DNN USED IN THE SEARCH FOR SCALAR TOP QUARK PAIR PRODUCTION IN THE TOP CORRIDOR REGION

Andrea Trapote

(On behalf of the CMS Collaboration)

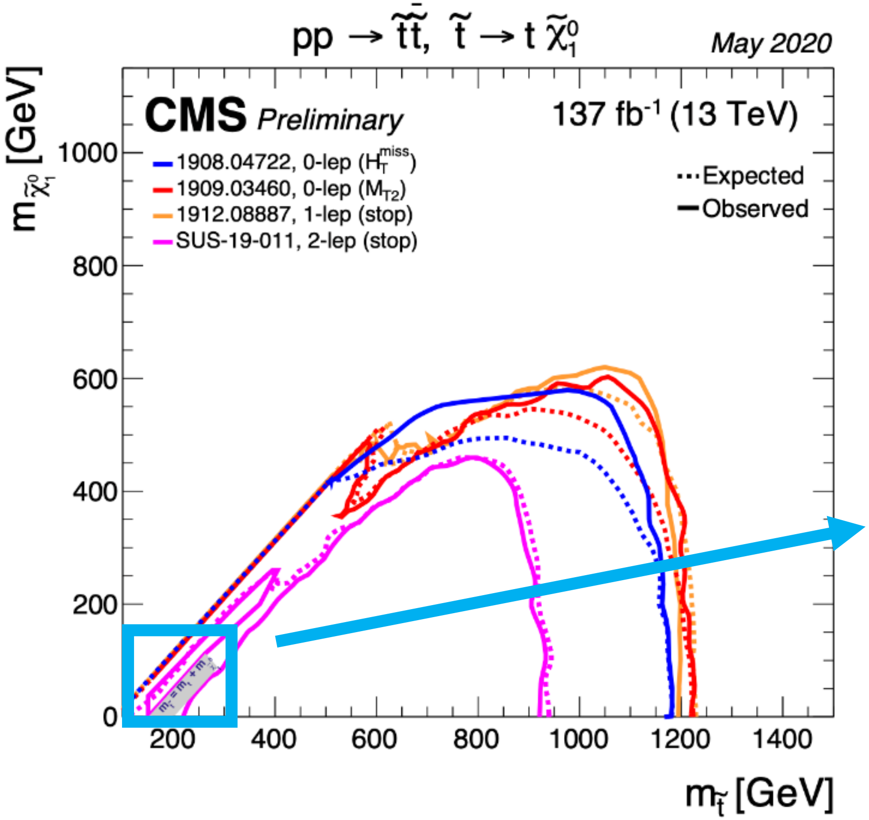
CMS-PAS-SUS-20-002

- I Workshop de Computing y Software de la Red Española de LHC -

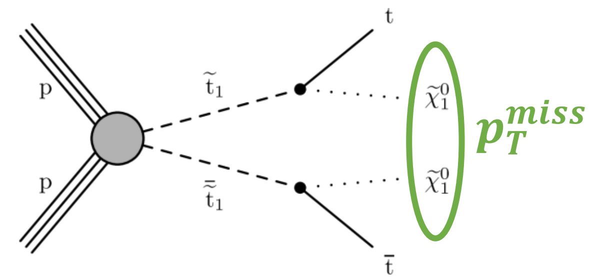
28-29 April 2021

STOP QUARK SEARCHES

- The stop quark plays an essential role in understanding the SUSY models.
- Several searches with the full **Run 2 dataset** have been performed by the CMS Collaboration excluding stop masses up to 1.2 TeV, but most of these searches are not sensitive in the so-called "top corridor".



Simplified Model Spectra "T2tt"



TOP CORRIDOR

- The mass difference between stop and neutralino is close to the **top mass**.
- Signal and $t\bar{t}$ background have **similar kinematics**, especially at low neutralino masses.
- Signal events can only be detected as an **excess on the $t\bar{t}$ cross section**.
- The **accurate estimation of $t\bar{t}$** process is very important to have sensitivity.

NEW REGION: SIGNAL REGION

BASELINE SELECTION

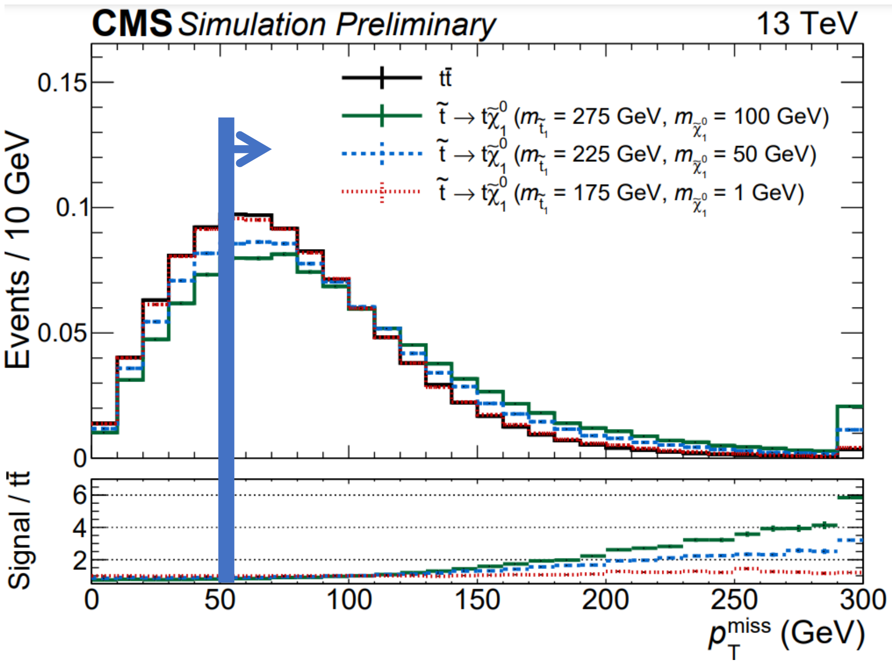
$\ell\ell$, $\geq$ **2 jets,** $\geq$ **1 b-tagged jet**

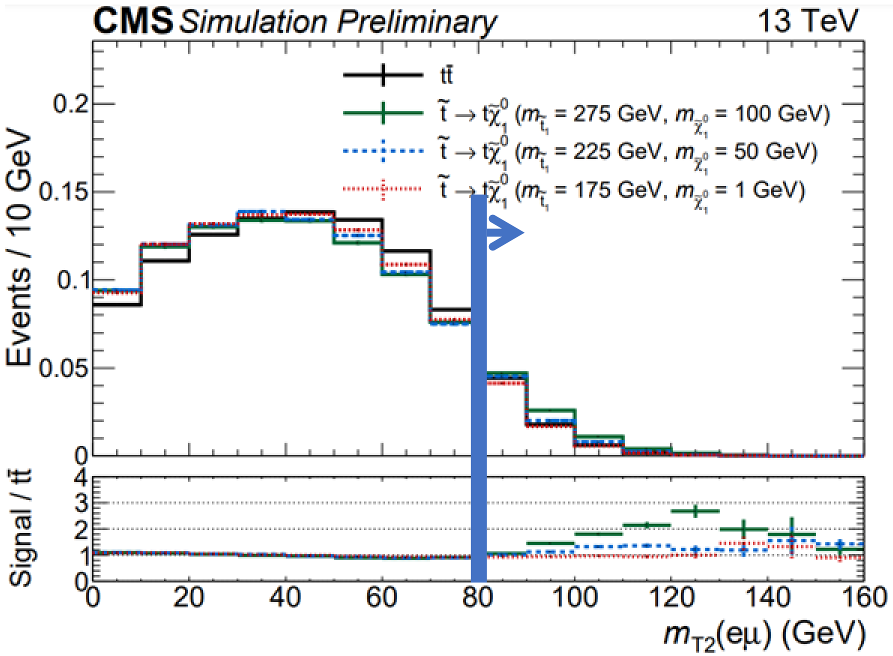
Signal with massive neutralinos tends to have larger p_T^{miss} and m_{T2}

