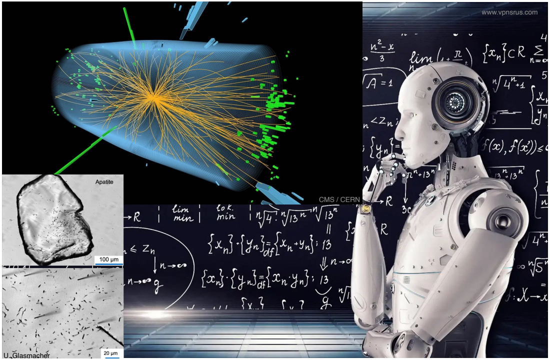
Using ML to search for scalar lepton partners at the LHC

Based on work with B. Dutta, T. Ghosh, A. Horne, J. Kumar, S. Palmer, P. Sandick, M. Snedeker and J. W. Walker, Phys. Rev. D 109, no.7, 075018 (2024) arXiv:2309.10197



Jožef Stefan Institute
Ljubljana, Slovenia

SMASH
machine learning for science and humanities postdoctoral program

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101081355.

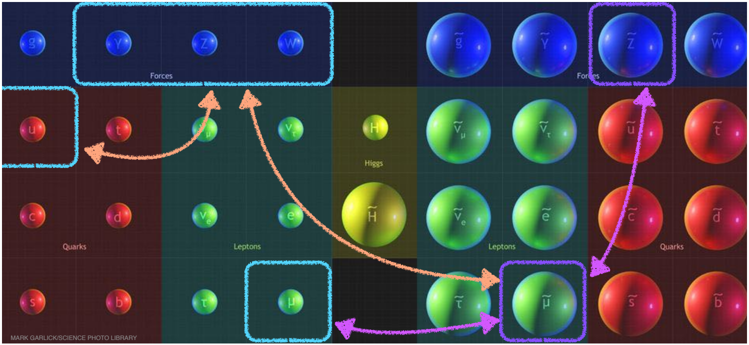
What do we really know about Dark Matter?

What we (typically) assume

- Non-electromagnetically interacting particle
- Must be cold and stable
- Not in the Standard Model

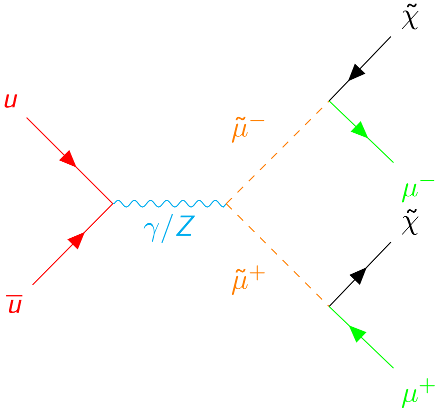
Weakly interacting massive particle

- Produced in early universe
- Weak scale mass for relic density
- Predict interactions with SM in e.g. **charged mediator models**

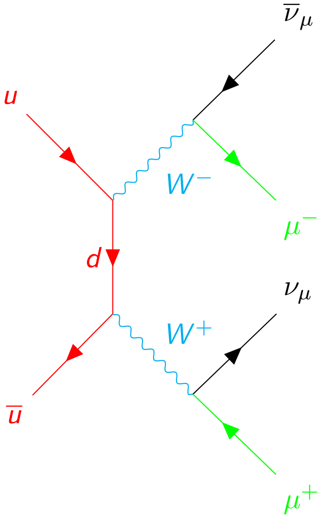


Charged mediator signals in proton collisions at LHC

- Protons composed of quarks, gluons
- SM interactions with mediators
- Decay to leptons and **invisible** DM



SM background process



100x-1000x more likely

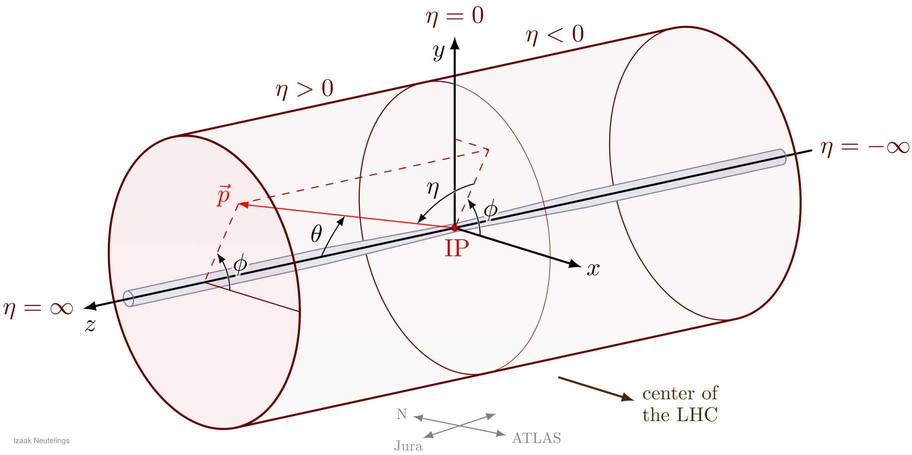
Simulation chain for new physics at LHC

Define Model
- Particle spectrum
- Feynman rules
- Parameter space

Parton Level
- Matrix elements
- Event simulation
- Sufficient statistics?

