Search for single production of vector-like B quarks decaying to a top quark and a W boson in semi-leptonic final states

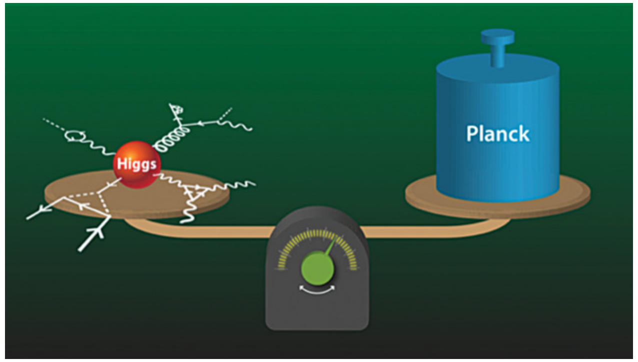
Xiaohe Shen (Brown), Julie Hogan (Bethel)

DPF 2024

May 16, 2024

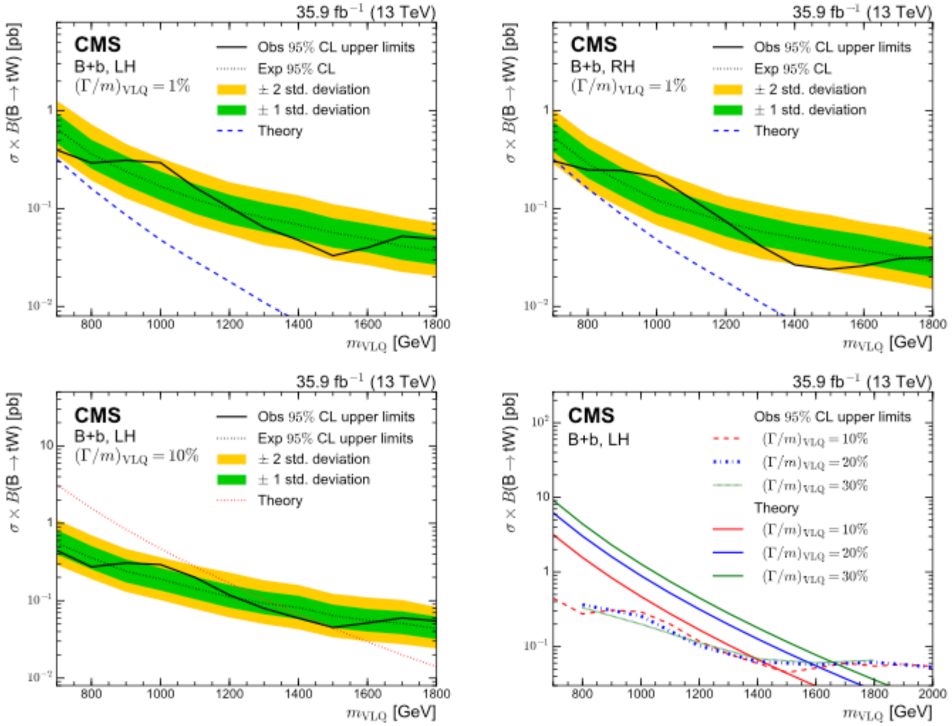
Vector-like quarks

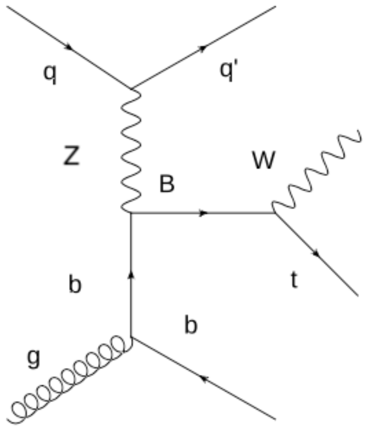
- Left-handed components and right-handed components transform the same way under the SM electroweak symmetry (the weak interaction couples to the vector current)
- Hierarchy Problem
- Mass: $\mathcal{O}(1\ \text{TeV})$
- Not constrained by current Higgs boson cross section measurements



Previous search at CMS

- Search has been done at CMS with 2016 data
- Single production: 2016 VLQ->tW search (arXiv:1809.08597)
  - Decay width 1, 10, 20, 30%
  - Semi-leptonic final states



