Unsupervised ML & Bayesian Inference @ four-tops

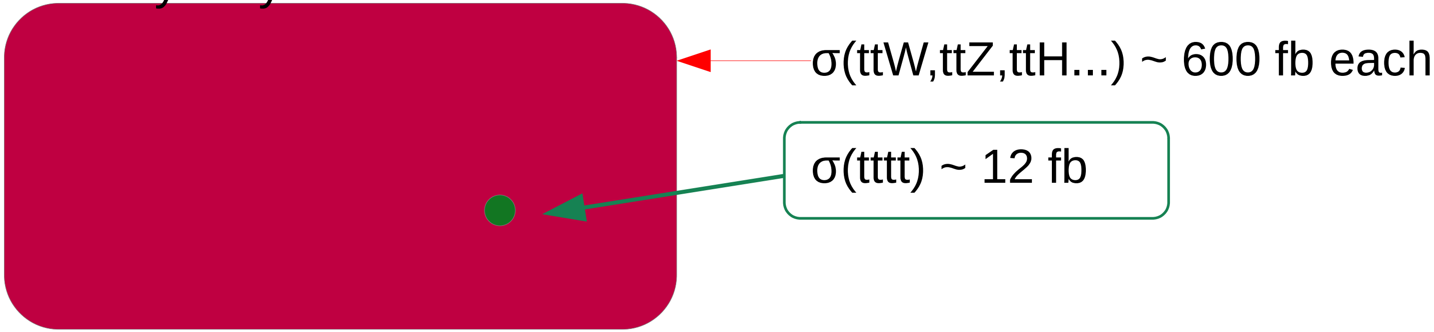
Ezequiel Álvarez[1]
sequi@unsam.edu.ar

In collaboration with
B.Dillon[2], D.Faroughy[3], J.Kamenik[4], F.Lamagna[1] & M.Szewc[1]

1) ICAS & IB (Argentina); 2) Heidelberg U. (Germany); 3) Zurich U. (Switzerland); 4) Jozef Stefan I. & Ljubljana U. (Slovenia)

Four-tops @ LHC

Already very small...



Measured cross-sections

ATLAS: 25^{+7}_{-6} fb

CMS: $12.6^{+5.8}_{-5.2}$ fb

(Details in previous talk by Kong)

Four-tops @ LHC

Already very small...

σ(ttW,ttZ,ttH...) ~ 600 fb each

σ(tttt) ~ 12 fb

Very sensitive window for top-philic NP

Measured cross-sections

ATLAS: 25^{+7}_{-6} fb

CMS: $12.6^{+5.8}_{-5.2}$ fb

(Details in previous talk by Kong)

Four-tops @ LHC

Already very small...

$\sigma(ttW, ttZ, ttH...) \sim 600$ fb each

$\sigma(tttt) \sim 12$ fb

Very sensitive window for top-philic NP

Measured cross-sections

ATLAS: 25^{+7}_{-6} fb

CMS: $12.6^{+5.8}_{-5.2}$ fb

(Details in previous talk by Kong)

General NP

2105.03372,1910.09581, 1906.09703, 1805.10835 1804.05598,1611.05032, 1206.3064,1203.5862, 1112.3778,1107.4616 1101.1294, 1008.3562, hep-ph/9507411

EFT

2104.09512, 2011.15060, 2010.05915,1708.05928, 1010.6304

Machine Learning

J. Phys.Conf.Ser.**1380** 012069
1911.09699

Four-tops @ LHC

Already very small...

σ(ttW,ttZ,ttH...) ~ 600 fb each

σ(tttt) ~ 12 fb

Very sensitive window for top-philic NP

Measured cross-sections

ATLAS: 25^{+7}_{-6} fb

CMS: $12.6^{+5.8}_{-5.2}$ fb

(Details in previous talk by Kong)

General NP

2105.03372,1910.09581, 1906.09703, 1805.10835 1804.05598,1611.05032, 1206.3064,1203.5862, 1112.3778,1107.4616 1101.1294, 1008.3562, hep-ph/9507411

EFT

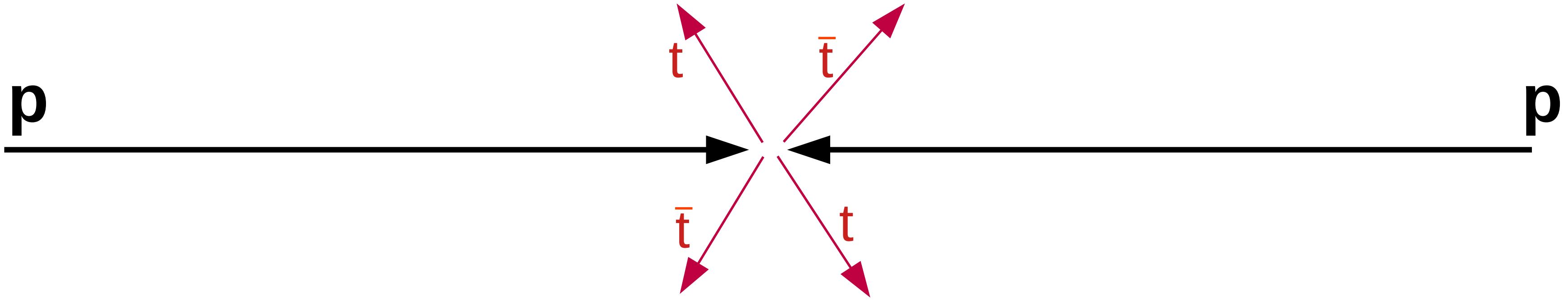
2104.09512, 2011.15060, 2010.05915,1708.05928, 1010.6304

