Measurements of single top quark production in association with a W boson

Rui Zhang

UCAS seminar

31.08.2018

In this talk …

- Physics motivation – Single-top-quark production and tW channel
- Inclusive cross-section measurement
  - Systematic uncertainties and maximum likelihood fit method
- Differential cross-section measurements
- Adversarial neural network

Part 1

- Top-quark production and decay
- Single-top-quark production
- tW & $t\bar{t}$ interference

Top-quark production

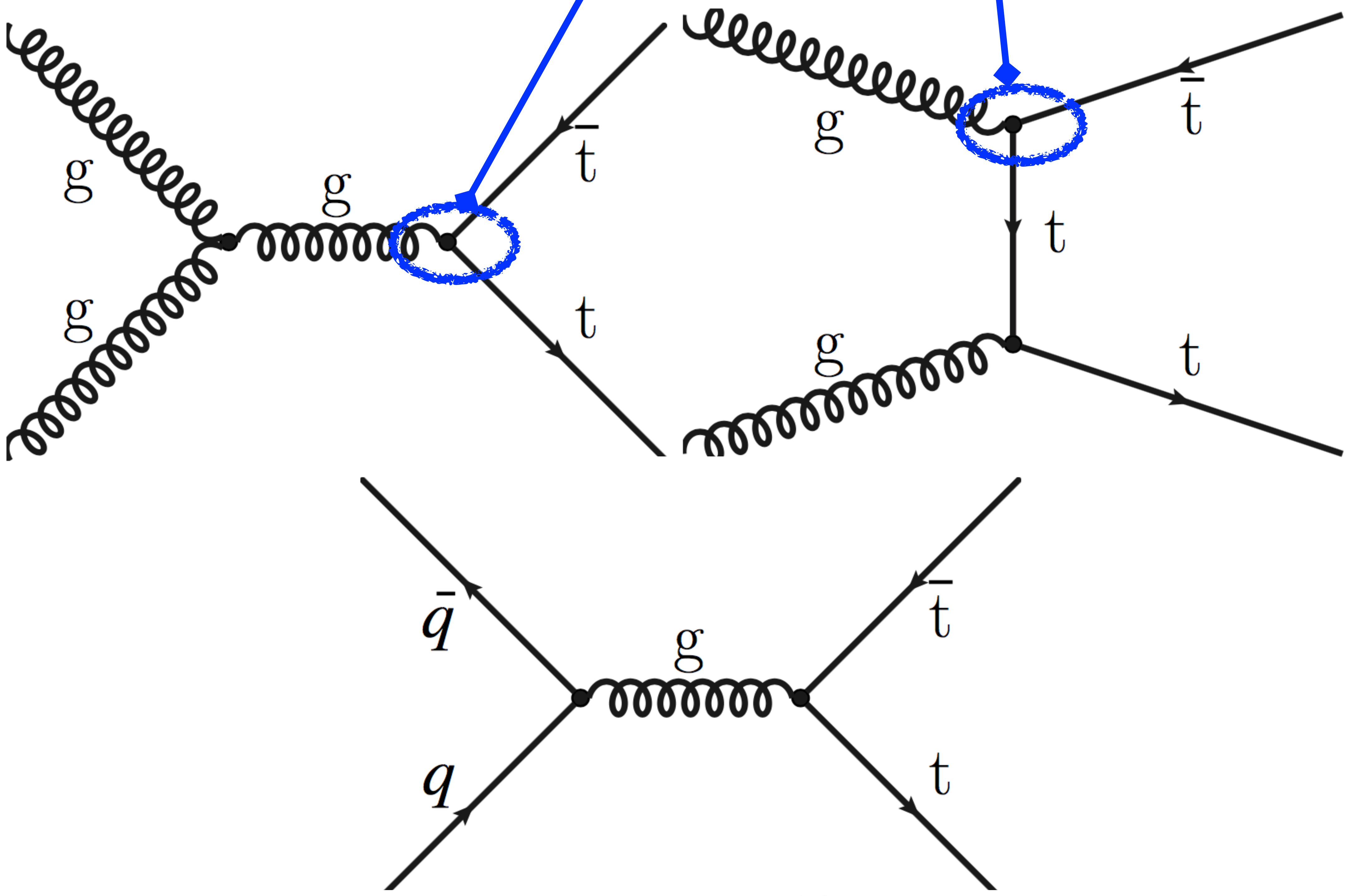
UNIVERSITÄT BONN

- Top-quark: m ~ 173GeV, τ ~ 0.5×10^{-25} < τ_{QCD} decay before hadronisation
- Production at LHC:

Strong interaction

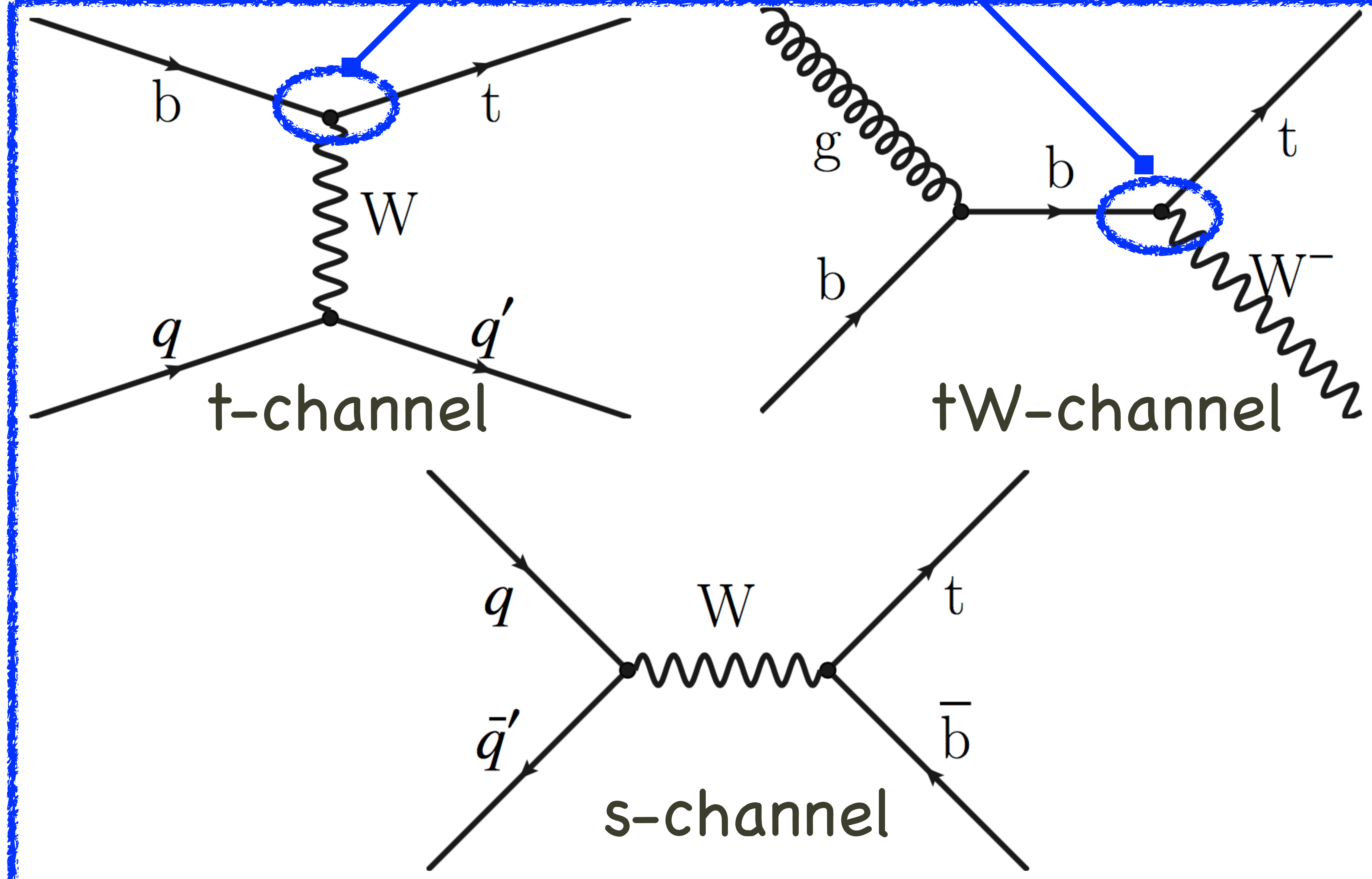
Weak interaction $\propto |V_{tb}|^2$

Pair production: $t\bar{t}$



$$\sigma_{t\bar{t}}(13\,\mathrm{GeV}) = 831.76^{+19.77}_{-29.20} \pm 35.06\,\mathrm{pb}.$$

Single production



$$\sigma_{tW\text{-channel}} = 71.7 \pm 1.8 \pm 3.4\,\mathrm{pb}$$

Top-quark production

UNIVERSITÄT BONN

- Top-quark: m ~ 173GeV, τ ~ 0.5x10^{-25} < τ_{QCD} decay before hadronisation
- Production at LHC:

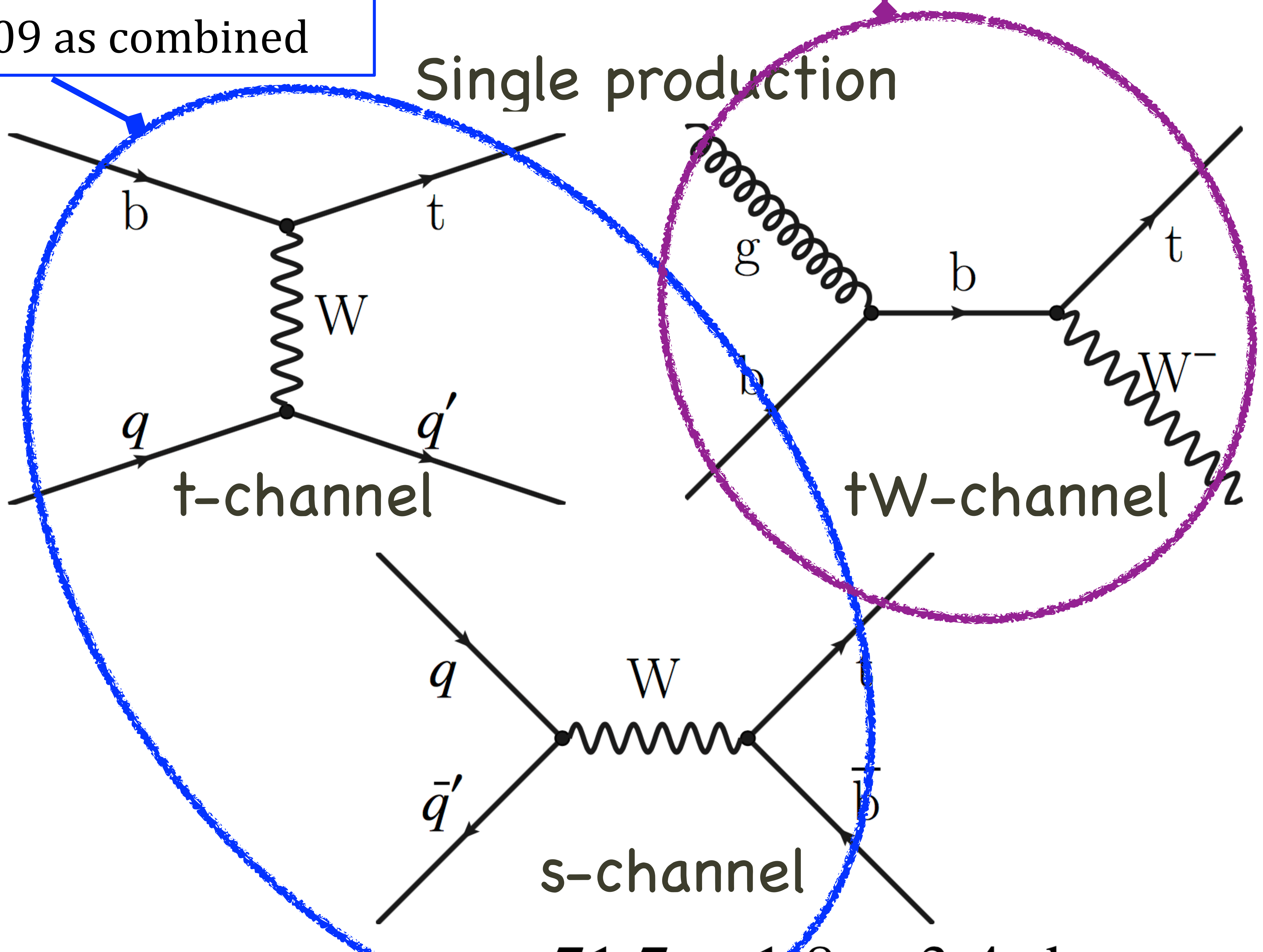
Pair production: $t\bar{t}$

$$\sigma_{t\bar{t}}(13\,\mathrm{GeV}) = 831.76\,^{+19.77}_{-29.20} \pm 35.06\,\mathrm{pb}.$$

Discovered at Tevatron in 2009 as combined

LHC is the unique place to discover it

Single production



$$\sigma_{tW\text{-channel}} = 71.7 \pm 1.8 \pm 3.4\,\mathrm{pb}$$

tW production and dileptonic decay

- Br(t→Wb) ≈ 100%
- tW & $t\bar{t}$ interference
  - Same final states: llbbνν
  - tW become even ill-defined in higher order

All-hadronic	Single-lepton	Dilepton
46.2 %	43.5 %	10.3 %

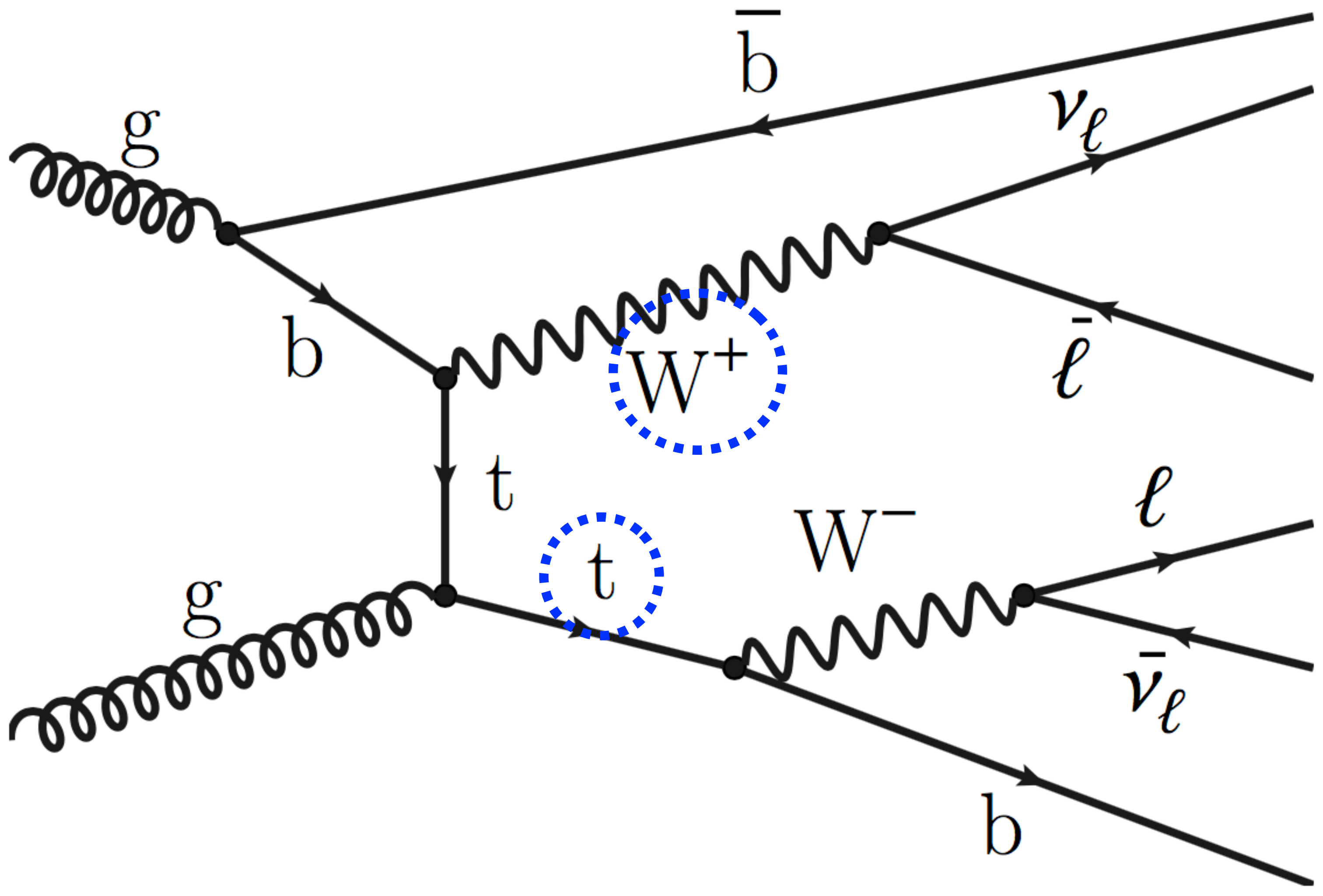
In theory:

Amplitude: $\mathcal{A} = \mathcal{A}_{tW} + \mathcal{A}_{t\bar{t}}$,

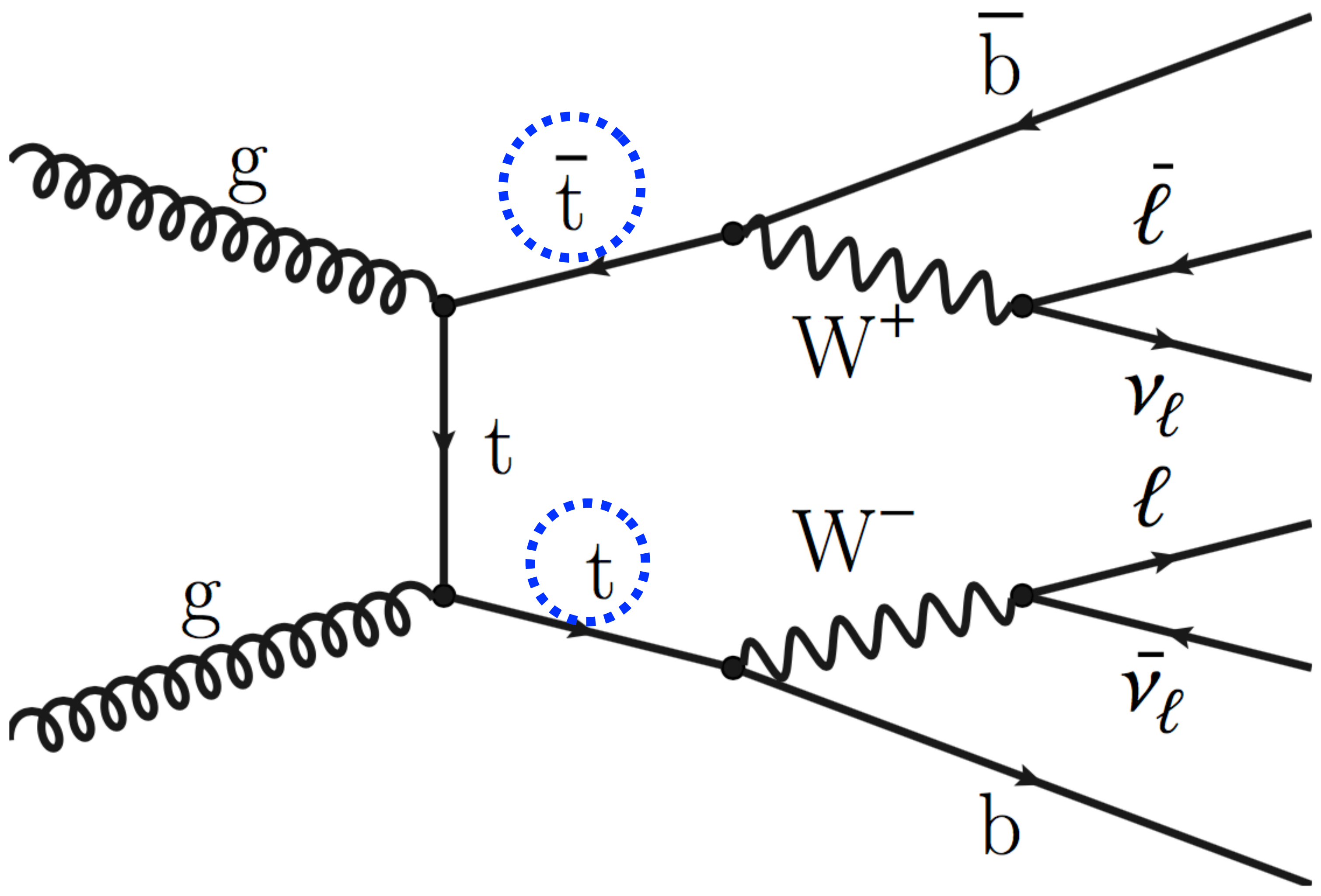
Cross-section:

$$\sigma \propto |\mathcal{A}|^2 = |\mathcal{A}_{tW}|^2 + 2\Re(\mathcal{A}_{tW}\mathcal{A}^{\dagger}_{t\bar{t}}) + |\mathcal{A}_{t\bar{t}}|^2$$
$$= \mathcal{S} + \mathcal{I} + \mathcal{B}.$$

tW-channel



$t\bar{t}$

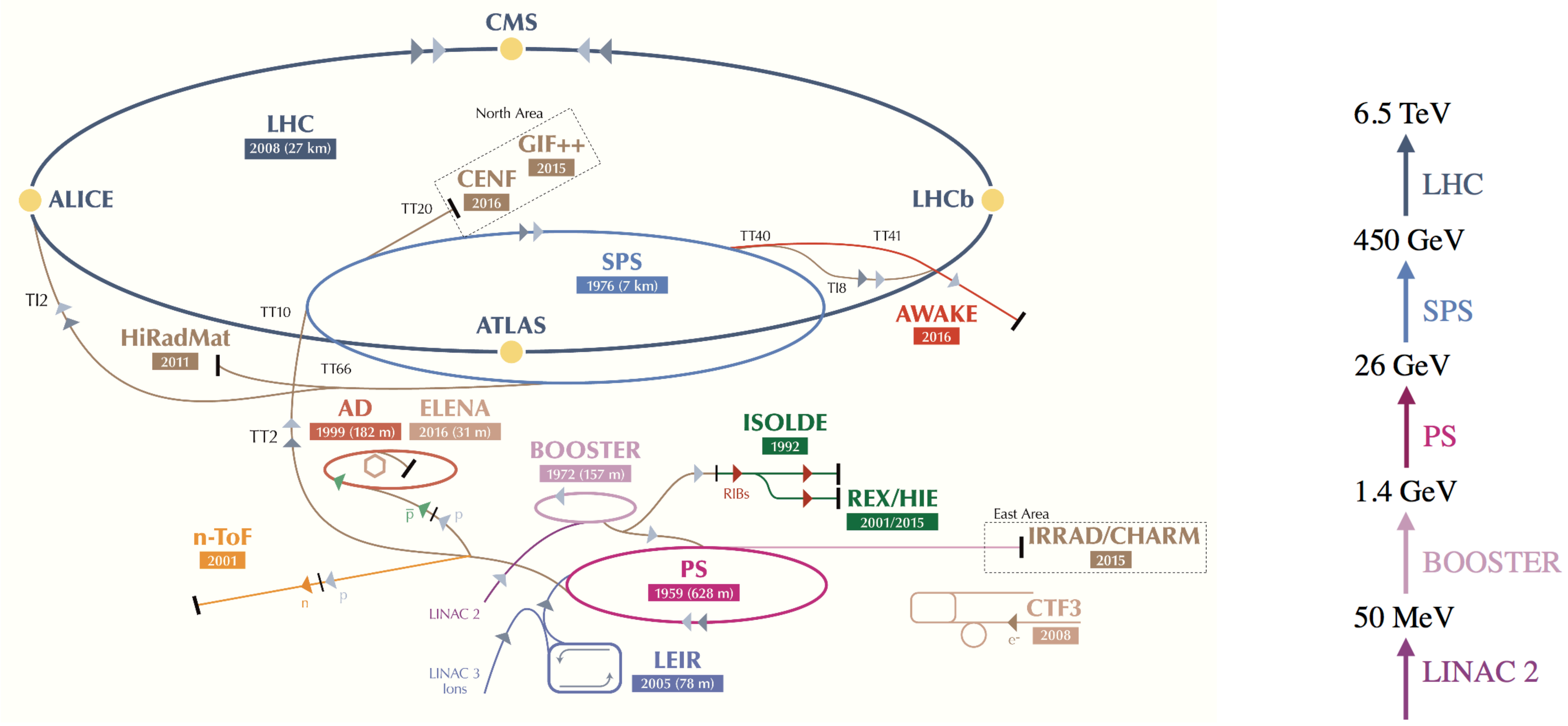


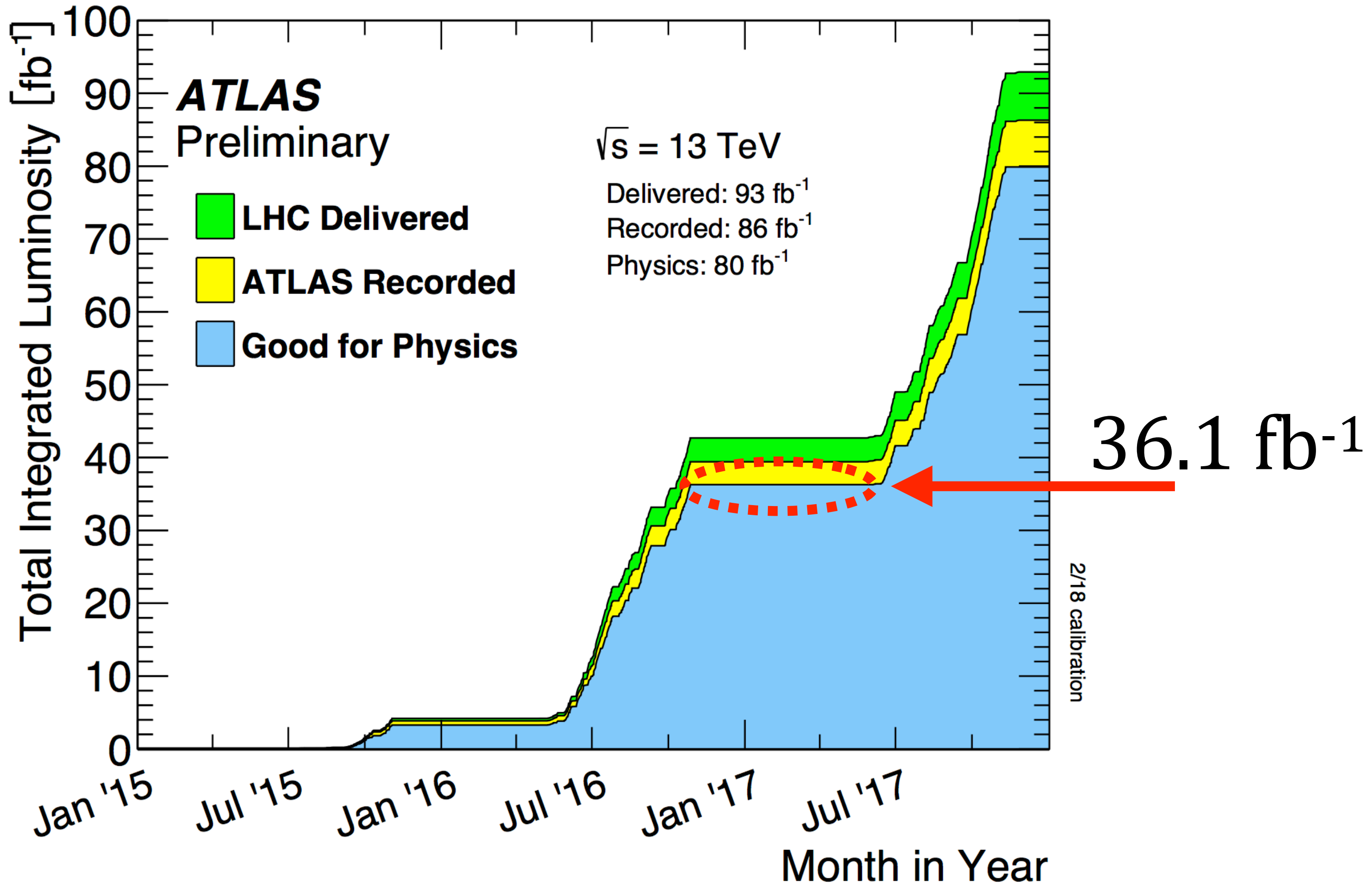
In MC generation, two schemes:

