NORMALIZING FLOW FOR DATA-DRIVEN BACKGROUND ESTIMATION

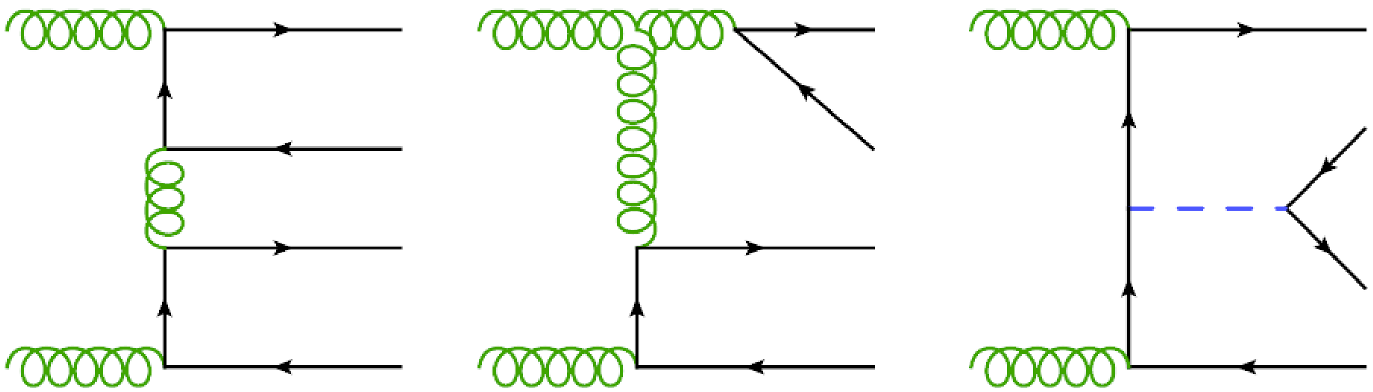
Suyong Choi
Korea University

CONTENTS

FOUR TOP QUARKS IN SM

Production

- Strong $O(\alpha_s^4)$
- Partially electroweak $O(\alpha_s^2 y_t^4)$, $O(\alpha_s^2 \alpha^2)$



Production cross sections

- At $\sqrt{s} = 13$ TeV pp: 9.7 fb (LO) 12.2 (NLO)
- At $\sqrt{s} = 14$ TeV pp: 17 fb (NLO)

BACKGROUNDS

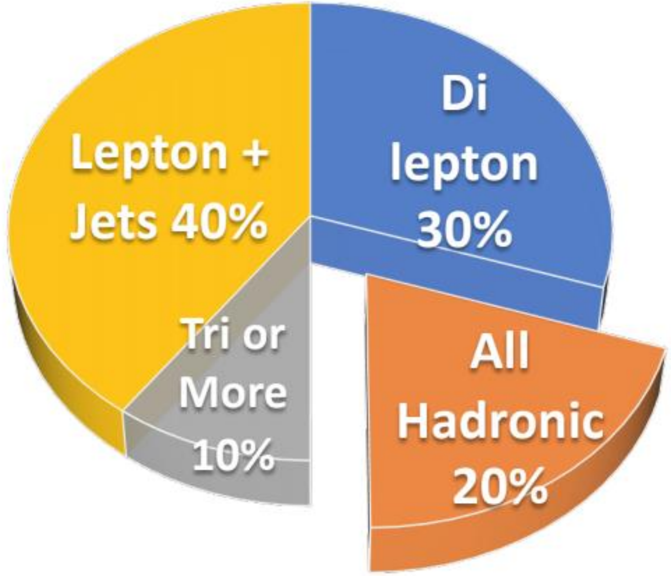
Backgrounds to $t\bar{t}t\bar{t}$ in the all-hadronic channel

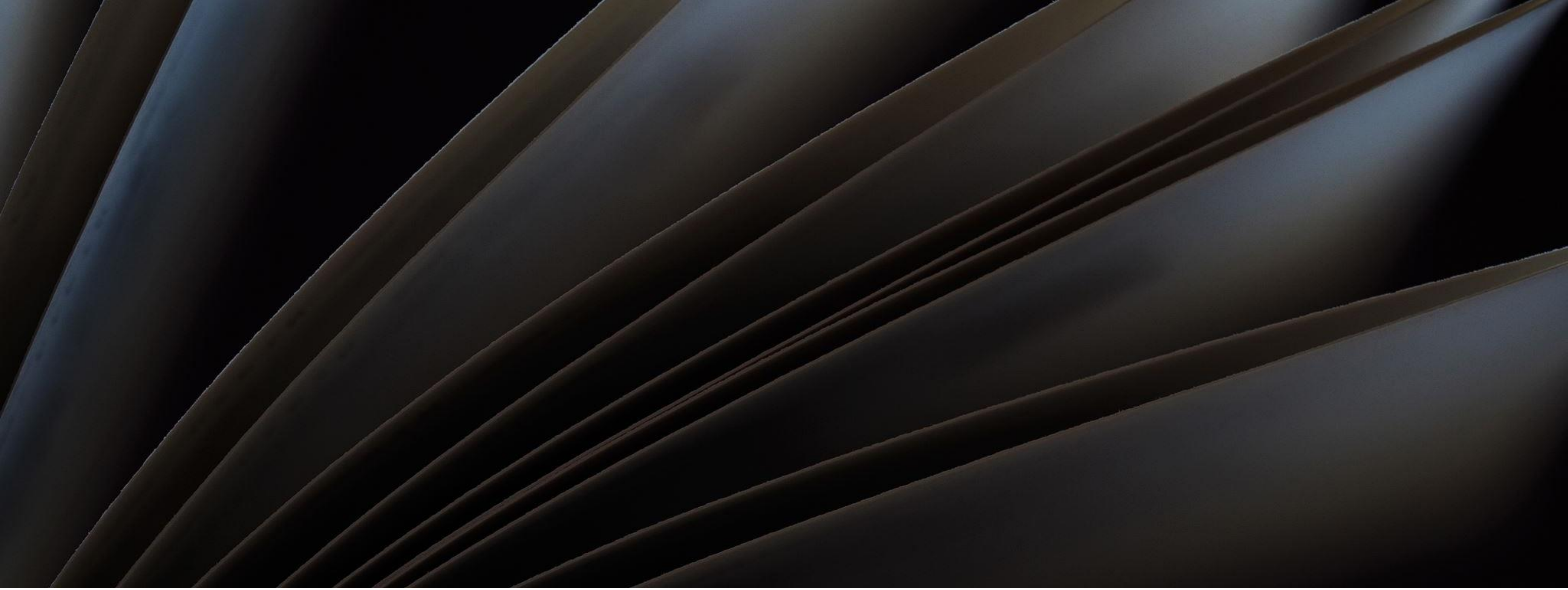
- Overwhelming $t\bar{t}+\text{jets}$ and QCD multijets

Challenges

- $t\bar{t}+jets$: reliability of modeling
- QCD : LO only. Cannot generate enough statistics.
- Cannot fully rely on MC simulations

