ML for BSM

A challenge and an opportunity

IOP, April 2024

Tomasz Procter
University of Glasgow

Outline

- **Challenge:** BSM searches increasingly depend on NN/BDT discriminants
  - Case Study 1: Vector-Like Quarks (VLQs) - MCBOT
    - Pheno study
    - ATLAS internal study
  - Case Study 2: SUSY
- **Opportunity:**
  - BSM signal grid reweighting

Challenge: reinterpretation

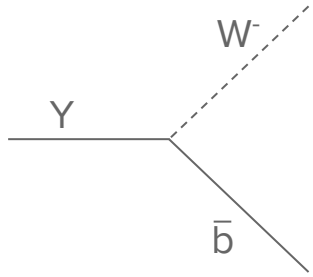
Neural Network Preservation/Re-interpretation

- New and exciting topic!
  - Pheno recasting tools need to run models against a fast, simple preservation of the analysis
  - More and more analyses rely on neural nets
  - How do pheno/recasting community keep up?
- ATLAS SUSY group have largely led the way -
  - Published ONNX files for some of their analyses.
- Large topic of discussion at the recent Reinterpretation Forums[1]:
  - Talks discussed experienced in variety of recasting tools.
  - Status summarised in the “Les Houches Guidelines”[2]
- Neural Nets typically trained on data that has gone through full sim:
  - How valid is its use in Rivet (and similar - GAMBIT, MadAnalysis, CheckMate...), which use some form of smearing fastsim - Delphes or similar ?
  - Which variables are particularly affected?
- New ONNX interfaces in Rivet and Gambit

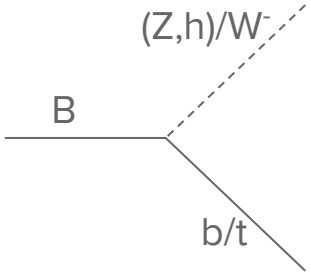
[1] https://indico.cern.ch/event/1197680/
[2] arXiv:2312.14575v2

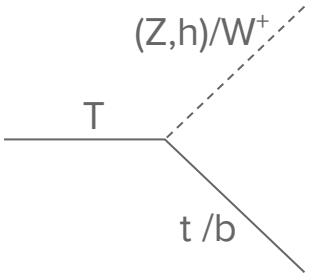
VLQs - MCBOT

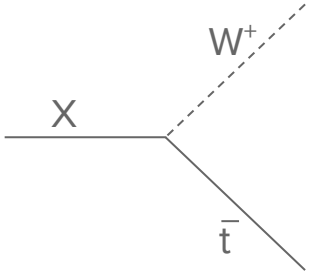
VLQs:
MCBOT - “Multi-Class Boosted Object Tagger”



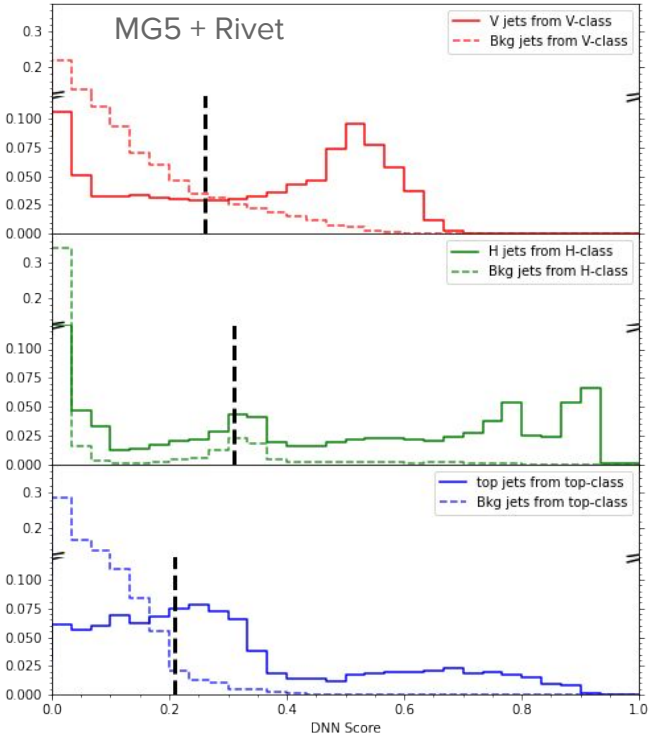
- Started off as an external reinterpretation of arXiv:2210.15413 and arXiv:1806.01762 (Atlas VLQ searches)
- Designed to tag reclustered (RC) jets as originating from Vector(W/Z)/Higgs/Top for use in VLQ analyses
  - RC Jets are large radius jets made by reclustering anti-kt R=0.4 jets.
- DNN with 18 inputs...:
  - RC jet pT, mass, number of subjets.
  - pT, η, ϕ, E, *b*-tag for 3 leading (highest pT) subjets
  - N.b. *b*-tag is a potentially complicated input
- ... and 4 outputs:
  - Probability of originating from Vector/Higgs/Top/Background
- Trained on variety of VLQ jets + QCD Multijet background.

