MEASUREMENT OF THE TOP YUKAWA COUPLING WITH THE ATLAS DETECTOR AT LHC

Merve Nazlim AGARAS
Supervisor: Djamel BOUMEDIENE

JRJC, 18 Oct,2018

Outline

- Motivation
- Signatures
- 2LSS signature
  - Selection and background
  - Background estimation
  - BDT discriminant
- Fit
- Results
- MEM on going study

Standard Model

- The Standard Model (SM) of particle interactions describes the structure of ordinary matter and the fundamental interactions of nature
  - Re-normalisable, Lorentz inv QFT build upon local gauge symmetries of the Lagrangian
- The internal symmetries of the symmetry group

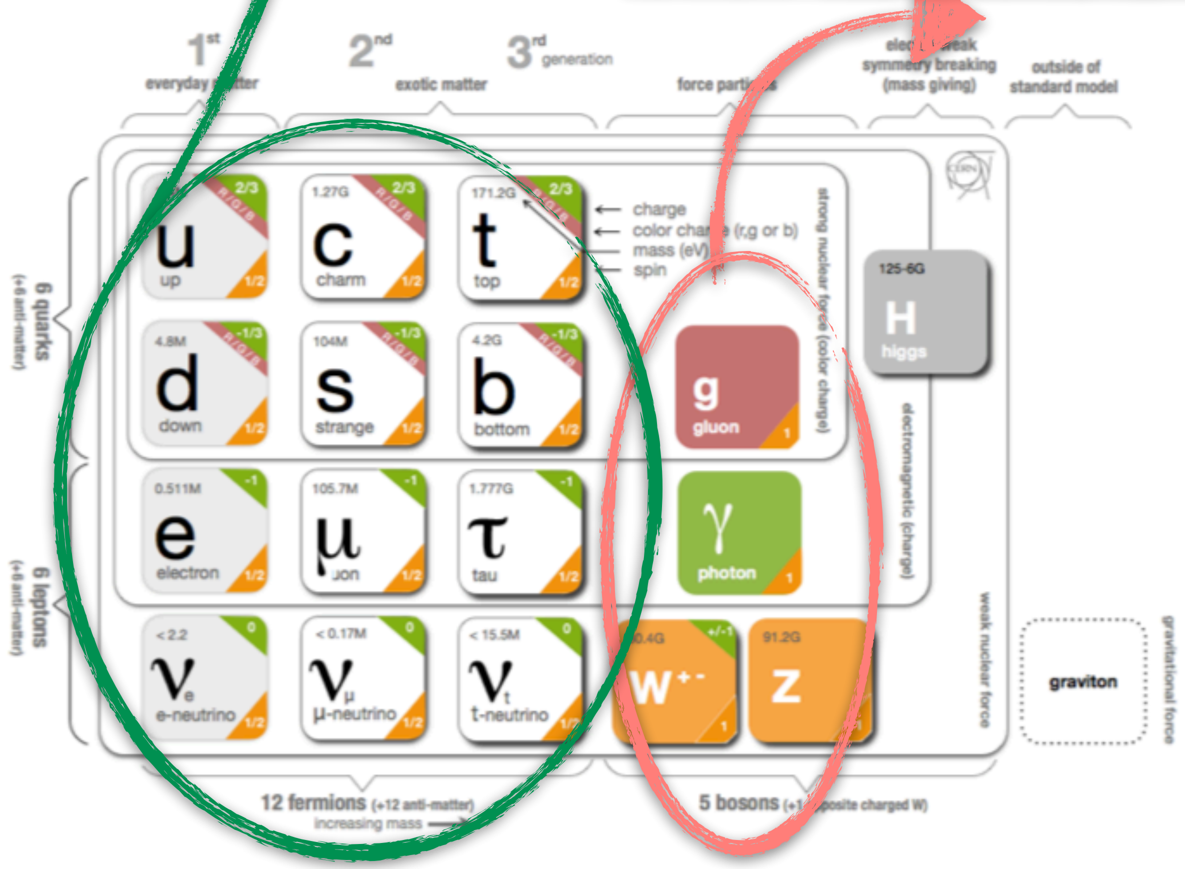
$$\mathrm{SU}\left(3\right)_C \times \underbrace{\mathrm{SU}\left(2\right)_L \times \mathrm{U}\left(1\right)_Y}$$

Strong inter.
QCD

The unified electromagnetic and weak interactions

Fundametal particles; quark & leptons

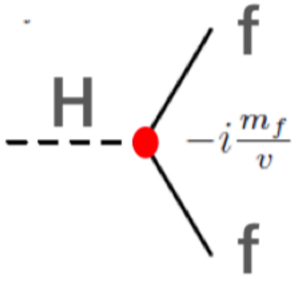
Interactions -> exchange of Gauge bosons

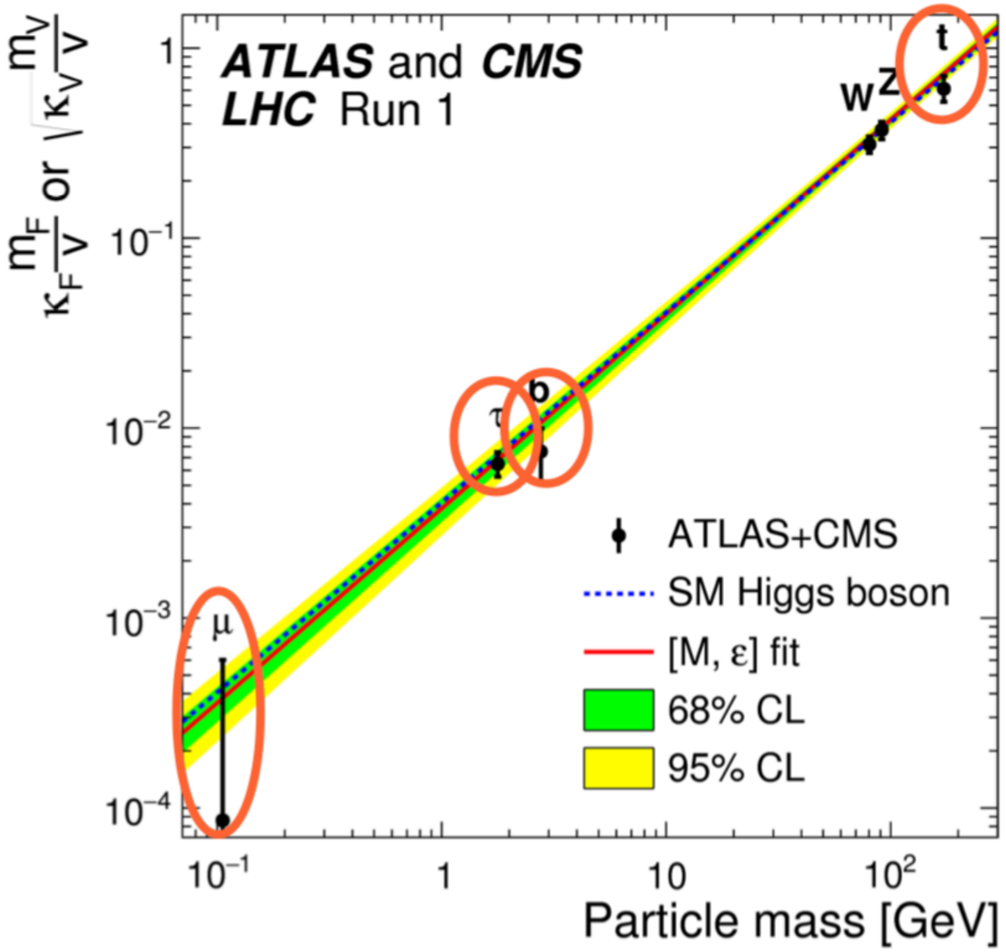


Higgs Couplings

- To give gauge boson masses in the SM without losing gauge invariance in the weak interactions, a complex scalar field is introduced (4 dof)
- fermions can also couple to the scalar field and acquire mass in a way which preserves gauge invariance in the weak interaction
  - Yukawa interaction, describe the strength of the coupling of the fermion to the Higgs field
  - proportional to the mass of the fermion
  - arbitrary and can only be measured from experiments

$$\frac{g_f}{\sqrt{2}} = \frac{m_f}{v}$$



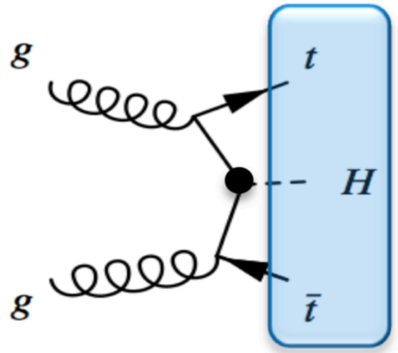


Top Yukawa Coupling at LHC

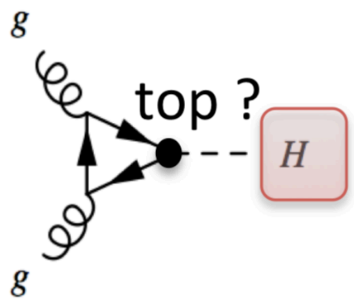
- Higgs-Top quark Yukawa coupling has a strong impact on the Theory, eg. Predicted Vacuum Stability or Instability depends strongly on y_t
- Can be determined:
  - Indirectly obtained through measurement of top quark mass or observed through SM Higgs decaying in two photons and production of Higgs by gluon-gluon fusion

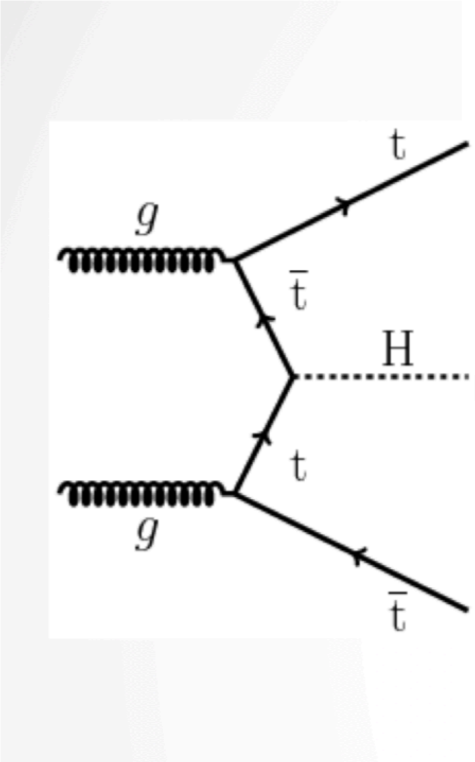
- Direct measurement possible through ttH production by calculating the x-section of the process

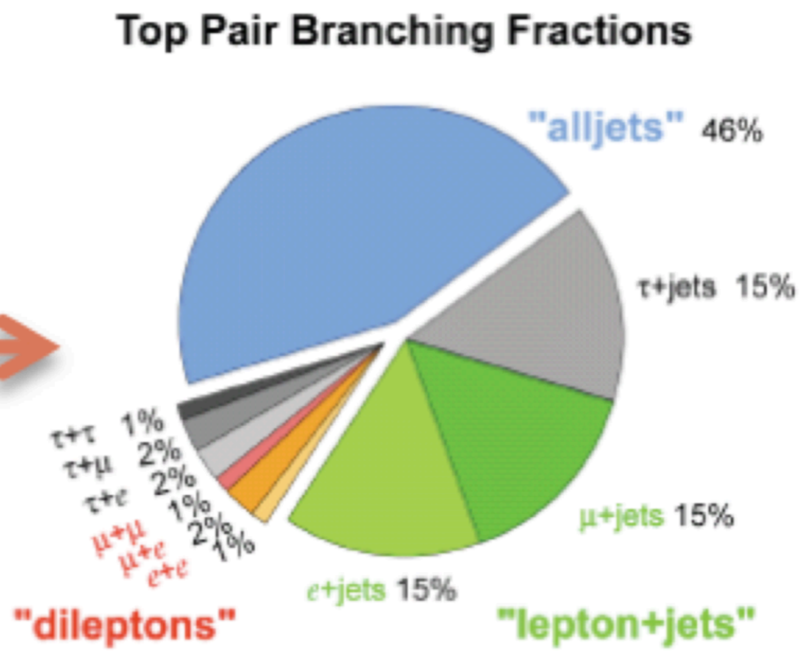
ttH (0.9%)



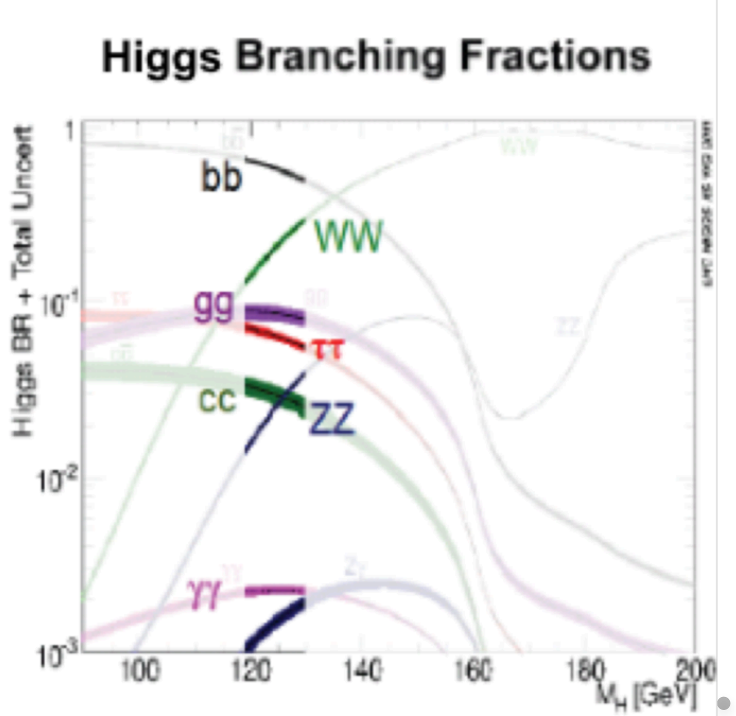
Gluon fusion (88% @13TeV pp)





