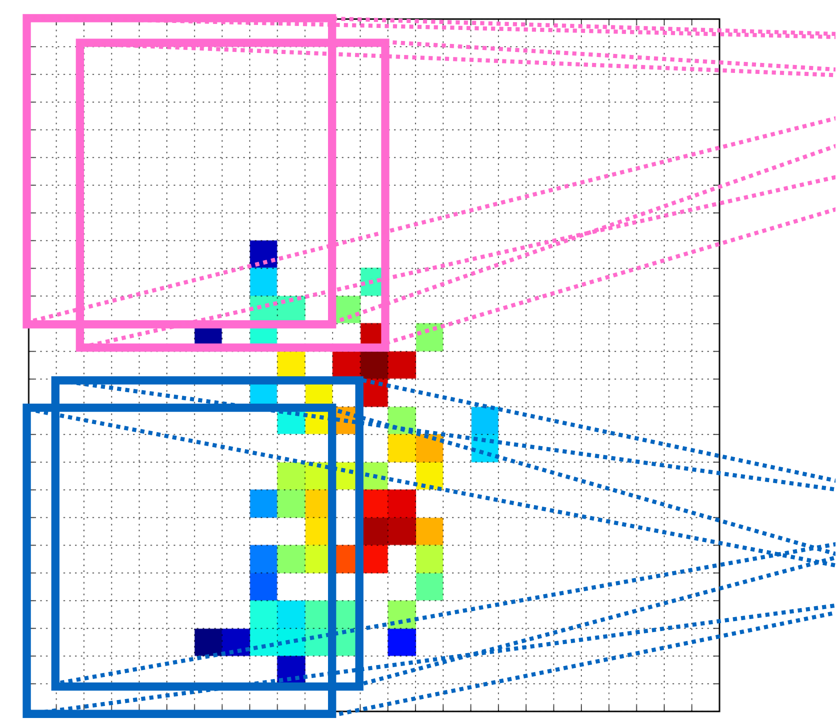
ML4Jets 2020 @ NYU

https://indico.cern.ch/event/809820

Benjamin Nachman

Lawrence Berkeley National Laboratory



BERKELEY
EXPERIMENTAL
PARTICLE
PHYSICS

ATLAS LBNL
Group Meeting
Feb. 25, 2020

Brief history

Machine Learning for Jet Physics

11-13 December 2017
Lawrence Berkeley National Laboratory
US/Pacific timezone

Machine Learning for Jet Physics

14-16 November 2018
Fermilab
America/Chicago timezone

ML4Jets2020

15-17 January 2020
Kimmel Center for University Life
America/New_York timezone

In many ways, jet physics has been leading the adaptation and development of advanced ML in HEP

This workshop series was born at LBNL in 2017

Community challenges

One of the outcomes of the 2017 workshop was the need for a community comparison study of top tagging algorithms.

Community challenges

One of the outcomes of the 2017 workshop was the need for a community comparison study of top tagging algorithms.

SciPost Physics

Submission

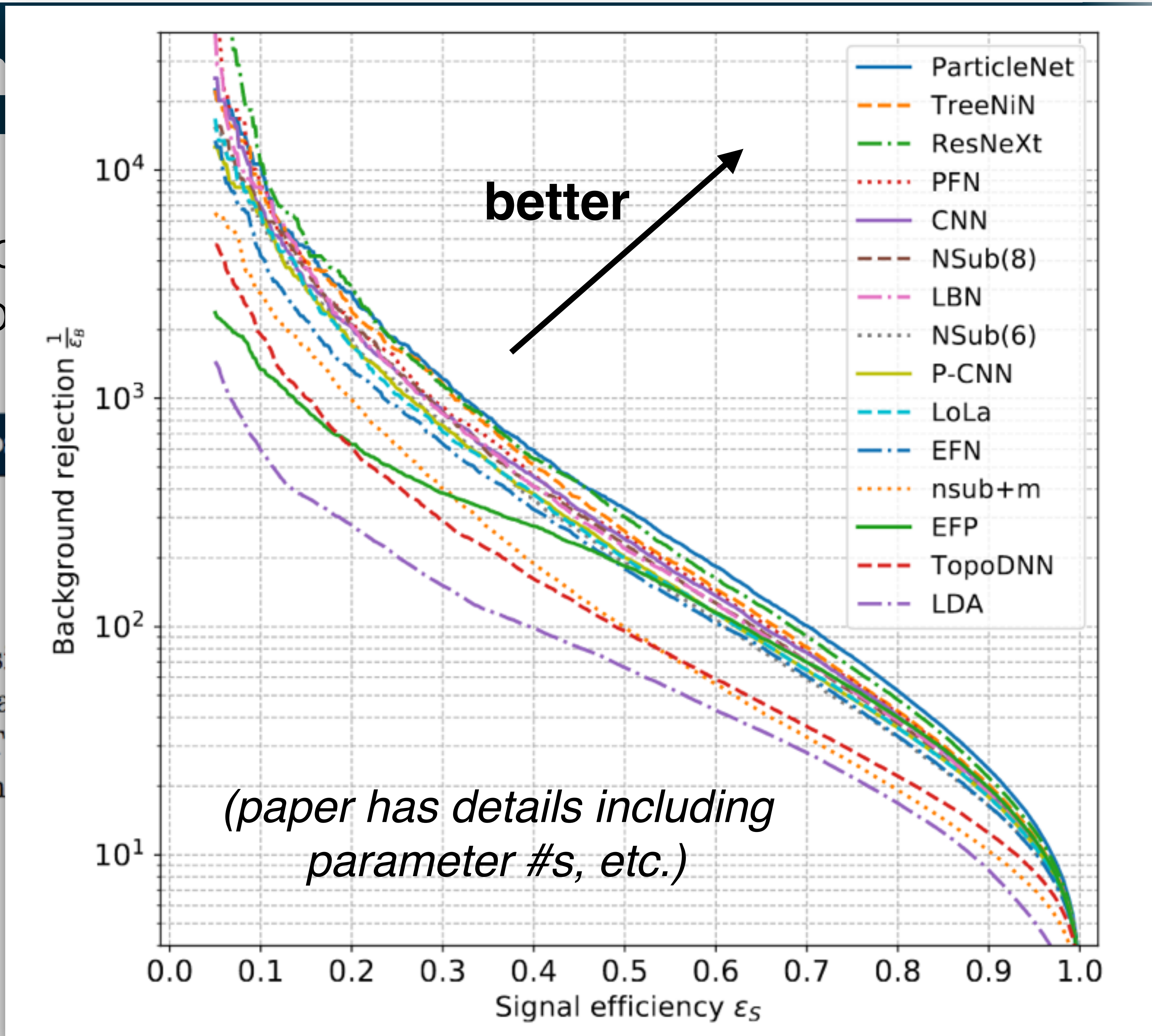
The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

Con

One c ... ed for
a co ... hms.

SciPo ... ission



G. Kas ... Dillon[5],
M. Fa ... nik[5,9],
P. T ... re[11],
B. Nach ... . Shih[4],

Community challenges

One of the outcomes of the 2017 workshop was the need for a community comparison study of top tagging algorithms.

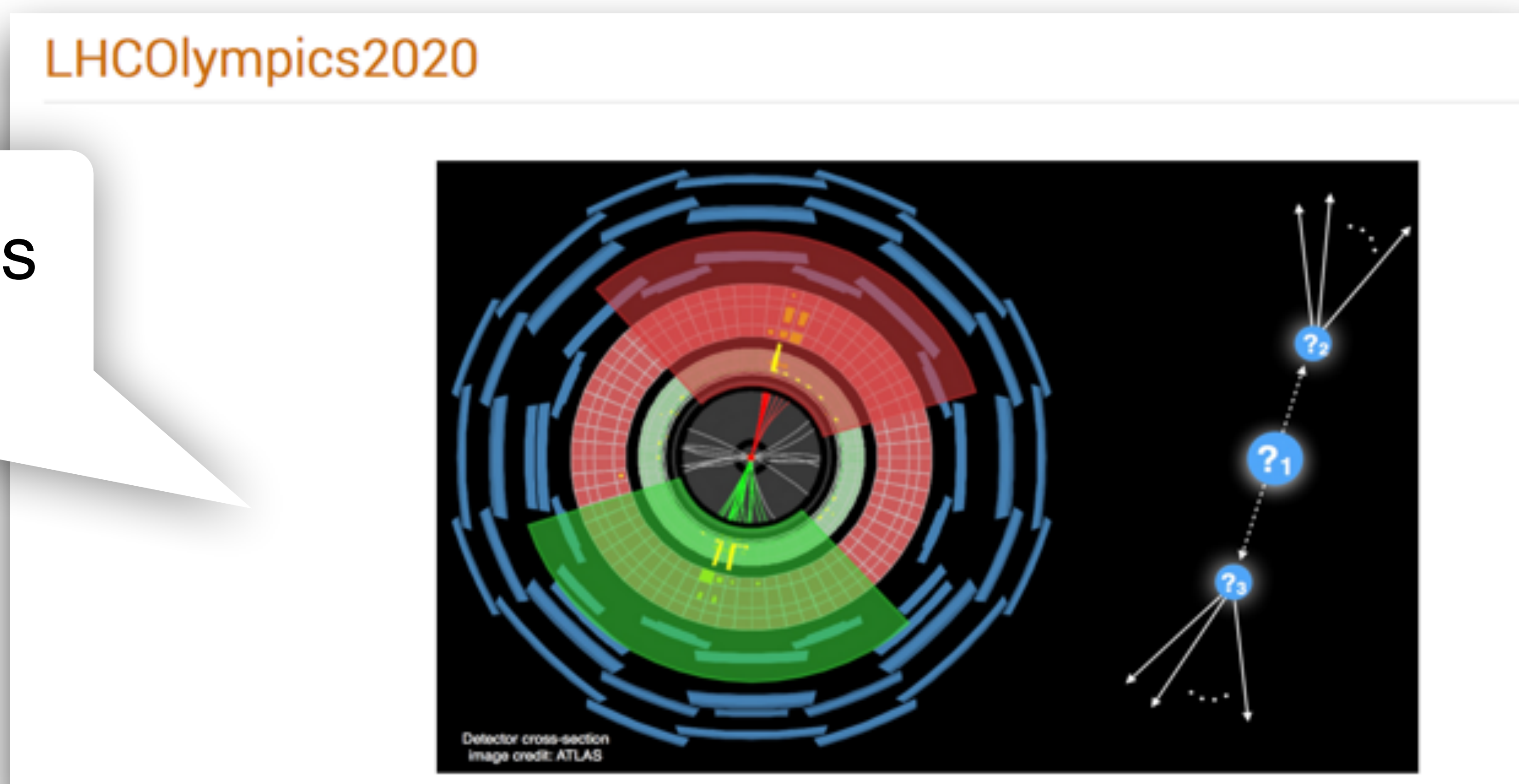
One of the outcomes of the 2018 workshop was the need for a community challenge for anomaly detection.

Community challenges

One of the outcomes of the 2017 workshop was the need for a community comparison study of top tagging algorithms.

One of the outcomes of the 2018 workshop was the need for a community challenge for anomaly detection.

More on this in a bit…



Community challenges

One of the outcomes of the 2017 workshop was the need for a community comparison study of top tagging algorithms.

One of the outcomes of the 2018 workshop was the need for a community challenge for anomaly detection.

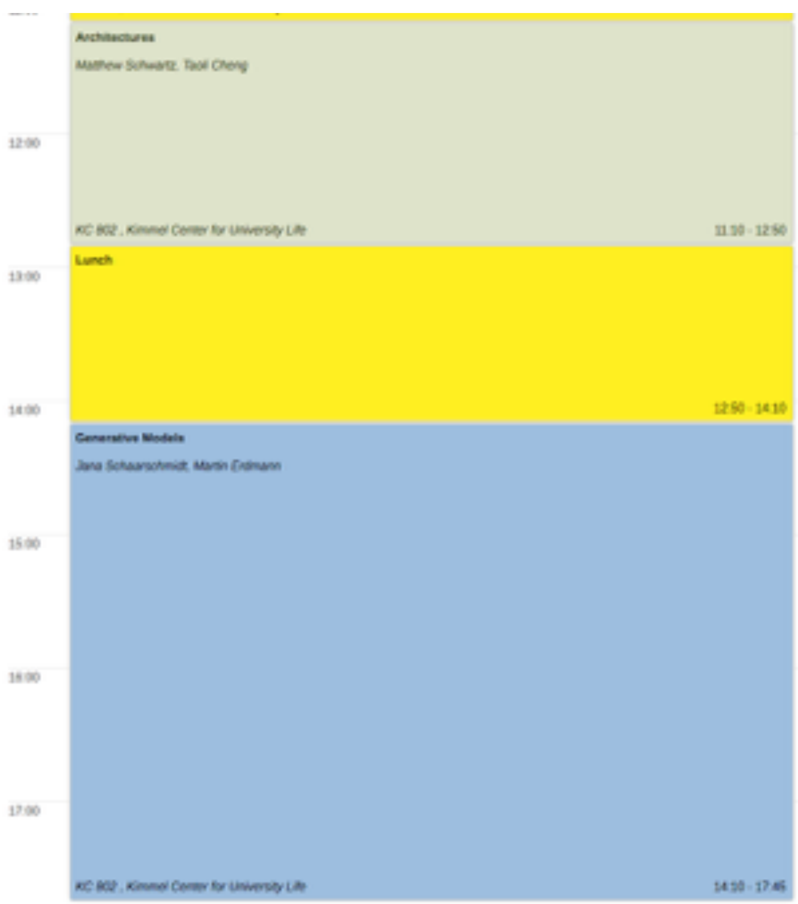
One of the outcomes of the 2020 workshop was the need for a community challenge for unfolding.