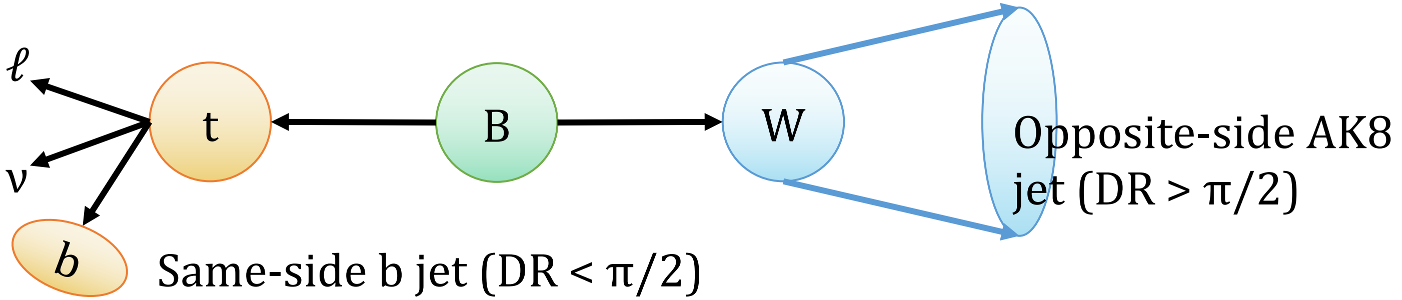


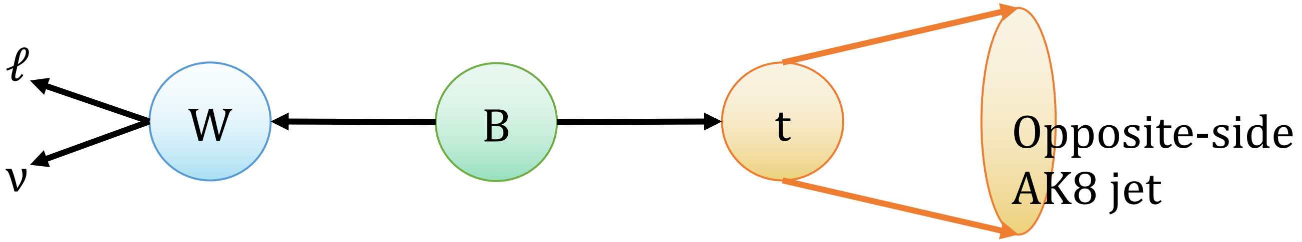
- Mass exclusion limit: up to 1660 GeV for large decay widths
- Not enough sensitivity for narrow decay width

Analysis strategy

$B \rightarrow tW: 1l$

- Select 1 lepton + MET + 1 "opposite-side" fatjet
- Build W from lepton + MET, then study the jets involved:





Reconstruction divided into 4 "cases":

- Leptonic W:
  - Top-tagged jet (case 1)
  - Not top-tagged (case 4)
- Leptonic t:
  - W-tagged jet (case 2)
  - Not W-tagged (case 3)
- Build B 4-vector from fatjet + leptonic particle

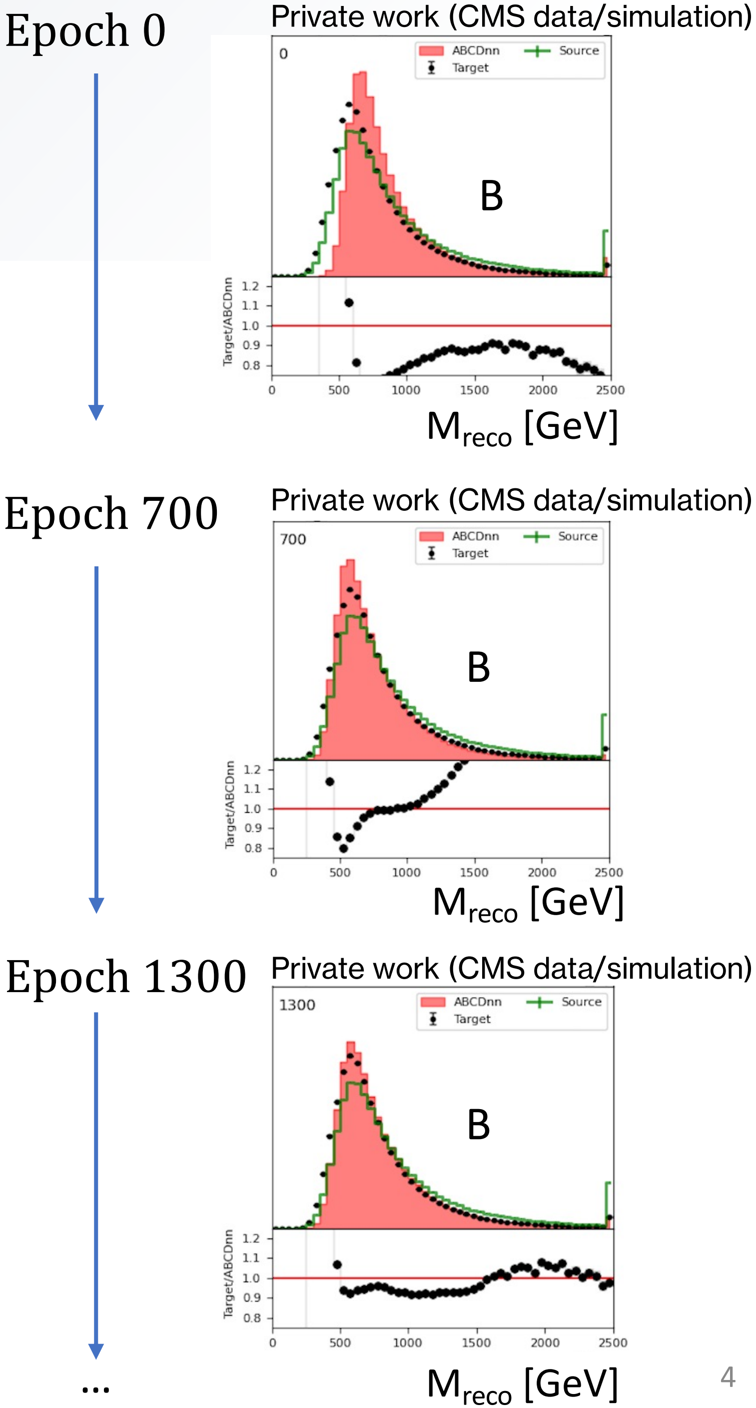
Background estimation: Overview

- Yield estimate + shape modelling
- Shape modelling: ABCDnn method
  - General data-driven shape estimation method
  - Learns the shape of the background through a DNN
  - Key components:

    2 control variables + 2 transform variables
  - 6 regions defined:

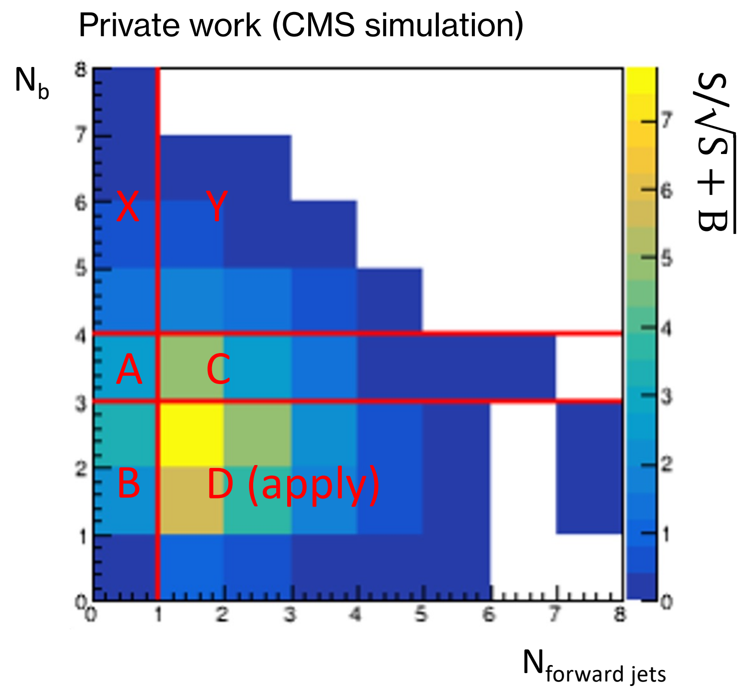
    5 control region (training) + 1 "signal" region (application)
  - Aim for reduced uncertainty over MC modelling

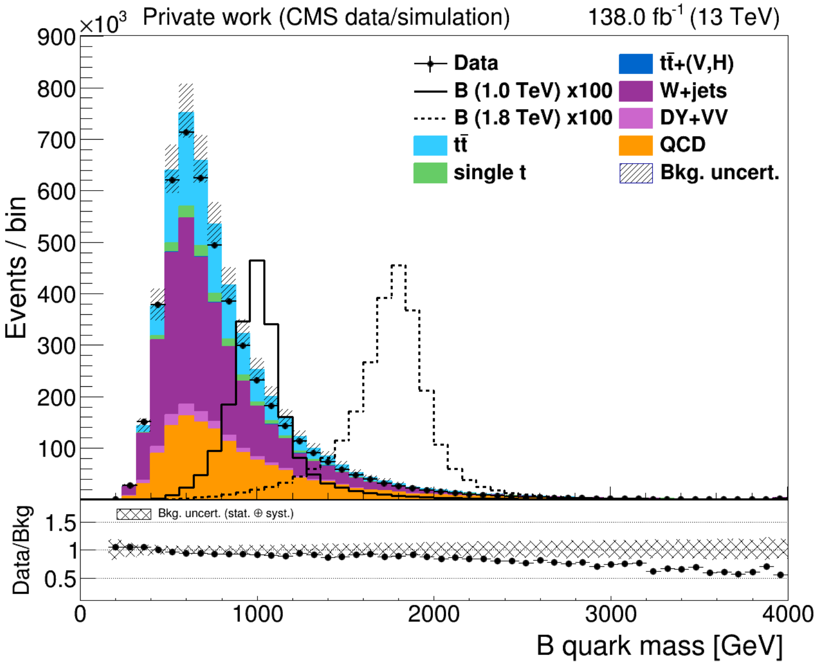
X	Y
A	C
B	D

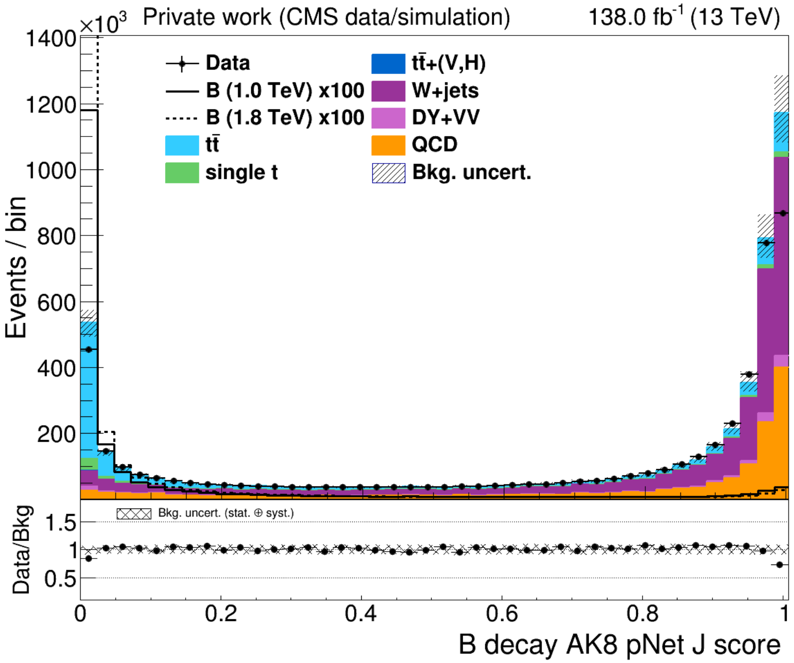


Background estimation: ABCDnn method

- Control variables: $N_{\text{forward jets}}$, N_b
- Transform variables: B mass, pNetJ score of leading fatjet