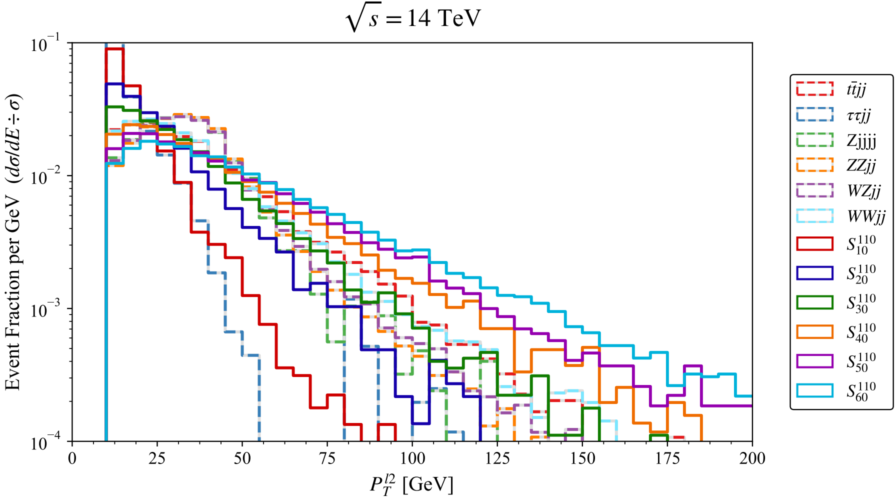
Hadron Level
- Parton showering
- Hadronization
- Jet Matching

Detector Level
- Fast simulation
- Reconstruction
- Phase space

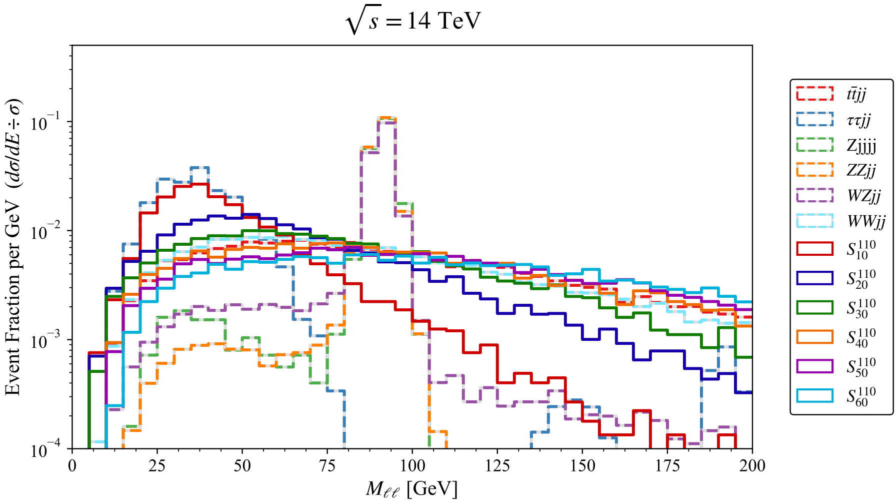


Izaak Neutelings

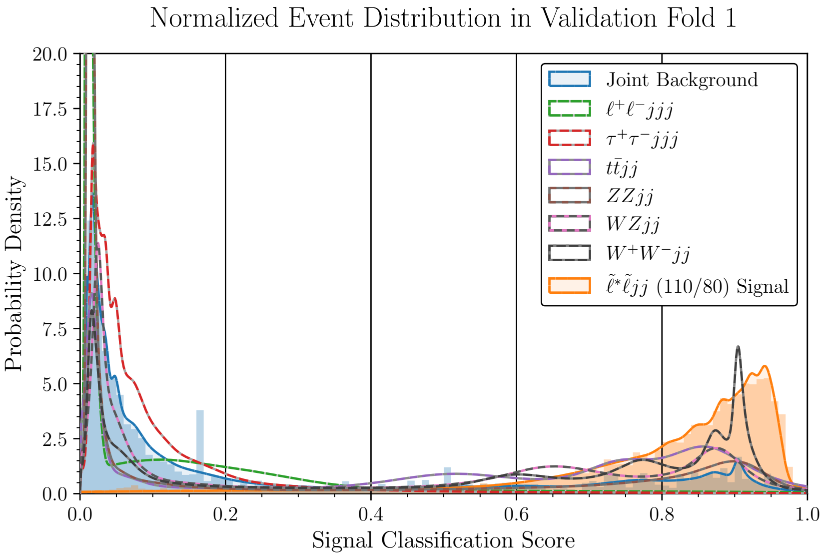
Search $pp \to \ell^+\ell^- \not{E}_T$ phase space for charged mediators