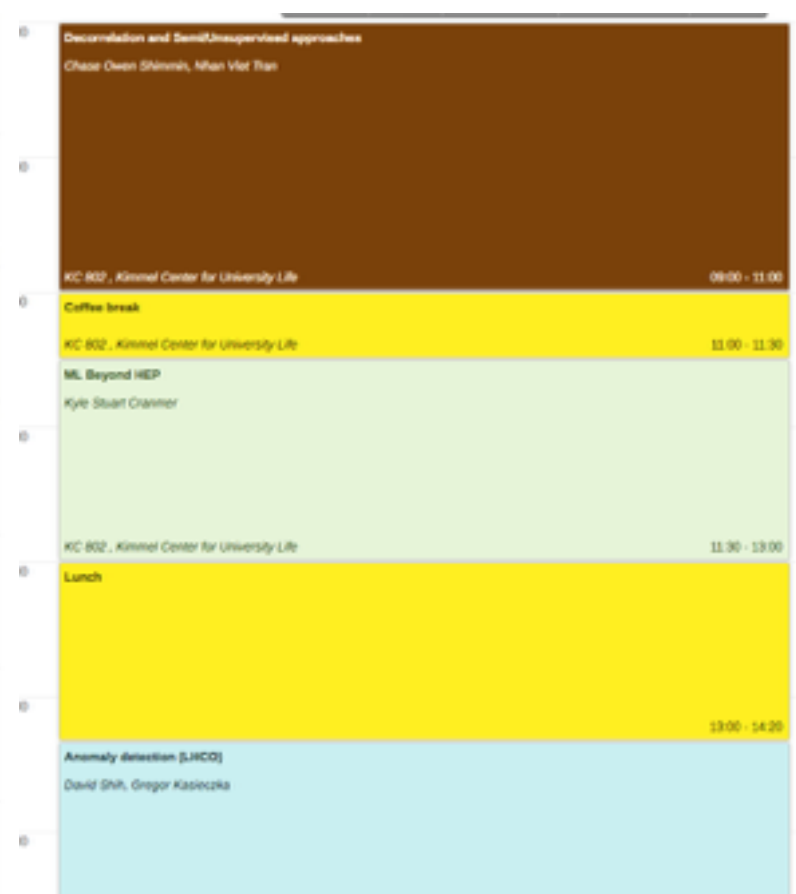
Stay tuned!

2020 Edition: Overview

There was a packed agenda with three very full days of interesting talks!

I won't review every talk … please take a look at the slides for all the contributions and for details. These slides are some personal highlights.



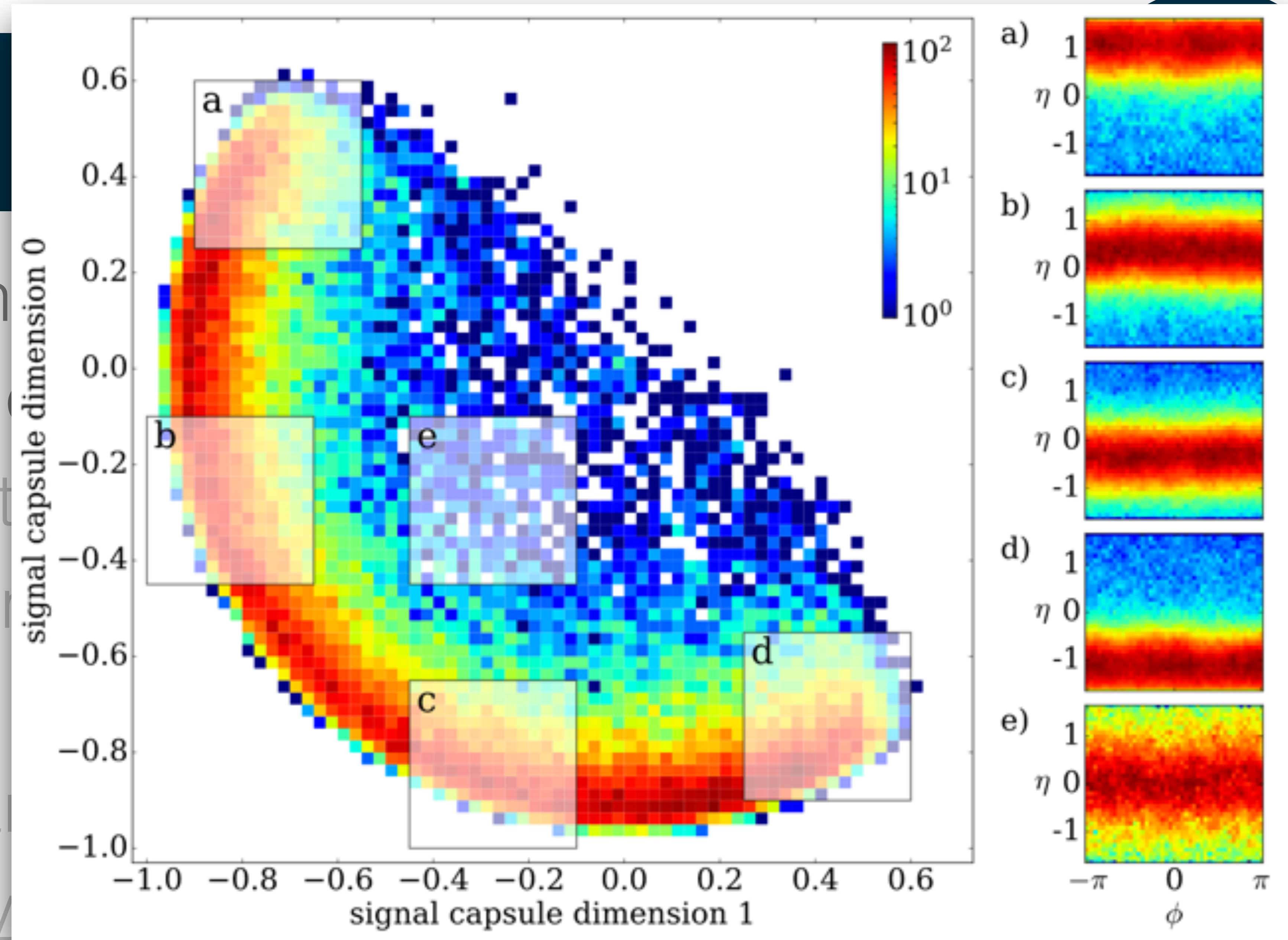


Architectures - Highlights

- Lorentz covariant networks (more on next slide)
- Yet another particle cloud architecture
  - Jet constituents are permutation invariant thanks to QM … need an architecture that acts on sets. There are now a few of these based on Deep Sets (MIT group) and various graph networks (UCSB and now also University of Zurich)
- A first study of “capsule networks” in HEP. These networks try to learn directions orthogonal to classification directions in feature space. Supposed to be interpretable.

Architectures -

- Lorentz covarian
- Yet another parti
    - Jet constituent
      QM … need a
      are now a few
      group) and va
      also University



- A first study of "capsule networks" in HEP. These networks try to learn directions orthogonal to classification directions in feature space. Supposed to be interpretable.

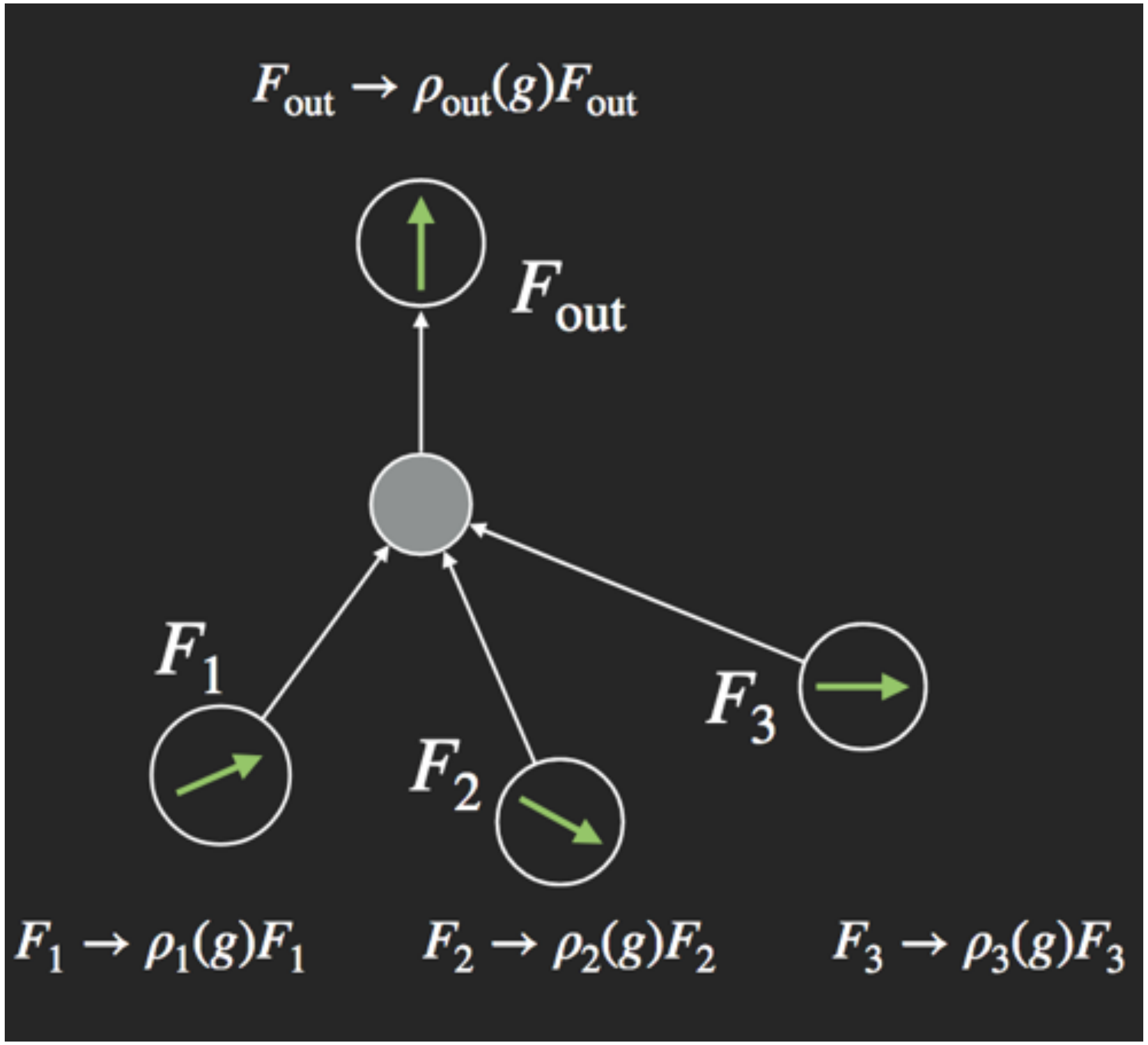
Classifier = length in capsule space

Architectures - Lorentz Covariance

New architecture that is Lorentz covariant - output is in a representation of the Lorentz group and transforms with the input

…e.g. for classification, output is a scalar (Lorentz invariant)

No need to preprocess as it is already invariant under translations along η (boosts in z), etc.

