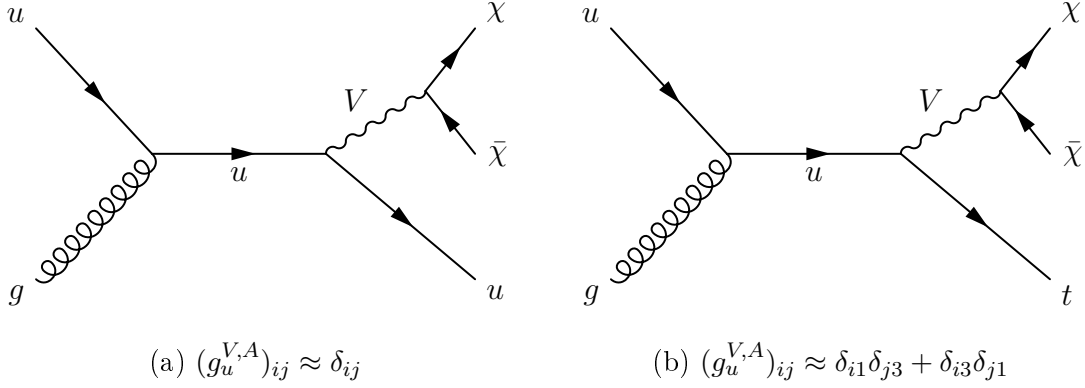
Figure 1.2: Possible DM production at the LHC, assuming a simplified spin-1 extension to the SM.

In the second benchmark model, the charged, colored scalar ϕ couples to down-type quarks, or to a fermionic DM candidate ψ and a top quark. The interaction terms of the Lagrangian is given by:

$$\mathcal{L}_{\text{int}} = \phi \bar{d}_i^C [(a_q)^{ij} + (b_q)^{ij}\gamma^5]d_j + \phi\bar{t}[a_\psi + b_\psi\gamma^5]\psi + \text{h.c.} \tag{1.2}$$

There are 16 free parameters in this model, broadly organized in three categories:

- The masses m_ϕ and m_ψ. (2)

- The couplings at the $\phi\bar{t}\psi$ vertex a_ψ and b_ψ, which respectively control the strength of the scalar and pseudoscalar interactions. (2)

- The couplings at the $\phi\bar{d}_i d_j$ vertex a_q^{ij} and b_q^{ij} where $i,j = 1,2,3$. Again, a and b refer the scalar and pseudoscalar couplings, respectively. (12)

In this model, mono-top production primarily occurs through the resonant decay of ϕ to ψ and t, as shown in Figure 1.3.

The two benchmark models show markedly different spectra in Figure 1.4, motivating their use to test different modes of mono-top production. The FCNC produces a falling p_T^{miss} distribution, whereas the scalar resonance produces a p_T^{miss} distribution peaking at approximately $m_\phi/2$.
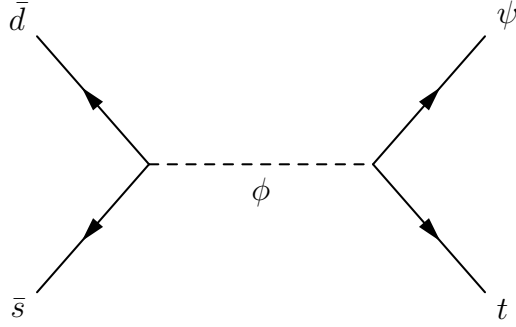
Figure 1.3: Possible DM production at the LHC, assuming the existence of a charged, color scalar that couples to DM and the top quark.

(a) FCNC

(b) Scalar resonance
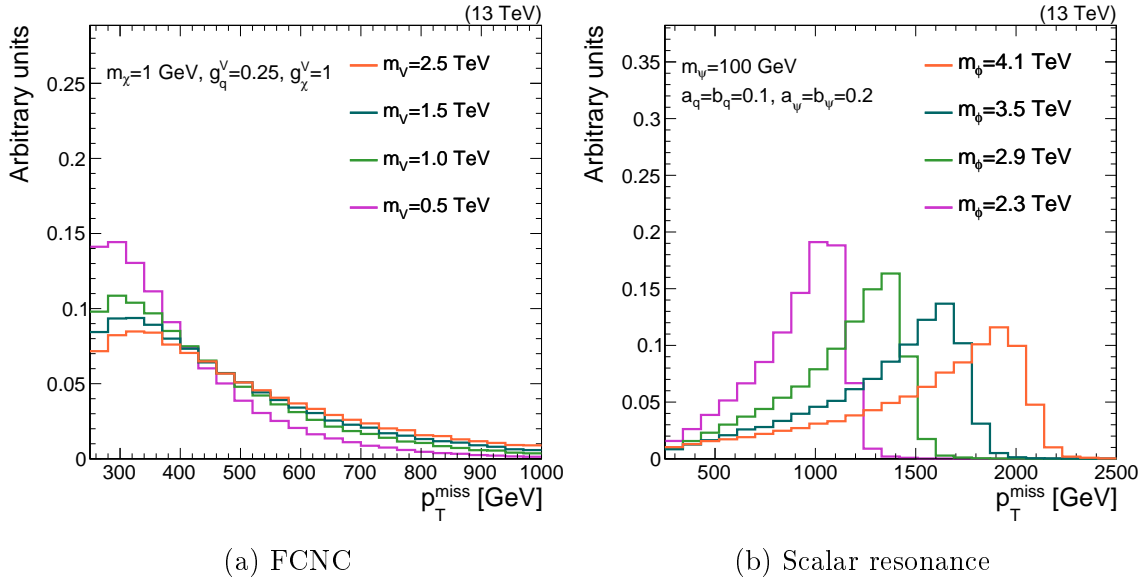
Figure 1.4: Spectra of DM (missing) momentum under various signal hypothesis.

(a) Missing momentum　　　　　　　(b) Jet p_{T}
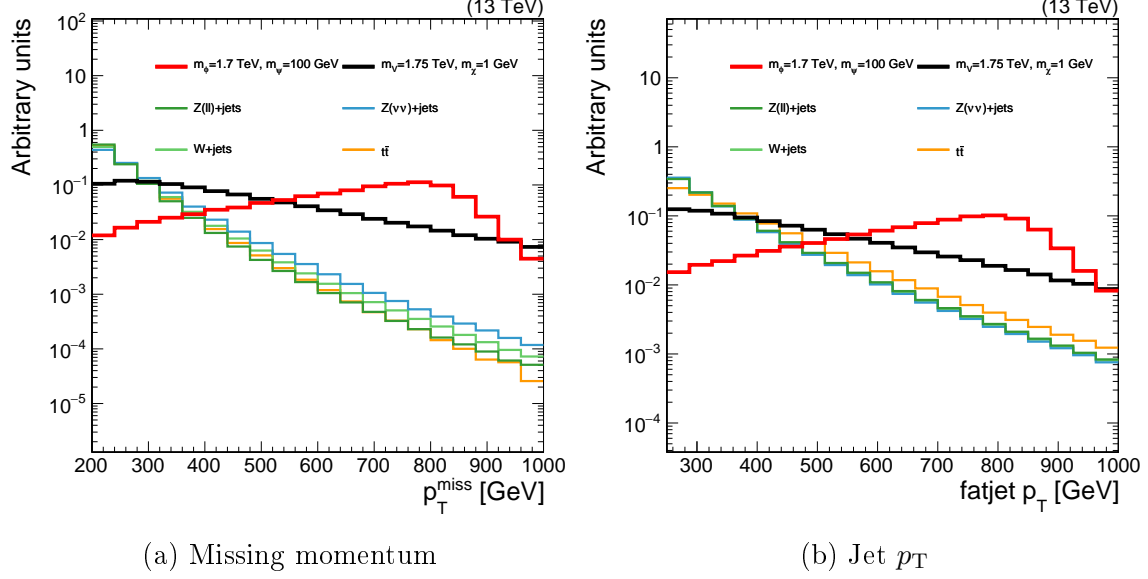
Figure 1.5: Comparison of missing and jet momenta in various backgrounds and signal models.

1.1 Signal selection

When looking at events that pass a simple set of criteria (moderate $p_{\mathrm{T}}^{\mathrm{miss}}$ and one CA15 jet), it is clear (Figure 1.5) that the highest signal sensitivity is found in regions of high $p_{\mathrm{T}}^{\mathrm{miss}}$ and jet p_{T}. The three primary background processes are:

- $Z \to \nu\nu$. When the Z is produced in association with one or more jets, the jet system can (with low probability) pass the criteria used to select a top jet. The neutrinos manifest as $p_{\mathrm{T}}^{\mathrm{miss}}$.

- $W \to \ell\nu$. As in the case of the Z, additional jets can mimic the signature of a top jet. Typically, the charged lepton in the final state is vetoed, but if it is out of acceptance (e, μ) or fails ID criteria (τ_{h}), then it is not identified.

- $t\bar{t} \to bq\bar{q}' + \bar{b}\ell\nu$. As in the case of the W, a charged lepton in the final state may not be properly identified. Unlike the previous two processes, a semi-leptonic $t\bar{t}$ event contains a real hadronic top quark decay.

Events in the signal regions (SRs) are then selected according to Table 1.1, chosen to be consistent with the signal topology while mitigating the aforementioned SM

Table 1.1: Criteria used to select events for the mono-top search signal regions. Note that two SRs are defined, based on the BDT score.

Criterion	Notes
$p_T^{miss} > 250$ GeV	Signal events should have large missing momentum. Exact threshold is chosen to maximize online trigger efficiency.
1 CA15 jet with $p_T > 250$ GeV	Top quark candidate. Recoils against p_T^{miss}, so threshold is set at 250 GeV.
CA15 jet $110 < m_{SD} < 210$ GeV	Consistency with top quark mass.
At least one b-tagged sub-jet	Identifying B hadron produced from top decay/hadronization.
No b-tagged narrow jets	Rejecting semi-leptonic $t\bar{t}$ decays.
No identified e, μ, τ_h	Suppress W+jet and $t\bar{t}$ processes.
No identified γ	Suppress γ+jet processes.
$\min_{jets} \Delta\phi(jet, p_T^{miss}) > 0.5$	Remove events with large p_T^{miss} caused by mismeasured jets.
CA15 jet BDT	Identifying top decay structure. If the jet passes the tight WP, it is placed in the "tight" SR. Otherwise, if it only passes the loose WP, it is placed in the "loose" SR.

backgrounds. As described in Section **??**, two working points (WPs) are defined for the top ID BDT. The signal events (passing all other selection criteria) are partitioned into a "loose" SR and a "tight" SR on the basis of which WP the top candidate jet satisfies.