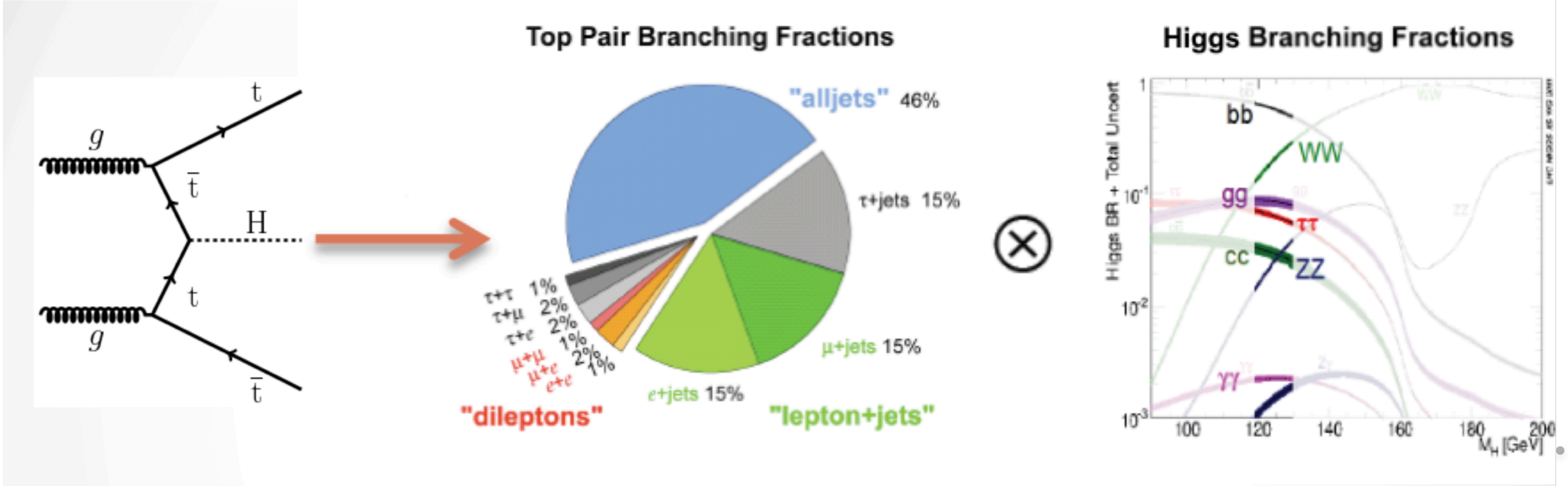


⊗



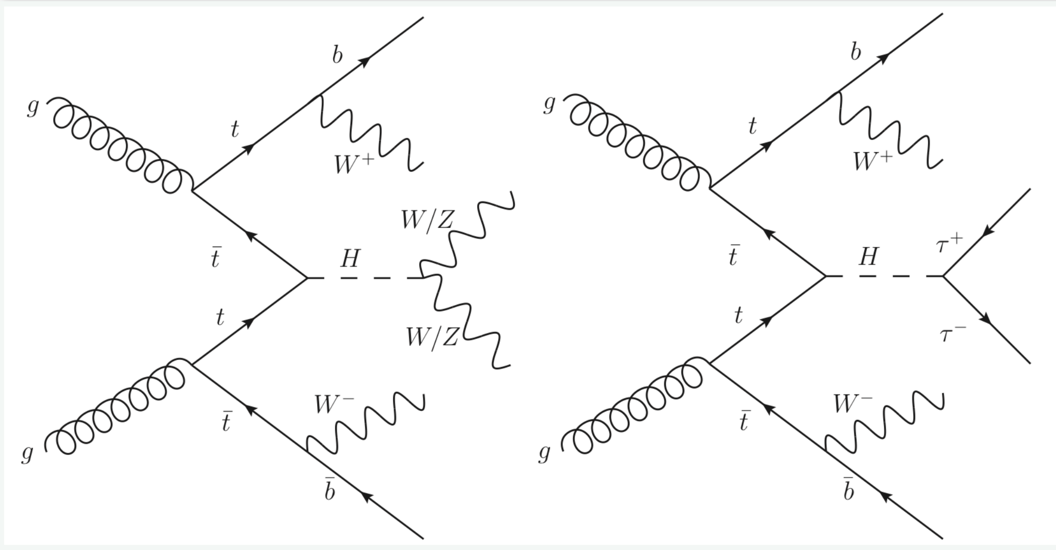
Top Yukawa Coupling at LHC

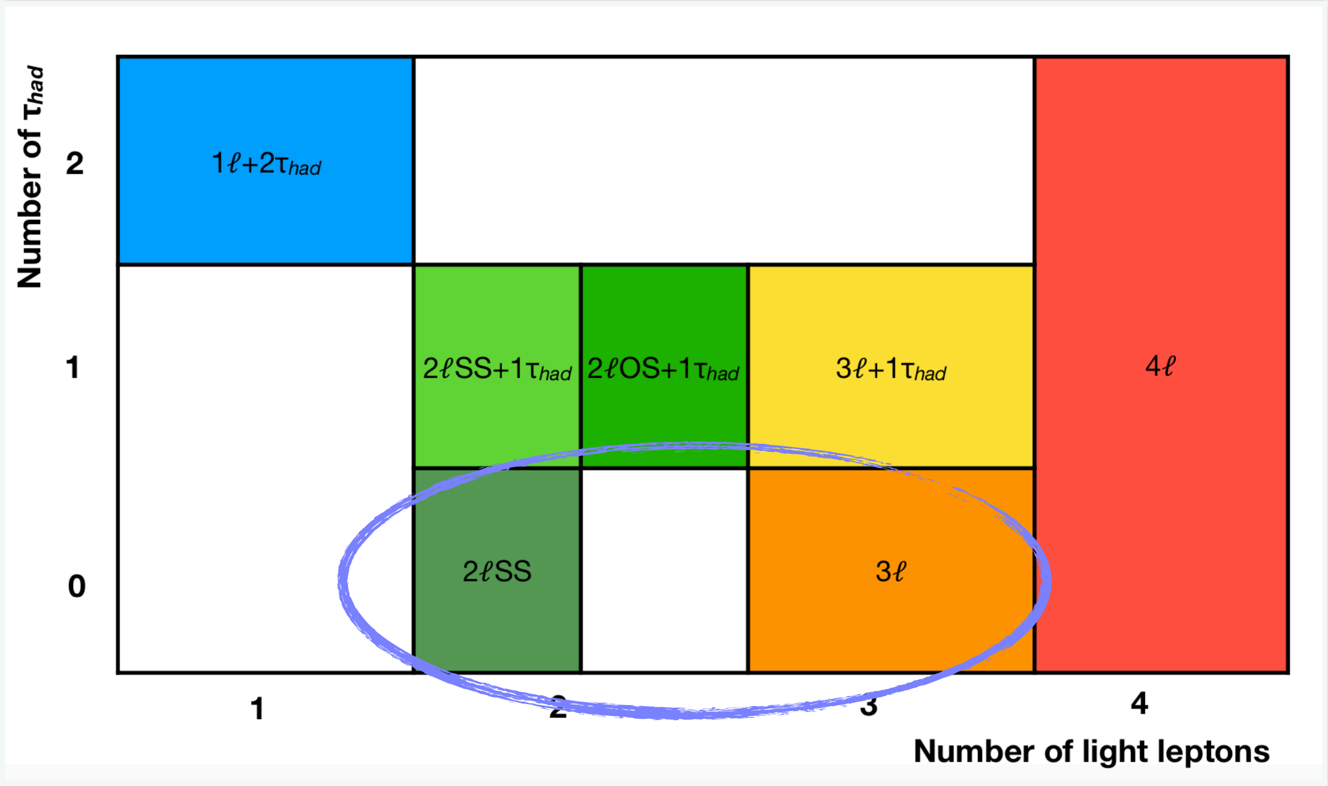
- Complex final states, with many objects: jets, b-jets, light leptons (l), hadronic taus (τhad), photons
- ttH final state combines top pair decay signature and Higgs decay signature → large number of possible final states
  - $H \rightarrow b\bar{b}$: 4b + 2W —>Final state with largest BR from Higgs but with very large background contribution ($t\bar{t}$ +jets)
  - H → WW,ττ,ZZ : 2b + multileptons —>Less background contamination
  - H → γγ : 2b + 2γ —>Very rare decay but very pure signal

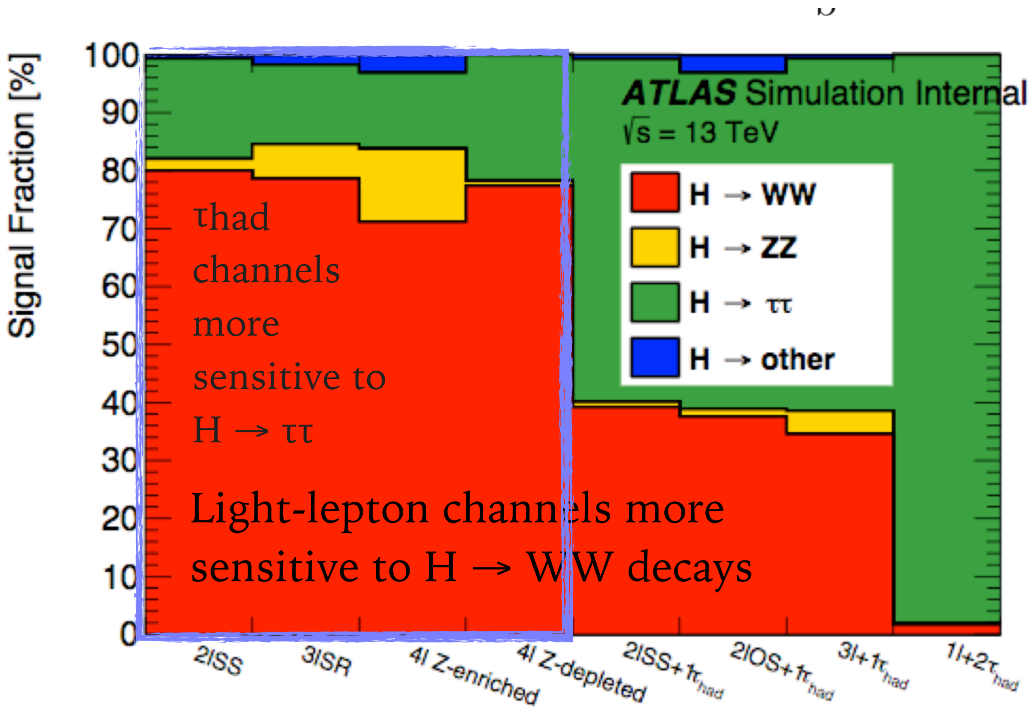


ttH-Multilepton

- Targets Higgs decays to WW, ZZ and $\tau\tau$ with ≥ 2 (1light) lepton in their final state
- Analysis channels are defined wrt light leptons (l) and hadronic taus (τ_{had}) multiplicity (7 orthogonal channels)
- High lepton multiplicity and charge requirements are chosen to suppress backgrounds
- MVA (multivariate analysis techniques) in lepton definitions to reject fakes/non-prompt lepton
- Event classified in the different regions using MVA







τhad channels more sensitive to H → ττ

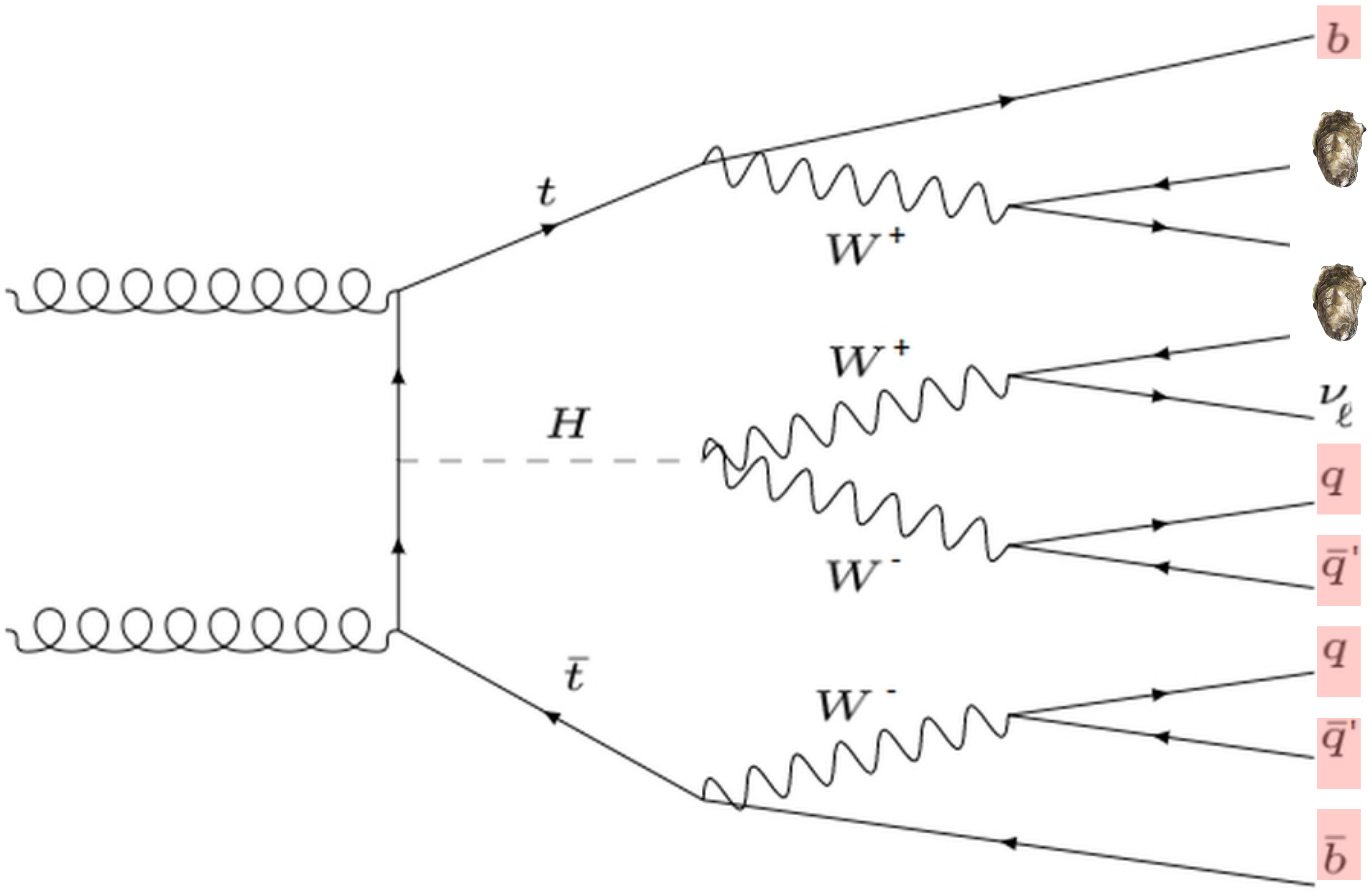
Light-lepton channels more sensitive to H → WW decays

Personal Contributions

- Coordinator of productions of the reconstructed data/MC samples to the group See backup slides..
- Fit contact of the team
- Contributions to 2LSS analysis
  - Fakes group
  - Developing a new method to discriminate the signal from background (MEM)

2L same-sign channel

- Two reconstructed light leptons with the same electric charge
- 4 jets 2 b-jets 2 leptons



✦ **Irreducible backgrounds:**

➡ **From MC (with dedicated CRs)**

- **ttW, ttZ, diboson**

✦ **Reducible backgrounds;**

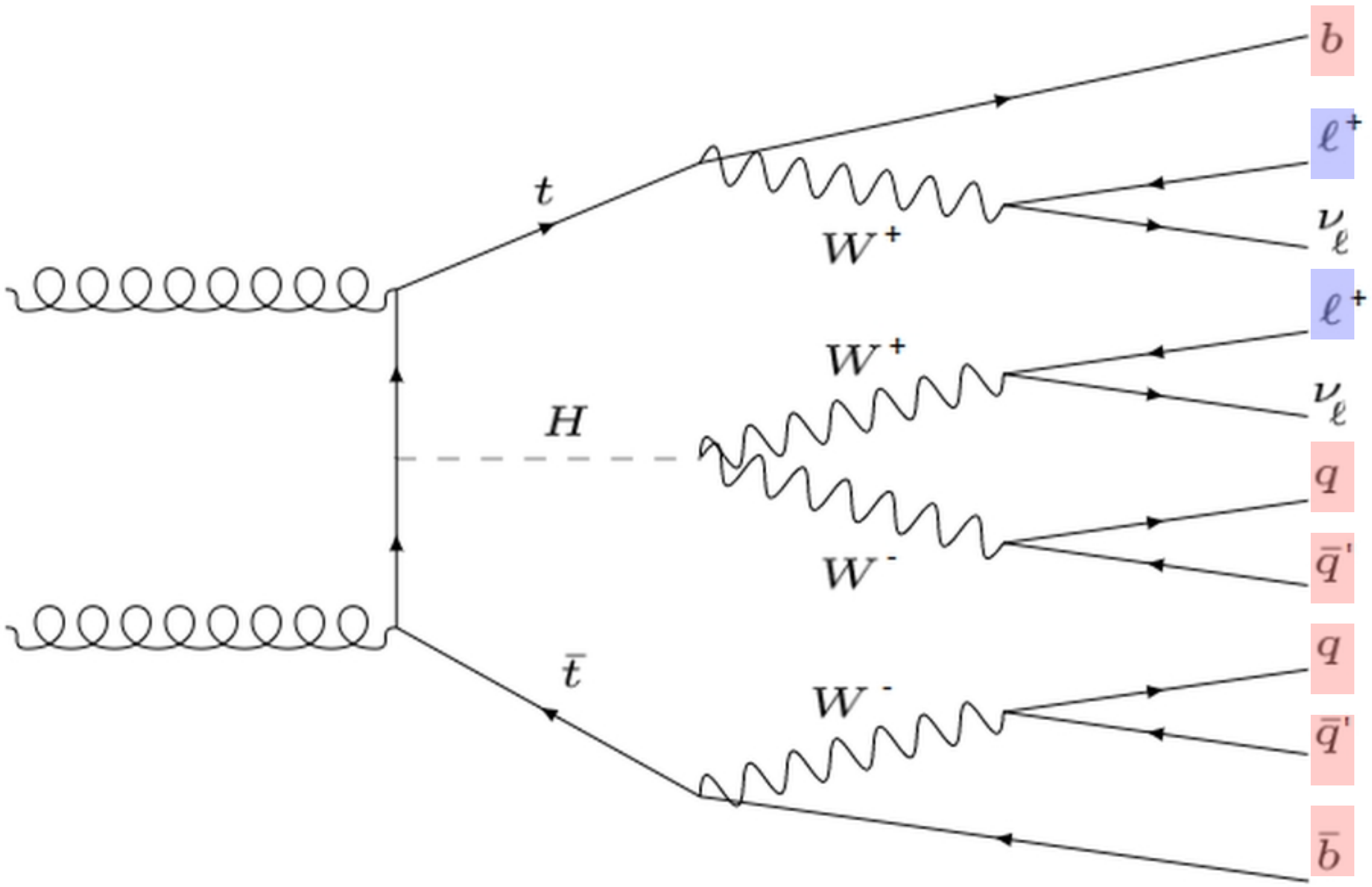
➡ **Data driven**

- Fake/**Non-prompt leptons**
- **Fake hadronic**
- Electron charge mis-identification

M.Nazlim Agaras

2L same-sign channel

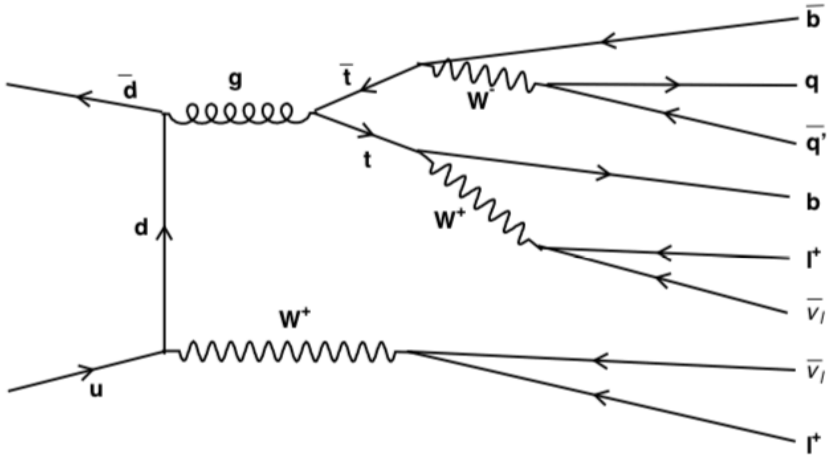
- Two reconstructed light leptons with the same electric charge
- 4 jets 2 b-jets 2 leptons