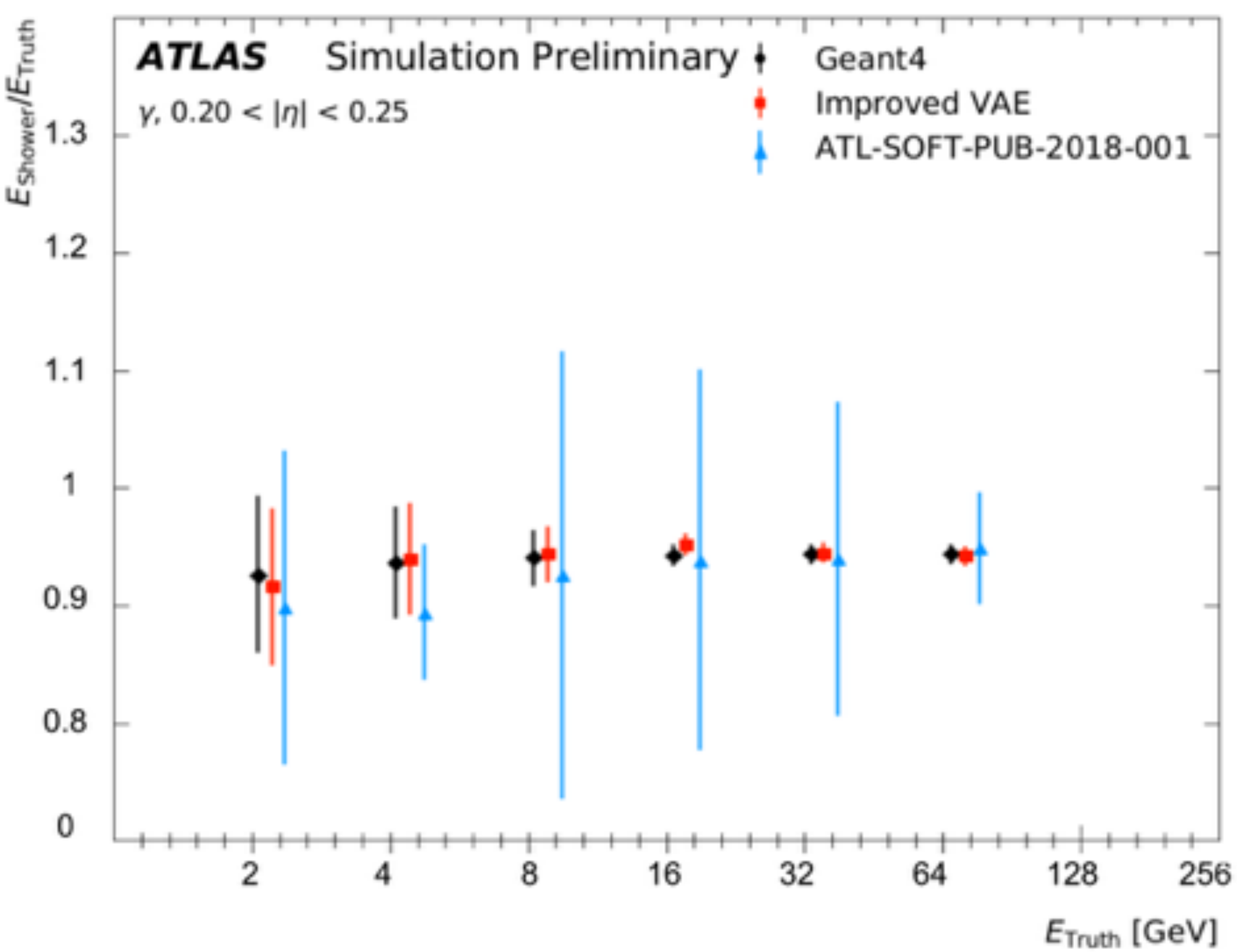


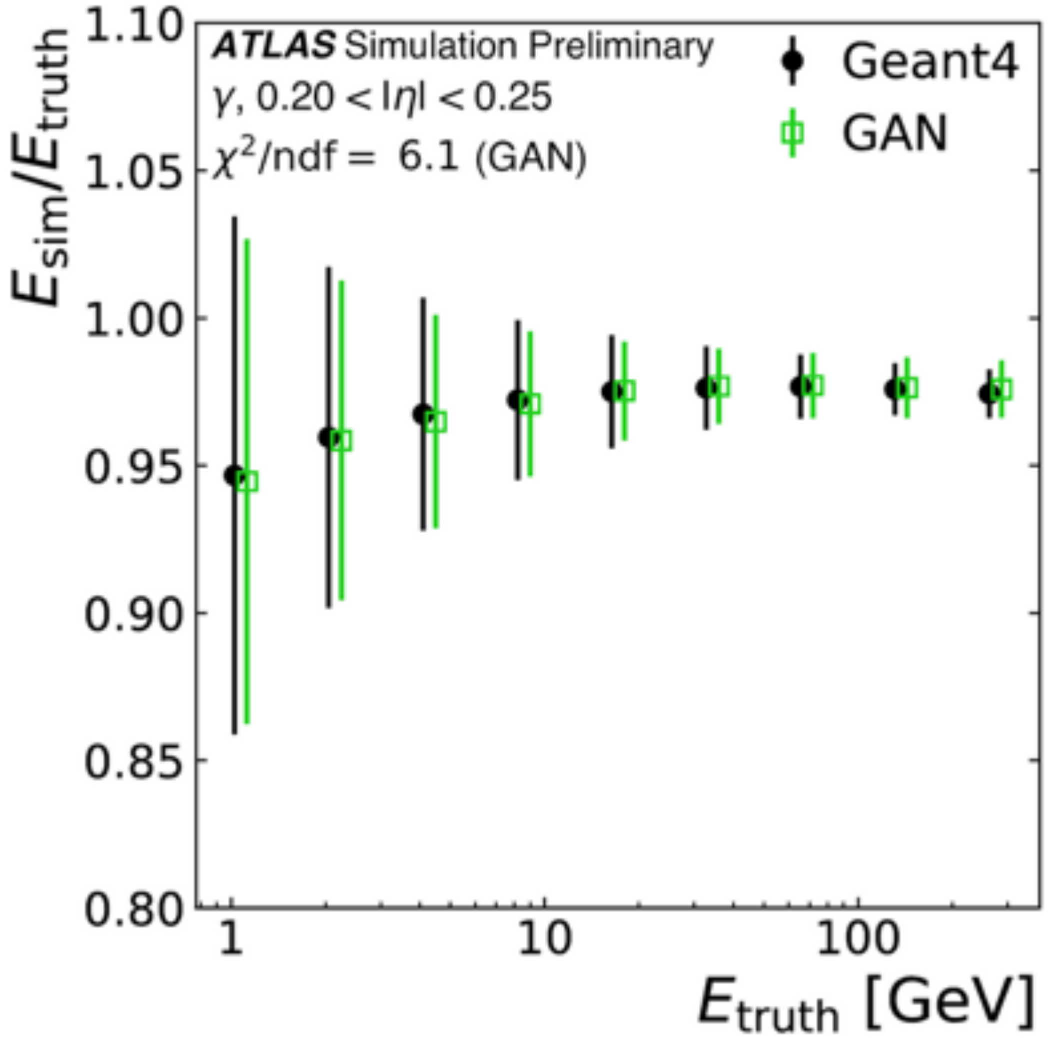
Demonstrated to be in the same ballpark as other top tagging algorithms, but with **way fewer parameters**.

A. Bogatskiy, B. Anderson, R. Kondor, D. Miller, J. Offermann, M. Roussi

Generative models - Highlights

- Calorimeter simulation from ATLAS and CALICE
  - Rapid improvements in fidelity

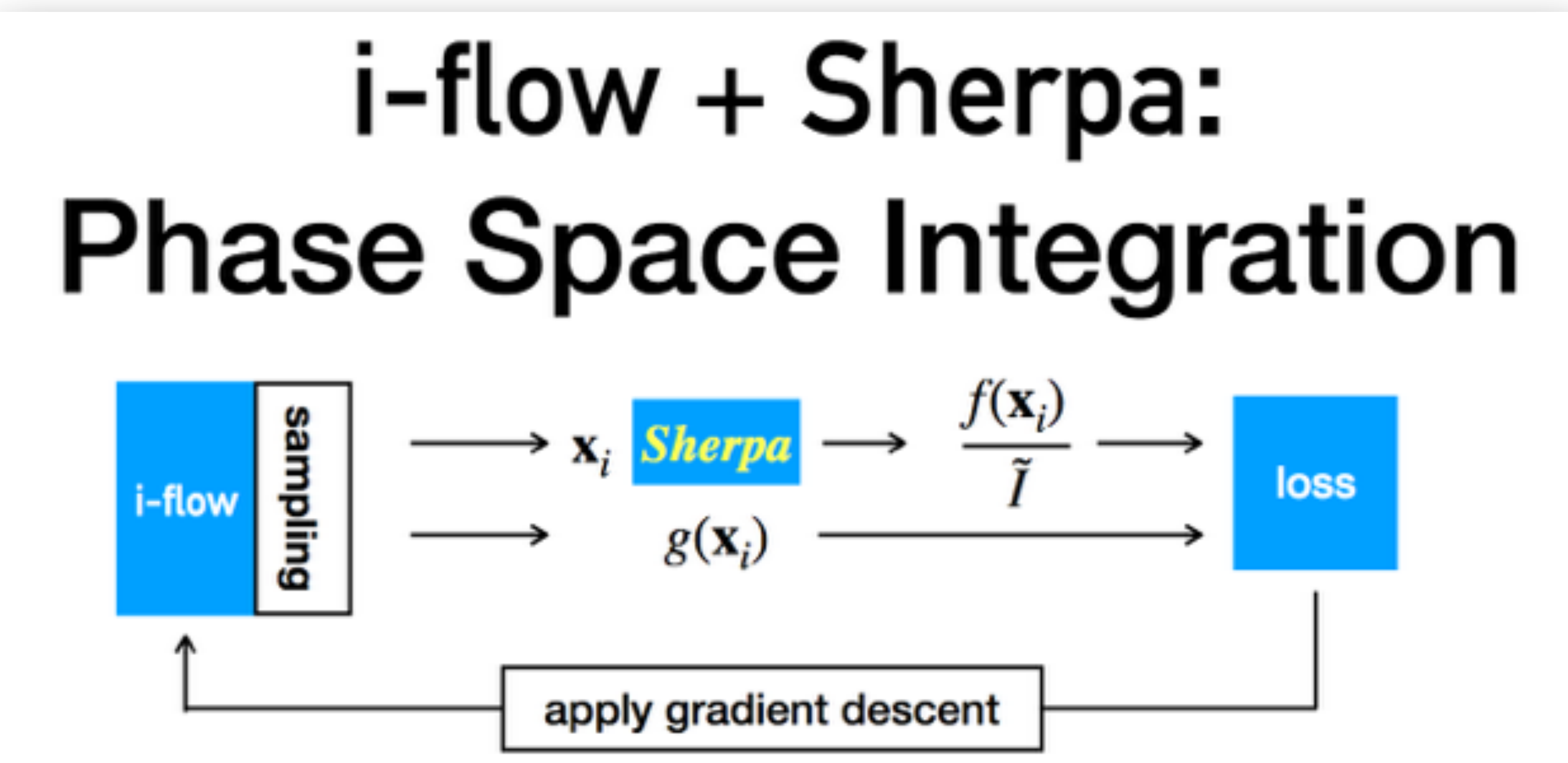




(two different deep generative models)

Generative models - Highlights

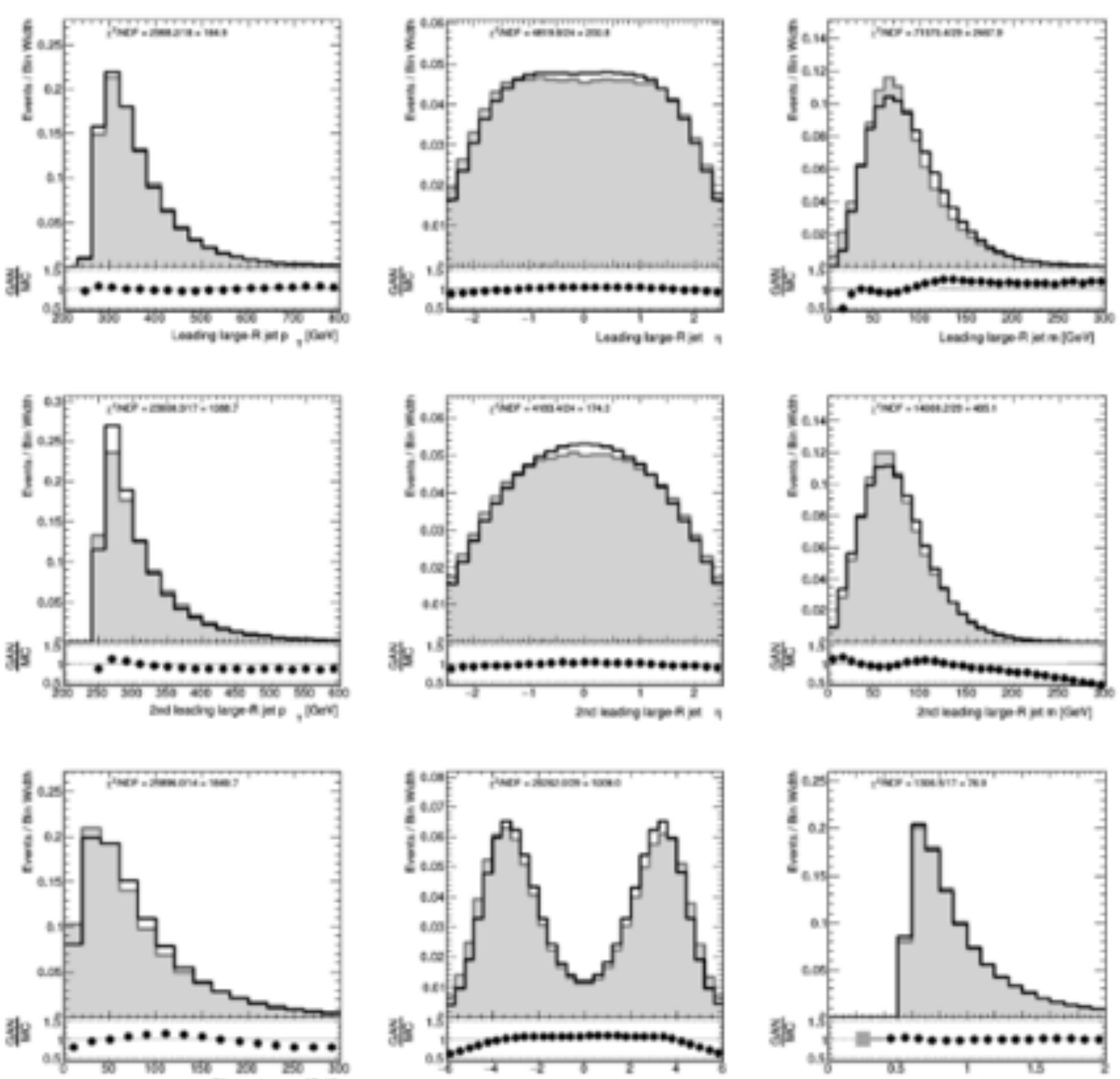
- Calorimeter simulation from ATLAS and CALICE
  - Rapid improvements in fidelity
- As a “non-parameteric” fitting function (next slide)
- For unbinned event subtraction (2 slides from now)
- For phase space integration

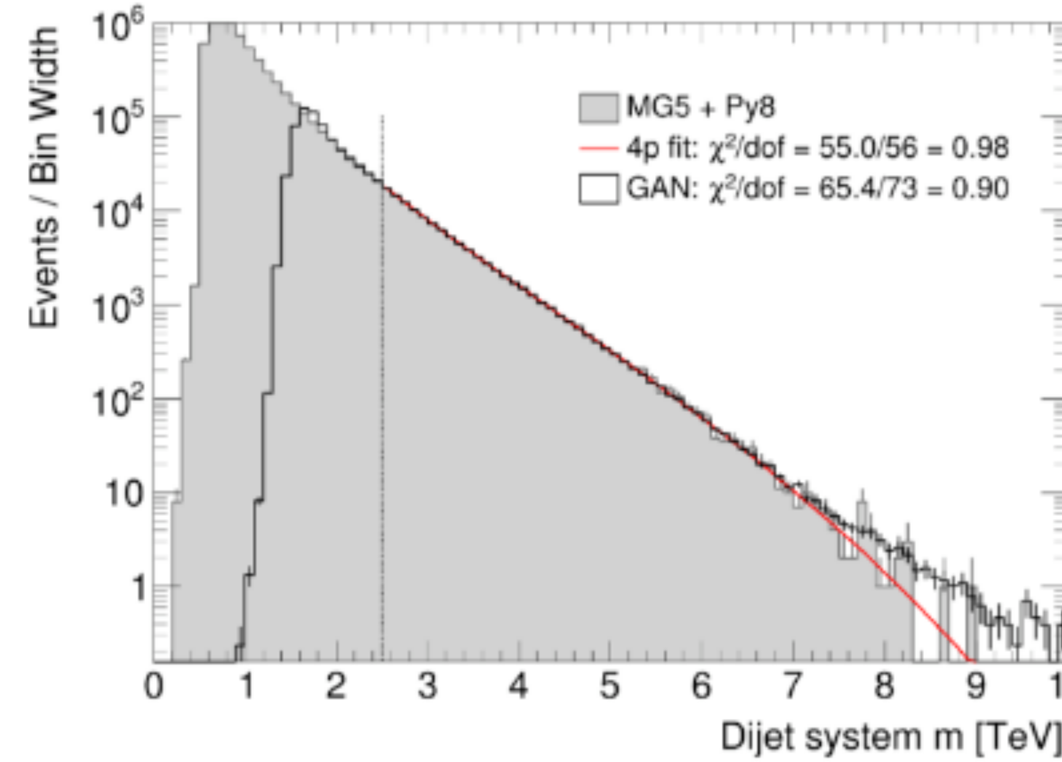


Preliminary results for unweighting efficiency are promising compared to state-of-the-art (VEGAS and FOAM)

Generative models - Dijet GAN

Generative NN's are good at interpolating - can they be used to learn good ~non-parameteric fitting functions?