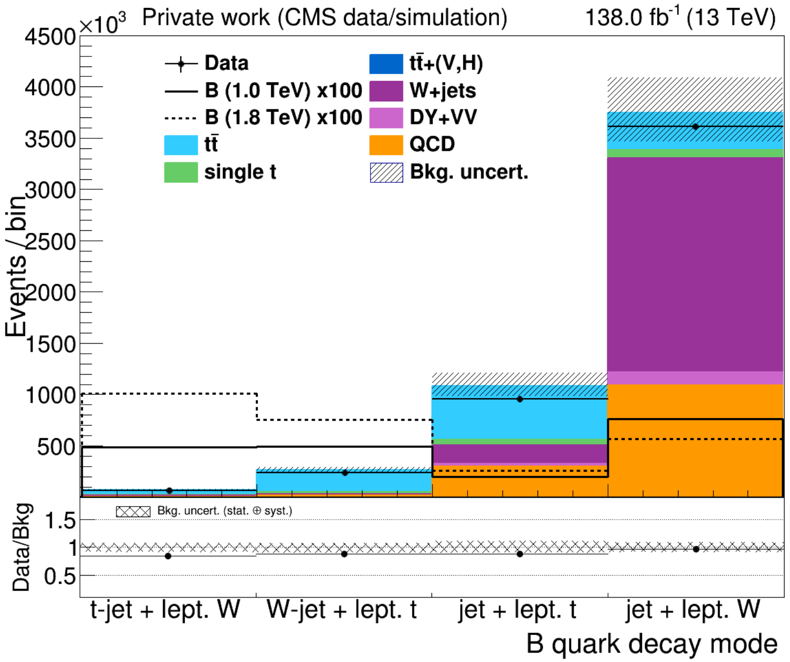






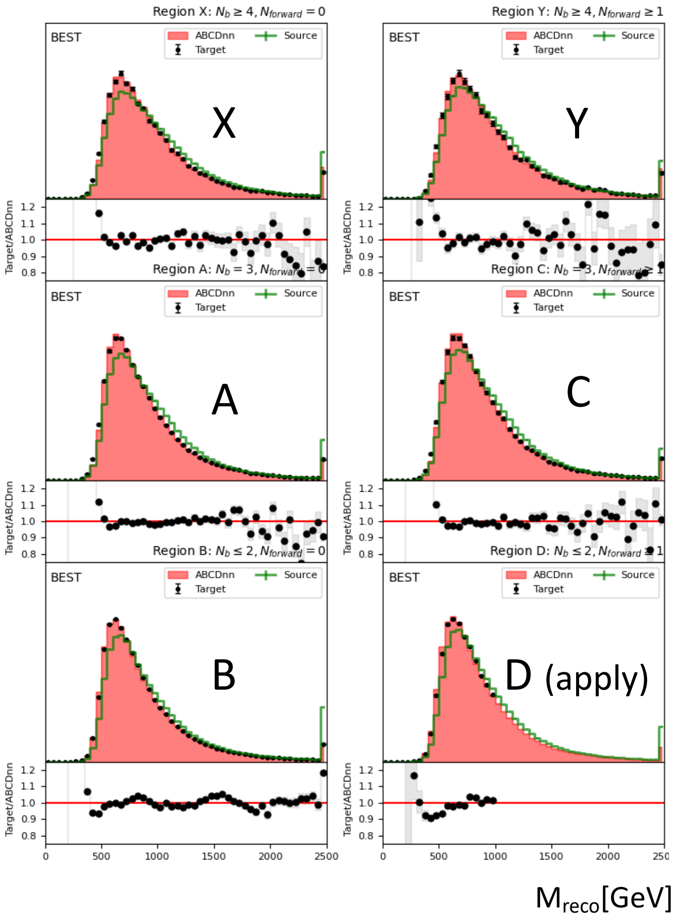
Background estimation: ABCDnn method

- Train two DNNs separately:
  - leptonic W (case 1+4)
  - leptonic T (case 2+3)
- Replace major background MC with ABCDnn estimation
- Minor backgrounds still modelled by MC

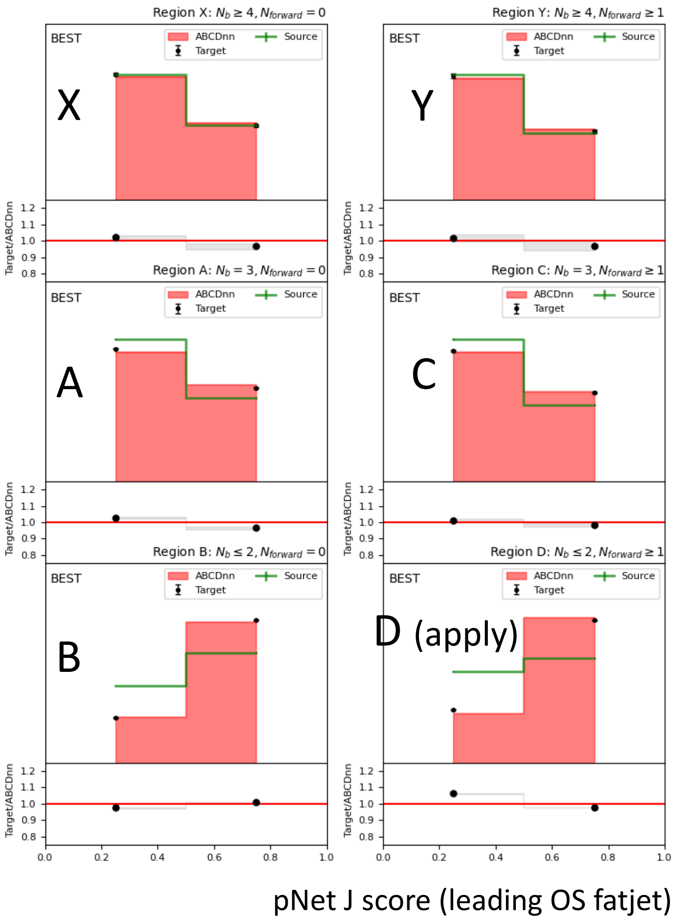


ABCDnn training results for leptonic top (case 2+3):