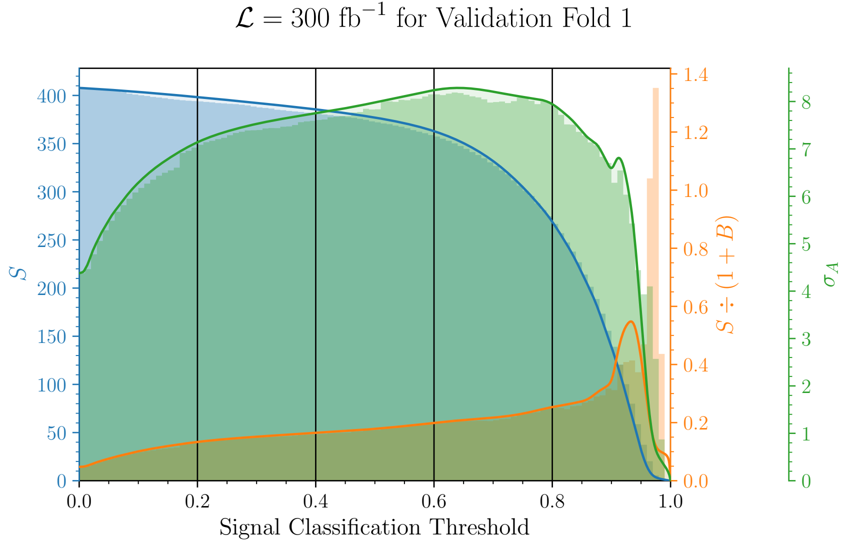
Construct higher level features



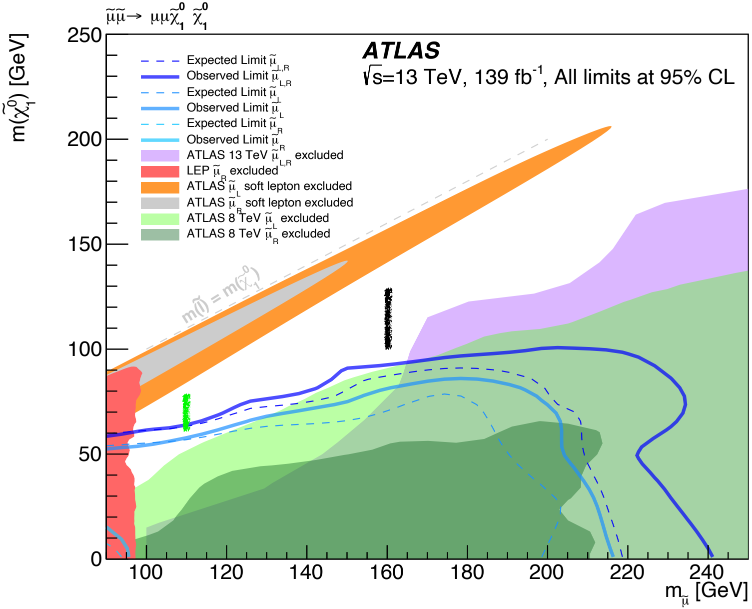
After precuts, train BDT to classify signal and background



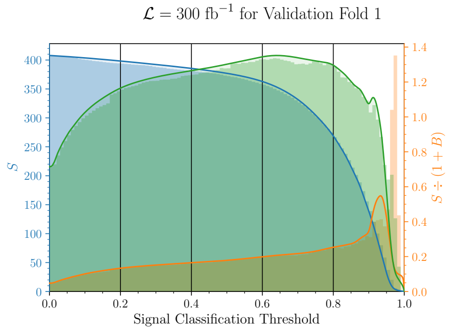
Significance $\gtrsim 6\sigma$ for $m_{\tilde{\mu}_L} = 110\,\mathrm{GeV}$ and $m_\chi = 80\,\mathrm{GeV}$

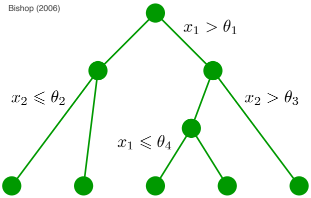


Discover $m_{\tilde{\mu}_L} \gtrsim 110\,\mathrm{GeV}$ and exclude $m_{\tilde{\mu}_L} \lesssim 160\,\mathrm{GeV}$



Using ML to probe the nature of Dark Matter





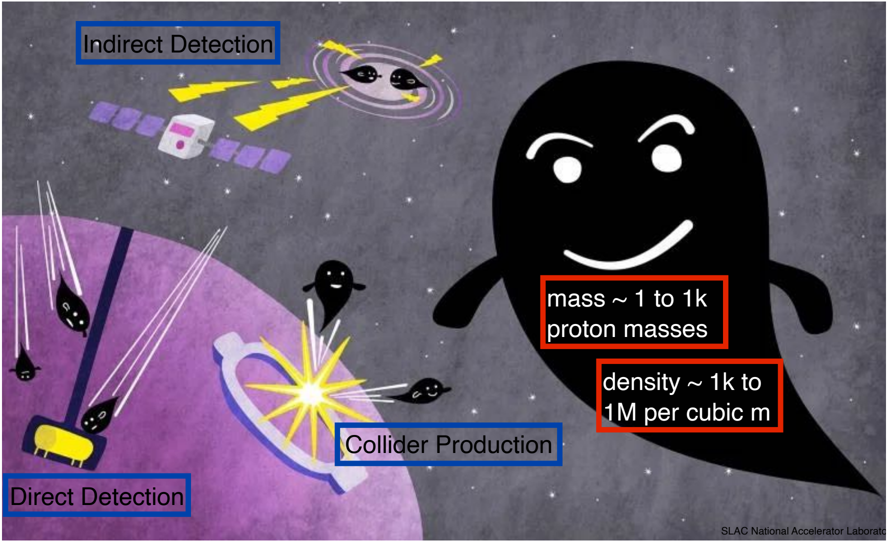
Improve on cut-and-count analysis for scalar lepton searches at LHC

- Sensitivity to $m_{\tilde{\mu}_L} \lesssim 160\,\mathrm{GeV}$
- Systematics $S/B \sim 0.15 - 0.40$
- Kinematic tranching to increase sampling at tails of distributions
- Precuts to bring signal and backgrounds (closer) to parity

Additional ML techniques

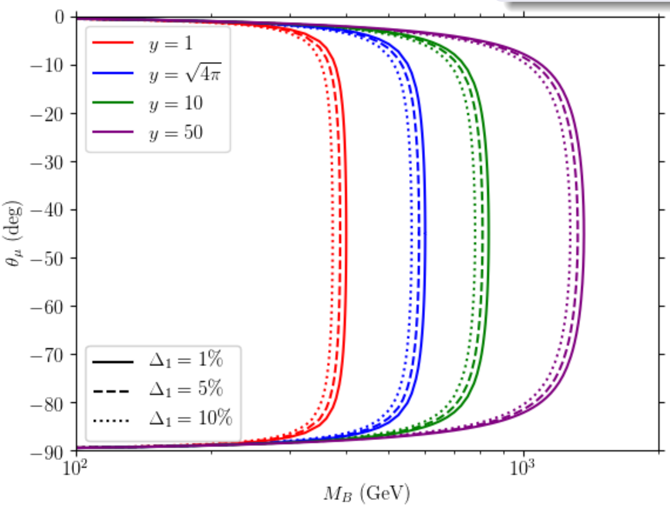
- Deep neural networks
- Convolutional neural networks
- Adversarial neural networks