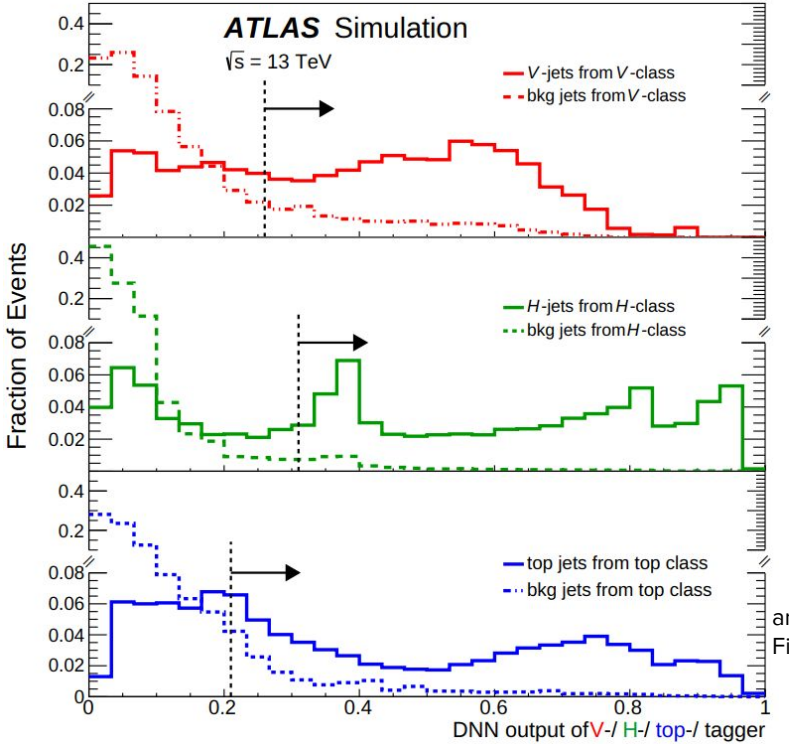






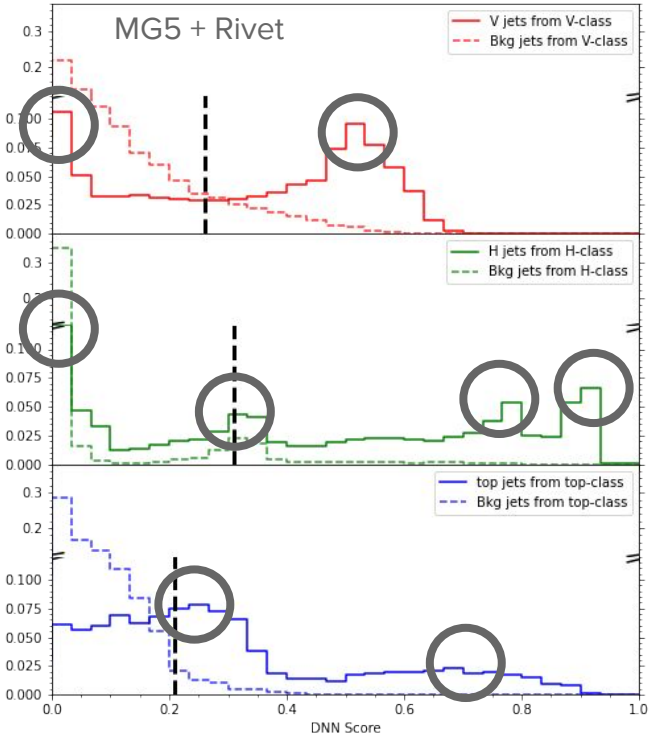
Replicating MCBOT validation plots - 2022 DNN score plots

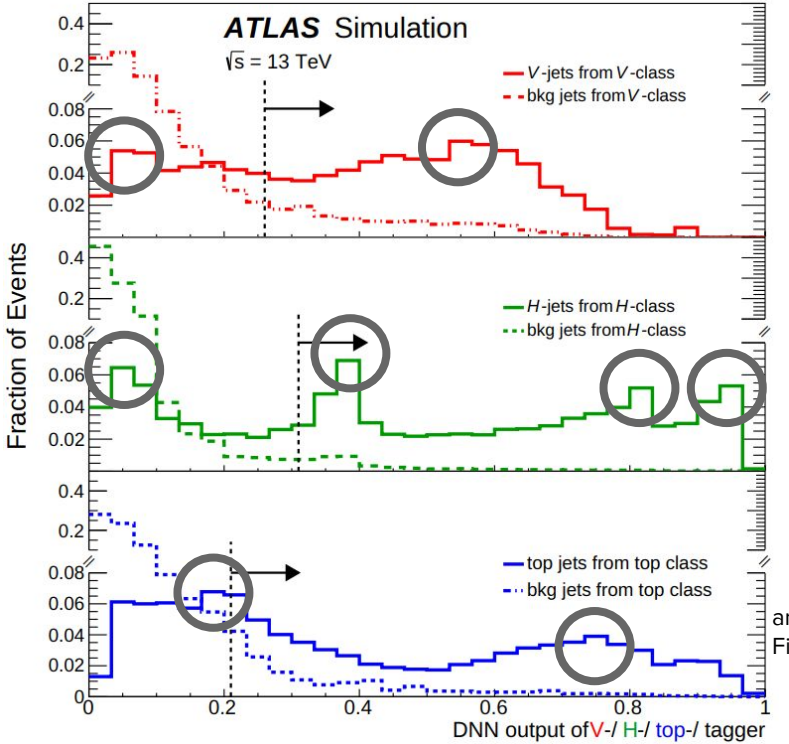




arXiv:2210.15413
Fig 3.

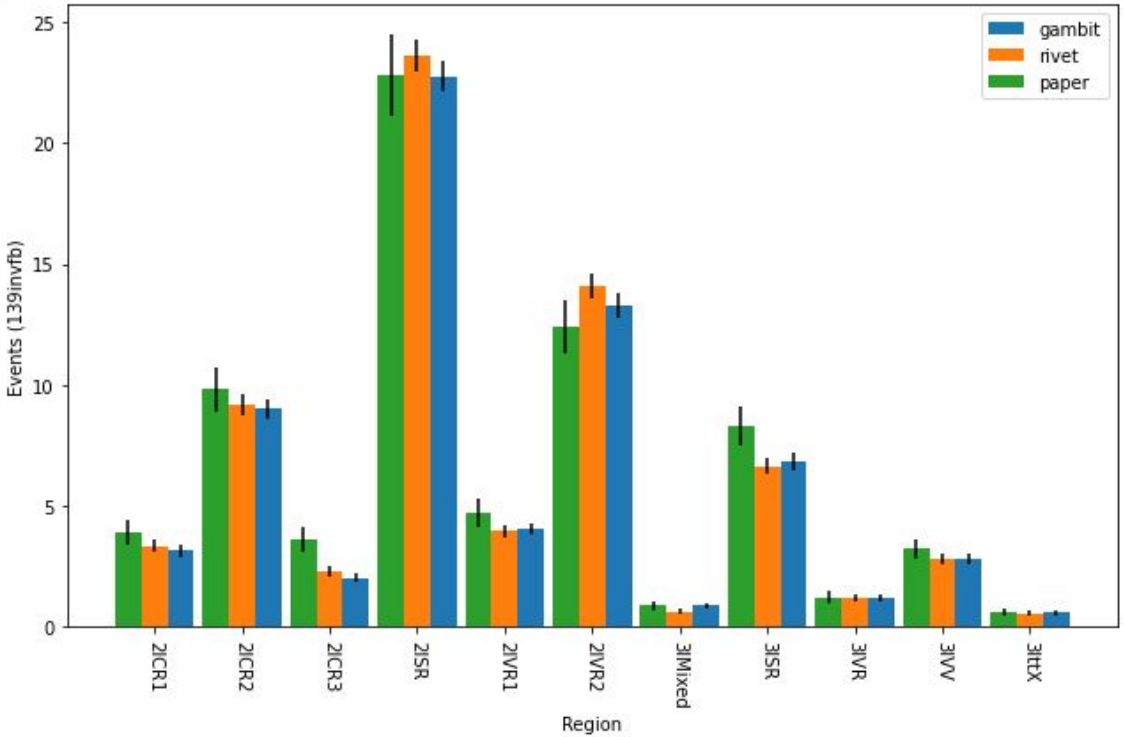
Replicating MCBOT validation plots - 2022 DNN score plots





arXiv:2210.15413
Fig 3.