Machine Learning

J. Phys.Conf.Ser.**1380** 012069 1911.09699

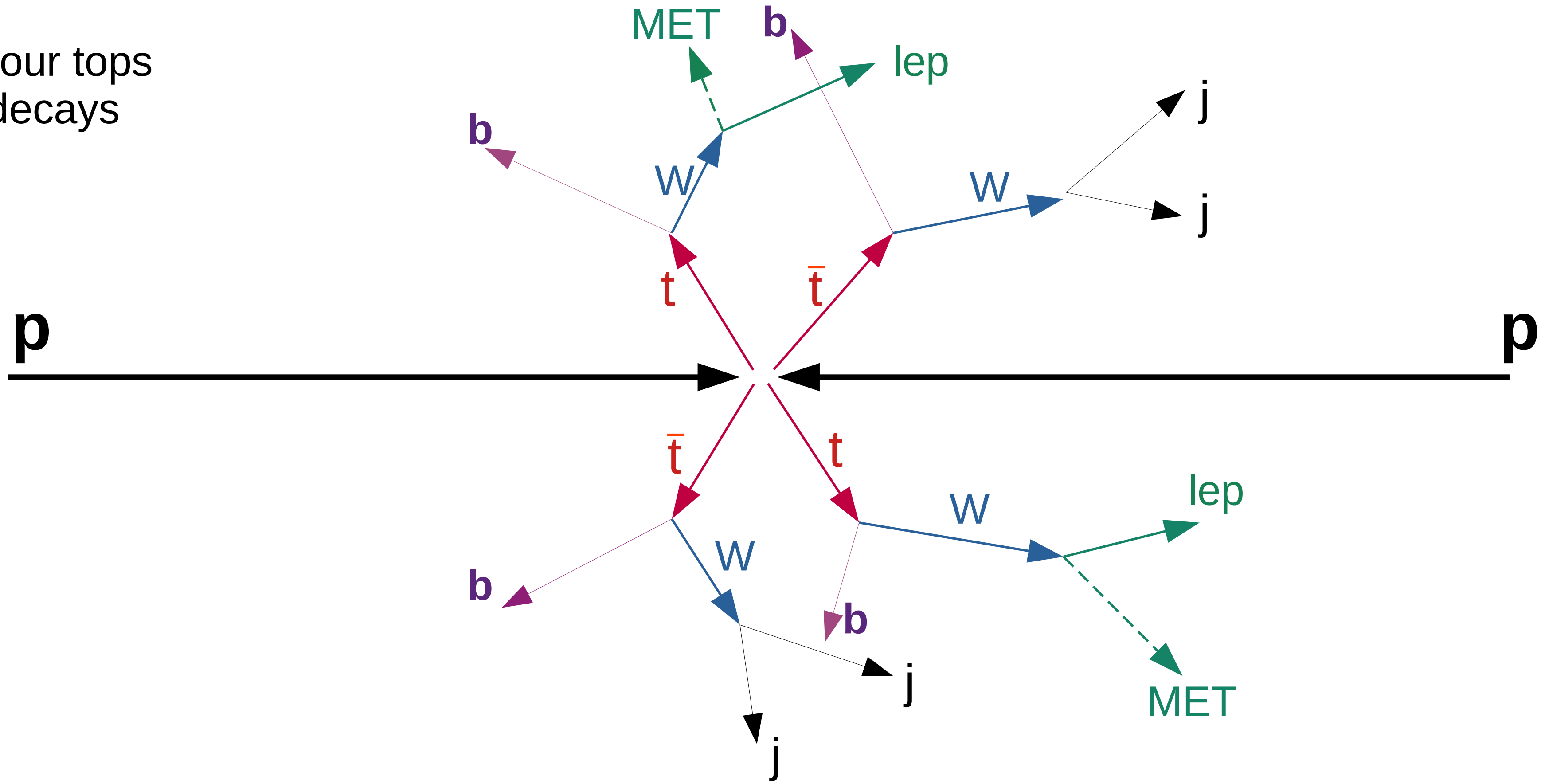
Four-top production

Four tops



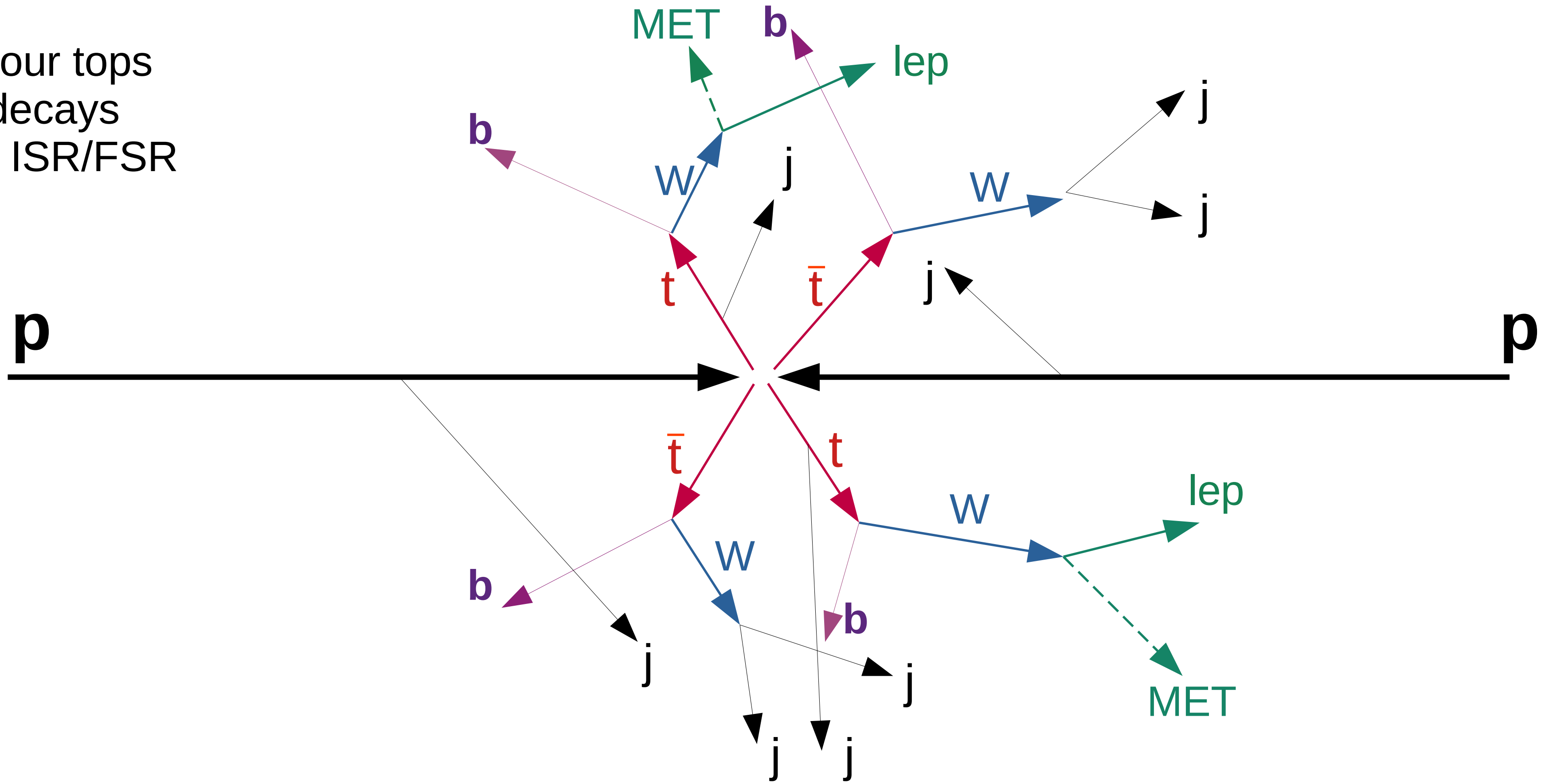
Four-top production

Four tops decays



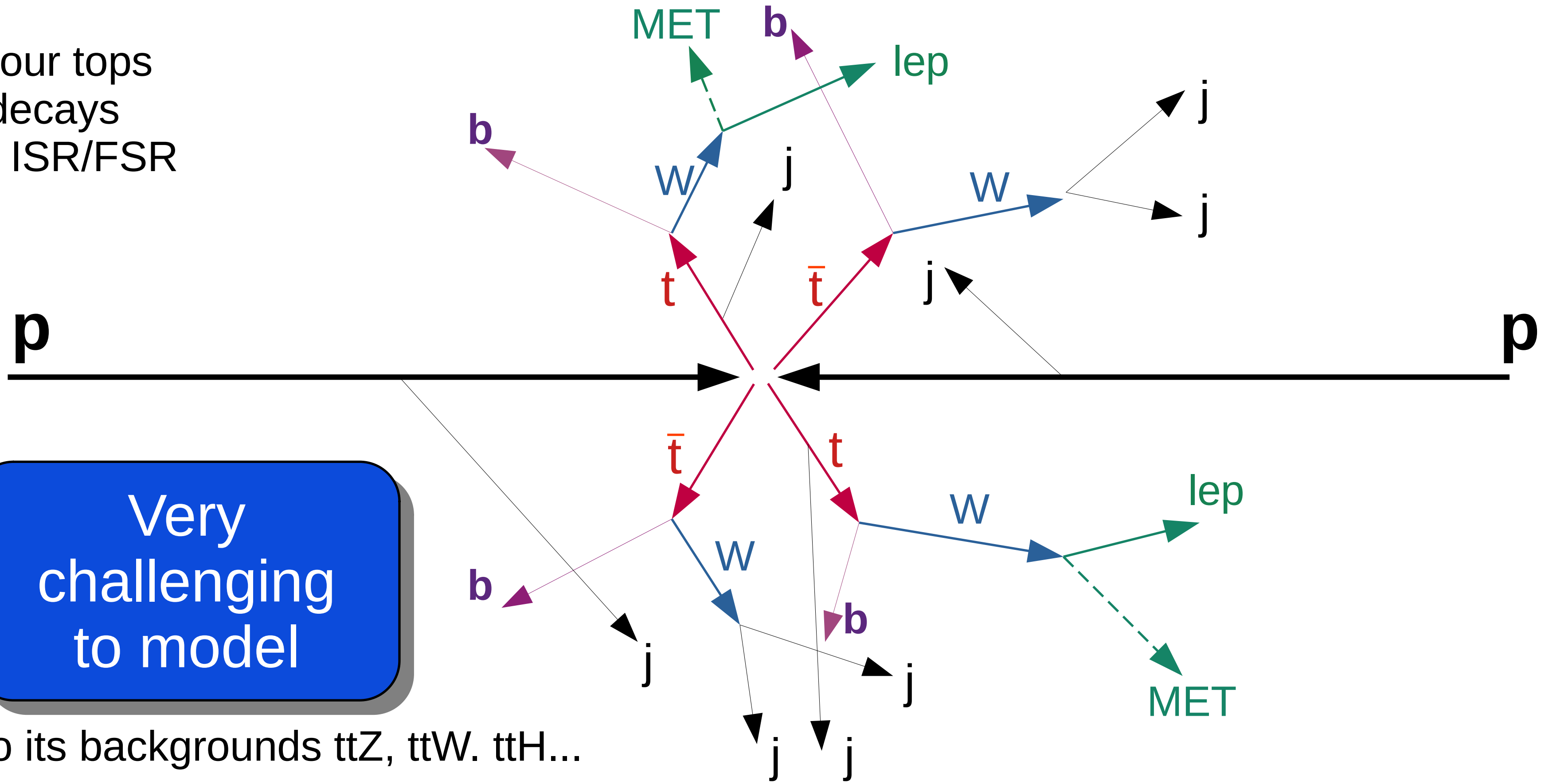
Four-top production

Four tops
decays
+ ISR/FSR



Four-top production

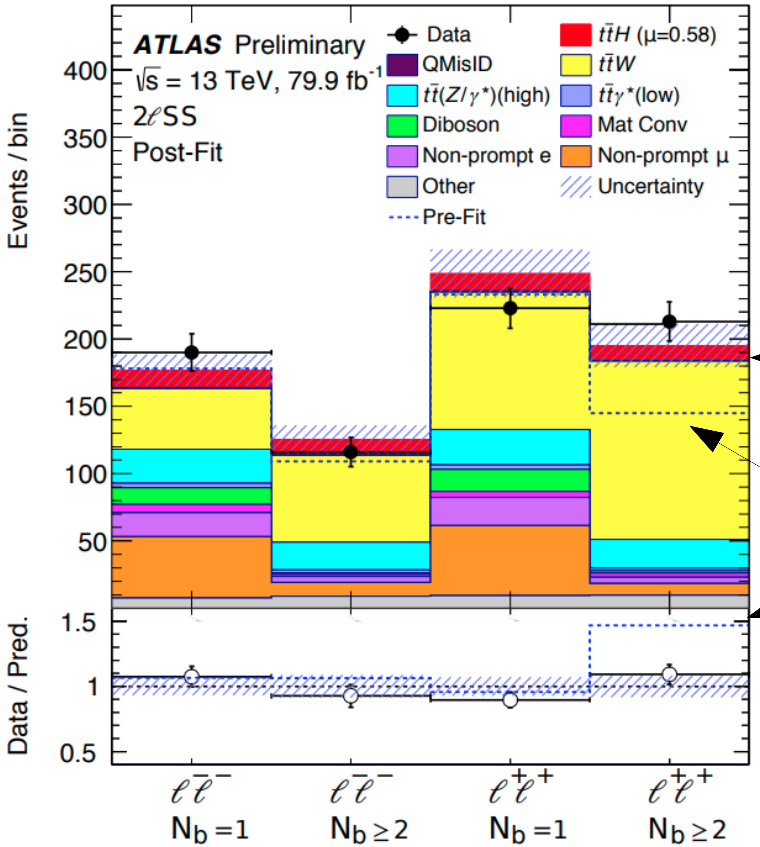
Four tops
decays
+ ISR/FSR



Very challenging to model

Also its backgrounds ttZ, ttW. ttH...

Browsing the data



ttH

ATLAS-CONF-2019-045

Post-fit

Pre-fit

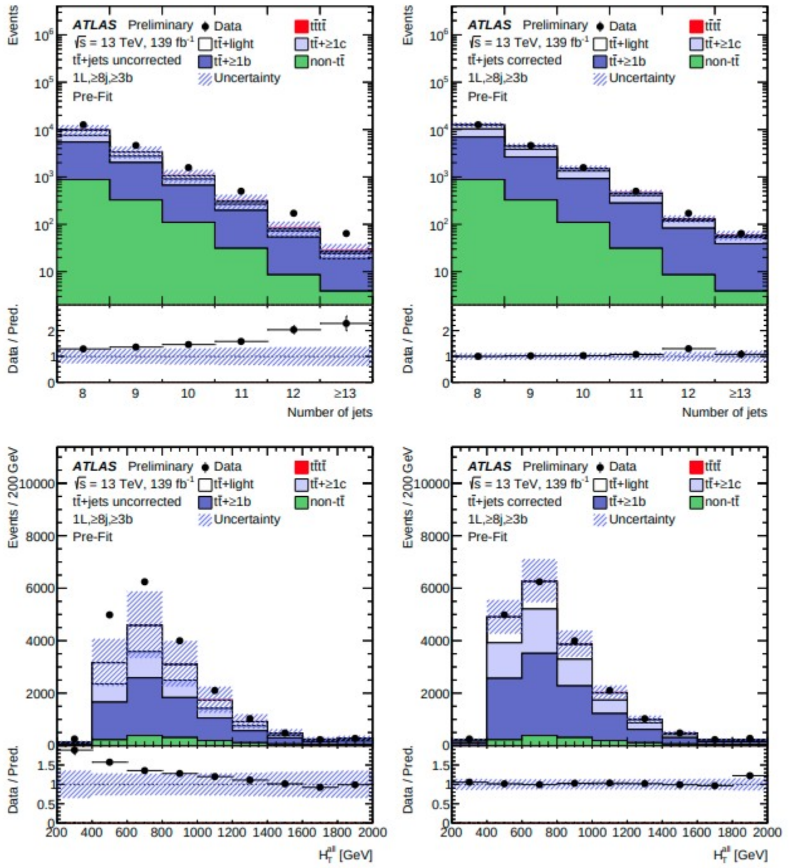
(See 2011.06514 for NP on this mismatch)

