Applying ML Techniques to Searches for Lepton-Partner Pair-Production at Intermediate Mass Gap at the LHC

Jason Kumar

University of Hawaii

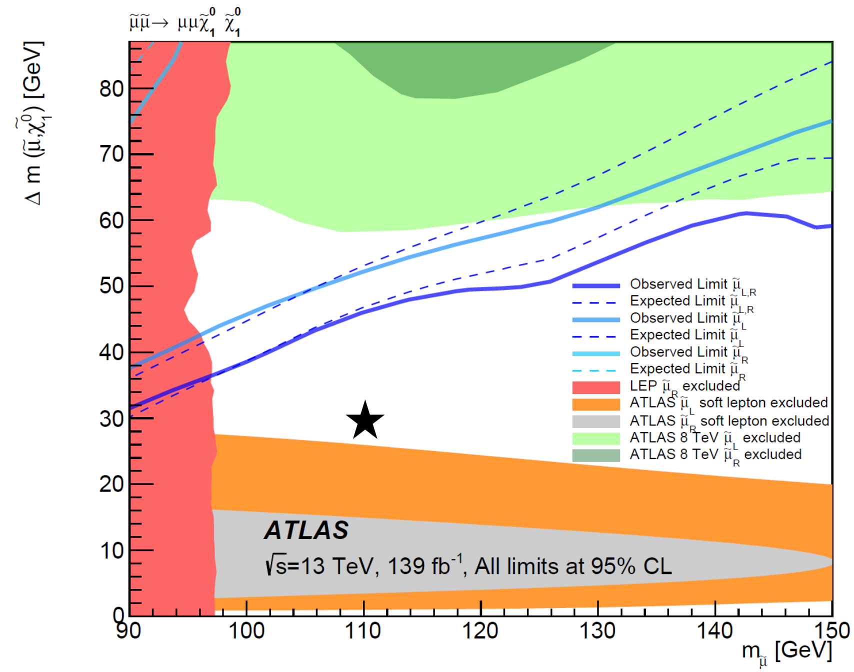
23xx.xxxxx

collaborators

- Bhaskar Dutta
- Tathagata Ghosh
- Alyssa Horne
- Sean Palmer
- Pearl Sandick
- Marcus Snedeker
- Patrick Stengel
- **Joel W. Walker**

lepton partners at intermediate Δm

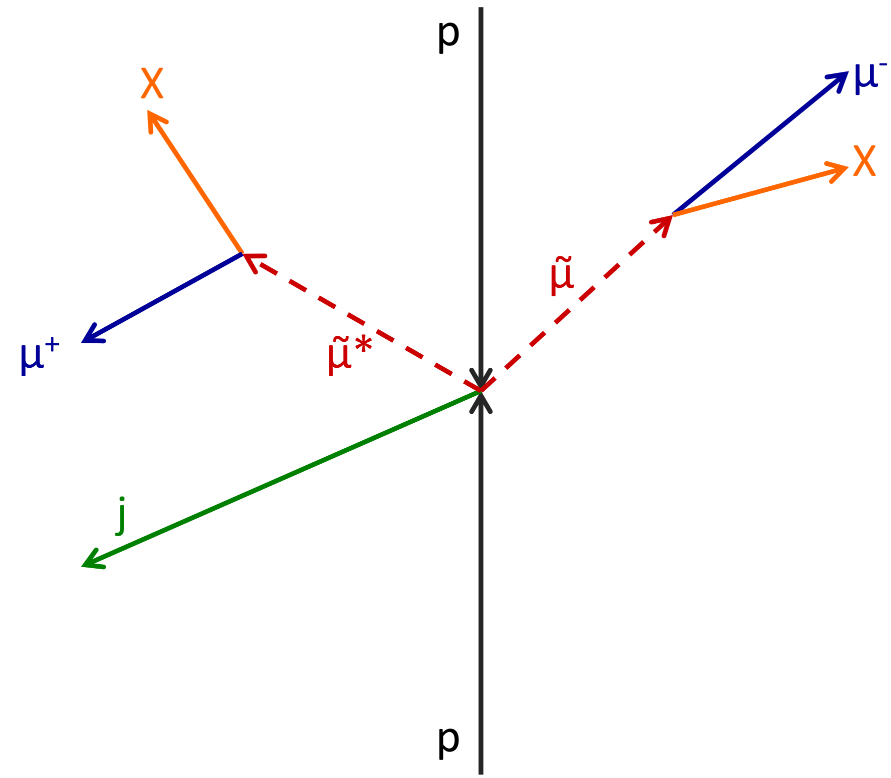
- searches at the LHC
  - pp → γ,Z → $\tilde{\ell}\ \tilde{\ell}^*$
  - $\tilde{\ell}$ → ℓ X
  - $\Delta m = m_{\tilde{\ell}} - m_X \sim$ 30 – 50 GeV
- many models, including MSSM
  - $\tilde{\ell}$ = slepton, X = bino LSP
- tough to constrain
  - leptons tend to be soft
  - EW bgds produce ℓ and MET (ν) with similar energies
- LHC mass reach no better than LEP for $\tilde{\ell} = \tilde{\mu}_R$