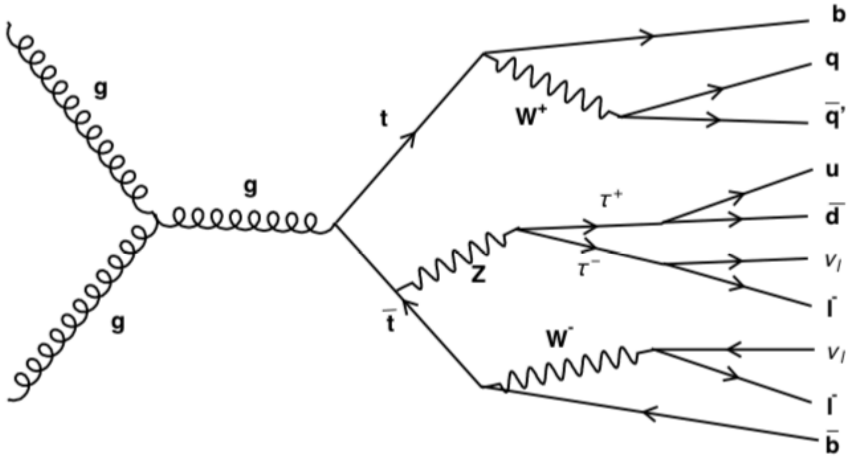
- ✦ **Irreducible backgrounds:**
  - ➡ **From MC (with dedicated CRs)**
  - **- ttW, ttZ, diboson**
- ✦ **Reducible backgrounds;**
  - ➡ **Data driven**
  - - Fake/**Non-prompt leptons**
  - - **Fake hadronic**
  - - Electron charge mis-identification

Irreducible Backgrounds

- Few SM processes with similar signatures
- True physical same-sign background: ttW , ttZ , VV estimated from MC simulation
- These background estimates are a crucial part of the analysis, because their final state and kinematics are similar to the signal



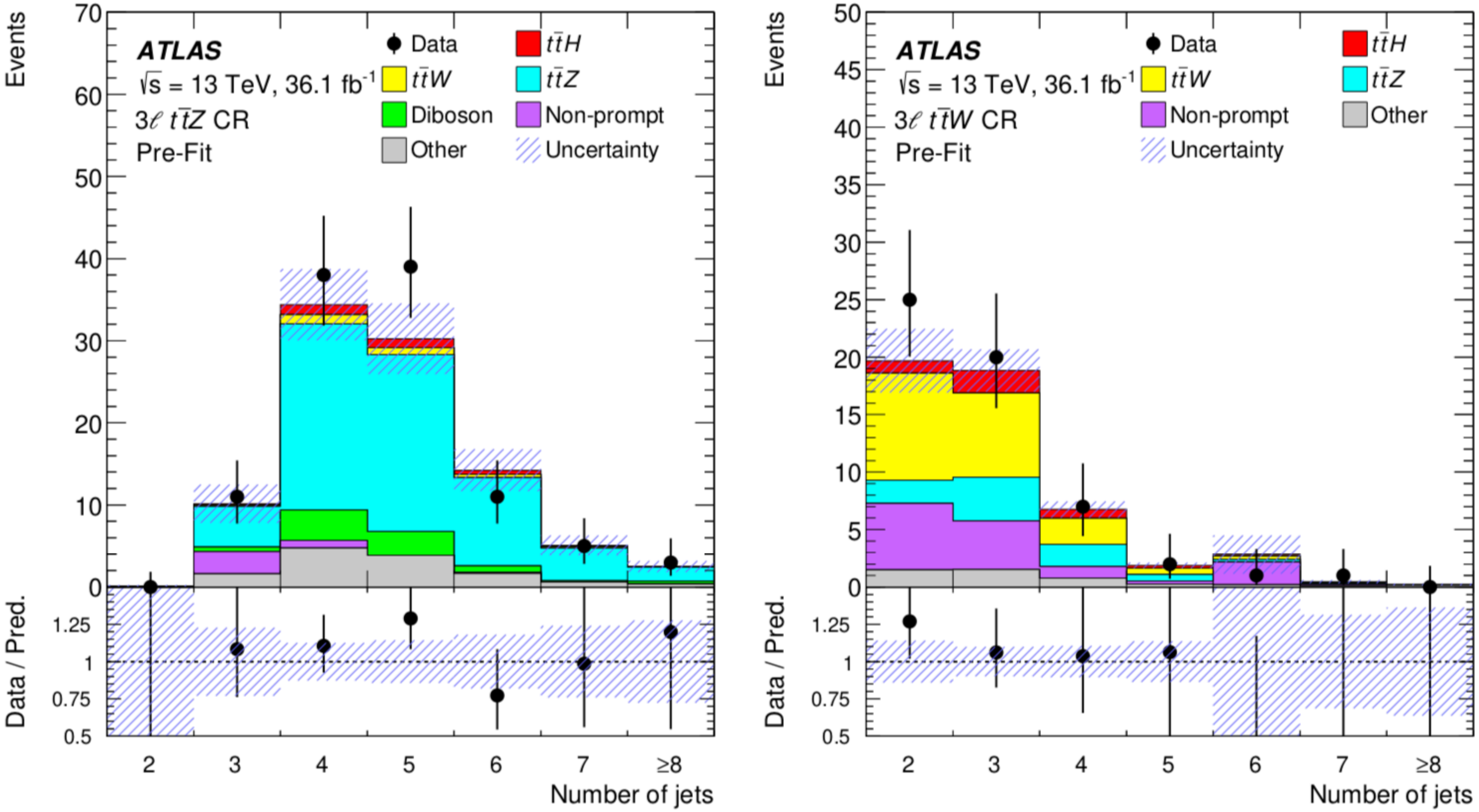
- $t\bar{t}W$ process



- $t\bar{t}Z$ process

Irreducible Backgrounds

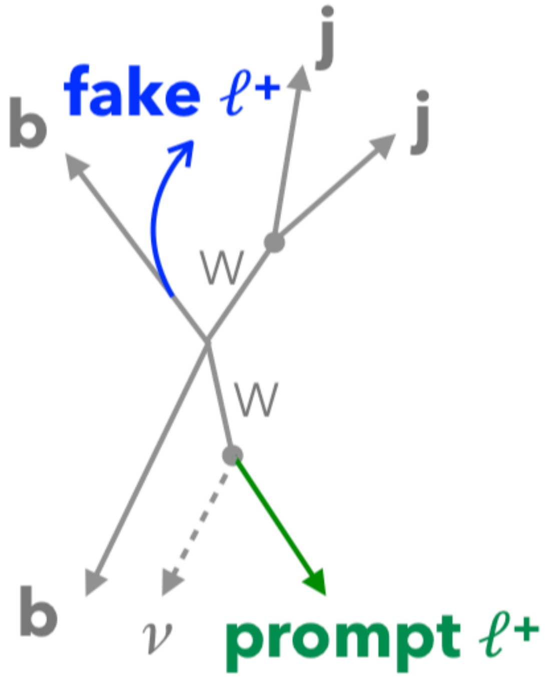
- Few SM processes with similar signatures
- True physical same-sign background: ttW , ttZ , VV estimated from MC simulation
- These background estimates are a crucial part of the analysis, because their final state and kinematics are similar to the signal



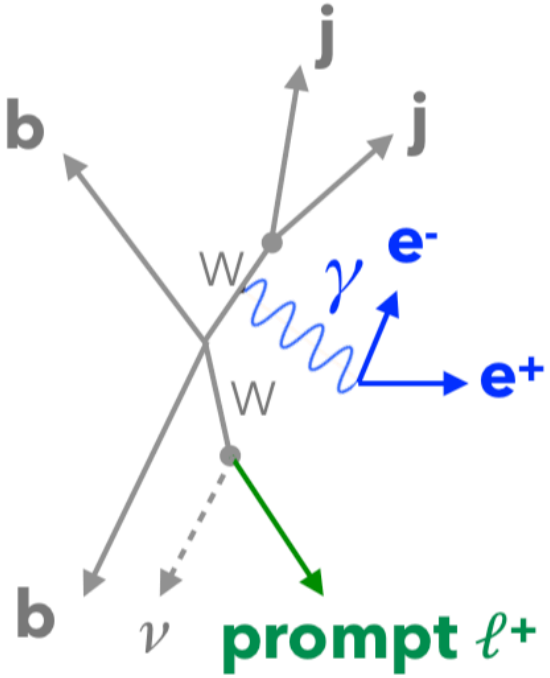
Reducible Backgrounds

- Data-driven methods are used to estimate the backgrounds with non-prompt light leptons, defining control regions enriched in such backgrounds and extrapolating the observed yields to the signal regions
- The non-prompt lepton background in the 2lSS channel is a mixture of leptons from semileptonic HF decays, conversions and charge mis-identification of electrons

Semileptonic b-decay



Photon conversions



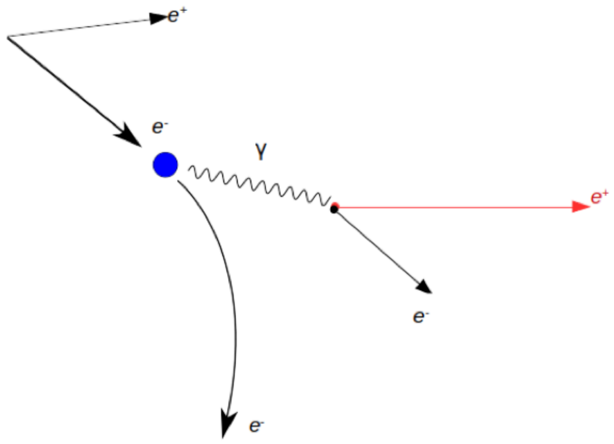
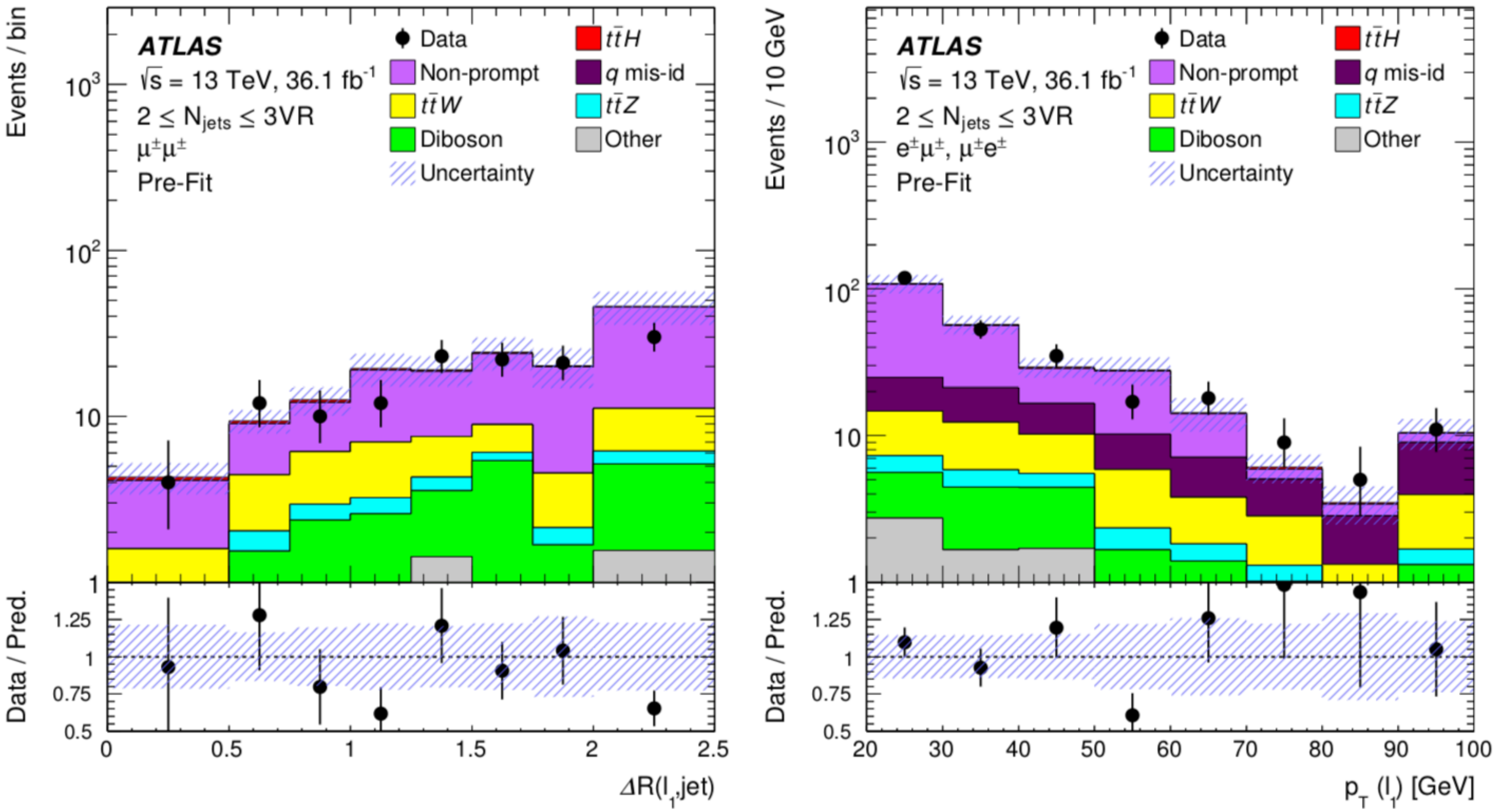


Figure: *Tight electron charge-flip rates.*

Reducible Backgrounds

- Data-driven methods are used to estimate the backgrounds with non-prompt light leptons, defining control regions enriched in such backgrounds and extrapolating the observed yields to the signal regions
- The non-prompt lepton background in the 2lSS channel is a mixture of leptons from semileptonic HF decays, conversions and charge mis-identification of electrons



Fake Estimates


Tight

Loose

- A data-driven method, called matrix method(MM)
  - estimates the number of non-prompt leptons in the signal region
    - splitting the events in four orthogonal categories (tight/anti-tight)
- The probabilities for both the loose prompt and non-prompt leptons to be tight are measured in control regions independent from SR
- These are used to estimate the number of non-prompt events in the signal regions (at least one fake lepton)

$$\begin{pmatrix} N^{TT} \\ N^{T\not{T}} \\ N^{\not{T}T} \\ N^{\not{T}\not{T}} \end{pmatrix} = \begin{pmatrix} \varepsilon_{r,1}\varepsilon_{r,2} & \varepsilon_{r,1}\varepsilon_{f,2} & \varepsilon_{f,1}\varepsilon_{r,2} & \varepsilon_{f,1}\varepsilon_{f,2} \\ \varepsilon_{r,1}\not{\varepsilon}_{r,2} & \varepsilon_{r,1}\not{\varepsilon}_{f,2} & \varepsilon_{f,1}\not{\varepsilon}_{r,2} & \varepsilon_{f,1}\not{\varepsilon}_{f,2} \\ \not{\varepsilon}_{r,1}\varepsilon_{r,2} & \not{\varepsilon}_{r,1}\varepsilon_{f,2} & \not{\varepsilon}_{f,1}\varepsilon_{r,2} & \not{\varepsilon}_{f,1}\varepsilon_{f,2} \\ \not{\varepsilon}_{r,1}\not{\varepsilon}_{r,2} & \not{\varepsilon}_{r,1}\not{\varepsilon}_{f,2} & \not{\varepsilon}_{f,1}\not{\varepsilon}_{r,2} & \not{\varepsilon}_{f,1}\not{\varepsilon}_{f,2} \end{pmatrix} \begin{pmatrix} N^{rr} \\ N^{rf} \\ N^{fr} \\ N^{ff} \end{pmatrix}$$

what we want

Promp-lepton eff. are measured as a function of p_t in CR dominated by ttbar

Fake lepton rates are measured as a function of p_T in CR corrected with the HF/Conv fraction in CR/SR from MC (aka q-scaling)

Fake Estimates


Tight

Loose

- A data-driven method, called matrix method(MM)
  - estimates the number of non-prompt leptons in the signal region
    - splitting the events in four orthogonal categories (tight/anti-tight)
- The probabilities for both the loose prompt and non-prompt leptons to be tight are measured in control regions independent from SR
- These are used to estimate the number of non-prompt events in the signal regions (at least one fake lepton)

Systematic Uncertainties

- Uncertainties of the subtracted background in the CRs
- Truth Closure
- Difference in the in the fraction of conversions from CR to SR

leptons passing the tight and loose-but-not-tight lepton selections

$$N^{f}_{TT} = w_{TT}N^{TT} + w_{\not{T}T}N^{\not{T}T} + w_{T\not{T}}N^{T\not{T}} + w_{\not{T}\not{T}}N^{\not{T}\not{T}}$$

Depend on the measured prompt and non-prompt lepton efficiencies

Charge mis-identification estimation

- Electron charge-flip in SS dilepton final states introduces background from OS events
- Two main mechanisms:
  - Trident process with an electron radiating a photon converting to a pair of electrons
  - Mis-reconstructed electron track in the Inner Detector. Becomes dominant at large pt
- Rate of QMisid computed from Z → e^+e^- mass peak region and used to reweight OS data using 3D likelihood method [pT, η, Tight/Loose]
- The contamination in the SR is estimated from the reconstructed OS data events passing SR criteria (except the SS requirement).