Browsing the data

- Flavour rescaling
- Sequential kinematic reweighting

Four-tops
ATLAS-CONF-2021-013

Before



After

N_j

H_T

See talk by
Albert Kong

Browsing the data

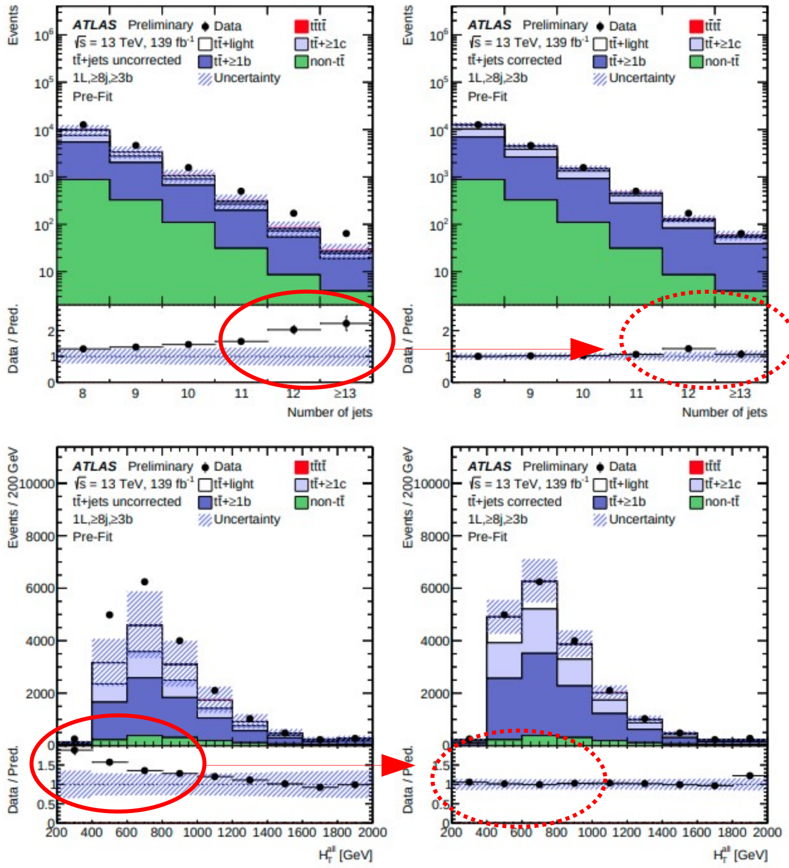
- Flavour rescaling
- Sequential kinematic reweighting

Four-tops

ATLAS-CONF-2021-013

N_j

Before



After

H_T

See talk by Albert Kong

Browsing the data

In multilepton and/or multijets/b, some results have tuning

→ Quite clever techniques
→ In principle do not modify results
→ Some times after having seen the data

Browsing the data

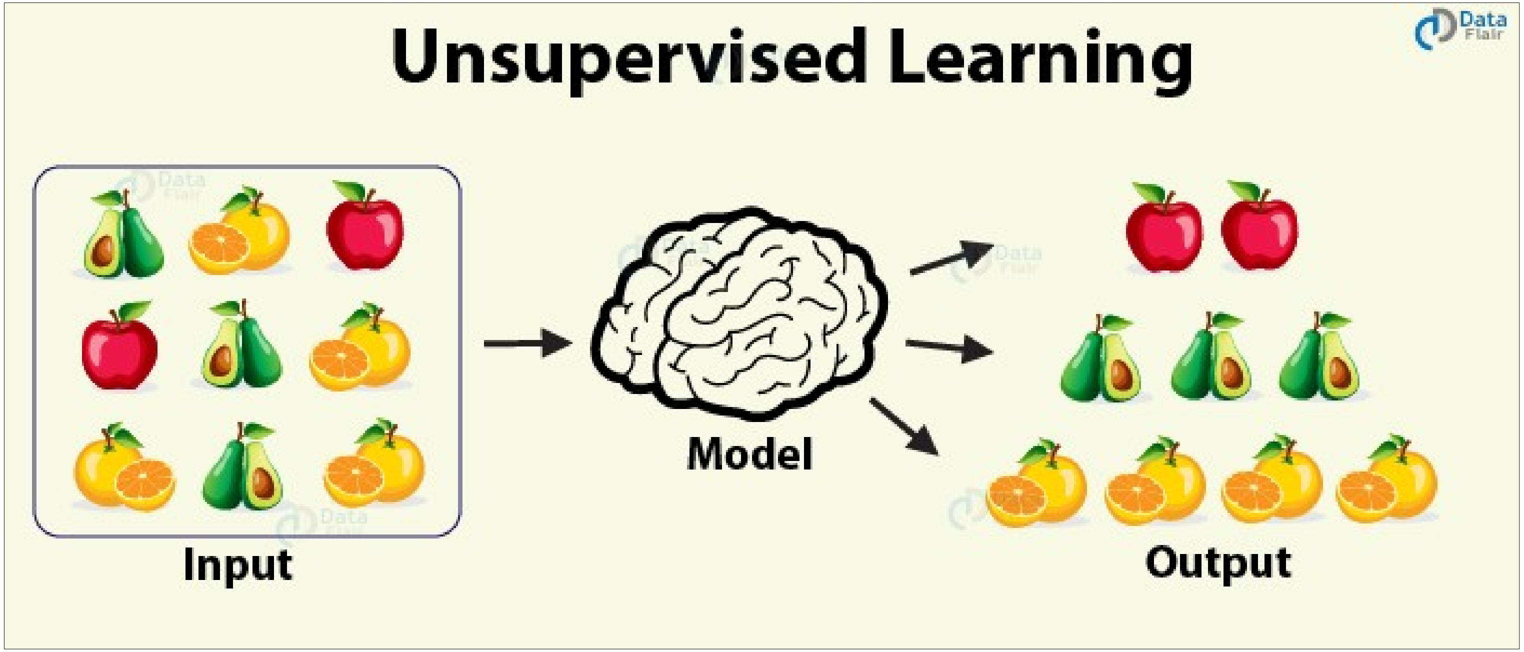
In multilepton and/or multijets/b, some results have tuning

→ Quite clever techniques
→ In principle do not modify results
→ Some times after having seen the data

But….

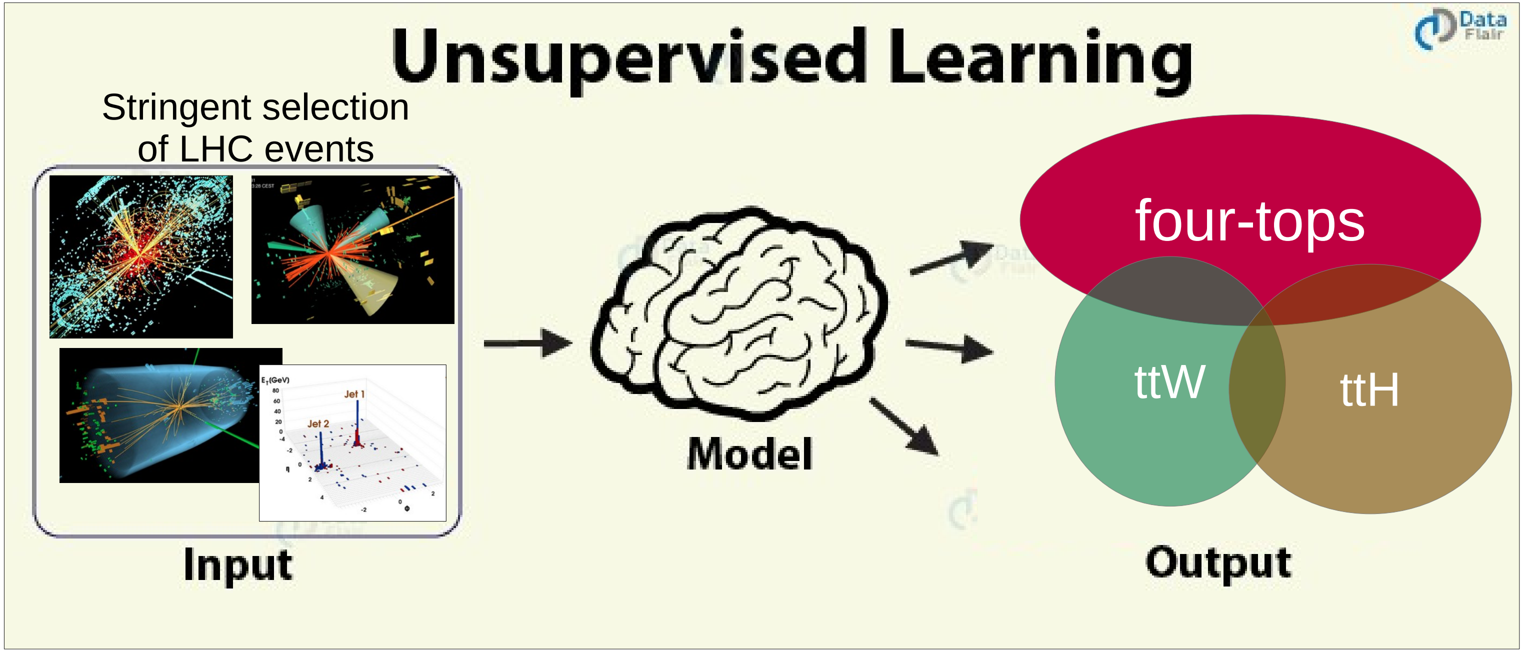
What if we are missing something ?