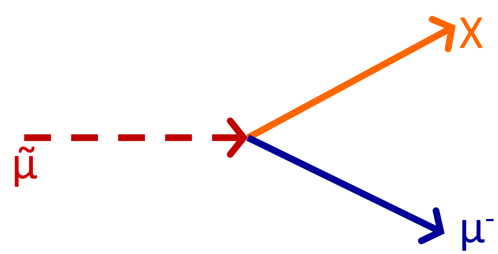


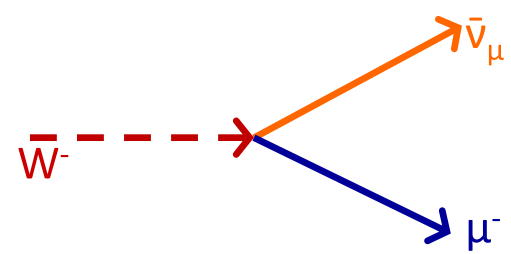
2209.13935

strategies and difficulties

- can demand a recoil jet to give leptons a transverse boost
- reduce threshold for lepton ID
- helps when mass splitting is small or large
- but hard when MPT of signal is similar to p_T of ν from W/Z decay
- rely on kinematic/angular dists. which distinguish parent, MET
- detailed cuts, but hard to find overarching principle







machine learning

- can we make progress with machine learning?
- we use a boosted decision tree (BDT)
  - will give us feedback on which kinematic variables are useful
  - help us reconstruct what the machine learned
- rotate training sample with analysis sample for reliability

- details of the BDT implementation ….
- use XGBoost
- depth = 5
- maximum number of trees = 50
- learning rate = ½

signal and background

- signal topology is $\mu^+ \mu^-$, 1 non-b-jet, MET (hadronic τ veto)
- signal benchmark
  - $m_{\tilde{\mu}R}$ =110 GeV, m_X = 80 GeV
- leading backgrounds
  - $\mu^+ \mu^- jjj$ (via Z,γ) (MET from missed jet)
  - $\tau^+ \tau^- jjj$ (via Z,γ) (MET, μ from τ decay)
  - $\bar{t}\, t\, jj$ ($t \rightarrow b\, W$, b missed/mistagged)
  - $W^+ W^- jj$ ($W \rightarrow \mu\, \nu$)
  - $Z\, Z\, jj$ ($Z \rightarrow \mu^+ \mu^-$, $Z \rightarrow \nu\nu$)
  - $W\, Z\, jj$ ($Z \rightarrow \mu^+ \mu^-$, $W \rightarrow \tau\, \nu$, τ mistag)

- details of the simulation ….
- generate events → MadGraph 5
- showering, hadronization → Pythia8
- detector simulation → DELPHES
- b-tag efficiency → 85%
- hadronic τ-tag efficiency → 85%

kinematic variables

- BDT uses high-level kinematic variables
- focus on variables distinguishing
  - mass/spin of parent
  - mass of invisible particle