SIGNAL REGION

$\ell\ell$, $\geq$ **2 jets,** $\geq$ **1 b-tagged jet (baseline selection)**

$\boldsymbol{p_T^{miss} \geq 50}$ **GeV,** $\boldsymbol{m_{\mathrm{T2}} \geq 80}$ **GeV**

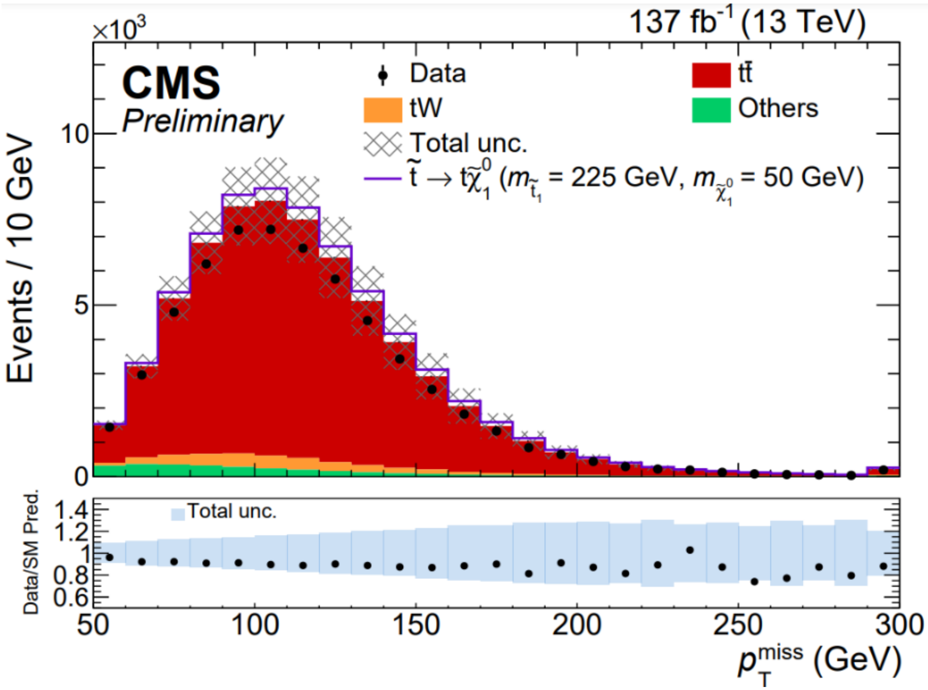


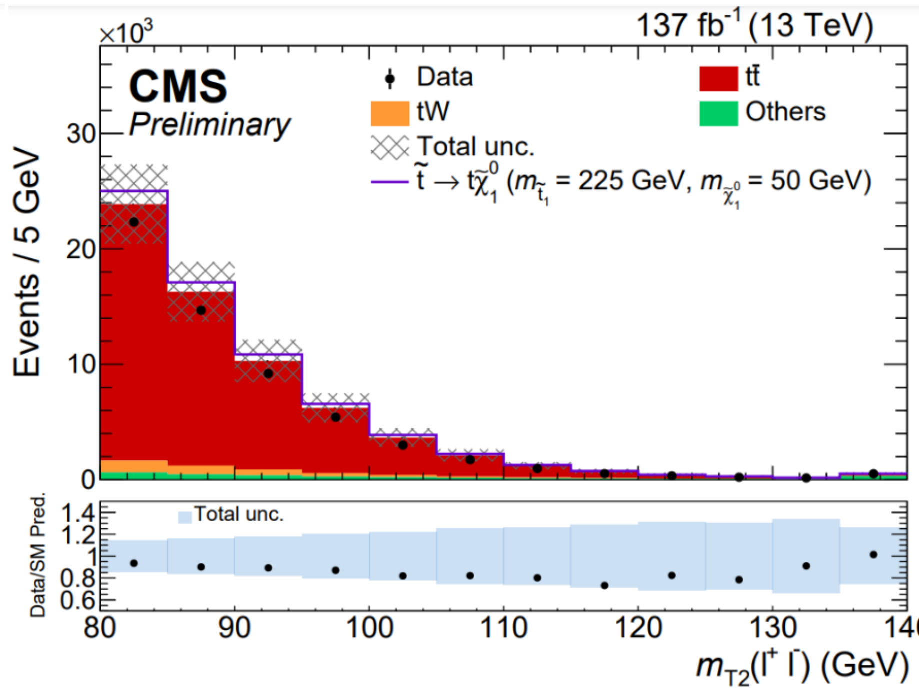


DATA/MC AGREEMENT

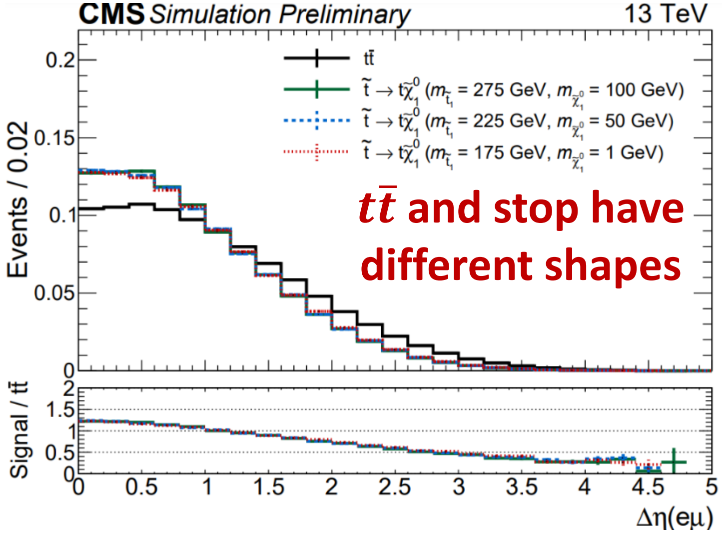
SR: BS, $p_T^{miss} \geq 50$ GeV, $m_{T2} \geq 80$ GeV

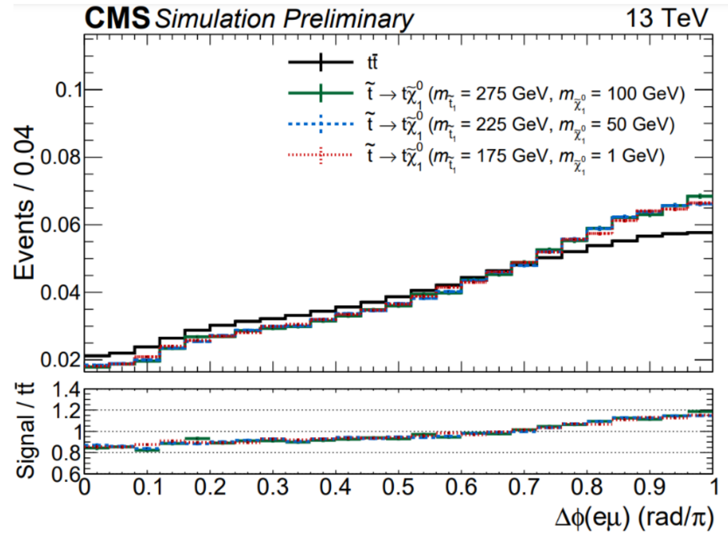
The amount of signal is much smaller than the background.