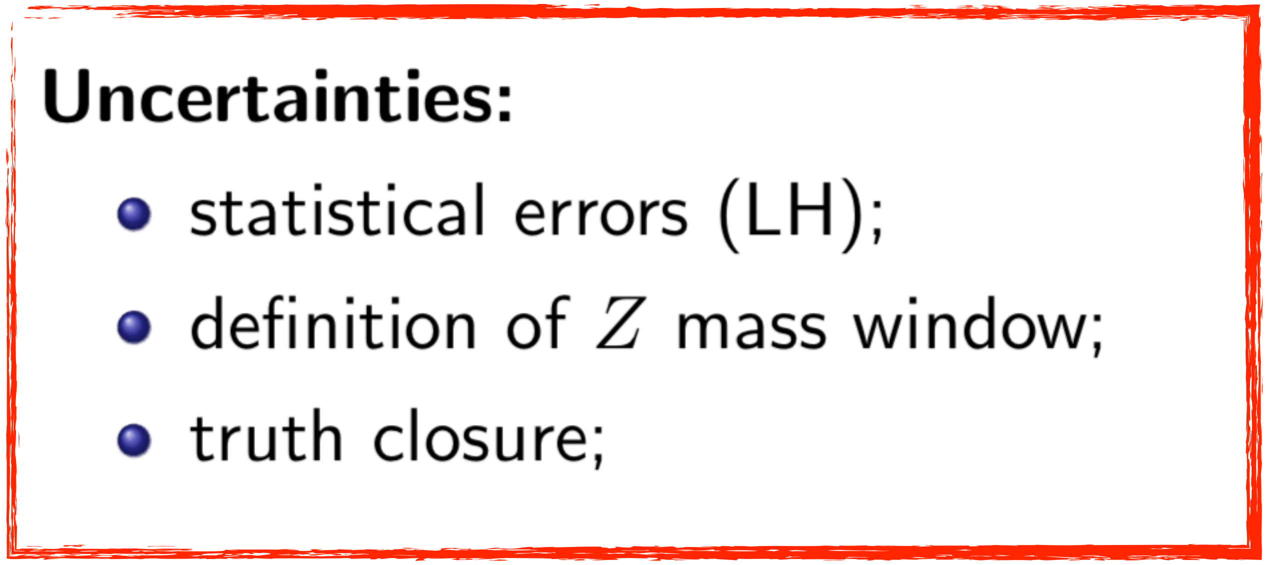
Likelihood:

$$L(\vec{\epsilon})=\prod_{i=1}^{N_{bins}}\prod_{j=1}^{N_{bins}} Poisson(N^{ss}_{Z,ij},(\epsilon_i+\epsilon_j-2\,\epsilon_i\,\epsilon_j)\cdot N^{ss+os}_{Z,ij})$$

where:

$$\epsilon_i\cdot(1-\epsilon_j)+\epsilon_j\cdot(1-\epsilon_i)=\epsilon_i+\epsilon_j-2\,\epsilon_i\cdot\epsilon_j$$

is the probability that the charge of exactly one electron is mis-identified.

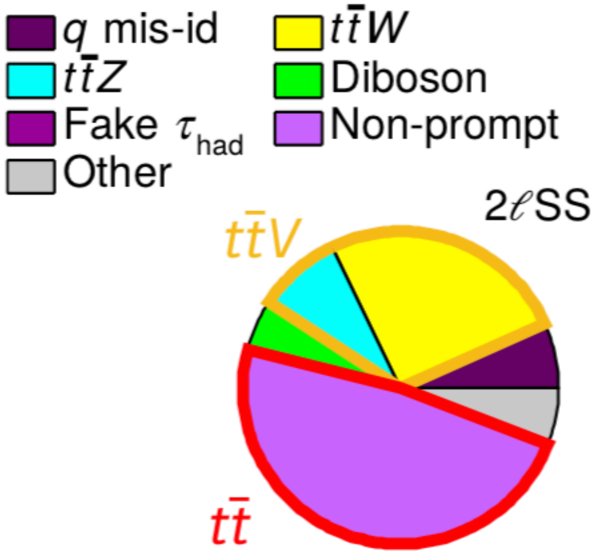


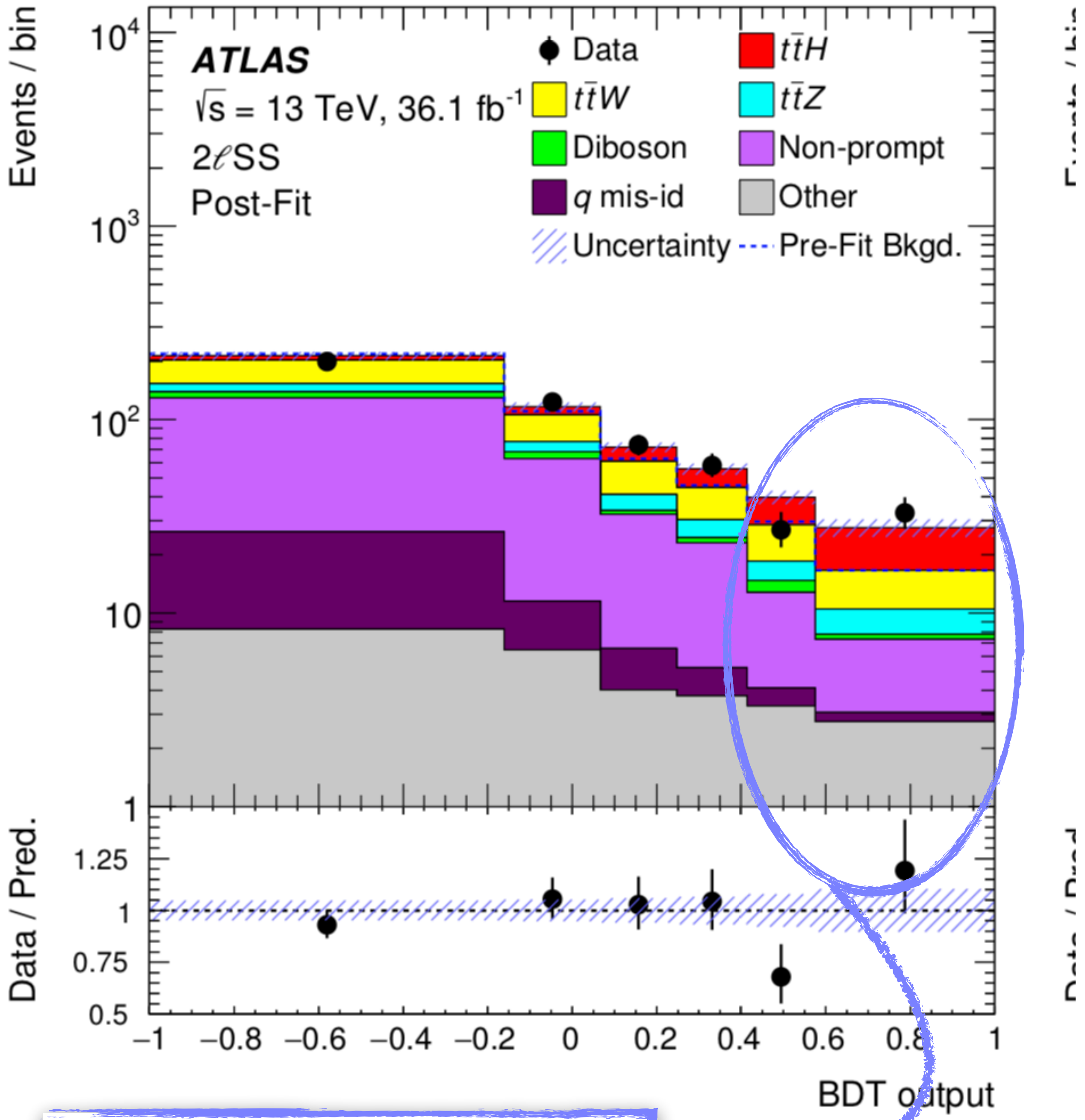
Uncertainties:

- statistical errors (LH);
- definition of Z mass window;
- truth closure;

2LSS Channel

- Use two independent event BDTs ttH vs ttV vs ttbar with input variables (average of the two BDTs is used)
  - Lepton properties
  - Jet and b-tagged jet multiplies
  - Angular distances
  - MET



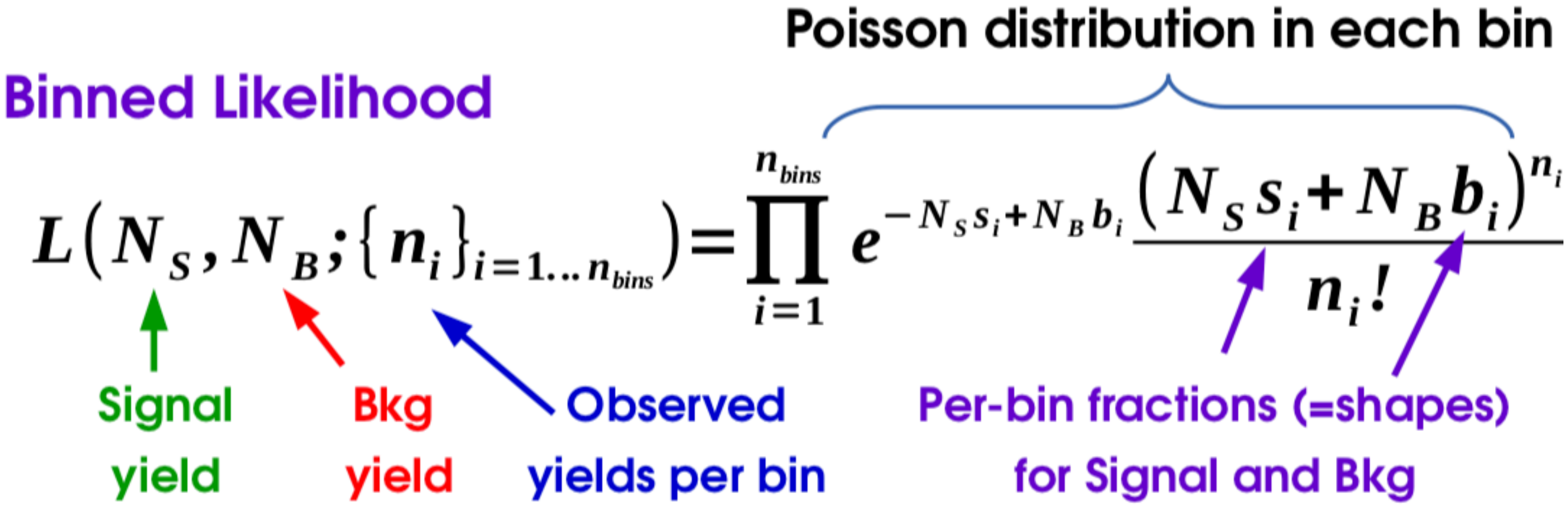


The discriminant BDTG is splitted in 6 bins

the highest bins more pure in ttH

Statistical Model

- A maximum-likelihood fit is performed on twelve categories (8 SR, 4 CR) simultaneously to extract the ttH signal strength (free parameter) $\mu_{t\bar{t}H} = \sigma/\sigma_{SM}$
- The statistical analysis of the data uses a binned likelihood function L(μ, θ), which is constructed from a product of Poisson probability distribution (the number of observed events in a given bin (n))



Binned Likelihood

$$L(N_S, N_B; \{n_i\}_{i=1\ldots n_{bins}}) = \prod_{i=1}^{n_{bins}} \overbrace{e^{-N_S s_i + N_B b_i} \frac{(N_S s_i + N_B b_i)^{n_i}}{n_i!}}^{\text{Poisson distribution in each bin}}$$

Signal yield

Bkg yield

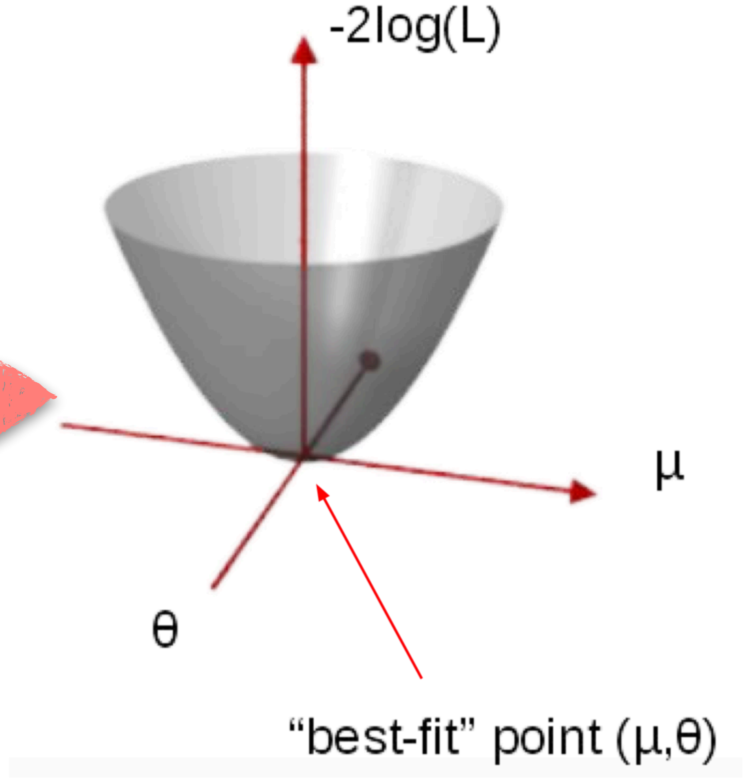
Observed yields per bin

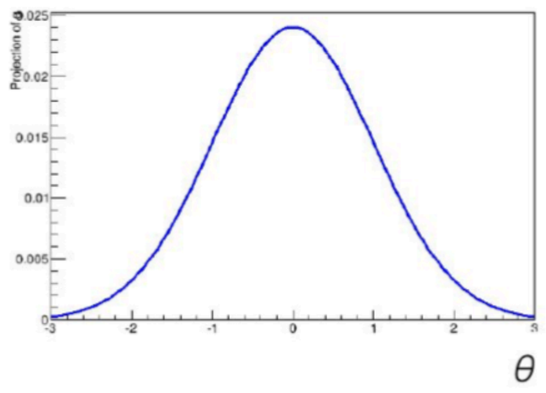
Per-bin fractions (=shapes) for Signal and Bkg

Systematics and Profile Likelihood

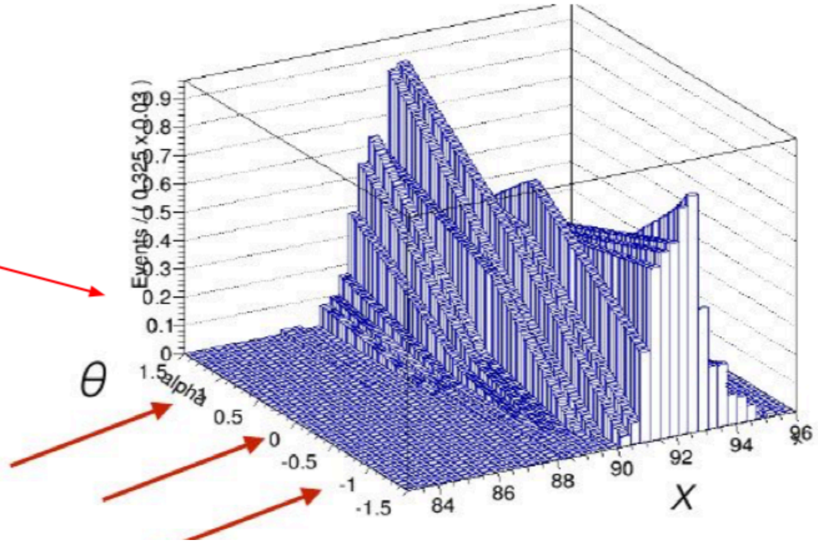
~300 NPs in the analysis

- Nuisance parameters (NPs), which encode all the uncertainties on quantities that can affect the model for signal and background
- NP probability density functions (Gaussian) are constrained by the auxiliary measurements of the parameters (unlike μ)
  - eg.
- N-dimensional likelihood maximisation (or negative-log-likelihood minimisation)





$$L(\boldsymbol{n}, \boldsymbol{\theta}^0 | \mu, \boldsymbol{\theta}) = \prod_{i \in bins} P(n_i | \mu S(\boldsymbol{\theta}) + B(\boldsymbol{\theta})) \times \prod_{j \in n.p.} G(\theta_j^0 | \theta_j)$$



Testing Model

- What values to use when defining the hypotheses ? → H(μ=0, θ=?) **Answer**: let the data choose the best-fit values
- Significance is given by the profile-likelihood ratio:

Profile likelihood ratio only dependent on μ

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$

Maximize L for a given μ 'conditional' likelihood

Maximize L 'unconditional' likelihood

- Construct Test statistics by using effective mu(how well the observed data agrees with the background-only hypothesis)

$$q_0 = \begin{cases} -2ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

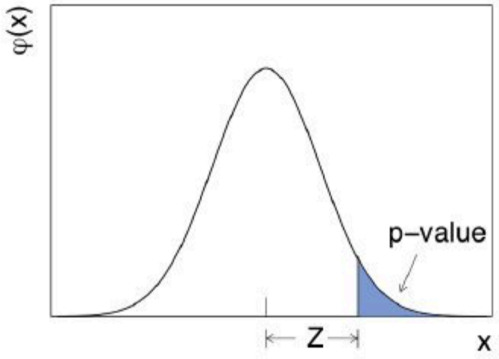
reject background-only

increasing level of incompatibility

- In particle physics, the rejection of the background-only hypothesis to claim for a discovery is conventionally achieved for a significance of Z ≥ 5, corresponding to p ≤ 2.87 × 10^{-7}