Replicating MCBOT validation plots - 2022 SR plots



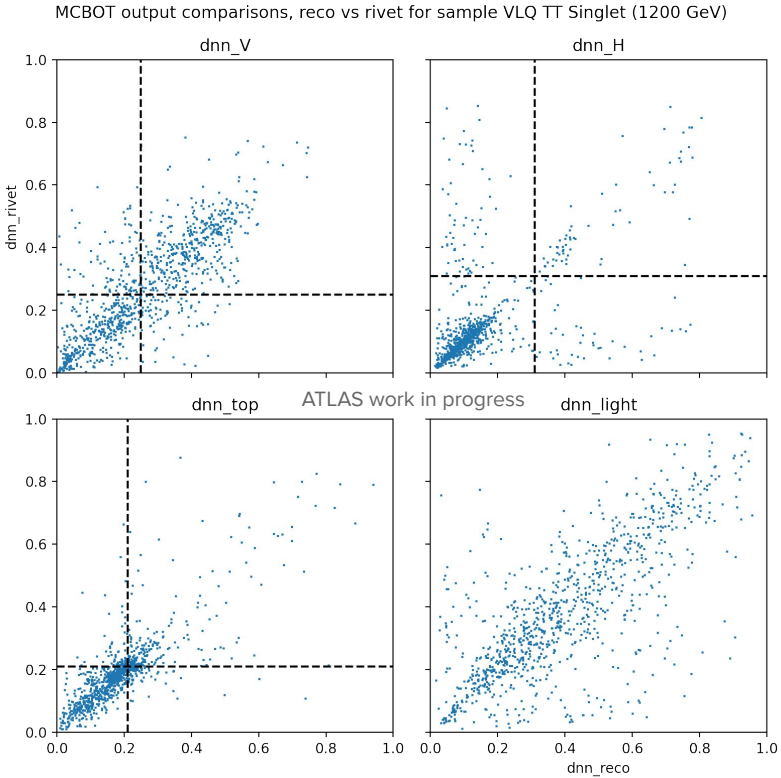
Truth-level study

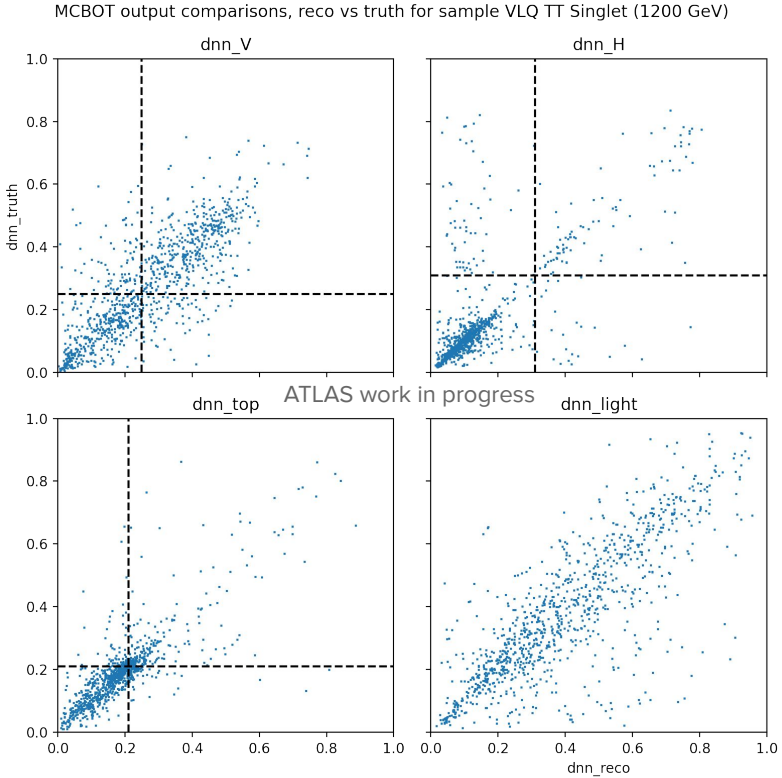
Quick comments on Truth-level study

- We compare truth/rivet to fully reconstructed ATLAS data:
  - Study carried out on very similar but more recent sample - can't promise all the triggering/calibration/etc is identical, but it will be **close**.
- Using cuts that went into NN training/validation, not the analysis signal cuts.
  - Better stats
  - Easier to compare to NN plots in the paper
- Good results - clearly Gaussian around y = x
- Truth (parton level) and smeared/emulated both perform similarly.