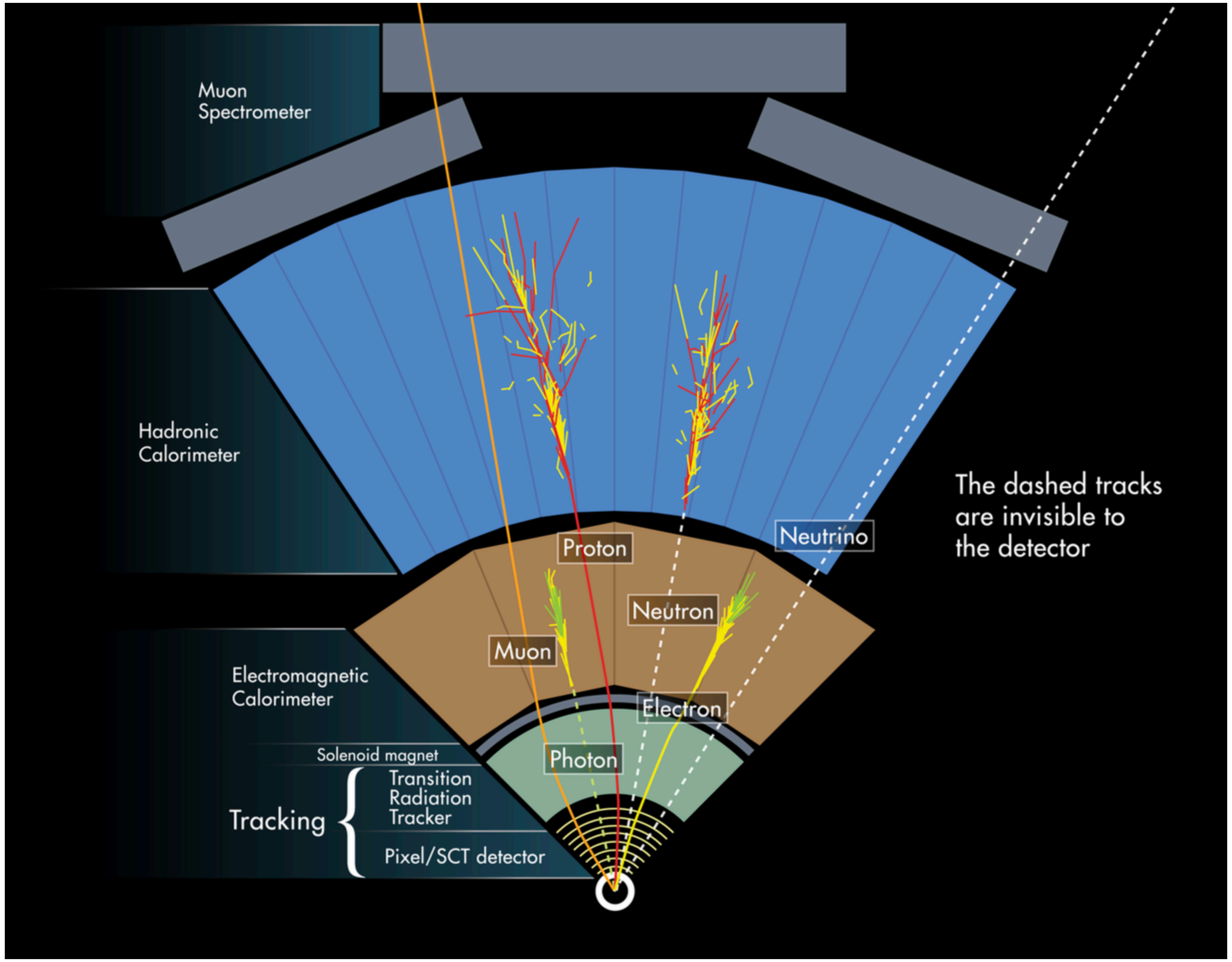
- Diagram removal: σ∝S
- Diagram subtraction: σ∝S+I+B−B′

As a systematic uncertainty

LHC and ATLAS



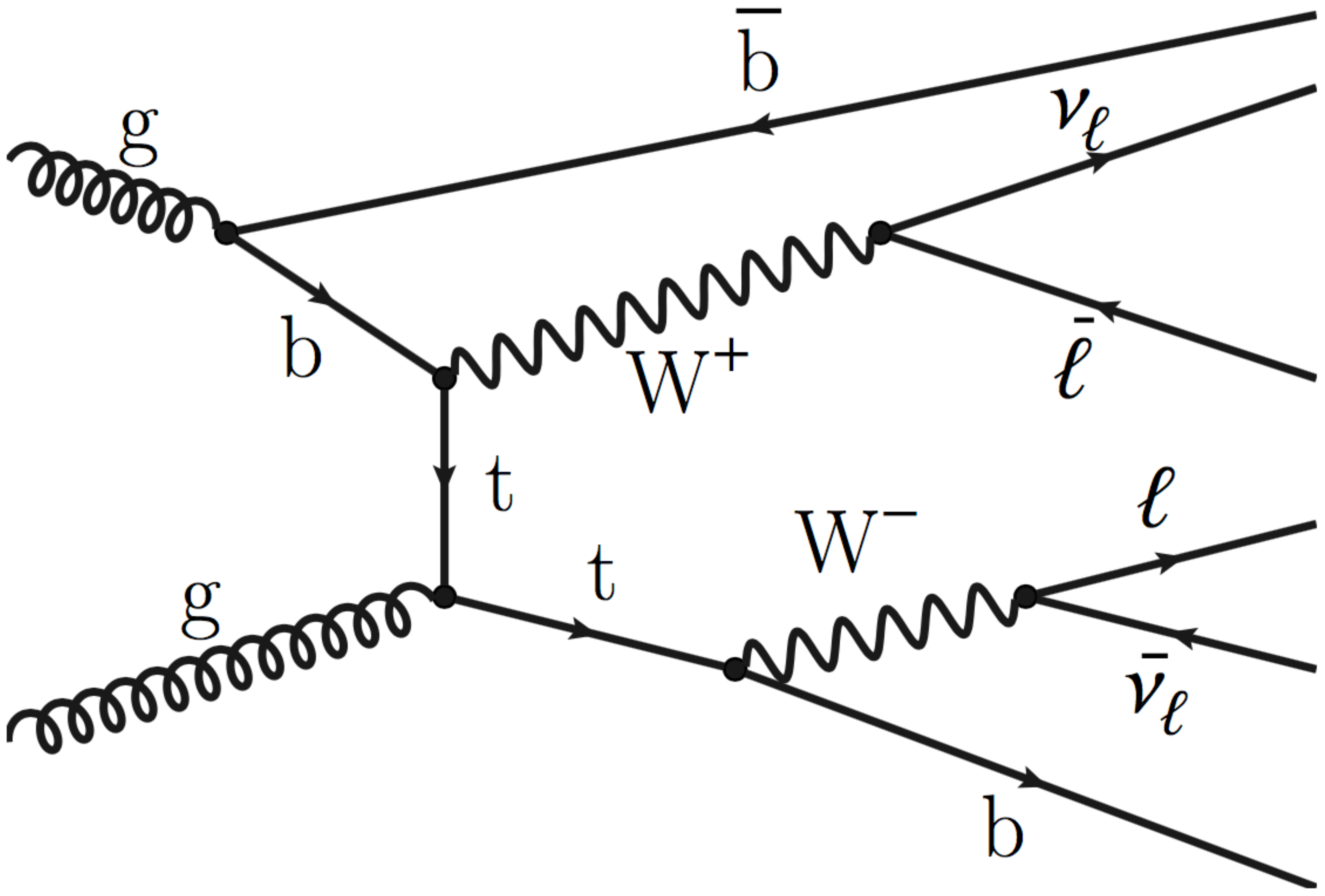




Part 2

- BDT technique
- Maximum likelihood fit method

Event selection and regions



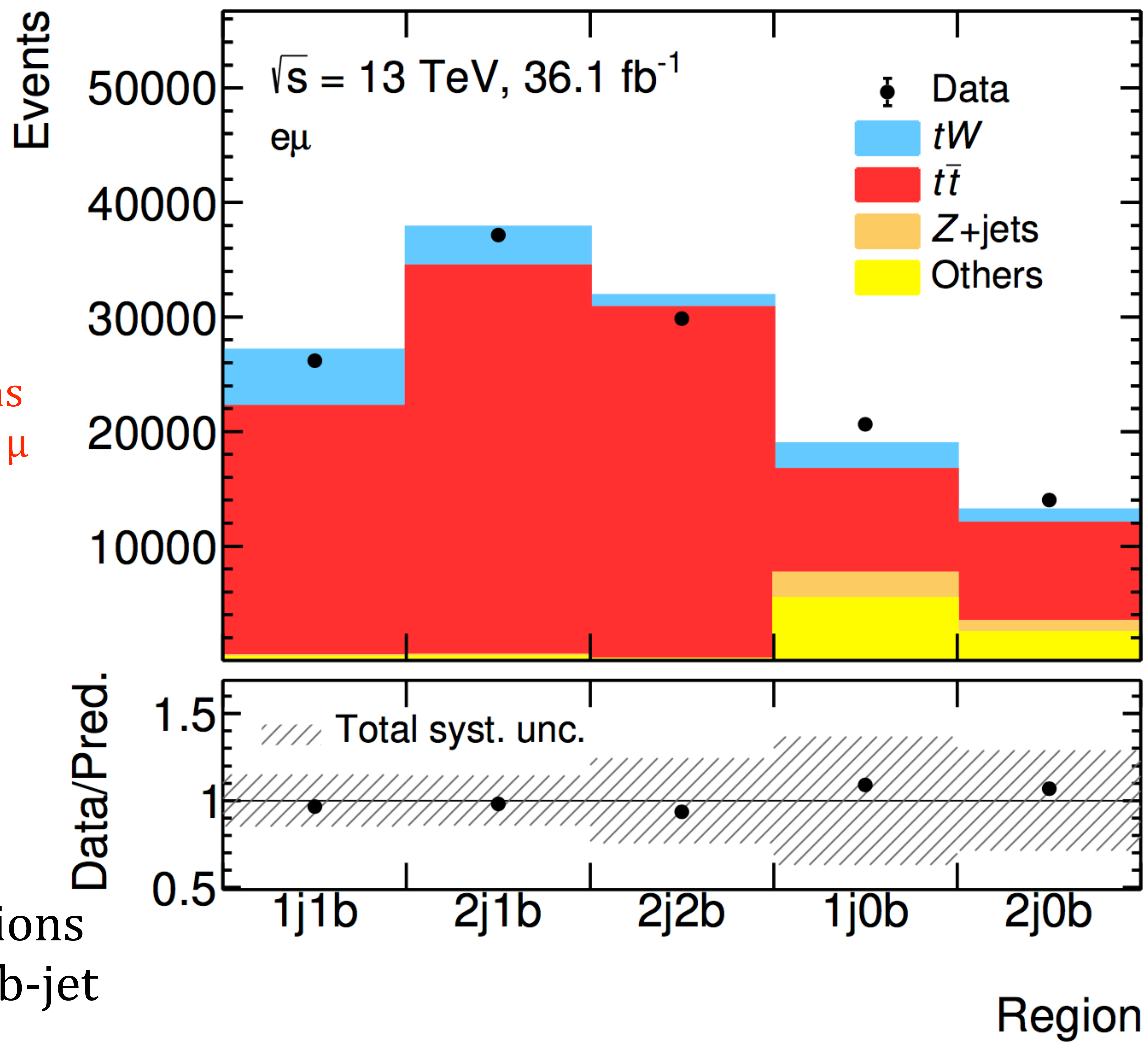
Usually out of detector

E_T^{miss}

Two leptons
Here e and μ

E_T^{miss}

One b-jet



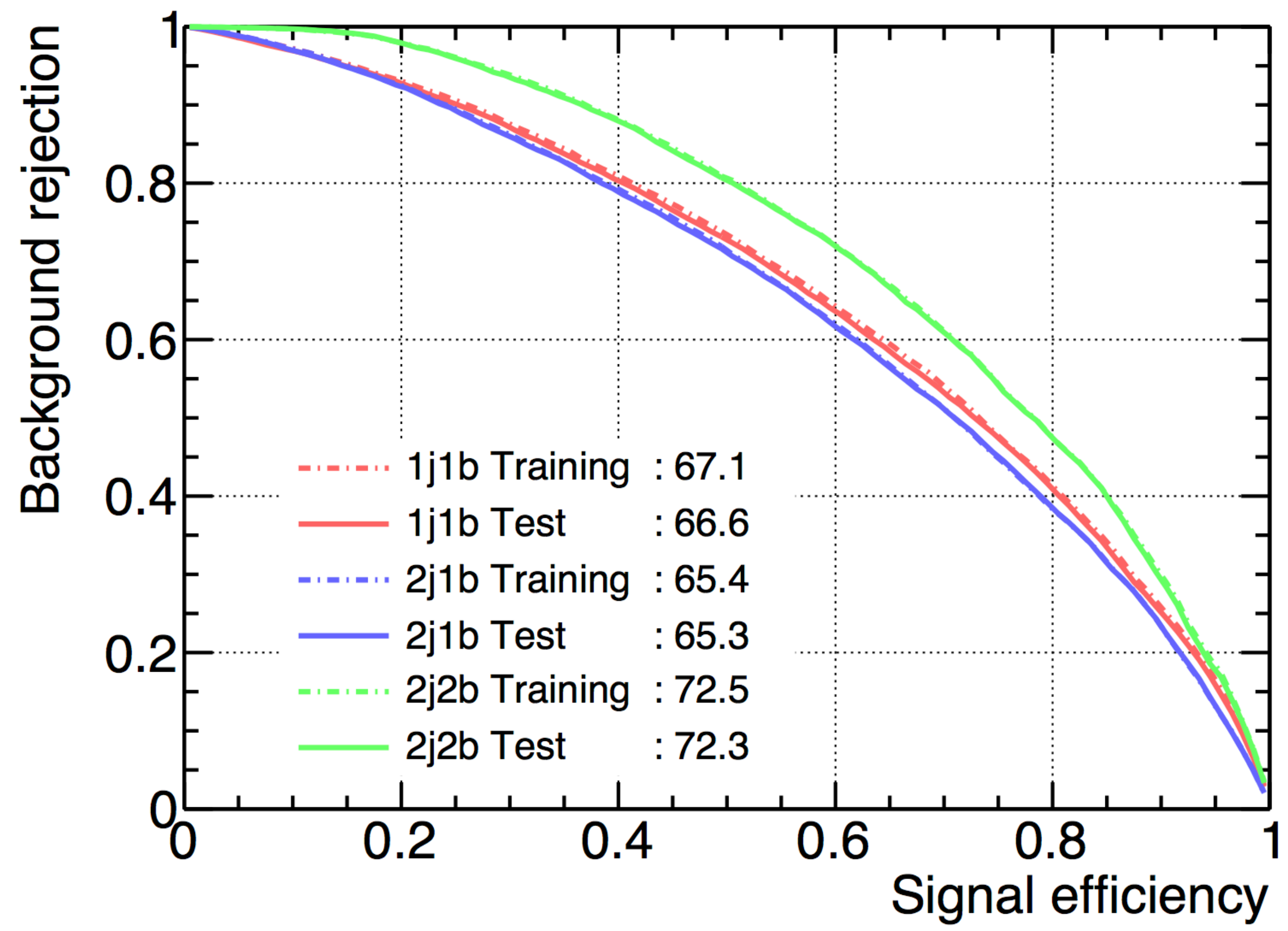
- Two opposite charged leptons
- Jets and b-jets
- E_T^{miss}

Split to different regions according to jet and b-jet multiplicities

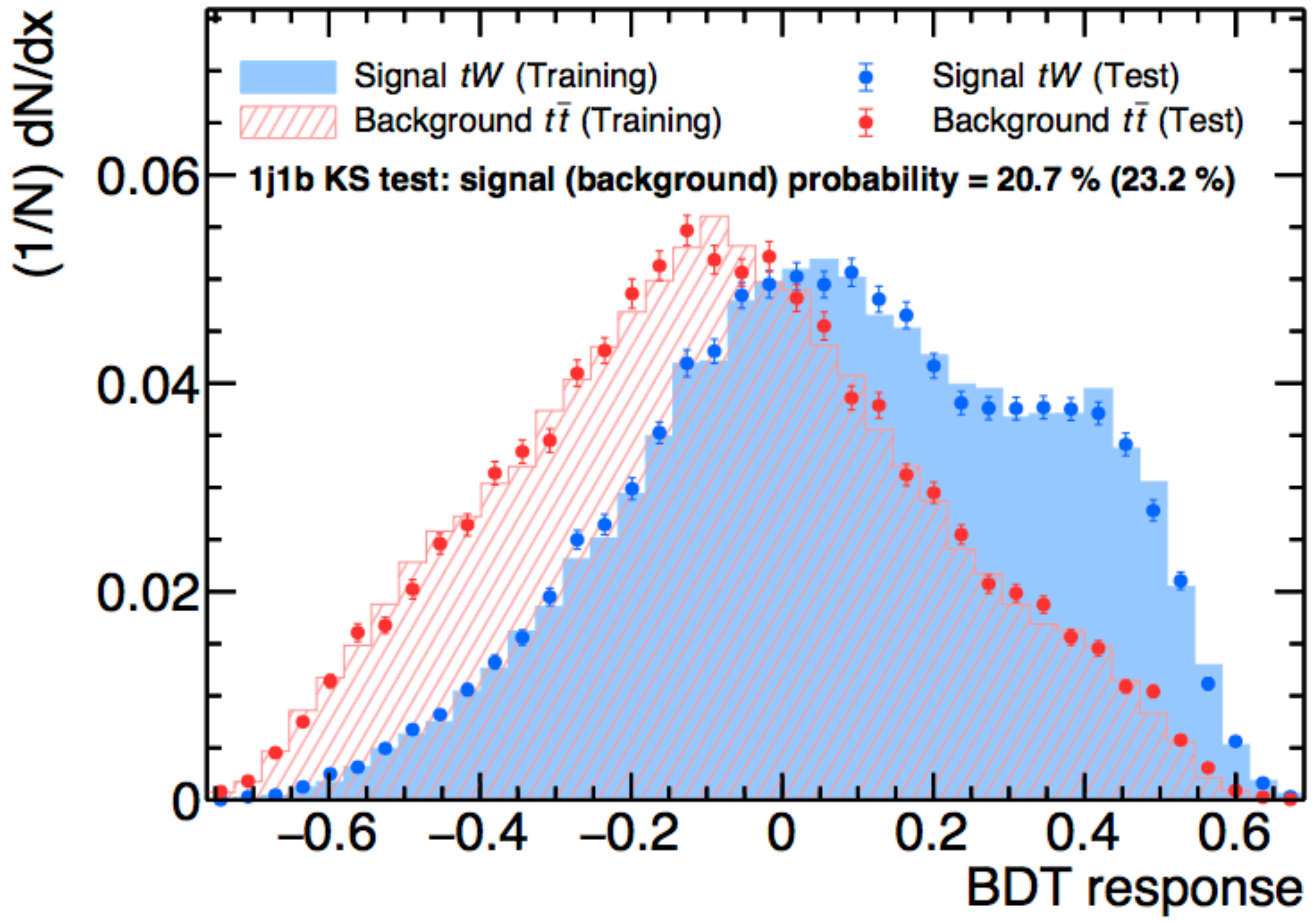
Separate signal from background - BDT

- 3 BDTs trained in 1j1b, 2j1b, and 2j2b
- Optimisation is done iteratively on variable list and hyperparameters of BDT (backup)

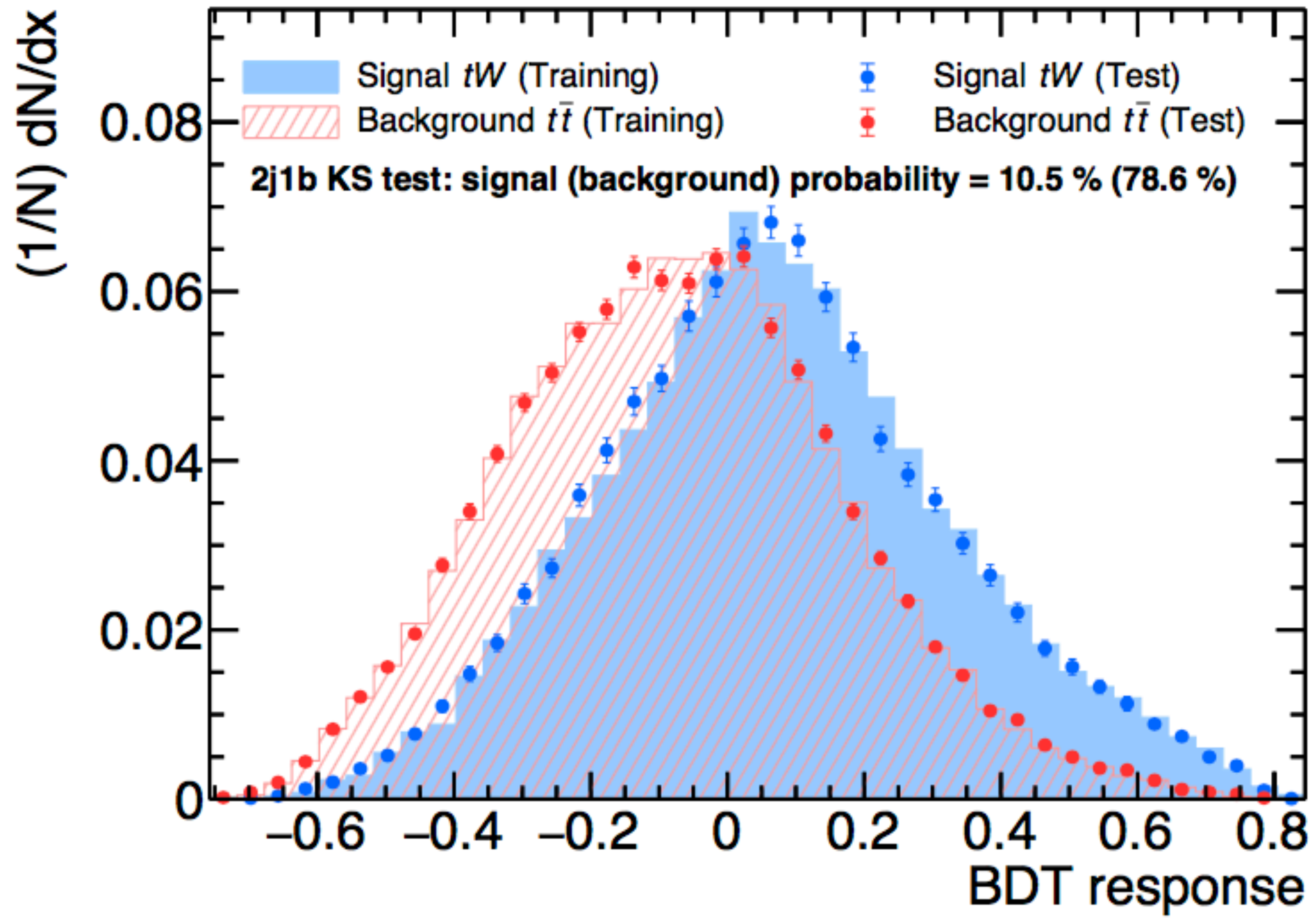
ROC curve

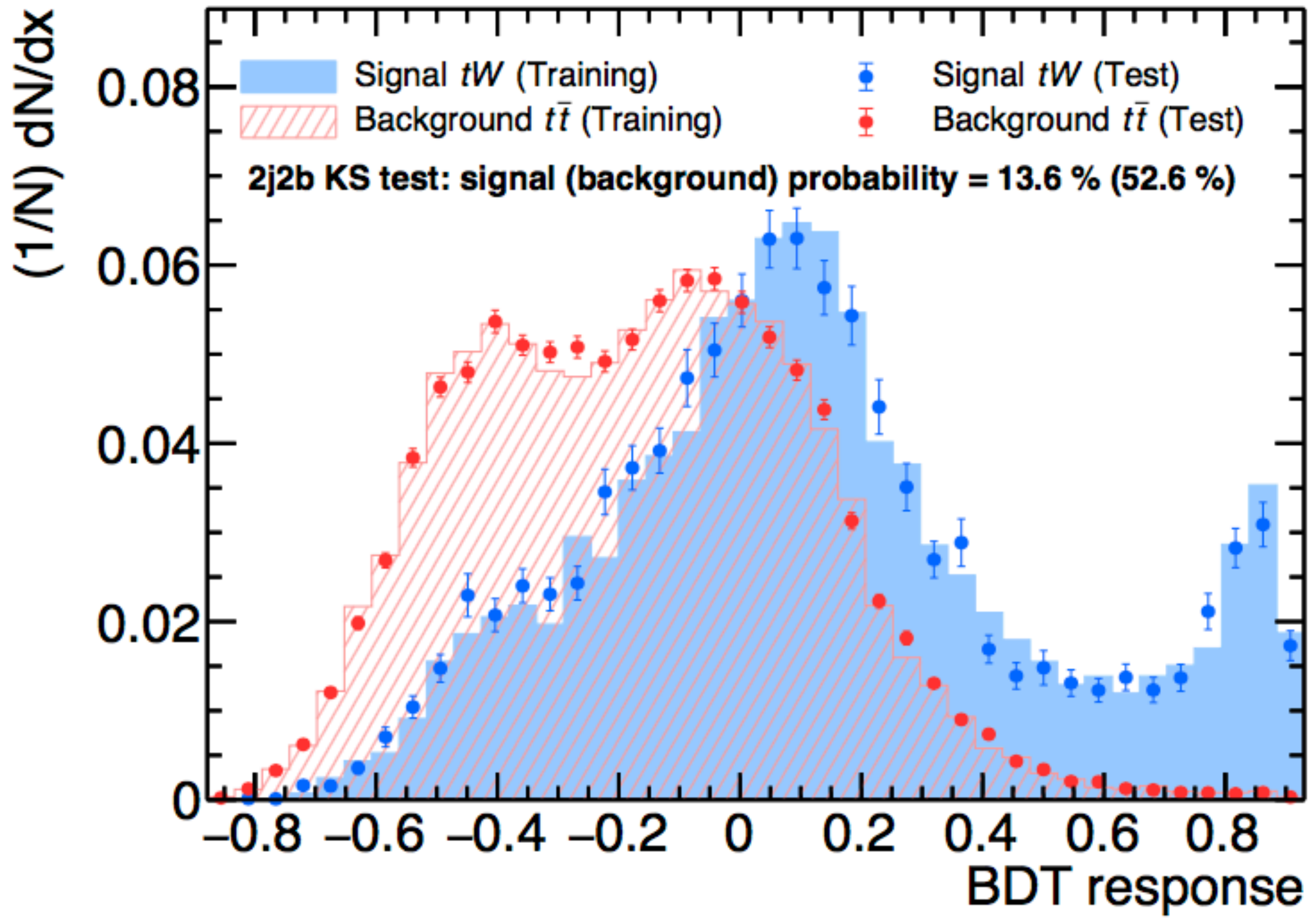


Overtraining check



Overtraining check



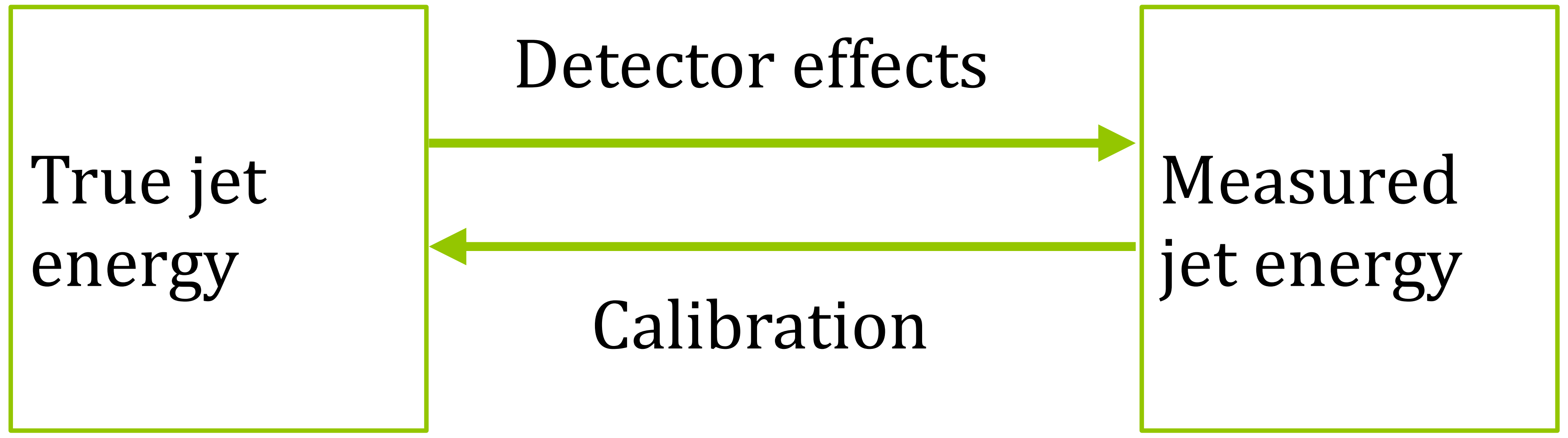


Uncertainties estimation

UNIVERSITÄT BONN

- Two categories: modelling and detector related
- Modelling related
  - Matrix element calculation (ME)/NLO subtraction
  - Parton shower (PS)/hadronisation
  - Initial- and final-state radiation (ISRFSR)
  - aDiagram removal/subtraction (DR vs DS)
  - Parton distribution function (PDF)
  - Backgrounds cross-sections
- Detector related
  - Lepton related
  - Jet energy scale (JES) and jet energy resolution (JER)
  - b-tagging
  - E_T^{miss}
  - Luminosity