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0$$

$$Z_0 = \Phi^{-1}(1 - p_0)$$



Testing Model

- What values to use when defining the hypotheses ? → H(μ=0, θ=?) **Answer**: let the data choose the best-fit values
- Significance is given by the profile-likelihood ratio:

Profile likelihood ratio only dependent on μ

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$

Maximize L for a given μ 'conditional' likelihood

Maximize L 'unconditional' likelihood

- Construct Test statistics by using effective mu(how well the observed data agrees with the background-only hypothesis)

$$q_0 = \begin{cases} -2ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}$$

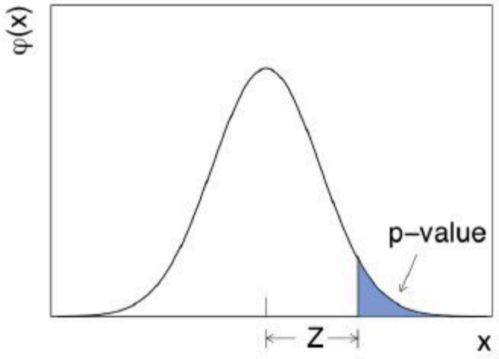
reject background-only

increasing level of incompatibility

- In particle physics, the rejection of the background-only hypothesis to claim for a discovery is conventionally achieved for a significance of $Z \geq 5$, corresponding to $p \leq 2.87 \times 10^{-7}$

$$p_0 = \int_{q_{0,\mathrm{obs}}}^{\infty} f(q_0|0)\, dq_0$$

$$Z_0 = \Phi^{-1}(1 - p_0)$$

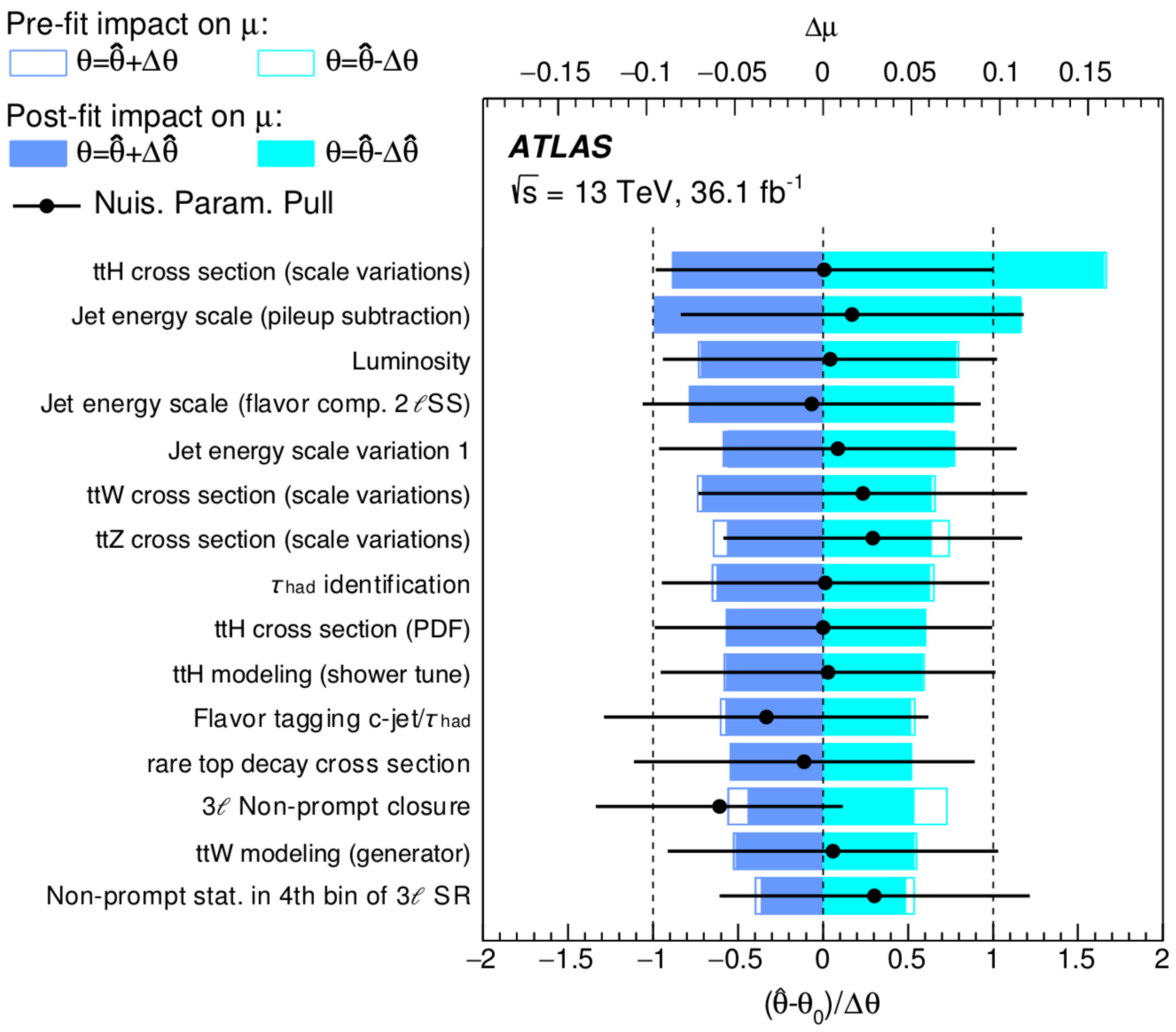


the sampling distribution for our test statistic—

>**Wilks' Theorem**: PLR also follows a χ^2 !

Pull/Impact Plot

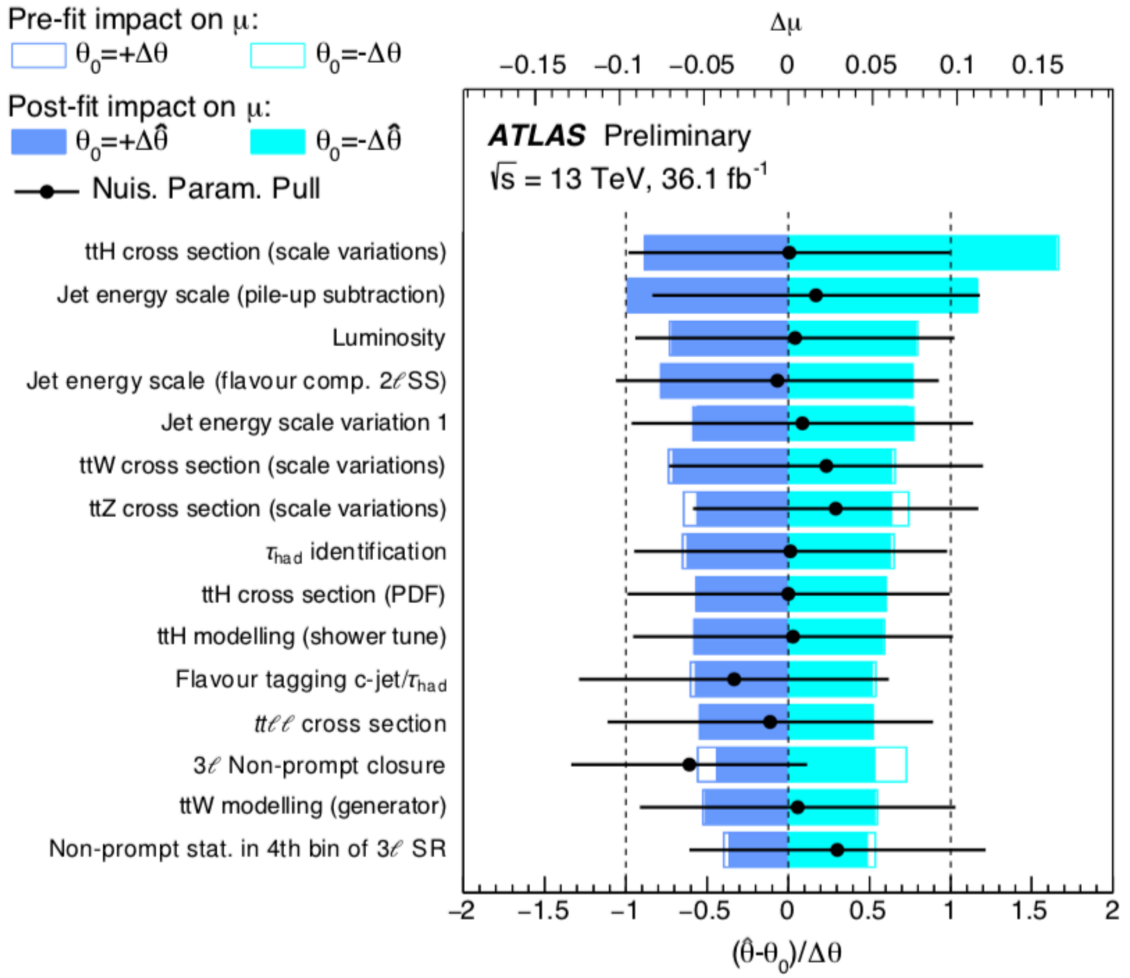
- Nominally systematics NPs have:
  - **Central value = 0** : i.e. the pre-fit expectation
  - **Uncertainty = 1** : NPs normalized to the value of the systematic
- From fit results:
  - **If central value /= 0:** it indicates the fit is correcting for a biased initial prediction of that parameter



the fit can constrain the estimated uncertainty —> more statistical sensitivity than the auxiliary measurement used to determine its prior uncertainty

Results

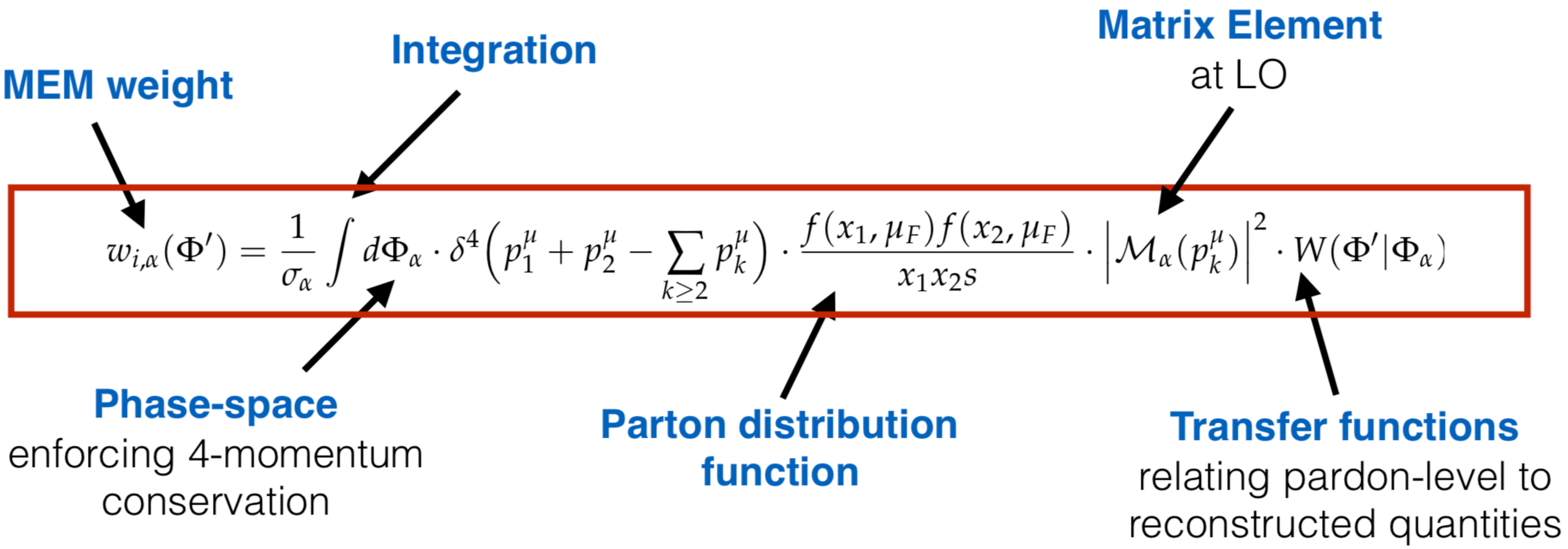
Uncertainty Source	$\Delta\mu$	
$t\bar{t}H$ modelling (cross section)	+0.20	−0.09
Jet energy scale and resolution	+0.18	−0.15
Non-prompt light-lepton estimates	+0.15	−0.13
Jet flavour tagging and τ_{had} identification	+0.11	−0.09
$t\bar{t}W$ modelling	+0.10	−0.09
$t\bar{t}Z$ modelling	+0.08	−0.07
Other background modelling	+0.08	−0.07
Luminosity	+0.08	−0.06
$t\bar{t}H$ modelling (acceptance)	+0.08	−0.04
Fake τ_{had} estimates	+0.07	−0.07
Other experimental uncertainties	+0.05	−0.04
Simulation statistics	+0.04	−0.04
Charge misassignment	+0.01	−0.01
Total systematic uncertainty	+0.39	−0.30



- Most relevant uncertainties on the signal strength:
  - Signal modelling (dominated by scale uncertainties)
  - Jet energy scale and resolution
  - Non-prompt ℓ estimation (with large contribution from limited CR statistics)

MEM

- The matrix element method (MEM) provides a way to calculate the likelihood that an event originates from a given production mechanism —> assign probability density value based on theory
- Use smart phase-space mappings to align peaks of integrand with coordinate axes (the structure of the integrand can be very complicated: integrand peaks coming from ME and TFs, and there can be many)
  - The phase-space can be organized in pieces or subsets of variables (blocks)



MEM weight

Integration

Matrix Element
at LO

$$w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha}\int d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k\geq 2} p_k^\mu\right) \cdot \frac{f(x_1,\mu_F)f(x_2,\mu_F)}{x_1 x_2 s} \cdot \left|\mathcal{M}_\alpha(p_k^\mu)\right|^2 \cdot W(\Phi'|\Phi_\alpha)$$

Phase-space
enforcing 4-momentum conservation

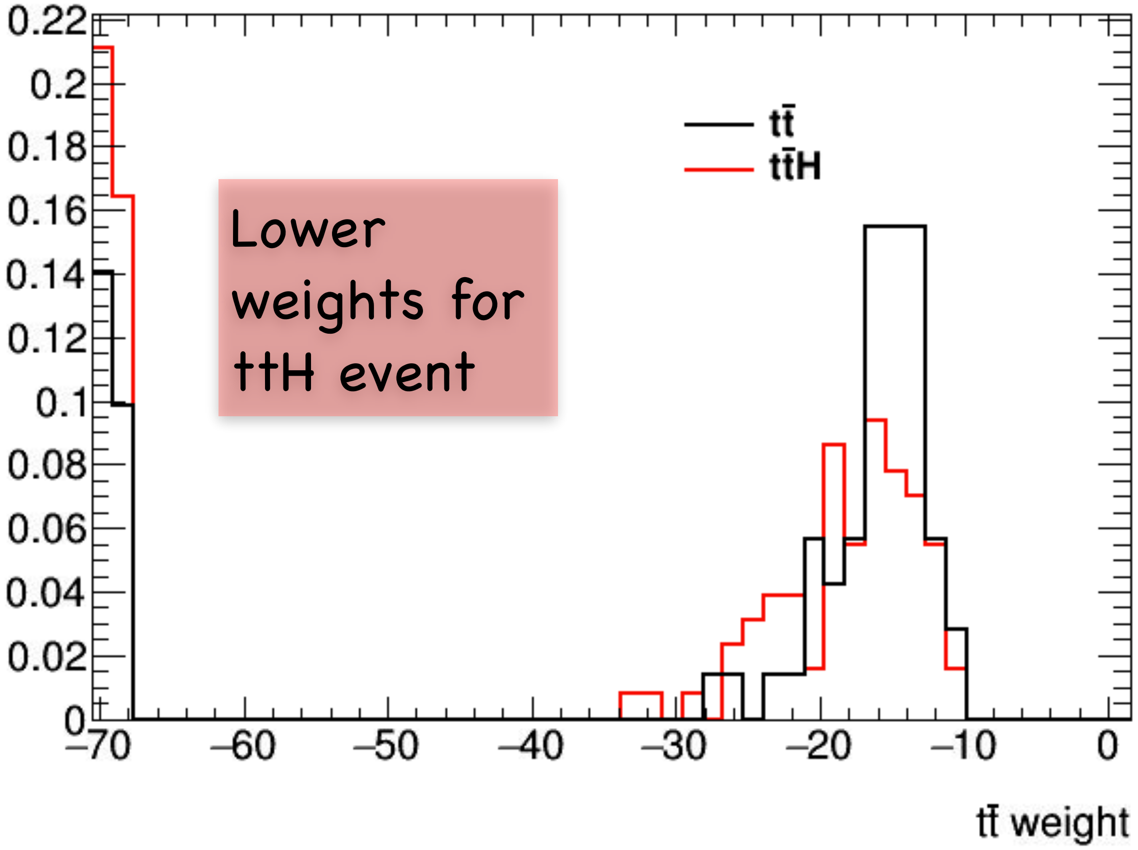
Parton distribution function

Transfer functions
relating pardon-level to reconstructed quantities

Interpretation: The MEM weight is the cross section, for a given hypothesis, evaluated at the phase space point of the event, convolved with the transfer functions