Look for WIMPs interacting around us or produce them

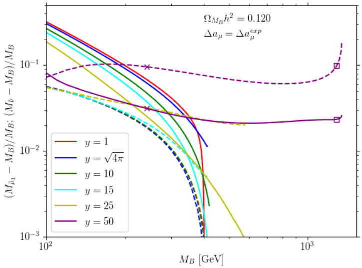


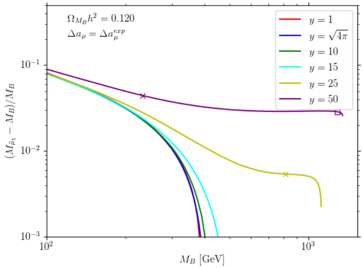
Motivate/constrain parameter space by requiring $g_\mu - 2$

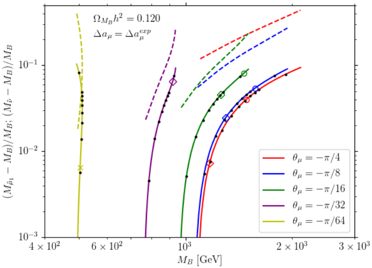
$$y = (m^2_{\tilde{\mu}_2} - m^2_{\tilde{\mu}_1}) \sin(2\theta_{\tilde{\mu}})/(4m^2_W)$$