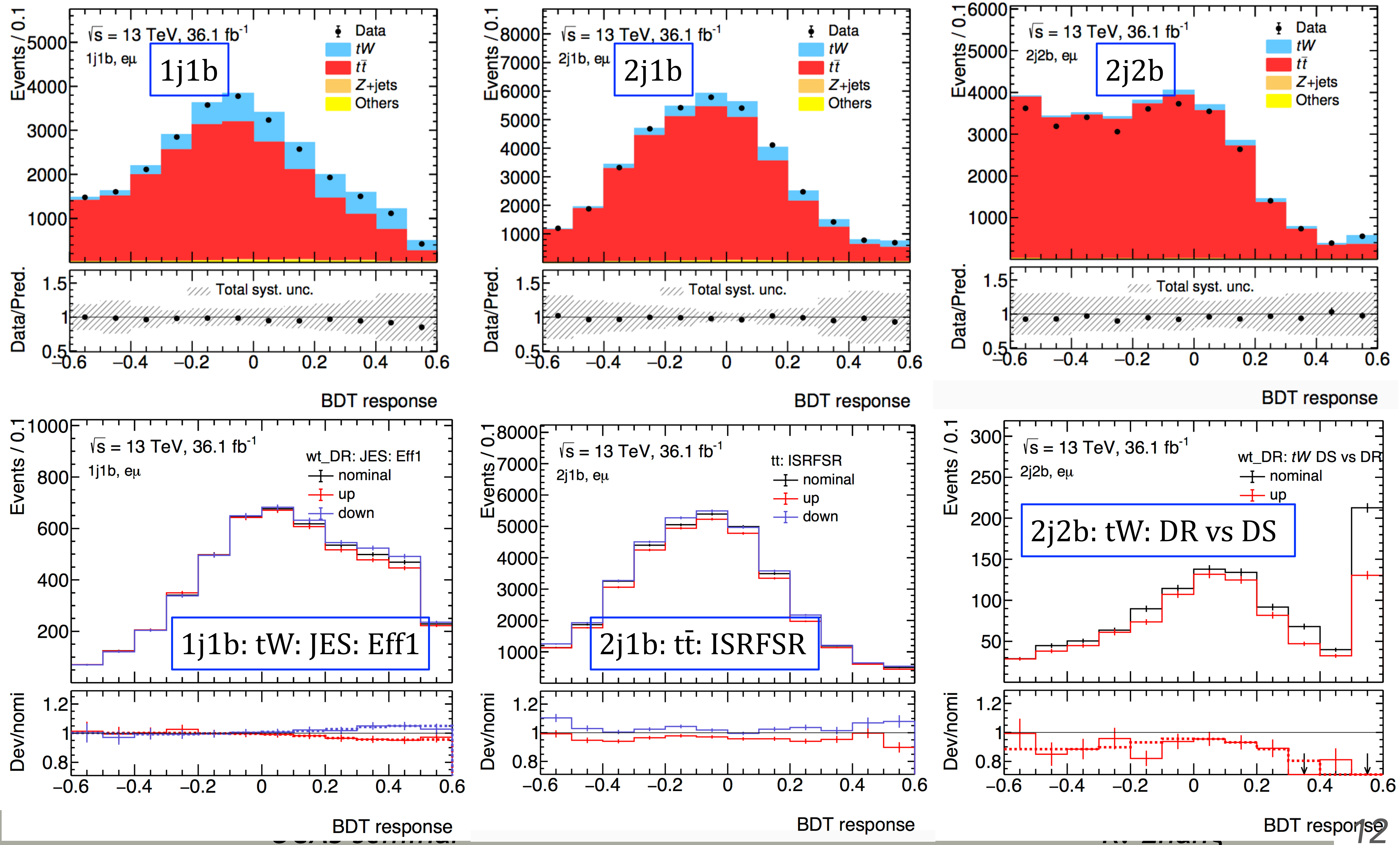
- Take JES as an example: (jet calibration and systematics in backup)



- Calibration is done in several steps; each step corrects some amount of energy.
- Each correction is parametrised by a so-called nuisance parameter (NP).
- In total 84 NPs are used in JES.
- To simplify analyses, NPs are further reduced to fewer.

Up to now, we have …

- BDT distributions for each region ☞
- Up/down variations from each systematics ☞
  - Both normalisation and shape uncertainties
- How to extract signal number/ yields/cross-section?



Maximum likelihood fit method

UNIVERSITÄT BONN

Binned likelihood model

$$\mathcal{P}(n|\lambda) = \lambda^n \frac{e^{-\lambda}}{n!}$$

Total event yields ~ Poisson

$$\mathcal{L}(\lambda) = \mathcal{P}(n|\lambda) \prod_{\text{event}}^{n} f(x)$$

Likelihood ~ Poisson×shape

$$\mathcal{L}(\lambda_r) = \prod_{r \in \text{regions}} \left[\mathcal{P}(n_r|\lambda_r) \prod_{\text{event}}^{n} f(x) \right]$$

Production in multiple regions

$$\mathcal{L}(\lambda_{r,b}) = \prod_{r \in \text{regions}} \prod_{b \in \text{bins}} \mathcal{P}(n_{r,b}|\lambda_{r,b})$$

"Binned" likelihood: no different treatment in region and bins

$$\ln \mathcal{L}(\mu) = \sum_{r \in \text{regions}} \sum_{b \in \text{bins}} \ln \mathcal{P}(n_{r,b}|\mu\lambda_{\text{sig}}^{\text{pred.}} + \lambda_{\text{bkg}}^{\text{pred.}})$$

Use "signal strength" μ as ratio of predicted and observed signal yields

Incorporating systematic uncertainties

$$\mathcal{L}(\mu, \vec{\theta}) = \prod_{r \in \text{regions}} \prod_{b \in \text{bins}} \mathcal{P}(n_{r,b}|\lambda_{r,b}(\hat{\mu}, \hat{\vec{\theta}})) \cdot \prod_{s \in \text{systematics}} \mathcal{G}(\hat{\theta}_s|\theta_s)$$

Likelihood is a function of "signal strength" and nuisance parameters (NP)

Yields can be modified by NPs as well

Can be Gaussian, Poisson, log-normal

NP θ =
0: nominal
1: up variation
-1: down variation

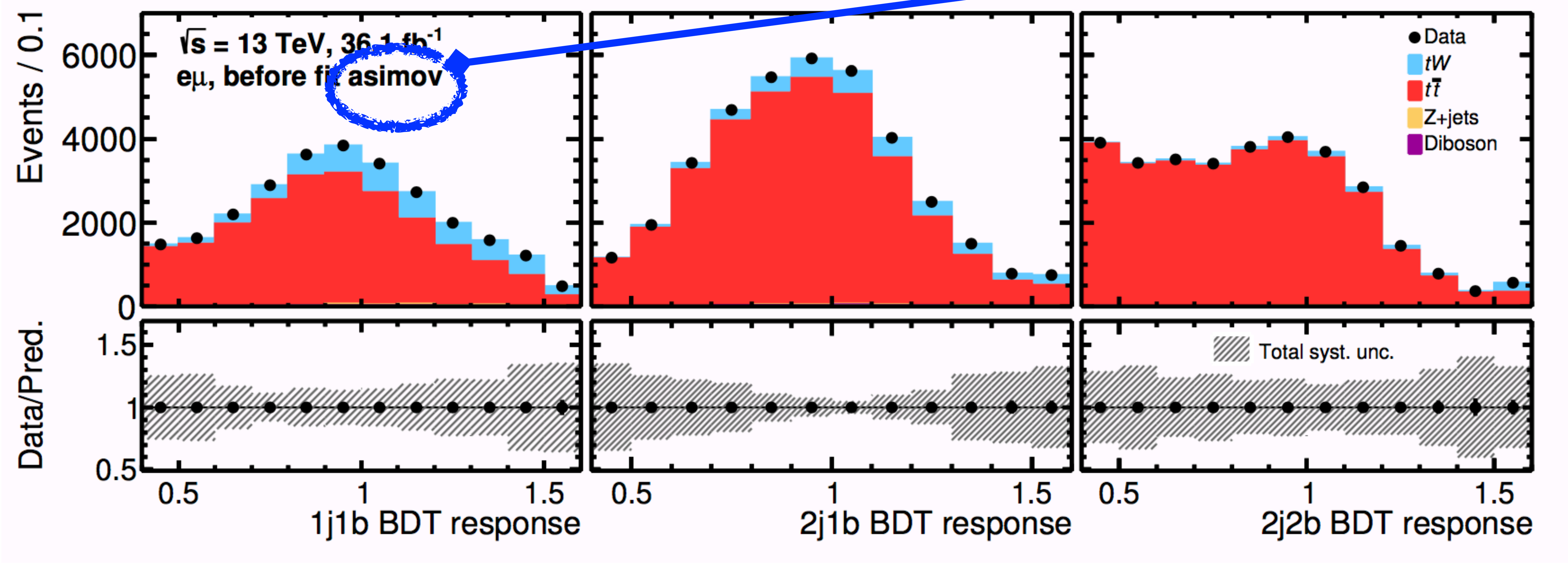
Interpolation and extrapolation

- Knowledge of what happens when $|\theta| \neq 0/\pm 1$

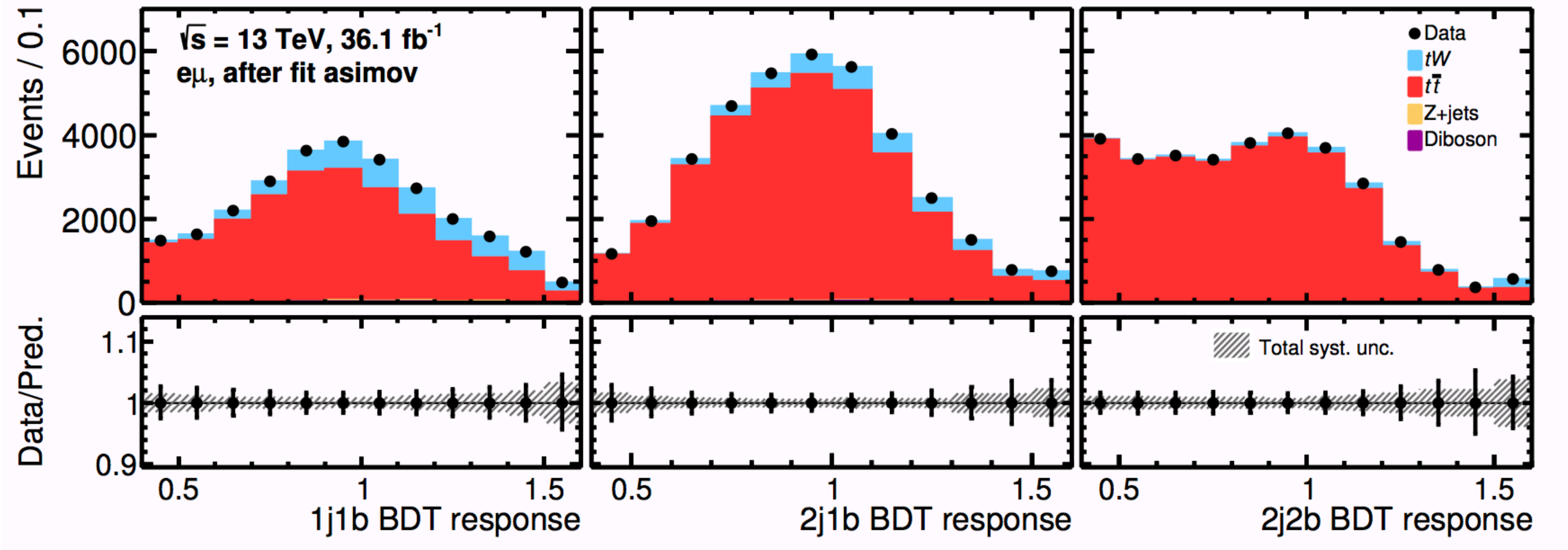
Fit results

- Asimov dataset to check fit machinery (more checks on NPs in backup)

"MC fit to MC"

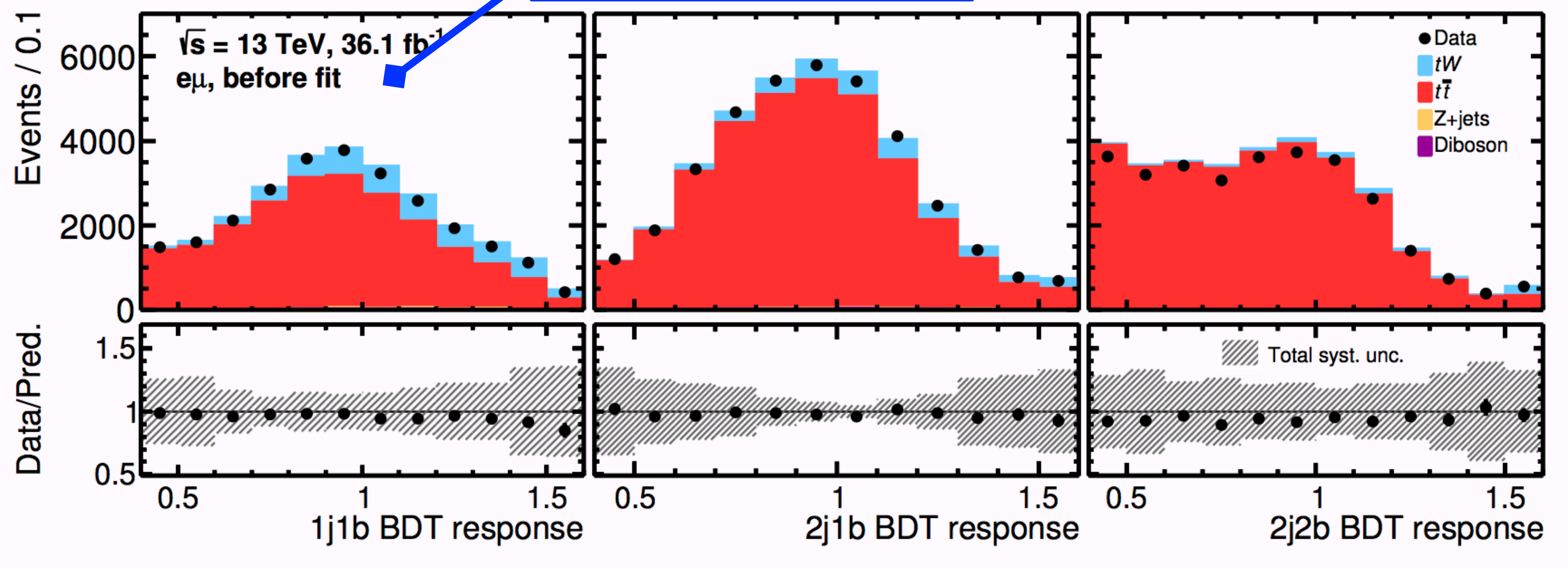


(a)

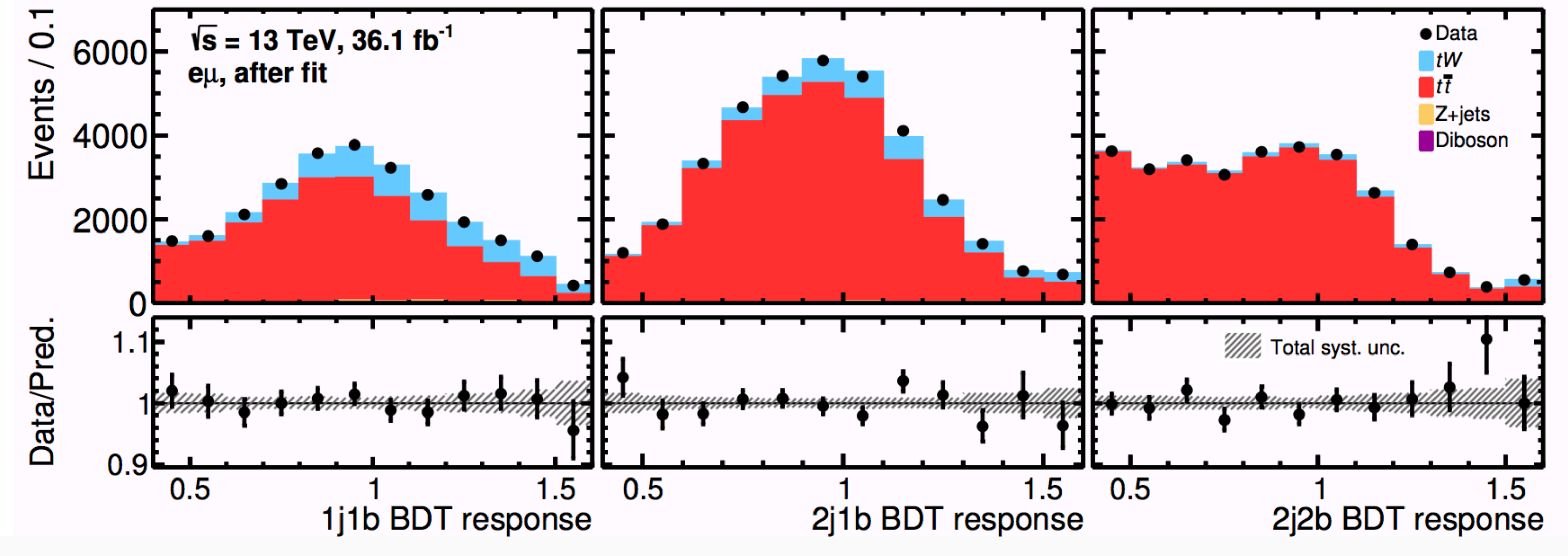


- Real dataset

"MC fit to data"



(a)



Fit results

UNIVERSITÄT BONN

	Expected	Observed
μ_{sig}	$1.00^{+0.20}_{-0.17}$	$1.14^{+0.26}_{-0.22}$
$\mu_{t\bar{t}}$	$1.000^{+0.032}_{-0.031}$	$1.004^{+0.033}_{-0.032}$

Source	$\Delta\mu_{sig}$[%]
ME	4.2
PS	3.6
ISRFSR	11
DR vs DS	19
PDF	$\lesssim$0
Non-$t\bar{t}$ background normalisation	$\lesssim$0
Lepton efficiency, energy scale and resolution	4.1
JES	21
JER	8.9
b-tagging	8.2
E_T^{miss} calculation	9.3
Luminosity	4.2
Total systematic uncertainty	23
Data statistics	3.4
Total uncertainty	23

- Measured cross-section:

$$\sigma_{tW} = 82 \pm 2.4\,(\text{stat.})^{+18}_{-15}\,(\text{syst.}) \pm 3.0\,(\text{lumi.})\,\text{pb}$$

- Theory prediction:

$$\sigma_{tW\text{-channel}} = 71.7 \pm 1.8 \pm 3.4\,\text{pb}$$

- Observed (expected) significance of 10σ (11σ)

These are estimated by fixing systematic sources to their best fit values in each category, re-fitting, and subtracting refitted uncertainty in quadrature from the total uncertainty.

Part 3

- Differential cross-section measurements

Differential cross-sections

UNIVERSITÄT BONN

- To check theoretical calculations in more details, cross-section are measured differentially depending on some kinematic observables, e.g. p_T, rapidity, etc.
- tW dilepton channel

2 neutrinos => 6 free

1 top mass, 2 W mass, E_T^{miss} x,y components => 5 constraints

→ Under-constrained system

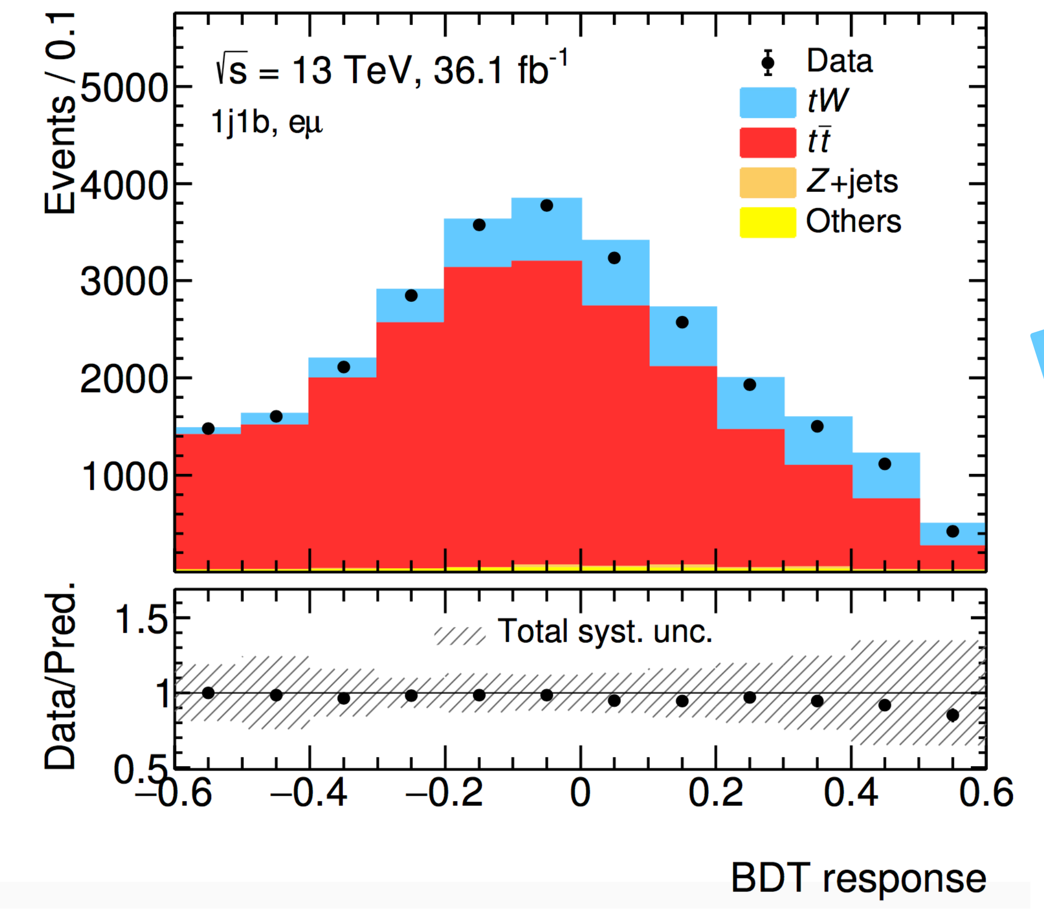
→ Unfold final-state variables

- However, still interesting to measure dependence on final observables to access top polarisation property
- Remark: best channel for differential is single leptonic tW, while suffered from huge background

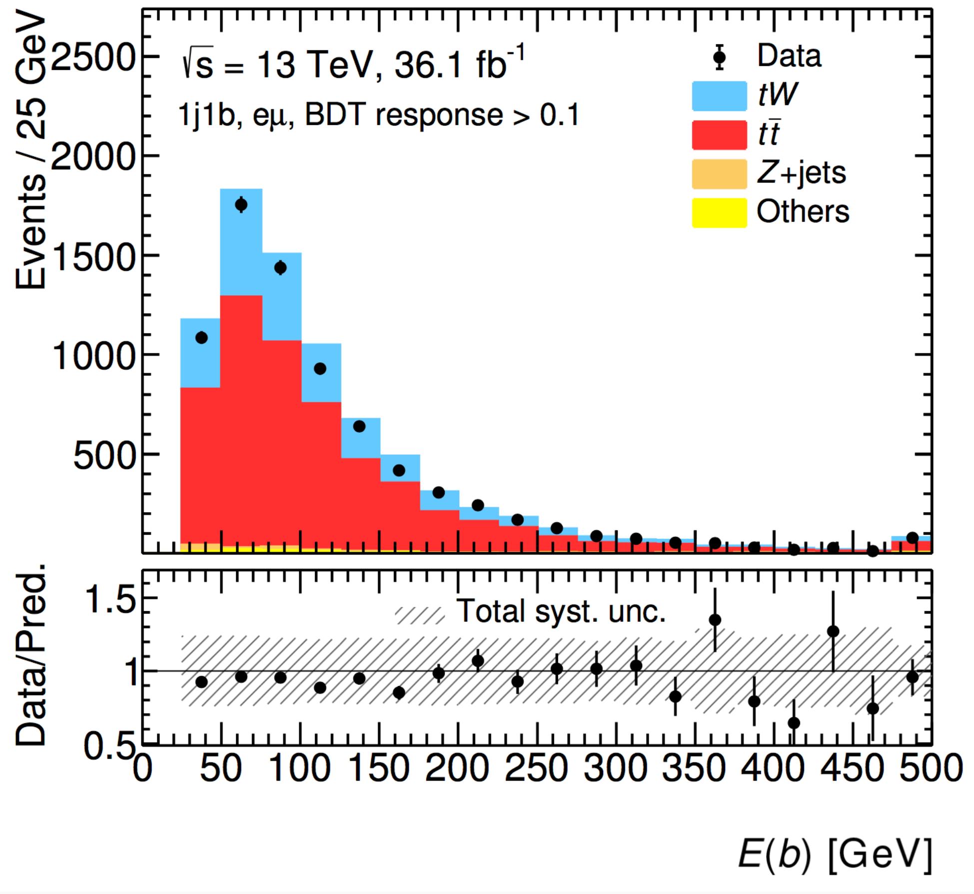
- the energy of the *b*-jet, $E(b)$;
- the mass of the leading lepton and *b*-jet, $m(\ell_1 b)$;
- the mass of the sub-leading lepton and the *b*-jet, $m(\ell_2 b)$;
- the energy of the system of the two leptons and *b*-jet, $E(\ell\ell b)$;
- the transverse mass of the leptons, *b*-jet and neutrinos, $m_T(\ell\ell\nu\nu b)$; and
- the mass of the two leptons and the *b*-jet, $m(\ell\ell b)$.