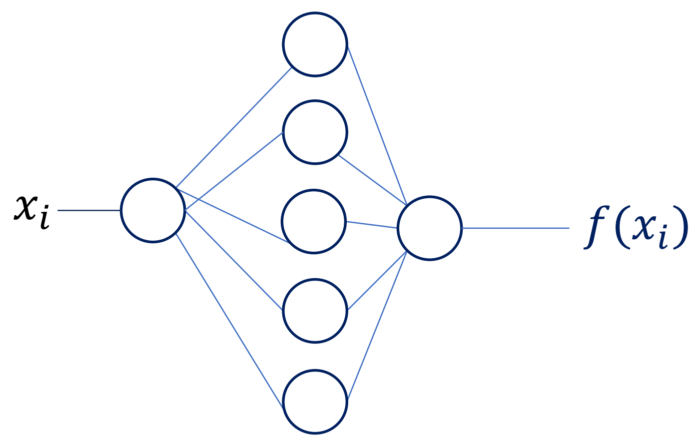
STRATEGY FOR SIGNAL EXTRACTION



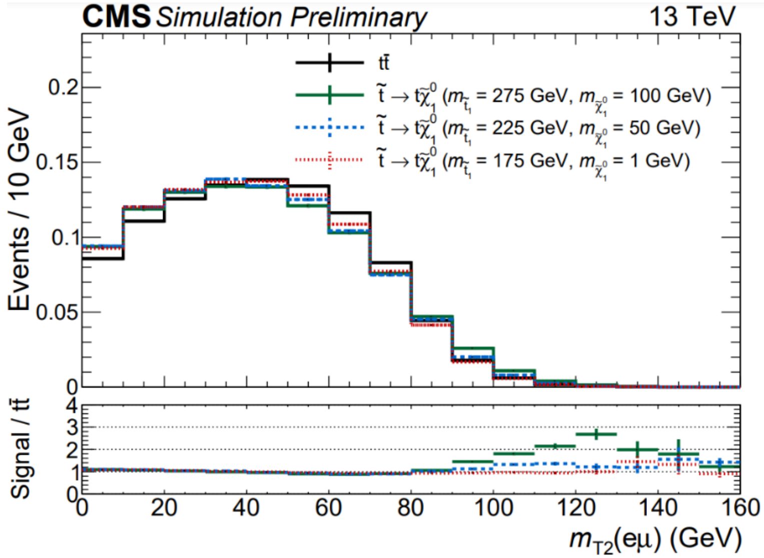


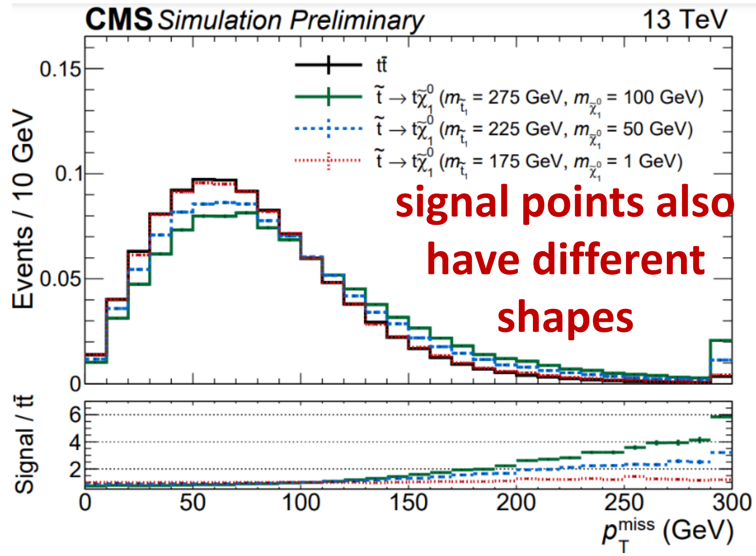
Separate signal from background is a complex work, but there are some variables where we can see differences. To achieve a better result we should take into account all these variables. Therefore, the best option is a **multivariate analysis:**

Deep Neural Network



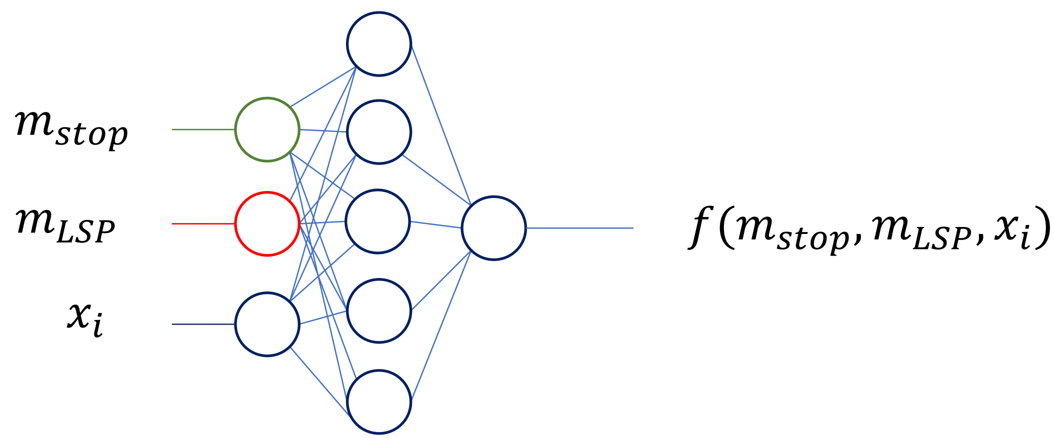
PARAMETRIC DEEP NEURAL NETWORK





In addition, some variables show that signal points have a different behaviour. We should have one model for each signal point or, what is better, we can have **one** model with a different result for each signal point:

PARAMETRIC DNN



With one DNN we have an specific model for each mass point!

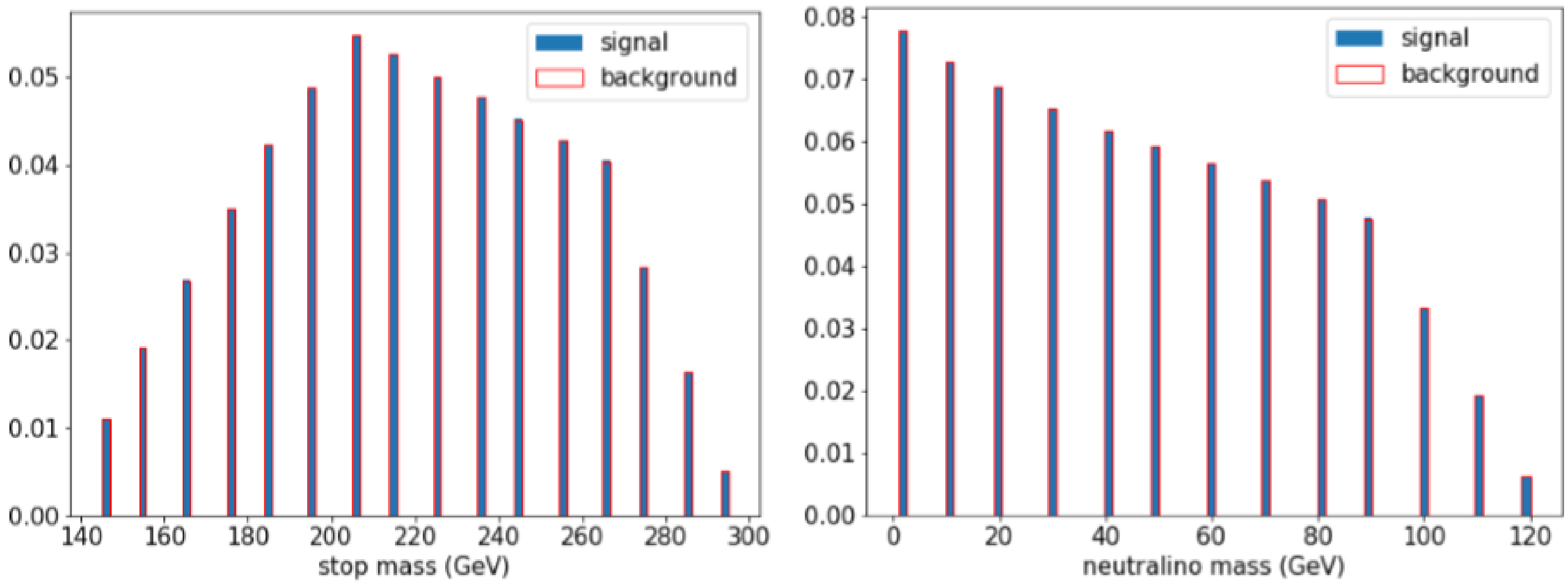
STRATEGY FOR SIGNAL EXTRACTION: PARAMETRIC DNN

A Parametric Deep Neural Network is used to separate signal from background.