- MET
- $m_{\mu\mu}$
- $\cos\theta^*_{\mu1\mu2}$
- m_j
- $M_{T2}^{0,100}$
- $(M_{T2}^{100}$ -100 GeV$)/ M_{T2}^{0}$
- $\Delta\phi_{(j,\mu1,\mu2,MET)}$
- M_{eff}
- H_T
- $m_{\tau\tau}$
- $p_T^{j,\mu1,\mu2}$
- $p_T^{j,\mu1,\mu2}$ / MET
- $\eta_{\mu1,\mu2,j}$
- $\tanh|\Delta\eta_{(\mu1,\mu2,j)}|$

the trouble with backgrounds

- several bgds are larger than signal, but have large hierarchies
- largest bgds are easy to remove
- pp → Z j(jj) → $\mu^+\mu^-$ j MET is the largest bgd by far …
  - but easily removed using $m_{\mu\mu}$
- but can't proceed unless we can kill the harder backgrounds also
- should we curate data to focus the BDT on the hard tasks?

- before precuts …

process	cross section (fb)
signal	12.3
$\mu^+\mu^-$jjj	12500
$\tau^+\tau^-$jjj	589
$\bar{t}t$ jj	65.6
W^+W^- jj	73.5
WZ jj	46.8
ZZ jj	26.6


What does WOPR recommend, Mr. McKittrick?
A dilepton mass veto around the Z-pole.
Ha ... I need some machine to tell me that?

precuts and logic

- veto $m_{\mu\mu} \in 91 \pm 10$ GeV
  - kill $Z \rightarrow \mu^+\mu^-$
- require MET > 75 GeV
  - kill MET via jet mismeasurement
- require $\cos\theta^*_{\mu\mu} < 0.5$
  - prefers a spin-0 parent

- rough goal
  - precuts we understand
  - all bgds have roughly comparable cross sections, and …
  - … not much more than 10-100 times larger than signal

- after precuts …

process	cross section (fb)
signal	3.27
$\mu^+\mu^-$jjj	15.7
$\tau^+\tau^-$jjj	48.7
$\bar{t}t$ jj	15.3
W^+W^- jj	13.7
WZ jj	0.876
ZZ jj	0.512