Analysis strategy

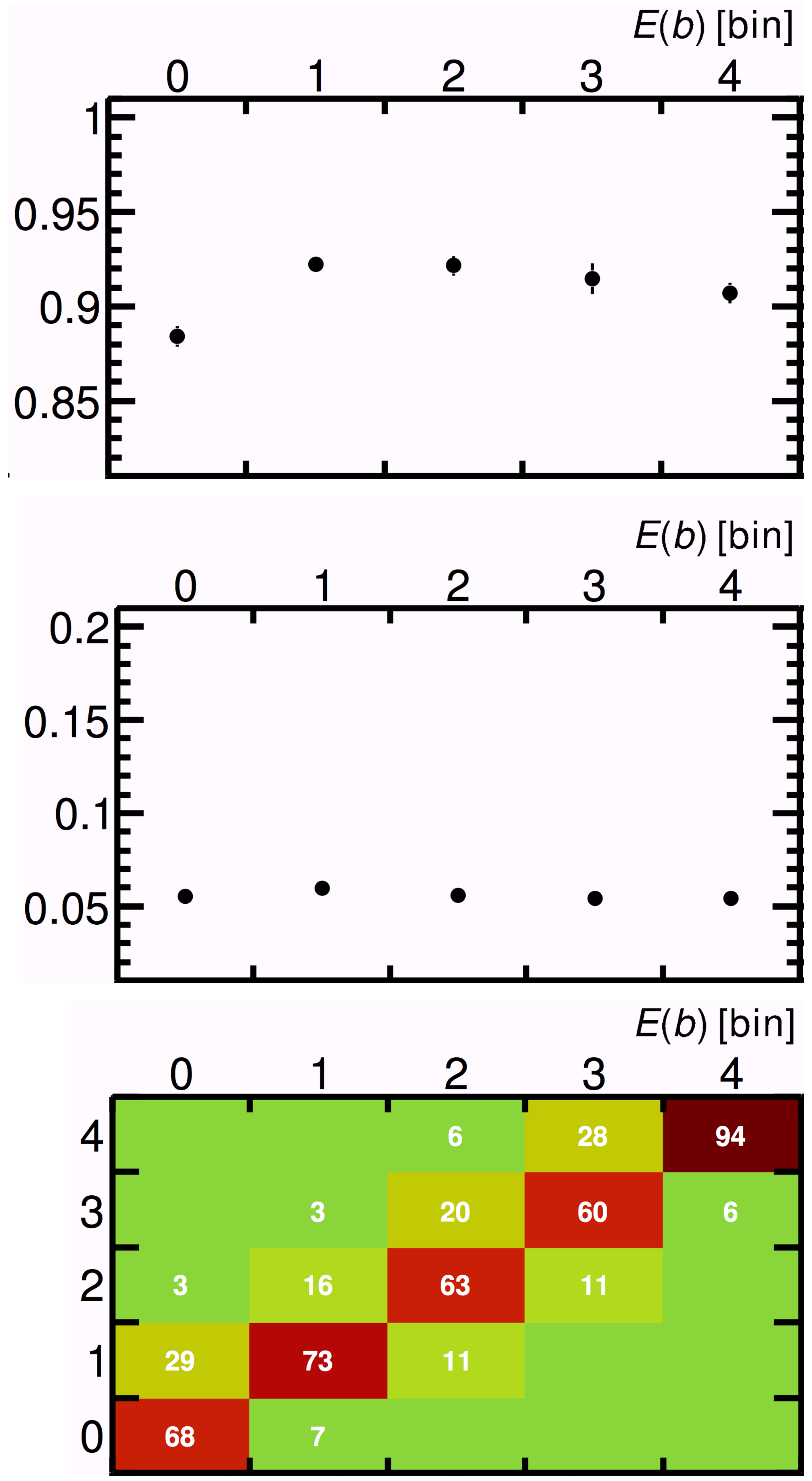
Cut on BDT response



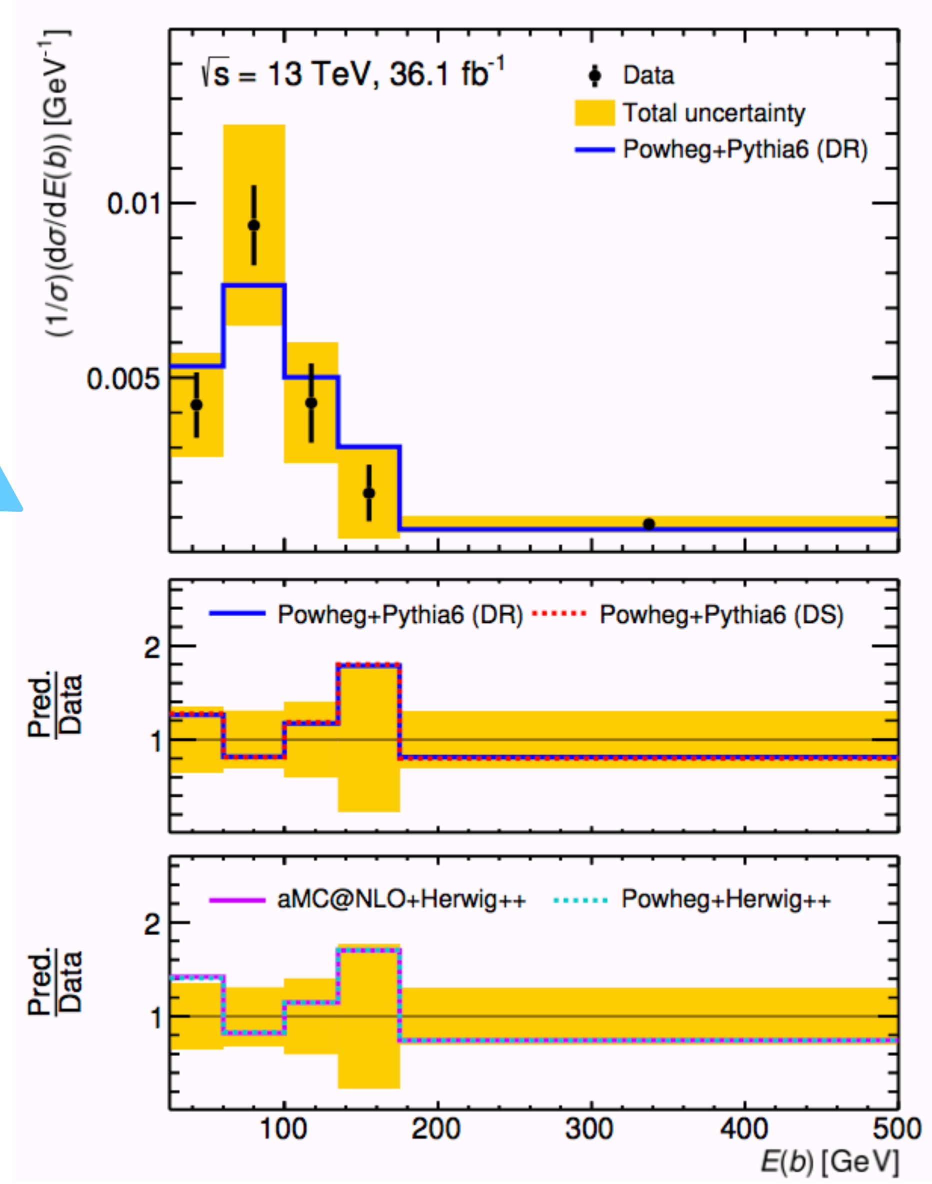
Subtract background



Calculate response matrix and correction factors



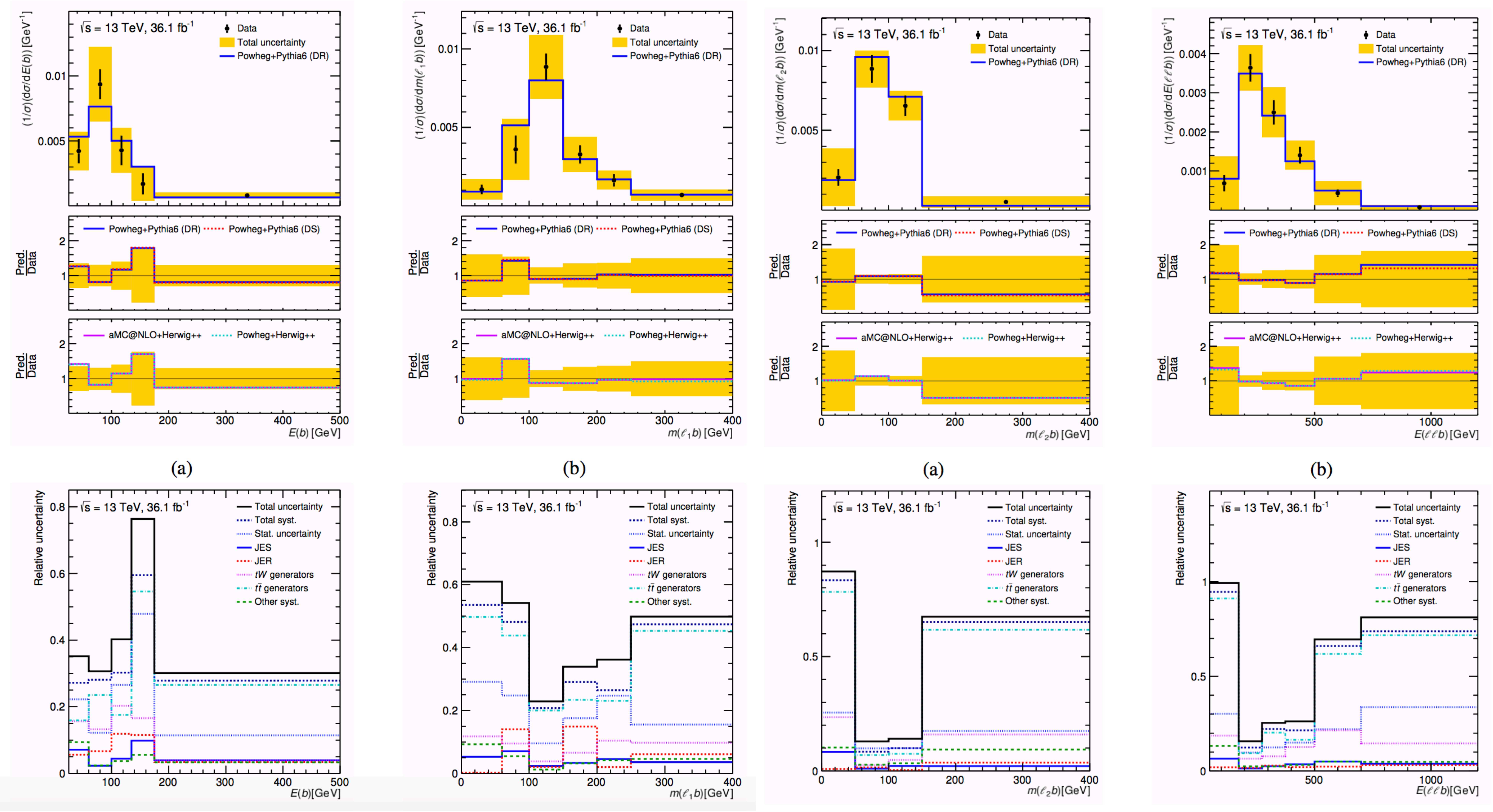
Cross-section



- Systematic uncertainty
  - Response matrix etc. from tW nominal
  - Unfold data -> differential cross-section
  - Unfold Syst -> syst = (unf-truth)/truth

Results

- Predictions agree with observation



Part 4

- Adversarial neural network

Systematic treatment in analysis

UNIVERSITÄT BONN

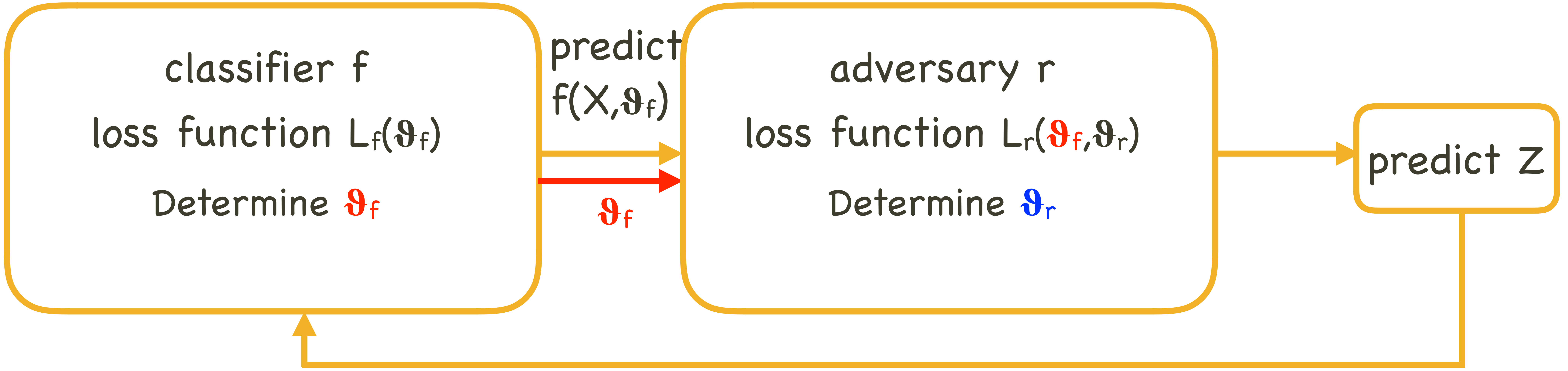
- Signal-background separation
  - MVA combines multiple observables to one classifier to gain separation
  - Training procedure done on large Monte-Carlo samples that target is known
  - Usually nominal distributions enter training, then apply to systematic variation
- Maximum likelihood fit
  - Nuisance parameter (NP) assigned for each systematic uncertainty
  - NP together with signal strength determined by fitting
- Pros
  - Easy to interpret
  - Fast training
- Cons
  - Absence of cross-talk between systematics and training

Adversarial networks

UNIVERSITÄT BONN

- Adversarial neural networks (ANNs)
  - Two nets, pitting one against the other
  - An analogy: an art forger & an art expert
  - Forger/generator produces realistic image while expert/discriminator receives both forgeries and real (authentic) images, and aims to tell them apart
  - Trained simultaneously
- ANN's huge potential
  - Can learn to mimic any distribution of data
  - Incorporate systematic uncertainties (next slide)

ANN as interpreter of systematic uncertainty

- Build a classifier f as usual, minimising a loss function $L_f(\vartheta_f)$, giving prediction $f(X,\vartheta_f)$, where X is the input of the classifier
- Pit f against an adversary network r, minimising a joint loss function $L_r(\vartheta_f,\vartheta_r)$ producing as output function a function $p(Z|f(X,\vartheta_f) = s)$, where Z is the NP value
  - If the adversary can predict the NP from the classifier's output, then it means that some information about the NP is carried through it: the classifier is dependent on the systematics
- 

Gilles Louppe

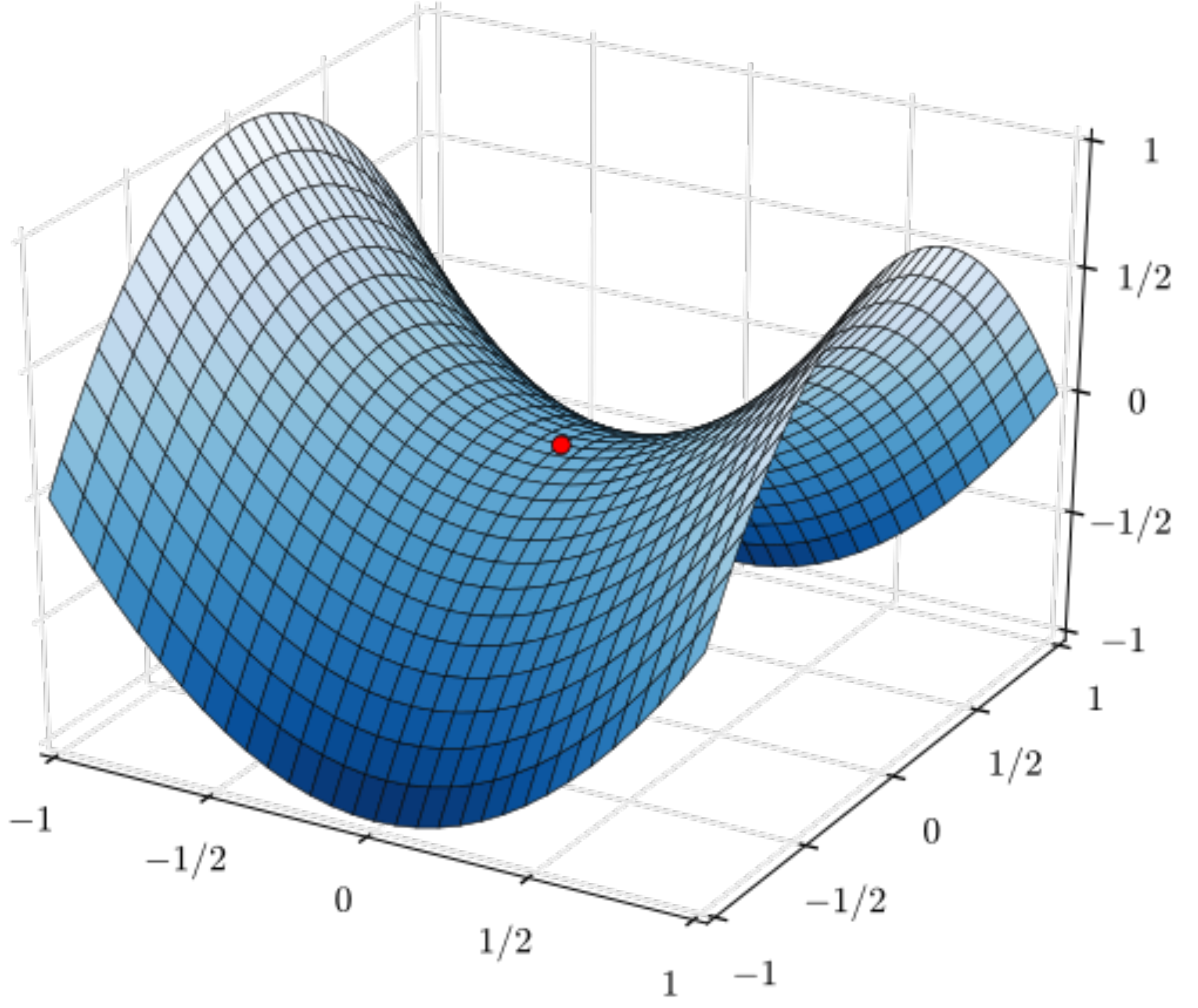
Adversarial training

UNIVERSITÄT BONN

- What if the classifier forces the adversary to perform worse by simultaneously maximising L_r? It should reduce its dependence on the nuisance parameter
  - minmax problem -> find the saddle point
- Constrain loss function
  - $E(\vartheta_f,\vartheta_r) = L_f(\vartheta_f) - L_r(\vartheta_f,\vartheta_r)$
- optimise it by finding the minimax solution

$$\hat{\theta}_f, \hat{\theta}_r = \arg\min_{\theta_f}\max_{\theta_r} E(\theta_f, \theta_r)$$

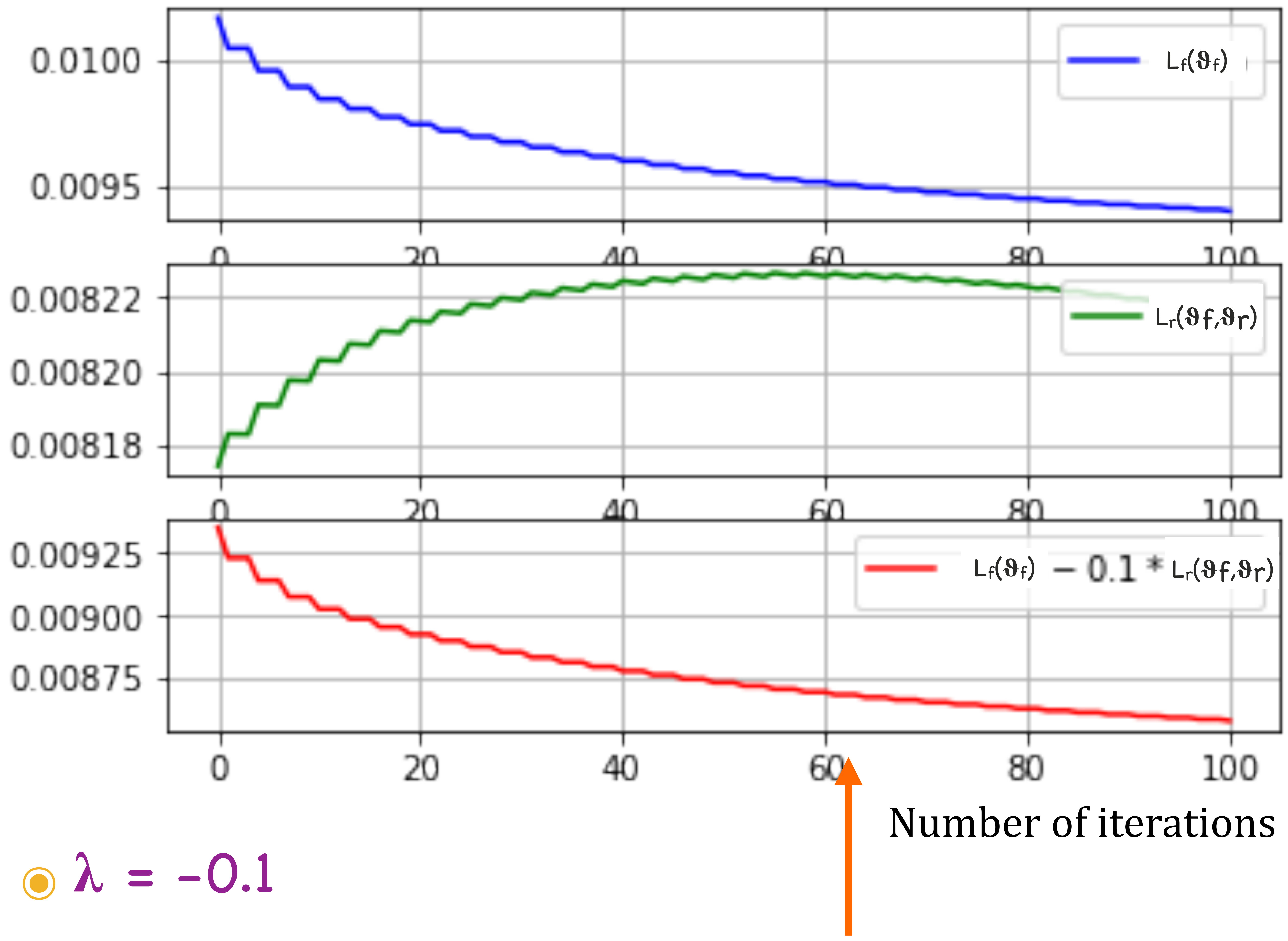
- Accuracy versus robustness trade-off: rewrite loss function:
  - $E(\vartheta_f,\vartheta_r) = L_f(\vartheta_f) - \lambda L_r(\vartheta_f,\vartheta_r)$



A saddle point (in red) on the graph of $z=x^2-y^2$

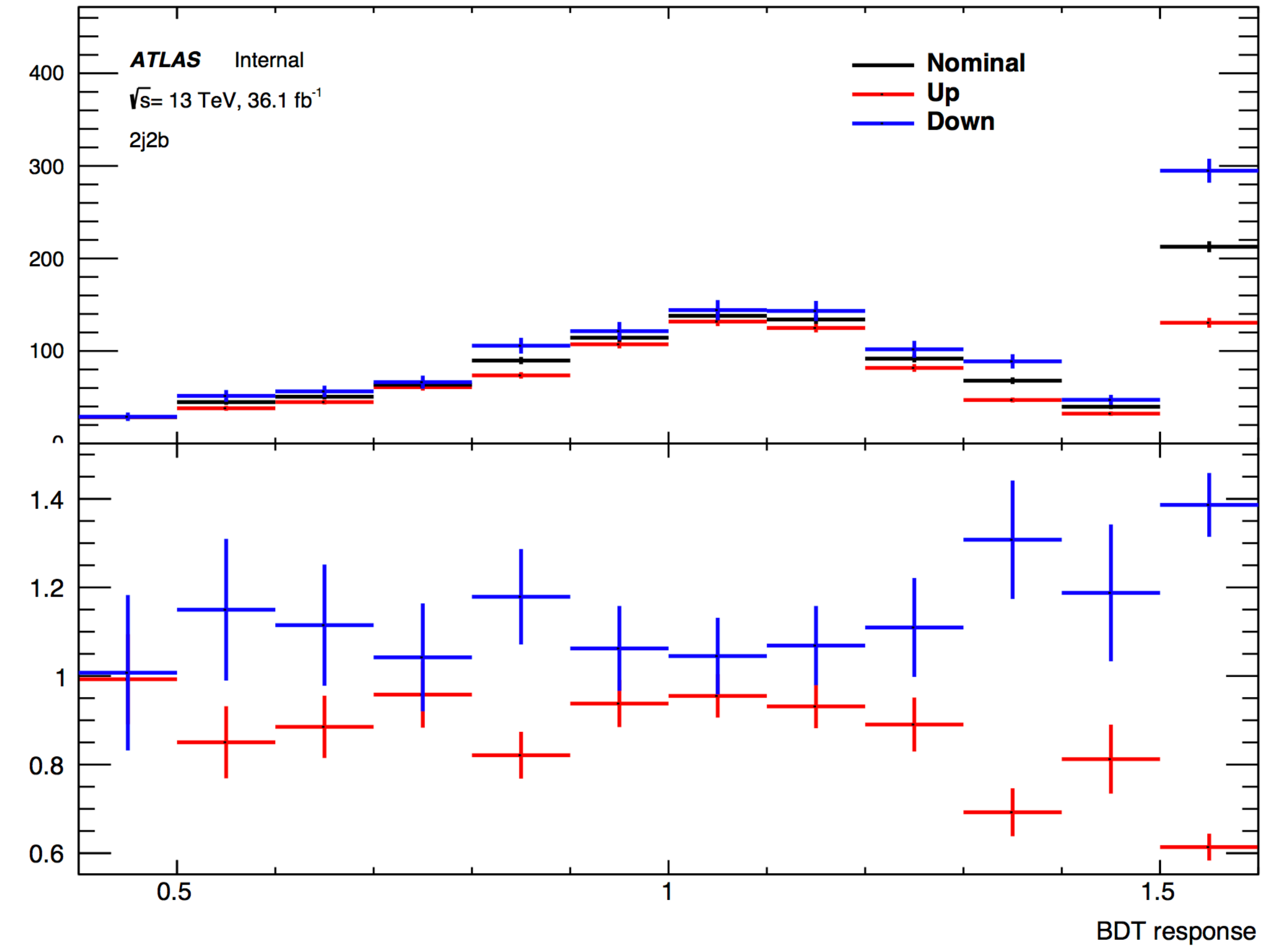
- where λ is a hyper-parameter controlling the trade-off between the performance of f and its independence with respect to the NP
  - Setting λ to a large value enforces f to be pivotal
  - Setting λ close to 0 constraints f to be optimal