- Only **$t\bar{t}$ events** are taken as background for the training.
- We are introducing the **3 years** in the training.
- **13 training variables**:

$$m_{stop},\, m_{LSP},\, p_T^{e\mu},\; \Delta\phi,\; \Delta\eta,\; p_T(l_0),\; \eta(l_0),\; p_T(l_1),\; \eta(l_1),\; p_T^{miss},\; m_{e\mu},\; m_{T2}(e\mu),\; H_T$$

- $\boldsymbol{m_{stop}}$ and $\boldsymbol{m_{LSP}}$ **values randomly assigned** to background events from the signal distributions to avoid introducing correlations.



DNN TRAINING

BS: $e\mu$, ≥ 2 jets, ≥ 1 b-tagged jet

SR: BS, $p_T^{miss} \geq 50$ GeV, $m_{T2} \geq 80$ GeV

Several training options were tested:

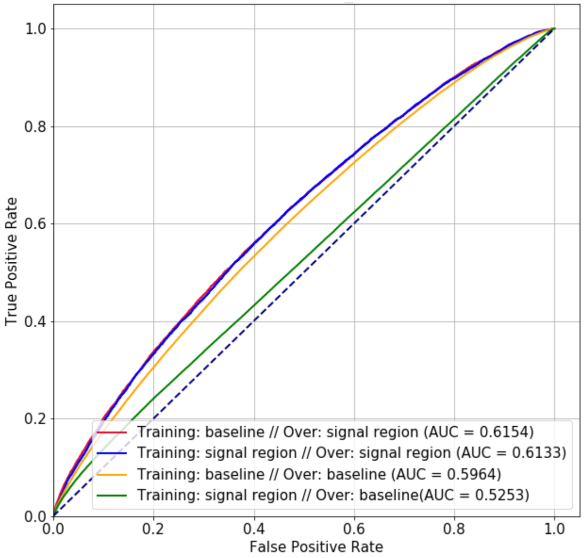
Model training in the BS and applied in the SR

Model training in the SR and applied in the SR

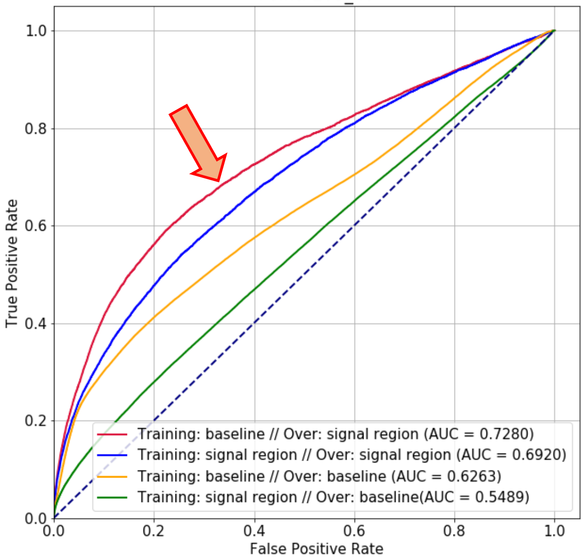
Model training in the BS and applied in the BS

Model training in the SR and applied in the BS

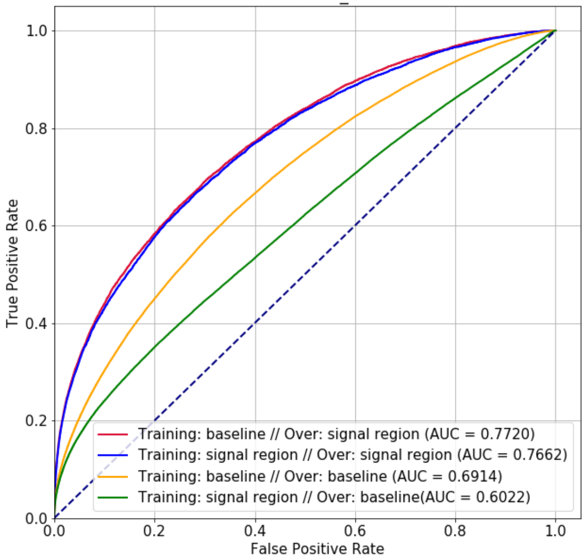
Point: 205 , 20



Point: 205 , 40



Point: 275 , 110



The best option is a model trained in the baseline selection and applied in the signal region.

DNN TRAINING: HYPERPARAMETERS

Layer (type)	Output Shape		Param #
dense_1 (Dense)	(None, 300)	ReLU	4200
dense_2 (Dense)	(None, 200)	ReLU	60200
dense_3 (Dense)	(None, 100)	ReLU	20100
dense_4 (Dense)	(None, 100)	ReLU	10100
dense_5 (Dense)	(None, 100)	ReLU	10100
dense_6 (Dense)	(None, 100)	ReLU	10100
dense_7 (Dense)	(None, 10)	ReLU	1010
dense_8 (Dense)	(None, 2)	Softmax	22

```
Total params: 115,832
Trainable params: 115,832
Non-trainable params: 0
```

The hyperparameters were optimized using *GridSearch* and manually.

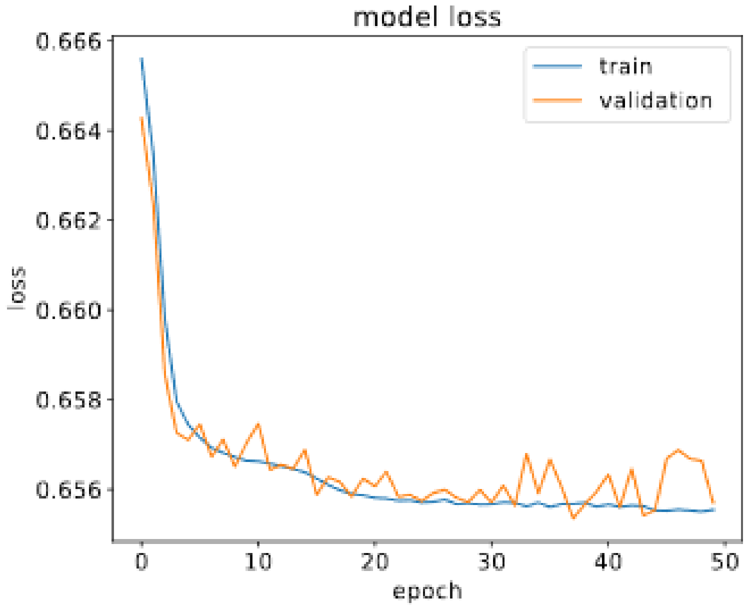
- **Optimizer**: Adam (lr=0.0001)
- **Loss function**: binary crossentropy
- **Epochs**: 50

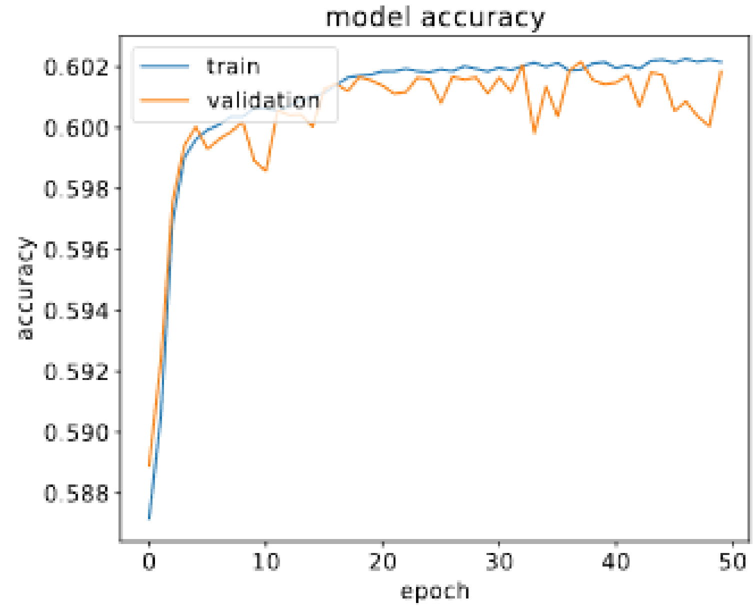
- **15 million events**:
8 from $t\bar{t}$ and 7 from signal