Data-driven background estimation for normalization and shape





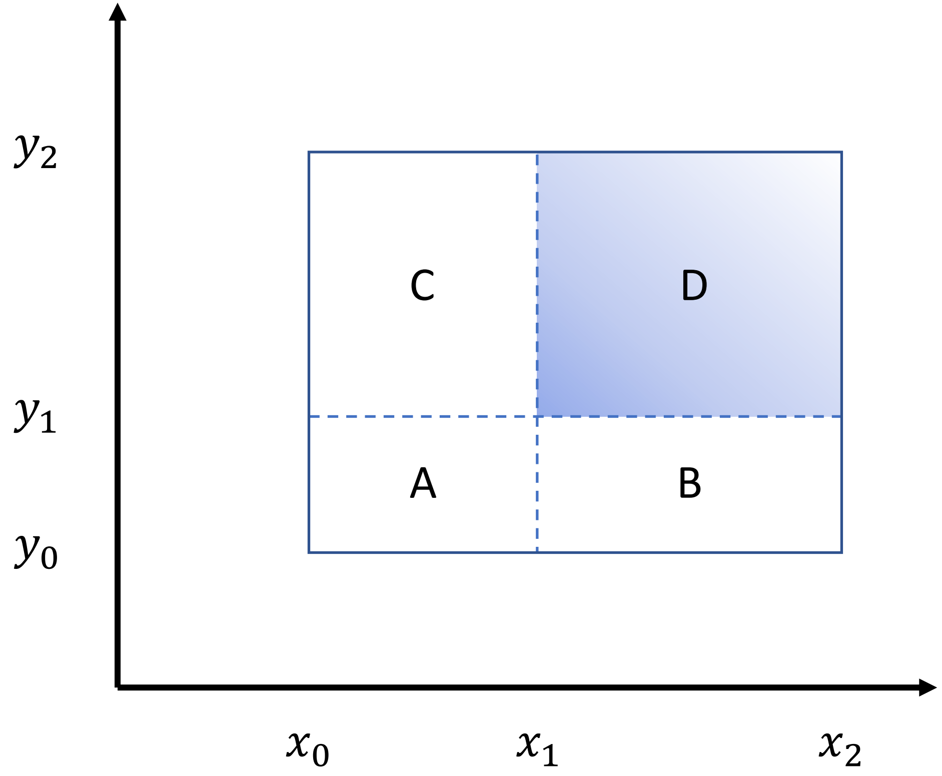
EXTENDED ABCD

ABCD METHOD

Estimate # of events in D from control regions through extrapolation

$$\hat{N}_D = \frac{N_C}{N_A} N_B$$

- Assumes independence of x and y



EXTENDED ABCD METHODS

Extended ABCD method with additional control regions A' and B'

ABCD method

$$\hat{N}_D = \frac{N_{A'}}{N_{C'}} \frac{N_C}{N_A} \cdot \frac{N_C}{N_A} N_B$$

- Weak correlations between x and y can be considered, unlike the original ABCD method
  - *Eur. Phys. J. C* **81**, 643 (2021), https://arxiv.org/abs/1906.10831

ABCD method-like NN's possible – case dependent

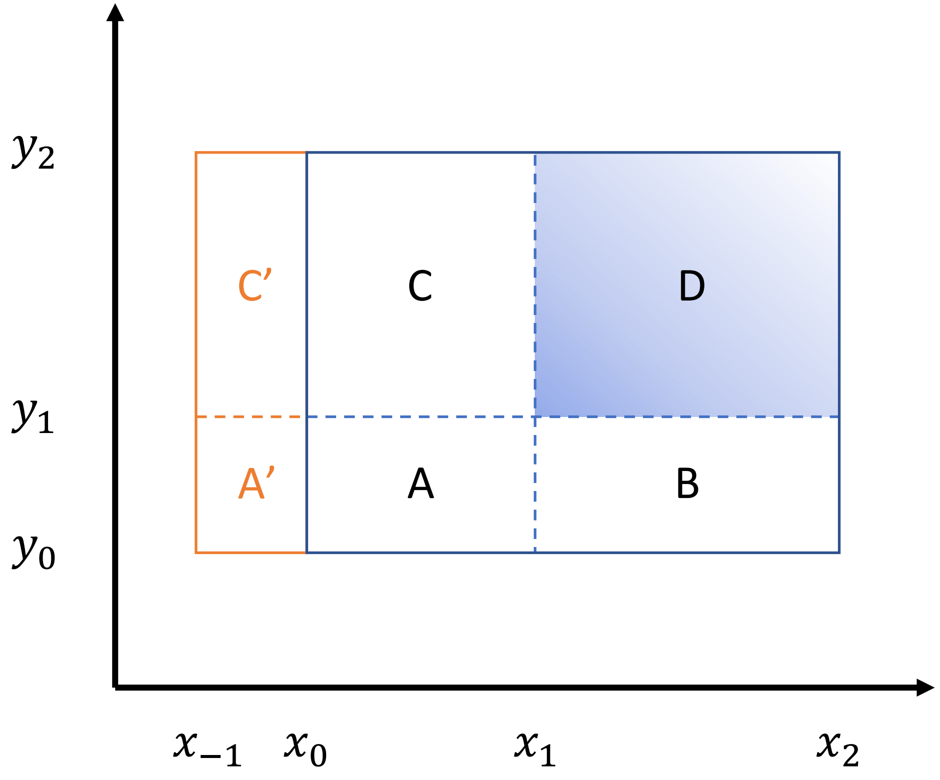


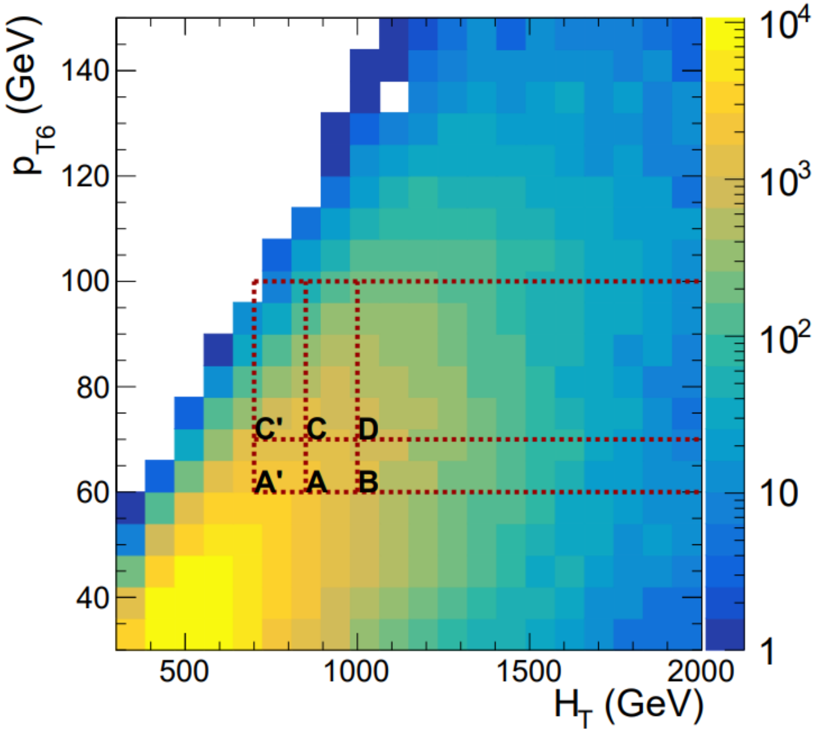
ILLUSTRATION OF THE EXTENDED ABCD METHOD

$t\bar{t} + jj$

- Generated with MG5 and Delphes
- 7 or more jets, 2 or more b-jets
- 6th jet p_T vs H_T
  - Correlation coefficient: 0.66

ABCD	Ext. ABCD	Truth
4802 ± 122	9976 ± 488	9288

Extended ABCD prediction closer to the truth, but larger statistical uncertainty



p_{T6} (GeV)	H_T (GeV) 700 – 850	850 – 1000	> 1000
60 – 70	6319 (A')	4479 (A)	4343 (B)
70 – 100	3364 (C')	4953 (C)	9288 (D)

NORMALIZING FLOW

NORMALIZING FLOW (NF)

Probability density estimation for multi-dimensional distributions

$$P_e(\vec{y}) = \left|\frac{d\vec{x}}{d\vec{y}}\right| P_h(\vec{x})$$

Jacobian

- Transform hard-to-evaluate PDF (P_h) to an easy PDF (P_e: usually normal distr. – hence "normalizing")
- Transform with **invertible** bijective function $\vec{y} = g(\vec{x})$: $\mathbb{R}^n \to \mathbb{R}^n$
- Efficient evaluation of Jacobian desired

g learned by minimizing Kullback-Leibler (KL) divergence between desired target P_e and $\tilde{P}_e$ obtained

USING NORMALIZING FLOW

1. Generator
   - Generate $\vec{y}$ to follow normal distribution
   - $\vec{x} = g^{-1}(\vec{y})$: follows P_h
   - Need to evaluate g^{-1}