Private work (CMS data/simulation)

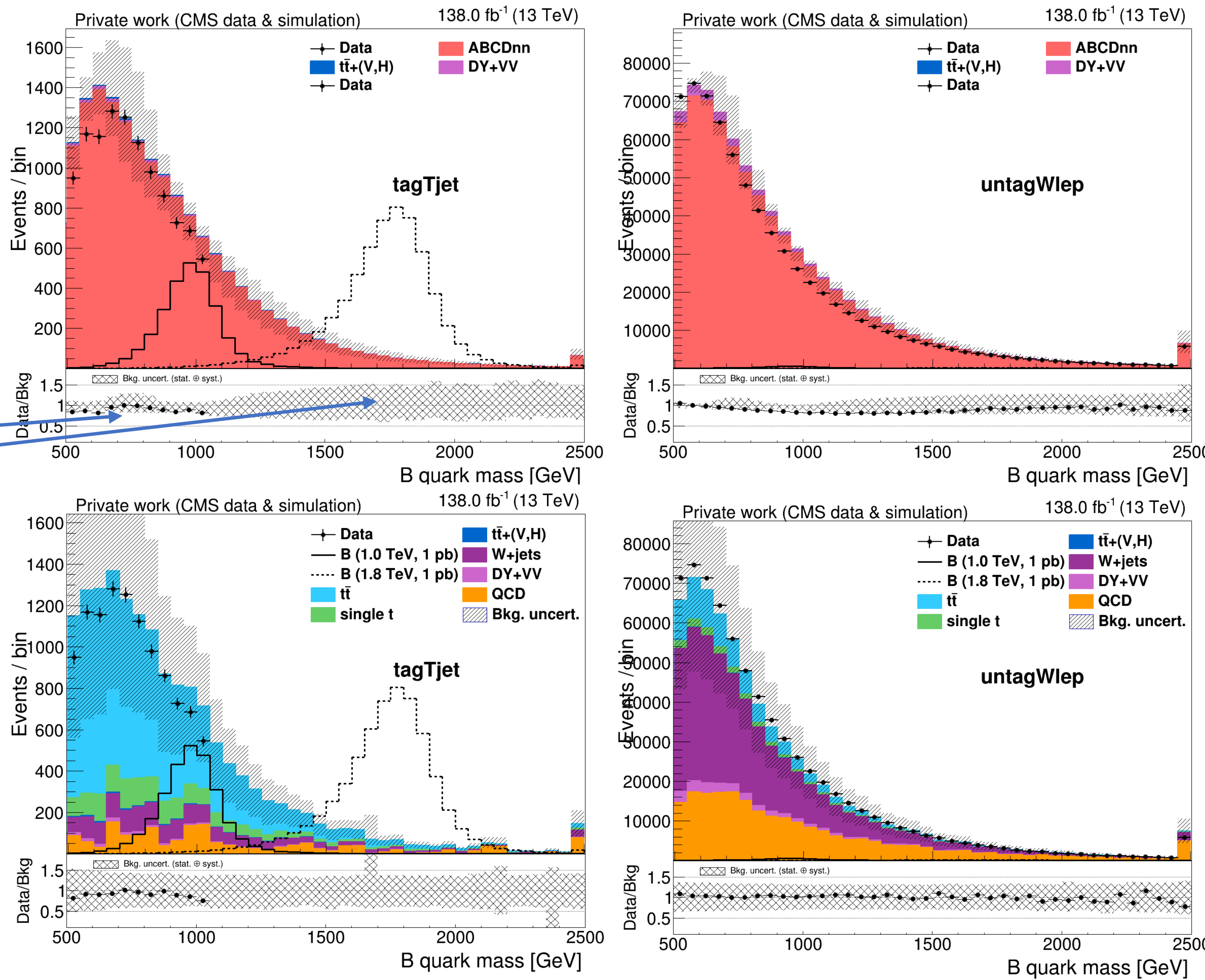


Private work (CMS data/simulation)



Background estimation: Apply ABCDnn

- Data-to-prediction agreement from ABCDnn in the background-heavy regions are comparable to MC modelling
- ABCDnn predictions carry uncertainties on the peak position, tail position and closure/yield
- ABCDnn uncertainties are often smaller in magnitude than MC-related uncertainties