Figure 1.6 shows the p_T^{miss} distributions, as predicted by MC and as observed in collected data, in the two signal regions.

(a) Loose SR (b) Tight SR
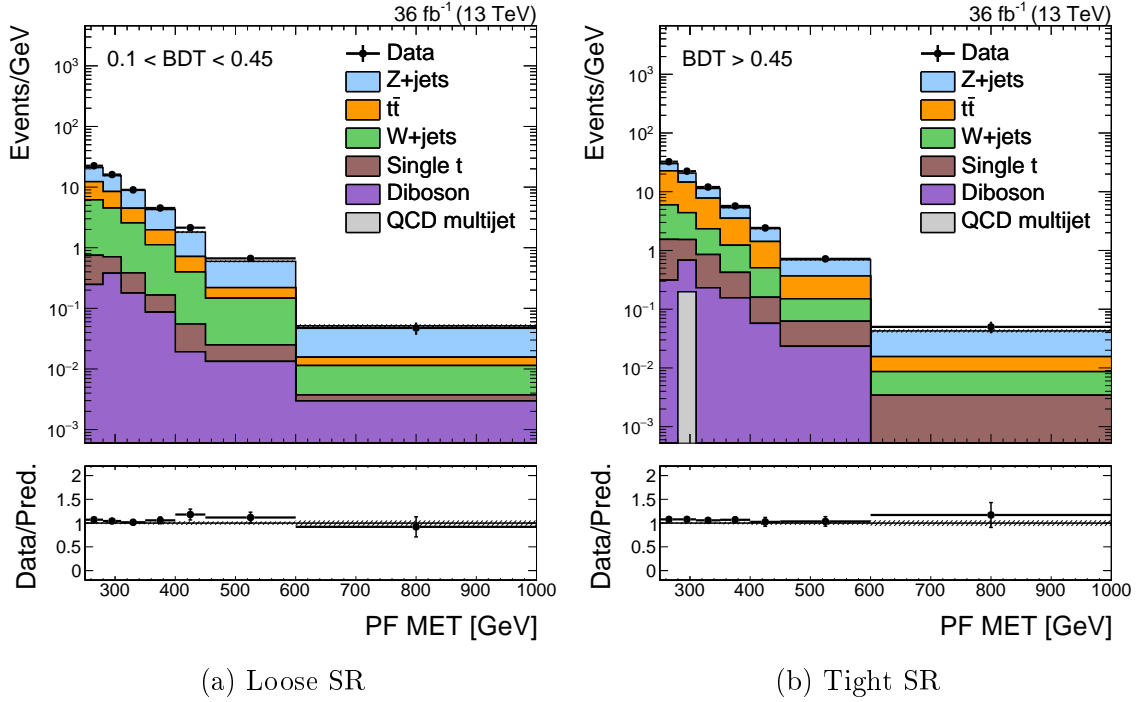
Figure 1.6: p_T^{miss} distributions in the two mono-top signal regions. The bottom section of each figure shows the ratio of the data and the prediction. The only uncertainties plotted in these figures are those arising from Poisson fluctuations in data (black bars) and MC (grey band).

1.2 Background estimation

Searching for DM amounts to looking for an excess of data events over the SM prediction at large values of $p_{\mathrm{T}}^{\mathrm{miss}}$. Therefore, the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution of the three primary SM backgrounds described in Section 1.1 must be predicted with small uncertainty. The MC simulation provides a reasonable description of the data, but the theoretical uncertainties inherent in the MC (primarily due to higher-order QCD effects) can range up to 20%. To reduce the prediction uncertainty further, a "data-driven" approach is used to estimate the SM processes in the SR. In this context, "data-driven" refers to the use of control data (i.e. data that cannot contain signal) to directly estimate or supplement the estimation of SM processes in the SR.

1.2.1 Visible final states to constrain invisible final states

As a starting point, let us tackle the estimation of $Z \to \nu\nu$ in the SR. Since the momentum imbalance (up to experimental effects) in a $Z \to \nu\nu$ event is just the transverse momentum of the Z boson (p_{T}^Z), we must estimate p_{T}^Z. To good approximation, the p_{T}^Z distribution is independent of the decay mode of the Z boson. Therefore, it is natural to estimate $p_{\mathrm{T}}^{\mathrm{miss}}(Z \to \nu\nu)$ by measuring $p_{\mathrm{T}}^Z(Z \to \mu\mu)$, as muons are easily identifiable and reconstructible.

However, there is one important distinction between $\nu\nu$ and $\mu\mu$ events. In the latter, p_{T}^Z can be directly measured, whereas in the former it must be inferred through a momentum imbalance. Effects like jet energy scale and acceptance can impact $p_{\mathrm{T}}^{\mathrm{miss}}$, but not $p_{\mathrm{T}}^{\mu\mu}$. Therefore, instead of directly measuring $p_{\mathrm{T}}^{\mu\mu}$ in $\mu\mu$ events, we define and use the hadronic recoil U:

$$\vec{U} = \vec{p}_{\mathrm{T}}^{\ \mathrm{miss}} + \sum_\mu \vec{p}_{\mathrm{T}}^{\ \mu} + \sum_e \vec{p}_{\mathrm{T}}^{\ e} + \sum_\gamma \vec{p}_{\mathrm{T}}^{\ \gamma} \tag{1.3}$$

In the SR (where there are no e, μ, γ), $U = p_{\mathrm{T}}^{\mathrm{miss}}$. In $Z \to \mu\mu$ events, U mimics the momentum imbalance, if we had pretended the identified muons did not exist when computing $p_{\mathrm{T}}^{\mathrm{miss}}$. Therefore, U is an exact analogy for $p_{\mathrm{T}}^{\mathrm{miss}}$ in the SR. Figure 1.7

(a) $Z \to \nu\nu$ (b) $Z \to \mu\mu$
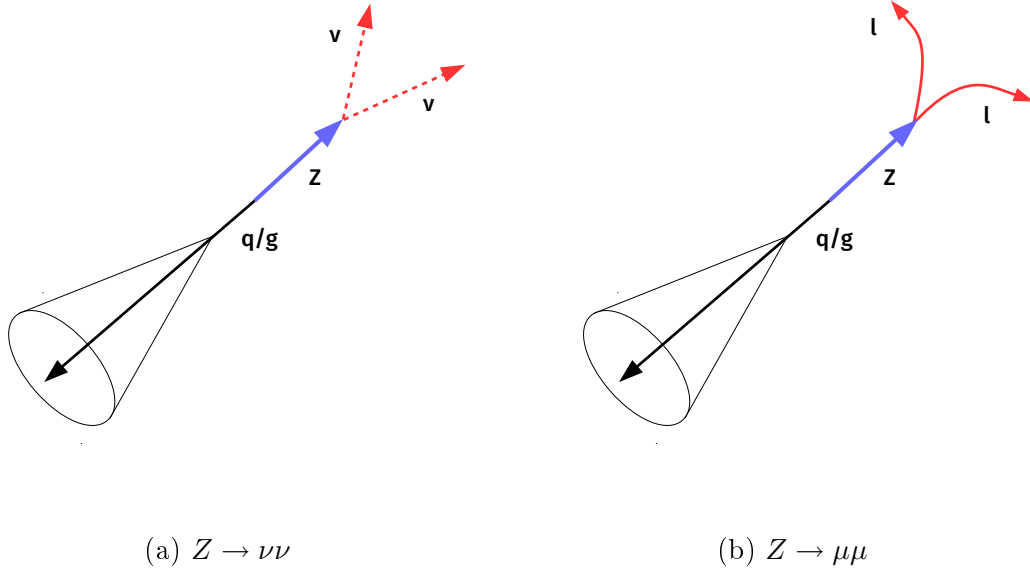
Figure 1.7: Schematic representation of two Z decay modes: to neutrinos (as in the SR) and to muons (as in the CRs). Note that in both cases, U is sensitive to the same effects arising from the measurement of the jet recoiling against the Z boson, whereas $p_{\mathrm{T}}^{\mu\mu}$ is largely independent of the jet.

makes the same argument in a schematic fashion.

Table 1.2 describes the criteria used to define events in the "$\mu\mu$" control regions (CRs). Figure 1.8 shows the distribution of U in these CRs, as well as the $m_{\mu\mu}$ and p_{T}^{μ} distributions.

The control data is used to constrain the SR prediction by means of "transfer factors" $T_{Y,i}^{X}$, where X refers to a particular CR (e.g. $\mu\mu$), Y refers to a particular process (e.g. Z), and i refers to a particular bin in the CR (e.g. $200 < U < 250$ GeV in the tight category). Formally:

$$T_{Z,i}^{\mu\mu} = \frac{N_i^{\mathrm{SR}}(Z \to \nu\nu)}{N_i^{\mu\mu}(Z \to \mu\mu)} \tag{1.4}$$