2. Probability density estimation

$$P_h(\vec{x}) = P_e(g(\vec{x}))\left|\frac{dg(\vec{x})}{d\vec{x}}\right|$$

   - Look for rare events

GENERAL BIJECTIVE FUNCTION

General n-dim bijective function with n number of 1-D invertible functions

$$(y_1, \dots, y_n) = \vec{g}(x_1, \dots, x_n | \vec{c})$$
$$y_1 = g_1(x_1 | \{\vec{c}\})$$
$$y_2 = g_2(x_2 | \{x_1, \vec{c}\})$$
$$y_3 = g_3(x_3 | \{x_1, x_2, \vec{c}\})$$
$$\dots$$

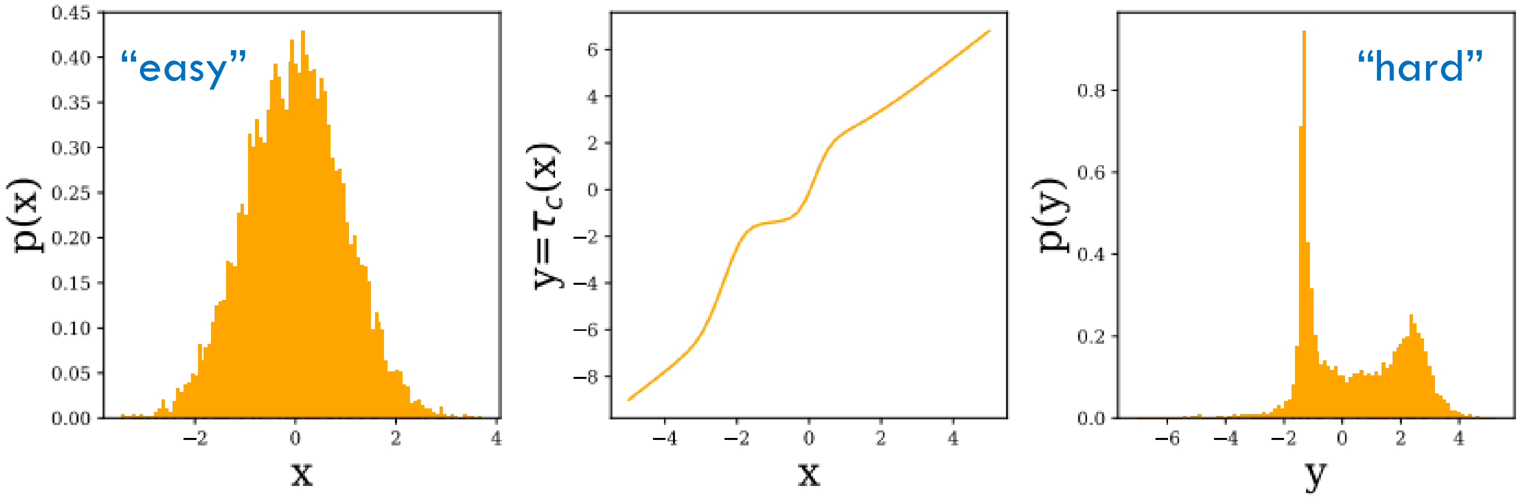
- $g_i(x_i | \{x_1 \dots x_{i-1}, \vec{c}\})$: monotonic function of x_i

Autoregressive

- Jacobian calculation $\sim O(n)$
- g^{-1} calculation feasible

NEURAL AUTOREGRESSIVE FLOW (NAF)

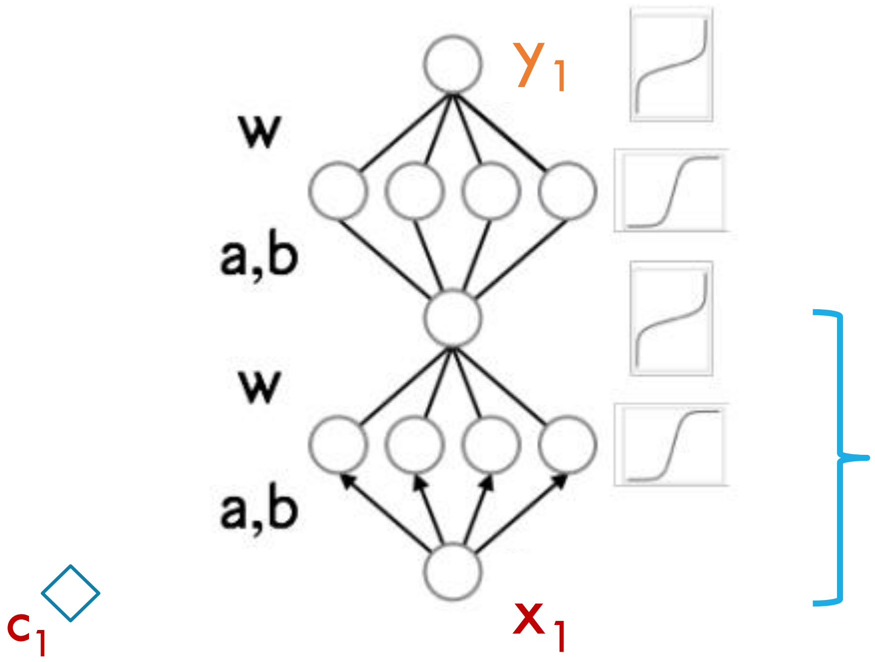
Neural autoregressive flow (arXiv:1804.00779)



Invertible transformation is unique

NEURAL AUTOREGRESSIVE FLOW

Neural autoregressive flow (arXiv:1804.00779), applying sigmoid and inverse sigmoidal allows for arbitrary transformation



Deep Sigmoidal Flow (DSF) Architecture

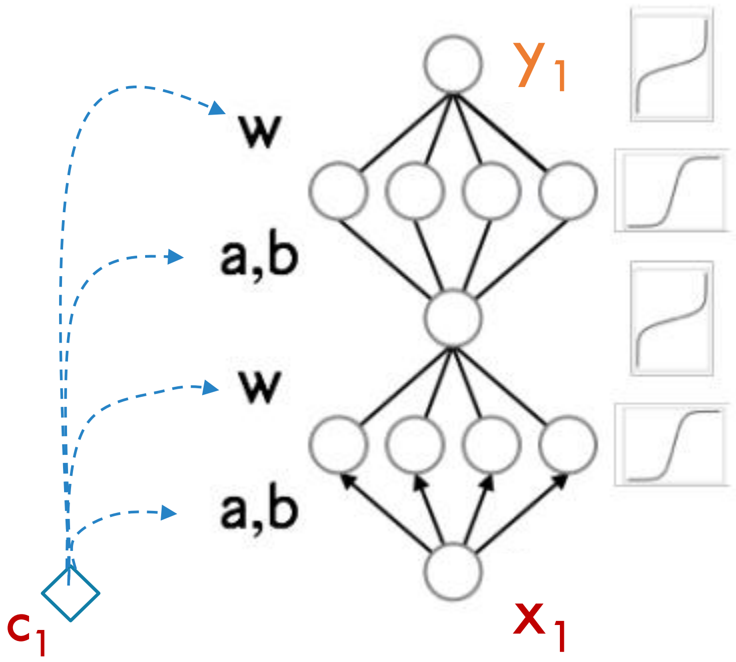
Sigmoid function

$$\sigma^{-1}(\underbrace{w^T}_{1\times d} \cdot \sigma(\underbrace{a}_{d\times 1} \cdot \underbrace{x_t}_{1\times 1} + \underbrace{b}_{d\times 1})))$$

For d hidden units

NEURAL AUTOREGRESSIVE FLOW

For an invertible transformation, the weight and scaling values should be constrained