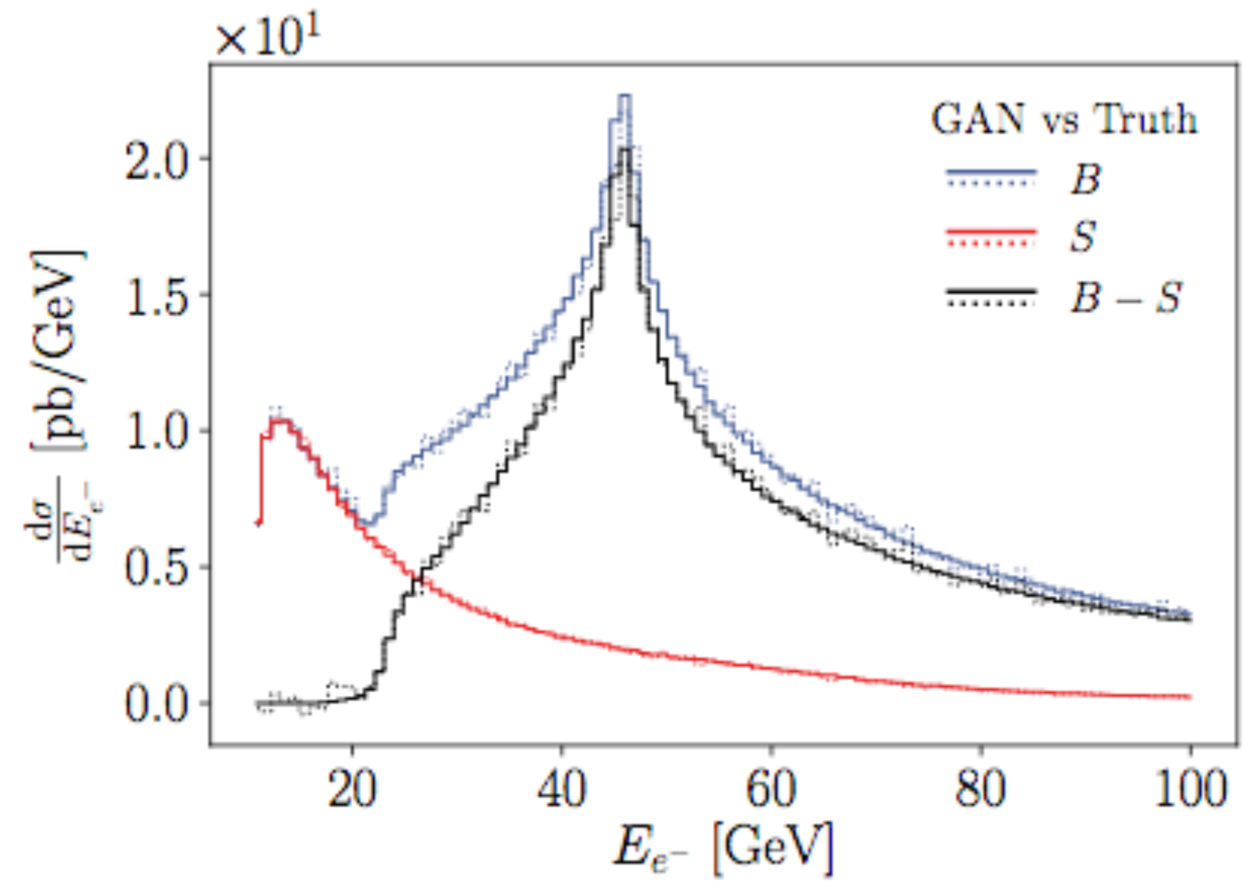
Seems to be effective at m_{jj} but not yet achieving precision in the additional jet kinematic features.

Perhaps could be useful for conditional generation as well, i.e. after making cuts.

R. Di Sipio, M. Giannelli, S. Haghighat, S. Palazzo

Generative models - Event subtraction

Proposal to use GANs to do unbinned event subtraction



Idea: learn to generate S and X where S and X have to sum in distribution to B. This will make X = B-S.

I'm not sure what the killer application is, but it seems like a very nice idea and maybe can be used for subtracting backgrounds for scale factors, unfolding, …

A. Butter, R. Winterhalder, T. Plehn

Decorrelation/Unsupervised - Highlights

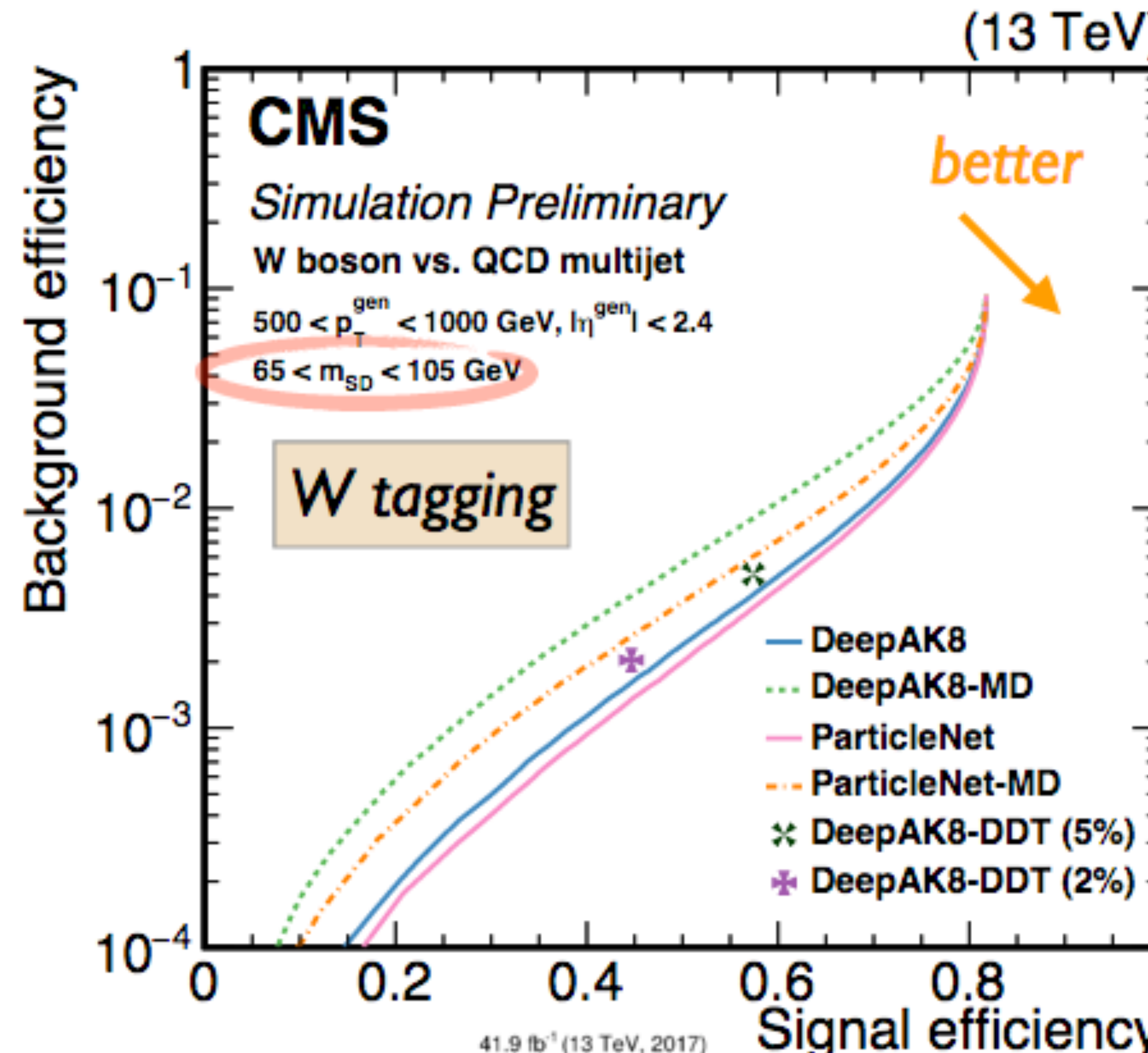
- Decorrelation
  - Why do we need to decorrelate? Might want a classifier to not sculpt bumps, so need classifier to be ~uncorrelated with e.g. m_J.
  - Decorrelation in CMS (next slide)
  - A new decorrelation scheme (2 slides)
- Optimal transport
  - How do define how close two jets are to each other?
  - Structure discovery with autoencoders (3 slides)
  - Generalization of energy movers' distance

Decorrelation - CMS

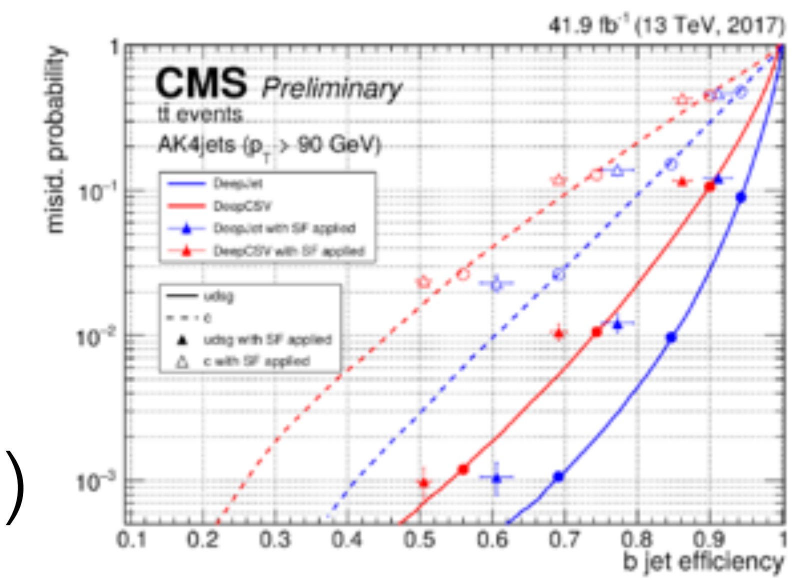
CMS presented many results with state-of-the-art taggers and various decorrelation schemes.

DDT + NN only works for a fixed WP
(can always decorrelate "by hand" in one place)

They are using rather complex networks compared to what we are doing - they have more power, but do they understand their uncertainties?