The transfer factors are estimated using MC simulation. To encode the effects of

23

(a) U in loose CR

(b) U in tight CR

(c) $m_{\mu\mu}$

(d) Leading muon p_T
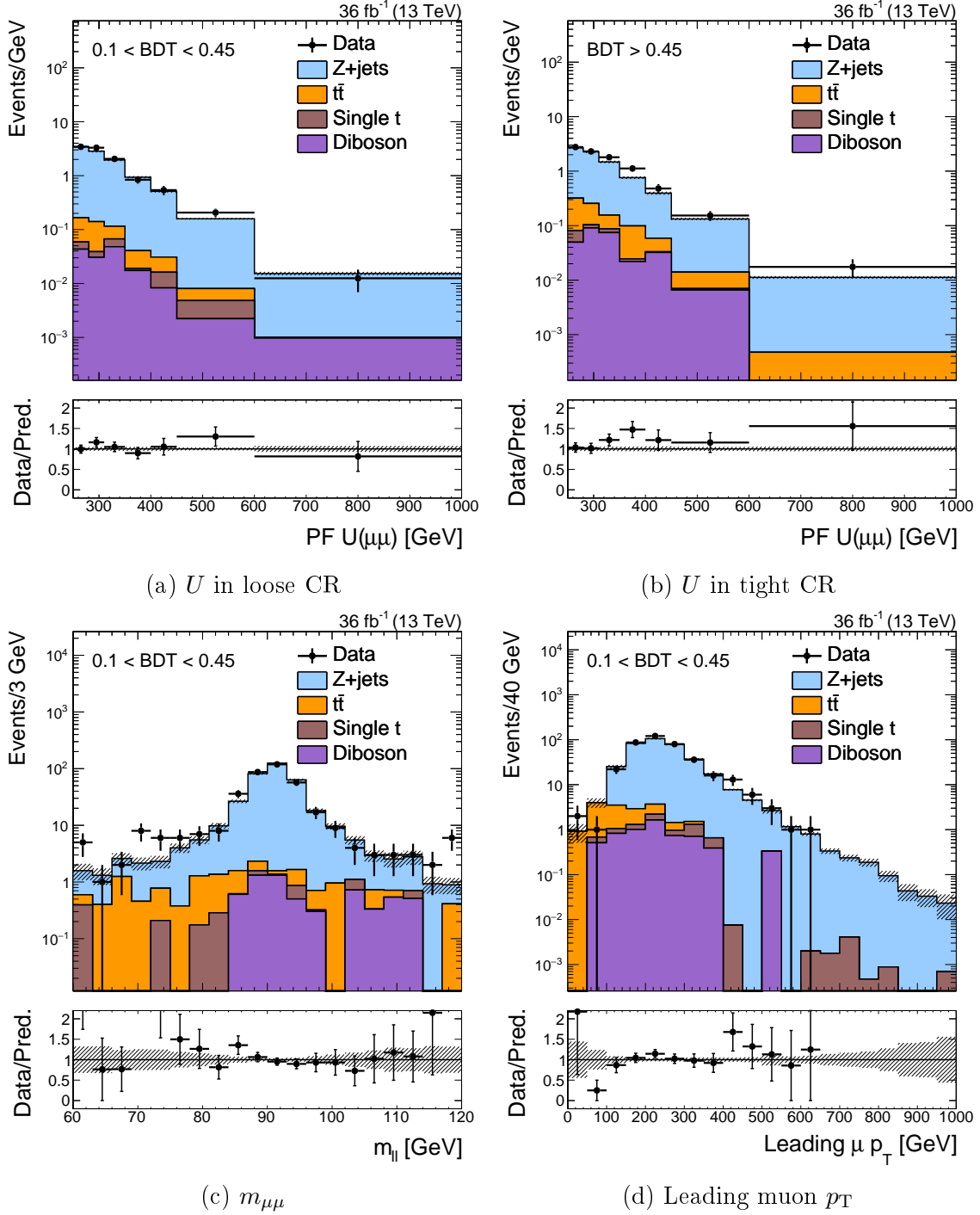
Figure 1.8: Various kinematic distributions in the two mono-top $\mu\mu$ CRs. Note the clearly discernable peak in the $m_{\mu\mu}$ distribution near m_Z

Table 1.2: Criteria used to select events for the mono-top $Z \to \mu\mu$ CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250$ GeV	Mimicking the selection in the SR; also constrained by trigger thresholds.
1 CA15 jet with $p_\mathrm{T} > 250$ GeV	Same as SR
CA15 jet $110 < m_\mathrm{SD} < 210$ GeV	Same as SR
Well-identified μ^-, μ^+ pair, with $|m_{\mu\mu} - m_Z| < 30$ GeV	Identifying the $Z \to \mu\mu$ resonance.
No identified e, τ_h	Same as SR.
No identified γ	Same as SR
$\min_\mathrm{jets} \Delta\phi(\mathrm{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

various uncertainties, we introduce nuisance parameters $\boldsymbol{\theta}$. That is:

$$T_{Y,i}^X \to T_{Y,i}^X(\boldsymbol{\theta}) \equiv T_{Y,i}^X \times \prod_{j=0}^{n_\theta}(1 + \theta_j) \tag{1.5}$$

$$\theta_j \sim p_j(\theta_j) \tag{1.6}$$

where n_θ is the number of nuisance parameters and $p_j(\theta_j)$ is some prior distribution for each nuisance (see below for how the priors are used). The priors are typically chosen to have a central value (e.g. mean, median) at 0, with a finite variance that encodes the uncertainty. In this chapter, we assume p_j is either a normal distribution centered at 0 or a log-normal distribution (in cases where negative values are undesirable). We will use the terms "uncertainty" and "nuisance parameter" interchangeably.

Let $\mathrm{Pois}(d|\lambda)$ refer to the Poisson probability of observing d with an expected mean of λ. In terms of these transfer factors, the likelihood for the data observed in the signal and $\mu\mu$ control regions is:

$$\mathcal{L}(\boldsymbol{d} \mid \mu, \boldsymbol{\mu}_\mathrm{SR}^{Z\to\nu\nu}, \boldsymbol{\theta}) = \prod_{i\in\mathrm{bins}} \Bigg[\mathrm{Pois}\left(d_i^\mathrm{SR} \mid \mu S_i^\mathrm{SR}(\boldsymbol{\theta}) + \mu_{\mathrm{SR},i}^{Z\to\nu\nu} + B_i^\mathrm{SR}(\boldsymbol{\theta})\right)$$
$$\times \mathrm{Pois}\left(d_i^{\mu\mu} \,\Bigg|\, \frac{\mu_{\mathrm{SR},i}^{Z\to\nu\nu}}{T_{Z,i}^{\mu\mu}(\boldsymbol{\theta})} + B_i^{\mu\mu}(\boldsymbol{\theta})\right) \Bigg] \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \tag{1.7}$$

where the following notation is used:

d_i^X : The data observed in bin i of region X. For now, $X = \text{SR}, \mu\mu$.

S_i^{SR} : The predicted number of signal events in bin i of the SR, under some fixed signal hypothesis.

μ : The "signal strength". Essentially an unconstrained nuisance parameter that scales up or down the total signal yield.

$\mu_{\text{SR},i}^P$: The expected number of events from process P in bin i of the SR. This is also an unconstrained nuisance parameter.

B_i^X : The predicted number of "minor" background events in bin i of region X. Here, "minor" refers to all SM processes that are not the signal and are not estimated using a data-driven method.

The signal and background yields $\boldsymbol{S}$ and $\boldsymbol{B}$ are estimated using MC. Note that the inclusion of the priors in the likelihood constrains the nuisance parameters to be close to their "nominal" values; moving a θ_j to fit the data incurs a large cost from the prior.

If we set $B_i = \mu = 0$ (the null hypothesis, ignoring small minor backgrounds), a simple picture emerges of how the likelihood is maximized. The parameters $\boldsymbol{\mu}_{\text{SR}}^{Z \to \nu\nu}$ float freely to satisfy $d_{\text{SR},i} \sim \mu_{\text{SR},i}^{Z \to \nu\nu}$ and $d_{\mu\mu,i} \sim \mu_{\text{SR},i}^{Z \to \nu\nu} / T_{Z,i}^{\mu\mu}(\boldsymbol{\theta})$. If both constraints cannot be satisfied simultaneously by scaling $\boldsymbol{\mu}_{\text{SR}}^{Z \to \nu\nu}$, the (constrained) nuisance parameters $\boldsymbol{\theta}$ modify the transfer factor $\boldsymbol{T}_Z^{\mu\mu}$. Table 1.6 shows the relevant uncertainties for $\boldsymbol{T}_Z^{\mu\mu}$, and Figure 1.9 shows the shape of uncertainties that evolve as a function of U.

The transfer factors are shown in Figure 1.10. The exact values of $\boldsymbol{T}_Z^{\mu\mu}(\boldsymbol{\theta})$ have two salient features:

1. The values are strictly greater than one. This is due to (a) $\mathcal{B}(Z \to \nu\nu) > \mathcal{B}(Z \to \mu\mu)$ and (b) a non-100% efficiency in reconstructing and identifying muons. This implies that the constraining power of the $\mu\mu$ CR is less than that of the SR, especially at high U (i.e. the Poisson uncertainties are larger).

Table 1.3: Uncertainties affecting the $\mu\mu \leftrightarrow \nu\nu$ extrapolation. "Shape" uncertainties have different priors for each bin, but are assumed to be correlated across bins.

Uncertainty	1 s.d.	Notes
μ ID	2%	
μ track	1%	
τ_h veto	3%	
Z+heavy flavor	3%	
Trigger	0-2%	Shape
b-tag	$\sim 0.5\%$	Shape
$udcsg$-mistag	3-4%	Shape

(a) Loose

(b) Tight
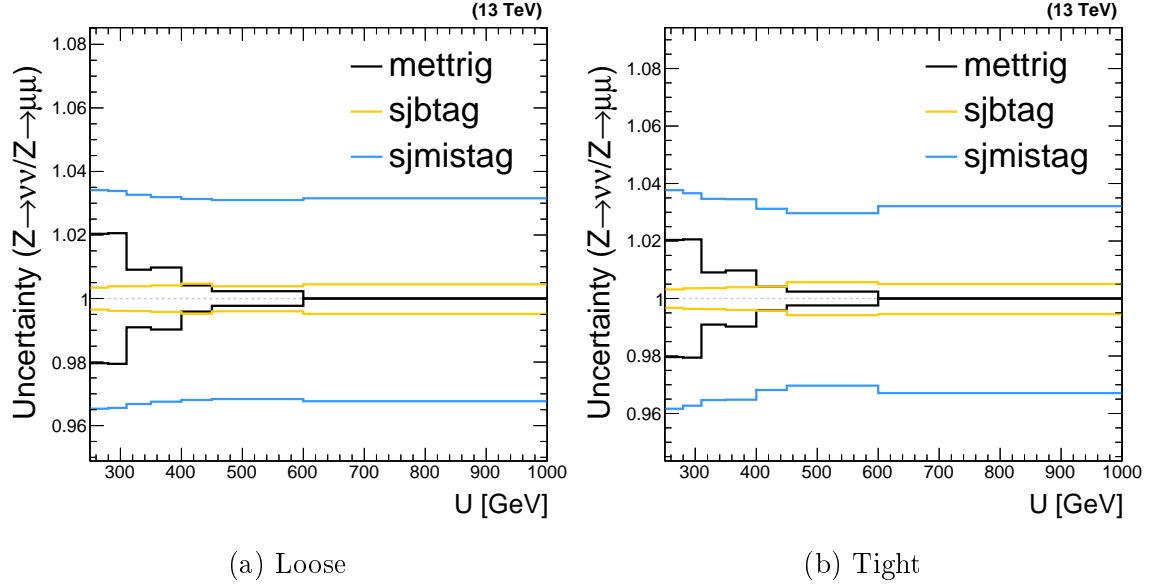
Figure 1.9: Shape uncertainties affecting $T_i^{\mu\mu}$ in both categories, as a function of U.