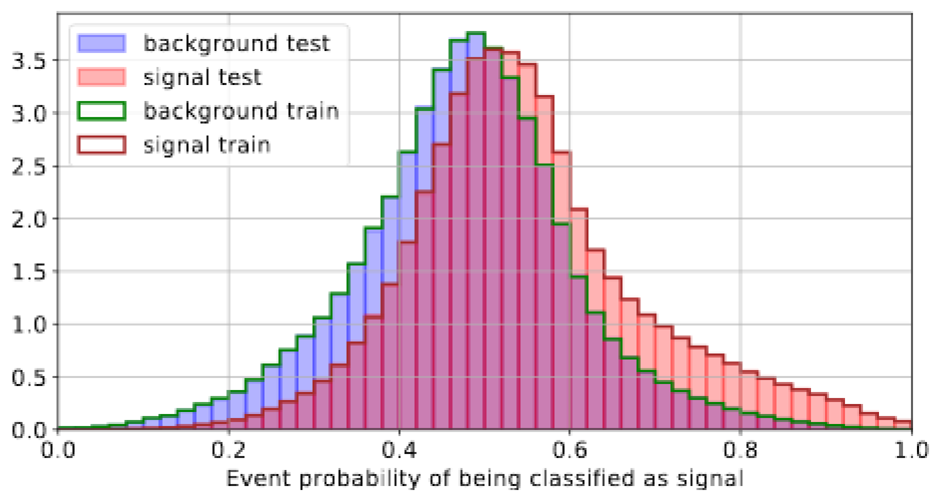
• 60% for train
• 15% for validation
• 25% for test

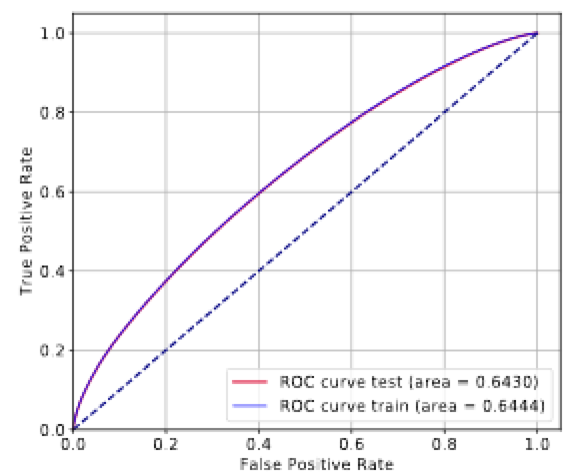
DNN TRAINING

The training was realized with **TensorFlow** using the **Keras** interface.





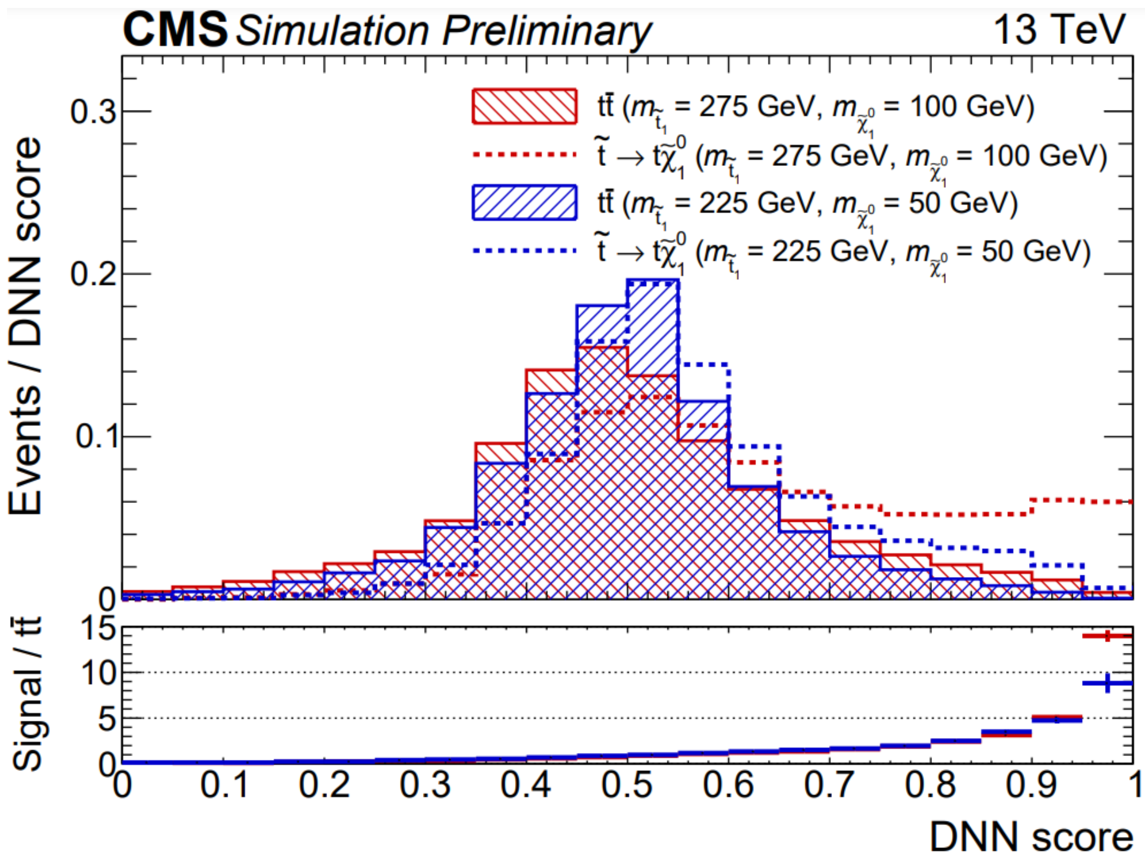




DNN OUTPUT

SR: BS, $p_T^{miss} \geq 50$ GeV, $m_{T2} \geq 80$ GeV

DNN output for two different mass points:

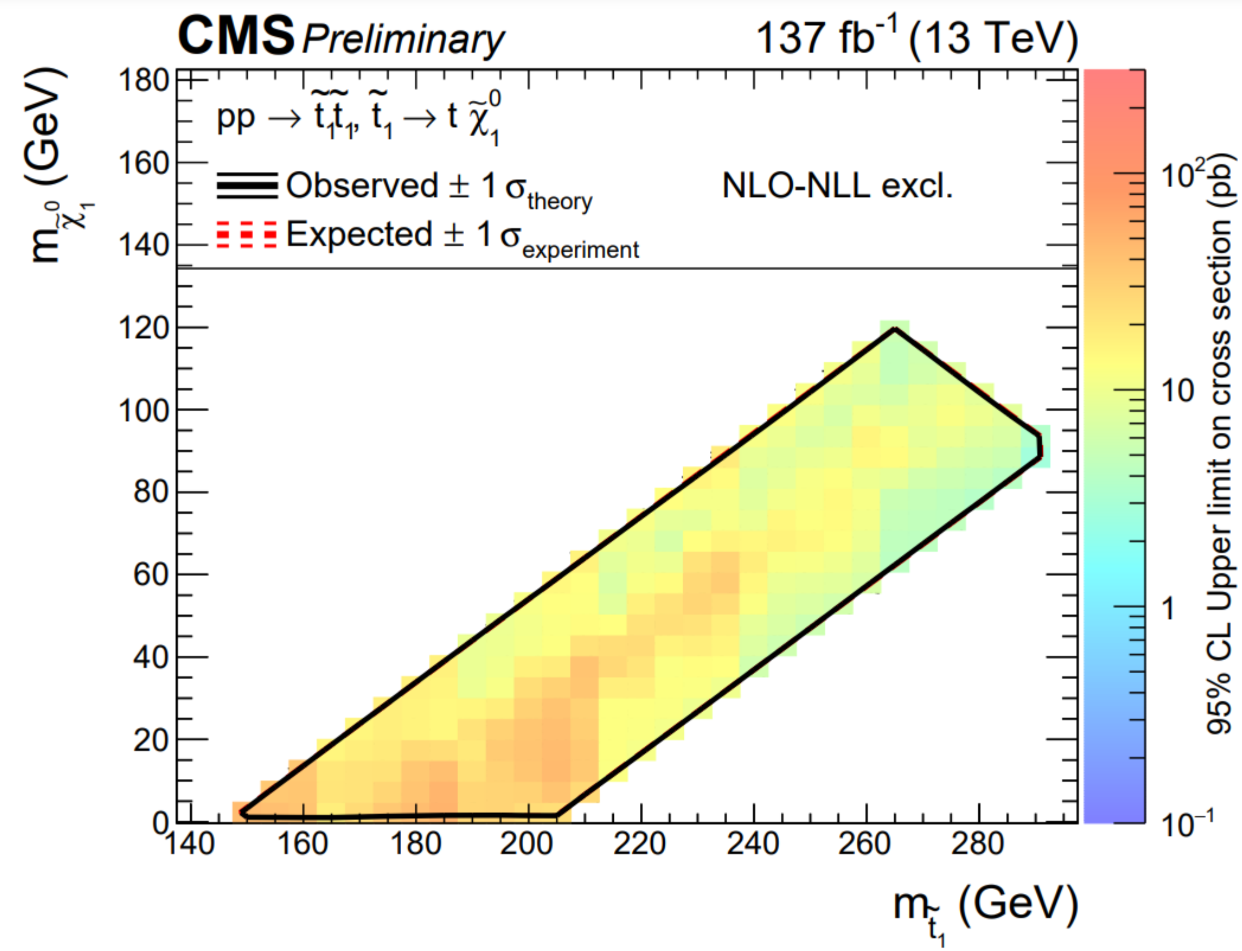


- The DNN score is different because of the parametric training, **we have one signal and one background distribution for each mass point**.

- **Good discrimination** is observed at high DNN score.

RESULTS

Upper limits on the production cross section of top squark pairs are calculated at 95% confidence level.



Full top corridor region is excluded!

SUMMARY

CMS-PAS-SUS-20-002

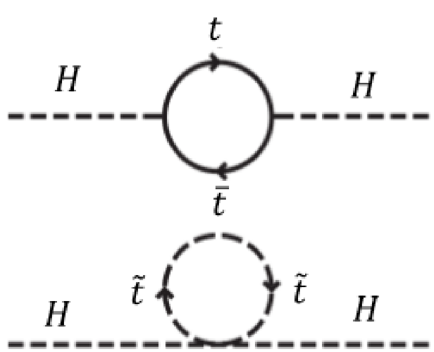
- Separate signal from $t\bar{t}$ background is very difficult in the "top-corridor" region because they have very similar kinematics.
- To consider all the small differences in several variables the best option is used a **Multivariate Analysis**.
- A **Parametric Deep Neural Network** was selected by introducing the stop and neutralino masses in the training. In this way, we have a specific model for each combination of masses.
- This strategy has been followed also by other analysis, such as the search for EWK SUSY with multileptons [CMS-PAS-SUS-19-012].

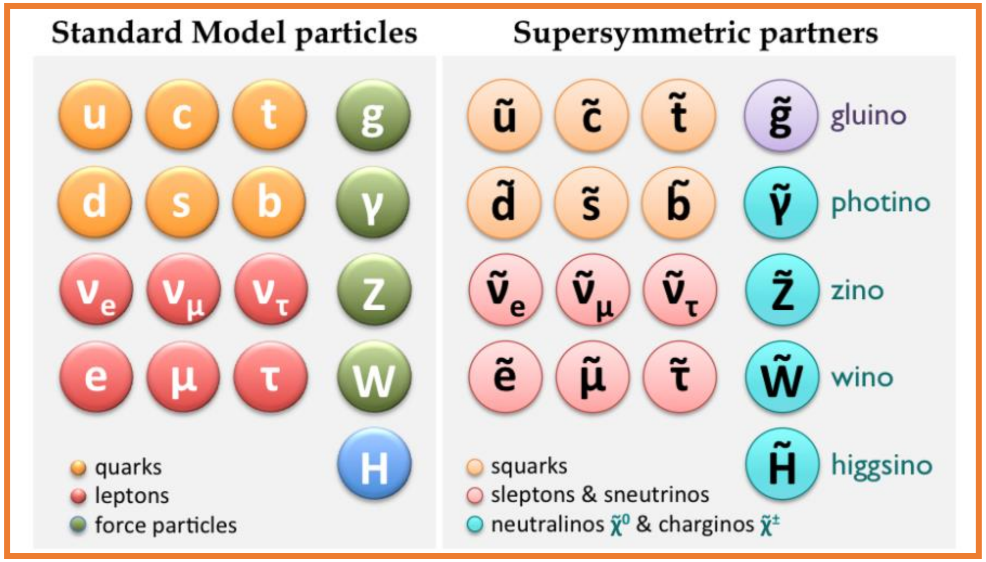
BACK UP

INTRODUCTION

- **Supersymmetry** is an extension of the SM that assigns a new particle (**superpartner**) to every SM particle differing only in ½ of spin.

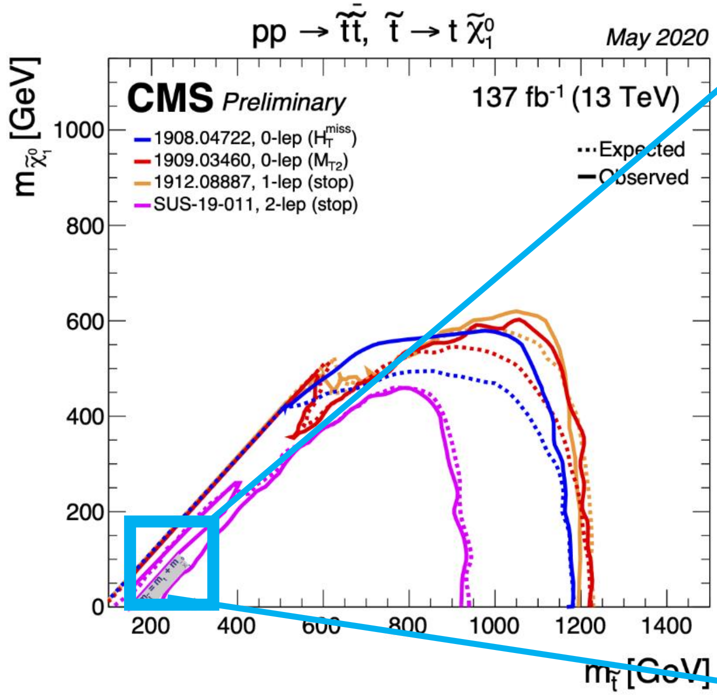
- This model can solve several shortcomings of the SM:
  - **Unification.**
  - If R-parity is conserved, the lightest supersymmetric particle (**LSP**) is stable and potentially massive, providing a good candidate for **Dark Matter**.
  - The **hierarchy problem** of the quantum loop corrections to the Higgs mass, due mainly to the top quark, can be compensated by the effect of the top quark superpartner.