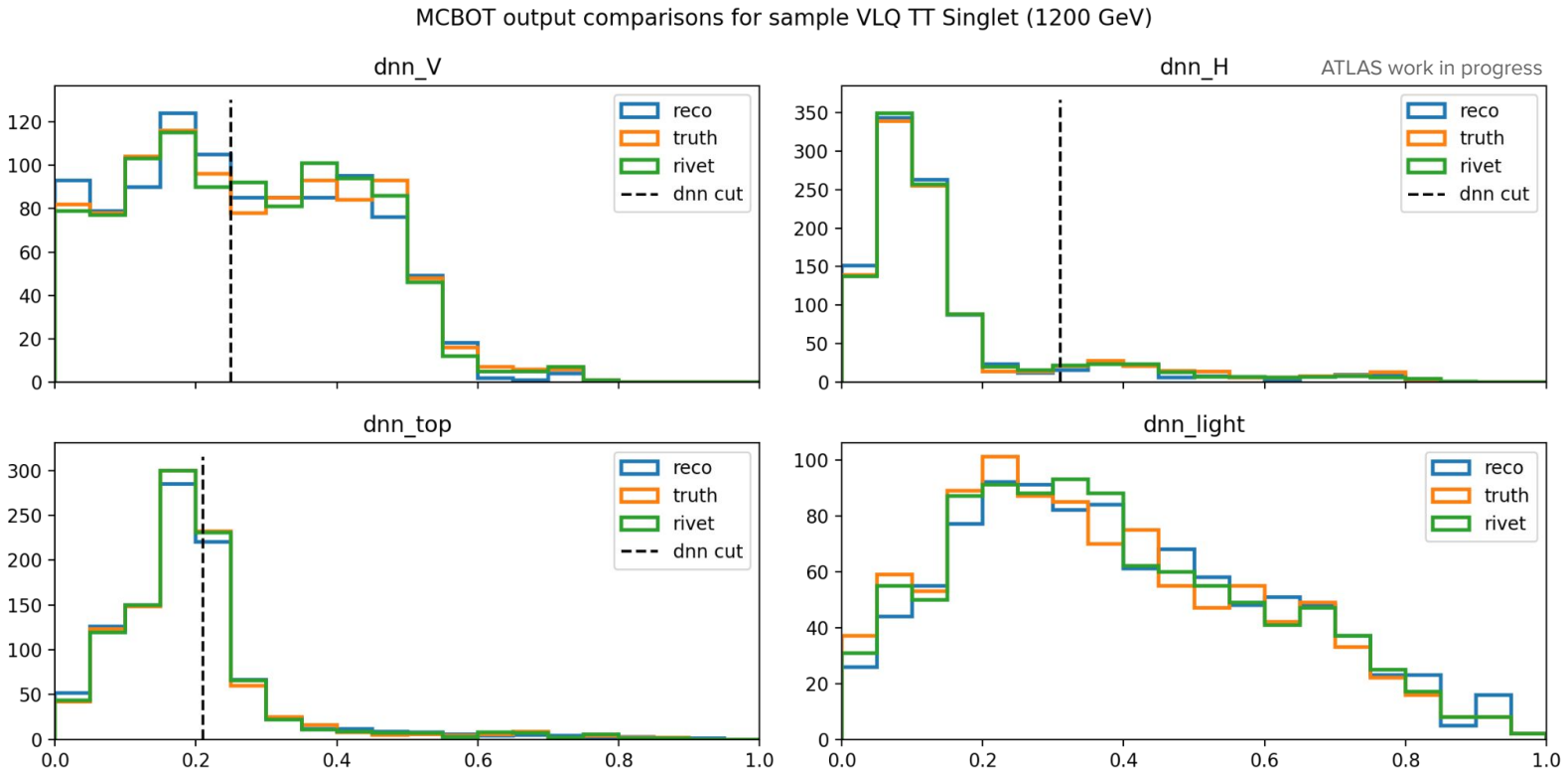
- This is the **best** possible test of if reusing the NN on truth data destroys any crucial information

Results (DNN output, VLQ)



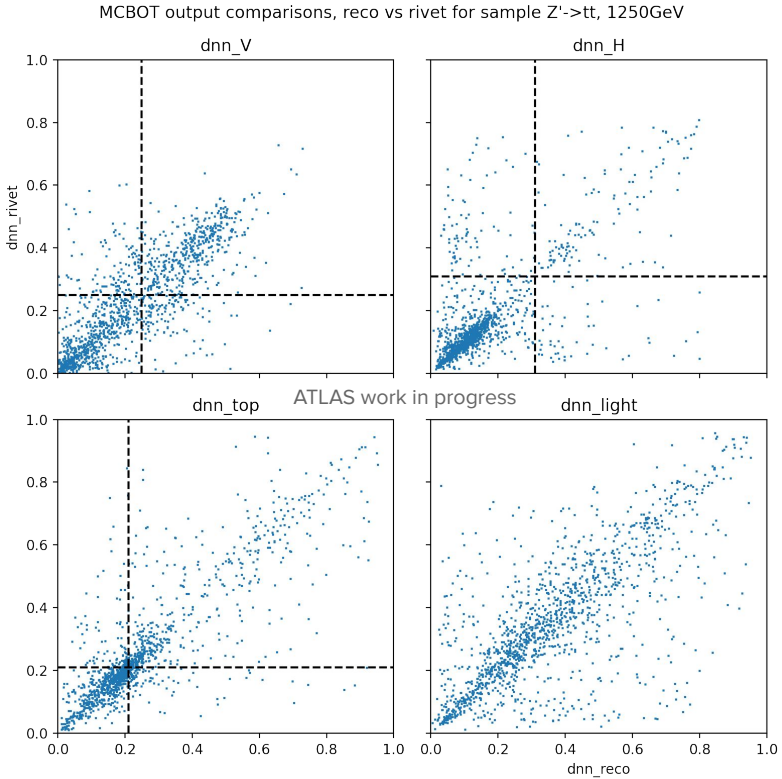


Results (DNN output binned, VLQ)



N.b “Rivet” is truth level with detector emulation

Results (DNN output Z' model),



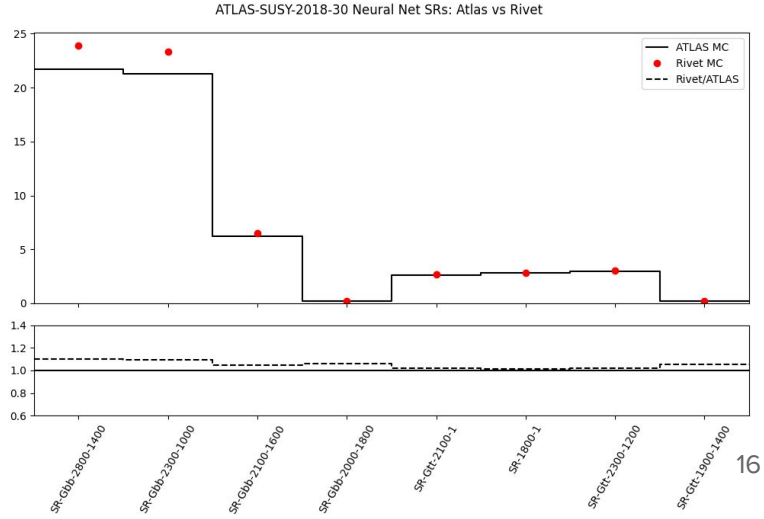
Reinterpreting SUSY-2018-30

SUSY-2018-30

- ATLAS search for SUSY in a final state with 3 b-jets, used a DNN.
  - Made public via SimpleAnalysis (script was incredibly helpful!)
- Became the benchmark test for reinterpretation tools (Rivet, Gambit, MA5, CheckMATE, ++)
- Required a little bit of extra development inside Rivet:
  - pT dependent b-tagging efficiencies
  - Improved Jet and electron reco-emulation.
- Small things - e.g. ϕ convention (0->2π vs -π -> π) can break everything:
  - Good documentation is essential!
- Good, reliable results in Rivet for both NN and Cut'n'count signal regions