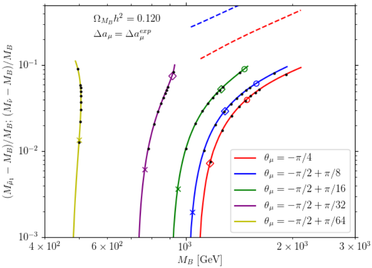


Parameter space for Δa_μ and $\Omega_{\rm DM}$ from co-annihilation

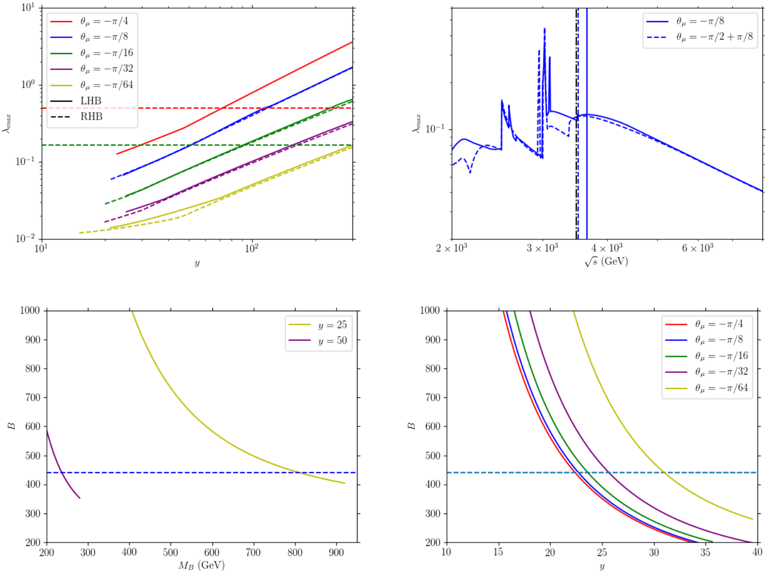






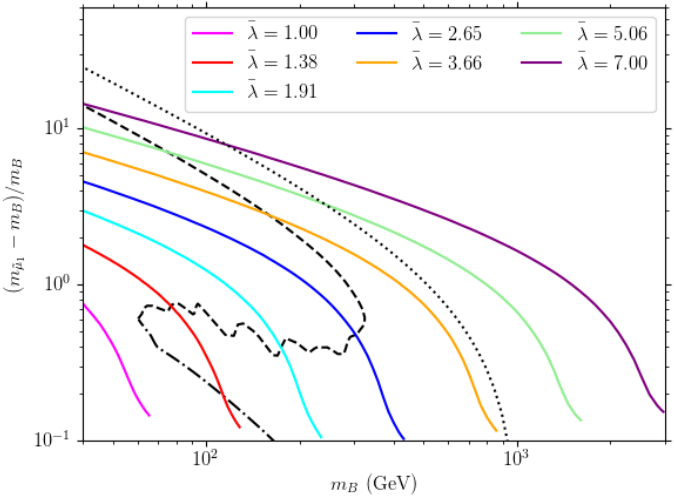


Perturbative unitarity and electroweak vacuum stability

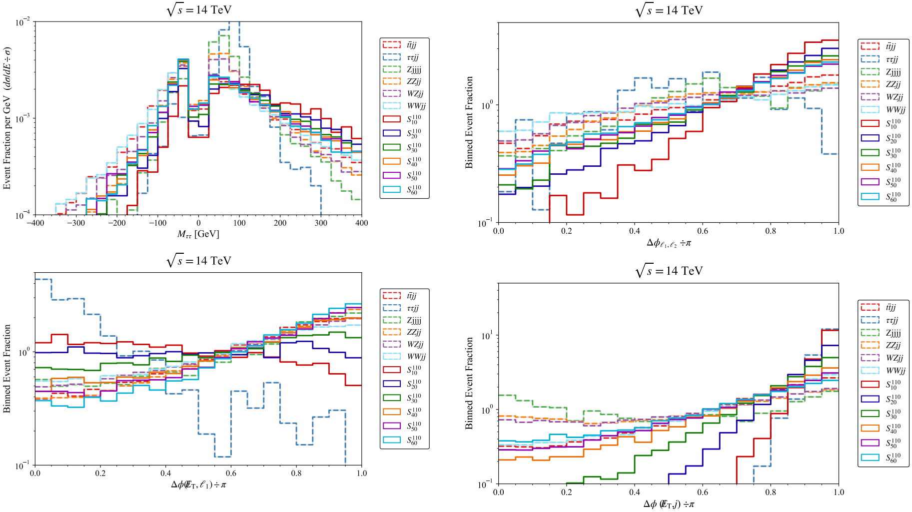


Generalize DM couplings to get $\Omega_{\rm DM}$ from DM annihilation

Fix $\theta_{\tilde{\mu}} = -45°$, $m_{\tilde{\mu}_2}/m_{\tilde{\mu}_1} = 1.25$

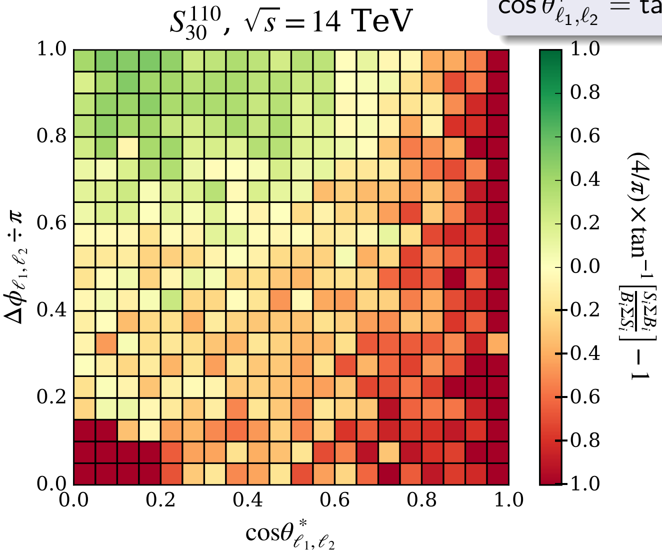


More kinematic distributions

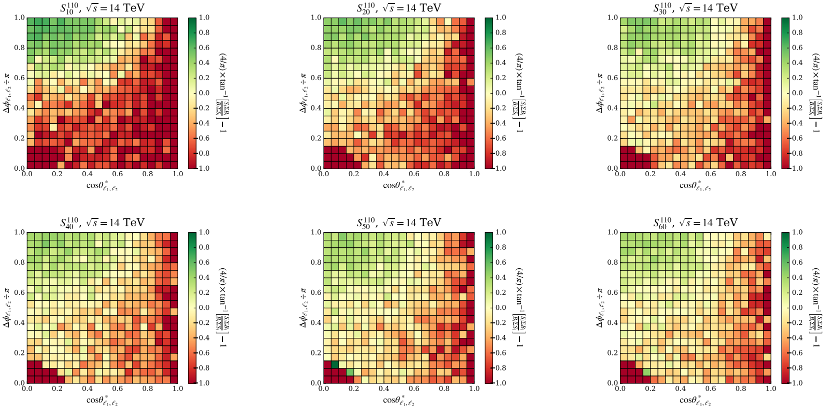


Consider multidimensional representations

$$\cos\theta^*_{\ell_1,\ell_2} = \tanh(\Delta\eta_{\ell_1\ell_2}/2)$$



2D histograms of angular kinematic distributions

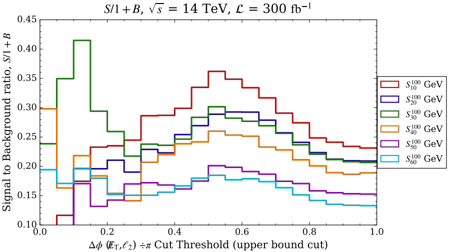


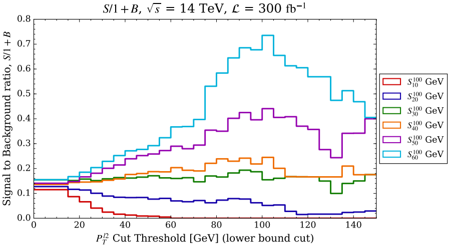
Residual cross sections (fb) for primary and secondary cuts

Primary Selection	$t\bar{t}jj$	$\tau\tau jj$	$Zjjjj$	$WWjj$	S_{30}^{110}	S_{40}^{110}
Matched Production	6.1×10^5	5.6×10^4	5.2×10^7	9.5×10^4	1.9×10^2	1.9×10^2
τ-veto	5.4×10^5	3.0×10^4	5.1×10^7	8.9×10^4	1.9×10^2	1.9×10^2
OSSF muon	3.5×10^3	4.3×10^2	6.0×10^5	5.1×10^2	8.1×10^1	8.8×10^1
exactly 1J $P_T > 30$	6.6×10^2	2.6×10^2	7.1×10^4	1.1×10^2	1.6×10^1	1.7×10^1
Jet b-veto	1.9×10^2	2.5×10^2	7.0×10^4	1.1×10^2	1.6×10^1	1.7×10^1
$\not{E}_T > 30$ GeV	1.6×10^2	1.8×10^2	8.9×10^3	9.2×10^1	1.3×10^1	1.4×10^1