Deep Sigmoidal Flow Architecture

Weight values a, b, w determined by another DNN with $\vec{c}$ as input

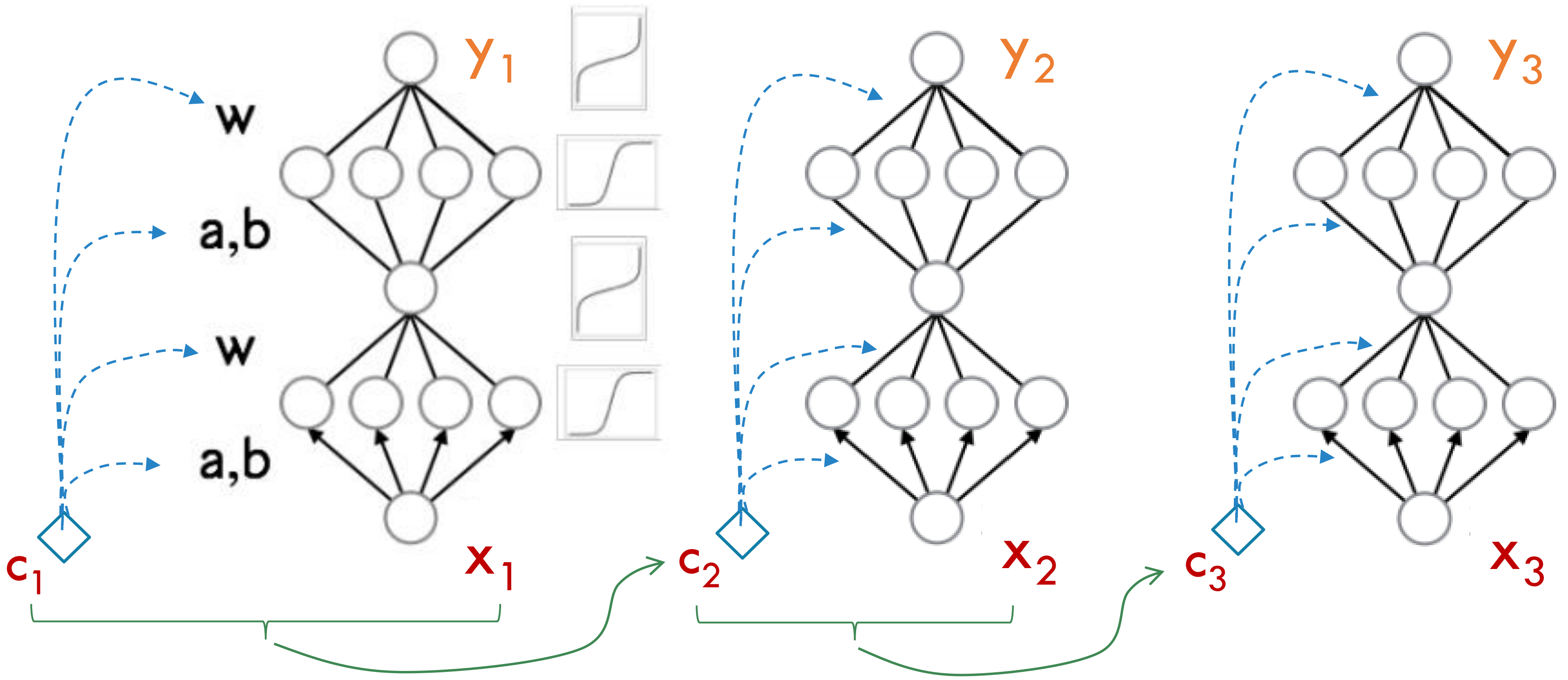
$$0 < w_{i,j} < 1, \sum_i w_{i,j} = 1, a_{s,t} > 0$$:ensures monotonicity

$$\sigma^{-1}(\underbrace{w^T}_{1\times d} \cdot \sigma(\underbrace{a}_{d\times 1} \cdot \underbrace{x_t}_{1\times 1} + \underbrace{b}_{d\times 1})))$$

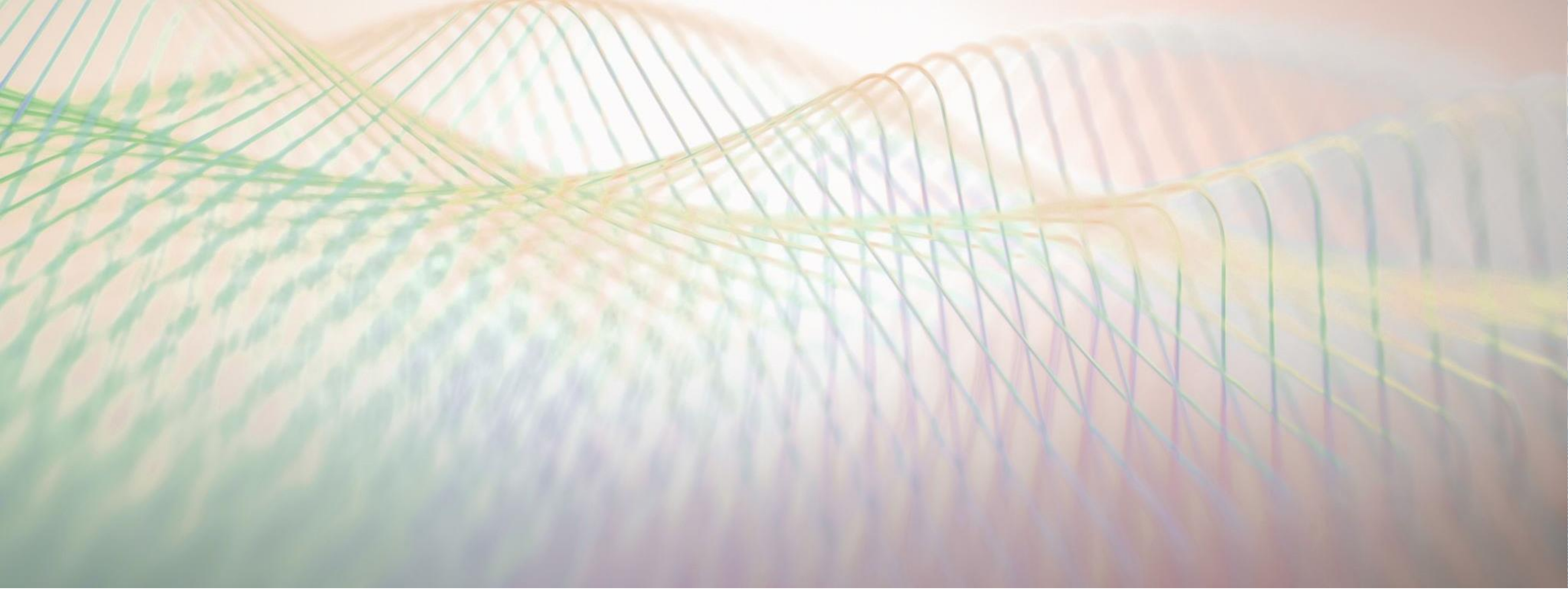
For d hidden units

NEURAL AUTOREGRESSIVE FLOW

To construct a n-dimensional bijective function, chain the DSF networks into a neural autoregressive flow architecture



$$y_3 = g_3(x_3|\{x_1, x_2, \vec{c}\})$$



ABCDNN

INTRODUCTION

Background shape prediction

- Simultaneous reweighting of events?
- Extended ABCD method or morphing techniques reweighting for one variable at a time

NF for shape prediction

- Find transformation to go from CR to SR

ABCDnn - https://arxiv.org/abs/2008.03636

CONDITIONAL FLOW

Transform distributions

$$\int \mathcal{T}(\vec{x}, \vec{x}_0|\vec{c}) f_{src}(\vec{x}_0)\, d\vec{x}_0 = f_{target}(\vec{x}|\vec{c})$$

- $\mathcal{T}$ dependent on conditional variable $\vec{c}$
- f_{src} and f_{target} not formally known – cannot use D_{KL}

Loss function – Maximum mean discrepancy (MMD, arXiv:1705.08584)

- Measure of similarity of two distributions estimated from finite samples.
- Positive definite. 0 if and only if two samples have the same distribution in the limit of infinite statistics.
- Gaussian kernels involved

TRAINING ABCDNN

Fully supervised

- Sample minibatch from the source
- Select randomly the region (as encoded by $\vec{c}$) and sample minibatch from target in that region
- Minimize MMD