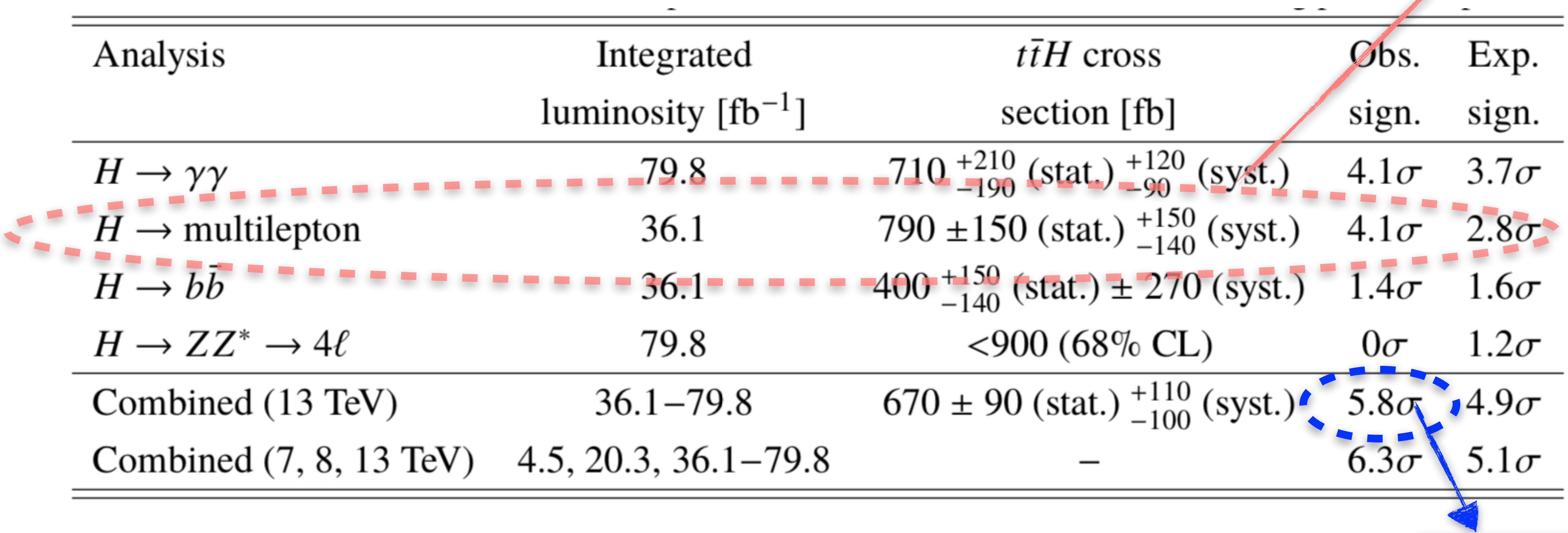
Studies on MEM

- MEM can be used to identify the process : Takes as input a set of kinematic observables associated to the diagram and calculates the weight for each event, can be calculated;
  - for all permutations of the selected objects or,
  - for the reconstructed particles (eg. by BDT).



Lower weights for ttH event

Conclusion



Analysis	Integrated luminosity [fb^{-1}]	$t\bar{t}H$ cross section [fb]	Obs. sign.	Exp. sign.
$H \to \gamma\gamma$	79.8	$710\ ^{+210}_{-190}$ (stat.) $^{+120}_{-90}$ (syst.)	4.1σ	3.7σ
$H \to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	4.1σ	2.8σ
$H \to b\bar{b}$	36.1	$400\ ^{+150}_{-140}$ (stat.) $\pm$ 270 (syst.)	1.4σ	1.6σ
$H \to ZZ^* \to 4\ell$	79.8	<900 (68% CL)	0σ	1.2σ
Combined (13 TeV)	36.1–79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	5.8σ	4.9σ
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1–79.8	–	6.3σ	5.1σ

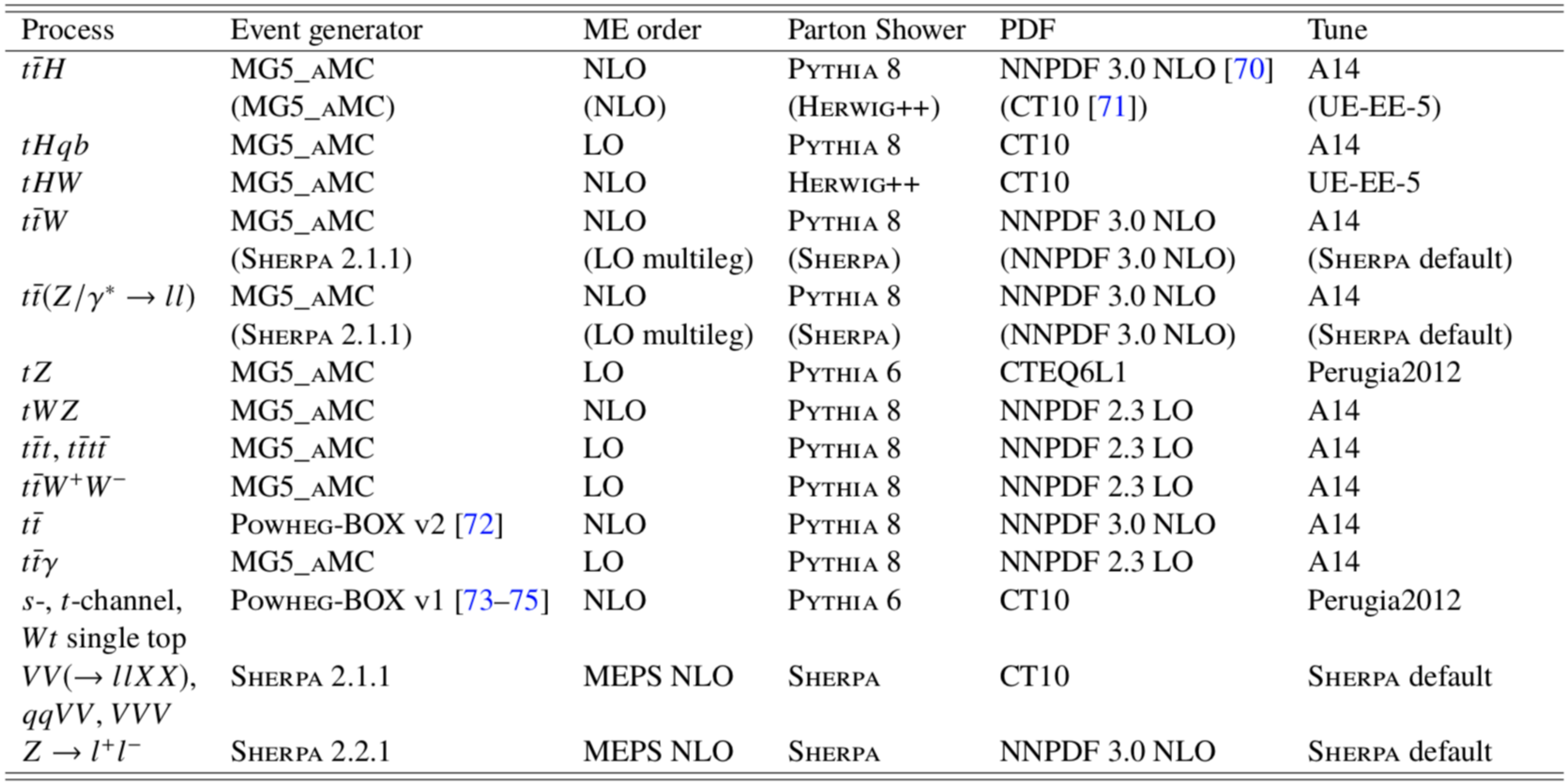
Evidence

Observation

- This year, the studies are on going to improve the ttH-Multilepton results
- Paper with 80/fb results for ttH-ML

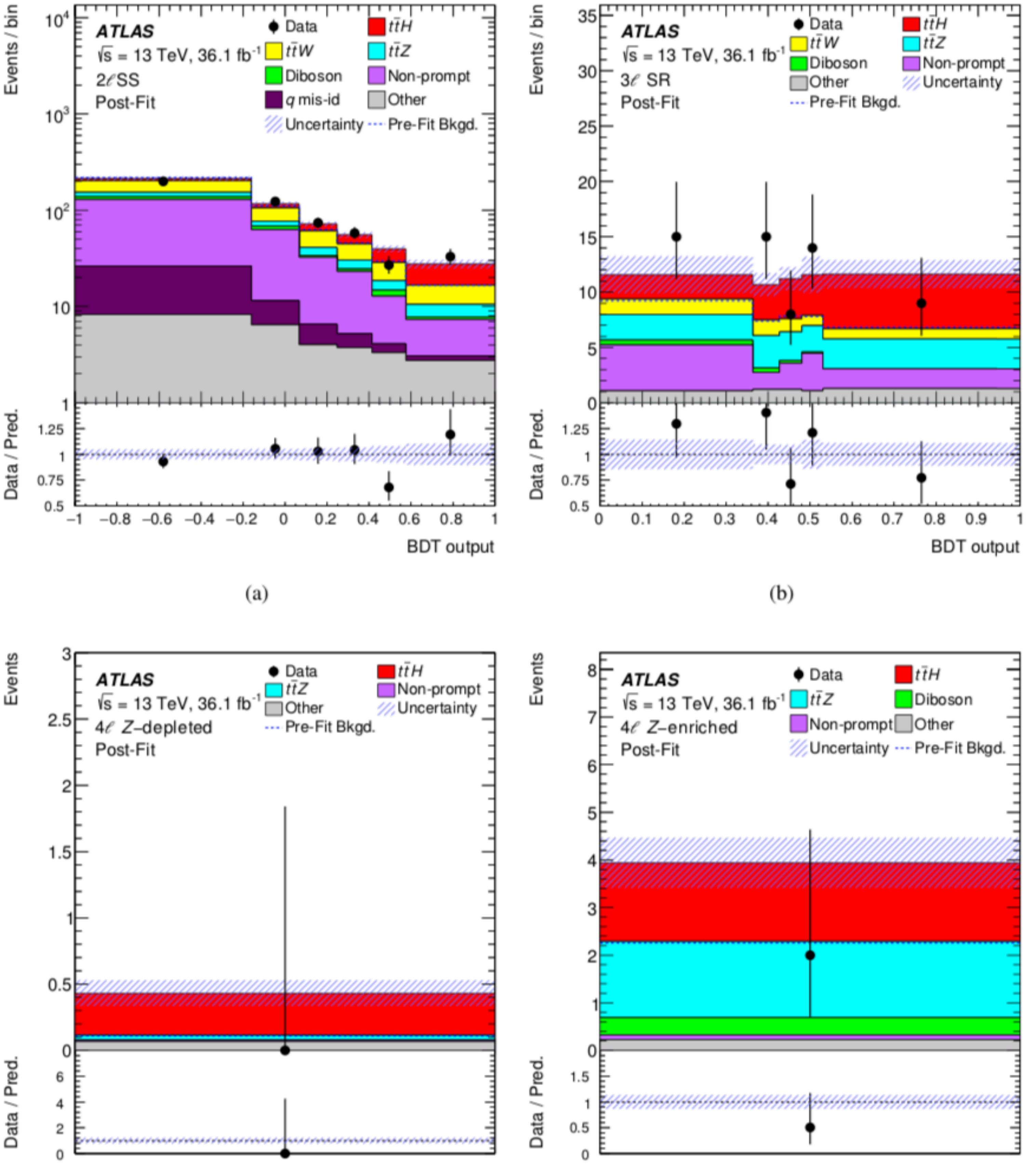
Backup

Backup



Process	Event generator	ME order	Parton Shower	PDF	Tune
$t\bar{t}H$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO [70]	A14
	(MG5_aMC)	(NLO)	(Herwig++)	(CT10 [71])	(UE-EE-5)
$tHqb$	MG5_aMC	LO	Pythia 8	CT10	A14
tHW	MG5_aMC	NLO	Herwig++	CT10	UE-EE-5
$t\bar{t}W$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO	A14
	(Sherpa 2.1.1)	(LO multileg)	(Sherpa)	(NNPDF 3.0 NLO)	(Sherpa default)
$t\bar{t}(Z/\gamma^* \to ll)$	MG5_aMC	NLO	Pythia 8	NNPDF 3.0 NLO	A14
	(Sherpa 2.1.1)	(LO multileg)	(Sherpa)	(NNPDF 3.0 NLO)	(Sherpa default)
tZ	MG5_aMC	LO	Pythia 6	CTEQ6L1	Perugia2012
tWZ	MG5_aMC	NLO	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}t$, $t\bar{t}t\bar{t}$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}$	Powheg-BOX v2 [72]	NLO	Pythia 8	NNPDF 3.0 NLO	A14
$t\bar{t}\gamma$	MG5_aMC	LO	Pythia 8	NNPDF 2.3 LO	A14
s-, t-channel, Wt single top	Powheg-BOX v1 [73–75]	NLO	Pythia 6	CT10	Perugia2012
$VV(\to llXX)$, $qqVV$, VVV	Sherpa 2.1.1	MEPS NLO	Sherpa	CT10	Sherpa default
$Z \to l^+l^-$	Sherpa 2.2.1	MEPS NLO	Sherpa	NNPDF 3.0 NLO	Sherpa default

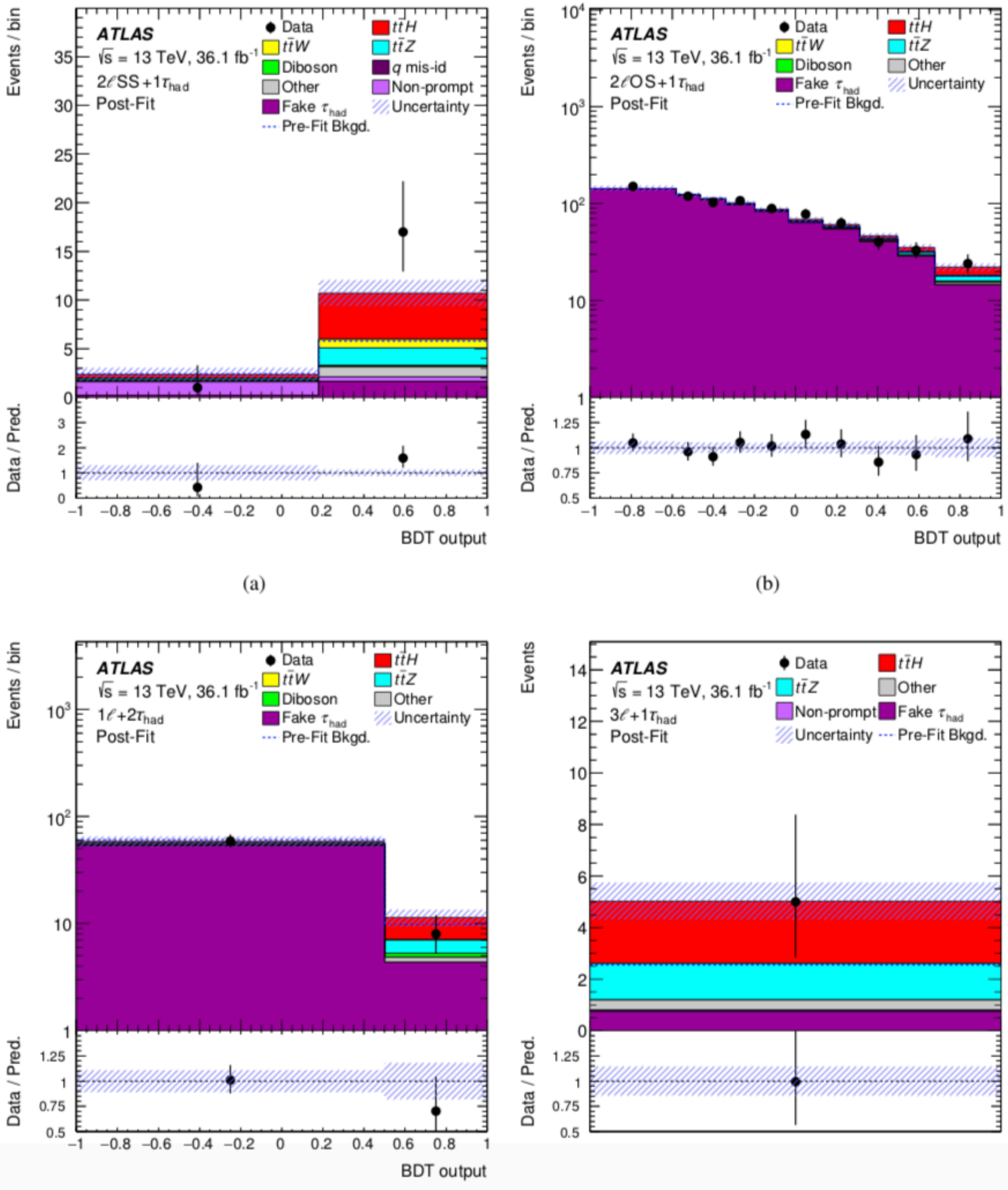
Backup


ATLAS
√s = 13 TeV, 36.1 fb⁻¹
2ℓSS
Post-Fit
Data
ttH
ttW
ttZ
Diboson
Non-prompt
q mis-id
Other
Uncertainty
Pre-Fit Bkgd.
Events / bin
Data / Pred.
BDT output
ATLAS
√s = 13 TeV, 36.1 fb⁻¹
3ℓ SR
Post-Fit
Events / bin
Data / Pred.
BDT output
ATLAS
√s = 13 TeV, 36.1 fb⁻¹
4ℓ Z–depleted
Post-Fit
Events
Data / Pred.
ATLAS
√s = 13 TeV, 36.1 fb⁻¹
4ℓ Z–enriched
Post-Fit
Events
Data / Pred.