Unsupervised ML @ four-tops



The algorithm recognizes similarities and differences and clusters the data

Unsupervised ML @ four-tops

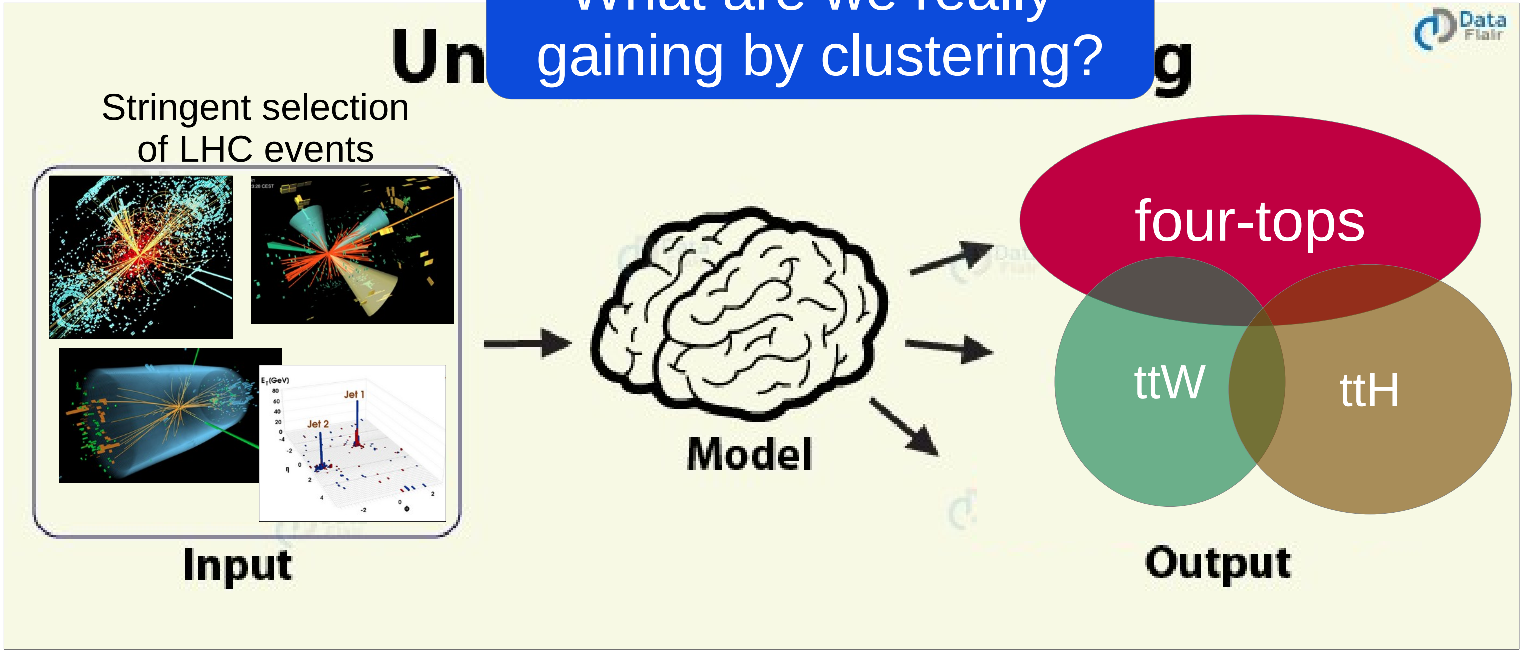


?

The algorithm recognizes similarities and differences and clusters the data

Unsupervised ML @ four-tops

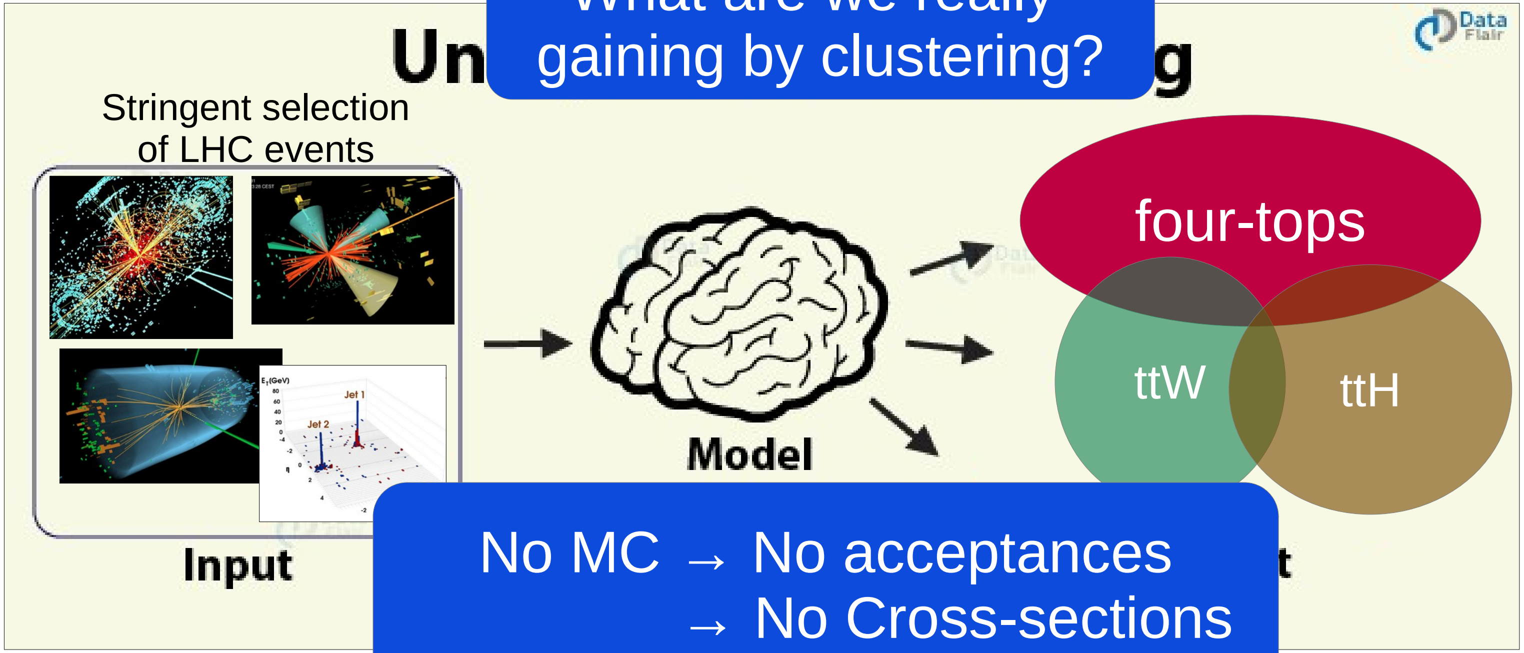
What are we really gaining by clustering?



?

Unsupervised ML @ four-tops

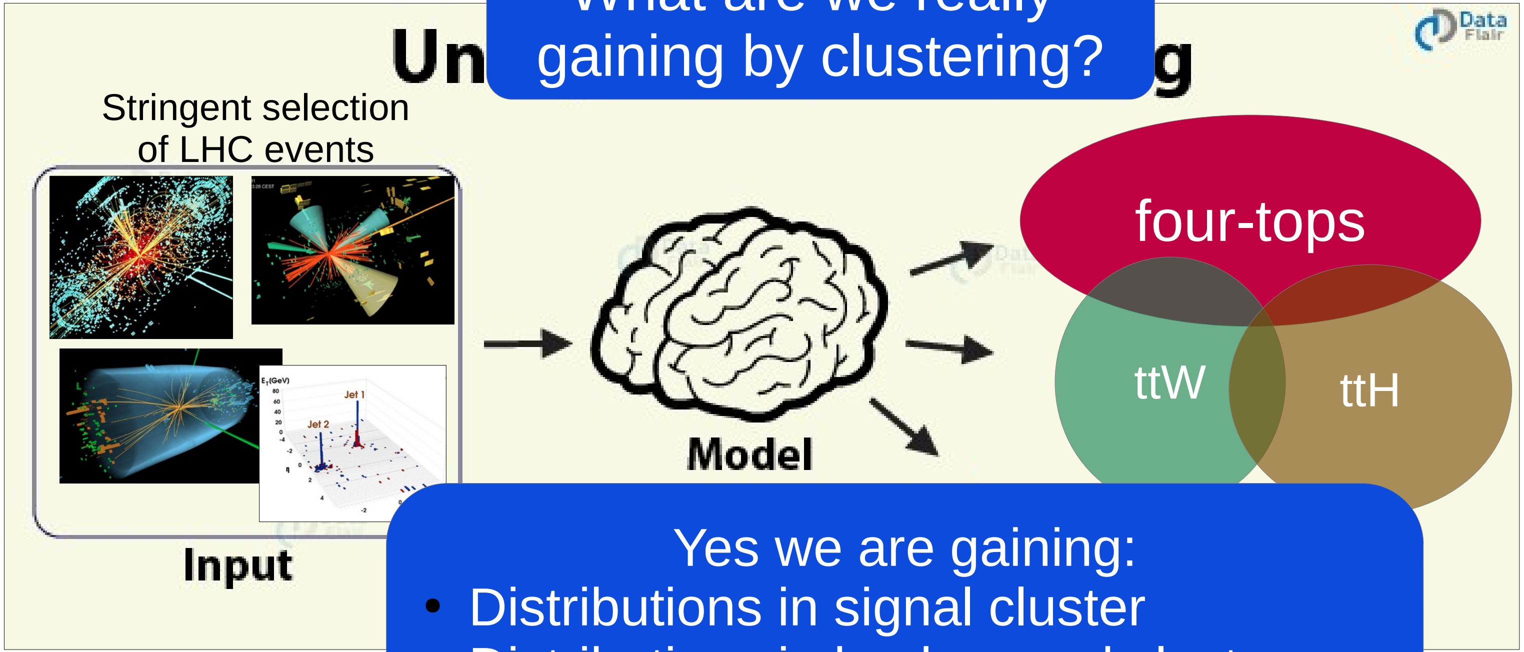
What are we really
gaining by clustering?



?

No MC → No acceptances
→ No Cross-sections

Unsupervised ML @ four-tops



What are we really gaining by clustering?

?