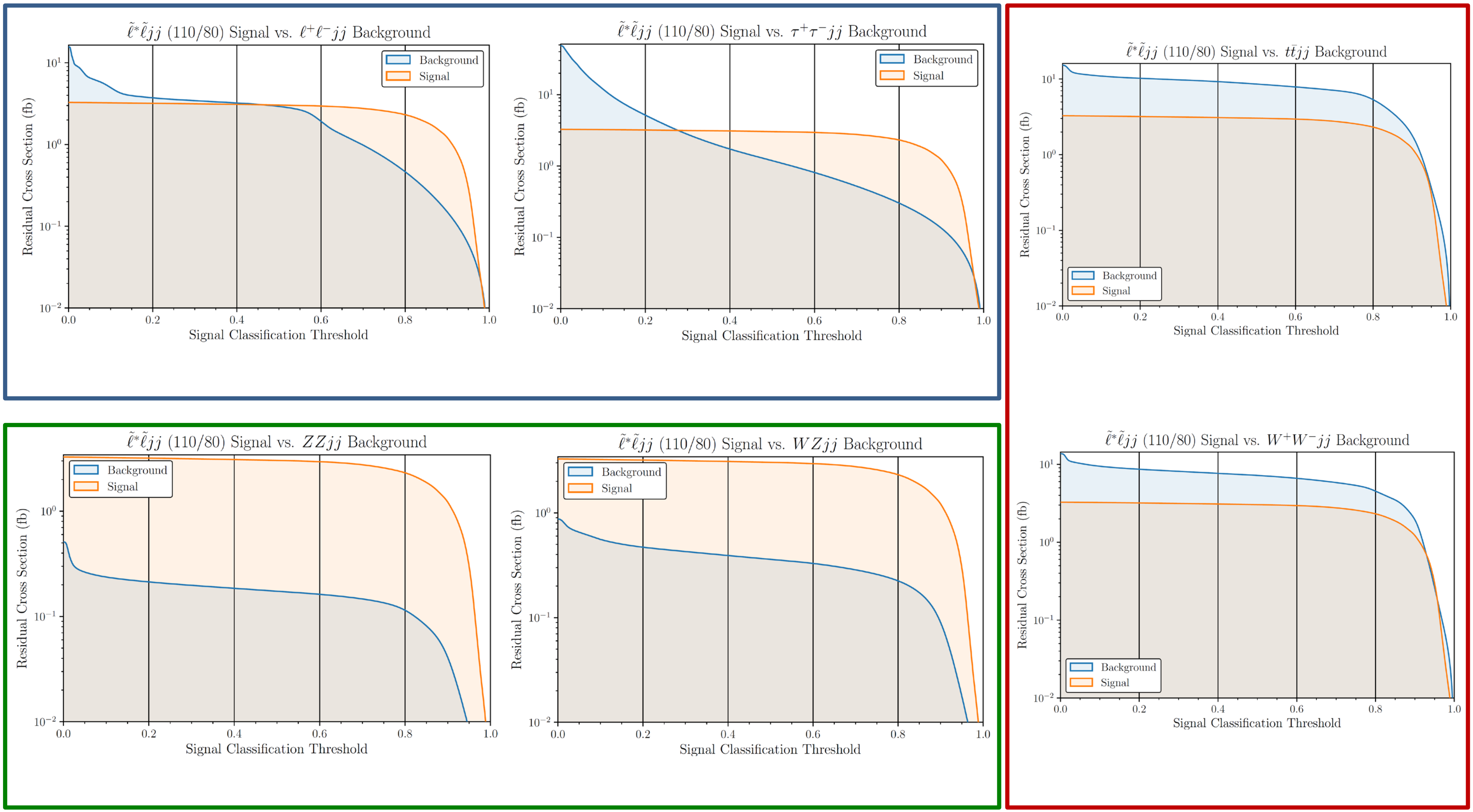
the trouble with simulation

- BDT works hardest on phase space regions w/ small cross sections

- easy to undersample these regions
  - not good if BDT training focused on only a few events

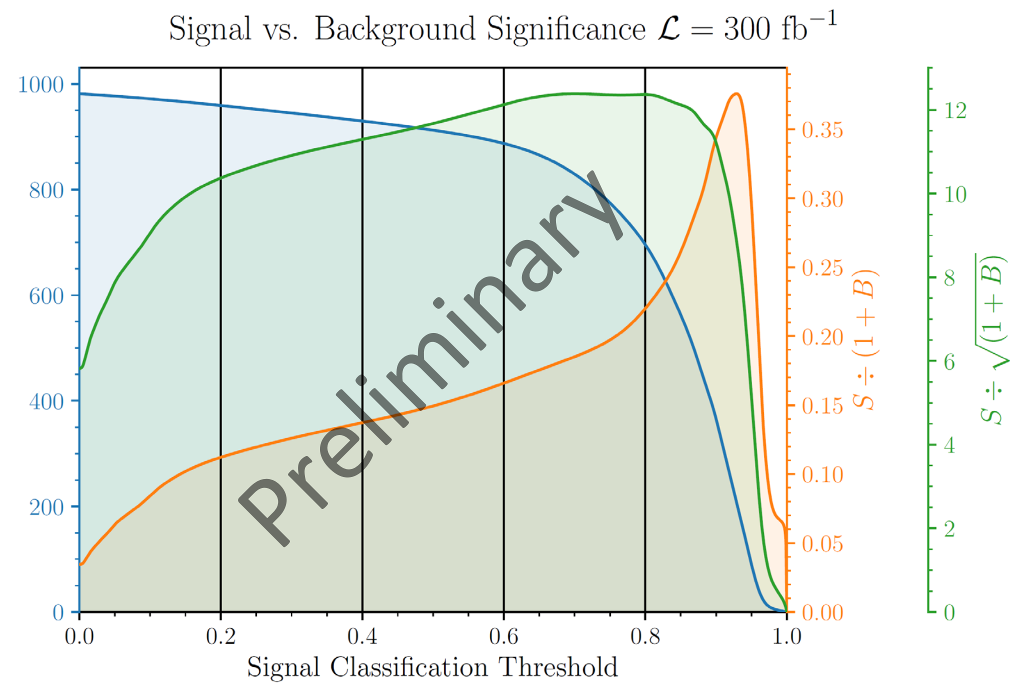
- generate signal and bgd simulation in kinematic tranches to ensure that tails are sufficiently sampled

discriminating signal from bgd.