(a) Loose

(b) Tight
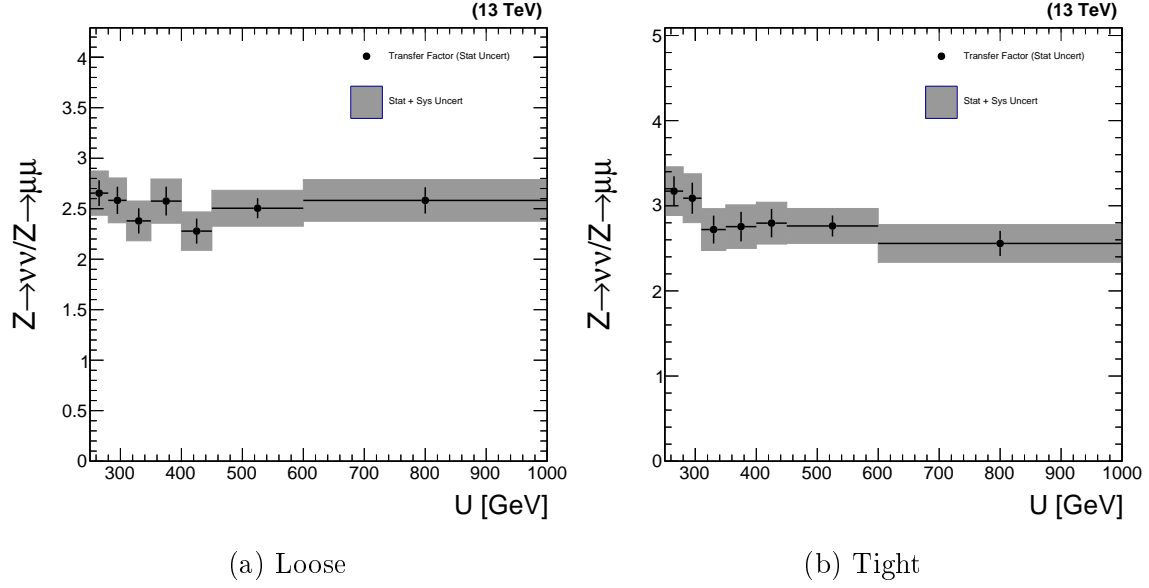
Figure 1.10: The transfer factors $T_{Z,i}^{\mu\mu}$ as a function of recoil and BDT score. The vertical black bars represent the Poisson uncertainties in the MC simulation, while the grey bands represent the sum of Poisson uncertainties and other, systematic, uncertainties. All uncertainties are represented as one standard deviation.

2. The one standard deviation variation of all uncertainties that impact $T_{Z,i}^{\mu\mu}$ are contained within a 10% envelope. This is already a factor of two smaller than the inherent $\sim 20\%$ uncertainties in the MC simulation.

To account for point (1), we can simply add more control data by also looking at $Z \rightarrow ee$ decays. Essentially all of the arguments used for the $\mu\mu$ CRs applies to the ee CRs. Figures 1.11-1.12 show the data/simulation agreement and the transfer factors for the new dielectron regions. A further set of statistical constraints to improve the estimate at high U (which is where the signals are most enhanced) is described in Section 1.2.2.

Similar methods are used to predict the W+jets and $t\bar{t}$ contributions in the SRs; these three backgrounds comprise at least 95% of the SM processes. In both cases, the momentum imbalance in the SR is a proxy for the momentum of the W boson, since the charged lepton is lost. A sketch of the event topologies is shown in Figure 1.13. Following the same arguments as used for the $Z \rightarrow \ell\ell$ CRs, we can use the hadronic recoil U in CRs that measure visible final states of W and $t\bar{t}$.

(a) U in loose CR   (b) U in tight CR   (c) m_{ee}
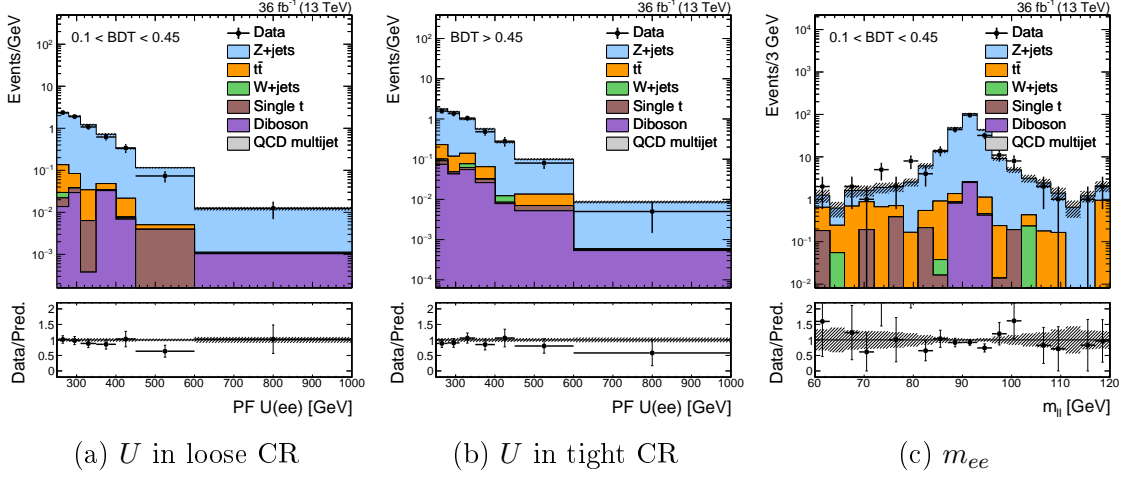
Figure 1.11: Various kinematic distributions in the two mono-top ee CRs.

(a) Loose   (b) Tight
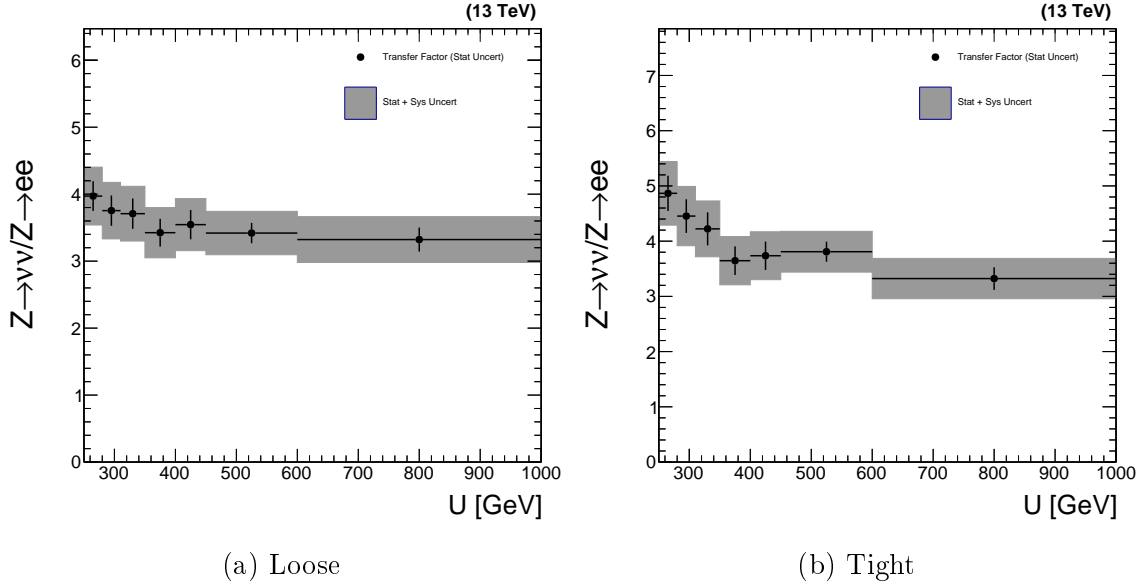
Figure 1.12: The transfer factors T_i^{ee} as a function of recoil and BDT score. The vertical black bars represent the Poisson uncertainties in the MC simulation, while the grey bands represent the sum of Poisson uncertainties and other, systematic, uncertainties. All uncertainties are represented as one standard deviation.

(a) $W \to \ell\nu$ (b) $t\bar{t} \to bqq' + b\ell\nu$
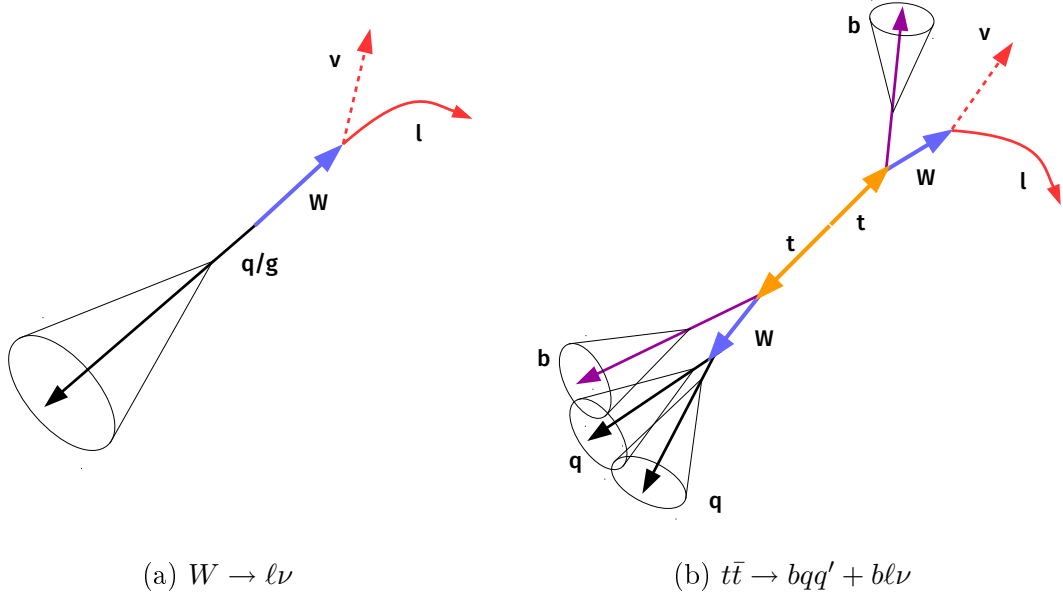
Figure 1.13: Schematic representation of the W and $t\bar{t}$ SM processes. In both cases, $U \approx p_T^W$. Furthermore, if the charged lepton is lost, $U = p_T^{\text{miss}} \approx p_T^W$.

Starting with muon final states (electrons will follow naturally), we define two sets of CRs based on the number of identified B hadrons. The selection for the $b\mu$ CRs (to measure $t\bar{t}$) is shown in Table 1.4. The selection for the μ CRs (to measure W) is shown in Table 1.5. Figures 1.14-1.15 show various kinematic distributions in these regions.