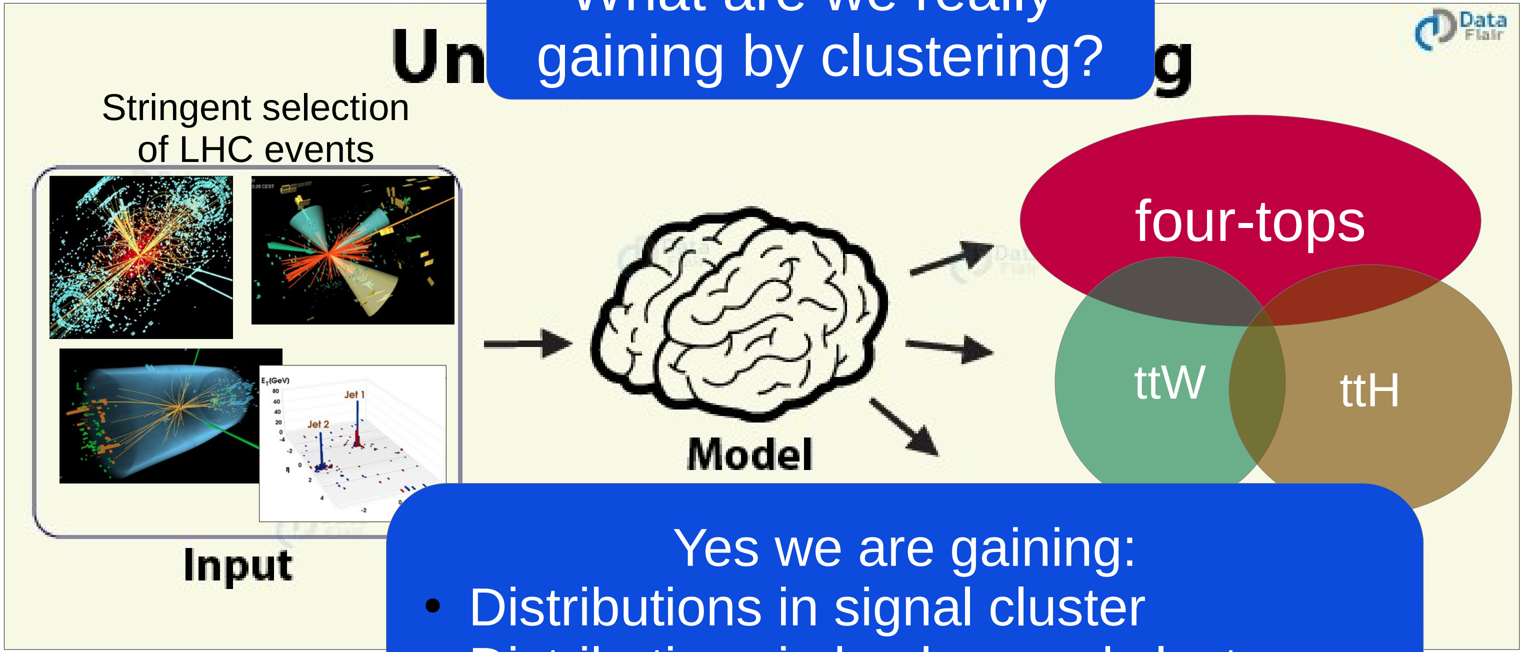
Yes we are gaining:

- Distributions in signal cluster
- Distributions in background cluster
- Tuning MC with background cluster in signal region

Many subtleties!

Unsupervised ML @ four-tops

What are we really gaining by clustering?



?

Yes we are gaining:

- Distributions in signal cluster
- Distributions in background cluster
- Tuning MC with background cluster in signal region

Many subtleties!

Unsupervised ML @ four-tops

We've started studying

- Latent Dirichlet Allocation (LDA)
- Autoencoders (AE)
- Variational Autoencoders (VAE)

Unsupervised ML @ four-tops

We’ve started studying

- Latent Dirichlet Allocation (LDA)
- Autoencoders (AE)
- Variational Autoencoders (VAE)

Dataset 2LSS++

Signal = tttt

Background = ttW+

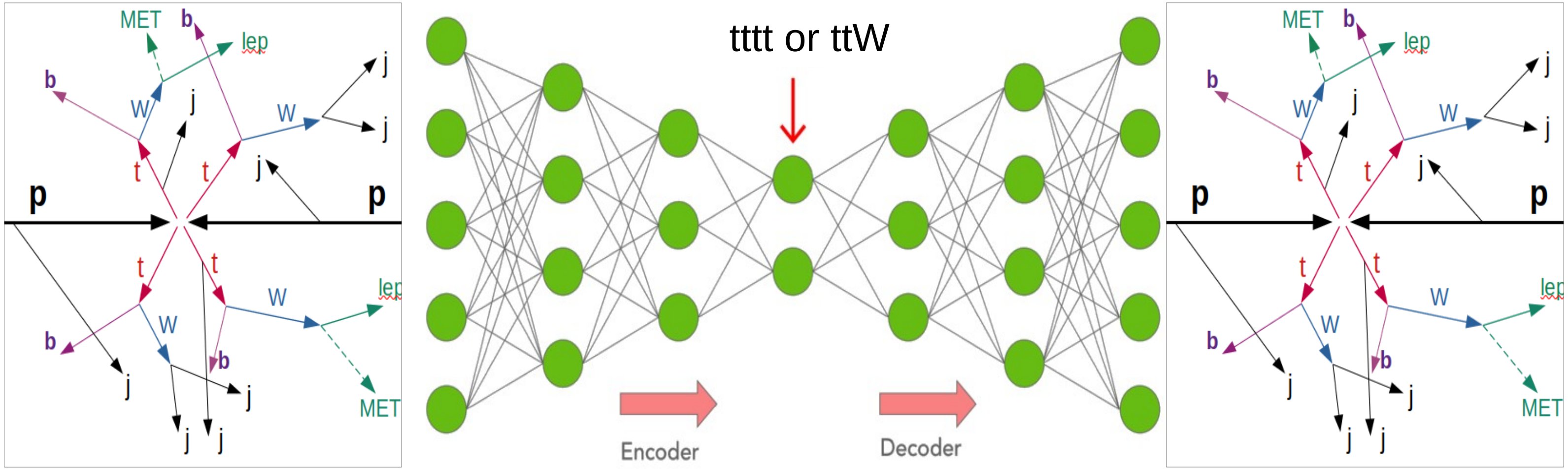
10 fb^{-1} / 600 fb^{-1} $\rightarrow$ ~ 1/1 events

Unsupervised ML @ four-tops

We've started studying

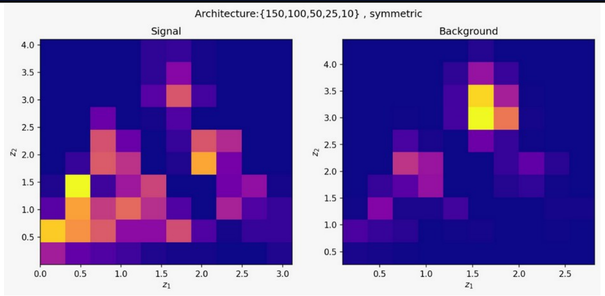
- Latent Dirichlet Allocation (LDA)
- Autoencoders (AE)
- Variational Autoencoders (VAE)

Dataset 2LSS++
Signal = tttt
Background = ttW+
10 fb^{-1} / 600 fb^{-1} → ~ 1/1 events

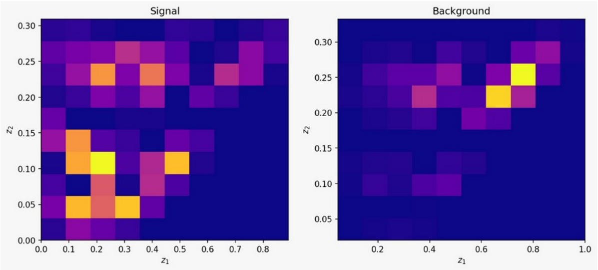


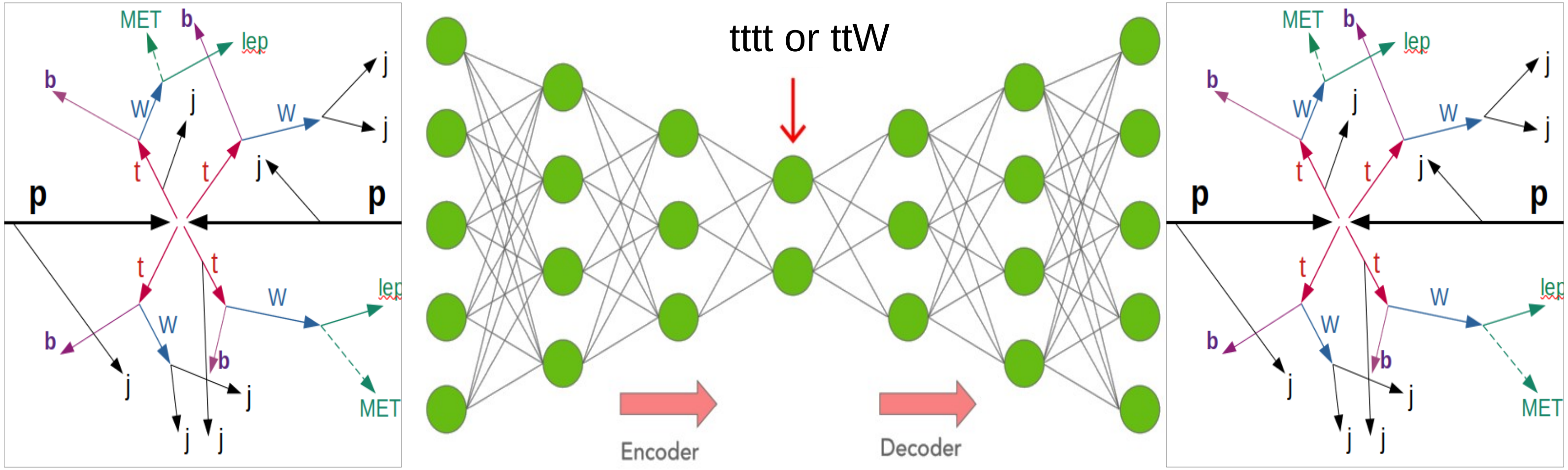
~ 50 input neurons

Unsupervised ML @ four-tops



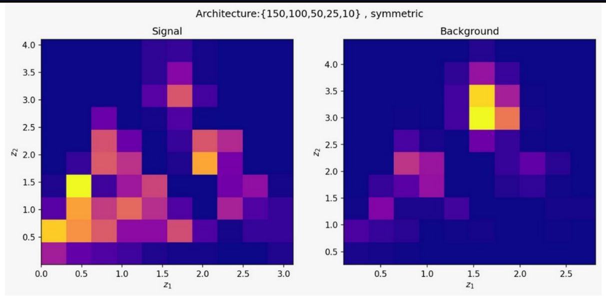
We find good S and B unsupervised clustering