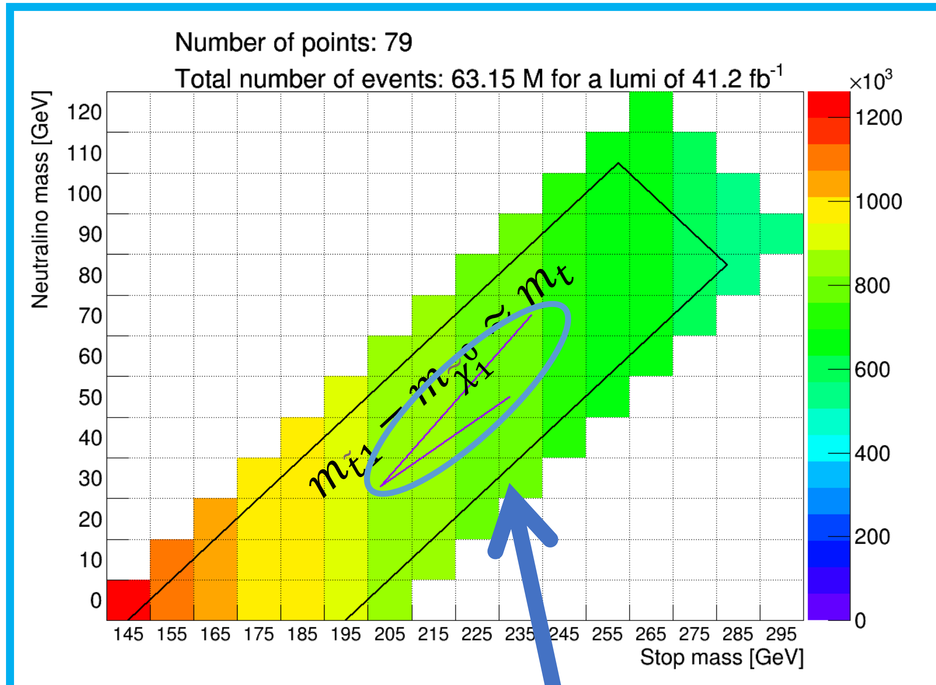




TOP CORRIDOR SEARCH

Goal: explore the non-excluded points (by SUS-18-003) in the top corridor with **Full Run 2** dataset using a **parametric DNN** to discriminate signal from $t\bar{t}$ background.





Changes wrt. SUS-18-003:

- FullSim signal samples from 2016 -> **Full Run 2**
- m_{T2} shape was used -> **MVA**
- 3 main diagonals -> **Full region tested**

Exclusion limits in SUS-18-003:

- $m(\tilde{t}_1)$ > 210 GeV for Δm = 175 GeV.
- $m(\tilde{t}_1)$ > 240 GeV for Δm = 175 ± 7.5 GeV.

$$m_{T2}(ll) = \min_{\vec{p}_{T,1}^{miss}+\vec{p}_{T,2}^{miss}=\vec{p}_T^{miss}} \left(\max[m_T(\vec{p}_T^{l1}, \vec{p}_{T,1}^{miss}), m_T(\vec{p}_T^{l2}, \vec{p}_{T,1}^{miss})]\right)$$

EVENT SELECTION

CMS-PAS-SUS-20-002

- Opposite-sign dilepton pair
- $m_{\ell\ell'} > 20$ GeV
- Leading lepton p_T > 25 GeV
- At least 2 jets
- At least 1 b-tagged jet
- $|m_{(ee,\mu\mu)} - m_Z| > 15$ GeV

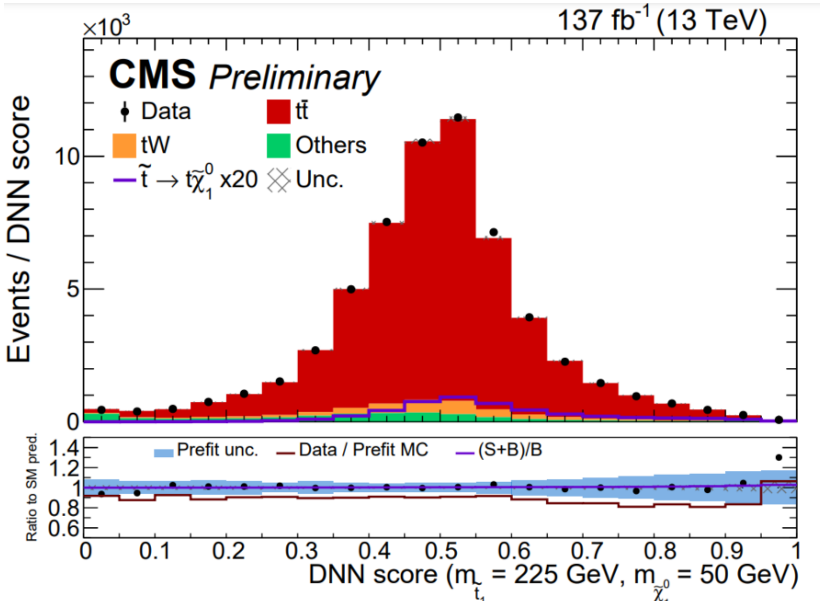
*** Standard selection for $t\bar{t}$ dilepton events.**