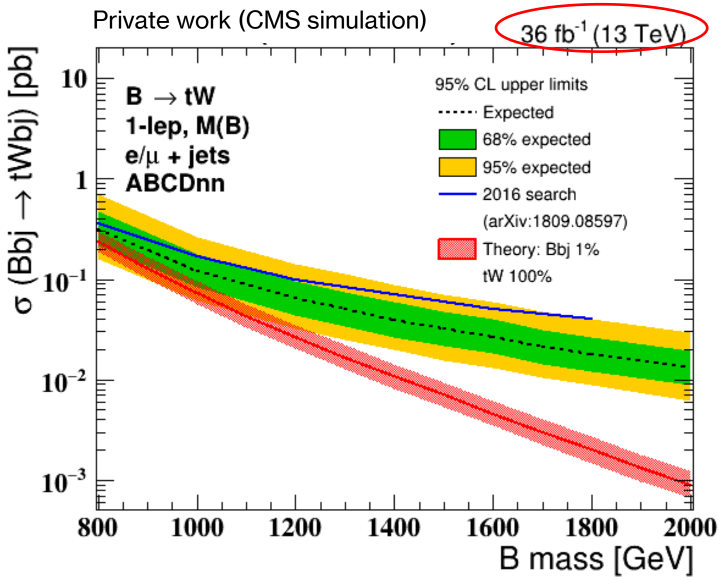


Expected limits

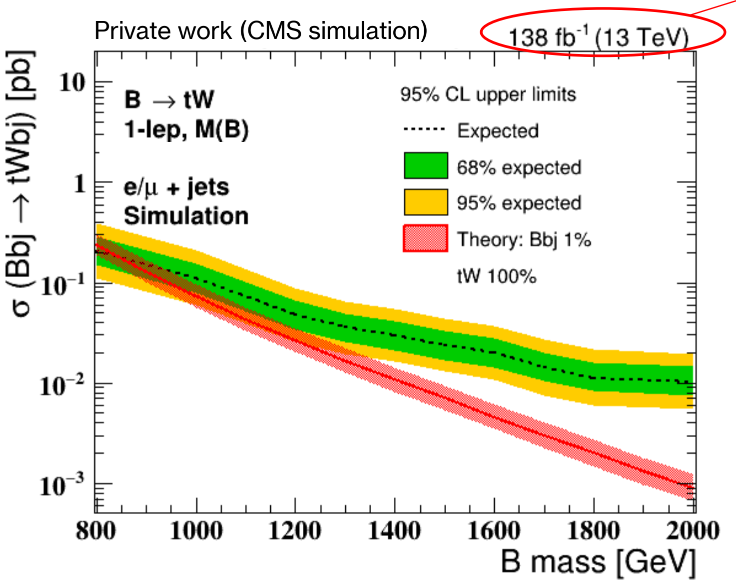
- Early analysis strategy studies showed promise for improvement over the previous search
- Full Run 2 expected limits are now tested comparing ABCDnn and MC for modelling major background
  - MC uncertainties included for experimental corrections and renormalization/factorization scale
  - Where used, ABCDnn uncertainties replace MC for major background

2016 only

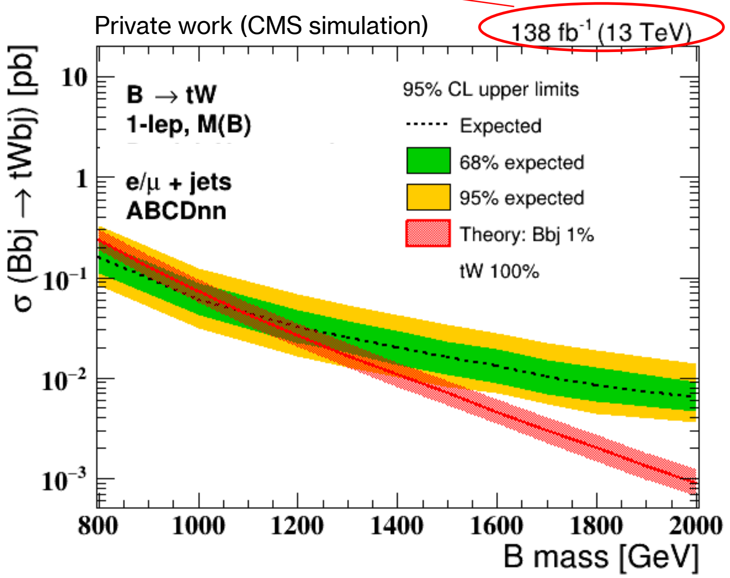
Full Run2 (2016-2018)



ABCDnn
Comparison with previous search



MC



ABCDnn

Summary & future work

- A search for single production of $B \to tW$ is in progress using CMS full Run2 data
- A new data-driven background estimation method (ABCDnn) is being explored and can be a good alternative to MC background modelling
- Expect an improved sensitivity using ABCDnn to model dominant backgrounds
- We hope to expand the search to include other single B production modes and widths, as well as Run3 data

Backup – Event selection

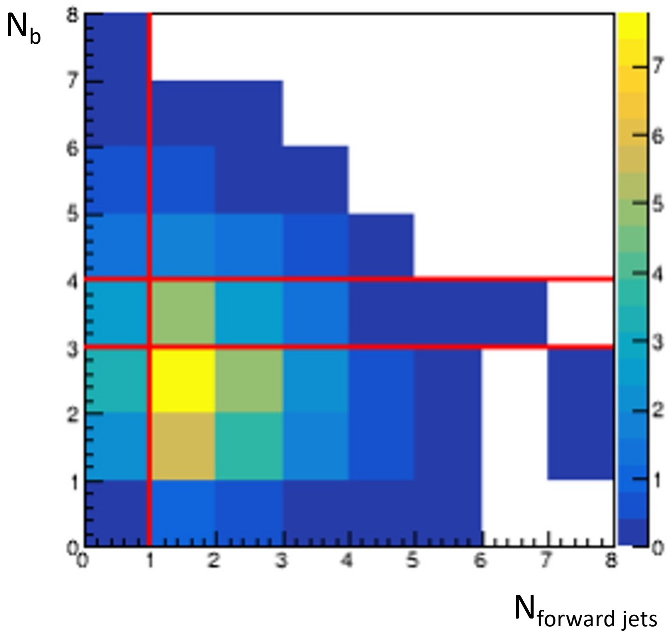
- Exactly 1 el/mu with high pT; no low pT leptons
- Lepton isolation requirement: loose
- AK4: min pT cut + central + tight ID + 2D cut against lepton
  - Loose b tagging
- AK8 jets: min pT cut + central + tight ID + minDR(lep, AK8) > 0.8
  - Top tagging 1% bkg WP, W tagging 5% bkg WP
- Require min AK4 HT cut + at least 1 AK8 with DR > π/2 from lepton
- QCD reduction
  - Get rid of low MET region
  - Upper limit on W m_T
  - Triangle Cut for electron events in $\Delta\varphi$-MET phase space