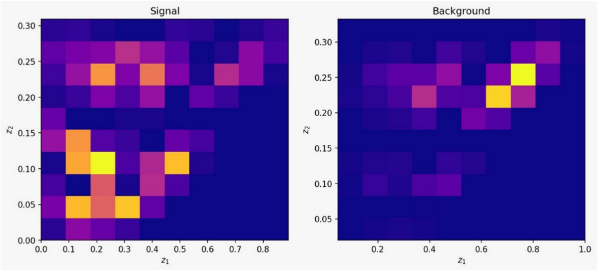


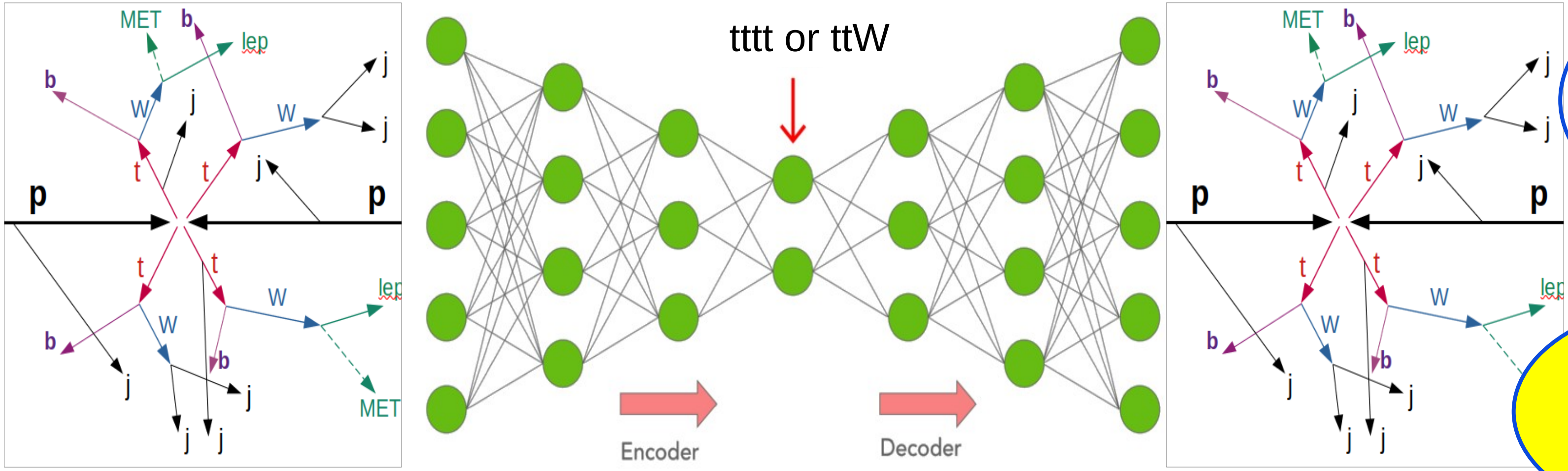
~ 50 input neurons

Unsupervised ML @ four-tops



We find good S and B unsupervised clustering

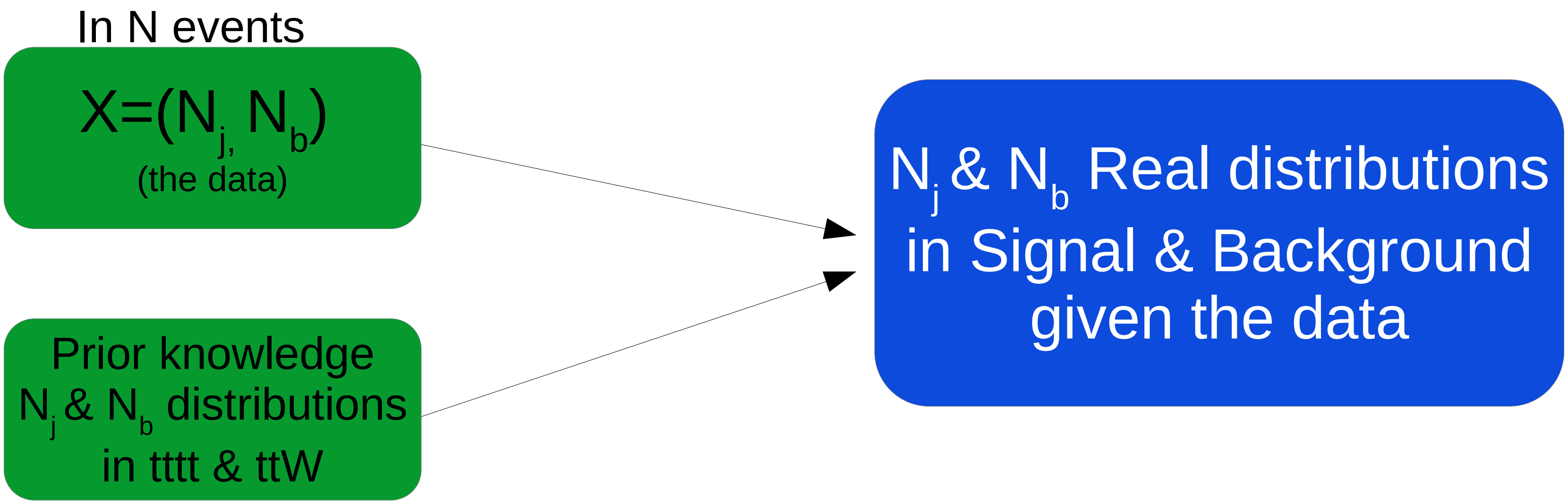




~ 50 input neurons

N_j and N_b relevants!

Bayesian Inference @ four-tops



Bayesian Inference @ four-tops

In N events

$X=(N_{j,} N_b)$
(the data)

Prior knowledge
N_j & N_b distributions
in tttt & ttW

Bayes

$$P(\theta|X) = \frac{P(X|\theta) \times Prior(\theta)}{P(X)}$$

Bayesian Inference @ four-tops

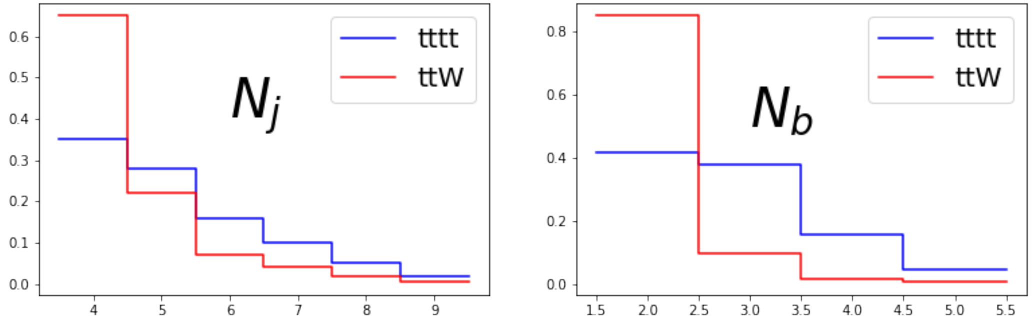
In N events

$X=(N_j, N_b)$
(the data)

Bayes

$$P(\theta|X) = \frac{P(X|\theta) \times Prior(\theta)}{P(X)}$$

Prior knowledge
N_j & N_b distributions
in tttt & ttW



Basic principles

Bayesian Inference @ four-tops

In N events

$X=(N_{j,} N_b)$

(the data)

Unknown mixture!??

Bayes

$$P(\theta|X) = \frac{P(X|\theta) \times Prior(\theta)}{P(X)}$$

Prior knowledge
N_j & N_b distributions
in tttt & ttW

Bayesian Inference @ four-tops

In N events

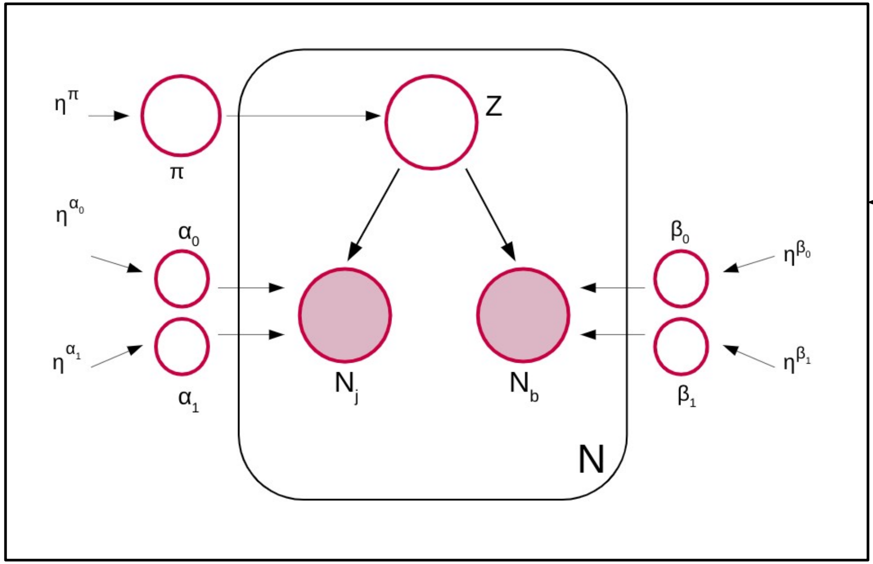
$X=(N_j, N_b)$

(the data)

Unknown mixture!??

Bayes

$$P(\theta|X) = \frac{P(X|\theta) \times Prior(\theta)}{P(X)}$$



Probabilistic *Graphical Model*

We only see (N_j, N_b), but we can infer all the parameters!

Bayesian Inference @ four-tops

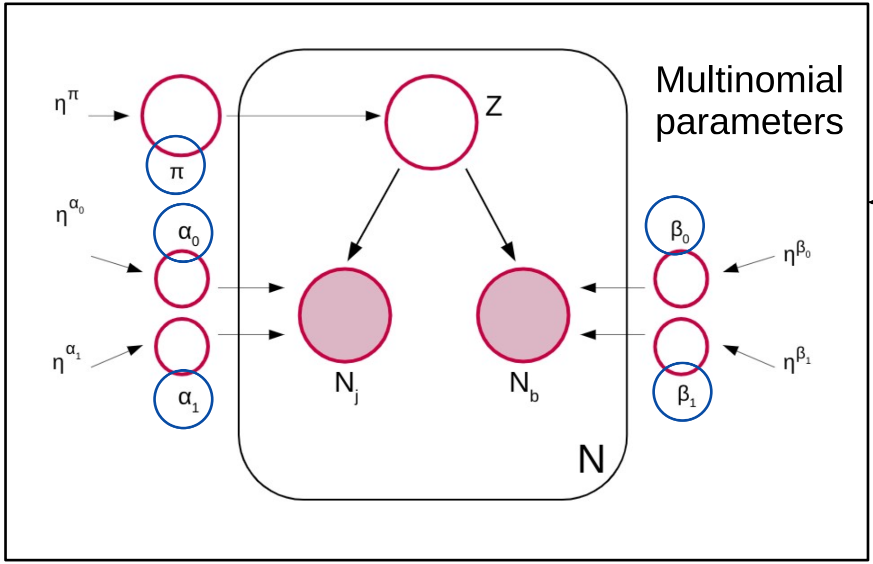
In N events

$X=(N_j, N_b)$
(the data)

Unknown mixture!??