Secondary Selection	$t\bar{t}jj$	$\tau\tau jj$	$Zjjjj$	$WWjj$	S_{30}^{110}	S_{40}^{110}
$m_{\ell\ell} \notin M_Z \pm 10$ GeV	1.4×10^2	1.8×10^2	6.2×10^2	7.9×10^1	1.1×10^1	1.2×10^1
$\cos\theta^*_{\ell_1,\ell_2} < 0.5$	8.1×10^1	1.6×10^2	4.7×10^2	4.5×10^1	8.0×10^0	9.0×10^0
$m_{\tau\tau} > 125$ GeV	2.7×10^1	2.3×10^1	8.7×10^1	1.4×10^1	3.6×10^0	3.9×10^0
$\not{E}_T > 125$ GeV	2.9×10^0	6.6×10^{-1}	0	2.3×10^0	6.6×10^{-1}	7.1×10^{-1}
Jet $P_T > 125$ GeV	1.1×10^0	6.6×10^{-1}	0	1.7×10^0	5.2×10^{-1}	4.6×10^{-1}

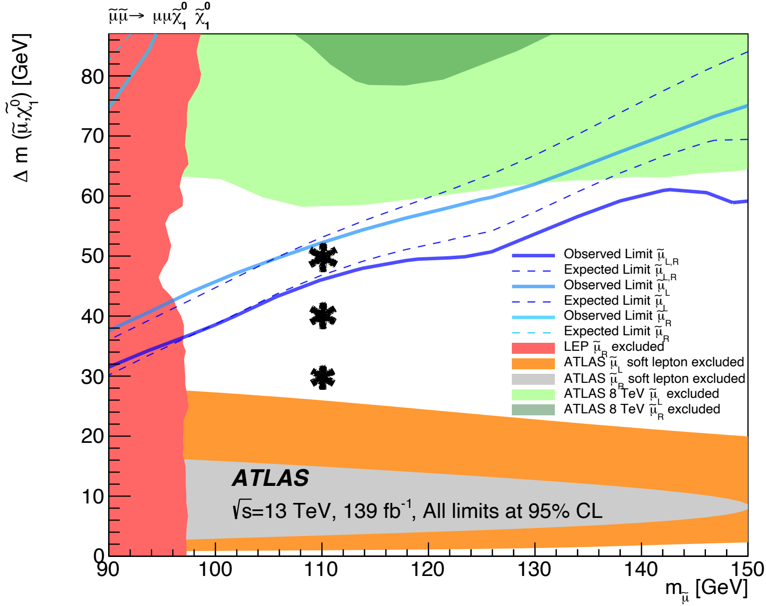
Tertiary cuts for optimized for intermediate mass gaps



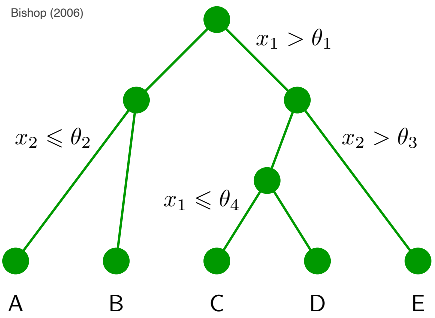


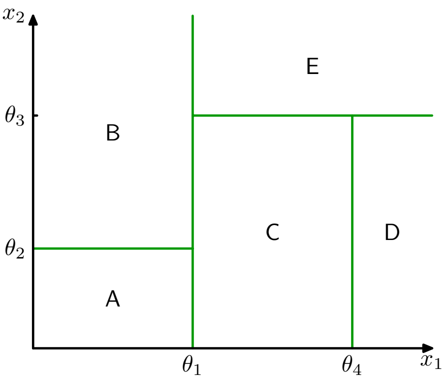
Tertiary Selection	$t\bar{t}jj$	$\tau\tau jj$	$WWjj$	S_{30}^{110}	S_{40}^{110}
$\Delta\phi(\ell_1, \ell_2) \div \pi > 0.5$	1.1×10^{0}	5.5×10^{-3}	1.3×10^{0}	4.4×10^{-1}	4.1×10^{-1}
$\Delta\phi(\not{E}_T, \ell_1) \div \pi < 0.6$	4.8×10^{-1}	5.5×10^{-3}	9.0×10^{-1}	3.3×10^{-1}	3.0×10^{-1}
$\Delta\phi(\not{E}_T, \ell_2) \div \pi < 0.6$	1.8×10^{-1}	0	5.1×10^{-1}	2.2×10^{-1}	2.0×10^{-1}
Events at $\mathcal{L} = 300\ \mathrm{fb}^{-1}$	52.8	0	151.7	66.0	60.0
$S \div (1 + B)$	-	-	-	0.30	0.27
$S \div \sqrt{1 + B}$	-	-	-	4.4	4.0

Project $\sim 3\sigma$ sensitivity to $m_{\tilde{\mu}_L} = 110\,\mathrm{GeV}$ at $\mathcal{L} = 300\mathrm{fb}^{-1}$



Trees partition final state phase space into decision regions





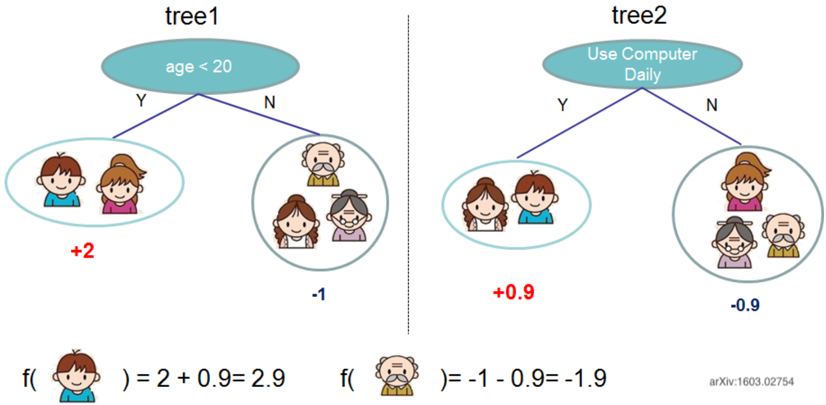
Split leaf nodes to minimize objective

$\text{obj} = \sum_{\text{data}} \ell(y_i, \hat{y}_i) + \omega(f)$, with $\hat{y}_i = f(\boldsymbol{x}_i)$ and regularization ω