Each CR gets a set of transfer factors to constrain the targetted process in the SR: $\boldsymbol{T}_{t\bar{t}}^{b\mu}$ and $\boldsymbol{T}_W^{\mu}$. In the tight μ CR (Figures 1.15b-1.15c), the stringent top ID requirement enhances the $t\bar{t}$ and suppresses the W contribution. Since we cannot create a pure W in the tight category, we introduce an additional set of transfer factors $\boldsymbol{T}_{t\bar{t}}^{\mu}$. This extra constraint uses the $b\mu$ CRs to estimate the $t\bar{t}$ component in the μ CRs, thereby leaving only one large degree of freedom in the μ CRs. These three sets of transfer factors, and the corresponding uncertainties, are shown in Figure 1.16.

As we added the ee CRs to complement the $\mu\mu$ CRs, we also add be (e) CRs to augment the statistical power of the $b\mu$ (μ) CRs, especially at high recoil. Figures 1.17 and 1.18 respectively show some kinematic distributions and the transfer factors corresponding to these electron constraints.

Table 1.4: Criteria used to select events for the mono-top $b\mu$ CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250$ GeV	Mimicking the selection in the SR; also constrained by trigger thresholds.
1 CA15 jet with $p_{\mathrm{T}} > 250$ GeV	Same as SR
CA15 jet $110 < m_{\mathrm{SD}} < 210$ GeV	Same as SR
At least one b-tagged sub-jet	Identifying B hadron produced from hadronic top decay.
Exactly one b-tagged narrow jet	Identifying B hadron produced from leptonic top decay.
Exactly one well-identified μ	Produced from $W \to \mu\nu$
No identified e, τ_{h}	Same as SR.
No identified γ	Same as SR
$\min_{\mathrm{jets}} \Delta\phi(\mathrm{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

Table 1.5: Criteria used to select events for the mono-top μ CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250$ GeV	Mimicking the selection in the SR; also constrained by trigger thresholds.
1 CA15 jet with $p_{\mathrm{T}} > 250$ GeV	Same as SR
CA15 jet $110 < m_{\mathrm{SD}} < 210$ GeV	Same as SR
No b-tagged sub-jets	Suppressing semi-leptonic $t\bar{t}$ decays
No b-tagged narrow jets	Suppressing semi-leptonic $t\bar{t}$ decays
Exactly one well-identified μ	Produced from $W \to \mu\nu$
No identified e, τ_{h}	Same as SR.
No identified γ	Same as SR
$\min_{\mathrm{jets}} \Delta\phi(\mathrm{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

(a) U in loose CR   (b) U in tight CR   (c) m_{SD}

Figure 1.14: Various kinematic distributions in the two mono-top $b\mu$ CRs.

(a) U in loose CR   (b) U in tight CR   (c) m_{SD}
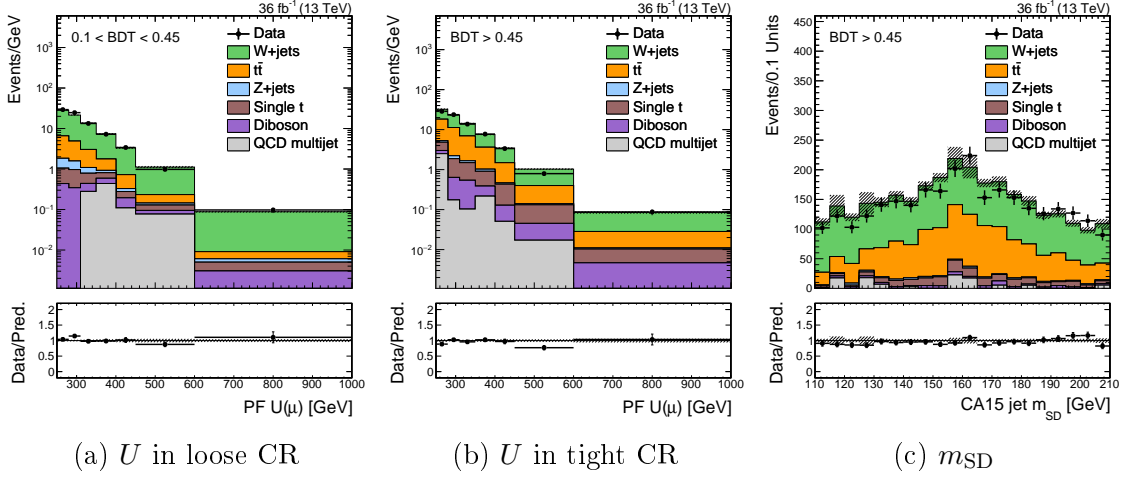
Figure 1.15: Various kinematic distributions in the two mono-top μ CRs.

Table 1.6: Uncertainties affecting the various single-muon extrapolations. "Shape" uncertainties have different priors for each bin, but are assumed to be correlated across bins.

Uncertainty	1 s.d. $(\boldsymbol{T}_{t\bar{t}}^{b\mu})$	1 s.d. $(\boldsymbol{T}_{W}^{\mu})$	1 s.d. $(\boldsymbol{T}_{t\bar{t}}^{\mu})$	Notes
μ ID	1%	1%	1%	
μ track	0.5%	0.5%	0.5%	
τ_{h} veto	3%	3%	3%	
W+heavy flavor		3%		
Trigger	0-2%	0-2%	0-2%	Shape
b-tag	2%	$\sim 0.5\%$	3-6%	Shape
$udcsg$-mistag	1%	5%	1%	Shape

(a) Loose $\boldsymbol{T}_{t\bar{t}}^{b\mu}$

(b) Tight $\boldsymbol{T}_{t\bar{t}}^{b\mu}$

(c) Loose $\boldsymbol{T}_{W}^{\mu}$

(d) Tight $\boldsymbol{T}_{W}^{\mu}$

(e) Loose $\boldsymbol{T}_{t\bar{t}}^{\mu}$

(f) Tight $\boldsymbol{T}_{t\bar{t}}^{\mu}$
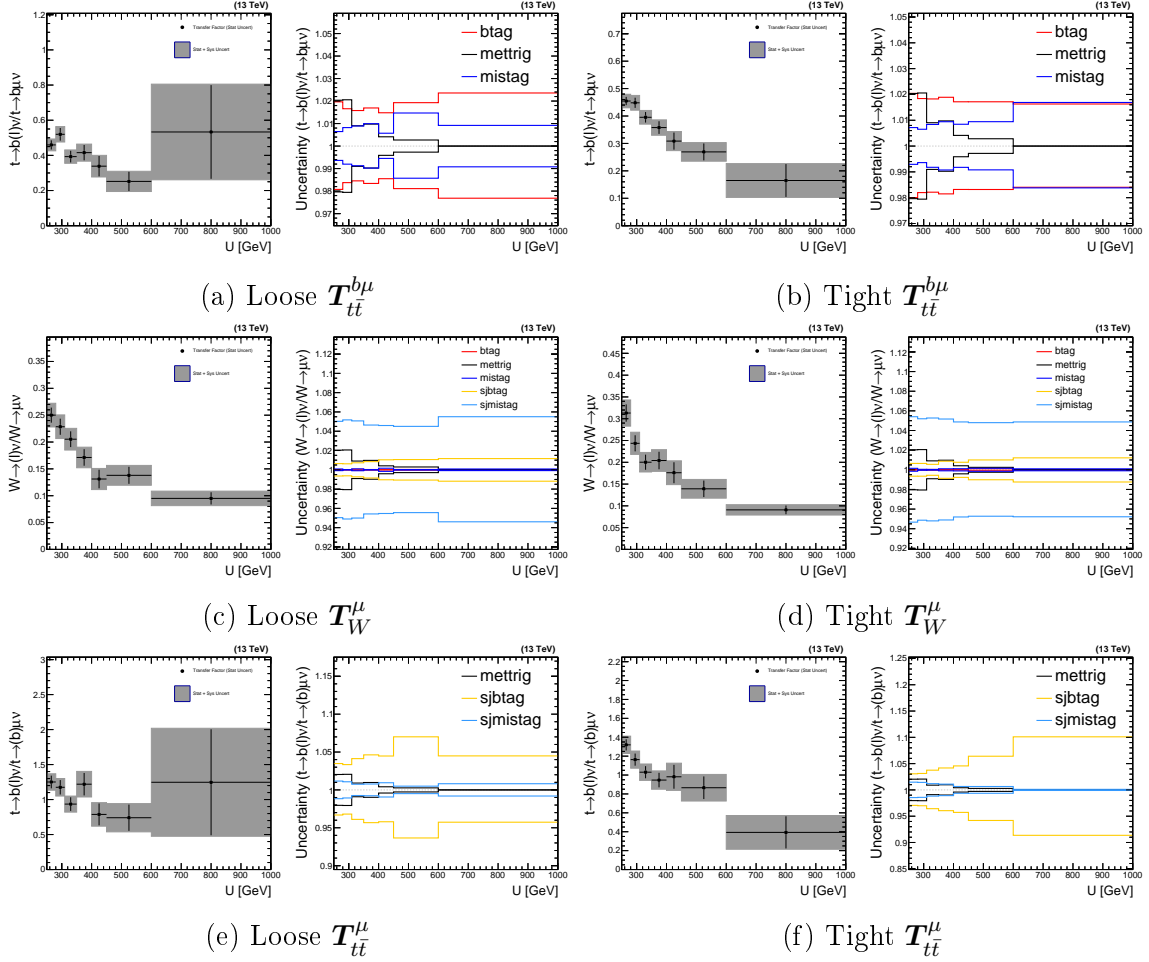
Figure 1.16: The transfer factors $\boldsymbol{T}_{t\bar{t}}^{b\mu}$, $\boldsymbol{T}_{W}^{\mu}$, and $\boldsymbol{T}_{t\bar{t}}^{\mu}$; and corresponding shape uncertainties.