results

- precuts kill WZ and ZZ
- BDT easily kills $\bar{\mu}\mu$, $\bar{\tau}\tau$
  - little loss of signal
- BDT earns its pay w/ W^+W^-, $\bar{t}t$
- roughly S = 300, S/B ~ 0.5, at best
- so expect maybe ~ 10 σ



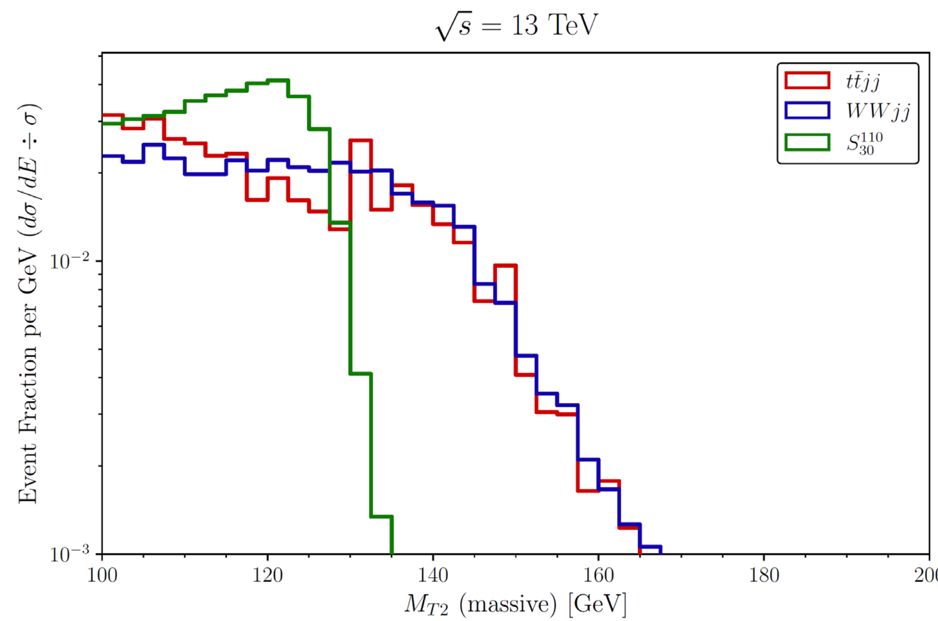
$m_{\tilde{\mu}R}$ =110 GeV
m_X = 80 GeV

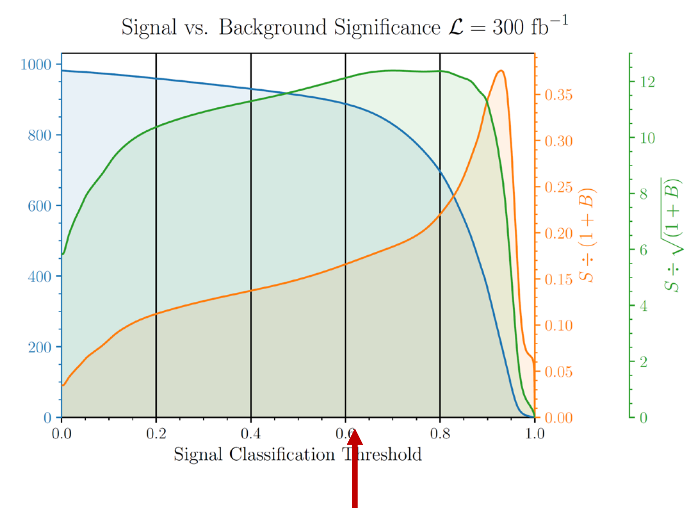
sensitivity scan upcoming!

what did the BDT learn?