Define tree by score on each leaf

$f(\boldsymbol{x}) = \boldsymbol{w}_{q(\boldsymbol{x})}$, vector of scores $\boldsymbol{w}$ with q assigning each $\boldsymbol{x}_i$ to a leaf

Ensembles of trees built iteratively using gradient boosting



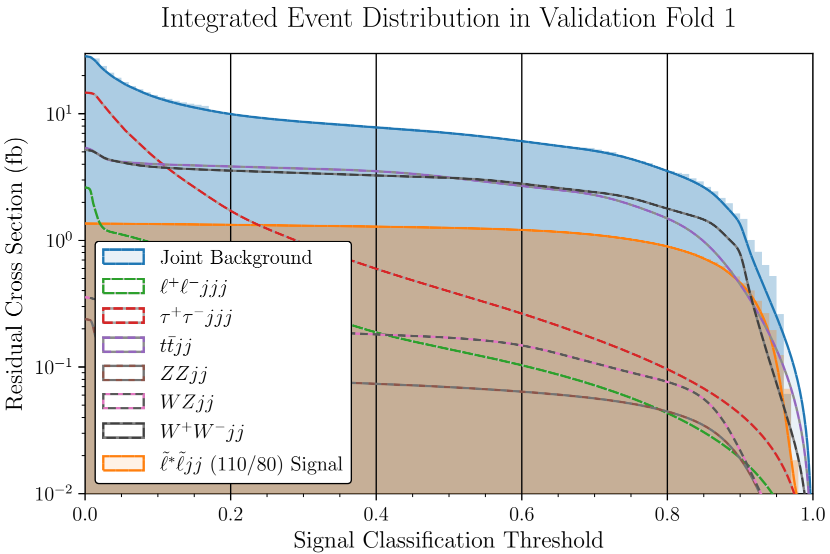
$$\hat{y}_i^{(t)} = \sum_{\text{trees}} f_j(\boldsymbol{x}_i) = \hat{y}_i^{(t-1)} + f_t(\boldsymbol{x}_i)$$

$$\text{obj} = \sum_{\text{data}} \ell(y_i, \hat{y}_i^{(t)}) + \omega(f_t)$$

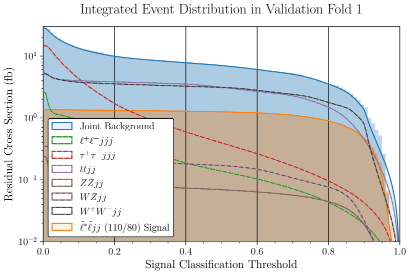
$$\Delta\ell \approx \sum_{\text{data}} \left[g_i f_t(\boldsymbol{x}_i) + h_i f_t^2(\boldsymbol{x}_i)/2 \right]$$

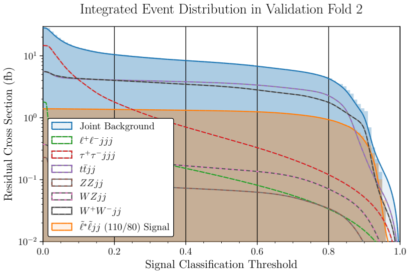
$$g_i, h_i = \partial^{1,2}_{\hat{y}_i^{(t-1)}} \ell(y_i, \hat{y}_i^{(t-1)})$$

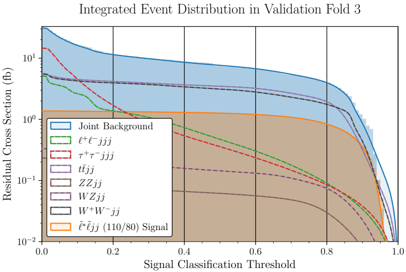
After precuts, train BDT to classify signal and background



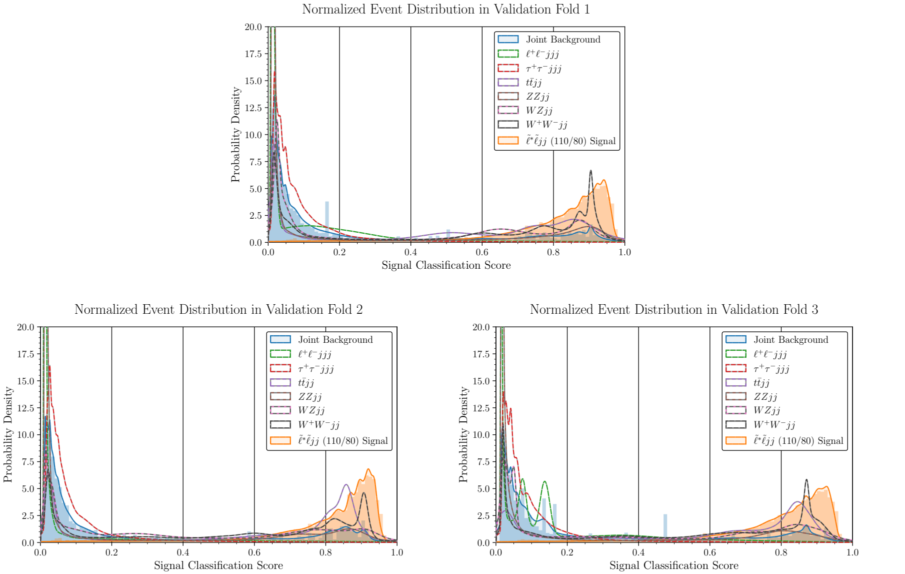
Additional folds for event distributions