Preliminary training

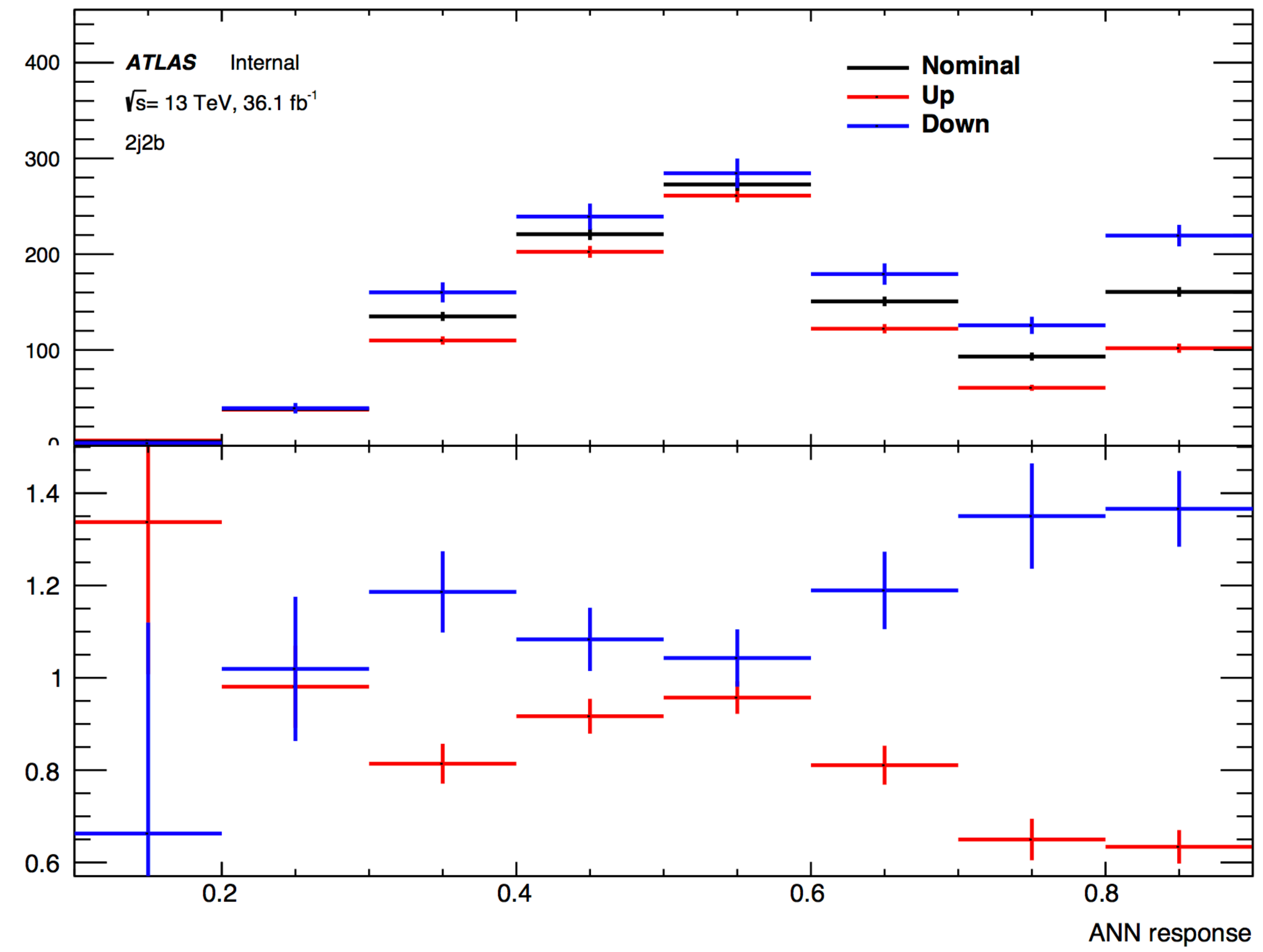


- $\lambda = -0.1$
- N = 61

BDT 2j2b



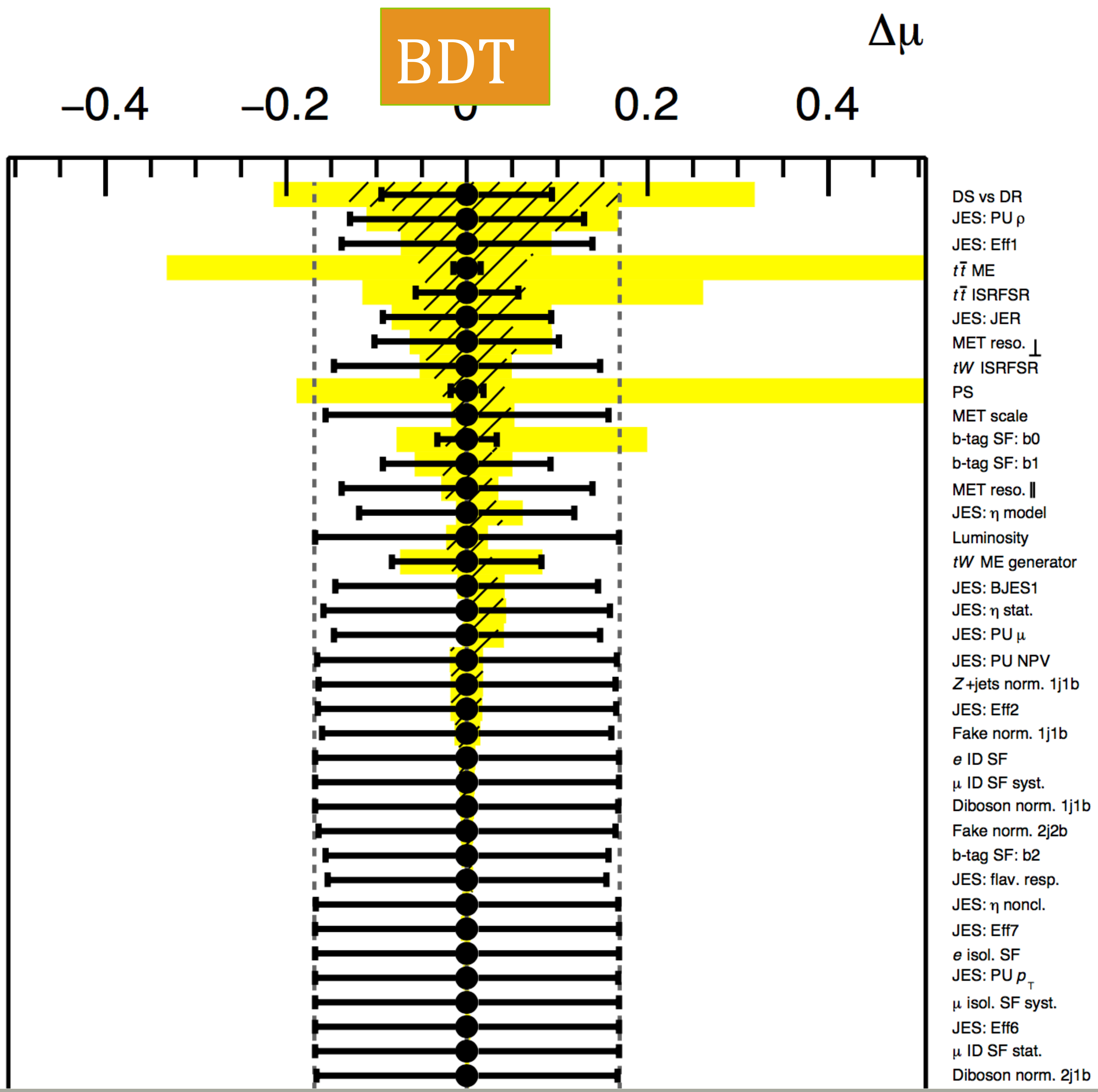
ANN 2j2b



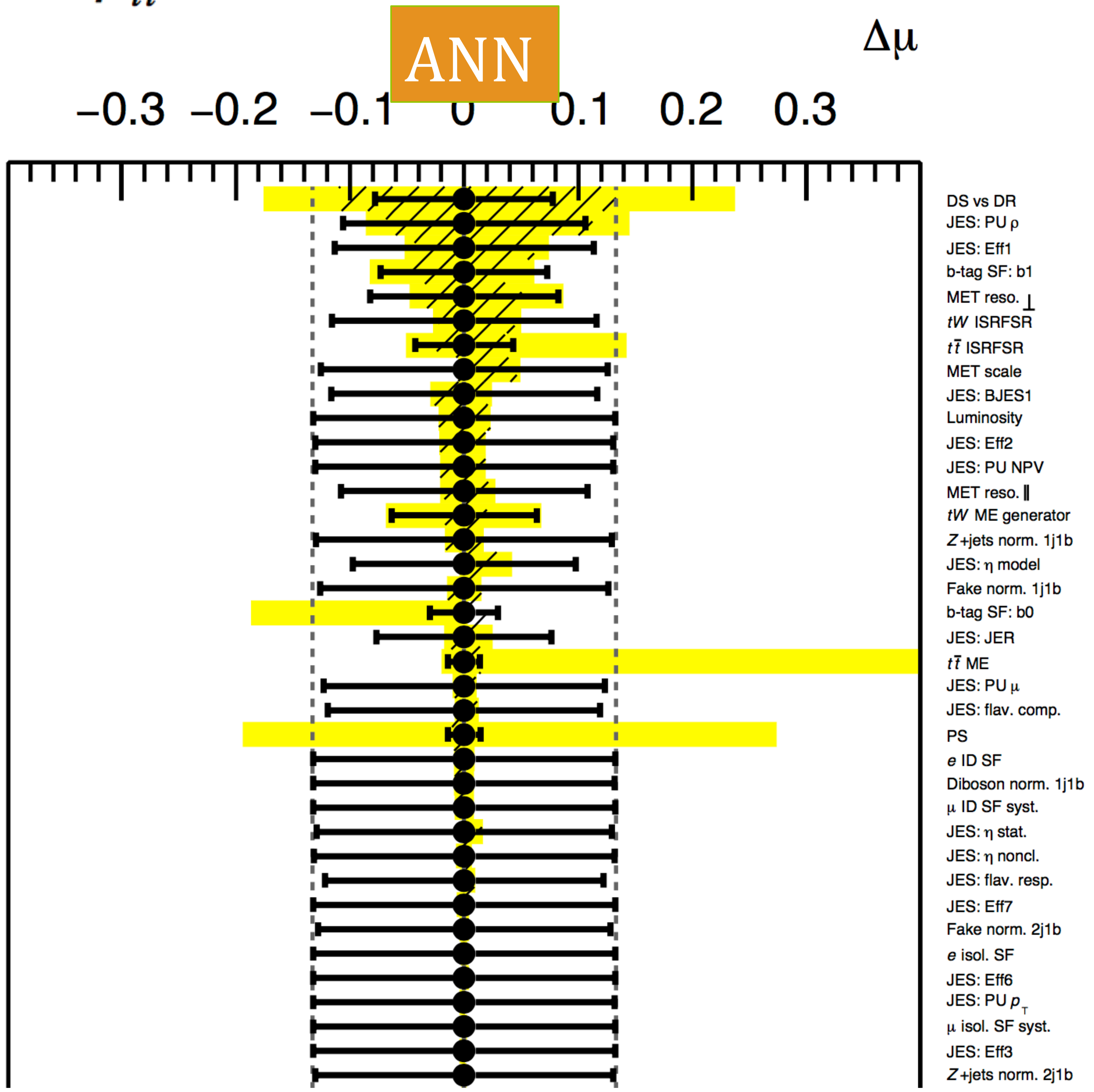
Impact on signal strength

UNIVERSITÄT BONN

- $\Delta\mu_{sig}$: +0.20,−0.166
- $\Delta\mu_{t\bar{t}}$: +0.0318,−0.0308



- $\Delta\mu_{sig}$: +0.19,−0.153
- $\Delta\mu_{t\bar{t}}$: +0.0324,−0.0311



Summary

UNIVERSITÄT BONN

- Single-top-quark production is sensitive to CKM matrix element V_{tb}
  - t-channel, tW-channel, s-channel
- BDT technique used to separate signal from background (mainly $t\bar{t}$)
- Likelihood fit to incorporate systematic uncertainties
- First tW-channel differential cross-sections measurement
- Investigation of adversarial neural network to reduce systematic uncertainties
- Publications:
  - Inclusive measurement (2015 data only) JHEP 01 (2018) 63, arXiv:1612.07231
  - Differential measurements Eur. Phys. J. C 78 (2018) 186, arXiv:1712.01602, ATLAS Physics Briefing, CERN Courier 58 (2018) p11

Thank you!

Backup

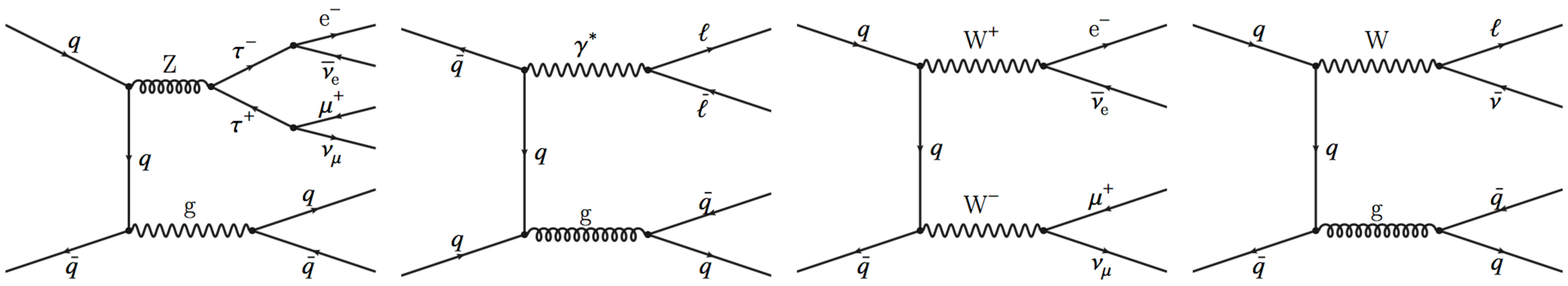
MC samples

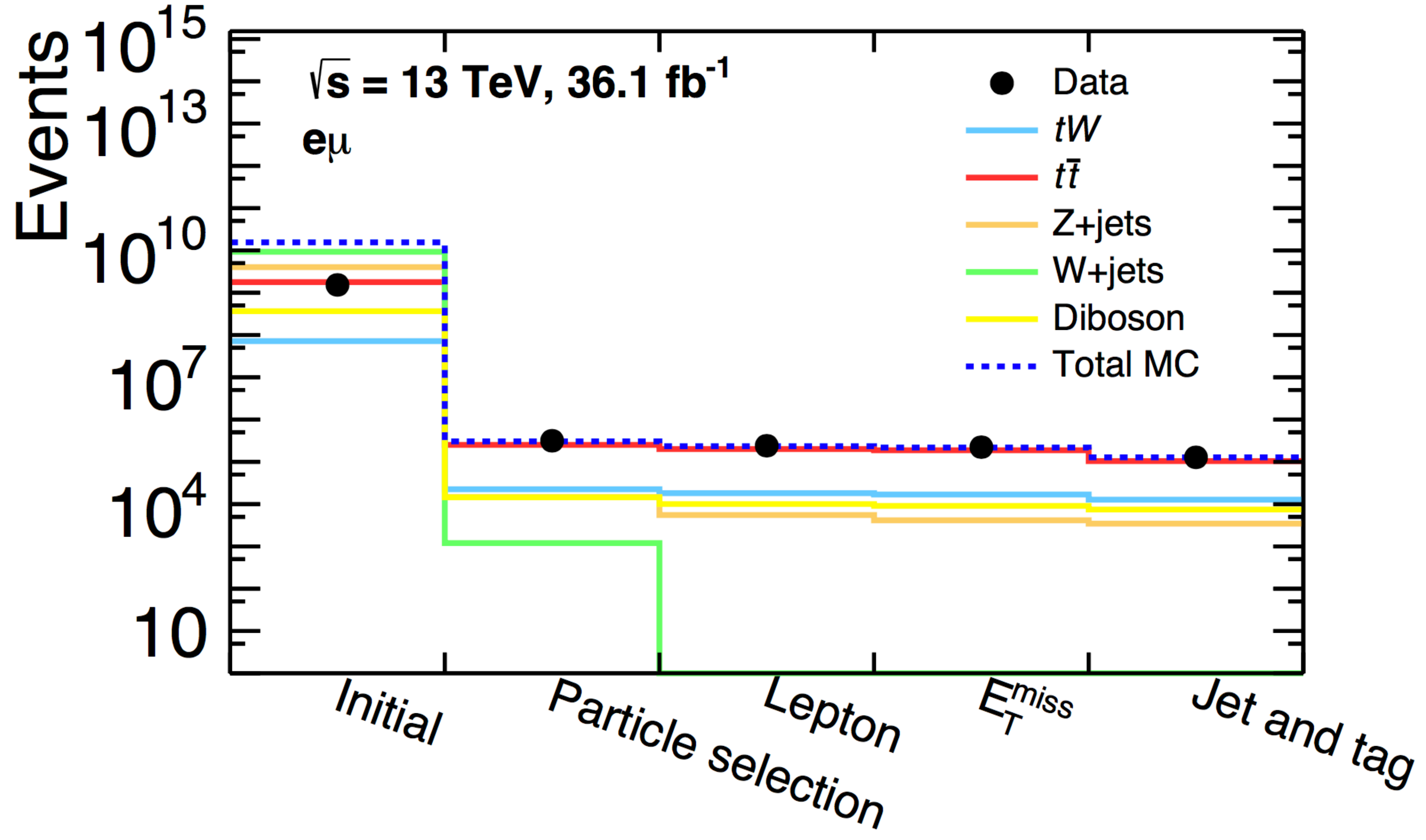
tW sample	ME + PDF		PS + PDF		Tune	PU + PDF + Tune	b/c decay	Detector	Scheme
nominal	POWHEG-BOX 1	CT10	PYTHIA 6.428	CTEQ6L1	P2012	PYTHIA 8.186 + MSTW2008LO + A2	EVTGEN v1.2.0	FULLSIM	DR
DS syst.	POWHEG-BOX 1		PYTHIA 6.428		P2012				DS
AF2 nominal	POWHEG-BOX 1		PYTHIA 6.428		P2012			AF2	DR
ME syst.	aMC@NLO 2.2.2		HERWIG 2.7.1		UE-EE-5				
PS syst.	POWHEG-BOX 1		HERWIG 2.7.1		UE-EE-5				
ISRFSR syst.	POWHEG-BOX 1		PYTHIA 6.428		P2012radHi				
ISRFSR syst.	POWHEG-BOX 1		PYTHIA 6.428		P2012radLo				

Table 4.2: Summary of generators used in variant simulated tW signal events. Sample "nominal" is used to estimate the central value and "syst." are used to access the systematic uncertainties.

$t\bar{t}$ sample	ME + PDF		PS + PDF		Tune	PU + PDF + Tune	b/c decay	Detector
nominal	POWHEG-BOX 2	CT10	PYTHIA 6.428	CTEQ6L1	P2012	PYTHIA 8.186 + MSTW2008LO + A2	EVTGEN v1.2.0	FULLSIM
AF2 nominal	POWHEG-BOX 2		PYTHIA 6.428		P2012			AF2
ME syst.	aMC@NLO 2.2.2		HERWIG 2.7.1		UE-EE-5			
PS syst.	POWHEG-BOX 2		HERWIG 2.7.1		UE-EE-5			
ISRFSR syst.	POWHEG-BOX 2		PYTHIA 6		P2012radHi			FULLSIM
ISRFSR syst.	POWHEG-BOX 2		PYTHIA 6		P2012radLo			

Backgrounds and cutflow





BDT optimisation

Variables selection

The separation power S of a variable x is defined by

$$S = \frac{1}{2} \int \frac{(Y_S(x) - Y_B(x))^2}{(Y_S(x) + Y_B(x))} dx,$$

The selection of the variables for the BDT training is done in the following steps:

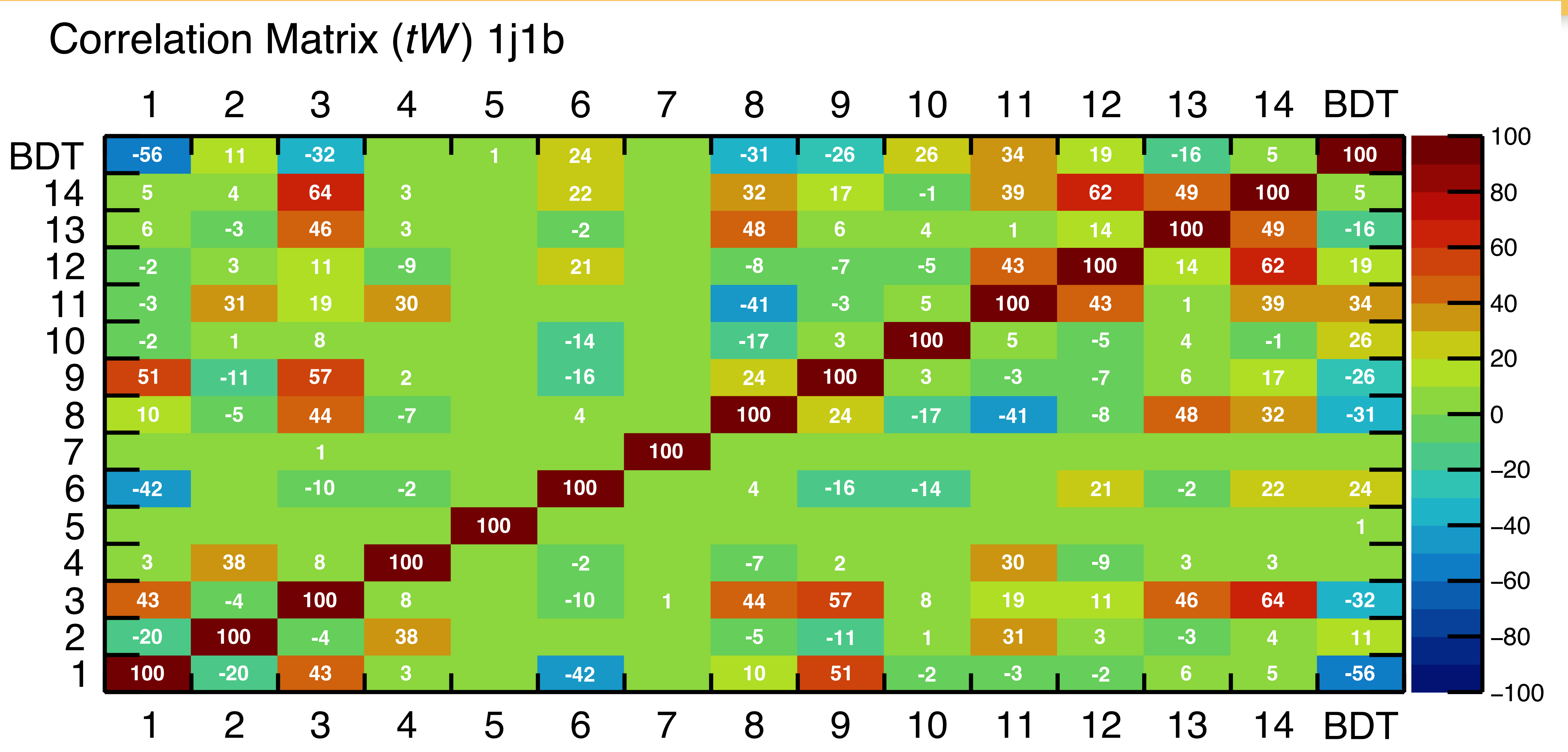
1. Rank the variables by descending separation power S.
2. Keep only one variable with the highest S if several variables are highly correlated (linear correlation above 70 %). An exception is accepted when the variable shows large correlation difference (> 10 %) between signal and background. After this step, the number of variables are largely reduced to 40 to 70.
3. Starting with the top one variable, add variables one by one and repeat training. If the S of the newly trained BDT response increases more than 1 % (2 %) of that of the previous response or it exceeds more than 5 % (10 %) of the maximum S ever emerged in `1j1b` (`2j`) region, the new variable is kept.
4. If it ends up with different collections giving similar S (difference smaller than 1 %), the collection with fewer variables is used.

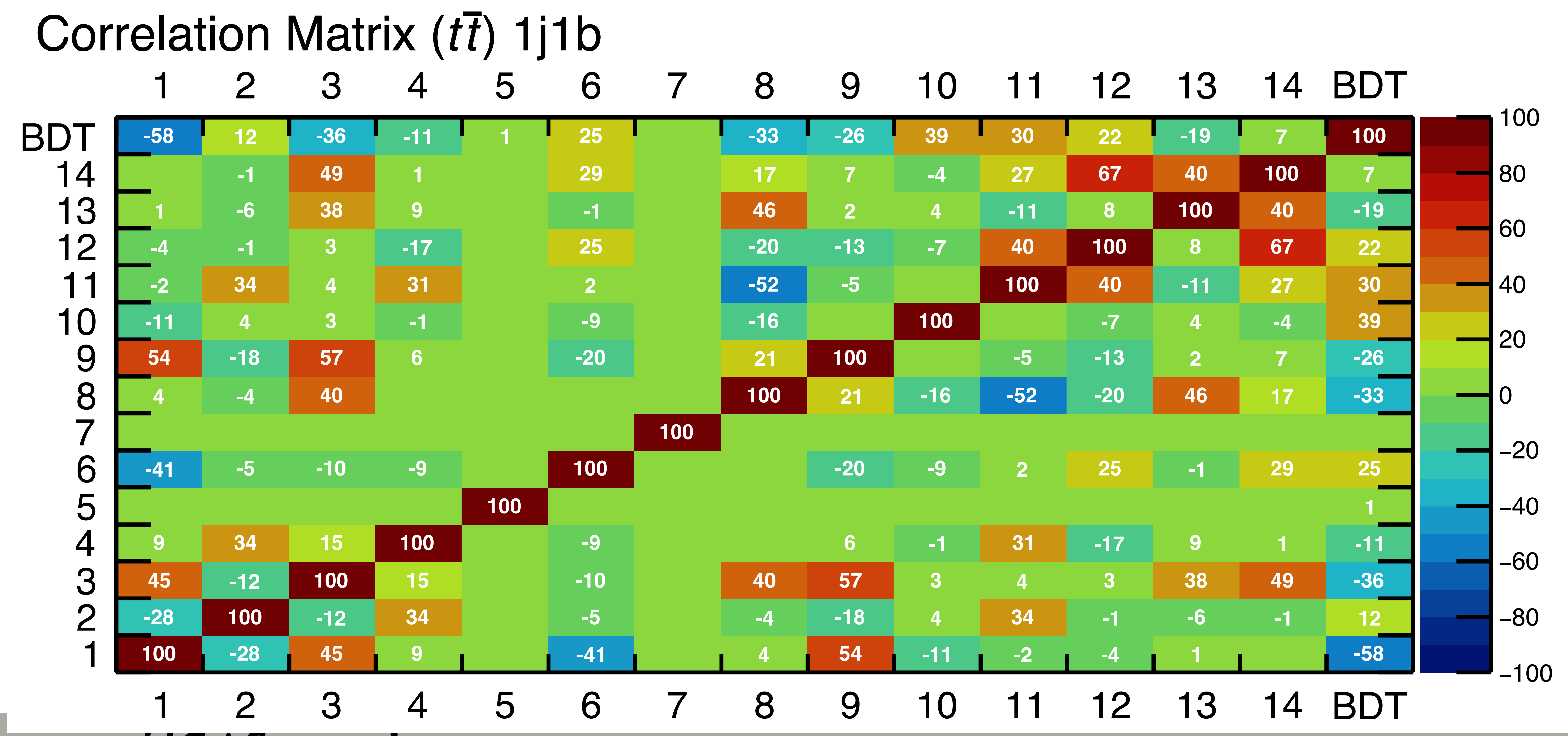
Hyperparameters optimised one by one

Region	NTrees	MinNodeSize [%]	Shrinkage	BaggedSample Fraction	nCuts	MaxDepth	No. of variables
Range	20–400	0.5–10	0.01–0.2	0.05–1	5–100	{2,3,4}	
Step	20	0.5	0.01	0.05	5	—	
`1j1b`	300	0.5	0.03	0.40	20	4	14
`2j1b`	280	2.0	0.14	0.15	5	3	11
`2j2b`	100	0.5	0.19	0.60	5	4	11

1j1b BDT variables and correlations

R	1j1b Variable	S $[10^{-2}]$
1	$p_{\mathrm{T}}(\ell_1\ell_2 j_1 E_{\mathrm{T}}^{\mathrm{miss}})$	4.8
2	$\Delta p_{\mathrm{T}}(\ell_1\ell_2, j_1 E_{\mathrm{T}}^{\mathrm{miss}})$	2.8
3	$\sum E_{\mathrm{T}}$	2.8
4	$\Delta p_{\mathrm{T}}(\ell_1\ell_2 j_1, E_{\mathrm{T}}^{\mathrm{miss}})$	2.7
5	$\Delta\phi(\ell_1\ell_2 j_1, E_{\mathrm{T}}^{\mathrm{miss}})$	1.7
6	$\Delta R(\ell_1\ell_2, j_1 E_{\mathrm{T}}^{\mathrm{miss}})$	1.6
7	$\eta(\ell_1\ell_2 j_1 E_{\mathrm{T}}^{\mathrm{miss}})$	1.4
8	$m(\ell_2 j_1 E_{\mathrm{T}}^{\mathrm{miss}})$	1.2
9	$p_{\mathrm{T}}(\ell_1 j_1 E_{\mathrm{T}}^{\mathrm{miss}})$	1.2
10	$C(\ell_1\ell_2)$	1.1
11	$\Delta p_{\mathrm{T}}(\ell_1\ell_2, E_{\mathrm{T}}^{\mathrm{miss}})$	1.0
12	$\Delta R(\ell_1, j_1)$	0.6
13	$m(j_1)$	0.5
14	$m(\ell_1 j_1)$	0.1
	BDT response	8.6





2j1b BDT variables and correlations

R	2j1b Variable	S $[10^{-2}]$
1	$\Delta p_T(\ell_1\ell_2, j_2)$	1.9
2	$\Delta R(\ell_1\ell_2, j_1j_2E_T^{miss})$	1.7
3	$m(\ell_1 j_2)$	1.4
4	$\Delta R(\ell_1\ell_2, j_1j_2)$	1.3
5	$p_T(j_2)$	1.0
6	$p_T(j_1j_2)$	0.9
7	$m(\ell_1 j_1 j_2)$	0.9
8	$\sigma(p_T)(\ell_1\ell_2 j_1 j_2 E_T^{miss})$	0.9
9	$p_T(\ell_2 j_1 j_2 E_T^{miss})$	0.8
10	$\sum E_T(j_2 E_T^{miss})$	0.6
11	$\Delta R(\ell_1, j_2)$	0.5
	BDT response	7.3

Correlation Matrix (tW) 2j1b

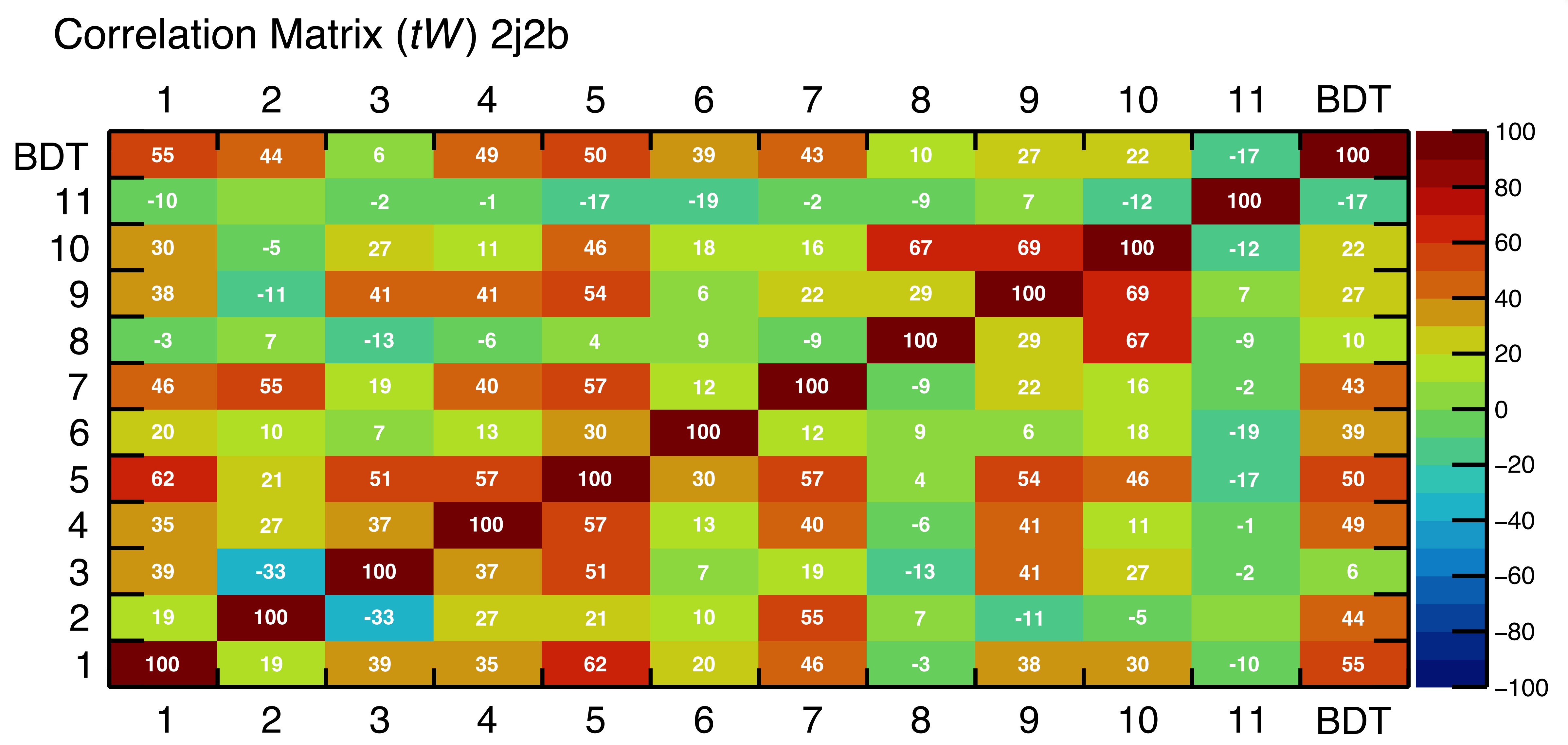
	1	2	3	4	5	6	7	8	9	10	11	BDT
BDT	51	38	43	36	-2	35	40	-35	42	4	28	100
11	8	16	62	22	-9	2	14	1		-10	100	28
10	-11	5	25	-10	65	58	47	1	25	100	-10	4
9	56		42	11	16	39	50	6	100	25		42
8	2	-36	-1	-5	-4	-4	-3	100	6	1	1	-35
7	22	29	63	37	55	61	100	-3	50	47	14	40
6	28	4	29	9	39	100	61	-4	39	58	2	35
5	-31	8	35	10	100	39	55	-4	16	65	-9	-2
4	6	59	30	100	10	9	37	-5	11	-10	22	36
3	19	23	100	30	35	29	63	-1	42	25	62	43
2	-9	100	23	59	8	4	29	-36		5	16	38
1	100	-9	19	6	-31	28	22	2	56	-11	8	51