(large scale factors →)



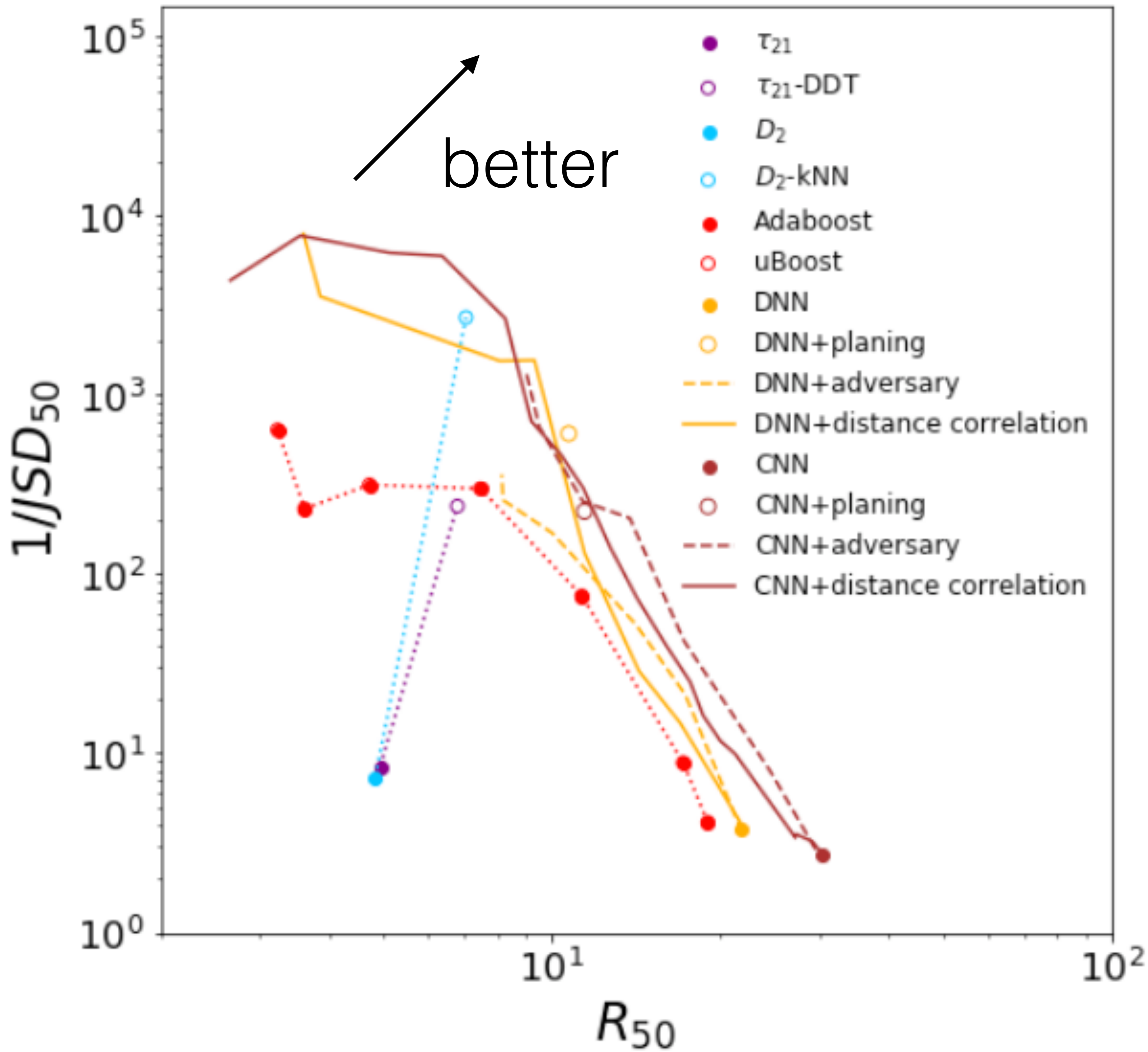
H. Qu, for CMS

Decorrelation - DisCo

New alternative to adversarial decorrelation using "Distance Correlation"

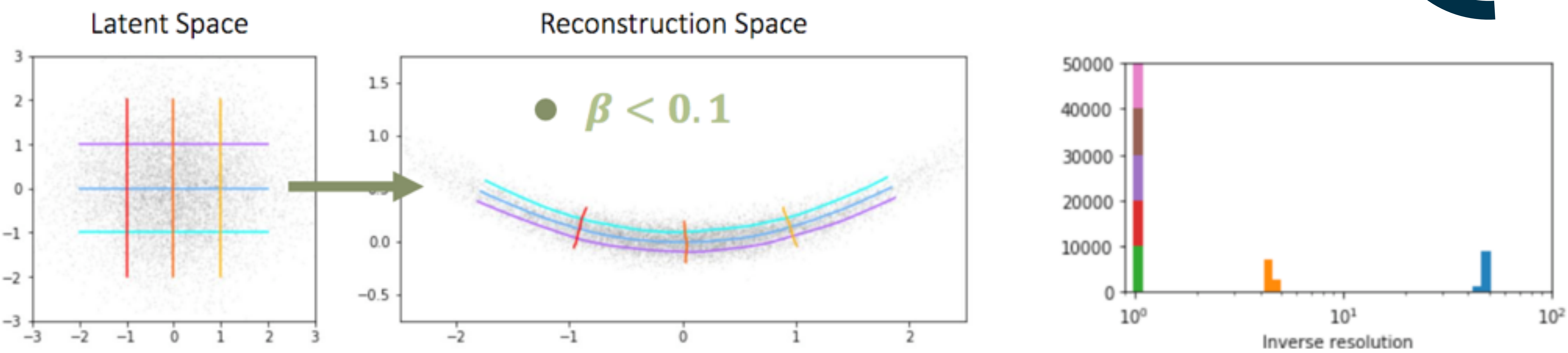
Much easier to train because not minimax … also has only one additional parameter (not a whole NN's worth)

They did a nice recast of the ATLAS result as their benchmark!



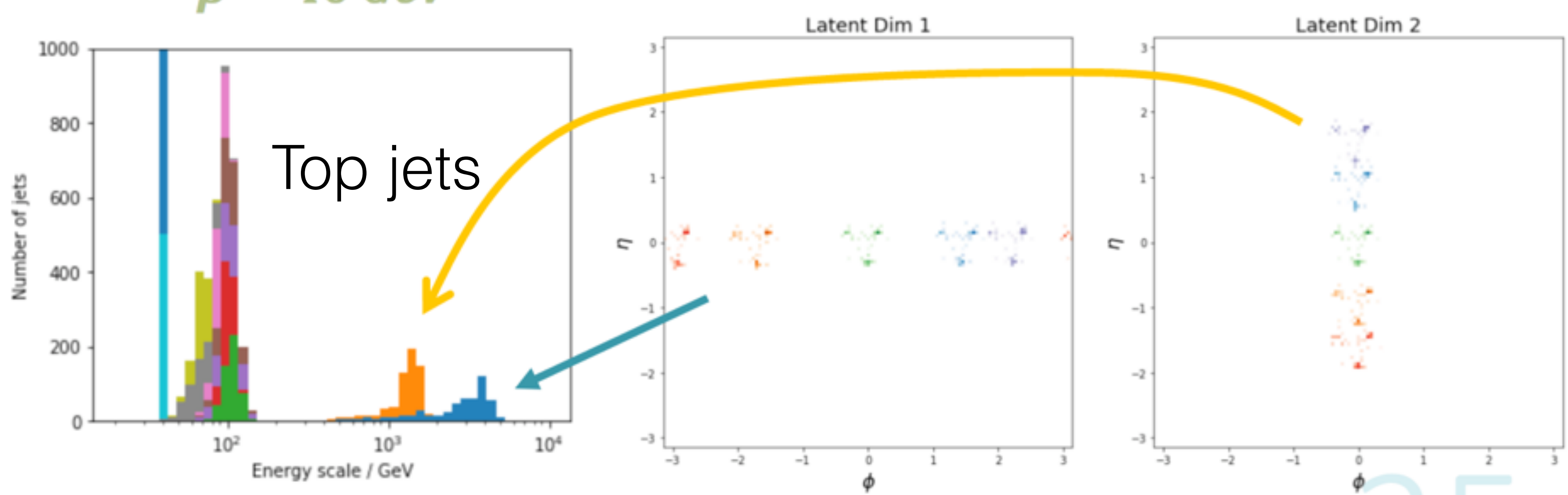
G. Kasieczka, D. Shih

Semi-supervised - VAEs to find structure



$\beta = 40\ GeV$

β is the cost for encoding information



Physical scales in a problem can be automatically discovered

Jack Collins

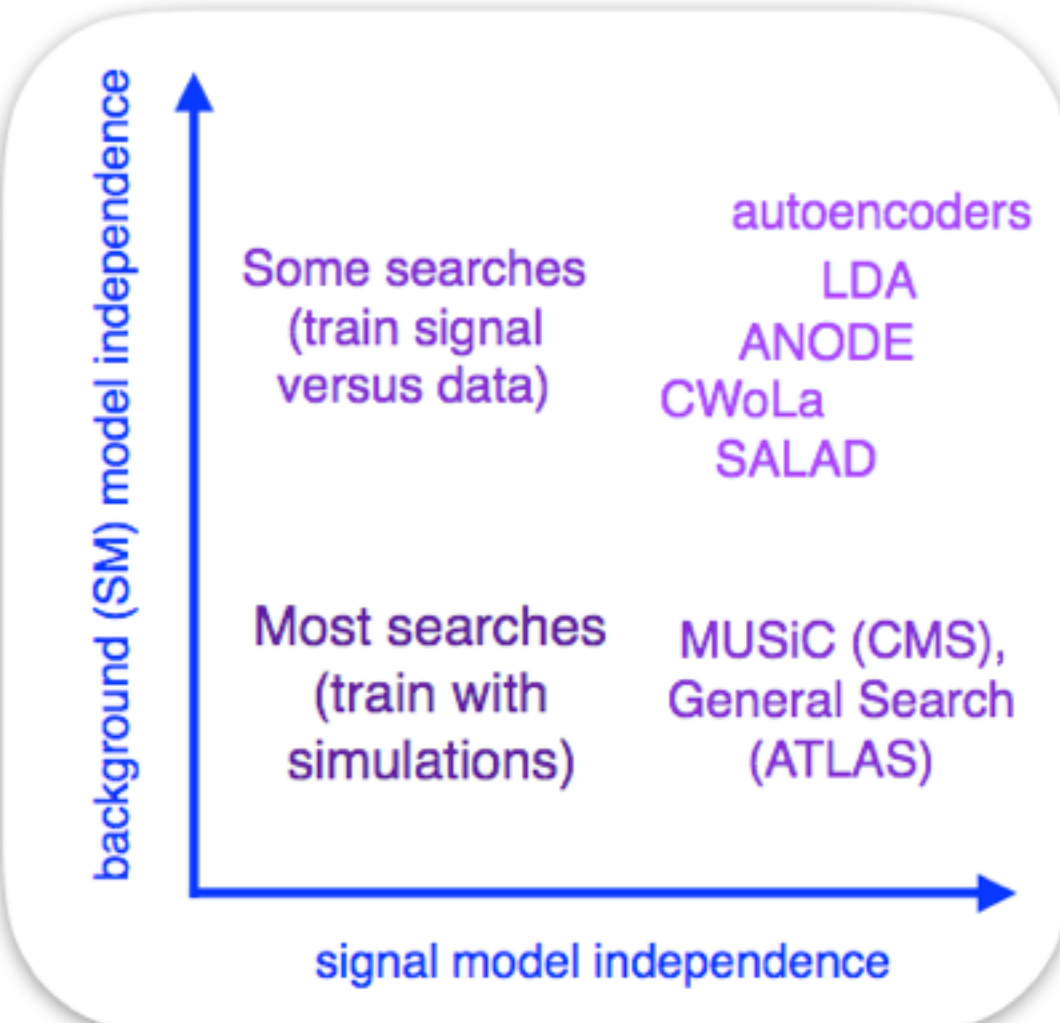
Anomaly detection - Highlights

- This session was special - we still had talks of new ideas, but additionally heard LHC Olympics solutions
- At the end of the session, we unveiled the first of three "black boxes" with simulations that may or may not have had added signal.
    - We heard very positive feedback especially from ATLAS/CMS experimentalists who told us that this was a great exercise for them to prepare for data analyses

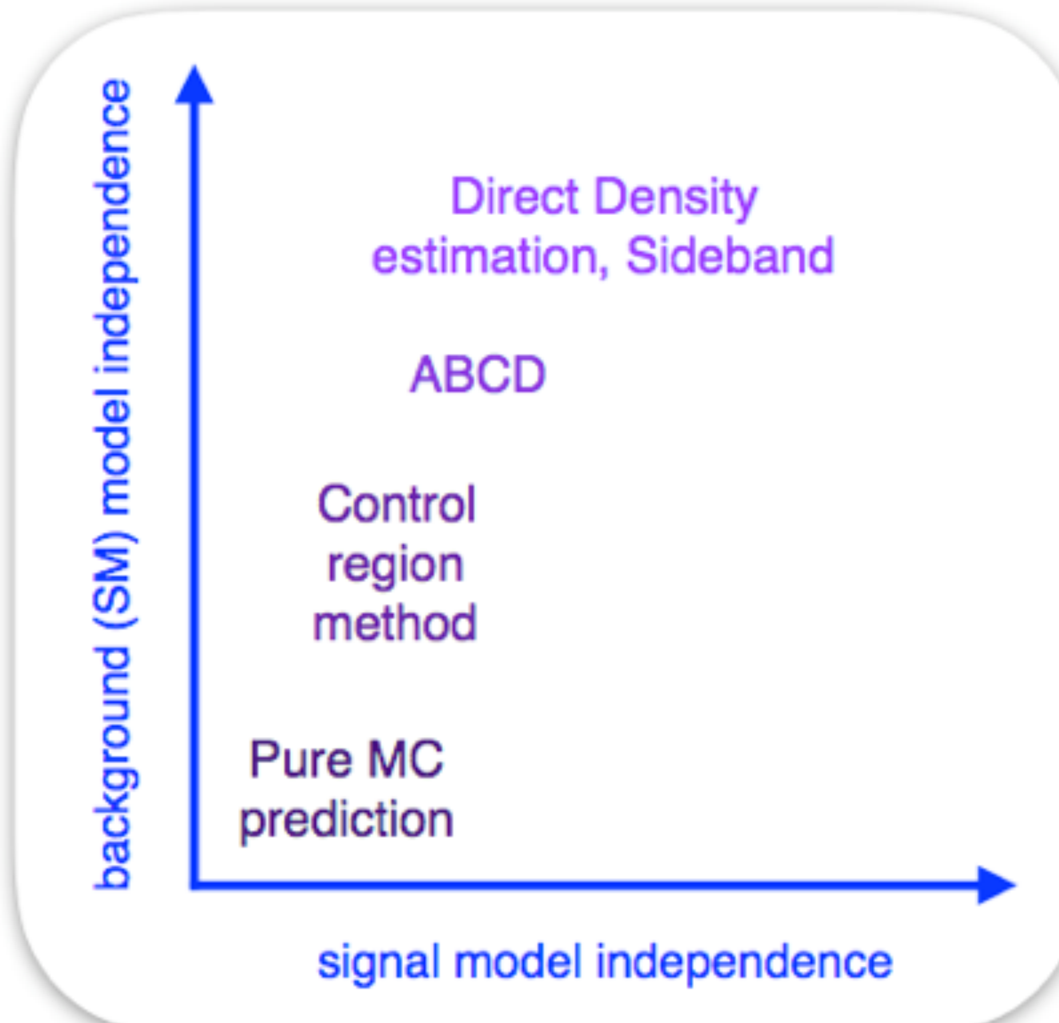
Anomaly detection - New methods

Many new methods were presented in the anomaly detection session!

from Nachman & DS 2001.04990



(a) Signal sensitivity



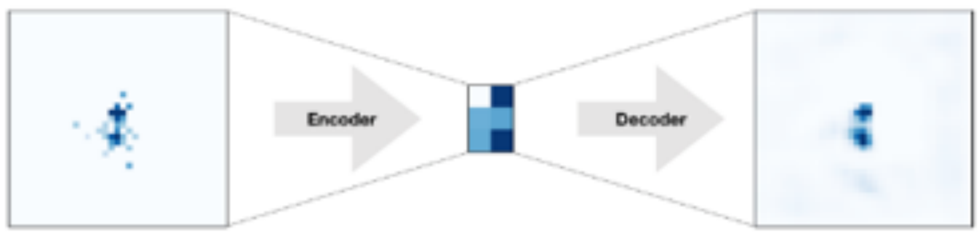
(b) Background specificity

D. Shih

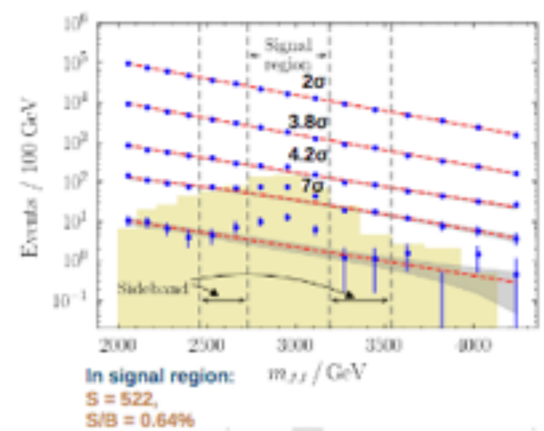
Anomaly detection - New methods

Many new methods were presented in the anomaly detection session!