- for W^+W^- and $\bar{t}t$ bgd, M_{T2}^{100} dominates the total gain
  - distinguishes mass of invisible particle
- assume we can kill all other bgd.
- 1000 signal, 4500 W^+W^-, 4500 $\bar{t}t$ after precuts
- cut on $m_{T2}^{100} < 130$ GeV
- 1000 signal, 6000 bgd
- S/B ~ 0.15, signif ~ 10 σ

- BDT improves on just cutting on important variables
  - doubles S/B





conclusion

- lepton partner searches at LHC difficult when splitting with invisible particles is 30-50 GeV
- tough to beat electroweak backgrounds
- confront with boosted decision tree (BDT)

- can get large improvements
- BDT identifies the important variables and correlations

Mahalo!

Backup Slides

event topology

- exactly 1 μ^+ and 1 μ^-
  - muon threshold → $p_T > 3$ GeV (generator level)
- exactly 1 jet, not b-tagged
  - jet threshold → $p_T > 30$ GeV
- MET
- no hadronic τ–tag

tranching

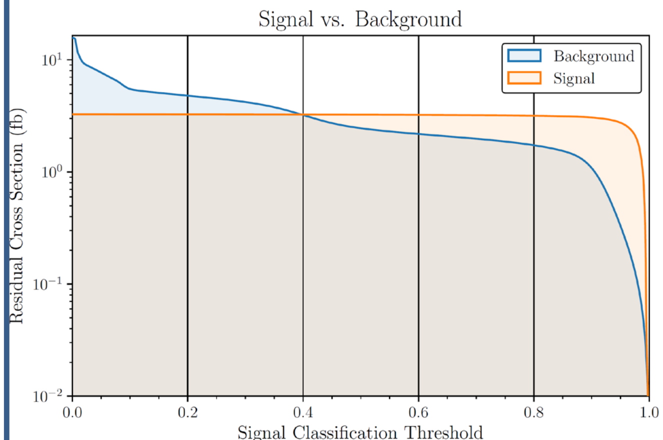
- variables to base tranching on …
- $\mu\mu$ jjj → $p_T^{\mu 1}$
- $\tau\tau$ jjj → $p_T^{\mu 1}$
- $\bar{t}t$ jj → → p_T^{t} (not $\bar{t}$)
- WW jj → $p_T^{\mu 1}$
- WZ jj → p_T^{Z}
- ZZ jj → decay one Z, tranche using p_T of remaining Z

- bounds on variable for each tranche …
- 0, 50, 100, 150, 200, 300, 400, 500, 750, 1000, 1500 GeV

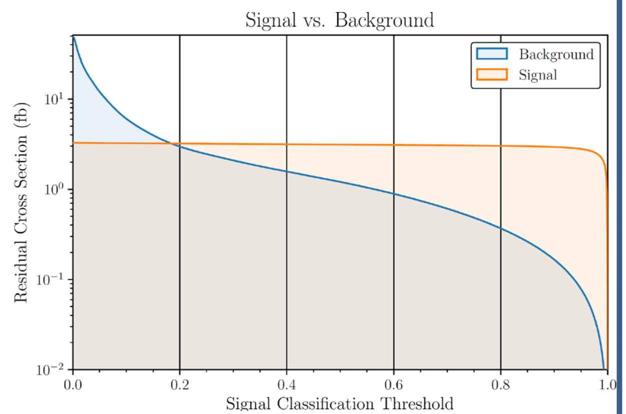
cos $\theta^*_{\mu\mu}$ and M_{T2}^{100}

$$\cos\theta^*_{\mu\mu} = \tanh\left|\Delta\eta_{\mu1\mu2} / 2\right|$$

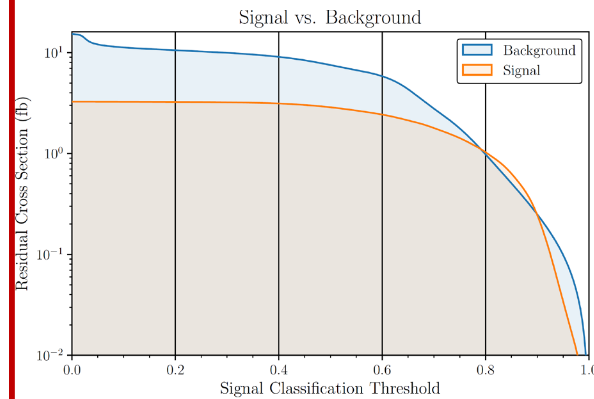
discriminating signal from bgd.