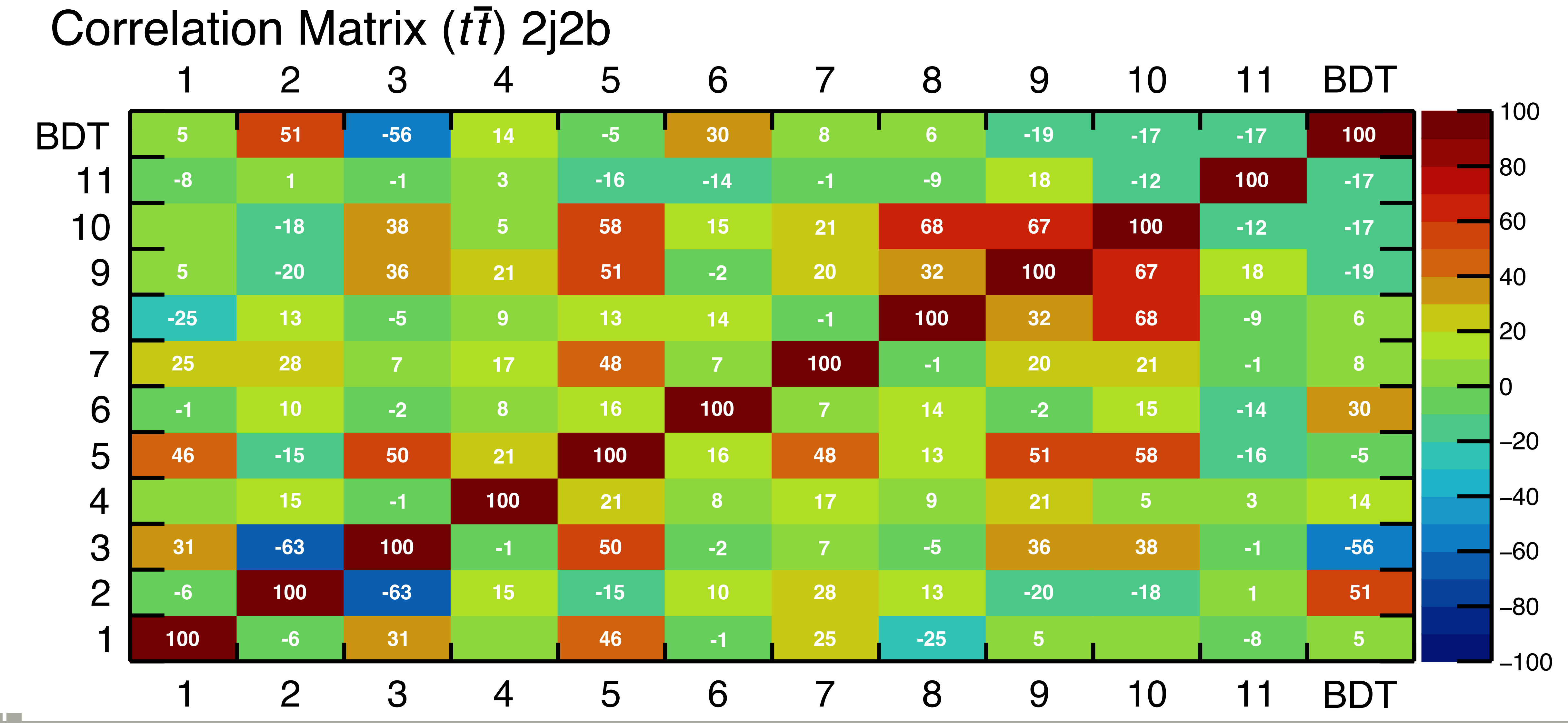
Correlation Matrix ($t\bar{t}$) 2j1b

	1	2	3	4	5	6	7	8	9	10	11	BDT
BDT	48	42	25	33	-33	19	15	-28	21	-16	27	100
11	10	13	66	13	-14	2	8	4	-6	-9	100	27
10	-30		20	-22	60	41	38		19	100	-9	-16
9	32	3	28	9	7	24	38	13	100	19	-6	21
8	7	-39	1	-1	-7		-3	100	13		4	-28
7	-3	27	54	31	52	43	100	-3	38	38	8	15
6	14	4	13	9	15	100	43		24	41	2	19
5	-54	2	26	1	100	15	52	-7	7	60	-14	-33
4	8	50	20	100	1	9	31	-1	9	-22	13	33
3	3	21	100	20	26	13	54	1	28	20	66	25
2		100	21	50	2	4	27	-39	3		13	42
1	100		3	8	-54	14	-3	7	32	-30	10	48

2j2b BDT variables and correlations

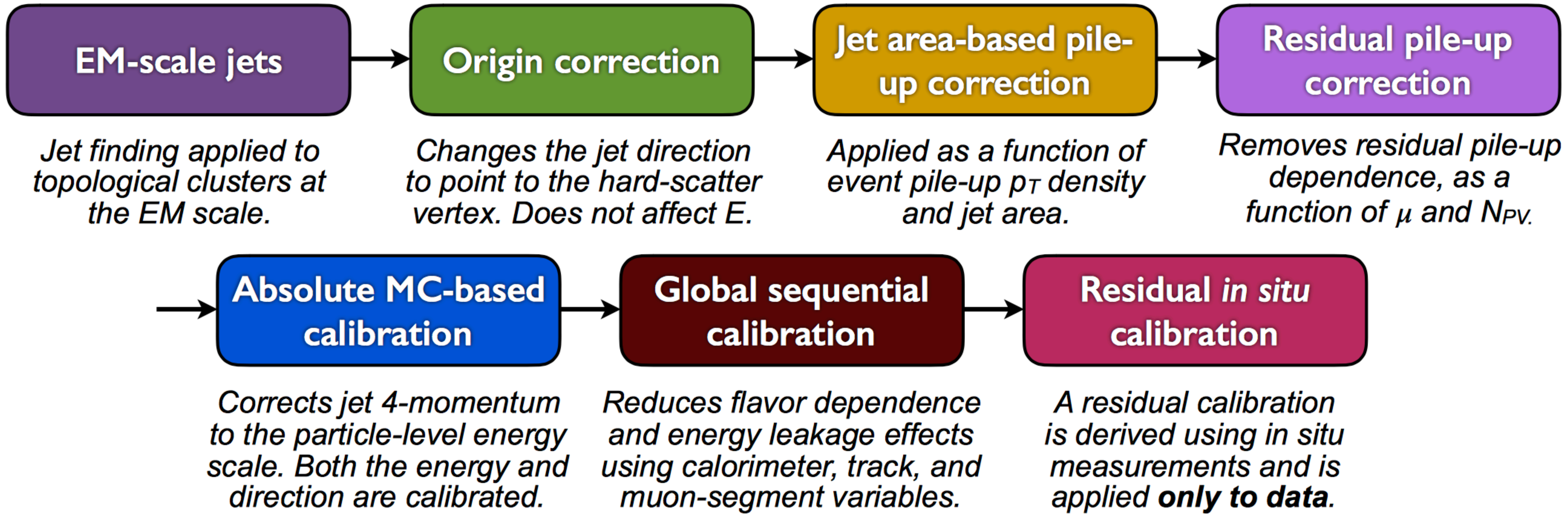
R	2j2b Variable	S [10^{-2}]
1	$m(\ell_1 j_2)$	4.0
2	$\Delta p_T(\ell_1\ell_2, j_2)$	3.4
3	$p_T(j_2)$	2.9
4	$p_T(j_1 j_2)$	2.9
5	$m(\ell_1\ell_2 j_1 j_2)$	2.4
6	$\Delta R(\ell_1\ell_2, j_1 j_2)$	2.2
7	$p_T(\ell_2 j_1 j_2 E_T^{miss})$	1.2
8	$\Delta R(\ell_2, j_2)$	1.1
9	$m(\ell_2 j_2 E_T^{miss})$	0.8
10	$m(\ell_2 j_2)$	0.7
11	$E/m(\ell_1\ell_2 j_1 j_2)$	0.6
	BDT response	16.2





JES calibration

- What do we have to take into account?
  - Non-compensated calorimeter: different E scales for HAD/EM showers.
  - Dead material: energy lost in the inactive areas.
  - Leakage: showers reaching the outer of the calorimeter (punch-through).
  - Energy deposited below noise thresholds: none cluster created.
- Good calibration scheme should provide jets with:
  - Good linearity in the response
  - Good energy resolution
  - Good position reconstruction



JES calibration

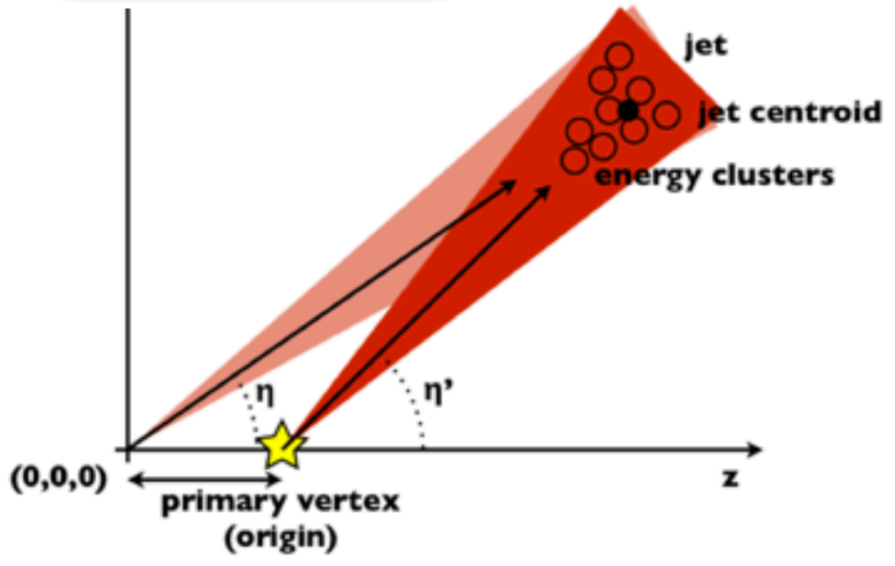
EM-scale jets → **Origin correction** → **Jet area-based pile-up correction** → **Residual pile-up correction**

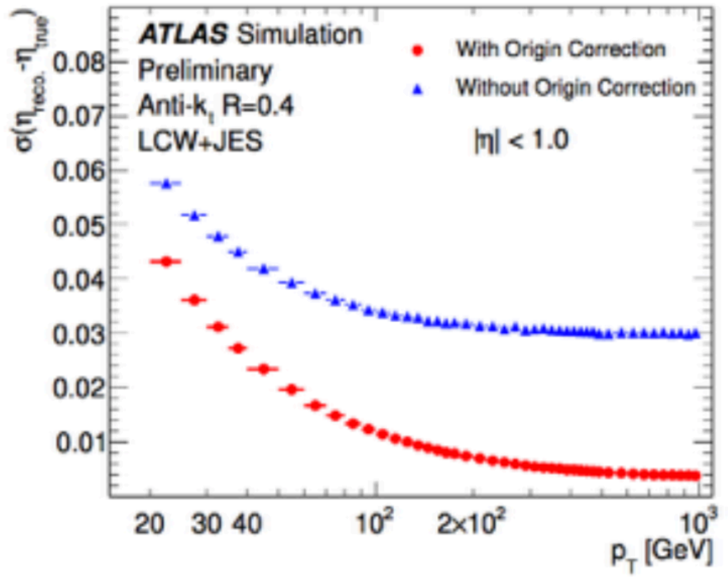
- **EM-scale jets**: *Jet finding applied to topological clusters at the EM scale.*
- **Origin correction**: *Changes the jet direction to point to the hard-scatter vertex. Does not affect E.*
- **Jet area-based pile-up correction**: *Applied as a function of event pile-up p_T density and jet area.*
- **Residual pile-up correction**: *Removes residual pile-up dependence, as a function of μ and N_{PV}.*

→ **Absolute MC-based calibration** → **Global sequential calibration** → ***Residual in situ calibration***

- **Absolute MC-based calibration**: *Corrects jet 4-momentum to the particle-level energy scale. Both the energy and direction are calibrated.*
- **Global sequential calibration**: *Reduces flavor dependence and energy leakage effects using calorimeter, track, and muon-segment variables.*
- **Residual in situ calibration**: *A residual calibration is derived using in situ measurements and is applied **only to data**.*

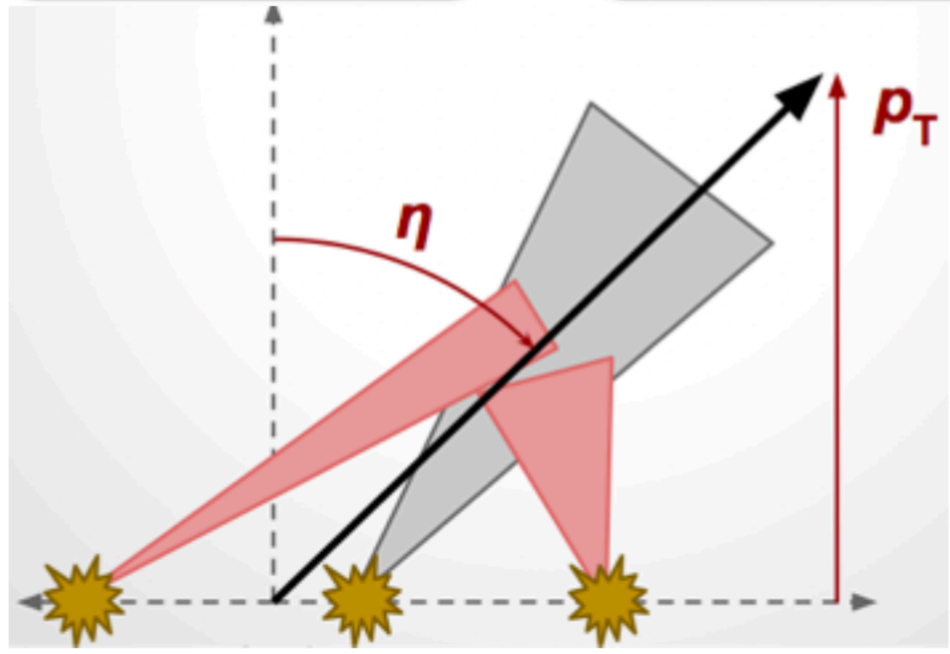
Origin Correction Jet corrected to point back to the primary vertex.





The origin correction improves the η angular resolution due to the length of the BS

Jet area based pile-up correction → **Residual pile-up correction** Decreases pile-up contamination.



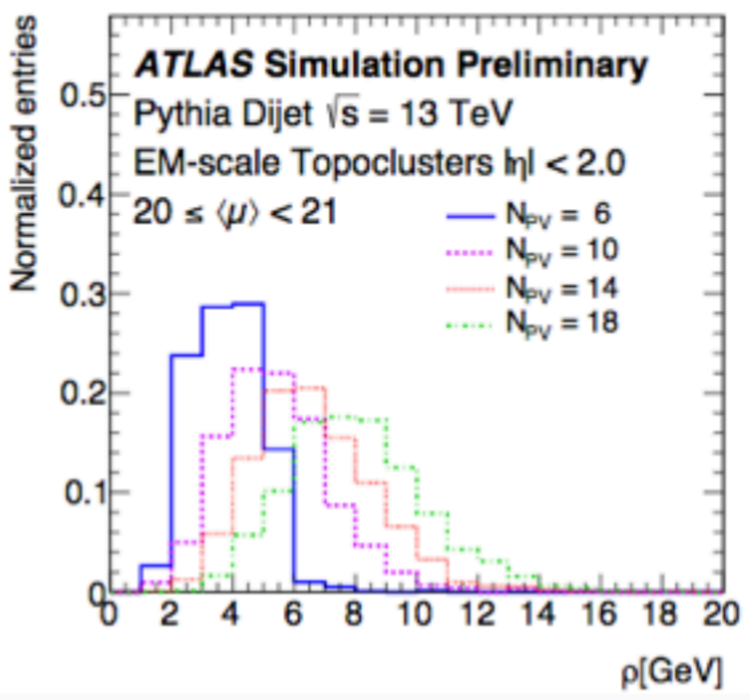
Jets at constituent level receive affected by:

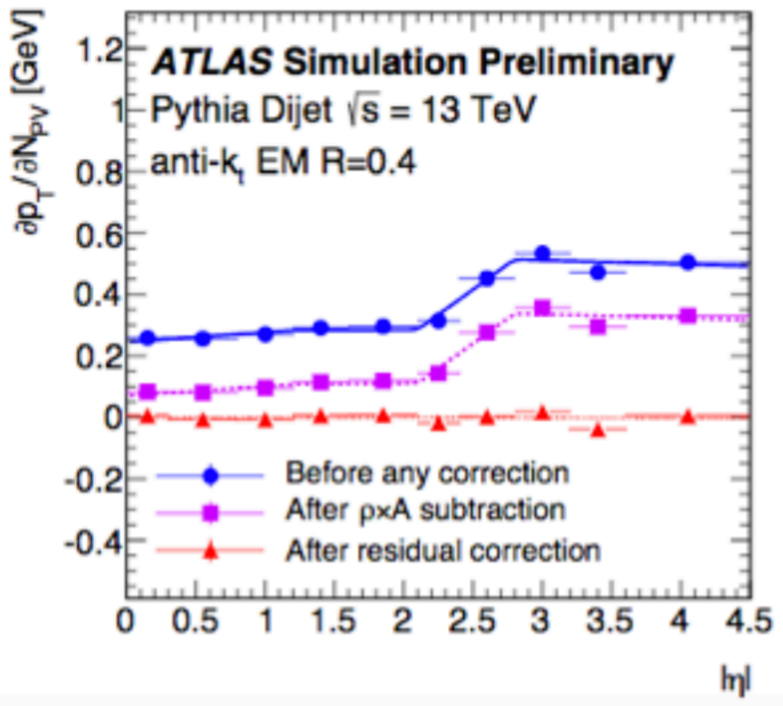
- *In-time pile-up*: additional interactions in the same bunch crossing (N_{PV}).
- *Out-of-time pile-up*: multiple interactions in surrounding bunch crossing (<μ>).

Jet area corr. — Residual correction

$$p_T^{corr} = p_T^{const} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times \langle\mu\rangle$$

- Area based subtraction:
  - ρ: pile-up energy density
  - **A**: area of the jet





- Residual pile-up correction

EM-scale jets → **Origin correction** → **Jet area-based pile-up correction** → **Residual pile-up correction**

- **EM-scale jets**: *Jet finding applied to topological clusters at the EM scale.*
- **Origin correction**: *Changes the jet direction to point to the hard-scatter vertex. Does not affect E.*
- **Jet area-based pile-up correction**: *Applied as a function of event pile-up p_T density and jet area.*
- **Residual pile-up correction**: *Removes residual pile-up dependence, as a function of μ and N_{PV}.*

→ **Absolute MC-based calibration** → **Global sequential calibration** → **Residual *in situ* calibration**

- **Absolute MC-based calibration**: *Corrects jet 4-momentum to the particle-level energy scale. Both the energy and direction are calibrated.*
- **Global sequential calibration**: *Reduces flavor dependence and energy leakage effects using calorimeter, track, and muon-segment variables.*
- **Residual *in situ* calibration**: *A residual calibration is derived using in situ measurements and is applied **only to data**.*

UNIVERSITÄT BONN

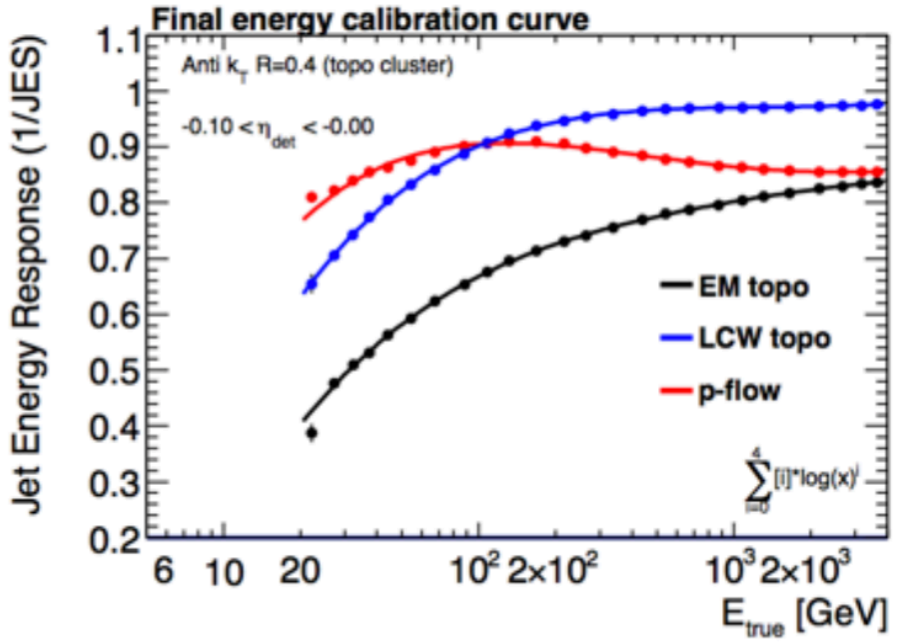
Absolute EtaJES Calibrate the jet energy and the η to the ***particle level scale***.

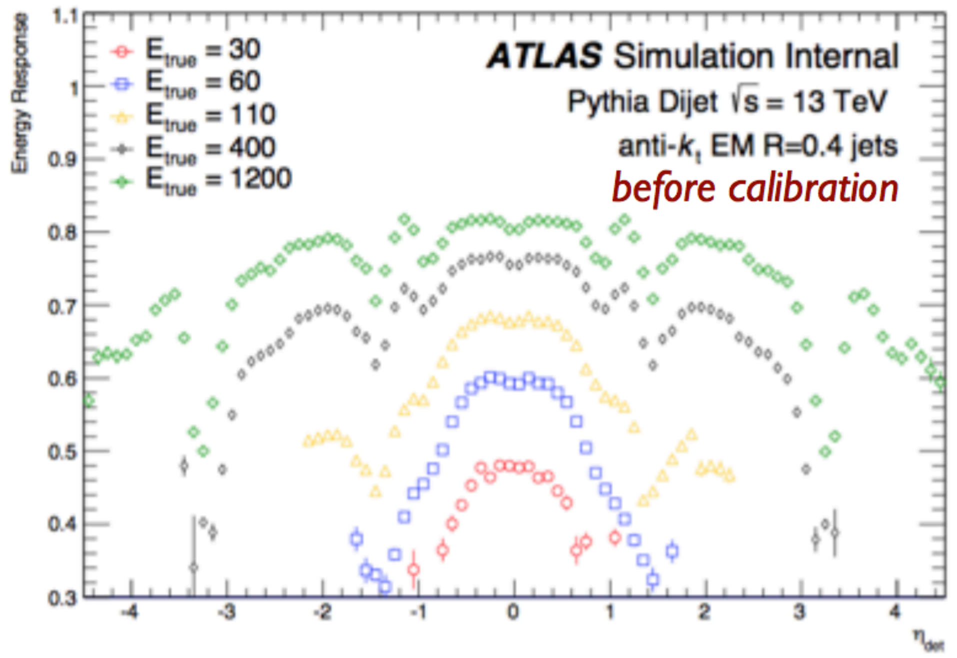
Functions are parametrised in η:

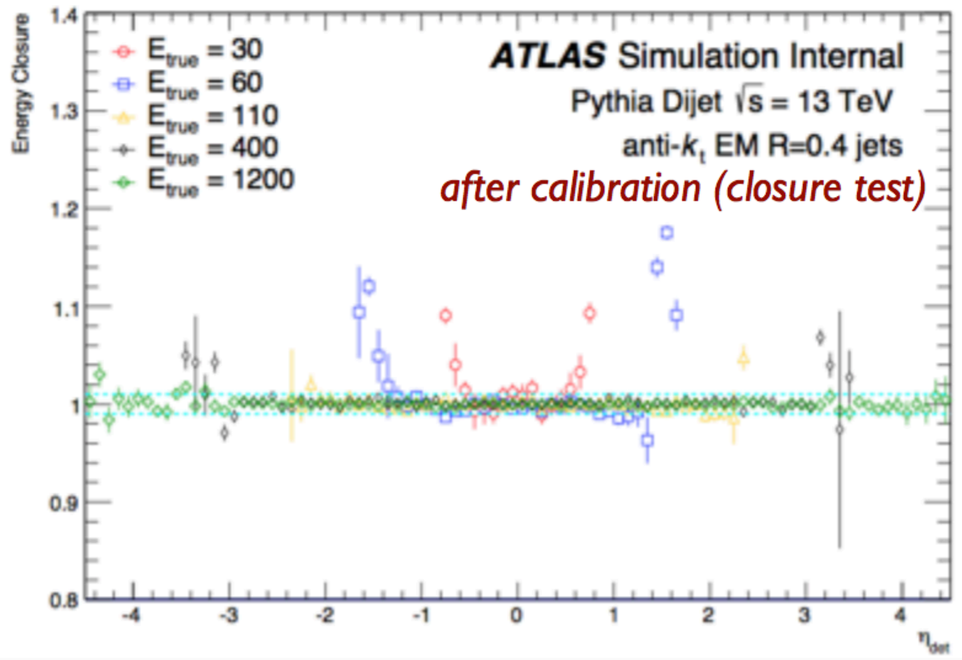
- *Energy response*: $R = \langle E^{reco}/E^{true} \rangle$ vs. E^{true}
- *η-response*: $\Delta\eta = \eta^{reco} - \eta^{true}$ vs. E^{true}

The energy of the jet is then corrected by multiplying by the inverse of the R:

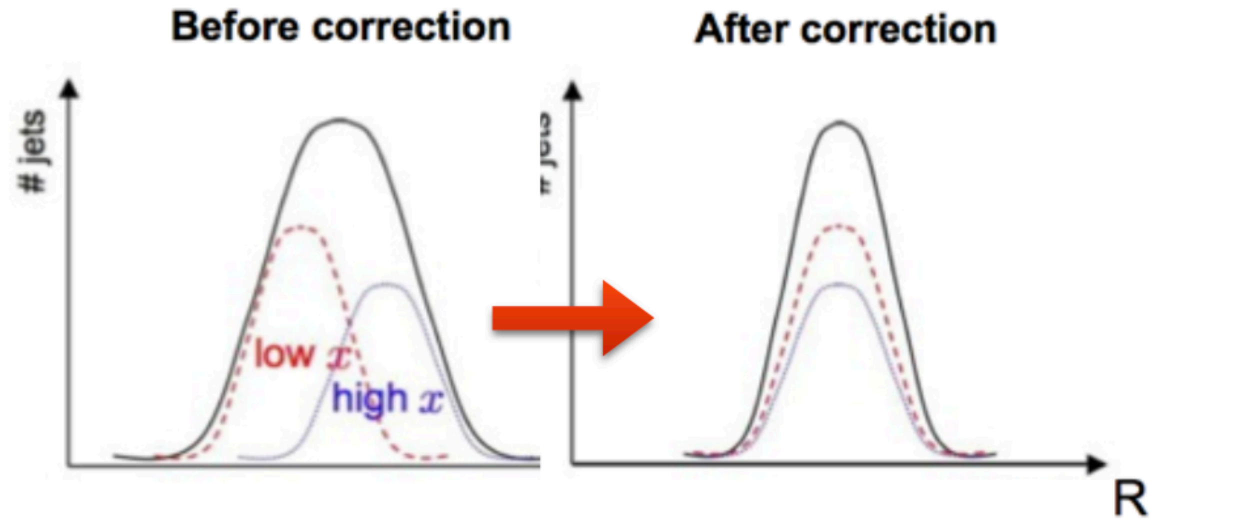
$$E^{corr} = R^{-1} \times E^{reco}$$





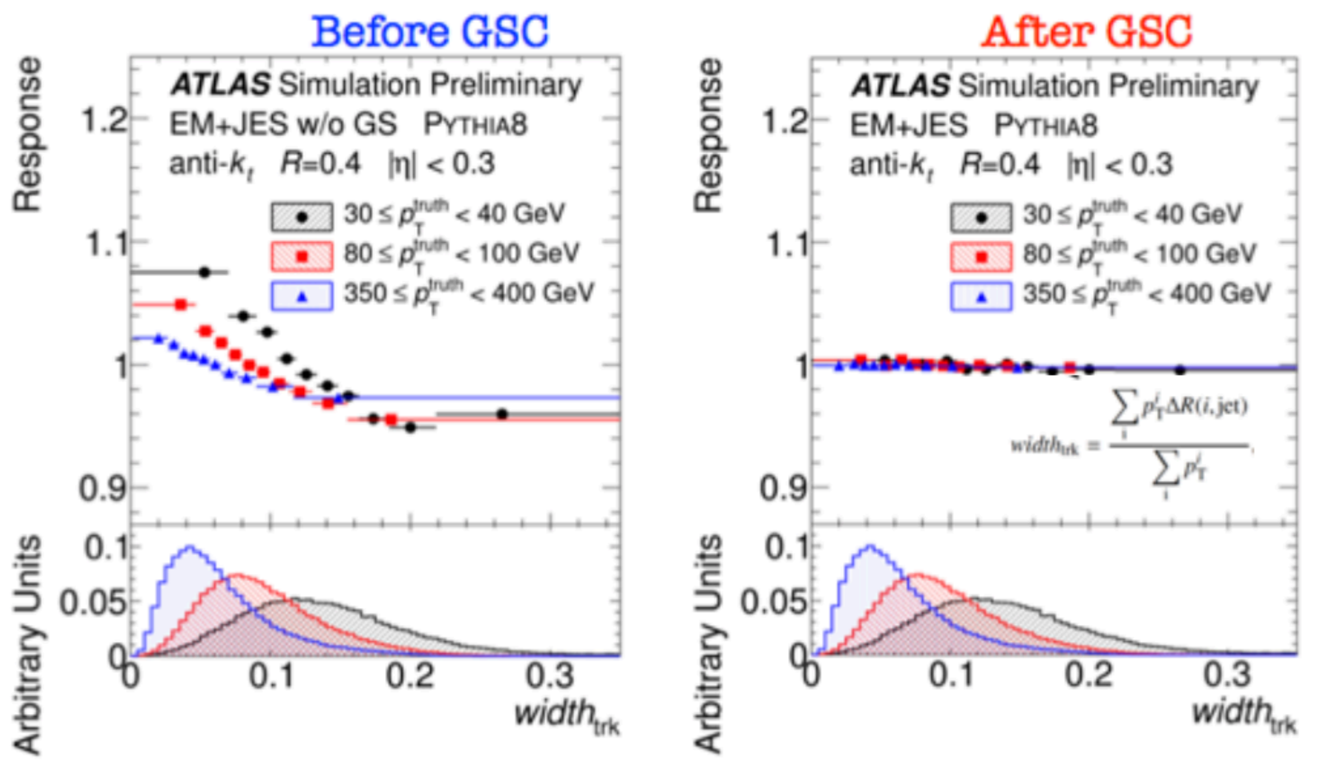


Global sequential calibration Reduce the differences in response between *quark-gluon jets* and also correct for the *jets not fully contained in the calorimeter*.