(a) (b)

(c) (d)

Backup

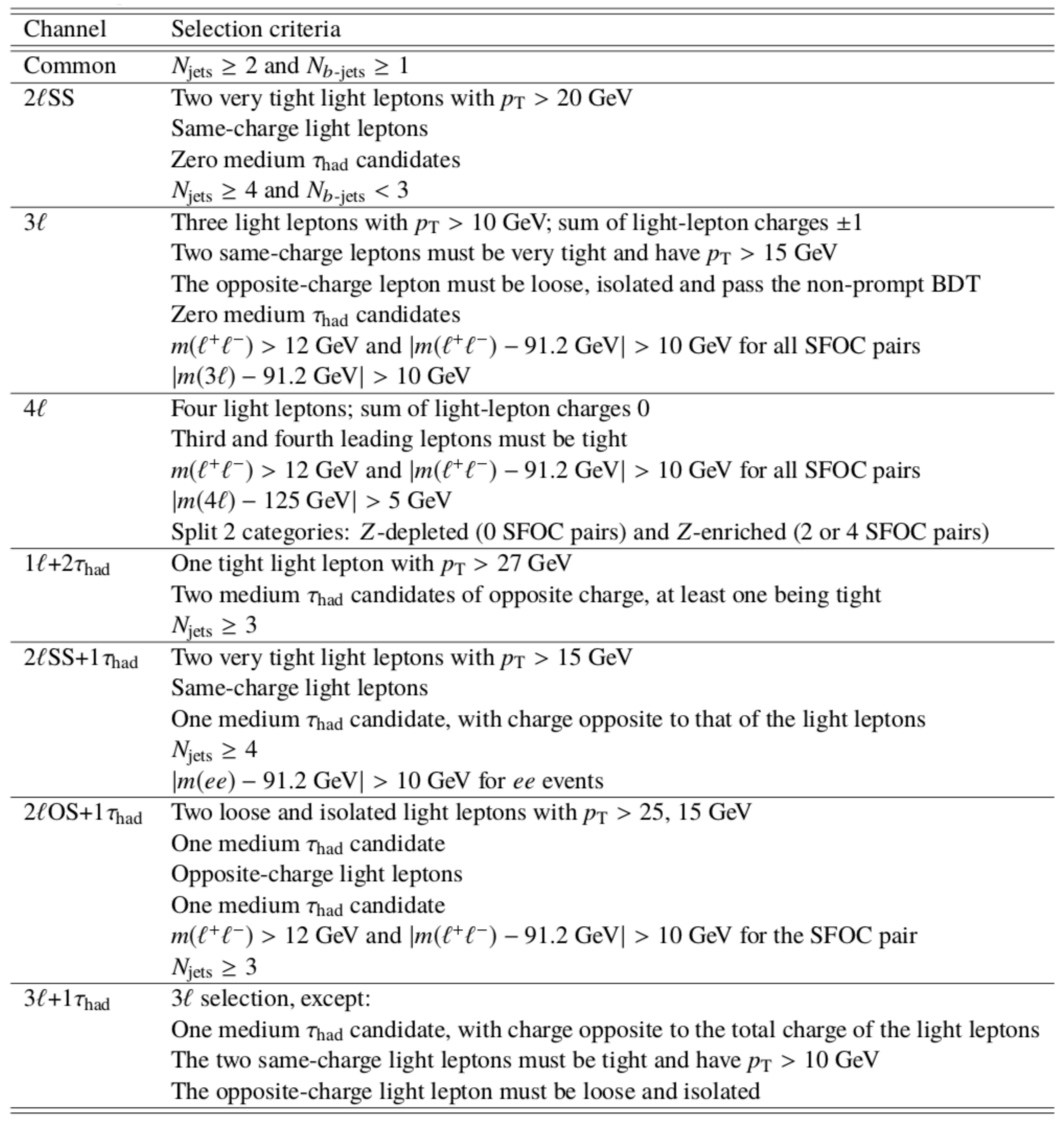


ATLAS
√s = 13 TeV, 36.1 fb⁻¹
2ℓSS+1τhad
Post-Fit
Data
ttW
Diboson
Other
Fake τhad
Pre-Fit Bkgd.
ttH
ttZ
q mis-id
Non-prompt
Uncertainty
Events / bin
Data / Pred.
BDT output
ATLAS
√s = 13 TeV, 36.1 fb⁻¹
2ℓOS+1τhad
Post-Fit
Data
ttW
Diboson
Fake τhad
Pre-Fit Bkgd.
ttH
ttZ
Other
Uncertainty
Events / bin
Data / Pred.
BDT output
ATLAS
√s = 13 TeV, 36.1 fb⁻¹
1ℓ+2τhad
Post-Fit
Data
ttW
Diboson
Fake τhad
Pre-Fit Bkgd.
ttH
ttZ
Other
Uncertainty
Events / bin
Data / Pred.
BDT output
ATLAS
√s = 13 TeV, 36.1 fb⁻¹
3ℓ+1τhad
Post-Fit
Data
ttZ
Non-prompt
Uncertainty
ttH
Other
Fake τhad
Pre-Fit Bkgd.
Events
Data / Pred.

(a)

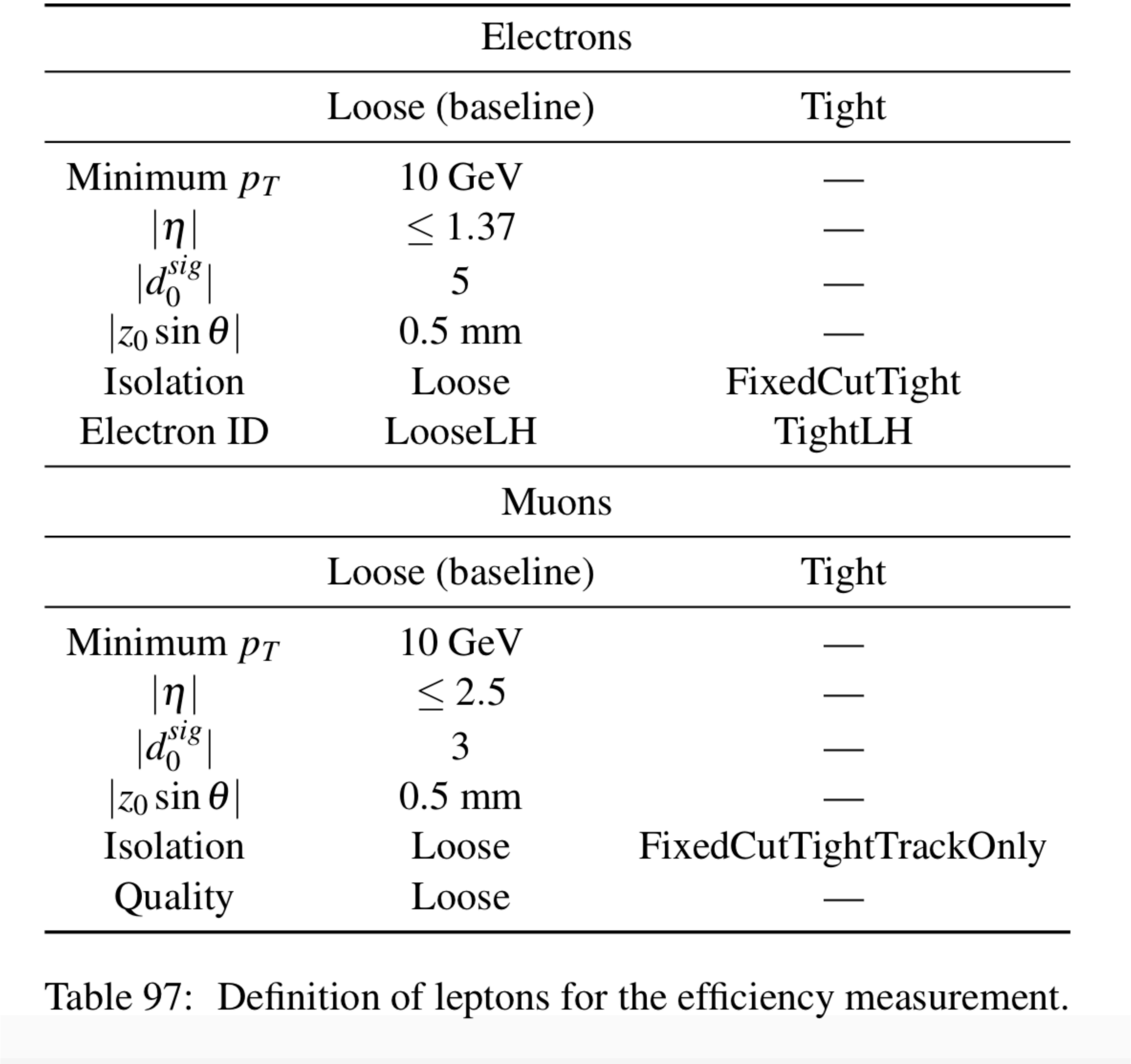
(b)

Backup



Channel	Selection criteria
Common	$N_{\text{jets}} \geq 2$ and $N_{b\text{-jets}} \geq 1$
2ℓSS	Two very tight light leptons with $p_{\text{T}} > 20$ GeV
	Same-charge light leptons
	Zero medium τ_{had} candidates
	$N_{\text{jets}} \geq 4$ and $N_{b\text{-jets}} < 3$
3ℓ	Three light leptons with $p_{\text{T}} > 10$ GeV; sum of light-lepton charges ±1
	Two same-charge leptons must be very tight and have $p_{\text{T}} > 15$ GeV
	The opposite-charge lepton must be loose, isolated and pass the non-prompt BDT
	Zero medium τ_{had} candidates
	$m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\text{ GeV}\rvert > 10$ GeV for all SFOC pairs
	$\lvert m(3\ell) - 91.2\text{ GeV}\rvert > 10$ GeV
4ℓ	Four light leptons; sum of light-lepton charges 0
	Third and fourth leading leptons must be tight
	$m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\text{ GeV}\rvert > 10$ GeV for all SFOC pairs
	$\lvert m(4\ell) - 125\text{ GeV}\rvert > 5$ GeV
	Split 2 categories: Z-depleted (0 SFOC pairs) and Z-enriched (2 or 4 SFOC pairs)
1ℓ+$2\tau_{\text{had}}$	One tight light lepton with $p_{\text{T}} > 27$ GeV
	Two medium τ_{had} candidates of opposite charge, at least one being tight
	$N_{\text{jets}} \geq 3$
2ℓSS+$1\tau_{\text{had}}$	Two very tight light leptons with $p_{\text{T}} > 15$ GeV
	Same-charge light leptons
	One medium τ_{had} candidate, with charge opposite to that of the light leptons
	$N_{\text{jets}} \geq 4$
	$\lvert m(ee) - 91.2\text{ GeV}\rvert > 10$ GeV for ee events
2ℓOS+$1\tau_{\text{had}}$	Two loose and isolated light leptons with $p_{\text{T}} > 25, 15$ GeV
	One medium τ_{had} candidate
	Opposite-charge light leptons
	One medium τ_{had} candidate
	$m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\text{ GeV}\rvert > 10$ GeV for the SFOC pair
	$N_{\text{jets}} \geq 3$
3ℓ+$1\tau_{\text{had}}$	3ℓ selection, except:
	One medium τ_{had} candidate, with charge opposite to the total charge of the light leptons
	The two same-charge light leptons must be tight and have $p_{\text{T}} > 10$ GeV
	The opposite-charge light lepton must be loose and isolated

Backup

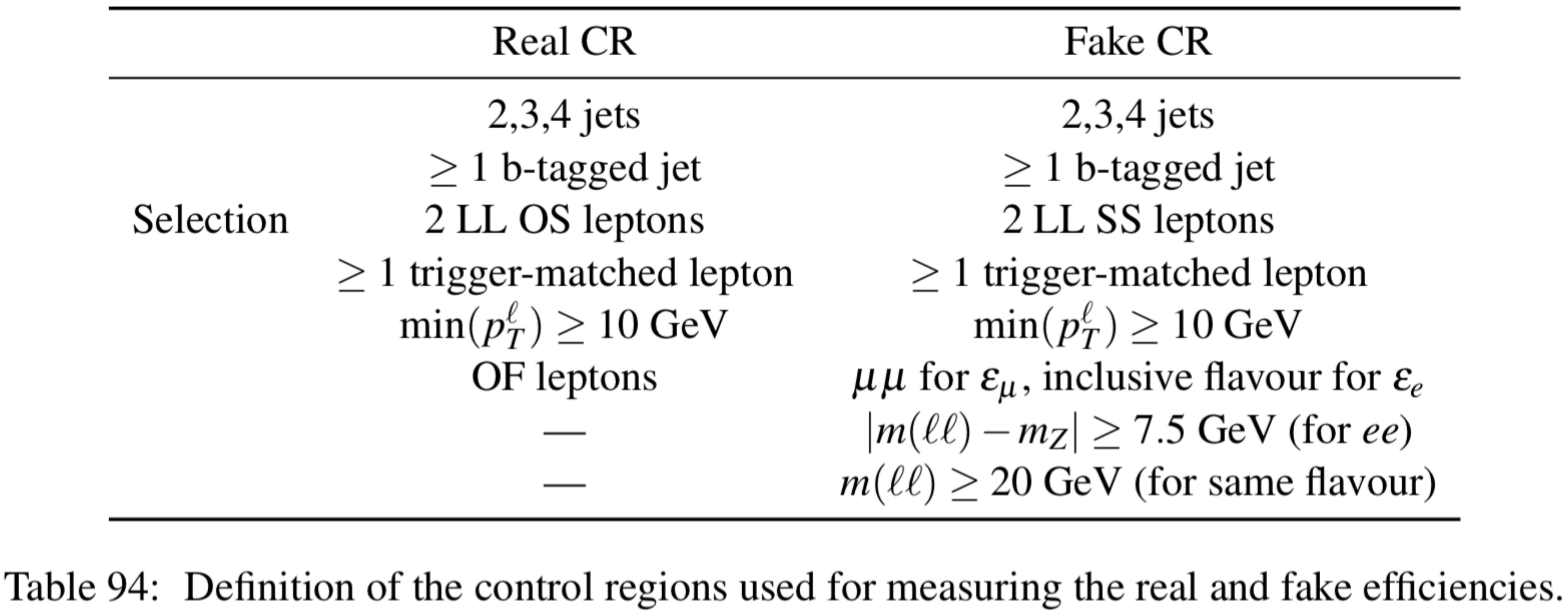


Electrons		
	Loose (baseline)	Tight
Minimum p_T	10 GeV	—
$\|\eta\|$	≤ 1.37	—
$\|d_0^{sig}\|$	5	—
$\|z_0 \sin\theta\|$	0.5 mm	—
Isolation	Loose	FixedCutTight
Electron ID	LooseLH	TightLH

Muons		
	Loose (baseline)	Tight
Minimum p_T	10 GeV	—
$\|\eta\|$	≤ 2.5	—
$\|d_0^{sig}\|$	3	—
$\|z_0 \sin\theta\|$	0.5 mm	—
Isolation	Loose	FixedCutTightTrackOnly
Quality	Loose	—

Table 97: Definition of leptons for the efficiency measurement.

Backup

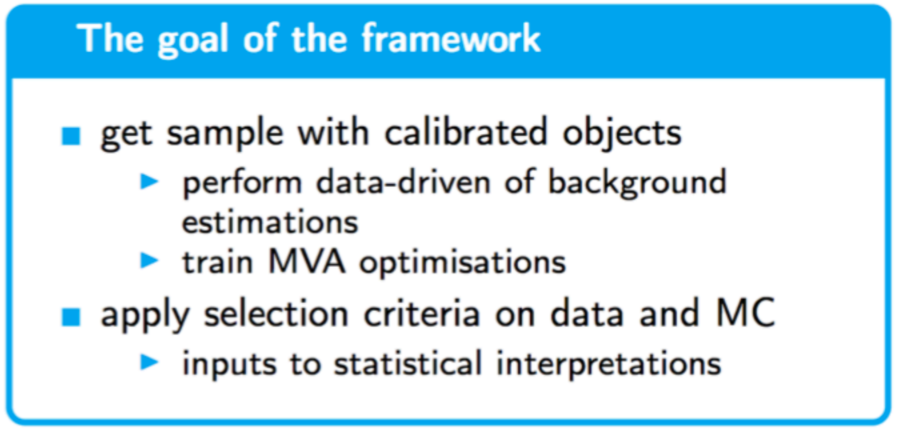


	Real CR	Fake CR
Selection	2,3,4 jets	2,3,4 jets
	≥ 1 b-tagged jet	≥ 1 b-tagged jet
	2 LL OS leptons	2 LL SS leptons
	≥ 1 trigger-matched lepton	≥ 1 trigger-matched lepton
	$\min(p_T^\ell) \geq 10$ GeV	$\min(p_T^\ell) \geq 10$ GeV
	OF leptons	$\mu\mu$ for ε_μ, inclusive flavour for ε_e
	—	$\lvert m(\ell\ell) - m_Z \rvert \geq 7.5$ GeV (for ee)
	—	$m(\ell\ell) \geq 20$ GeV (for same flavour)

Table 94: Definition of the control regions used for measuring the real and fake efficiencies.