Gbb Signal Model cutflow, (hepdata)

Cut	Paper	Rivet
0-lep	80.0	83.7
$\Delta\phi^{4j}_{min} \geq 0.6$	52.5	54.6
2800-1400 NN Cut	21.7	23.9
2300-1000 NN Cut	21.3	23.3
$\Delta\phi^{4j}_{min} \geq 0.4$	61.1	63.8
2100-1600 NN Cut	6.20	6.50
2000-1800 NN Cut	0.192	0.204



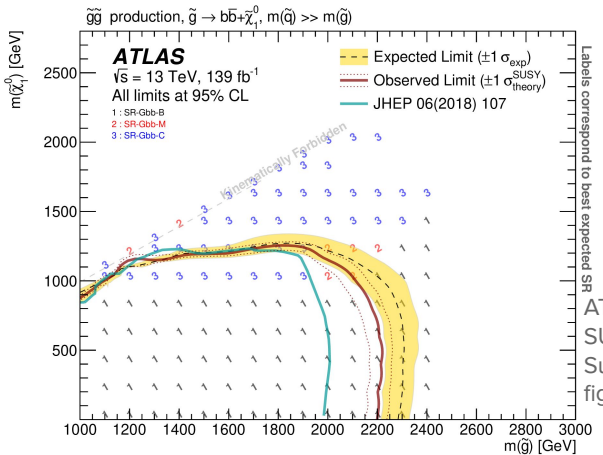
Opportunity

Reweighting BSM signal grids

- BSM searches often need big signal grids -
  - Computationally very expensive
- Possible solution: Generate a coarser grid, get to other points by reweighting.
- Enter the CARL method:
  - Use the classification score from a NN to obtain the likelihood ratio
  - Already used in some other contexts within ATLAS
  - Generate per-event weights - so ***all*** observables are available.

$$s(\vec{x}) = \frac{p_0(\vec{x})}{p_0(\vec{x}) + p_1(\vec{x})}$$

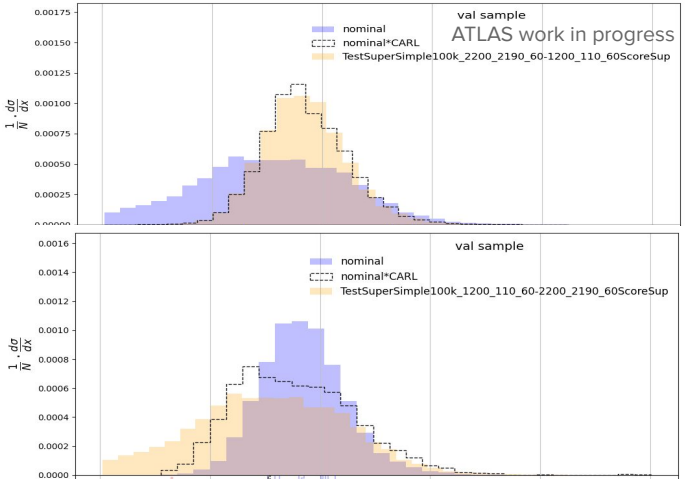
$$r(\vec{x}) = \frac{s(\vec{x})}{1 - s(\vec{x})}$$



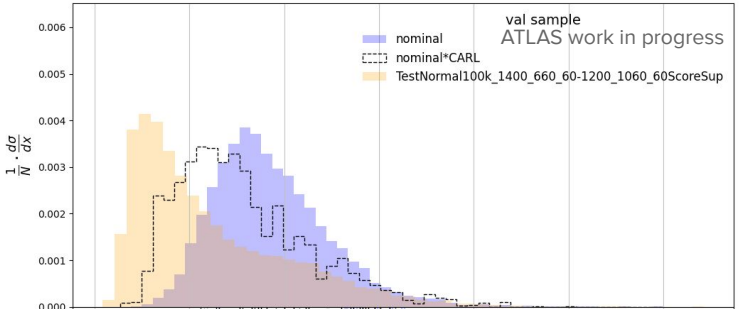
ATLAS SUSY-2018-30, Supplementary figure 5b

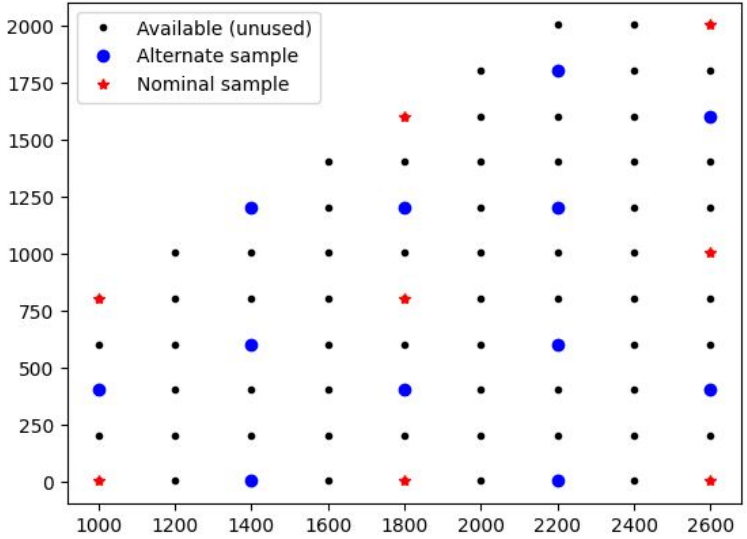
Reweighting BSM signal grids

- Ran initial tests using “point-to-point” reweighting
  - Highlight importance of covering the **entire domain**.