GSC variable (EMTopo jets):

- Longitudinal structure of the **energy depositions** within the calorimeters (f_{Tile} and f_{LAr3}) ➜ Improve the JER
- **Track information** associated to the jet (n_{trk} and $width_{trk}$) ➜ Correct for flavour differences in response between quarks and gluons jets.
- **Activity in the MS** behind the jets ($N_{\mu\_segments}$) ➜ Correct high p_T jets.



EM-scale jets → **Origin correction** → **Jet area-based pile-up correction** → **Residual pile-up correction**

- **EM-scale jets**: *Jet finding applied to topological clusters at the EM scale.*
- **Origin correction**: *Changes the jet direction to point to the hard-scatter vertex. Does not affect E.*
- **Jet area-based pile-up correction**: *Applied as a function of event pile-up p_T density and jet area.*
- **Residual pile-up correction**: *Removes residual pile-up dependence, as a function of μ and N_{PV}.*

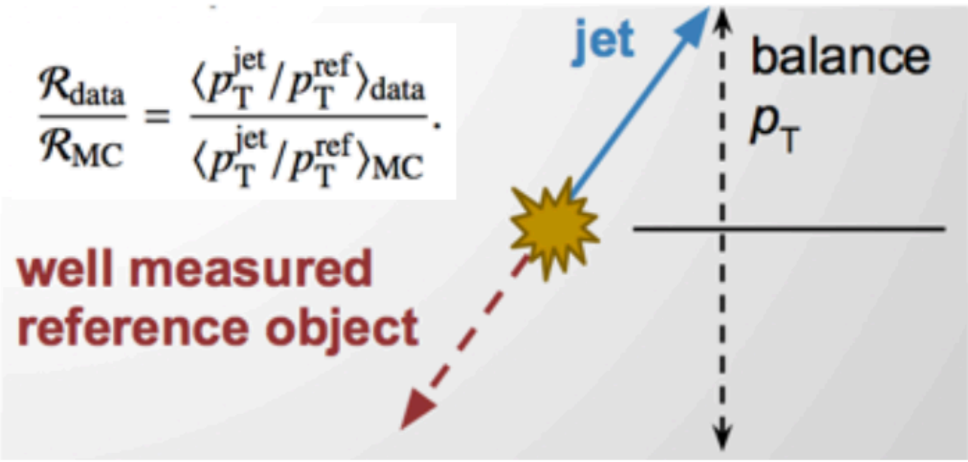
→ **Absolute MC-based calibration** → **Global sequential calibration** → **Residual *in situ* calibration**

- **Absolute MC-based calibration**: *Corrects jet 4-momentum to the particle-level energy scale. Both the energy and direction are calibrated.*
- **Global sequential calibration**: *Reduces flavor dependence and energy leakage effects using calorimeter, track, and muon-segment variables.*
- **Residual *in situ* calibration**: *A residual calibration is derived using in situ measurements and is applied **only to data**.*

Residual in-situ calibration

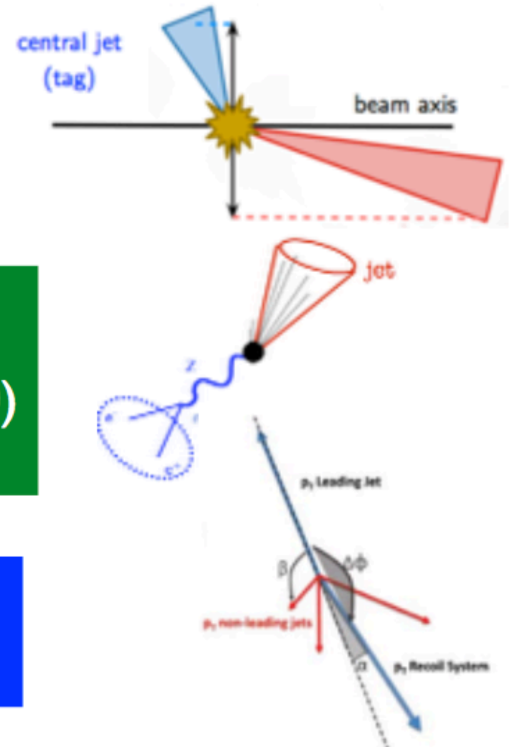
In-situ techniques employs the balance of physics objects in the transverse plane to correct residual Data/MC differences.

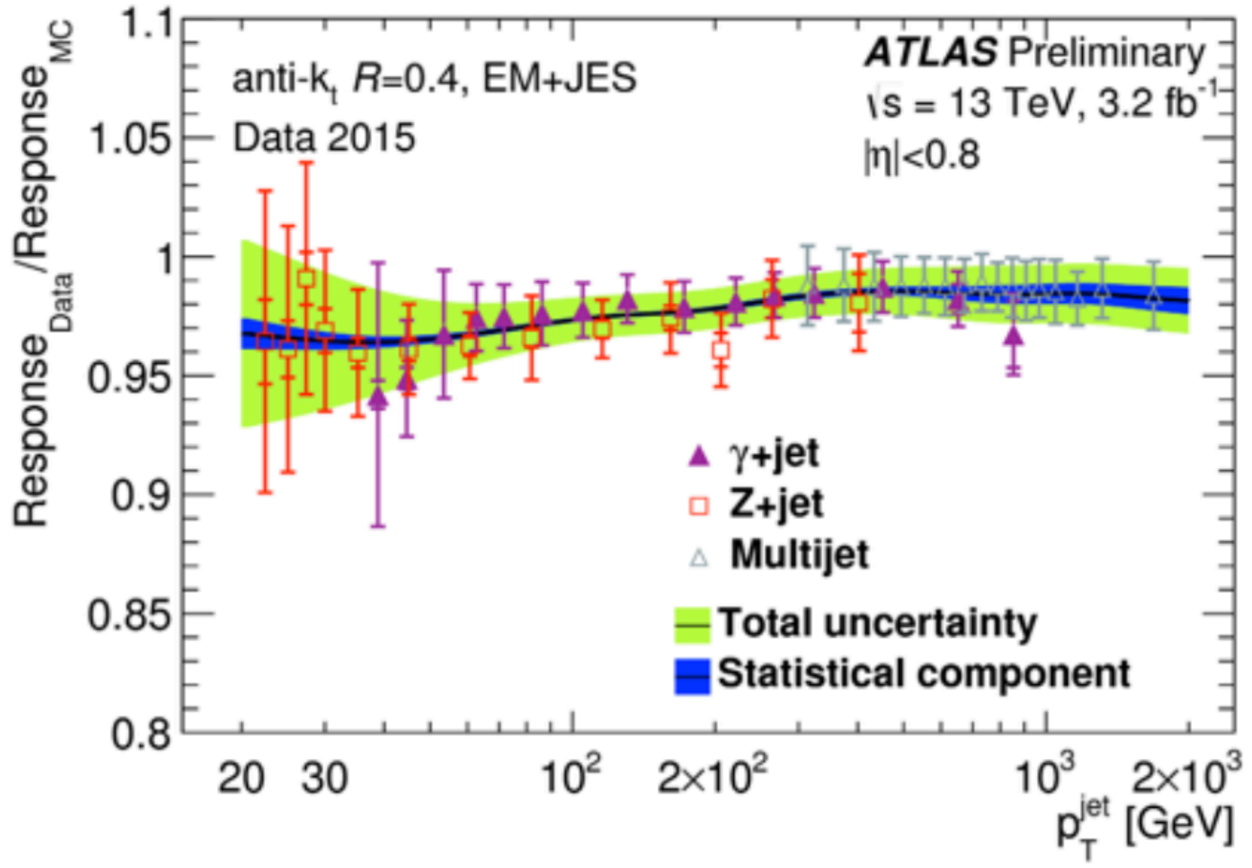
$$\frac{\mathcal{R}_{data}}{\mathcal{R}_{MC}} = \frac{\langle p_T^{jet}/p_T^{ref}\rangle_{data}}{\langle p_T^{jet}/p_T^{ref}\rangle_{MC}}.$$



Multiple references have been used

- *η-intercalibration (froward region)*
- *Z+jets / γ+jets* (20<p_T<200/30<p_T<800) [GeV]
- *Multijet balance* (300<p_T<1700) [GeV]



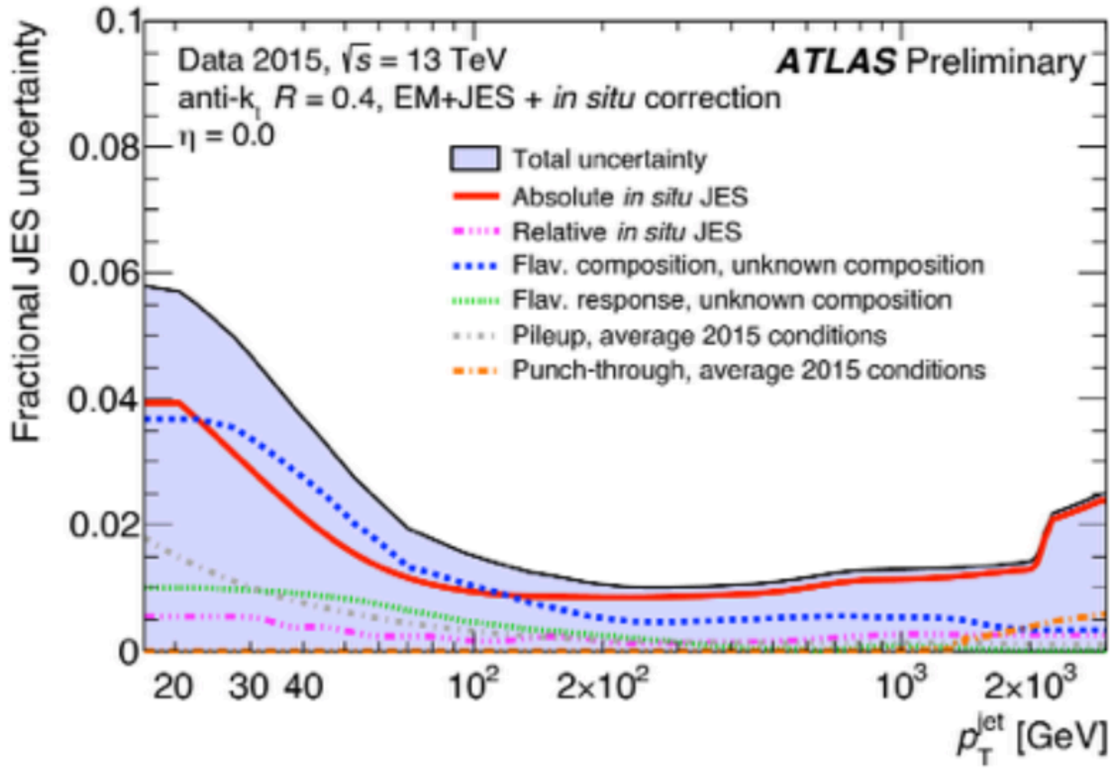


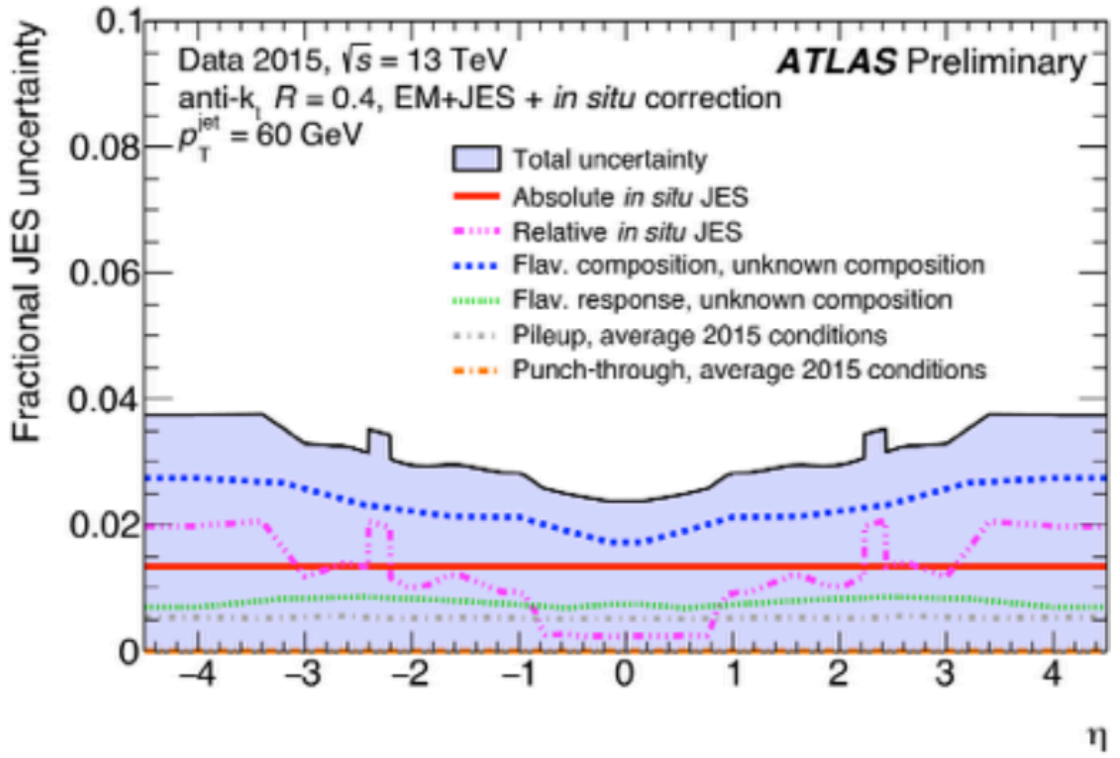
Jets are ready for physics!! …but the JES systematic uncertainty arrives!! 😖

JES uncertainty is one of the **dominant systematics** in many analyses.

Final JES 2015 (ICHEP):

- *In-situ analyses* (65NP)
- *η-intercalibration* (3NP)
- *Jet flavour:*
  - *Composition* (1NP)
  - *Response* (1NP)
  - *b-jets* (1NP)
- *Single hadron uncert.* (1NP)
- *Pile-up* (4NP)
- *Punch-through* (1NP)





Fit result

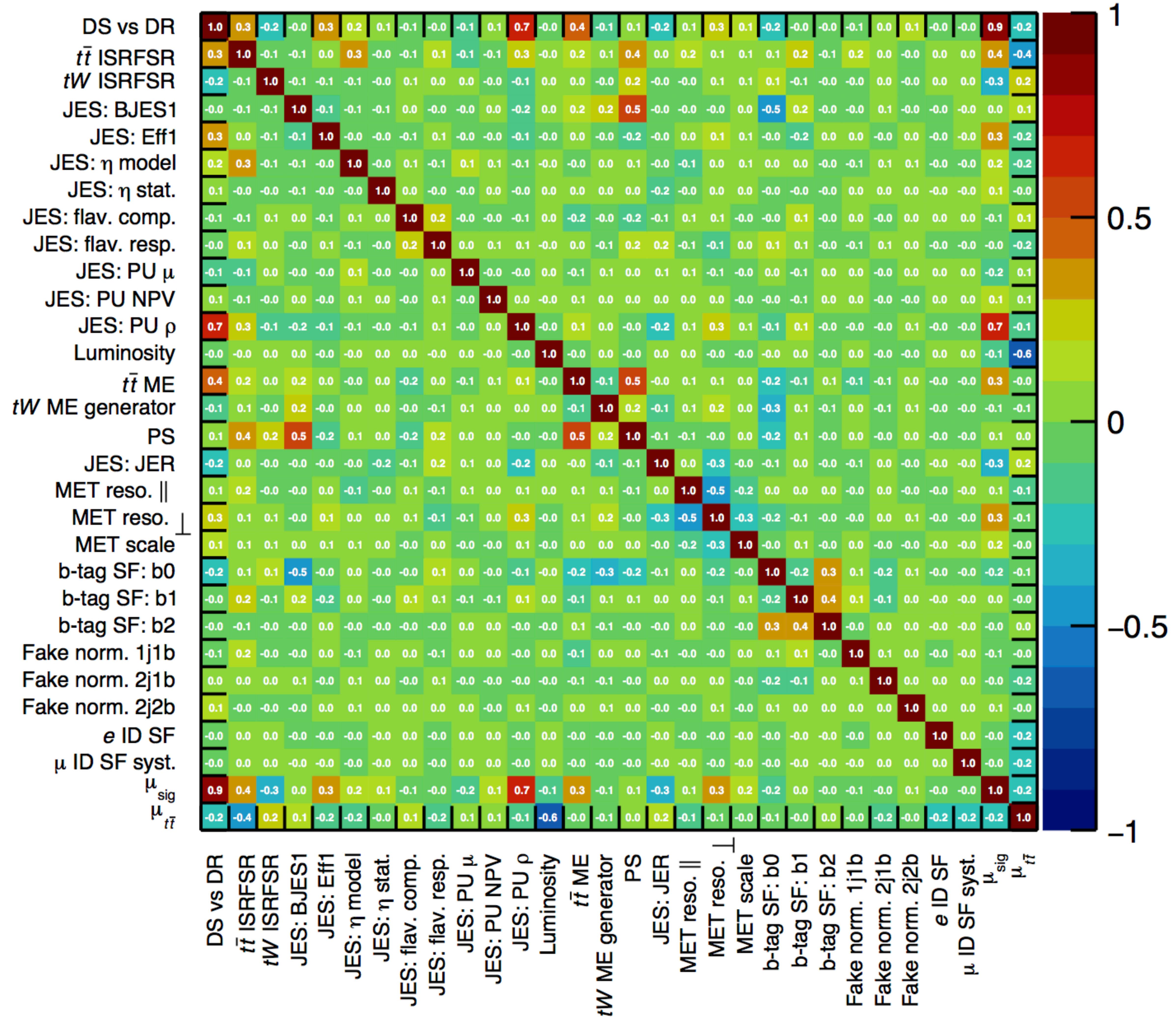
	Expected	Observed
μ_{sig}	$1.00^{+0.20}_{-0.17}$	$1.14^{+0.26}_{-0.22}$
$\mu_{t\bar{t}}$	$1.000^{+0.032}_{-0.031}$	$1.004^{+0.033}_{-0.032}$

Source	$\Delta\mu_{sig}$[%]
ME	4.2
PS	3.6
ISRFSR	11
DR vs DS	19
PDF	$\lesssim 0$
Non-$t\bar{t}$ background normalisation	$\lesssim 0$
Lepton efficiency, energy scale and resolution	4.1
JES	21
JER	8.9
b-tagging	8.2
E_T^{miss} calculation	9.3
Luminosity	4.2
Total systematic uncertainty	23
Data statistics	3.4
Total uncertainty	23

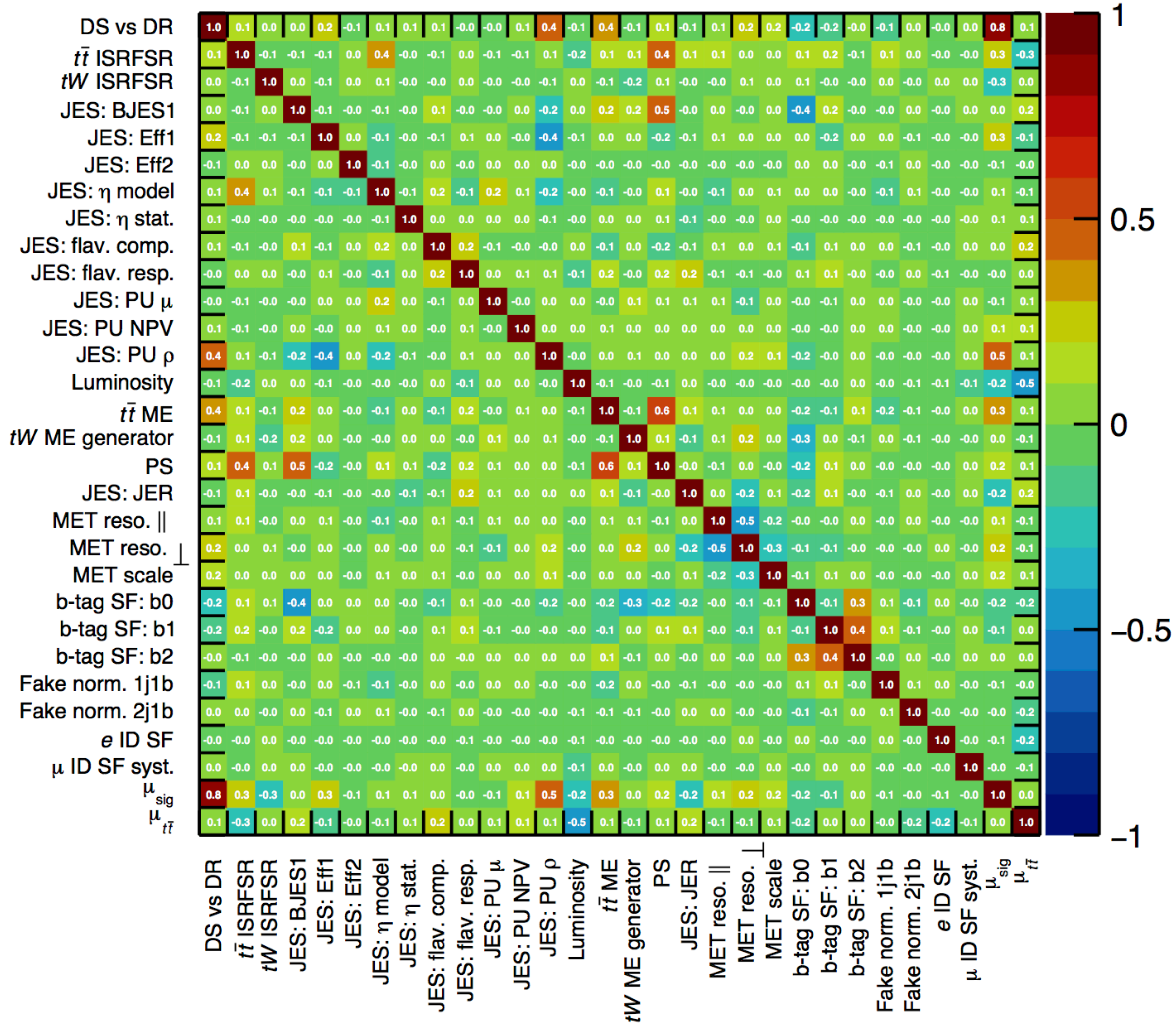
	1j1b	2j1b	2j2b
Pred. events	27100 ± 4400	37800 ± 5100	31900 ± 7700
Pred. tW events	4910 ± 800	3370 ± 640	1080 ± 290
Pred. $t\bar{t}$ events	21700 ± 3700	34000 ± 4900	30700 ± 7500
Pred. Z + jets events	152 ± 76	116 ± 58	10 ± 10
Pred. diboson events	169 ± 42	168 ± 42	9.5 ± 2.4
Pred. fake events	120 ± 120	180 ± 180	92 ± 92
Observed events	26171	37147	29874
Fitted events	26150 ± 200	37160 ± 230	29890 ± 200
Fitted tW events	5300 ± 1000	3900 ± 680	1050 ± 140
Fitted $t\bar{t}$ events	20300 ± 1000	32820 ± 760	28740 ± 270
Fitted Z + jets events	154 ± 75	99 ± 53	10 ± 10
Fitted diboson events	168 ± 42	162 ± 41	10.0 ± 2.4
Fitted fake events	130 ± 120	140 ± 140	79 ± 79

These are estimated by fixing systematic sources to their best fit values in each category, re-fitting, and subtracting refitted uncertainty in quadrature from the total uncertainty.