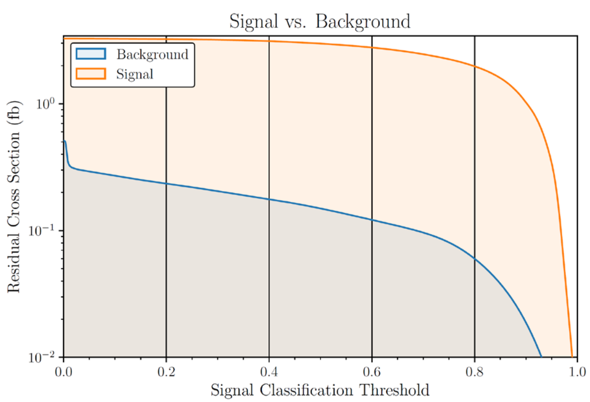
$\bar{\mu}\mu$



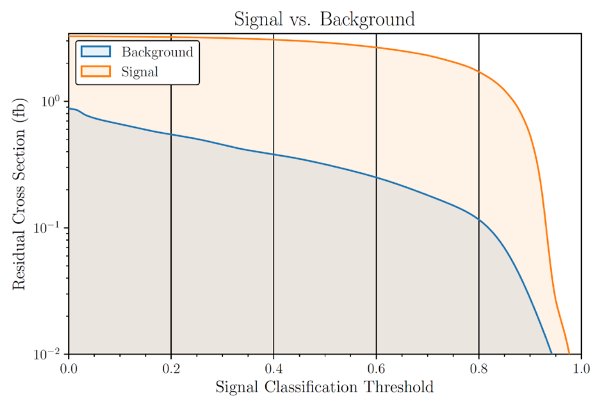
$\bar{\tau}\tau$



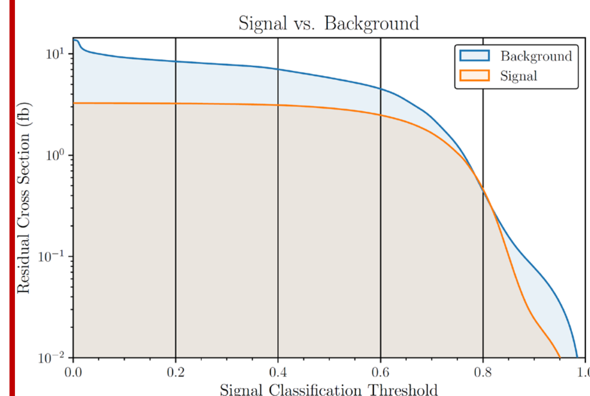
$\bar{t}t$



ZZ



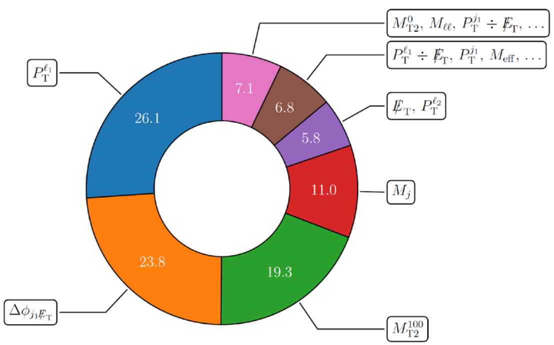
WZ



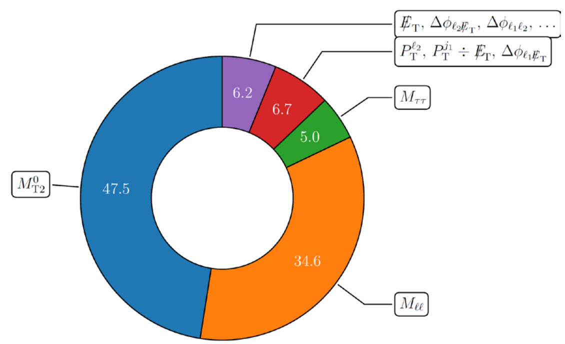
W^+W^-

gain

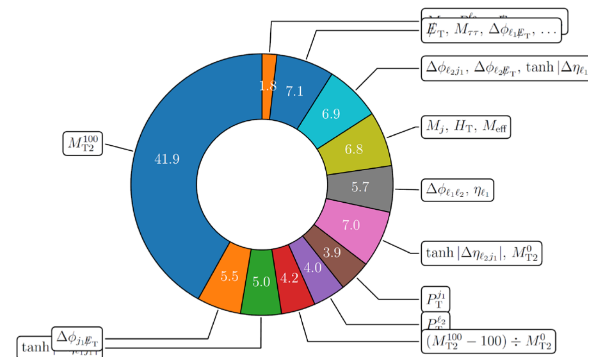
$\tilde{\ell}^{*}\tilde{\ell}jj$ (110/80) Signal vs. $\ell^{+}\ell^{-}jj$ Background



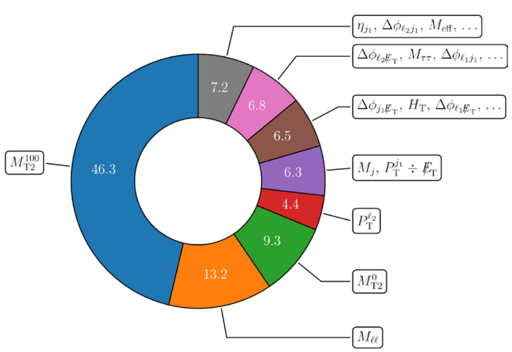