Reweighting wider -> narrower vs. narrower -> wider



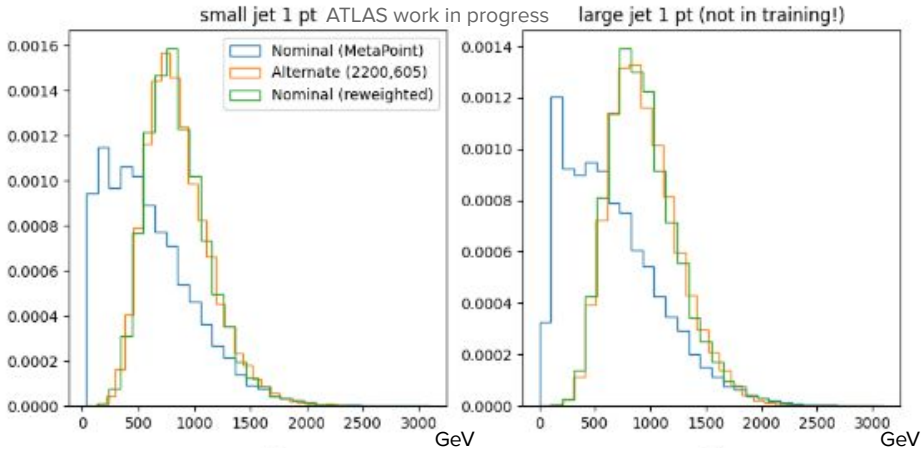


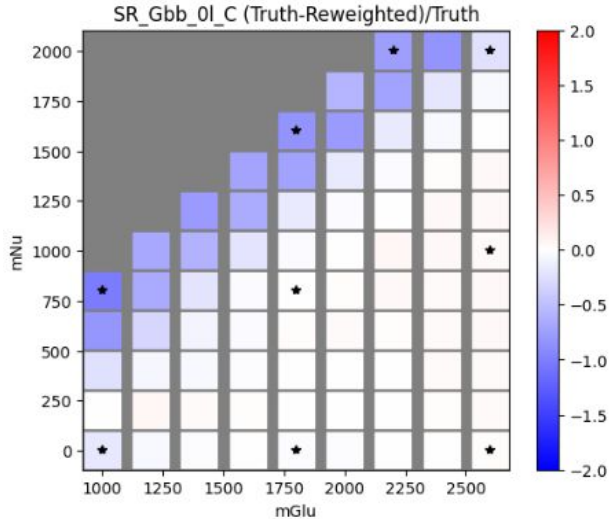
- Do we prioritise nearest neighbours?
- How do we ensure a broad distribution?
  - Use a nominal made up of points from ACROSS the distribution, let the network decide!

Promising early results

- Made a development workflow
  - Based on (Pythia->Rivet->ROOT)->CARL
- Accurate reweighting across a large signal grid
- Fewer than half the grid points involved in training.
- Good performance includes variables not used in training.
- Weakest performance in narrowly spiked observables and discrete variables

Reweighting nominal point to a point in the signal grid, comparing two observables





Reweighting from the nominal combination (black stars) to the entire grid, plotting the event count in one of the signal bins

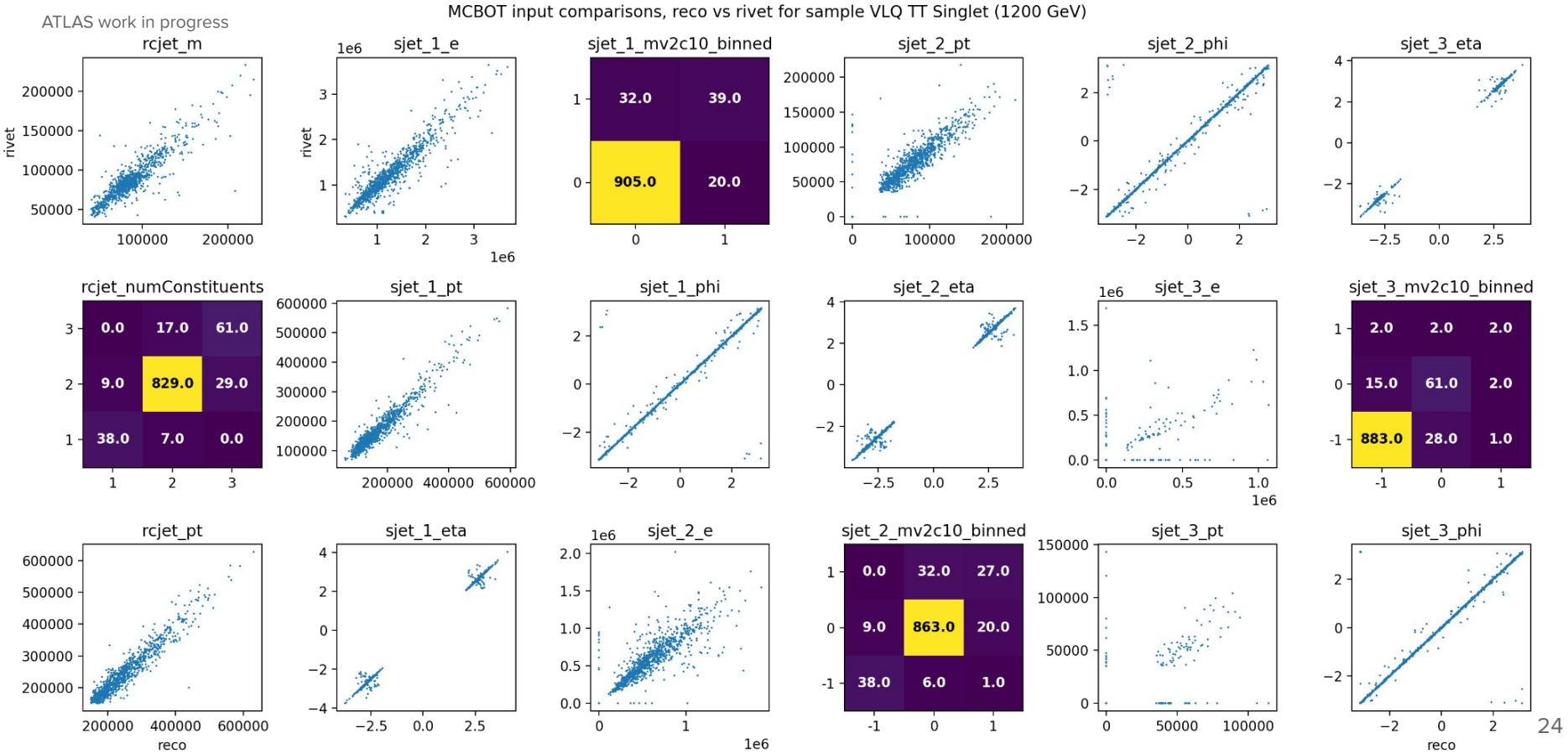
Conclusions

Conclusions

- Searches depending on ML are a challenge
  - But in most cases, one that can be overcome
    - ***If*** there is sufficient metadata, context and documentation
    - SimpleAnalysis/Rivet snippets are great for this!
    - See again the Les Houches guidelines
  - Reinterpretation tools are keen to try more examples
    - But the data (onnx/lwtnn files) needs to be public first!
- But - CARL based reweighting looks promising for reducing our computational load
  - May also have pheno applications?