Correlation Matrix Asimov

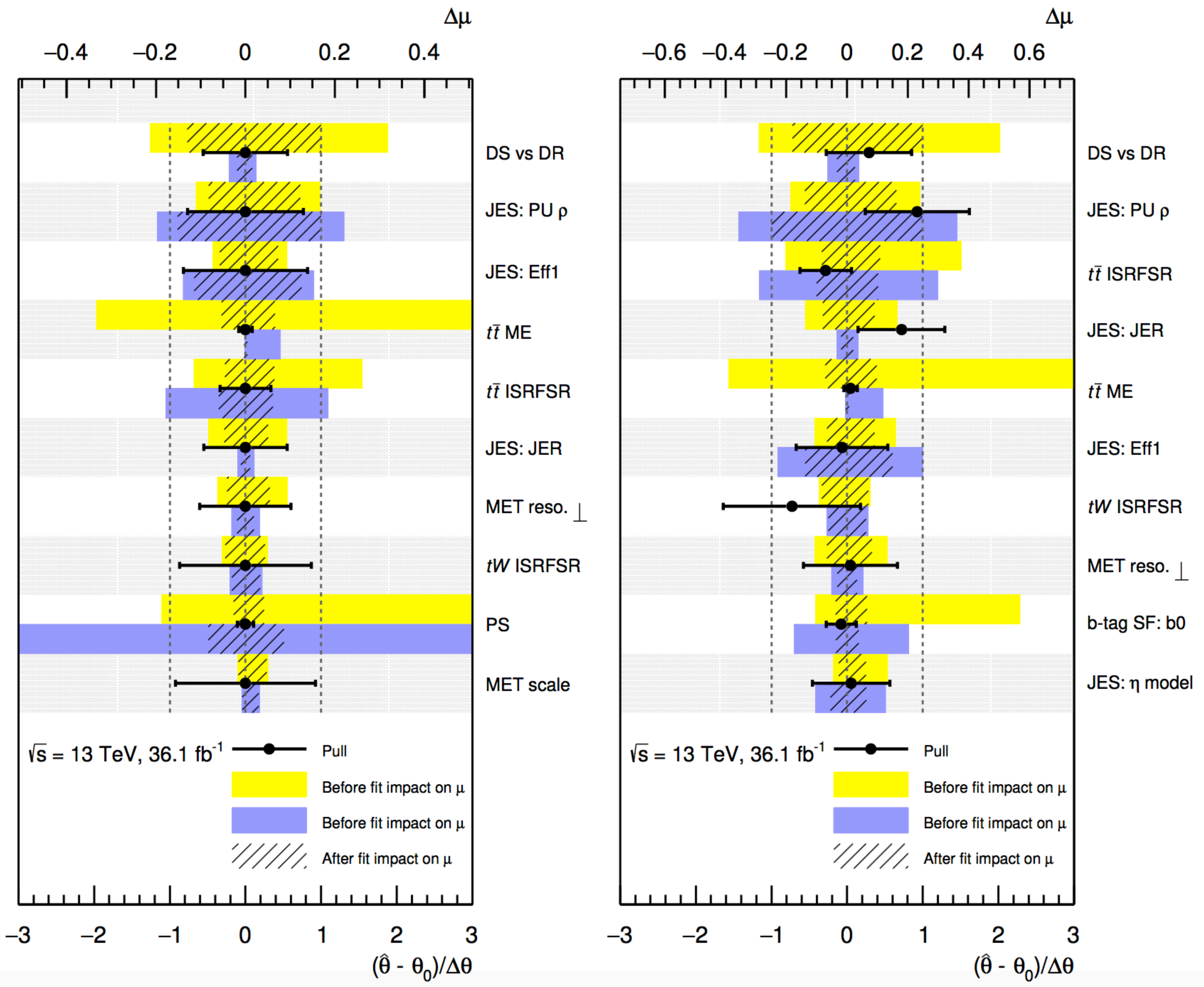


Correlation Matrix Real



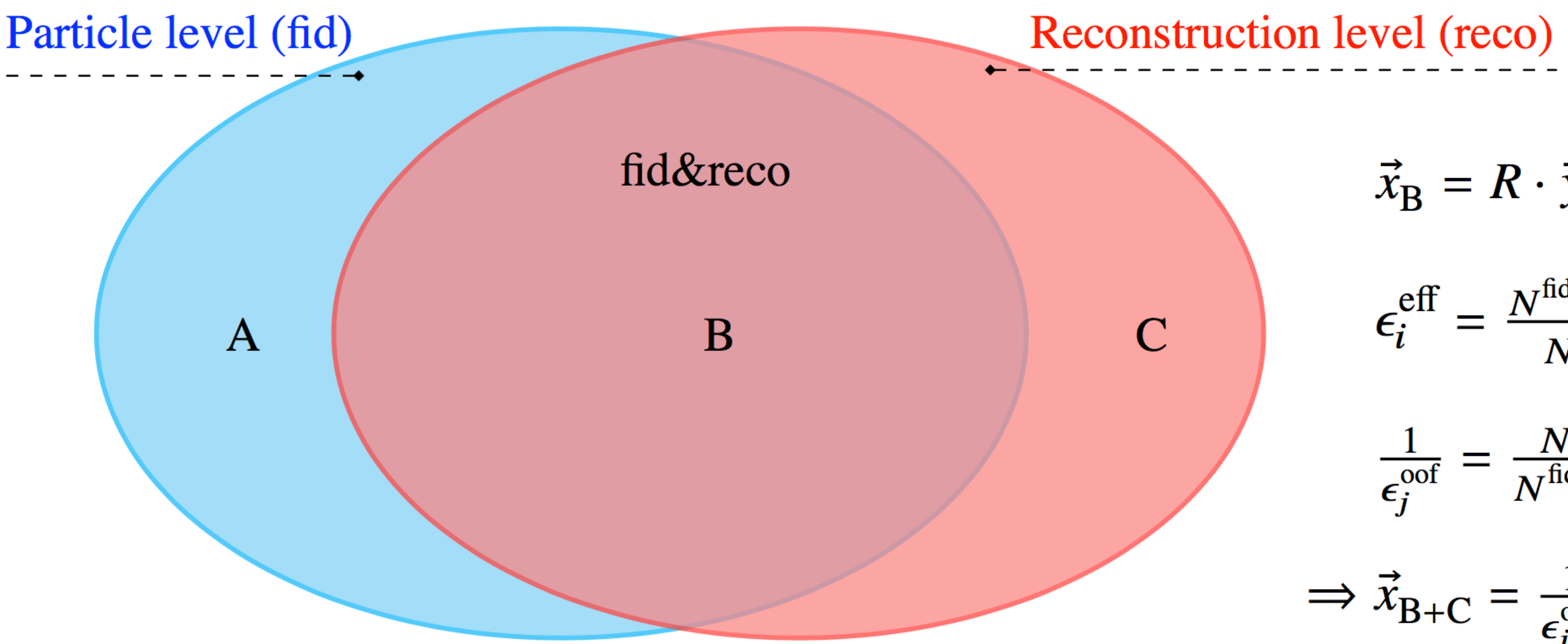
Impact

UNIVERSITÄT BONN



- Two ways to check impact of a NP on μ (upper scale)
  - Yellow: fix "the" NP to post/pre-fit value up/down, float others, float μ
  - Purple: fix "the" NP to post/pre-fit value up/down, fix others, float μ
- For example, two NPs and μ => 3 parameters
  - NP1: prefit: 0±1; postfit: 0.1±0.7
  - NP2: prefit: 0±1; postfit: -0.1±0.9
  - μ: postfit: 0.8
  - Yellow: fix NP1=1/-1/-0.6/0.8, fit with 2 parameters
  - Purple: fix NP1=1/-1/-0.6/0.8, fix NP2=0/0/-0.1/-0.1, fit with 1 parameter

Unfolding and RooUnfold package



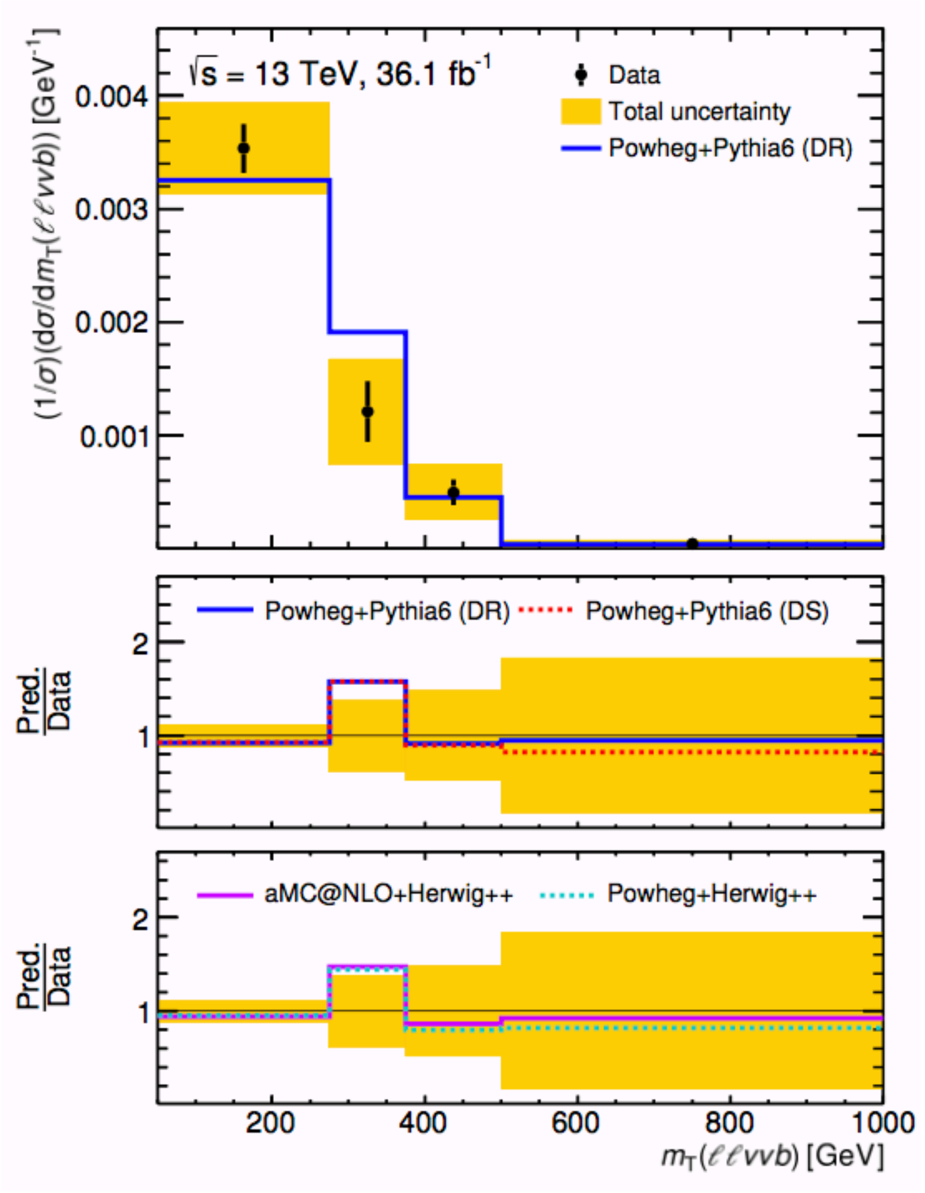
$$\vec{x}_{\mathrm{B}} = R \cdot \vec{y}_{\mathrm{B}}$$

$$\epsilon_i^{\mathrm{eff}} = \frac{N^{\mathrm{fid\&reco}}}{N^{\mathrm{fid}}} = \frac{\mathrm{B}}{\mathrm{A+B}}$$

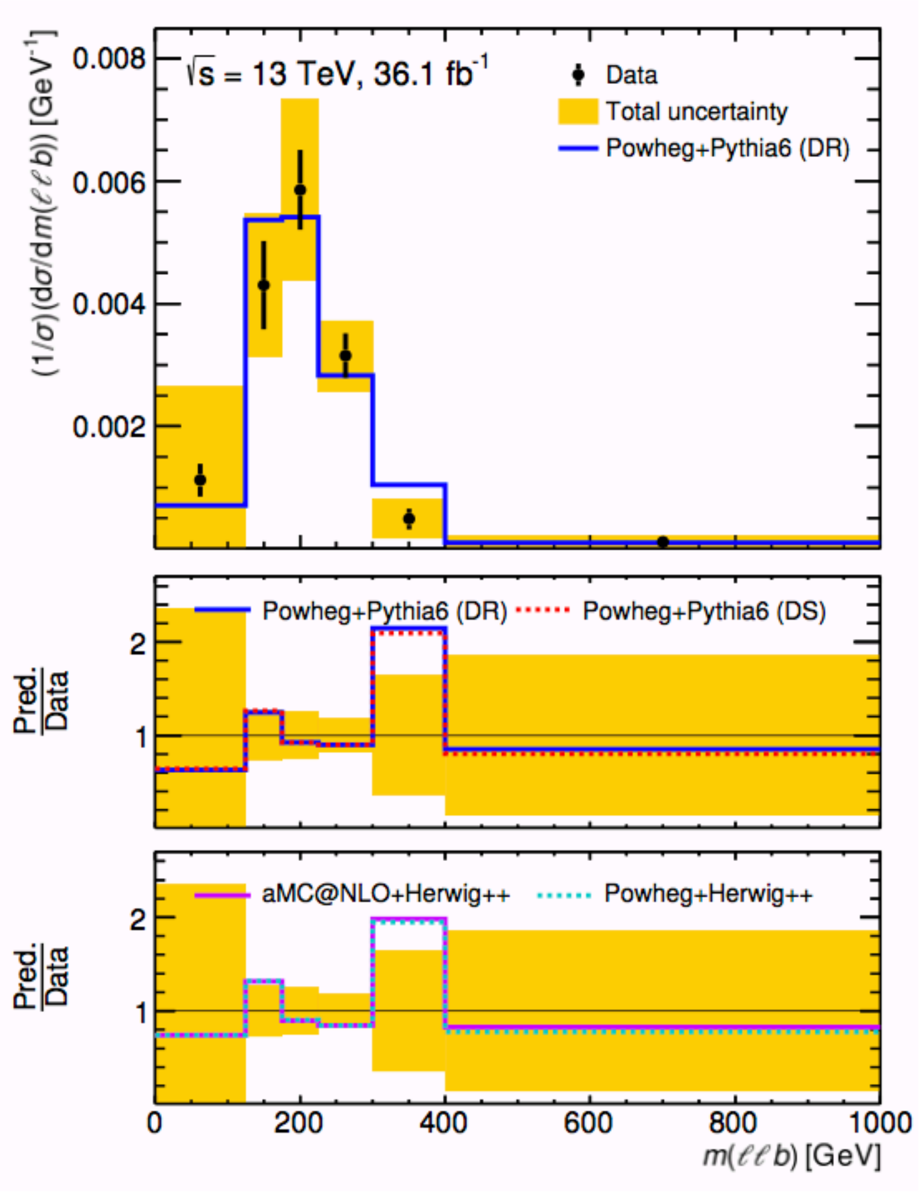
$$\frac{1}{\epsilon_j^{\mathrm{oof}}} = \frac{N^{\mathrm{reco}}}{N^{\mathrm{fid\&reco}}} = \frac{\mathrm{B+C}}{\mathrm{B}}$$

$$\Rightarrow \vec{x}_{\mathrm{B+C}} = \frac{1}{\epsilon_j^{\mathrm{oof}}} \cdot R \cdot \epsilon_i^{\mathrm{eff}} \vec{y}_{\mathrm{A+B}}$$

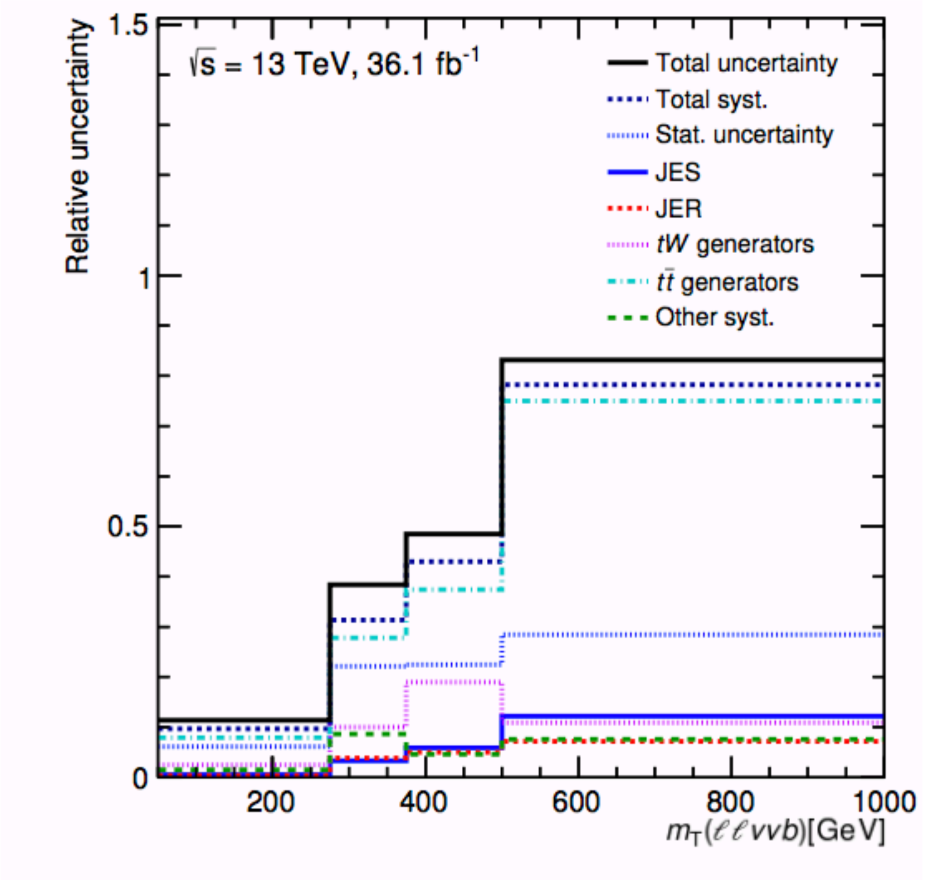
Results

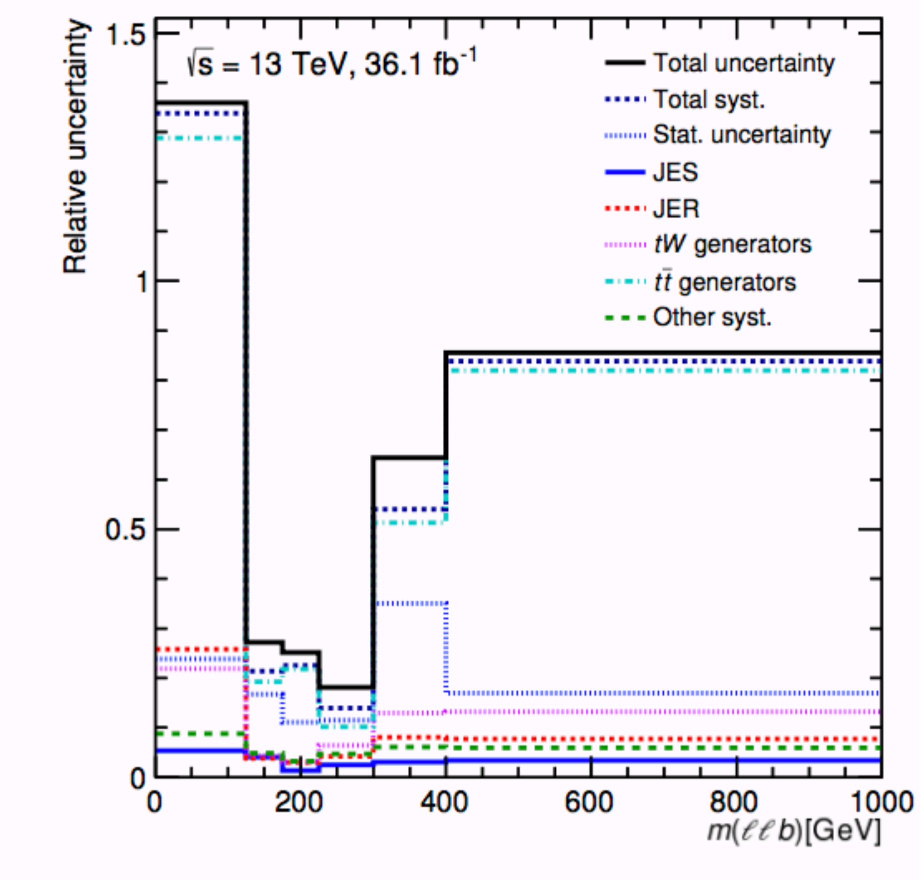


(a)

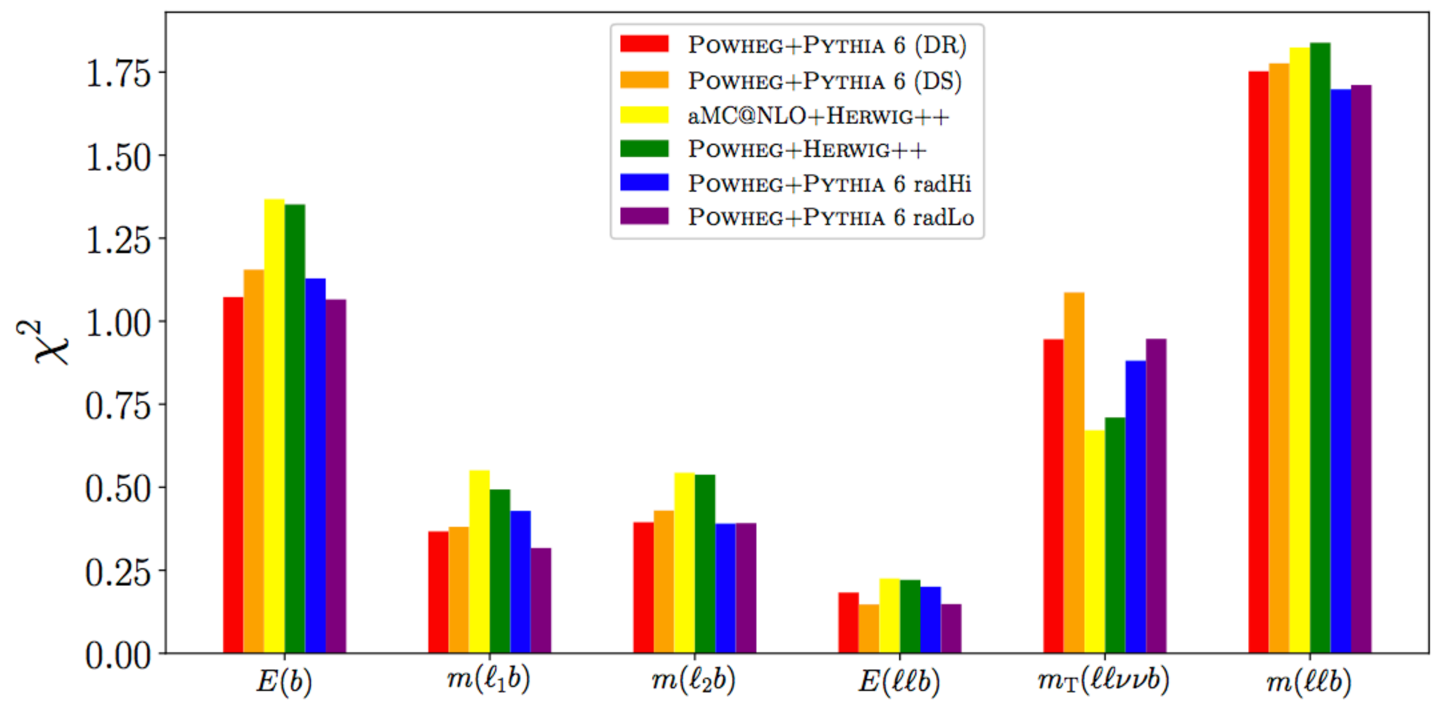


(b)

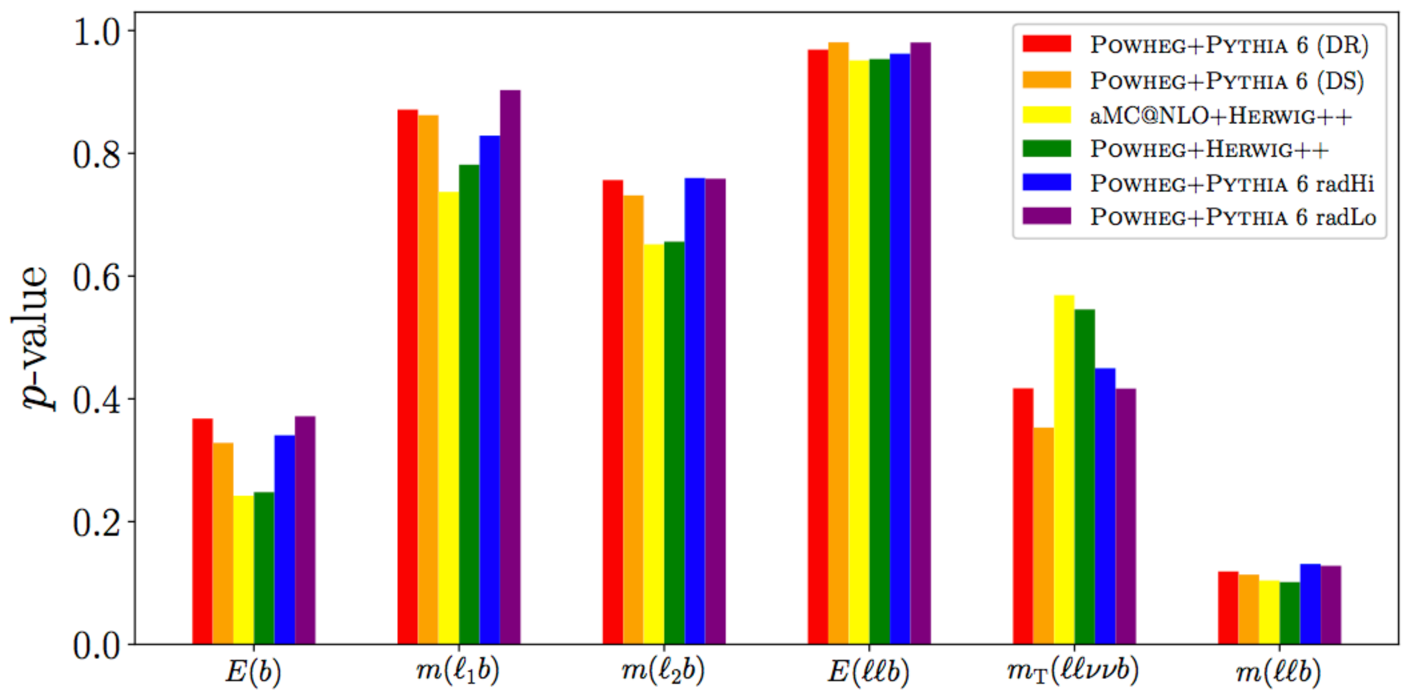




$E(b)$ bin [GeV]	[25, 60]	[60, 100]	[100, 135]	[135, 175]	[175, 500]	
$(1/\sigma)\ d\sigma/dx$ [GeV^{-1}]	0.00422	0.00937	0.00428	0.00169	0.000801	
Stat. uncertainty	22	12	27	48	11	
Total syst. uncertainty	27	28	30	59	28	
Total uncertainty	35	31	40	76	30	
$m(\ell_1 b)$ bin [GeV]	[0, 60]	[60, 100]	[100, 150]	[150, 200]	[200, 250]	[250, 400]
$(1/\sigma)\ d\sigma/dx$ [GeV^{-1}]	0.00105	0.0036	0.00887	0.00328	0.00163	0.000692
Stat. uncertainty	29	25	9.6	18	25	16
Total syst. uncertainty	54	48	21	29	26	47
Total uncertainty	61	54	23	34	36	50
$m(\ell_2 b)$ bin [GeV]	[0, 50]	[50, 100]	[100, 150]	[150, 400]		
$(1/\sigma)\ d\sigma/dx$ [GeV^{-1}]	0.00205	0.00885	0.00654	0.000512		
Stat. uncertainty	26	9.9	9.9	17		
Total syst. uncertainty	83	8.4	9.9	65		
Total uncertainty	87	13	14	67		
$E(\ell\ell b)$ bin [GeV]	[50, 175]	[175, 275]	[275, 375]	[375, 500]	[500, 700]	[700, 1200]
$(1/\sigma)\ d\sigma/dx$ [GeV^{-1}]	0.00069	0.00364	0.0025	0.0014	0.000435	7.39×10^{-5}
Stat. uncertainty	30	9.7	12	15	22	34
Total syst. uncertainty	95	12	22	21	66	74
Total uncertainty	99	16	25	26	70	81
$m_T(\ell\ell\nu\nu b)$ bin [GeV]	[50, 275]	[275, 375]	[375, 500]	[500, 1000]		
$(1/\sigma)\ d\sigma/dx$ [GeV^{-1}]	0.00353	0.00121	0.000498	4.25×10^{-5}		
Stat. uncertainty	6.1	22	22	28		
Total syst. uncertainty	9.6	31	43	78		
Total uncertainty	11	38	48	83		
$m(\ell\ell b)$ bin [GeV]	[0, 125]	[125, 175]	[175, 225]	[225, 300]	[300, 400]	[400, 1000]
$(1/\sigma)\ d\sigma/dx$ [GeV^{-1}]	0.00112	0.0043	0.00586	0.00315	0.000486	0.000112
Stat. uncertainty	24	17	11	12	35	17
Total syst. uncertainty	130	21	23	14	54	84
Total uncertainty	140	27	25	18	64	86



Powheg+Pythia 6 (DR)
Powheg+Pythia 6 (DS)
aMC@NLO+Herwig++
Powheg+Herwig++
Powheg+Pythia 6 radHi
Powheg+Pythia 6 radLo
χ^2
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
$E(b)$
$m(\ell_1 b)$
$m(\ell_2 b)$
$E(\ell\ell b)$
$m_T(\ell\ell\nu\nu b)$
$m(\ell\ell b)$



Powheg+Pythia 6 (DR)
Powheg+Pythia 6 (DS)
aMC@NLO+Herwig++
Powheg+Herwig++
Powheg+Pythia 6 radHi
Powheg+Pythia 6 radLo
p-value
0.0
0.2
0.4
0.6
0.8
1.0
$E(b)$
$m(\ell_1 b)$
$m(\ell_2 b)$
$E(\ell\ell b)$
$m_T(\ell\ell\nu\nu b)$
$m(\ell\ell b)$