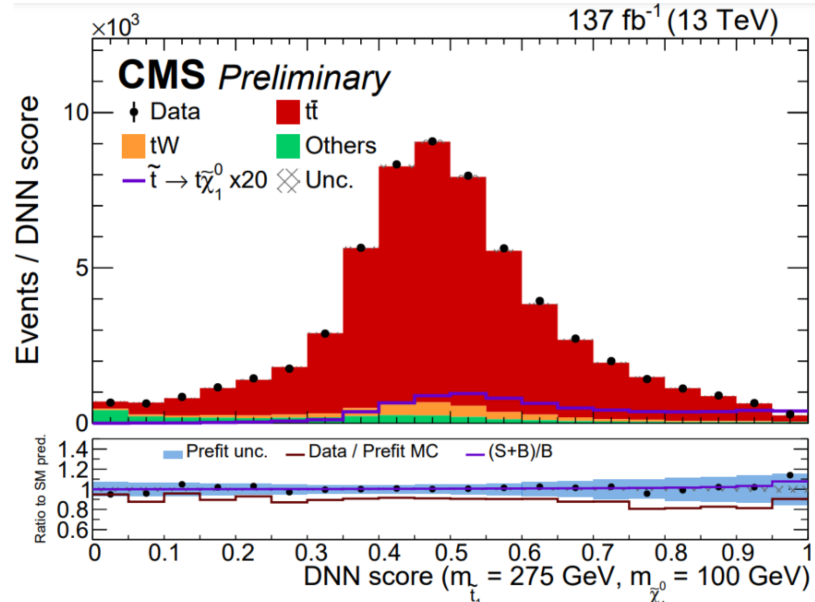
BASELINE SELECTION

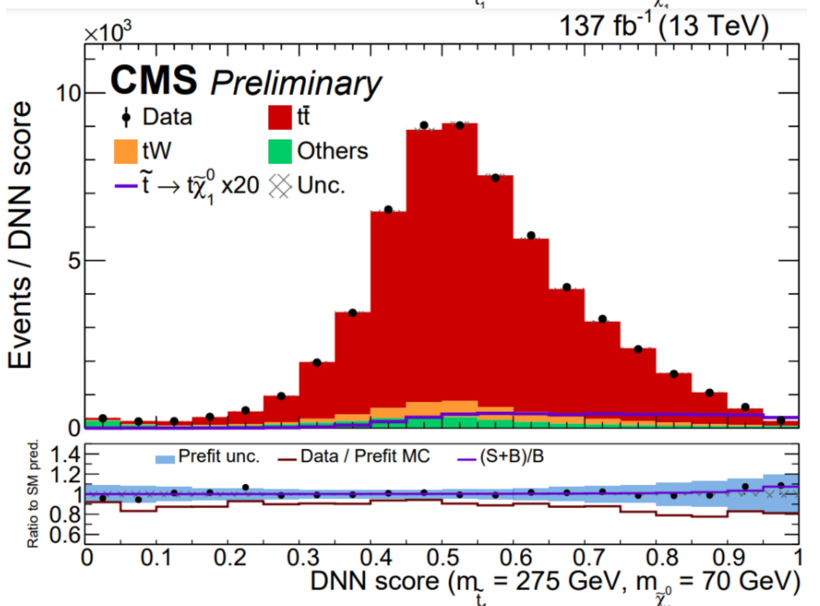
$\ell\ell$, ≥ 2 jets, ≥ 1 b-tagged jet

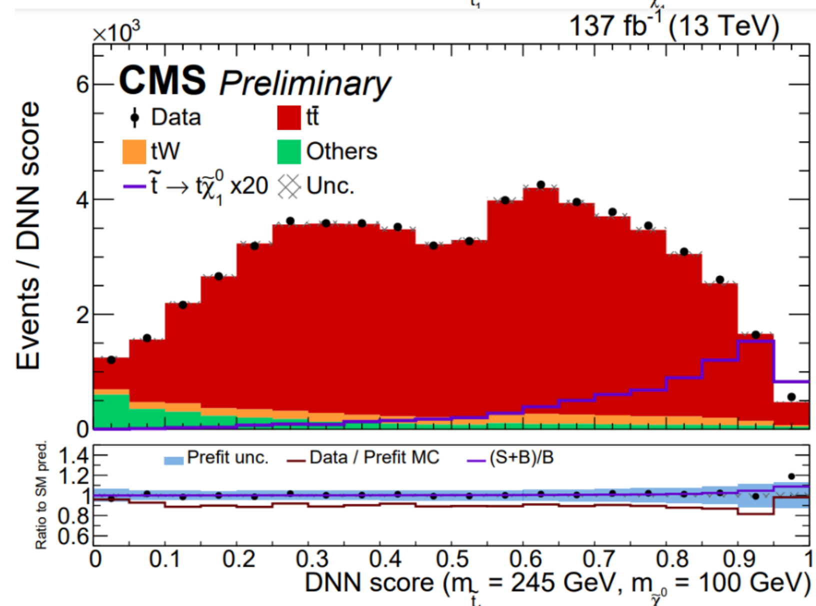
DNN POST-FIT DISTRIBUTIONS

SR: BS, $p_T^{miss} \geq 50$ GeV, $\mathrm{m_{T2}} \geq 80$ GeV





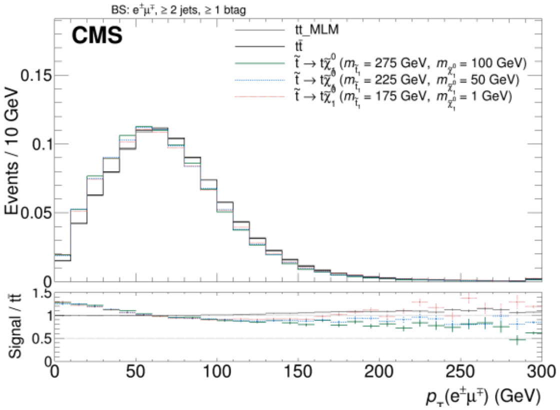


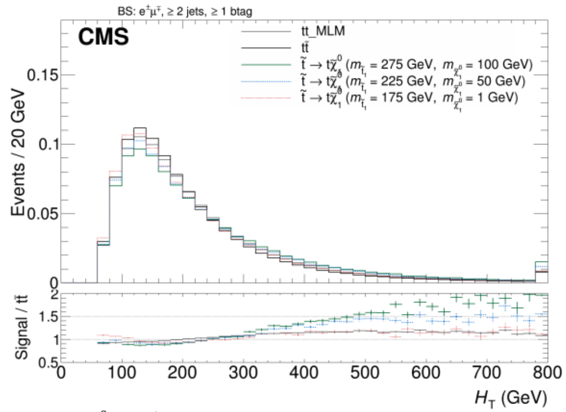


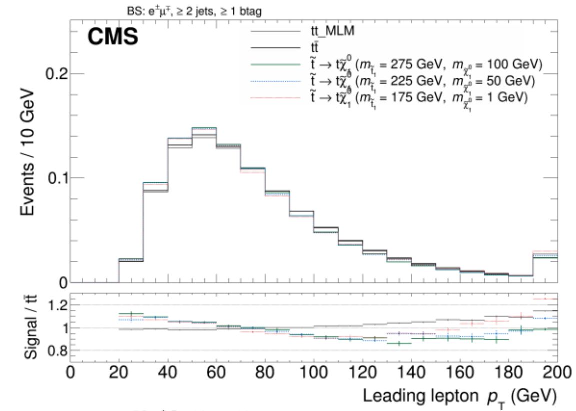
No excess is observed over the background prediction.

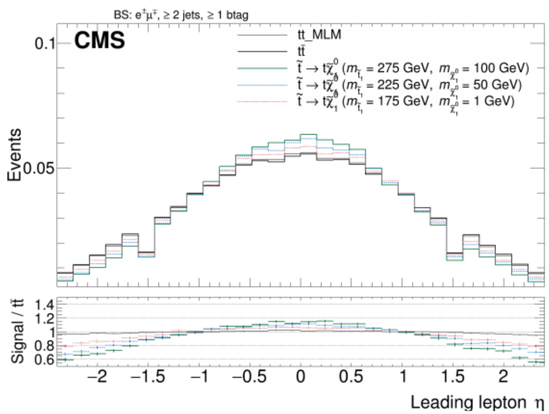
STRATEGY: PARAMETRIC DNN

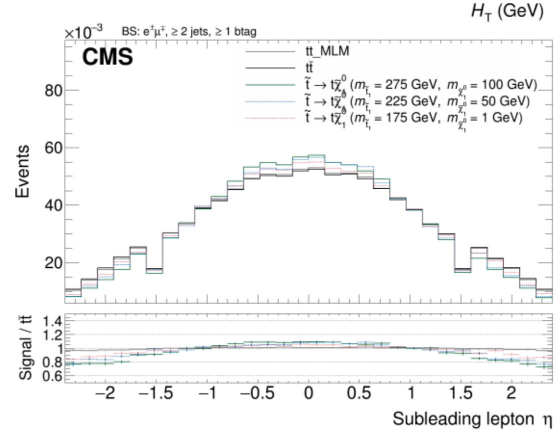
Training variables in baseline selection

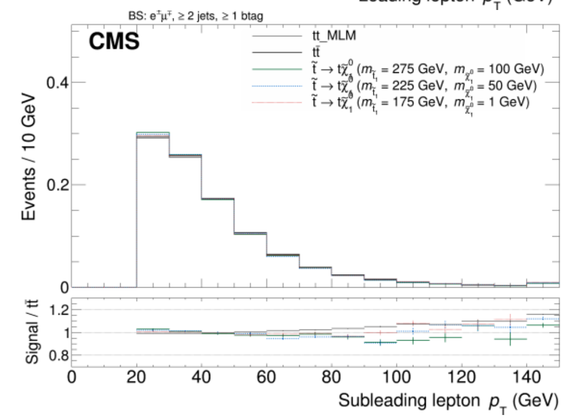


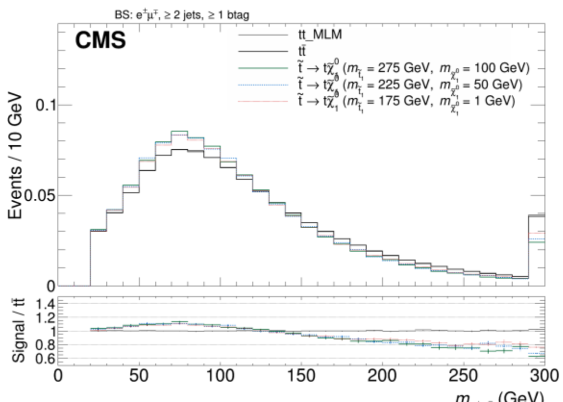












M_{T2} variable

$$M_{\mathrm{T2}} = \min_{\vec{p}_{\mathrm{T},1}^{\mathrm{miss}} + \vec{p}_{\mathrm{T},2}^{\mathrm{miss}} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}}} \left(\max \left[m_{\mathrm{T}}(\vec{p}_{\mathrm{T}}^{\ell 1}, \vec{p}_{\mathrm{T},1}^{\mathrm{miss}}), m_{\mathrm{T}}(\vec{p}_{\mathrm{T}}^{\ell 2}, \vec{p}_{\mathrm{T},2}^{\mathrm{miss}}) \right] \right)$$