(a) U in loose be CR   (b) U in tight be CR   (c) m_{SD} in tight be CR

(d) U in loose e CR   (e) U in tight e CR   (f) m_{SD} in tight e CR
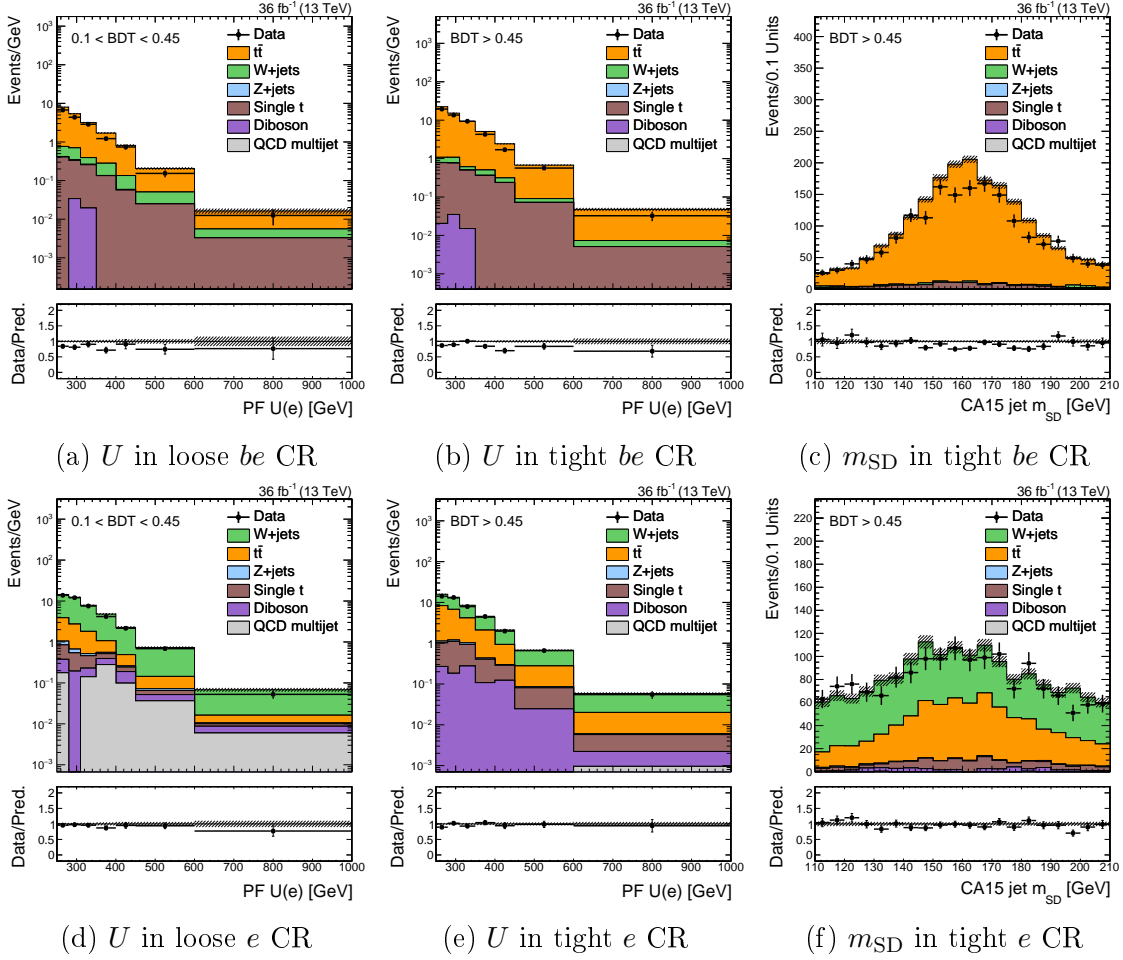
Figure 1.17: Various kinematic distributions in the mono-top be CRs (top) and e CRs (bottom).

(a) Loose $\boldsymbol{T}_{t\bar{t}}^{be}$

(b) Loose $\boldsymbol{T}_{W}^{e}$

(c) Loose $\boldsymbol{T}_{t\bar{t}}^{e}$

(d) Tight $\boldsymbol{T}_{t\bar{t}}^{be}$

(e) Tight $\boldsymbol{T}_{W}^{e}$

(f) Tight $\boldsymbol{T}_{t\bar{t}}^{e}$
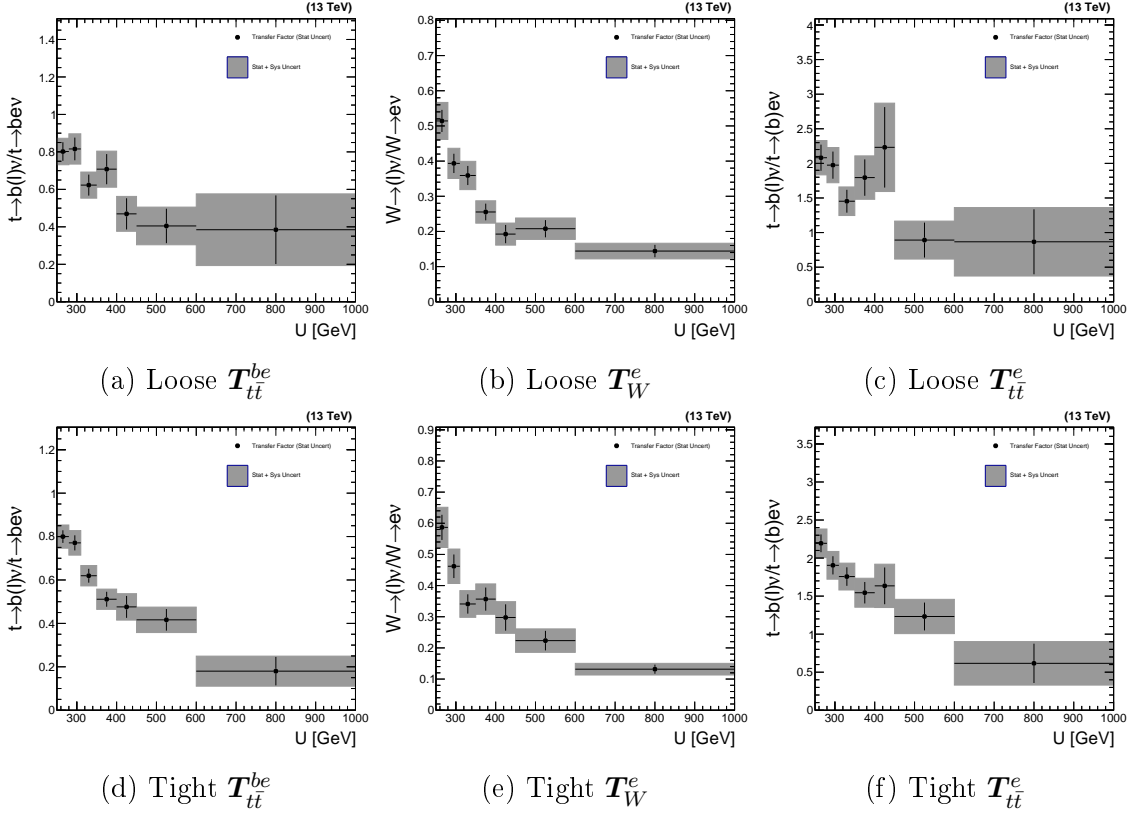
Figure 1.18: The transfer factors $\boldsymbol{T}_{t\bar{t}}^{be}$, $\boldsymbol{T}_{W}^{e}$, and $\boldsymbol{T}_{t\bar{t}}^{e}$

Having defined (almost all of) the CRs and transfer factors, we can write down a complete likelihood for the mono-top search:

$$
\mathcal{L}\left(\boldsymbol{d} \mid \mu, \boldsymbol{\mu}_{\mathrm{SR}}^{Z \to \nu\nu}, \boldsymbol{\mu}_{\mathrm{SR}}^{W \to \ell\nu}, \boldsymbol{\mu}_{\mathrm{SR}}^{t\bar{t}}, \boldsymbol{\theta}\right) =
$$

$$
\prod_{i \in \mathrm{bins}} \left[\mathrm{Pois}\left(d_i^{\mathrm{SR}} \,\Big|\, \mu S_i^{\mathrm{SR}}(\boldsymbol{\theta}) + \mu_{\mathrm{SR},i}^{Z \to \nu\nu} + \mu_{\mathrm{SR},i}^{W \to \ell\nu} + \mu_{\mathrm{SR},i}^{t\bar{t}} + B_i^{\mathrm{SR}}(\boldsymbol{\theta}) \right) \right.
$$

$$
\times \prod_{X=\mu\mu,ee} \mathrm{Pois}\left(d_i^X \,\Big|\, \frac{\mu_{\mathrm{SR},i}^{Z \to \nu\nu}}{T_{Z,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right)
$$

$$
\times \prod_{X=b\mu,be} \mathrm{Pois}\left(d_i^X \,\Big|\, \frac{\mu_{\mathrm{SR},i}^{t\bar{t}}}{T_{t\bar{t},i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right)
$$

$$
\left. \times \prod_{X=\mu,e} \mathrm{Pois}\left(d_i^X \,\Big|\, \frac{\mu_{\mathrm{SR},i}^{W \to \ell\nu}}{T_{W,i}^X(\boldsymbol{\theta})} + \frac{\mu_{\mathrm{SR},i}^{t\bar{t}}}{T_{t\bar{t},i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right) \right] \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \quad (1.8)
$$

1.2.2 Theoretically-limited extrapolations

Despite the combination of the $\mu\mu$ and ee regions, there are still large statistical uncertainties in the estimate of $Z \to \nu\nu$ at high U. This is apparent in Figure 1.11, in which exactly one event is observed in the last bin of the tight CR. The dilepton CRs are limited by $\sigma(pp \to Z \to \nu\nu) > \sigma(pp \to Z \to \ell^+\ell^-)$; accordingly, to alleviate this limitation, we look to a process with a much bigger cross-section. In similar regions of final-state phase space, $\sigma(pp \to \gamma + \mathrm{jets}) \sim 30 \times \sigma(pp \to Z(\to \nu\nu) + \mathrm{jets})$. Therefore, it is natural to use the γ+jet production spectrum as a way to estimate the Z+jet spectrum. As before, let us define another transfer factor:

$$
T_{\gamma,i}^\gamma = \frac{N_i^{\mathrm{SR}}(Z \to \nu\nu)}{N_i^\gamma(\gamma)} \quad (1.9)
$$

However, unlike the $\boldsymbol{T}_Y^X$ we have discussed so far (which correlate similar processes, e.g. $Z \to \nu\nu / Z \to \mu\mu$), $\boldsymbol{T}_\gamma^\gamma$ is highly sensitive to the theoretical predictions of the Z and γ spectra.