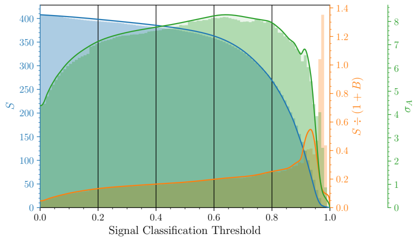


Additional folds for probability distributions

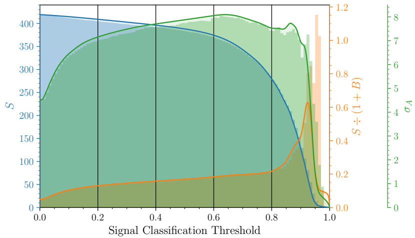


Additional folds for summary statistics

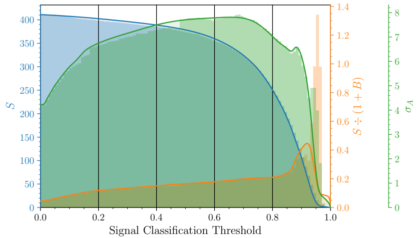
$\mathcal{L} = 300\ \text{fb}^{-1}$ for Validation Fold 1



$\mathcal{L} = 300\ \text{fb}^{-1}$ for Validation Fold 2

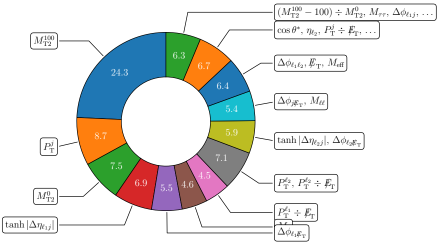


$\mathcal{L} = 300\ \text{fb}^{-1}$ for Validation Fold 3

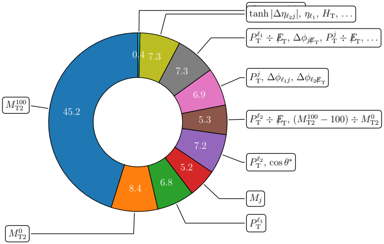


Features most important for BDT rejecting $t\bar{t}$, W^+W^-

$t\bar{t}jj$ Background in Training Fold 1



W^+W^-jj Background in Training Fold 1



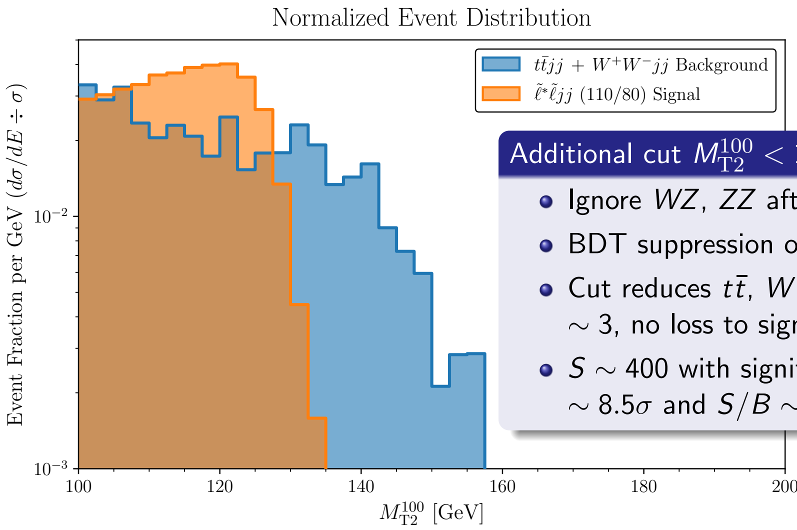
Relative contributions to reduction in ensemble objective function

- Sensitive to number of nodes
- Events through those nodes
- Weights carried by those events

M_{T2}^{100} dominates total gain for BDT trained for individual $t\bar{t}$, W^+W^-

Minimal mass of pair-produced parent to decay into $\ell + (\chi)(\nu)$ assuming $m_{\chi,\nu} = 100\,\mathrm{GeV}$

M_{T2}^{100} distribution for signal vs. $t\bar{t}$, W^+W^- after precuts

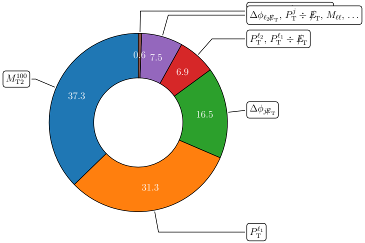


Additional cut $M_{\mathrm{T2}}^{100} < 130\,\mathrm{GeV}$

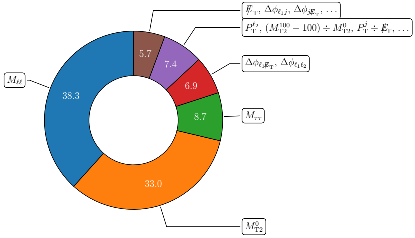
- Ignore WZ, ZZ after precuts
- BDT suppression of $\mu\mu$, $\tau\tau$
- Cut reduces $t\bar{t}$, W^+W^- by ~ 3, no loss to signal events
- $S \sim 400$ with significance $\sim 8.5\sigma$ and $S/B \sim 0.18$

Additional donut plots

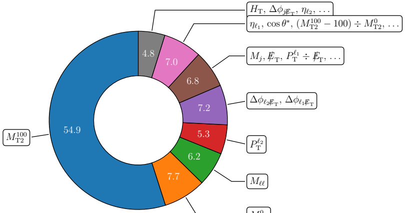
$\ell^+\ell^- jjj$ Background in Training Fold 1



$\tau^+\tau^- jjj$ Background in Training Fold 1



$ZZjj$ Background in Training Fold 1



$WZjj$ Background in Training Fold 1