ILLUSTRATION OF ABCDNN METHOD

All MC in CR combined – source

MC target

Transformed source

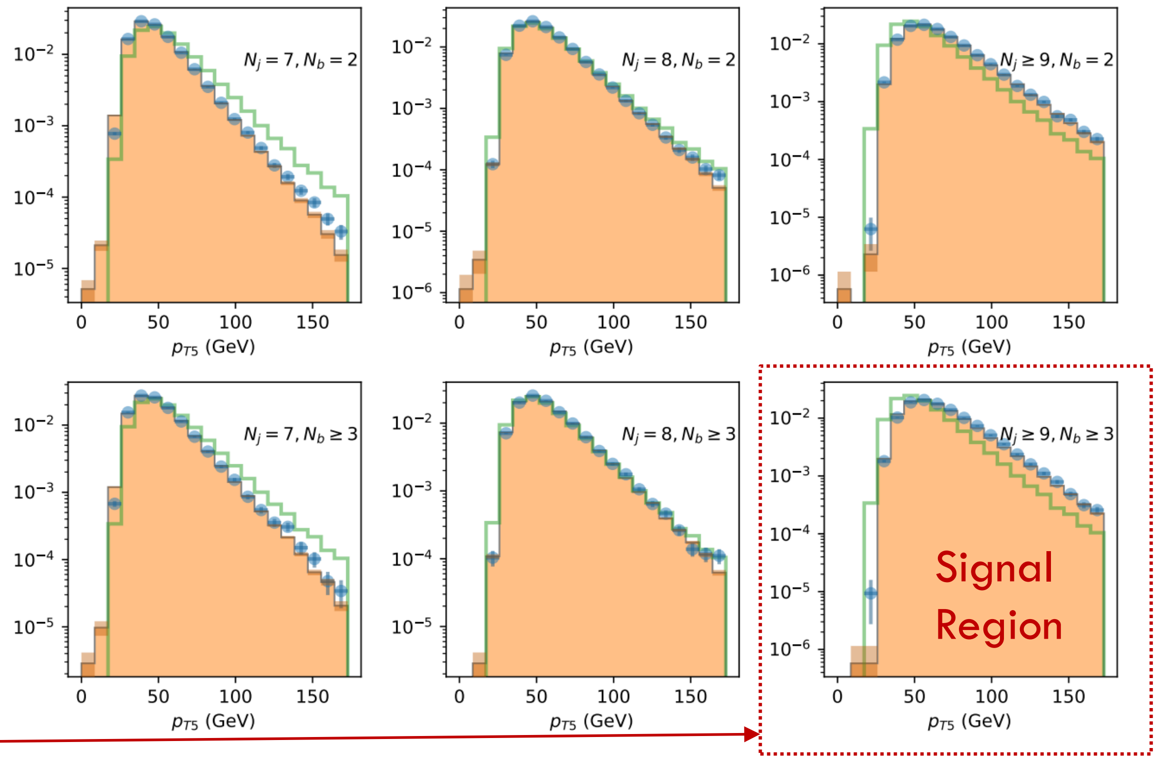
$t\bar{t}jj$ MG5+Delphes

- Control region is one-hot encoded - $\vec{c}$

	$N_j = 7$	$N_j = 8$	$N_j \geq 9$
$N_b = 2$	1 0 0 1 0	0 1 0 1 0	0 0 1 1 0
$N_b \geq 3$	1 0 0 0 1	0 1 0 0 1	0 0 1 0 1

- 5 CR data combined into source
- NAF learned to transform all MC in CR to each CR

The data in SR has not been used in training but when the region encoding condition 00101 is presented together with the source sample (from CRs)



Code in https://github.com/suyong-choi/ABCDnn

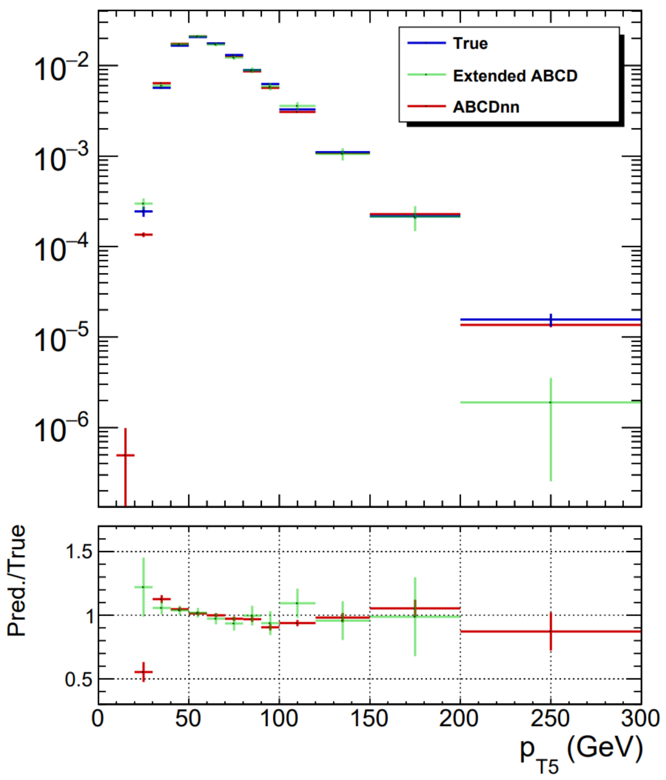
ABCDNN AND EXT. ABCD

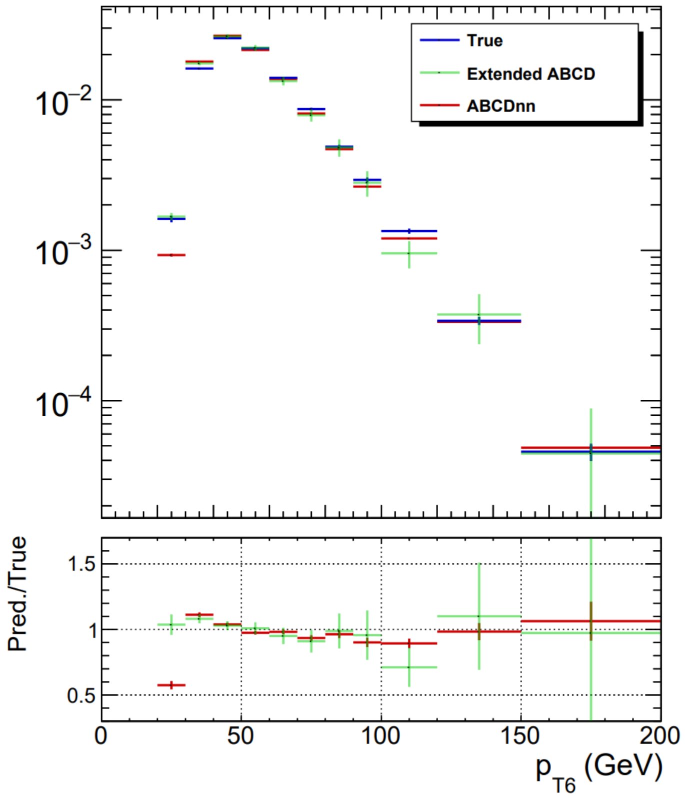
Comparisons

	ABCDnn	Ext. ABCD
Absolute normalization	✗	✓
Shape	Multiple variables	1 variable at a time
Stat. error	☺	

- ABCDnn can be used for shape prediction even in cases when ext. ABCD derivations are not available

Normalized distr.