To reduce the impact of higher-order effects on $\boldsymbol{T}_\gamma^\gamma$, we ensure that the numerator and denominator are predicted to as high an order as possible. Leading-order MC is used to simulate V+jet processes, because higher-order simulations are much more

computationally intensive. We therefore choose to produce a less accurate LO simulation, as opposed to a more accurate, but statistically-limited, NLO simulation. While producing sufficient NLO simulation for the analysis is prohibitive, we can compute certain inclusive distributions at NLO. Since $U \approx p_{\mathrm{T}}^V$ is the quantity of interest in this analysis, we want to ensure this distribution is accurately predicted. It is clear from Figure 1.19 that adding an additional QCD order induces large corrections, both at low and high p_{T}^V. In the LO simulation, we can obtain an estimate of the uncertainty due to NLO effects by varying the renormalization and factorization scales (μ_R, μ_F) by factors of two. This is represented by the red envelope and grey band in Figure 1.19 and clearly is insufficient to cover NLO effects. Therefore, we compute a simple correction for the NLO QCD effects, known as a k-factor:

$$k_{Z,\mathrm{QCD}}(p_{\mathrm{T}}^Z) = \frac{d\sigma_{\mathrm{NLO\ QCD}}(Z)/dp_{\mathrm{T}}^Z}{d\sigma_{\mathrm{LO}}(Z)/dp_{\mathrm{T}}^Z} \tag{1.10}$$

We include another k-factor, k_{EWK}, to correct for higher-order EWK effects. Unlike k_{QCD}, k_{EWK} is derived using a theoretical calculation [5, 6, 9] instead of NLO simulation. k_{EWK} covers NLO EWK terms, as well as large Sudakov logarithms that appear at high p_{T}^V in the NNLO expansion (NLL).

Figure 1.20 compares the k-factors for all three V+jet processes. While there are similar trends as a function of p_{T}^V, it is clear that the corrections are quite different for each process. Therefore, transfer factors like $\boldsymbol{T}_\gamma^\gamma$ are strongly sensitive to NLO effects, i.e.:

$$T_{\gamma,i}^\gamma = \frac{N_i^{\mathrm{SR}}(Z \to \nu\nu)}{N_i^\gamma(\gamma)} \neq \frac{N_i^{\mathrm{SR,LO}}(Z \to \nu\nu)}{N_i^{\gamma,\mathrm{LO}}(\gamma)} \tag{1.11}$$

The distributions in Sections 1.1-1.2 are all corrected using these k-factors. Figure 1.21 shows the equivalent for CRs that target γ+jet events, and Table 1.7 describes the selection used to define these CRs.

Now that we can describe $\boldsymbol{T}_\gamma^\gamma$ at NLO, we must assess the impact of unknown higher-order terms on the transfer factors. We account for variations caused by uncertainties in the PDF model by taking the RMS of the 100 parameter variations prescribed for the NNPDF3.0 set [1]. By varying μ_F and μ_R by factors of 0.5 and
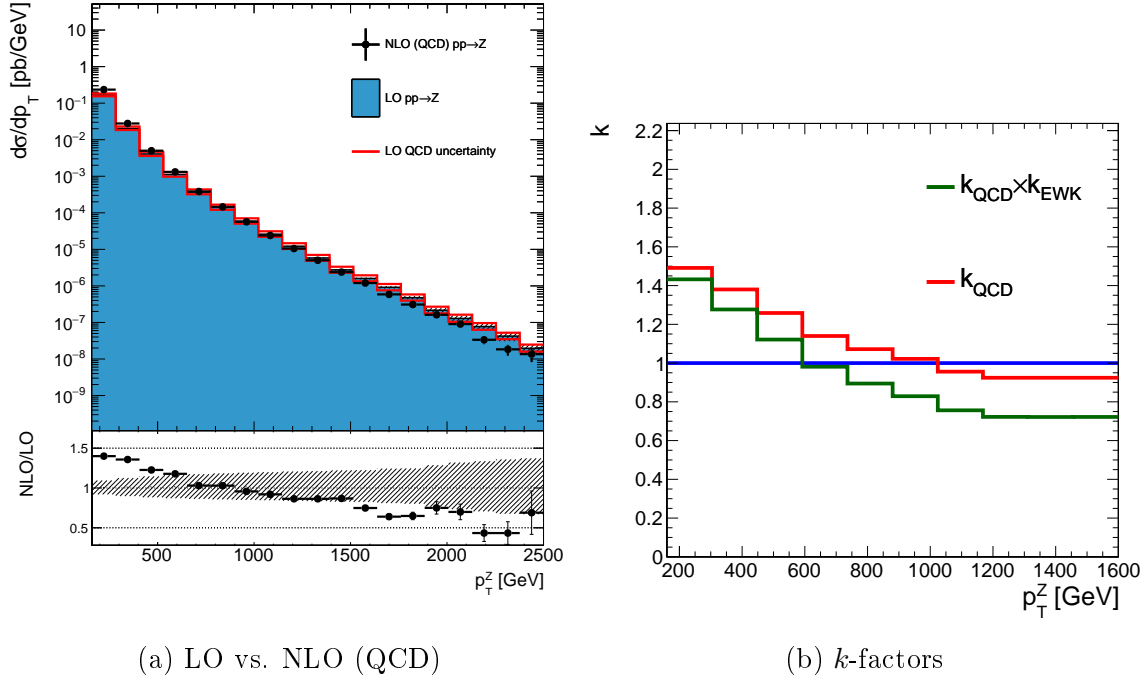
(a) LO vs. NLO (QCD)

(b) k-factors

Figure 1.19: Theoretical predictions for p_T^Z in Z+jet events and the corresponding k-factors. No detector simulation is applied in these figures; all quantities are directly from MC simulation of the physics process. "LO QCD uncertainty" refers to an estimate of the effect of the QCD renormalization and factorization scales on the LO simulation. The grey band in the ratio is the quadrature sum of the QCD and statistical uncertainties.
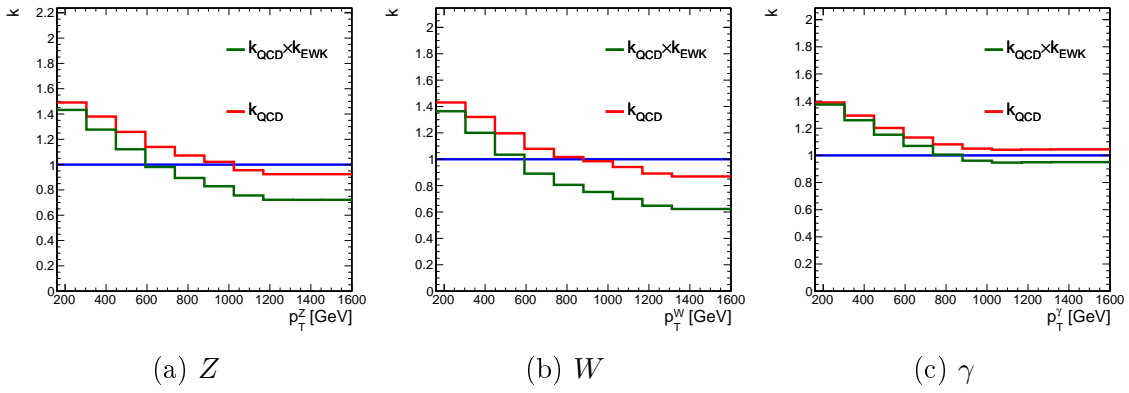

(a) Z

(b) W

(c) γ

Figure 1.20: Differential k-factors for each of the V+jet processes.

(a) U in loose CR   (b) U in tight CR   (c) p_T^γ
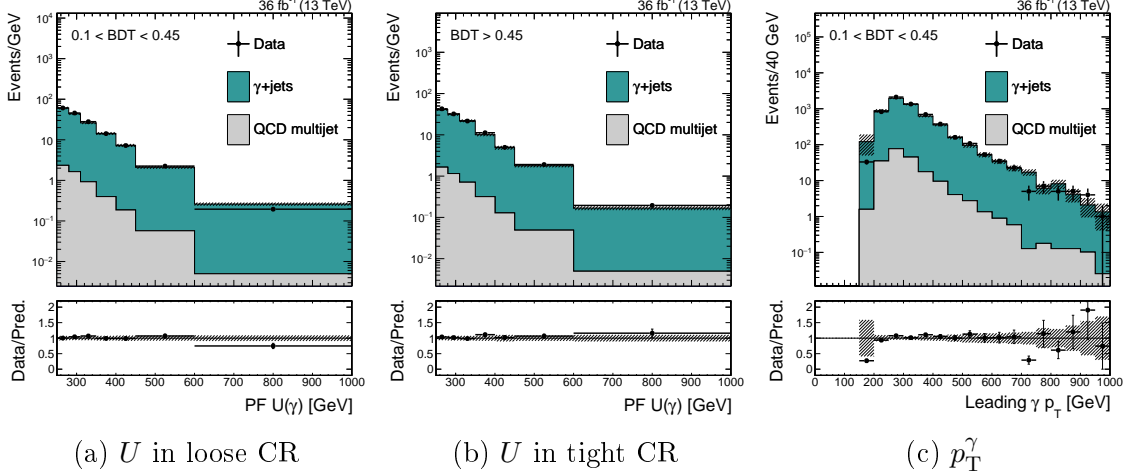
Figure 1.21: Various kinematic distributions in the two mono-top γ CRs.

Table 1.7: Criteria used to select events for the mono-top γ CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250$ GeV	Mimicking the selection in the SR
1 CA15 jet with $p_T > 250$ GeV	Same as SR
CA15 jet $110 < m_{SD} < 210$ GeV	Same as SR
No identified μ, e, τ_h	Same as SR.
Well-identified γ with $p_T^\gamma > 175$ GeV	High-p_T photon, set by trigger threshold
$\min_{jets} \Delta\phi(jet, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

39

2, we assess the effect of the integration scale choices on $\boldsymbol{T}$. These scale and PDF uncertainties cover all unknown QCD effects on the production of electroweak bosons. To be conservative, they are assumed to be uncorrelated between processes. However, the uncertainties are correlated between all bins (i.e. as a function of p_T^V). A second set of uncertainties is included for higher-order EWK effects, following what is suggested in References [2, 3, 4, 5, 6, 7, 8, 9, 10] and agreed upon in the LHC Dark Matter Working Group. These EWK uncertainties break down into three categories:

- Unknown Sudakov logarithms in the NLL correction. These uncertainties are correlated across processes (Z, W, γ).

- Missing NNLO EWK effects not covered by the NLL correction. These are not correlated across processes.

- The full difference between the NLL correction and an exponentiation of the NLO correction; also not correlated across processes.