Backup – Yield correction: alpha-ratio

- Extended ABCD overprediction: overestimate by 49% in some categories
- Alpha-ratio is a better alternative: off by less than 10% (less than 5% in most cases)

$$\frac{N_{D(pred)}}{N_{D(MC\ major)}} = \frac{N_{B(data-minor)}}{N_{B(MC\ major)}}$$

Private work (CMS simulation)



Backup – Neural Autoregressive Flow (NAF)

Problem of finding a transformation to a base distribution

Base distribution: $p(\vec{x}|\vec{c})$

Estimate: $p(\vec{x}'|\vec{c}')$

x: transform variable
c: control variable
$\vec{c}$: control region
$\vec{c'}$: signal region

$$p(\vec{x}'|\vec{c}') = \sum_{\vec{c}\neq\vec{c}'} \int \boxed{\mathcal{T}(\vec{x}'; \vec{x}|c'; \vec{c})}\, p(\vec{x}|\vec{c})\, d\vec{x}$$

$\mathcal{T}$ = transformation (to be learned by DNN)

Backup – Neural Autoregressive Flow (NAF)

Build complicated distributions from a base distribution by stacking simple transformations

$$f(x_1,\ldots,x_d) = f_1(x_1)f_2(x_2|x_1)\ldots f_t(x_d|x_1,\ldots,x_{d-1})$$

$$p(\vec{x}'|\vec{c}') = \sum_{\vec{c}\neq\vec{c}'} \int \boxed{\mathcal{T}(\vec{x}';\vec{x}|c';\vec{c})}\, p(\vec{x}|\vec{c})\, d\vec{x}$$

In neural network:

$$\begin{aligned} y_1 &= \hat{f}_1\left(x_1;\theta_1(\vec{c}_0)\right) \\ y_2 &= \hat{f}_2\left(x_2;\theta_2(\vec{c}_0,x_1)\right) \\ &\ldots \\ y_d &= \hat{f}_t\left(x_t;\theta_t(\vec{c}_0,x_1,\ldots,x_{t-1})\right) \\ t &= 1,\ldots,d \end{aligned}$$

$\hat{f}_i$: neural network
θ_i: weights and biases

source: $\vec{x}$ $\longrightarrow$ target: $\vec{y}$

One common choice of $\hat{f}_i$: $\hat{f}_i(x_i;\theta_i) = \sigma^{-1}\left[\vec{W}^T(\theta_i)\cdot\sigma(\vec{a}_i(\theta_i)x + \vec{b}_i(\theta_i))\right]$

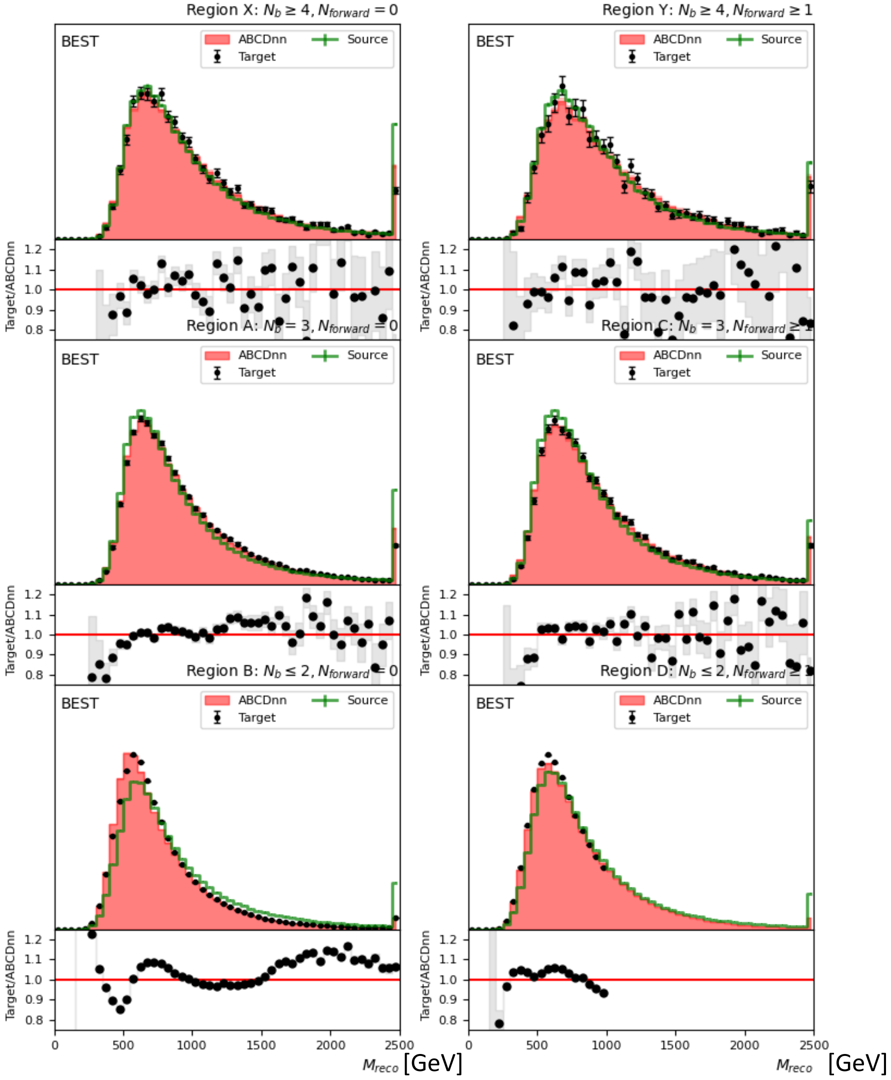
Deep Sigmoid Flow

(shown to be universal approximators to any bijective transformations in real space)