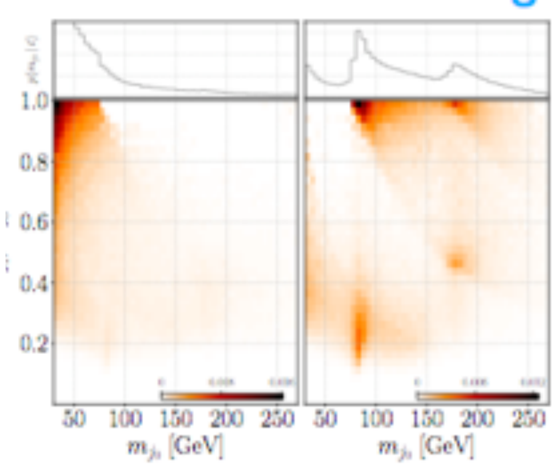
Autoencoders



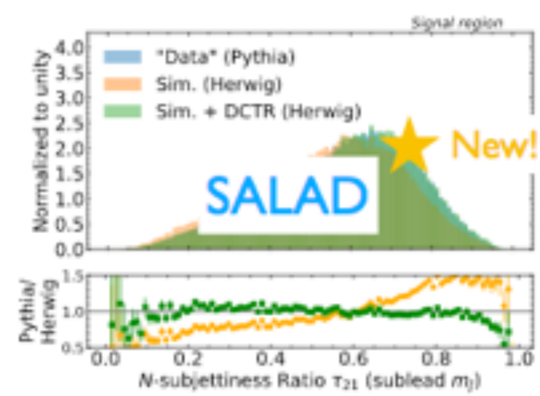
CWoLa

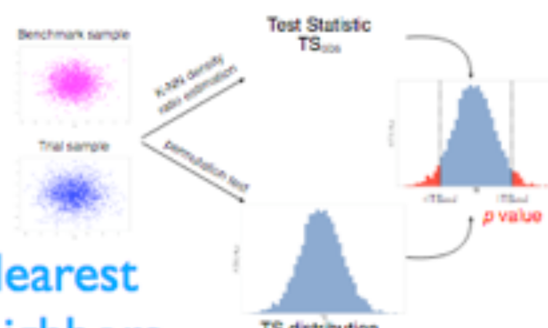


Probabilistic Modeling

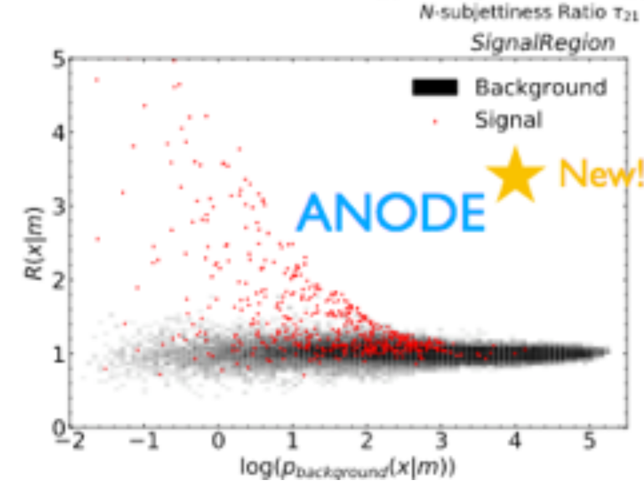


New approaches





Nearest Neighbors



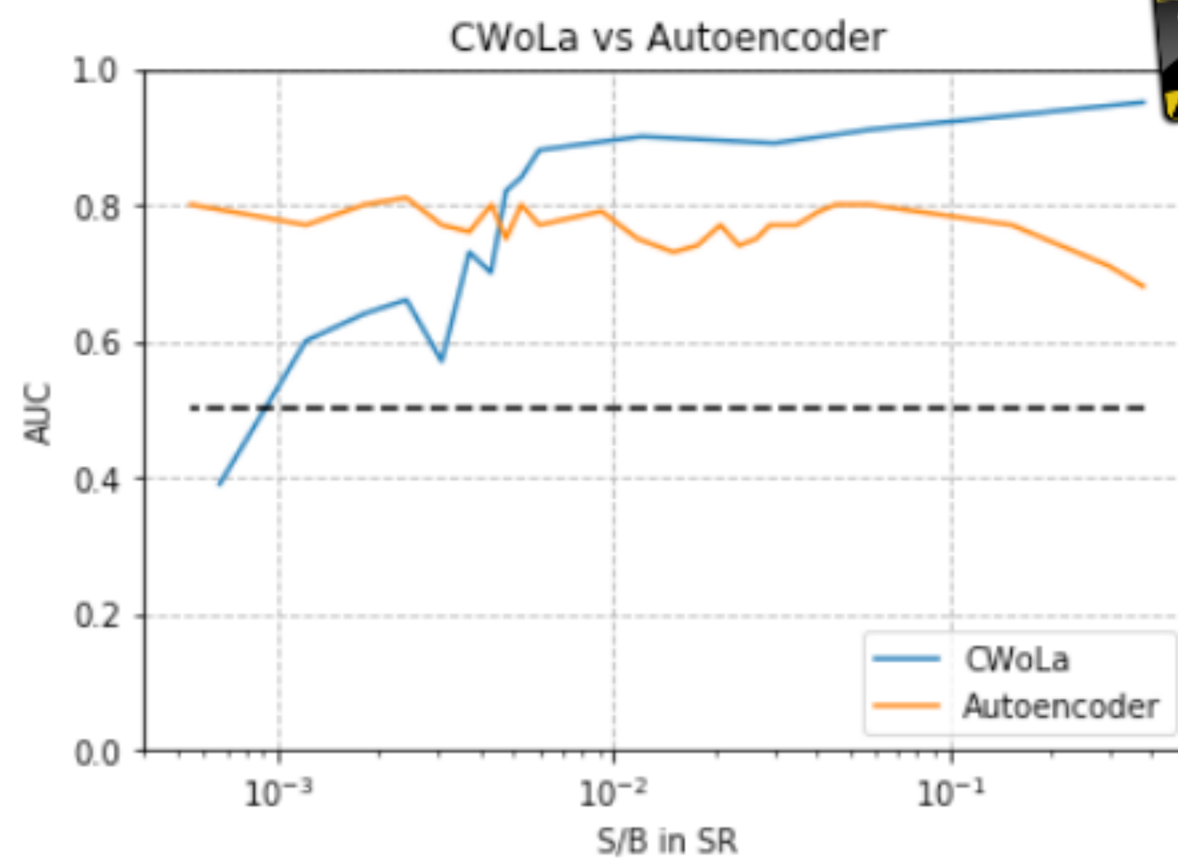
<u>D. Shih</u>

Anomaly detection - Tradeoffs

No algorithm will work best everywhere!

It is likely that we will need multiple approaches

This is just one plot that shows the complimentarily between a semi-supervised approach and an unsupervised approach

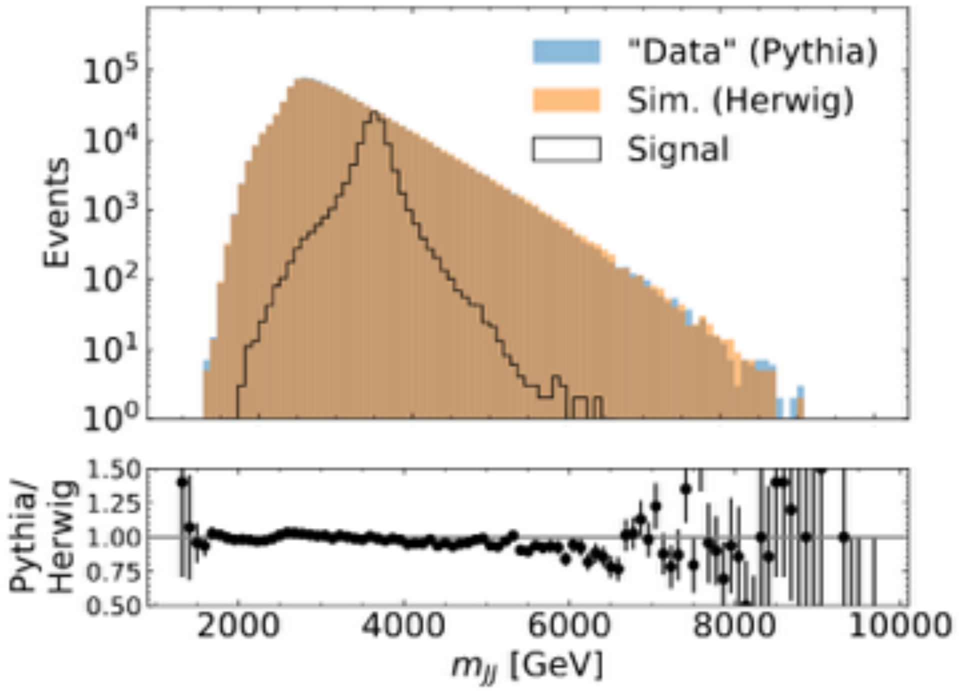


The semi-supervised approach does better when there is enough signal while the unsupervised one doesn't use signal at all so is independent of S/B.

P. Ramiro, J. Collins, B. Nachman, D. Shih

LHC Olympics 2020: Black Boxes

Organizers: Gregor Kasieczka, Ben Nachman & David Shih



Three black boxes of simulated data were prepared:

- 1 million events each
- 4-vectors of every reconstructed particle (hadron) in the event
- Particle ID, charge, etc not included
- Single R=1 jet trigger pT>1.2 TeV
- Black boxes are meant to be representative of actual data, meaning they are mostly background and may contain signals of new physics

In addition, a sample of 1M QCD dijet events (produced with Pythia8 and Delphes3.4.1) was provided as a background sample.

https://doi.org/10.5281/zenodo.3547721

Anomaly detection - LHC Olympics

Box 1

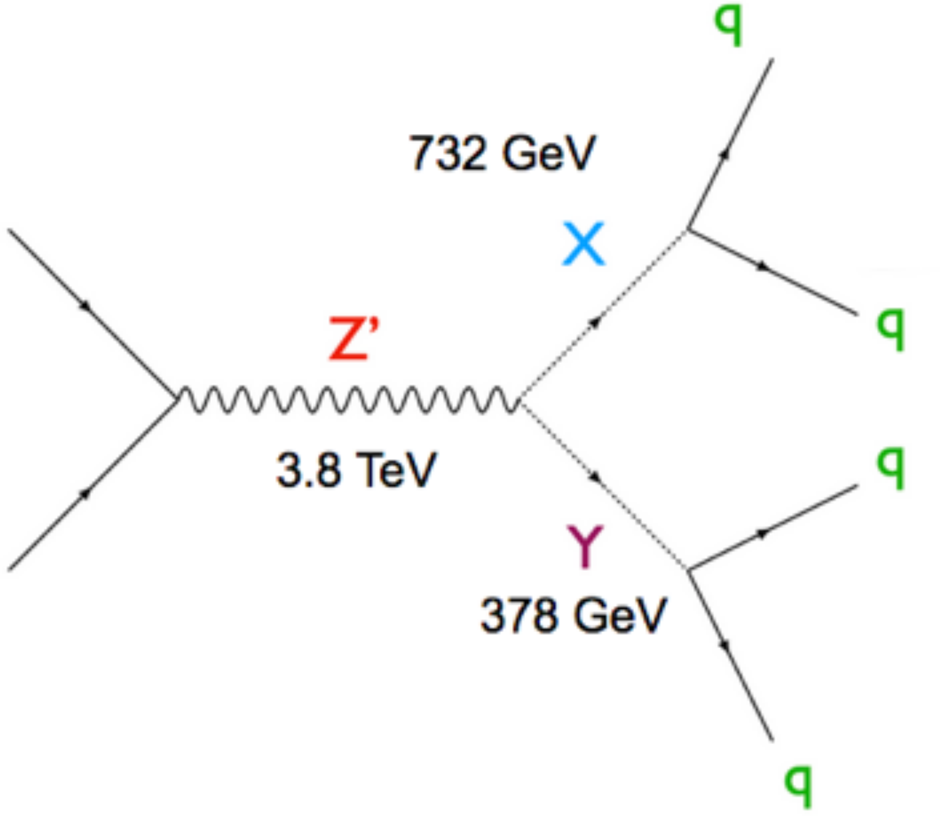
Signal: 834 events

Z'->XY; X,Y->qq
(same topology as R&D dataset)