HADRONIC FOUR TOP SEARCH

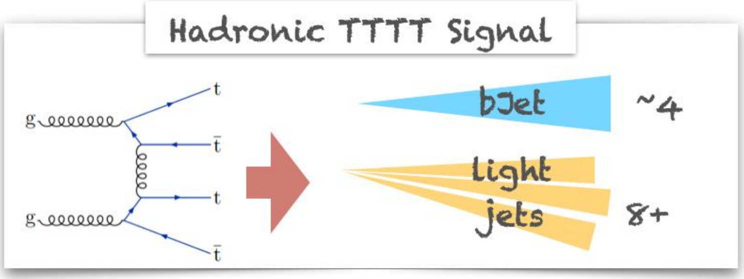
FOUR TOP SEARCH IN ALL HADRONIC CHANNEL AT CMS

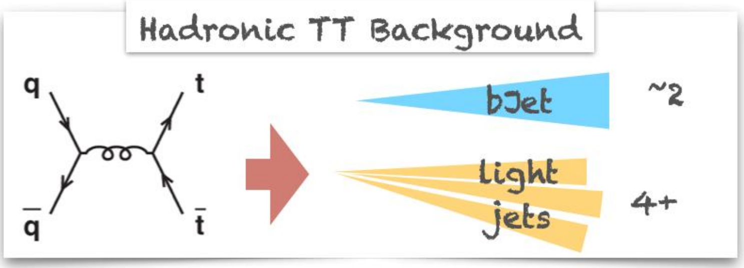
Event selection

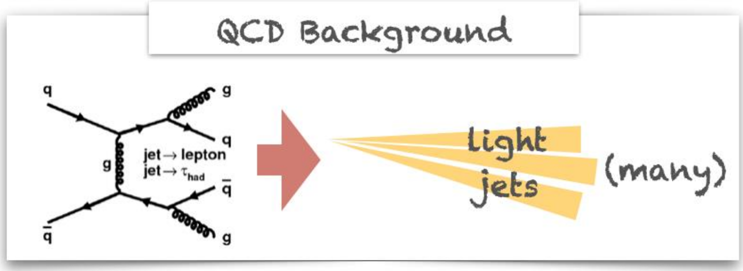
- 0 lepton
- $N_j \geq 7$, $N_b \geq 2$, $H_T > 700$ GeV
- Resolved top (RT) tagger, boosted top tagger
- $N_{RT} \geq 1$
- Main backgrounds: $t\bar{t} + jets$, multijets ($\mathrm{S/B}{\sim}10^{-5}$)

Analysis

- BDT for signal/background discrimination
- Categorize into different N_{RT}, N_{BT}, H_T regions
- CR's used to estimate background # of events and the BDT shape in the SR







EVENT SELECTION

Signal regions: baseline & $N_j \geq 9, N_b \geq 3, \mathrm{N_{RT}} \geq 1$

- Bins of different S/B

		HT bins (GeV)								
Top Bins	$N_{RT}=1$ $N_{BT}=0$	700≤ HT<800	800≤ HT<900	900≤ HT<1000	1000≤ HT<1100	1100≤ HT<1200	1200≤ HT<1300	1300≤ HT<1500	HT≥1500	SR1
	$N_{RT}=1$ $N_{BT}\geq1$	700≤ HT<1400				HT≥1400				SR2
	$N_{RT}\geq2$ $N_{BT}\geq0$	700≤HT<1100				HT≥1100				SR3

BACKGROUND PREDICTION

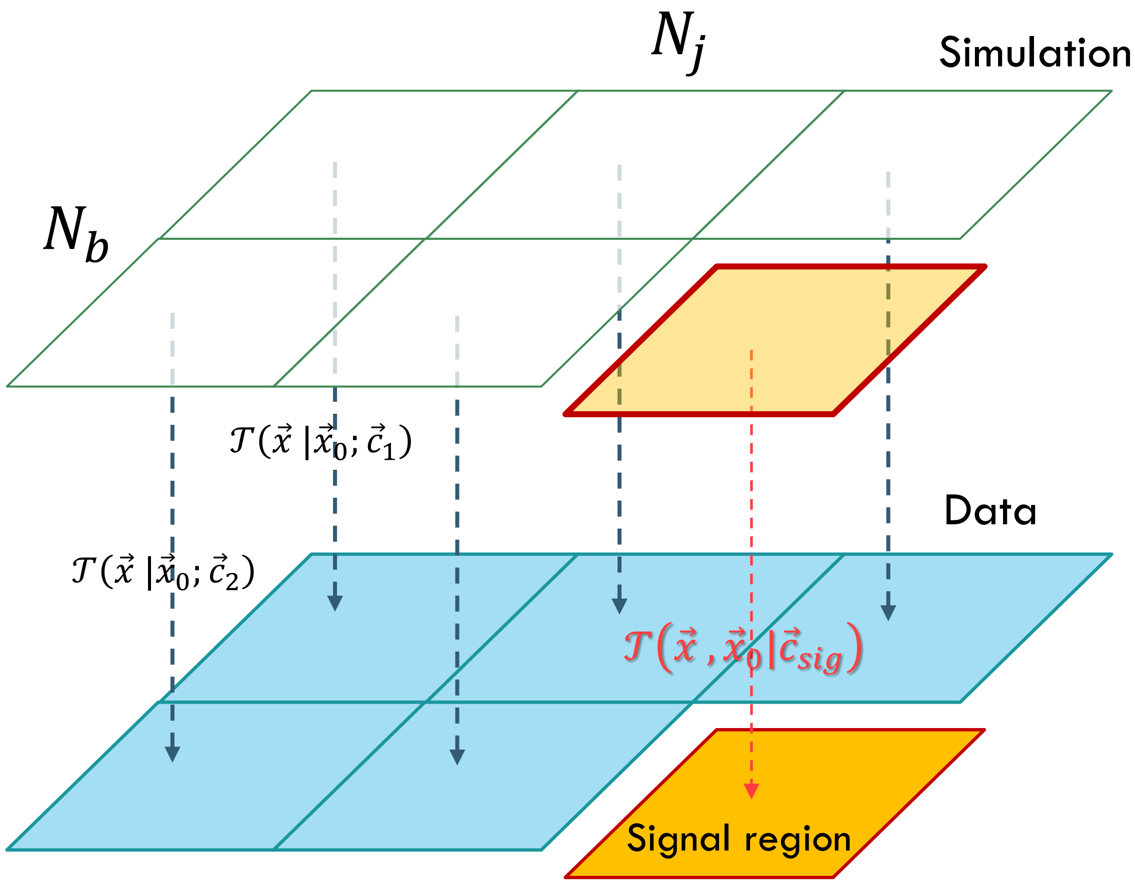
Absolute normalization from ext. ABCD

Shape from

$$\hat{f}_{Data}(\vec{x}|\vec{c}_i) = \int \mathcal{T}(\vec{x},\vec{x}_0|\vec{c}_i) f_{MC}(\vec{x}_0|\vec{c}_i)\, d\vec{x}_0$$

- $t\bar{t}$ inclusive MC
- Feature variables: BDT, H_T
- Target: data – minor bkgs