Backup – ABCDnn training results

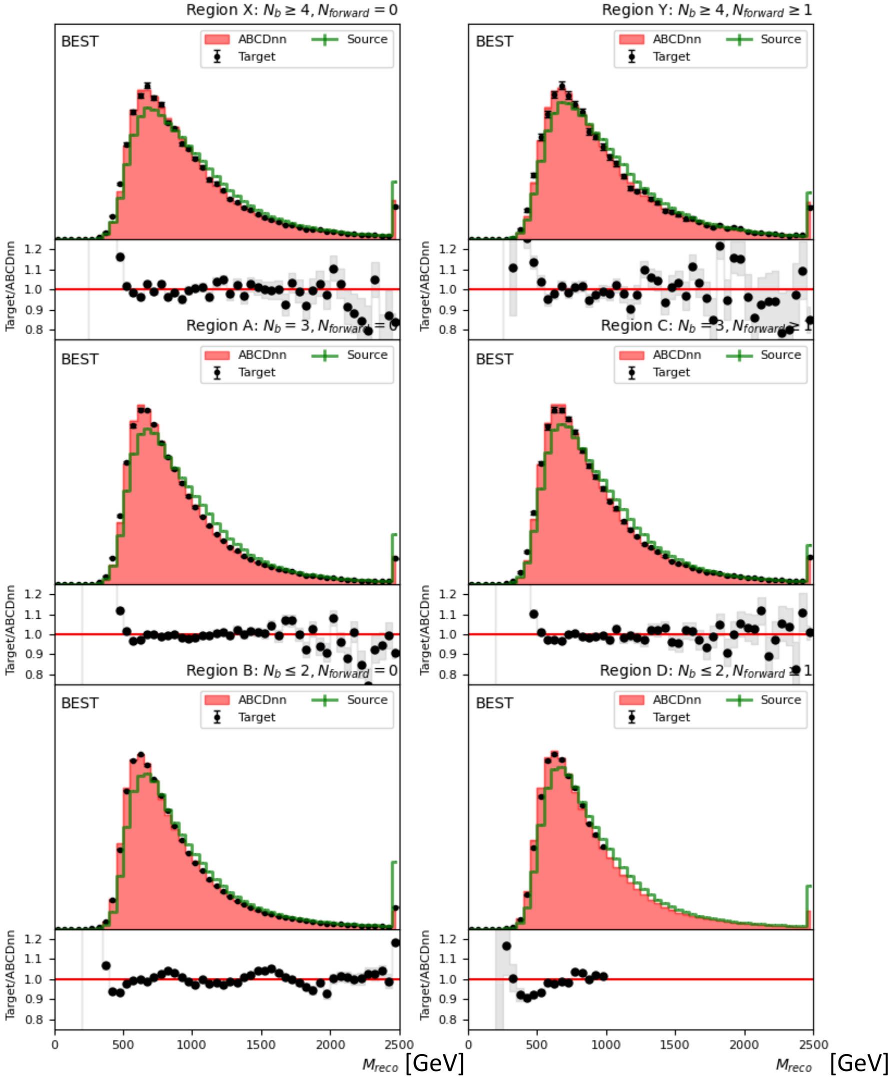
Private work (CMS simulation)

Leptonic W (case 1+4)



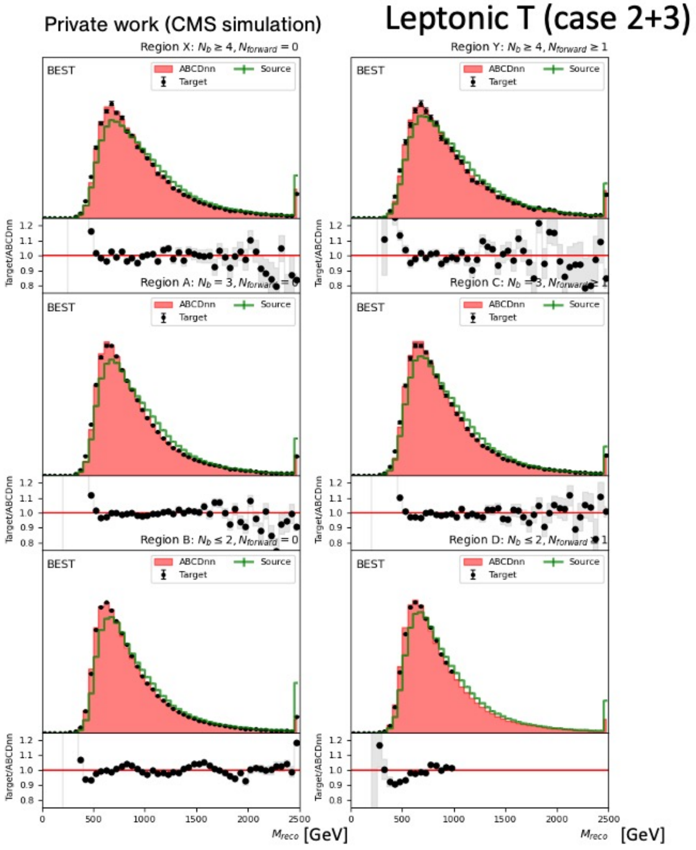
Private work (CMS simulation)

Leptonic T (case 2+3)



Backup – ABCDnn uncertainties

- Three sources of uncertainty
  - Tail variation $\delta y(x) = \pm\, c\, x \left(1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^2}\right)$, x: prediction, μ: prediction mean, σ: prediction std
  - Peak variation $\delta y(x) = \pm\, c\, x\, e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$
  - Closure variation $\delta y(x) = \pm\, c\, x\, e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$, x: input, μ: input mean, σ: input std



Backup: Background estimation results

Leptonic W (case 2+3)