mZ' = 3823 GeV

mX = 732 GeV

mY = 378 GeV



Anomaly detection - LHC Olympics

Box 1

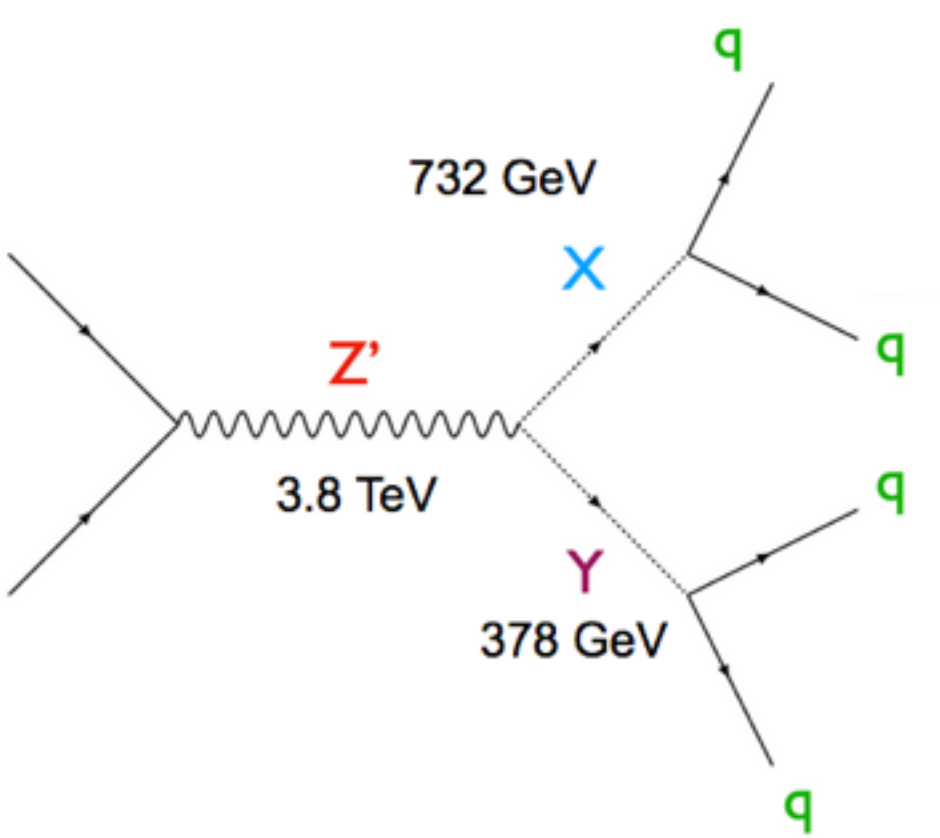
Signal: 834 events

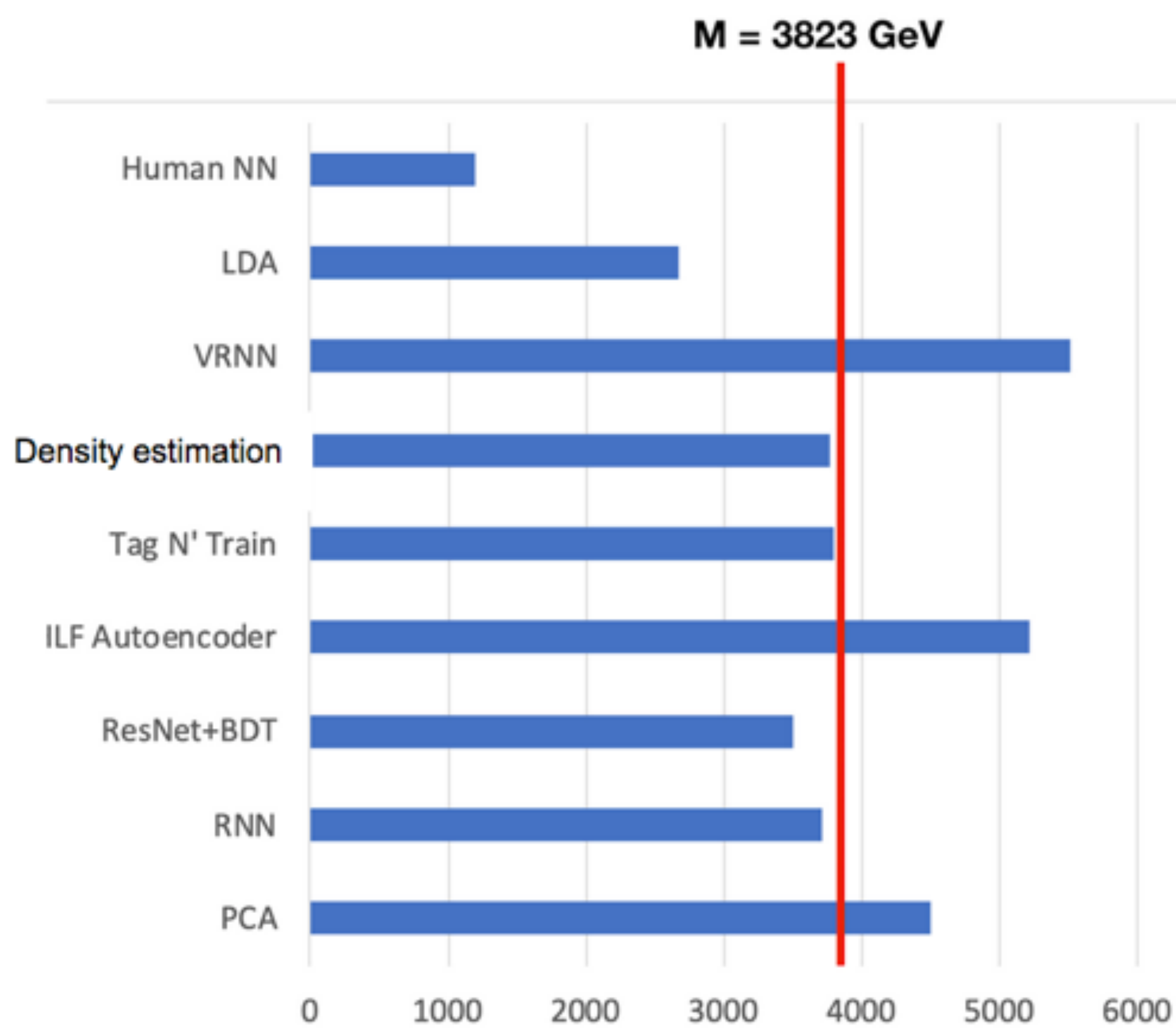
Z'->XY; X,Y->qq
(same topology as R&D dataset)

mZ' = 3823 GeV

mX = 732 GeV

mY = 378 GeV





Anomaly detection - LHC Olympics

Box 1

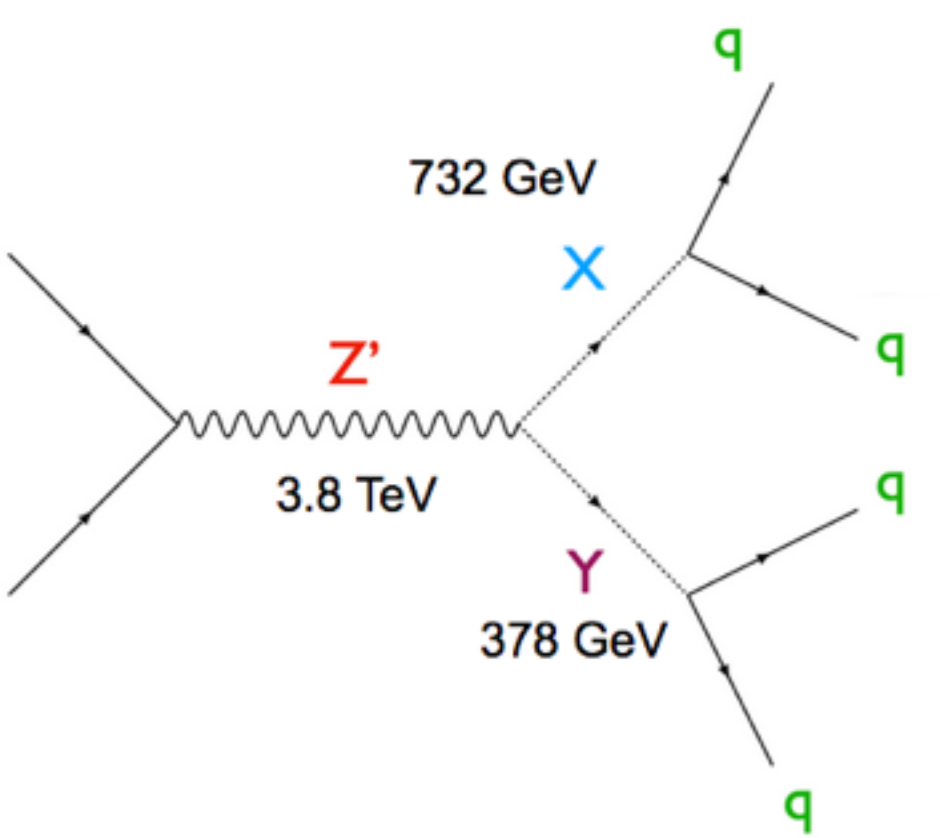
Signal: 834 events

Z'->XY; X,Y->qq
(same topology as R&D dataset)