$\tilde{\ell}^{*}\tilde{\ell}jj$ (110/80) Signal vs. $\tau^{+}\tau^{-}jj$ Background



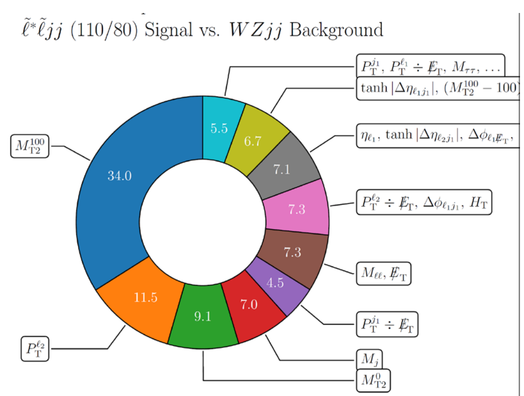
$\tilde{\ell}^{*}\tilde{\ell}jj$ (110/80) Signal vs. $t\bar{t}jj$ Background



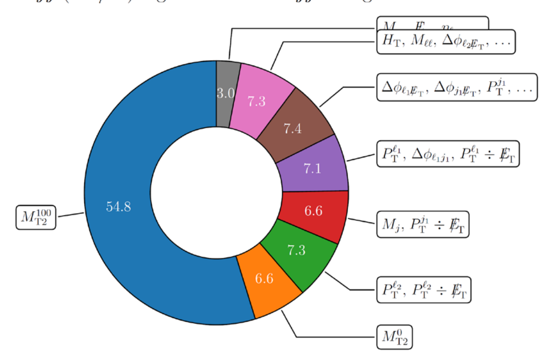
$\tilde{\ell}^{*}\tilde{\ell}jj$ (110/80) Signal vs. $ZZjj$ Background



$\tilde{\ell}^{*}\tilde{\ell}jj$ (110/80) Signal vs. $WZjj$ Background



$\tilde{\ell}^{*}\tilde{\ell}jj$ (110/80) Signal vs. $W^{+}W^{-}jj$ Background



sensitivity scan