Bayes

$$P(\theta|X) = \frac{P(X|\theta) \times Prior(\theta)}{P(X)}$$



Probabilistic *Graphical Model*

We only see (N_j, N_b), but we can infer all the parameters!

Bayesian Inference @ four-tops

In N events

$X=(N_{j,} N_b)$
(the data)

Prior knowledge
N_j & N_b distributions
in tttt & ttW

Bayes

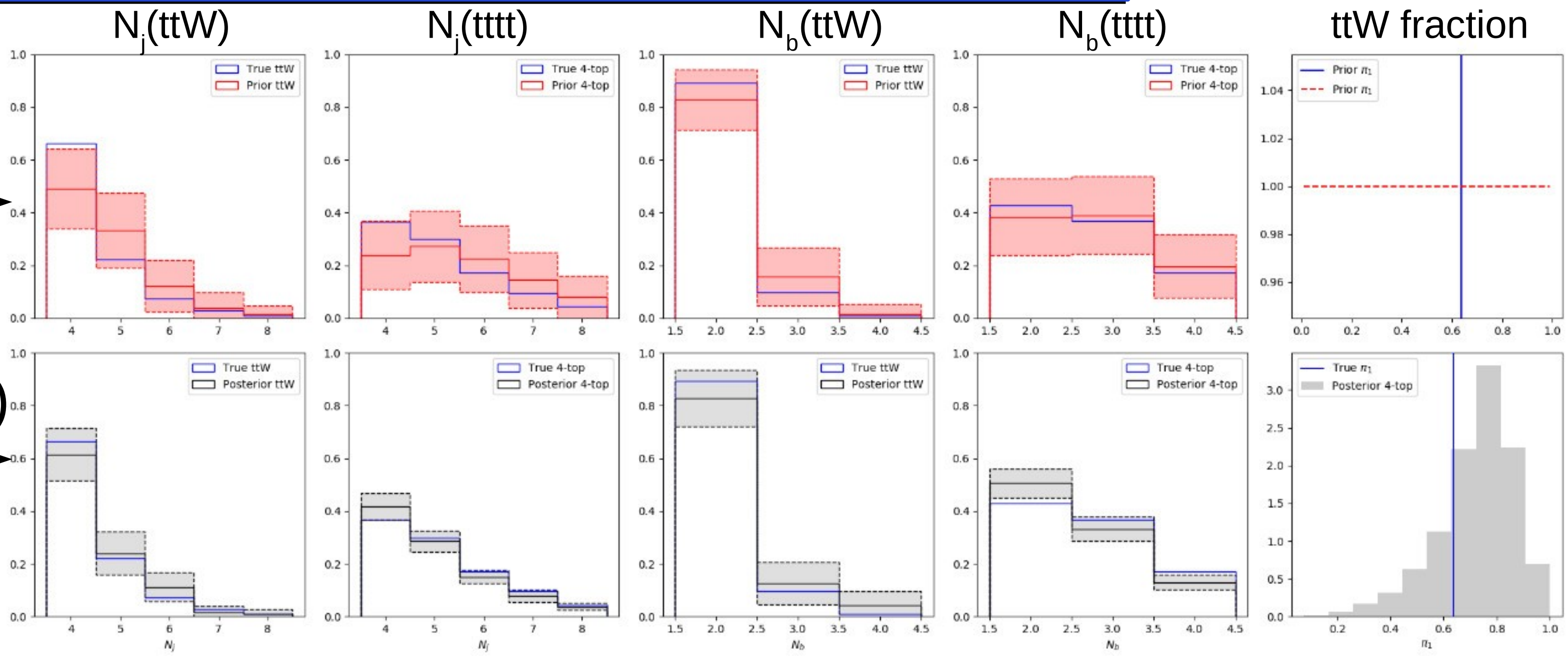
$$P(\theta|X) = \frac{P(X|\theta) \times \text{Prior}(\theta)}{P(X)}$$

Solve Bayesian inference numerically
using Gibbs Sampling
(also EMCEE and others)

Bayesian Inference Results

Prior(θ)

Posterior(θ|X)

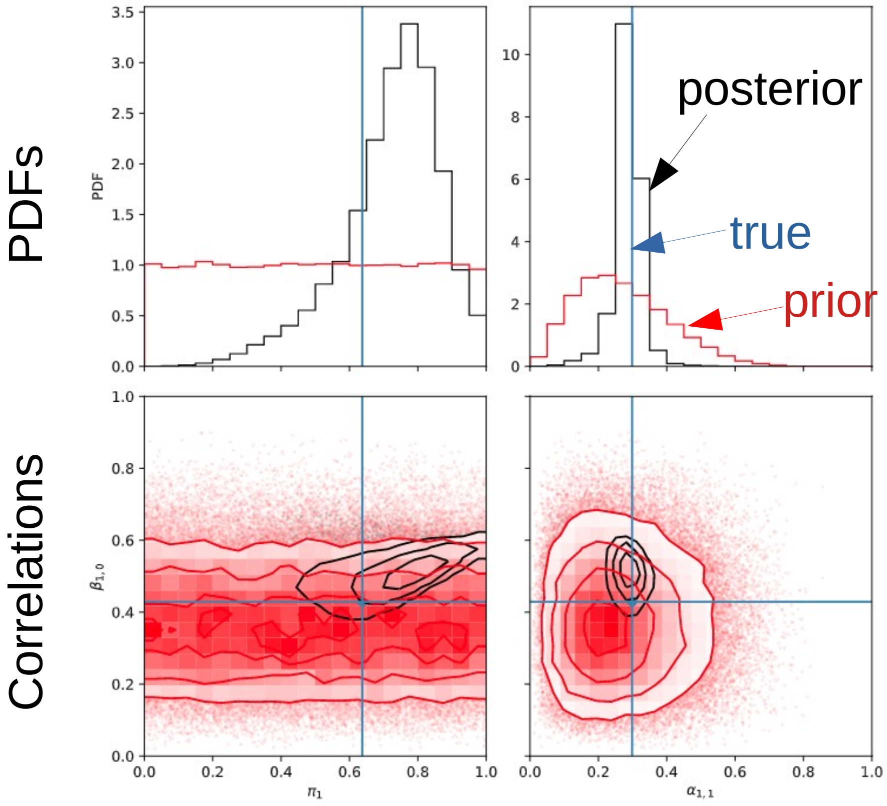


(500 fb^{-1})

$$P(\theta|X) = \frac{P(X|\theta) \times Prior(\theta)}{P(X)}$$

Inference correctly approaches true Values!

Bayesian Inference Results



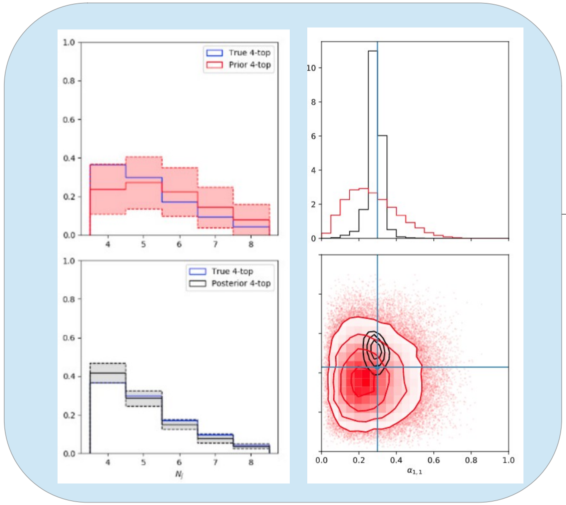
Each parameter approaches the true values with the posterior!

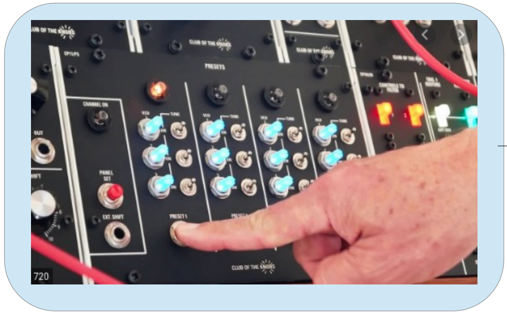
← Corner-plot panels

(500 fb^{-1})

Bayesian Inference @ four-tops

2LSS++





“Test” for NP!

- tttt xsection
- tttt distributions
- Sample fractions
- Etc….

Select a signal region sample

Bayesian Inference to get Signal and Background features

Tune Montecarlo using background features in signal region!

Conclusions

- Unsupervised ML & Bayes provide a new way of using data
- Four-tops is a very suitable physics case
- Presented algorithm: As simple as providing all (N_j,N_b) pairs!
- Tune MC with Background in Signal region
- New ways to test for SM & NP @ four-tops
- Subtleties and details in upcoming arXiv:2106.XXXX. Also more to explore & understand.

Conclusions

- Unsupervised ML & Bayes provide a new way of using data
- Four-tops is a very suitable physics case
- Presented algorithm: As simple as providing all (N_j, N_b) pairs!
- Tune MC with Background in Signal region
- New ways to test for SM & NP @ four-tops
- Subtleties and details in upcoming arXiv:2106.XXXX. Also more to explore & understand.

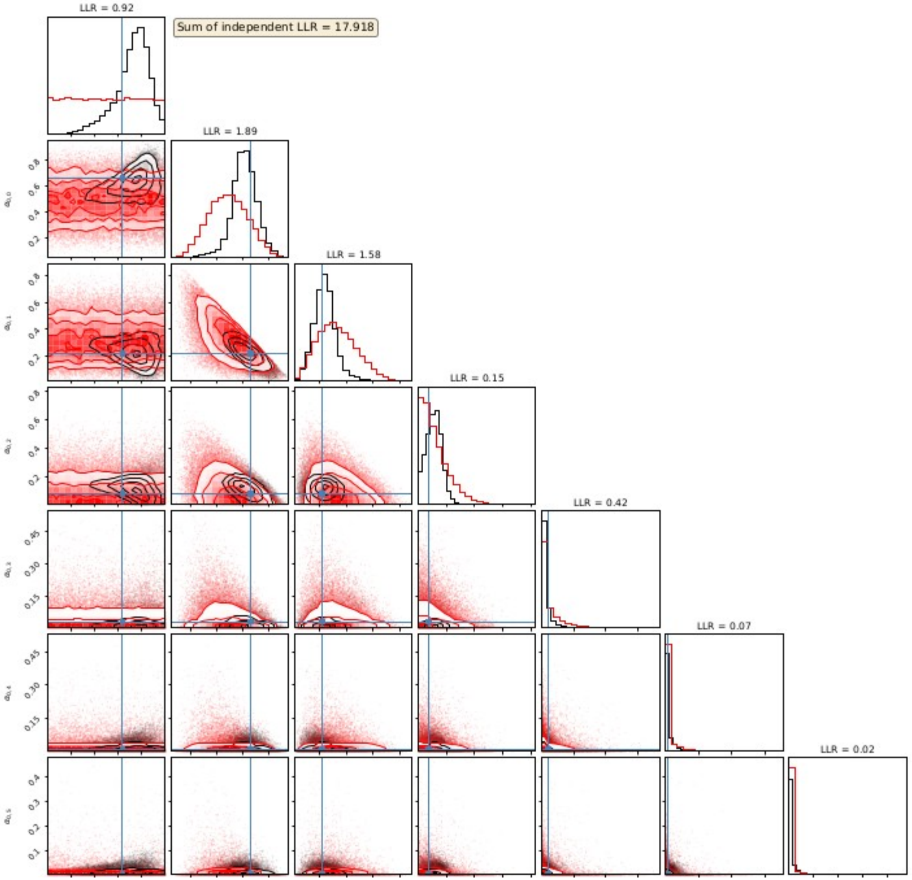
More discussion after this session:

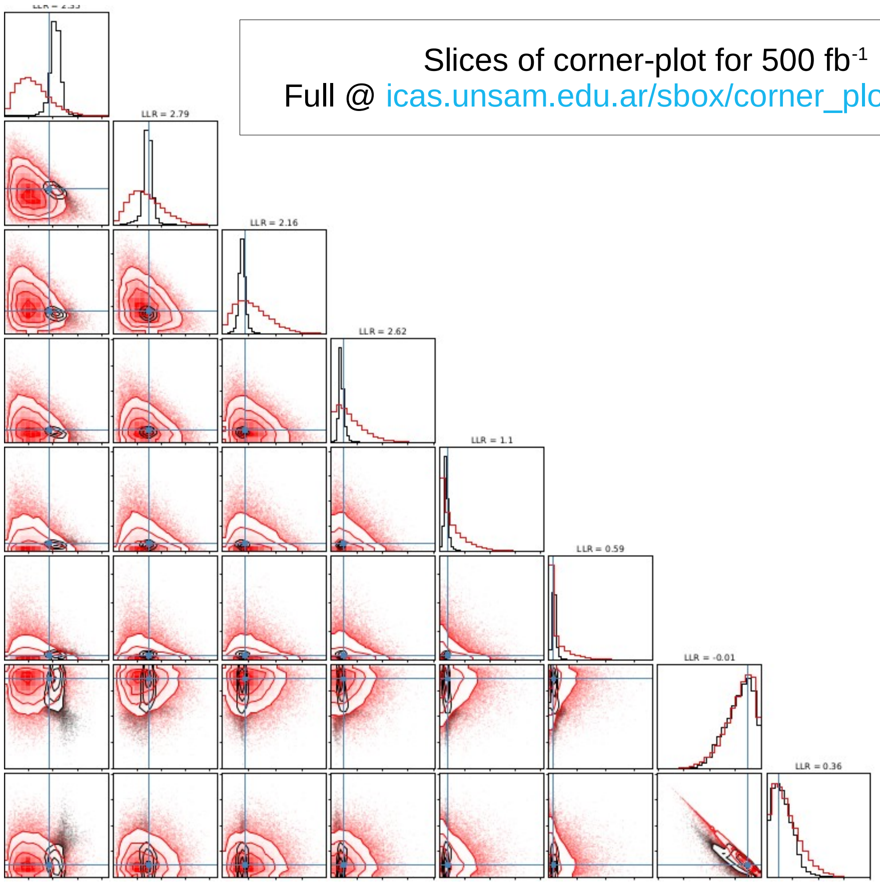
zoom.us/j/95915707476?pwd=NEdsekM5d05KVS9VZWdKOHEwY3l5QT09

pass: 46941614

sequi@unsam.edu.ar

Backup slides





Slices of corner-plot for 500 fb^{-1}
Full @ icas.unsam.edu.ar/sbox/corner_plot_bis.pdf

Backup slides

Four-tops
ATLAS-CONF-2021-013

7.2 Sequential kinematic reweighting

Following the flavour rescaling, a sequential reweighting is used to mitigate the kinematic mismodelling observed in $t\bar{t}$+jets MC. The reweighting corrects for the distributions of N_{jets}, the number of large-R jets ($N_{\text{LR-jets}}$), the scalar sum of all jet and lepton p_{T} in the event ($H_{\text{T}}^{\text{all}}$), and the average ΔR between any two jets ($\Delta R^{\text{ij}}_{\text{avg.}}$). These variables are related to the overall jet activities in the events and are observed to be mismodelled, especially the N_{jets} and $H_{\text{T}}^{\text{all}}$ spectra. These variables capture the most representative global kinematics of the events, as well as kinematic properties of the individual jets such as p_{T} and their angular distributions.

The $t\bar{t}$+jets events in $\geq 3b$ regions are reweighted according to the discrepancy between data and MC in the 2b regions. The reweighting factors are derived such that the overall MC prediction matches the data in the 2b regions. This is done based on the assumption that the deficiency of the radiation modelling in the parton shower is independent of the flavour of the radiated jets. Systematic variations on the $t\bar{t}$+jets modelling cover possible deviations from such assumption.

Backup slides

Four-tops
ATLAS-CONF-2021-013

7.1 $t\bar{t}$+jets flavour rescaling

The $t\bar{t}$+jets flavour rescaling corrects for the overall yields of $t\bar{t}$+light, $t\bar{t}+\geq 1c$ and $t\bar{t}+\geq 1b$ categories. The rescaling factors are derived from a dedicated profile likelihood fit to data using the event yields in the regions defined by various b-tagging requirements. Events with $\geq 8j$ in the 1L channel and $\geq 6j$ in the 2LOS channel are split into 2b, 3bL, 3bH and ≥4b regions, using the same criteria as defined in Table 1. The fit exploits the different $t\bar{t}$+jets flavour fractions in the eight fitted regions. The largest signal to background ratio in these regions is 2.5%, estimated using MC prior to the fit. Systematic uncertainties due to the tagging efficiency of b-jets and the mis-tag rate of c- and light-jets are considered as nuisance parameters in the profile likelihood fit. The measured rescaling factors for $t\bar{t}$+light, $t\bar{t}+\geq 1c$ and $t\bar{t}+\geq 1b$ are 1.0 ± 0.1, 1.6 ± 0.2 and 1.3 ± 0.1, respectively, where the quoted uncertainties are from the statistical uncertainty on data and from uncertainties on the b-tagging calibration.

Backup slides

ATLAS CONF Note

ATLAS-CONF-2019-045

16th October 2019

Minor revision: 24th August 2020

Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector

A search for the associated production of a top-quark pair with the Higgs boson ($t\bar{t}H$) in multilepton final states is presented. The search is based on a dataset of proton–proton collisions at $\sqrt{s}$ = 13 TeV recorded with the ATLAS detector at the CERN Large Hadron Collider and corresponding to an integrated luminosity of 80 fb^{-1}. Six final states, defined by the number and flavour of charged-lepton candidates, and 25 event categories are defined to simultaneously search for the $t\bar{t}H$ signal and constrain several leading backgrounds. The $t\bar{t}W$ background normalisation is left unconstrained in the statistical analysis and the resulting $t\bar{t}W$ normalisation is found to be higher than the theoretical prediction. An excess of events consistent with $t\bar{t}H$ production, over the expected background from Standard Model processes, is found with an observed significance of 1.8 standard deviations, compared to an expectation of 3.1 standard deviations. Assuming Standard Model branching fractions, the best-fit value of the $t\bar{t}H$ production cross section is $\sigma_{t\bar{t}H} = 294^{+182}_{-162}$ fb, which is consistent with the Standard Model prediction. The impact on the $t\bar{t}H$ cross section measurement of the assumptions made on the $t\bar{t}W$ background modelling is discussed.

Four-tops

ATLAS-CONF-2019-045

Backup slides

Eur. Phys. J. C (2016) 76:11
DOI 10.1140/epjc/s10052-015-3852-4

THE EUROPEAN PHYSICAL JOURNAL C

CrossMark

Regular Article - Experimental Physics

Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector

ATLAS Collaboration*

malised to the NNLO+NNLL result [32–37]. PYTHIA 8 offers several options for modelling $g \to b\bar{b}$ splittings in the final-state parton showers, which may be accessed by varying the TIMESHOWER:WEIGHTGLUONTOQUARK (wgtq) parameter [75]. Differences between the models arise by neglecting (wgtq5) or retaining (wgtq3, wgtq6) the mass-dependent terms in the $g \to b\bar{b}$ splitting kernels. Differences also arise with respect to the treatment of the high-$m_{b\bar{b}}$ region, with specific models giving an enhanced or suppressed $g \to b\bar{b}$ rate. The model corresponding to wgtq3 was chosen to maximise this rate. Finally, some of the models (wgtq5, wgtq6) offer the possibility to choose sgtq$\cdot m_{b\bar{b}}$ instead of the transverse momentum as the argument of α_S in the $g \to b\bar{b}$ vertices. Here sgtq refers to the TIMESHOWER:SCALEGLUONTOQUARK parameter, and is allowed to vary in the range $0.25 \leq$ sgtq ≤ 1, with larger values giving a smaller $g \to b\bar{b}$ rate and vice versa. For the model wgtq5, sgtq was set to 1, a combination that minimises the $g \to b\bar{b}$ rate, while for wgtq6, sgtq was set to 0.25.

(see discussion in 1701.04427)

Backup slides

Eur. Phys. J. C (2016) 76:379
DOI 10.1140/epjc/s10052-016-4105-x

THE EUROPEAN PHYSICAL JOURNAL C

CrossMark

Regular Article - Experimental Physics

Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV

CMS Collaboration*

11. CMS Collaboration, Measurement of the cross section ratio $\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$ in pp collisions at $\sqrt{s} = 8$ TeV. Phys. Lett. B **746**, 132 (2015). doi:10.1016/j.physletb.2015.04.060. arXiv:1411.5621

PYTHIA6 and HERWIG6. The normalization factors applied to the MADGRAPH and POWHEG predictions are found to be about 1.3 for results related to the leading additional b jet. The predictions from both generators underestimate the $t\bar{t}b\bar{b}$ cross sections by a factor 1.8, in agreement with the results from Ref. [11]. The normalization factors applied to MC@NLO are approximately 2 and 4 for the leading and subleading additional b jet quantities, respectively, reflecting the observation that the generator does not simulate sufficiently large jet multiplicities. All the predictions have slightly harder p_T spectra for the leading additional b jet than the data, while they describe the behaviour of the $|\eta|$ and m_{bb} distributions within the current precision. The predictions favour smaller ΔR_{bb} values than the measurement, although the differences are in general within two standard deviations of the total uncertainty.

(see discussion in 1701.04427)