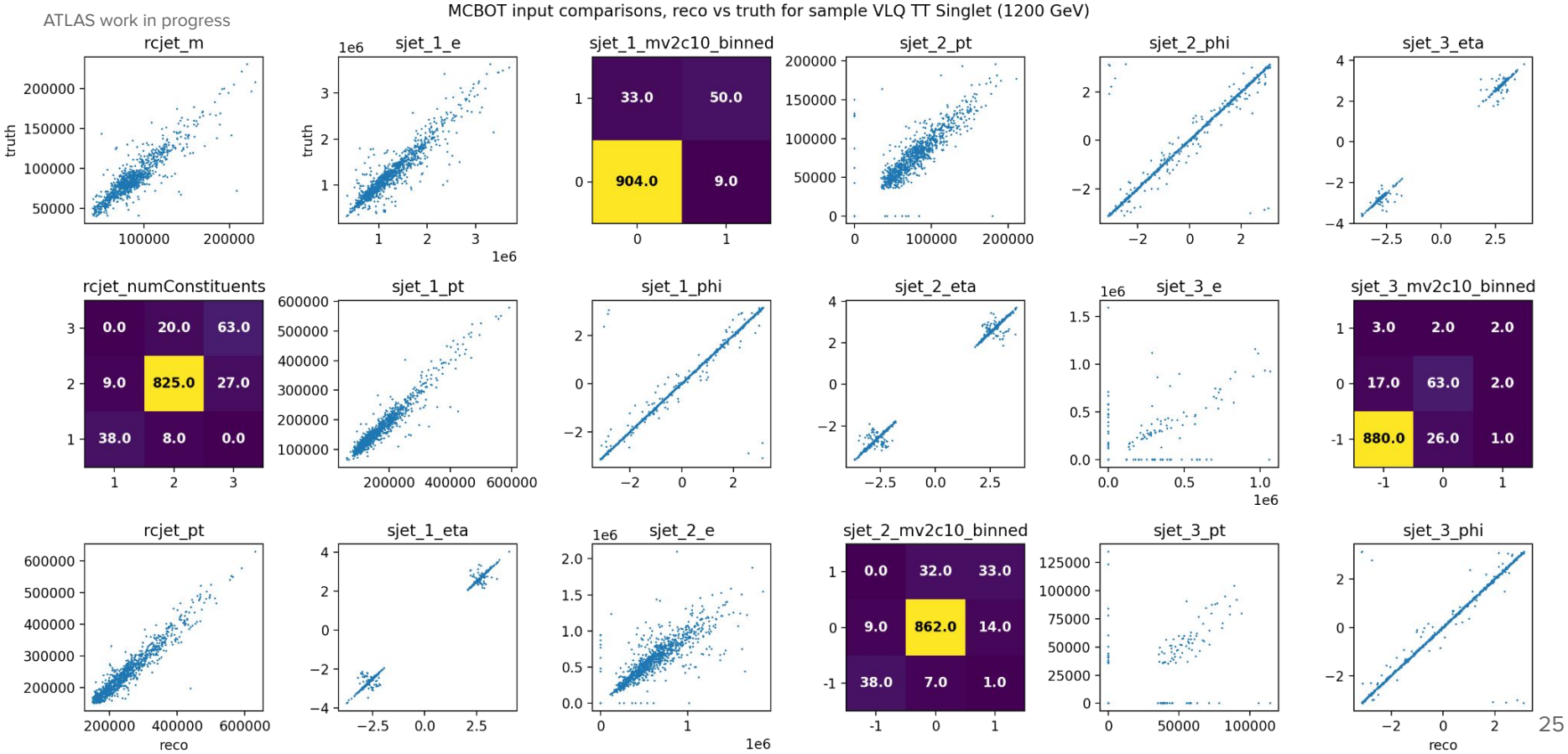
BONUS

Input comparison (rivet)

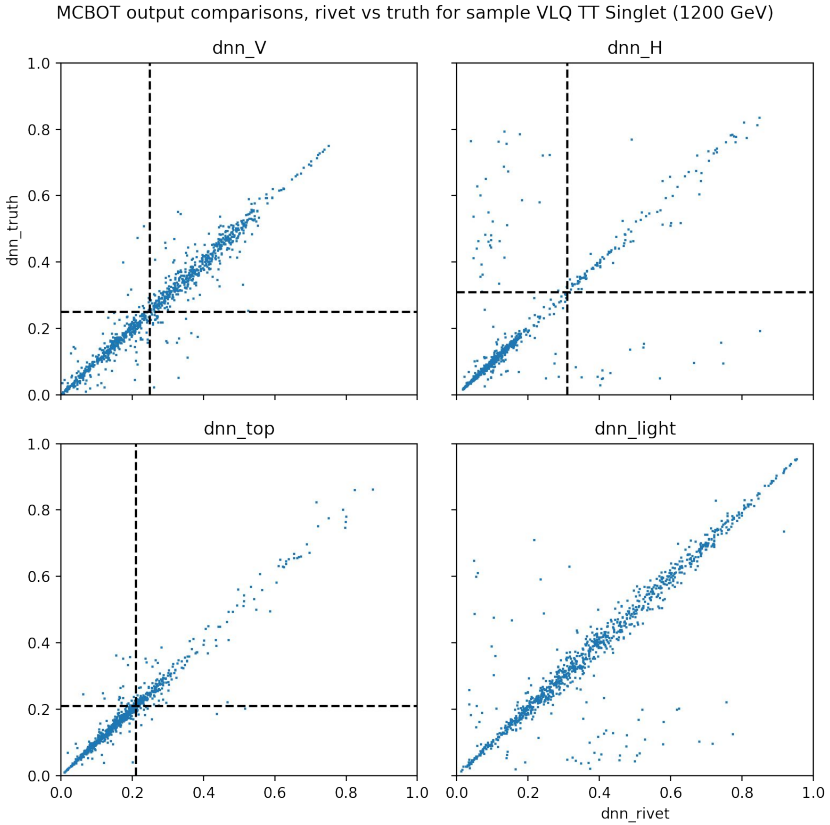


Input comparison (truth)