It should be stressed that while these uncertainties apply to the prediction of each V+jet processes, they do not affect transfer factors that correlate processes differing only in decay mode or acceptance. This is simply because these uncertainties primarily deal with the initial state or the production of an electroweak boson, which is not related to the description of the decay to leptons or the experimental identification of leptons and b-jets. That is:

$$T_{Z,i}^{\mu\mu} = \frac{N_i^{\mathrm{SR}}(Z \to \nu\nu)}{N_i^{\mu\mu}(Z \to \mu\mu)} \approx \frac{N_i^{\mathrm{SR,LO}}(Z \to \nu\nu)}{N_i^{\mu\mu,\mathrm{LO}}(Z \to \mu\mu)} \tag{1.12}$$

Now that we have tools to construct transfer factors of the form $N(V)/N(V')$ with reasonably small uncertainties (i.e. smaller than the statistical uncertainty of the data), it is natural to add another transfer factor to our toolbox:

$$T_{Z/W,i}^{\mathrm{SR}} = \frac{N_i^{\mathrm{SR}}(Z \to \nu\nu)}{N_i^{\mathrm{SR}}(W \to \ell\nu)} \tag{1.13}$$

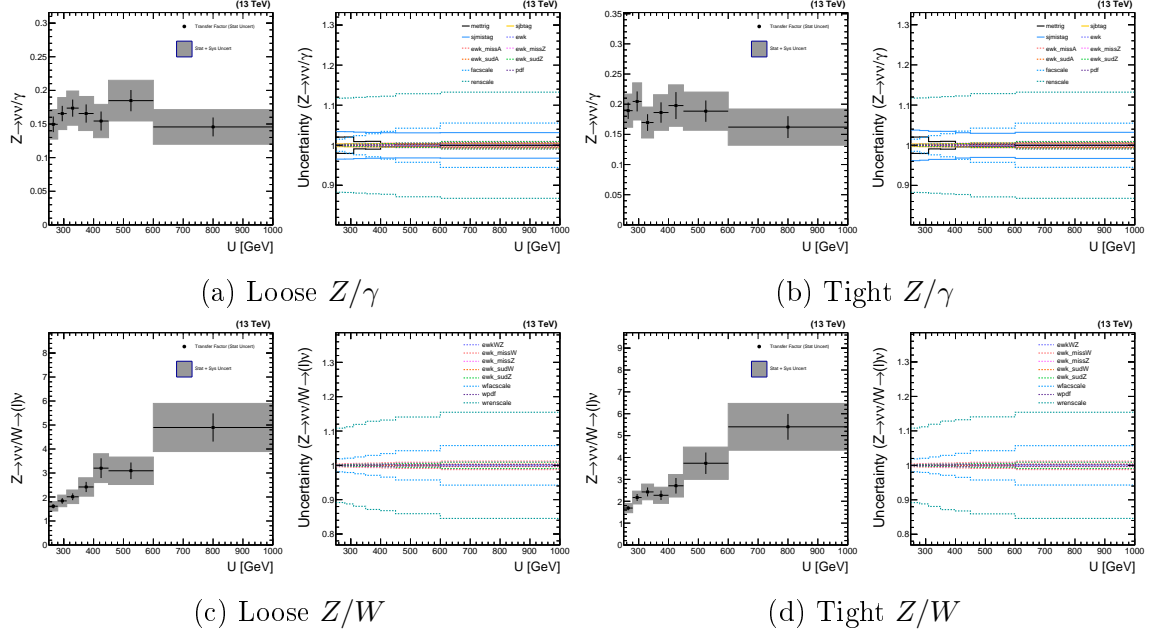This allows us to use the e, μ CRs (which target W+jet production) to further reduce

(a) Loose Z/γ (b) Tight Z/γ

(c) Loose Z/W (d) Tight Z/W

Figure 1.22: The transfer factors $\boldsymbol{T}_\gamma^\gamma$ and $\boldsymbol{T}_{Z/W}^{\mathrm{SR}}$; and corresponding shape uncertainties.

the uncertainty in the estimation of $Z \to \nu\nu$ in the SR. For technical reasons, the transfer factor is defined as the ratio Z/W in the SR. However, the SR and the μ CR are connected through a product of transfer factors:

$$N_i^{\mathrm{SR}}(Z \to \nu\nu) = T_{Z/W,i}^{\mathrm{SR}}(\hat{\boldsymbol{\theta}}) \times T_{W,i}^\mu(\hat{\boldsymbol{\theta}}) \times N_i^\mu(W \to \ell\nu) \qquad (1.14)$$

where $\hat{\boldsymbol{\theta}}$ is the maximum-likelihood estimate of $\boldsymbol{\theta}$.

Figure 1.22 shows these additional transfer factors and their shape uncertainties. It is clear from inspection that $\boldsymbol{T}_\gamma^\gamma \ll 1$, and the same holds for the effective transfer factor $\boldsymbol{T}_{Z/W}^{\mathrm{SR}} \times \boldsymbol{T}_W^\mu$. This indicates that the CR data to which the transfer factor is linked has greater statistical power than the SR data.

41

Having included all of these components, the likelihood can be written as:

$$\mathcal{L}\left(\boldsymbol{d} \mid \mu, \boldsymbol{\mu}_{\mathrm{SR}}^{Z \to \nu\nu}, \boldsymbol{\mu}_{\mathrm{SR}}^{t\bar{t}}, \boldsymbol{\theta}\right) =$$

$$\prod_{i \in \mathrm{bins}} \left[\mathrm{Pois}\left(d_i^{\mathrm{SR}} \mid \mu S_i^{\mathrm{SR}}(\boldsymbol{\theta}) + \mu_{\mathrm{SR},i}^{Z \to \nu\nu} + \frac{\mu_{\mathrm{SR},i}^{Z \to \nu\nu}}{T_{Z/W,i}^{\mathrm{SR}}(\boldsymbol{\theta})} + \mu_{\mathrm{SR},i}^{t\bar{t}} + B_i^{\mathrm{SR}}(\boldsymbol{\theta}) \right) \right.$$

$$\times \prod_{X=\mu\mu,ee} \mathrm{Pois}\left(d_i^X \mid \frac{\mu_{\mathrm{SR},i}^{Z \to \nu\nu}}{T_{Z,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right)$$

$$\times \prod_{X=b\mu,be} \mathrm{Pois}\left(d_i^X \mid \frac{\mu_{\mathrm{SR},i}^{t\bar{t}}}{T_{t\bar{t},i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right)$$

$$\times \prod_{X=\mu,e} \mathrm{Pois}\left(d_i^X \mid \frac{\mu_{\mathrm{SR},i}^{Z \to \nu\nu}}{T_{W,i}^X(\boldsymbol{\theta}) T_{Z/W,i}^{\mathrm{SR}}(\boldsymbol{\theta})} + \frac{\mu_{\mathrm{SR},i}^{t\bar{t}}}{T_{t\bar{t},i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right)$$

$$\times \mathrm{Pois}\left(d_i^{\gamma} \mid \frac{\mu_{\mathrm{SR},i}^{Z \to \nu\nu}}{T_{\gamma,i}^{\gamma}(\boldsymbol{\theta})} \right) \Bigg] \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \qquad (1.15)$$

The discussion in this section has largely relied on arguments from simulation and calculation. We can, however, validate that our estimates of $\boldsymbol{T}$ and the corresponding uncertainties are reasonable by using CR data and appropriate proxies. For example, to test $\boldsymbol{T}_\gamma^\gamma \sim N(Z \to \nu\nu)/N(\gamma)$, we can look at $N(Z \to \mu\mu)/N(\gamma)$. Up to differences in branching ratio and muon identification, these ratios are identical. Figure 1.23 shows some examples of these proxy ratios. In particular, we see that the Z/γ, Z/W and W/γ ratios are well-described the MC, especially as compared to the systematic uncertainties that are assigned.

1.3   Results

1.3.1   Constraints on spin-1 FCNCs

1.3.2   Constraints on scalar resonances

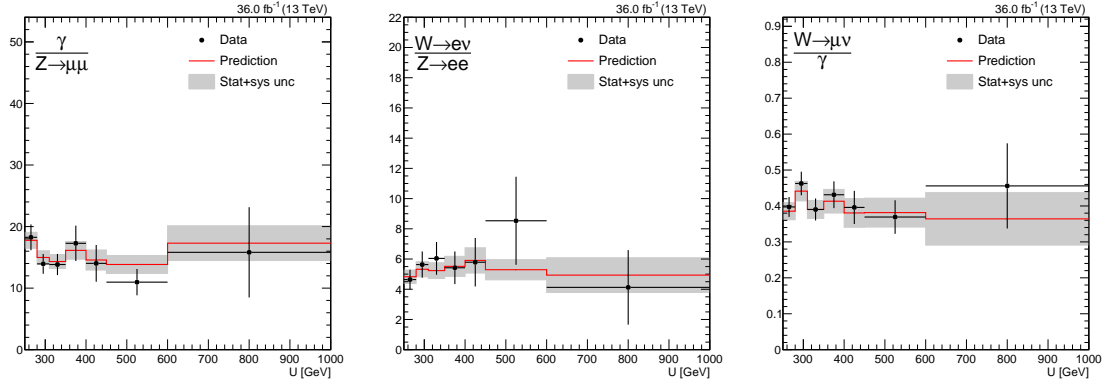### 1.3.3   Extending to new BSM models

Figure 1.23: Data validation of CR-to-CR transfer factors in the loose category. Only ratios with theoretically-limited systematic uncertainties are shown.

Bibliography

[1] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[2] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Muck. Electroweak corrections to W + jet hadroproduction including leptonic W-boson decays. *JHEP*, 08:075, 2009.

[3] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Muck. Electroweak corrections to dilepton + jet production at hadron colliders. *JHEP*, 06:069, 2011.

[4] Ansgar Denner, Stefan Dittmaier, Tobias Kasprzik, and Alexander Mück. Electroweak corrections to monojet production at the LHC. *Eur. Phys. J. C*, 73:2297, 2013.

[5] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhöfer, Stefano Pozzorini, and Marek Schönherr. NLO electroweak automation and precise predictions for W+multijet production at the LHC. *JHEP*, 04:012, 2015.

[6] Stefan Kallweit, Jonas M. Lindert, Stefano Pozzorini, Marek Schönherr, and Philipp Maierhöfer. NLO QCD+EW automation and precise predictions for V+multijet production. In *50th Rencontres de Moriond on QCD and High Energy Interactions La Thuile, Italy, March 21-28, 2015*, 2015.

[7] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Logarithmic electroweak corrections to hadronic Z+1 jet production at large transverse momentum. *Phys. Lett. B*, 609:277, 2005.

[8] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. One-loop weak corrections to hadronic production of Z bosons at large transverse momenta. *Nucl. Phys. B*, 727:368, 2005.

[9] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Electroweak corrections to hadronic photon production at large transverse momenta. *JHEP*, 03:059, 2006.

[10] Johann H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze. Electroweak corrections to hadronic production of W bosons at large transverse momenta. *Nucl. Phys. B*, 797:27, 2008.