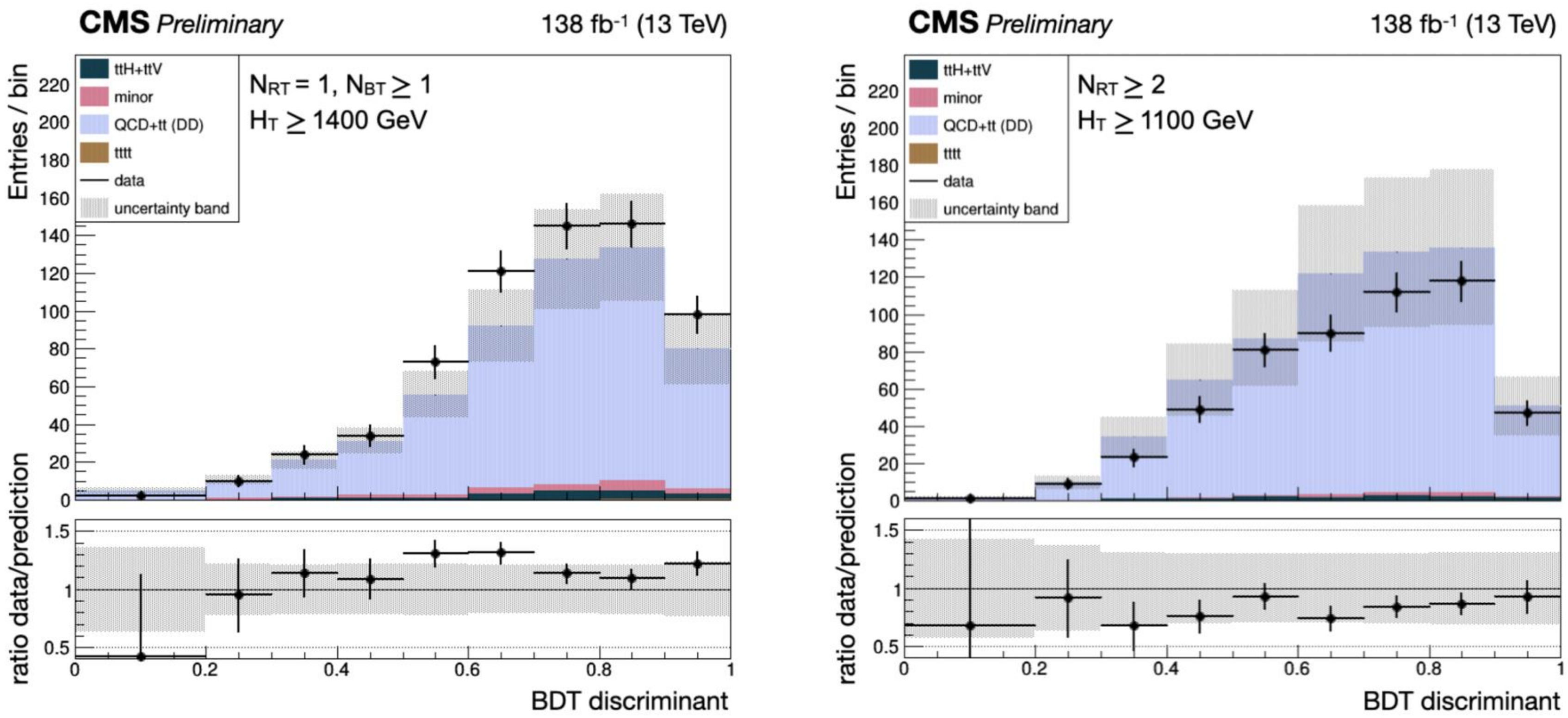
Tests of method with various MC's



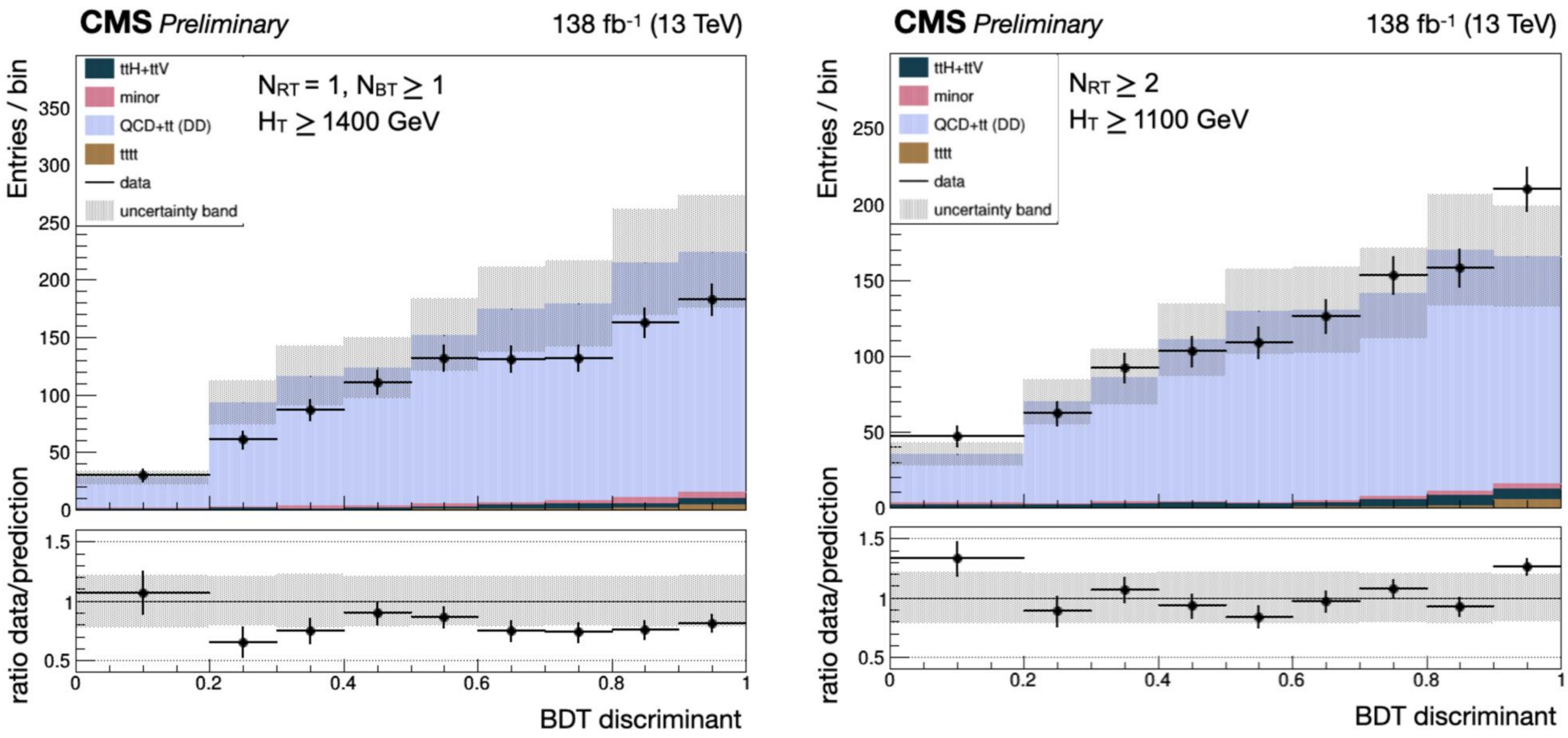
VALIDATION OF THE BACKGROUND ESTIMATION METHOD

Validation regions ($N_j = 8$)