ATLAS work in progress



DNN outputs, Rivet vs Truth

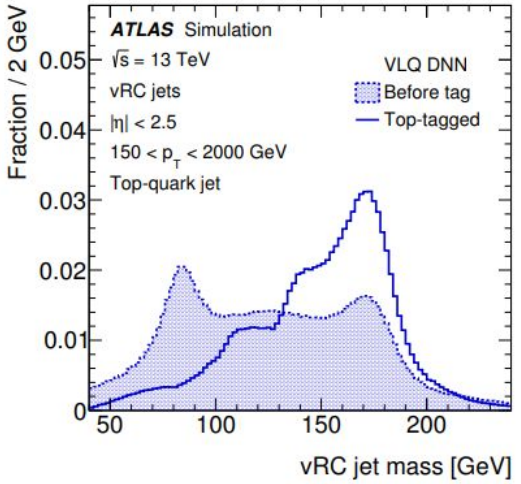


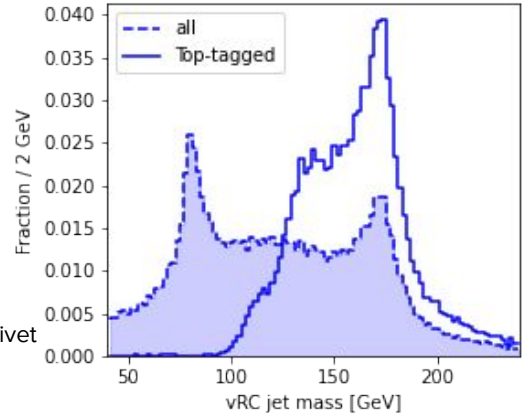
Efficiencies

Why not just use efficiencies?

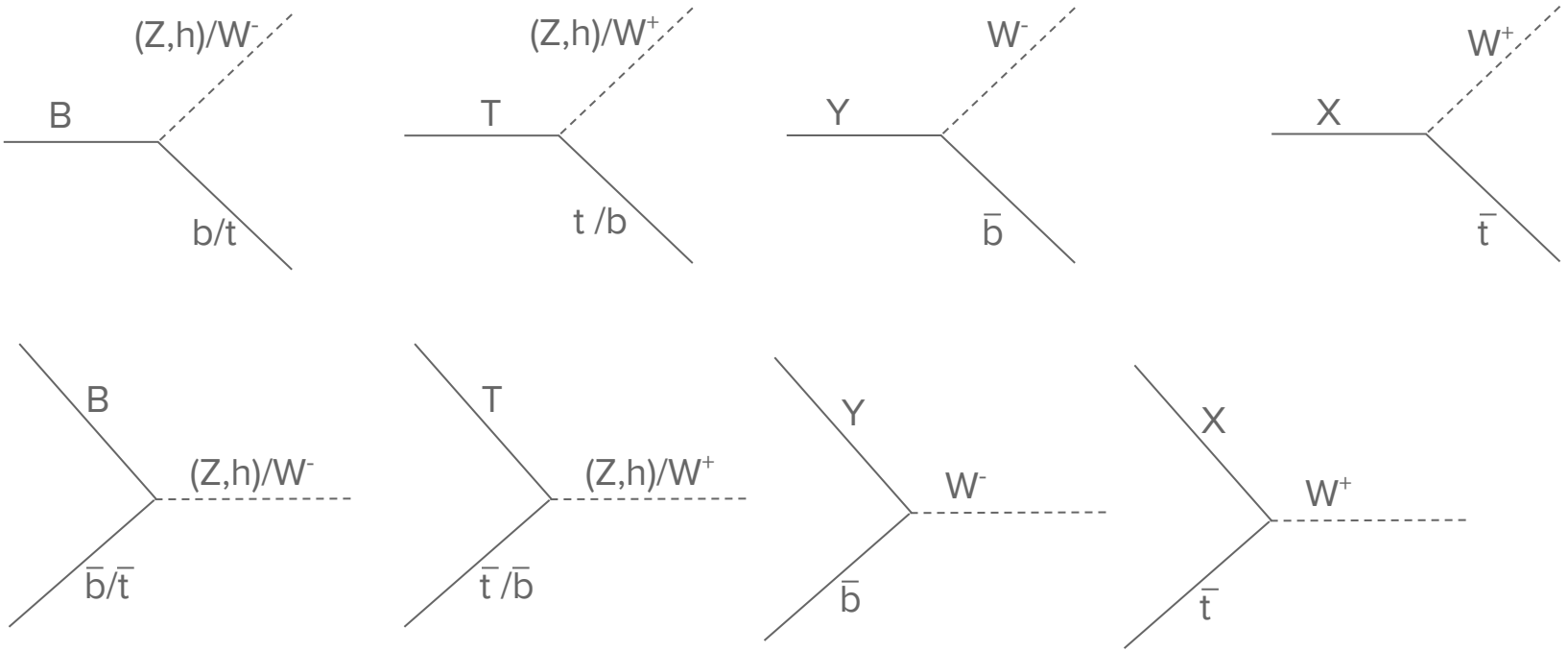
- Efficiencies don't capture kinematics
  - (seen even in the 2018 paper)
  - This becomes a serious problem if the NN is not the final cut.
- Ambiguities aplenty:
  - What to do in case of truth multi-tag?
  - Are miss-tag rates significant?
  - What even is a top quark (partonic tops)?
- Some variability across different new physics models (particularly in top tag)
- => Using the Net gives much better performance.
- *But* providing a detailed efficiency breakdown is still very useful
  - (especially if the net can't be provided)

arXiv:1808.01771
Fig 4.





MG5 + Rivet


(Z,h)/W-
B
b/t
(Z,h)/W+
T
t /b
W-
Y
b̄
W+
X
t̄
B
(Z,h)/W-
b̄/t̄
T
(Z,h)/W+
t̄ /b̄
Y
W-
b̄
X
W+
t̄