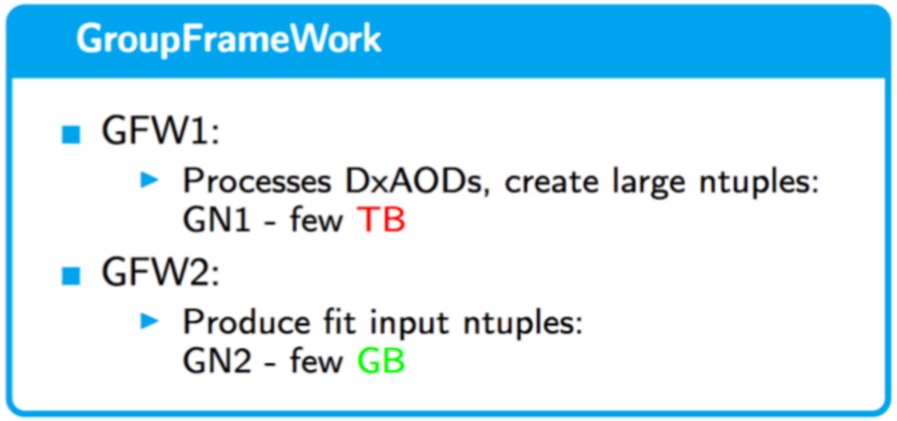
Backup

- ttH-ML analysis workflow = two main group frameworks + statistical tool framework
- Other FW used for:
  - Tau fake estimates [TBC]
  - Matrix Method efficiencies calculation



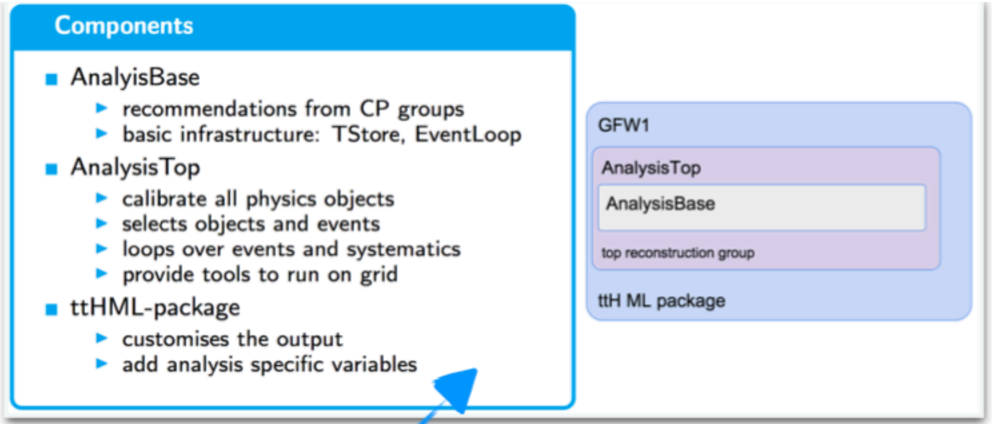
The goal of the framework

- get sample with calibrated objects
  - perform data-driven of background estimations
  - train MVA optimisations
- apply selection criteria on data and MC
  - inputs to statistical interpretations



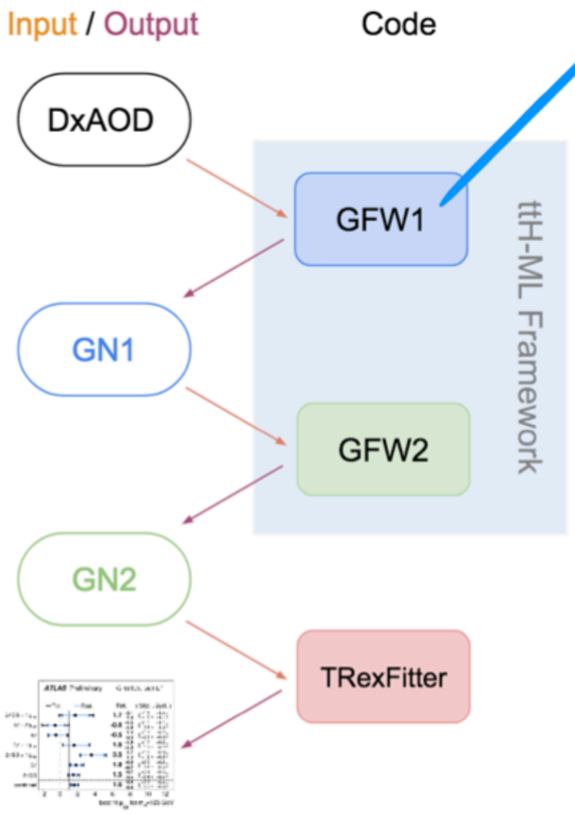
GroupFrameWork

- GFW1:
  - Processes DxAODs, create large ntuples: GN1 - few TB
- GFW2:
  - Produce fit input ntuples: GN2 - few GB

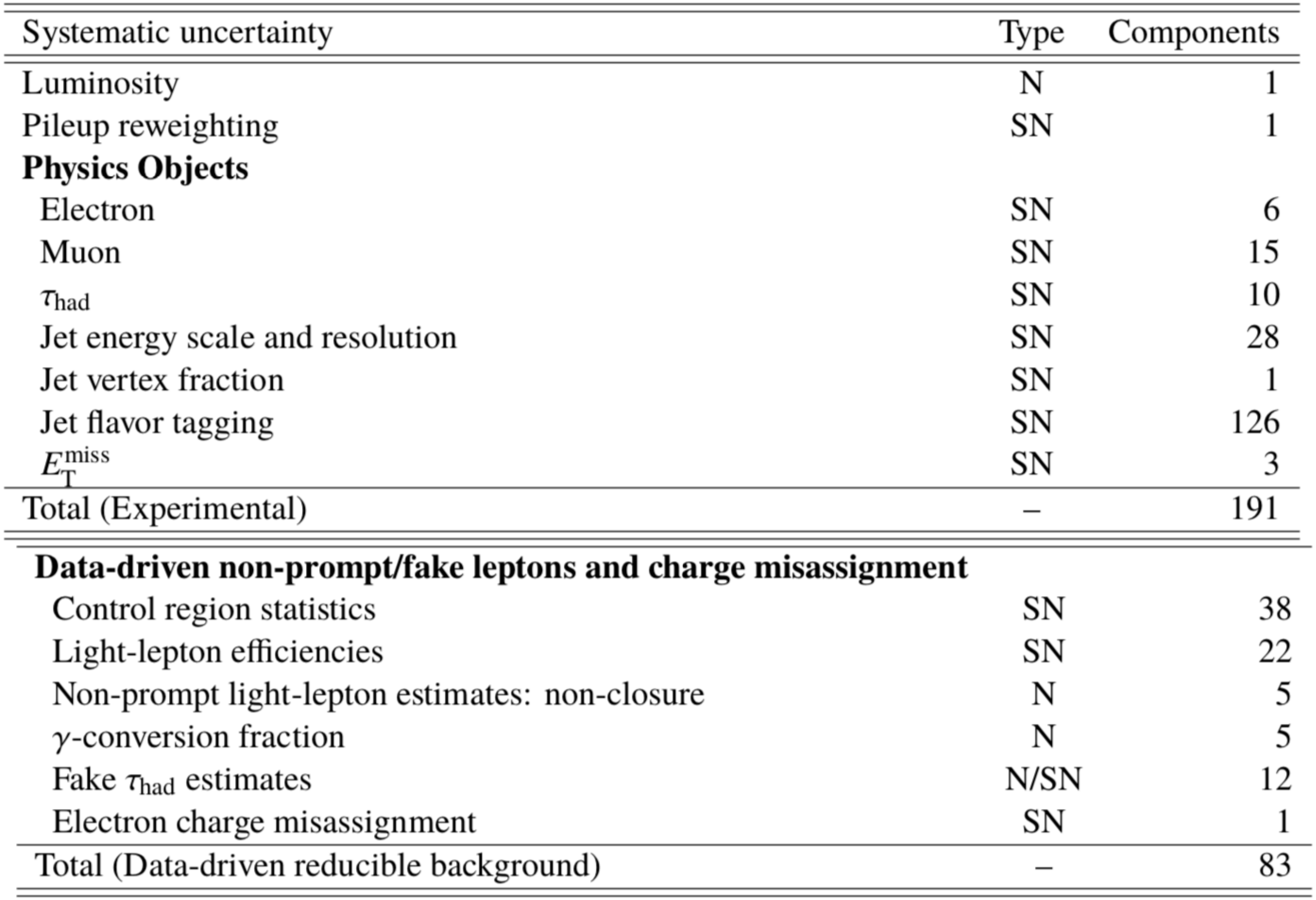


Components

- AnalysisBase
  - recommendations from CP groups
  - basic infrastructure: TStore, EventLoop
- AnalysisTop
  - calibrate all physics objects
  - selects objects and events
  - loops over events and systematics
  - provide tools to run on grid
- ttHML-package
  - customises the output
  - add analysis specific variables

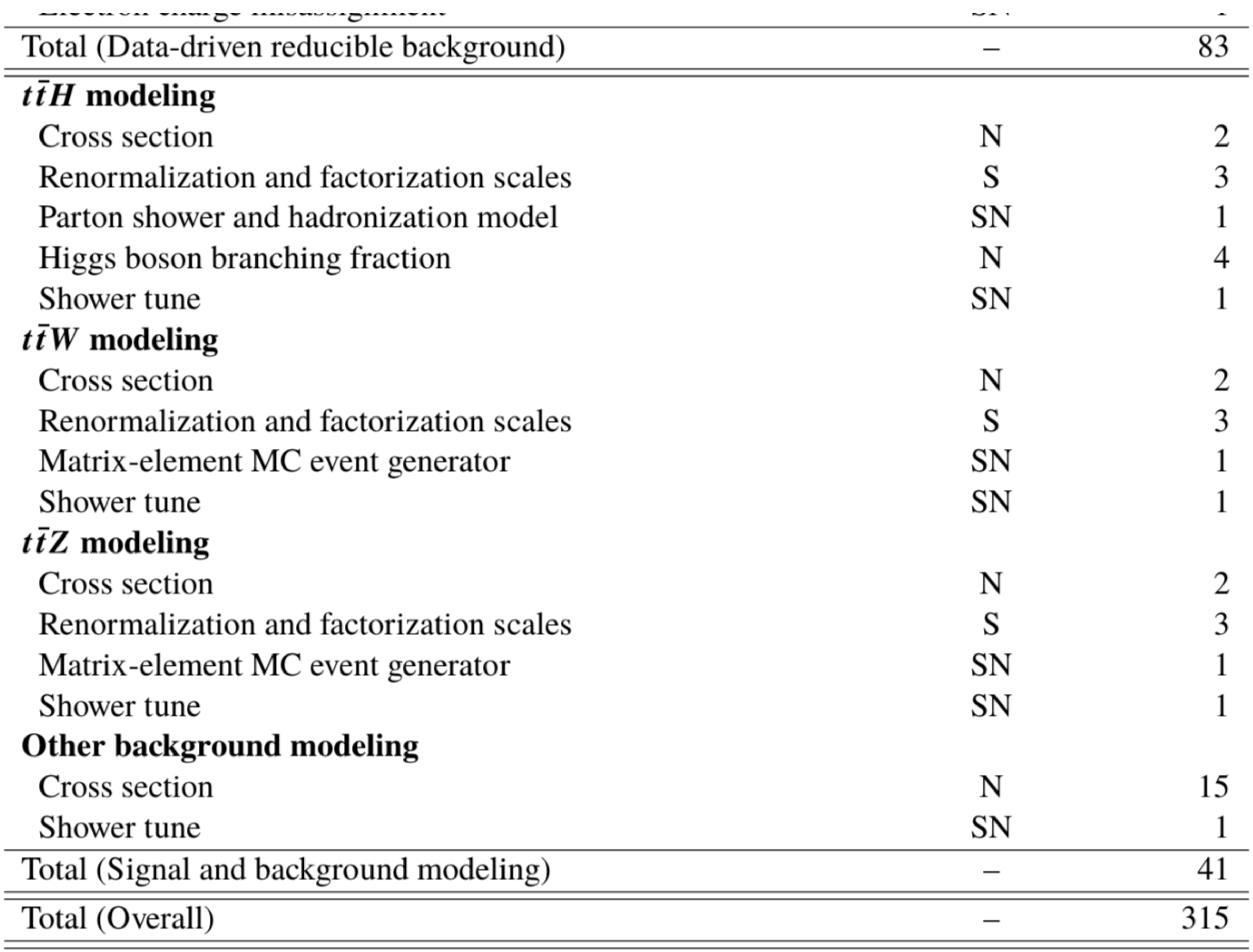


Backup



Systematic uncertainty	Type	Components
Luminosity	N	1
Pileup reweighting	SN	1
Physics Objects		
Electron	SN	6
Muon	SN	15
τ_{had}	SN	10
Jet energy scale and resolution	SN	28
Jet vertex fraction	SN	1
Jet flavor tagging	SN	126
$E_{\text{T}}^{\text{miss}}$	SN	3
Total (Experimental)	–	191
Data-driven non-prompt/fake leptons and charge misassignment		
Control region statistics	SN	38
Light-lepton efficiencies	SN	22
Non-prompt light-lepton estimates: non-closure	N	5
γ-conversion fraction	N	5
Fake τ_{had} estimates	N/SN	12
Electron charge misassignment	SN	1
Total (Data-driven reducible background)	–	83

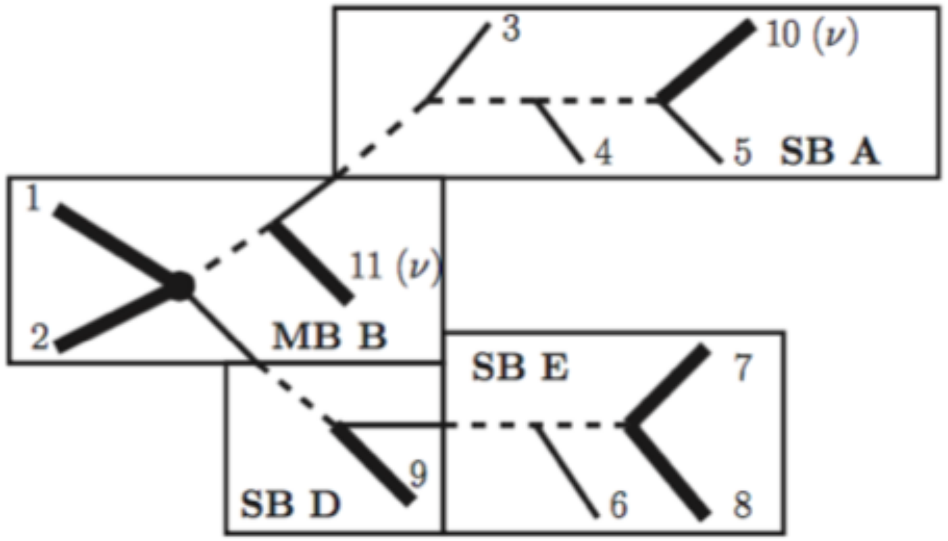
Backup



Total (Data-driven reducible background)	–	83
***$t\bar{t}H$* modeling**		
Cross section	N	2
Renormalization and factorization scales	S	3
Parton shower and hadronization model	SN	1
Higgs boson branching fraction	N	4
Shower tune	SN	1
***$t\bar{t}W$* modeling**		
Cross section	N	2
Renormalization and factorization scales	S	3
Matrix-element MC event generator	SN	1
Shower tune	SN	1
***$t\bar{t}Z$* modeling**		
Cross section	N	2
Renormalization and factorization scales	S	3
Matrix-element MC event generator	SN	1
Shower tune	SN	1
Other background modeling		
Cross section	N	15
Shower tune	SN	1
Total (Signal and background modeling)	–	41
Total (Overall)	–	315

Backup

- The method consists in the calculation of an integral, where:
  - PDF → available from several collaborations (eg. LHAPDF)
  - Matrix Element → from MC (eg. MadGraph)
  - Transfer Functions → can be parametrized from MC
  - Once you have all the terms, integrate! (eg. VEGAS, **Cuba**)
- Use smart phase-space mappings to align peaks of integrand with coordinate axes (the structure of the integrand can be very complicated: integrand peaks coming from ME and TFs, and there can be many)
  - The phase-space can be organized in pieces or subsets of variables (blocks)



$p_1 \to p_2 p_3 \Longleftrightarrow dE_2\, d\theta_2\, d\phi_2\, dE_3\, d\theta_3\, d\phi_3$, where particle3 is a neutrino

Apply change of variables to align the ME propagator:

$\Longrightarrow dE_2\, d\theta_2\, d\phi 2\, ds_1\, d\theta_3\, d\phi_3 \times jac.$

$\Longrightarrow$ Propagator for p_1 aligned with grid!

Standard phase-space parametrization:

$$d\Phi(y) \propto \prod_i \frac{|p_i|^2 d|p_i| d\phi_i sin\theta_i d\theta_i}{2E_i} \delta^4(P_{in} - P_{fin})$$