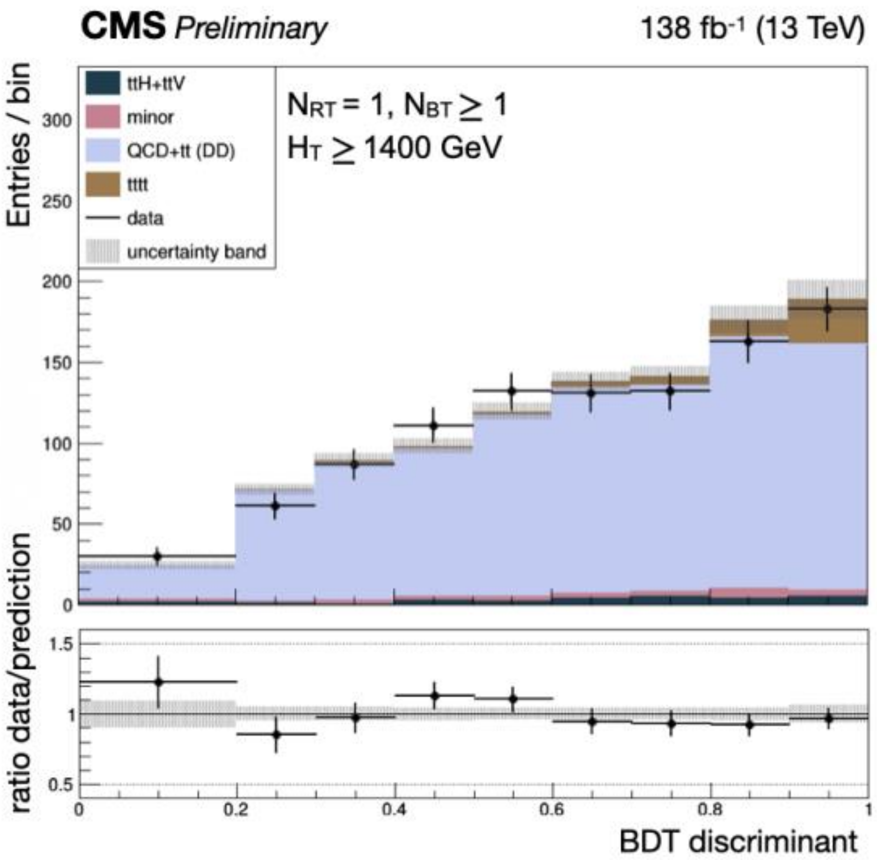
Normalization and shape uncertainties

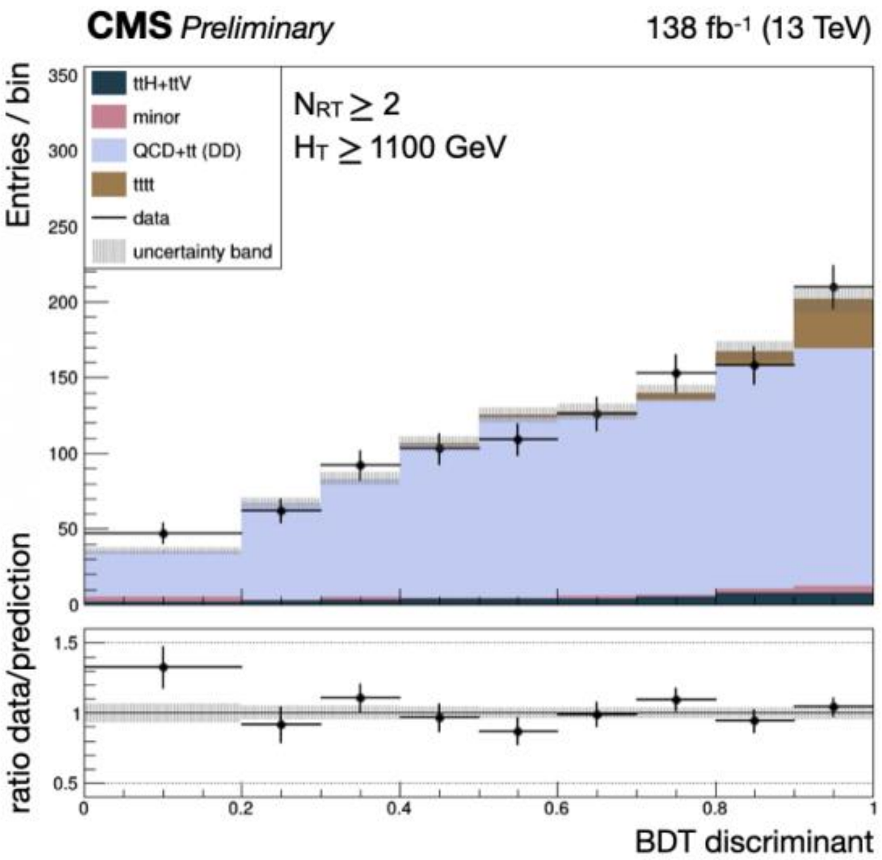


ALL HADRONIC PRE-FIT



ALL HADRONIC POST-FIT

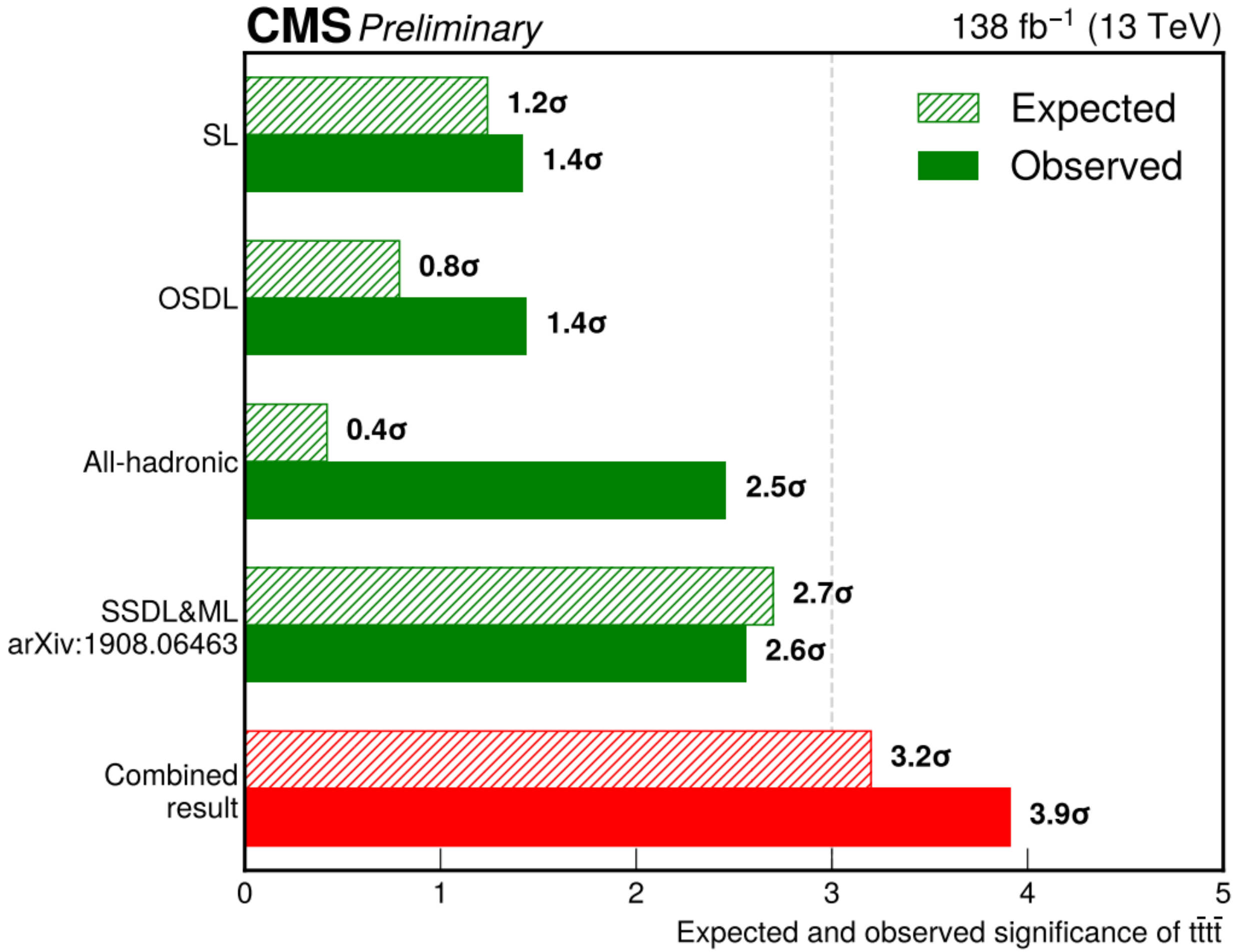




SIGNIFICANCE

Sensitivity dominated by

- Statistics
- $t\bar{t}H$ theoretical cross section
- $t\bar{t}b\bar{b}$ x-section and modeling



RESULTS

Analysis	Signal strength		Significance		Cross section
	exp	obs	exp	obs	obs (fb)
OSDL	$1^{+1.6}_{-1.6}$	$2.8^{+1.8}_{-1.6}$	0.6	1.8	37^{+21}_{-20}
Single lepton	$1^{+0.9}_{-0.8}$	$1.2^{+0.9}_{-0.9}$	1.2	1.4	15^{+13}_{-11}
All-hadronic	$1^{+2.5}_{-2.4}$	$5.8^{+2.5}_{-2.4}$	0.4	2.5	70^{+30}_{-29}
New results only	$1^{+0.7}_{-0.7}$	$2.5^{+0.7}_{-0.7}$	1.5	3.7	38^{+13}_{-11}
SSDL & multilepton [26]	$1^{+0.4}_{-0.4}$	$1.0^{+0.5}_{-0.4}$	2.7	2.6	13^{+6}_{-5}
All CMS 2016–2018	$1^{+0.4}_{-0.3}$	$1.4^{+0.4}_{-0.4}$	3.2	3.9	17^{+5}_{-5}

SUMMARY

Normalizing flow provides an elegant way to transform distributions

Methods of background estimation with extrapolation

Indispensable in the search for $t\bar{t}t\bar{t}$ in all-hadronic channel at CMS