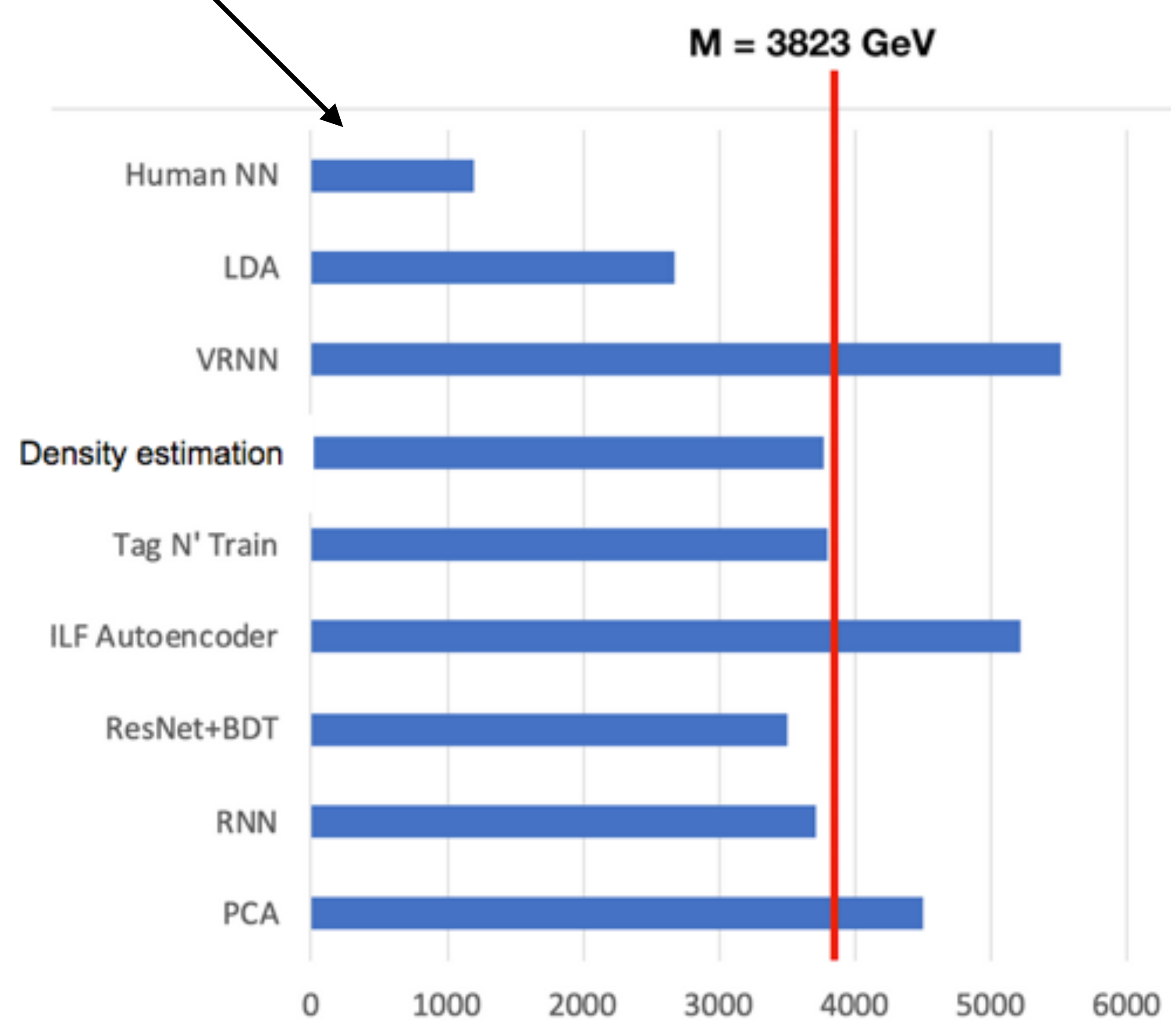
mZ' = 3823 GeV

mX = 732 GeV

mY = 378 GeV



Look at 1000 histograms



Anomaly detection - LHC Olympics

Box 1

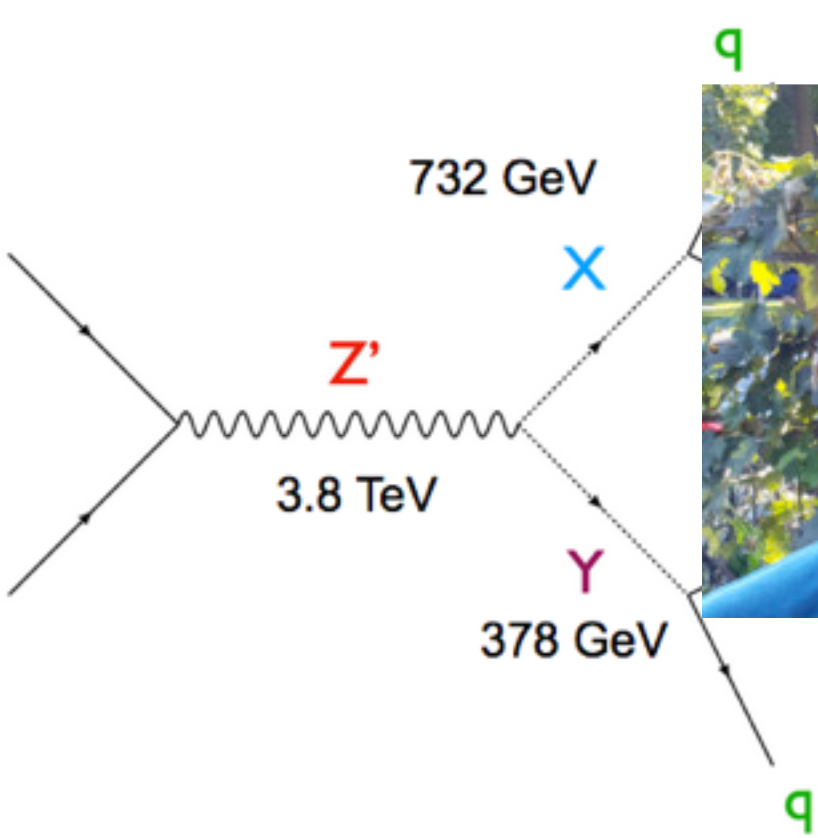
Signal: 834 events

Z'->XY; X,Y->qq
(same topology as R&D dataset)

mZ' = 3823 GeV

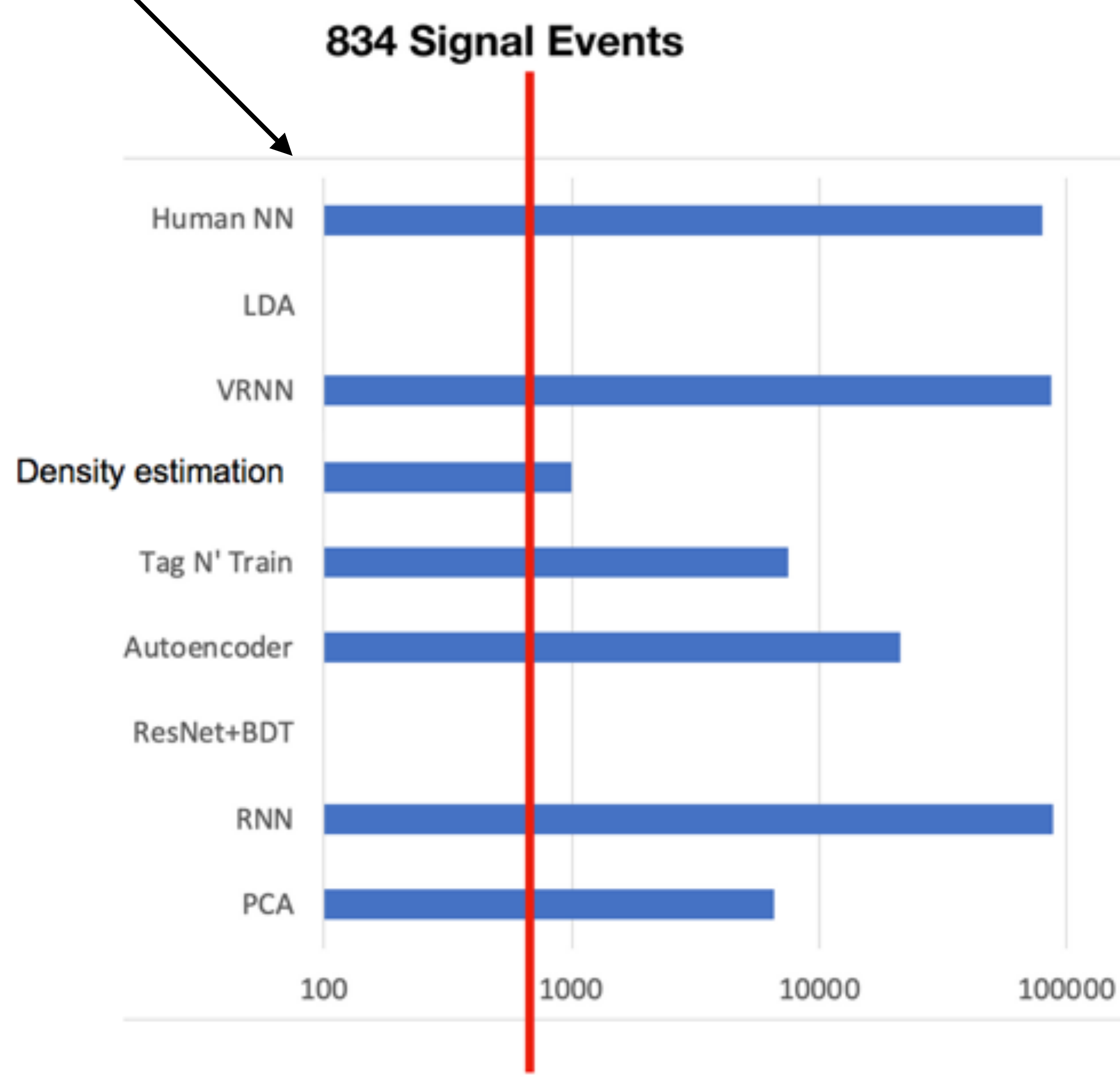
mX = 732 GeV

mY = 378 GeV



UCB Cosmology team

Look at 1000 histograms



LHC Olympics: next steps

- There will be another mini workshop just before BOOST where boxes 2 and 3 will be unveiled.

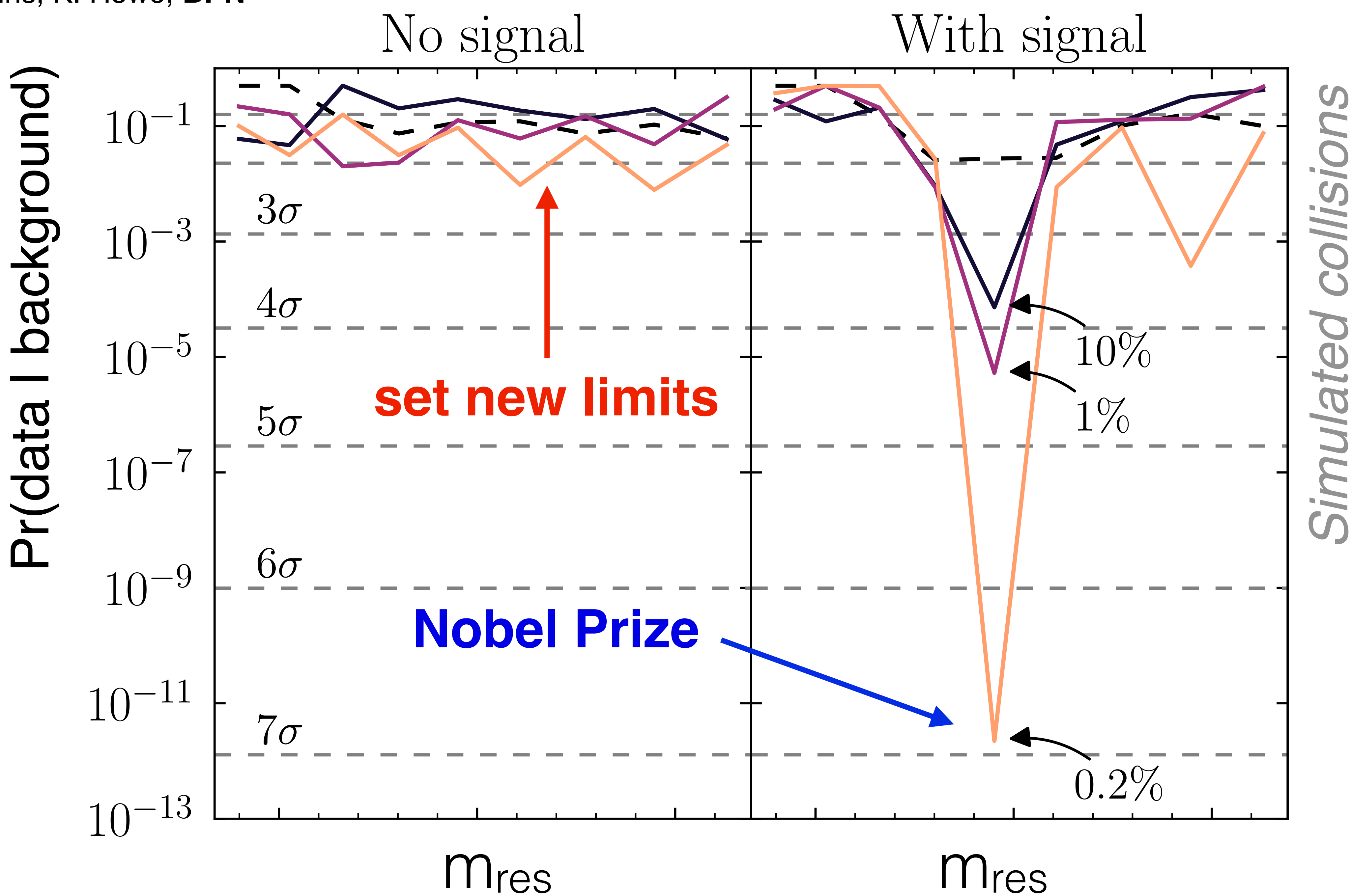
Anomaly Detection Mini-Workshop -- LHC Summer Olympics 2020

July 18, 2020
INF/AP Lecture Hall
Europe/Berlin timezone

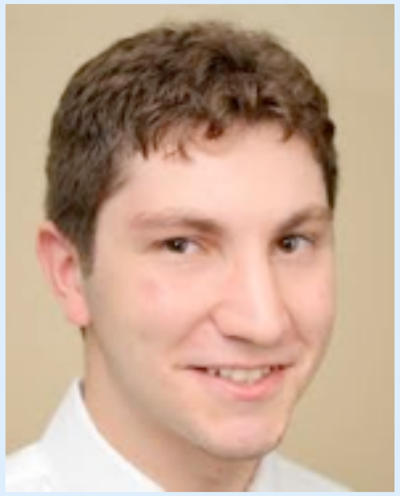
Anomaly detection - Bonus!

Phys. Rev. Lett. 121 (2018) 241803

J. Collins, K. Howe, **BPN**



Aviv Cukierman



*Our first data result from **ATLAS** will come out this spring!*

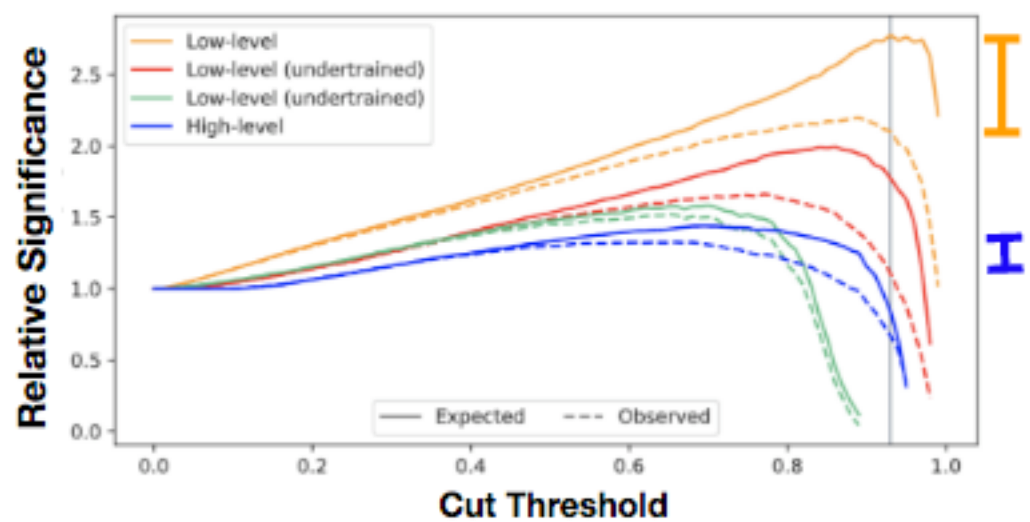
Inference and Interp. - Highlights

- Bayesian neural networks
  - Can encode some forms of uncertainty
- High-dimensional uncertainties with AI Safety (next slide)
- Unfolding
  - A generalization of iterative Bayesian unfolding (2 slides)
  - Conditional generative model (basically learn true given reco)
- Benchmarking (3 slides)

Machine Learning Inference and Interp.

We show that is possible to find systematic mismodellings $g(J) \mapsto J'$, that confuse NN classifiers

- These effects are subtle, remaining undetected in control/validation regions
- Susceptibility is reduced, but not entirely, when using fewer and higher-level inputs



High-dimensional uncertainties are a real challenge for deploying deep networks on low-level features!

What is AI Safety?

✓ We can all agree, we should try not to get murdered by robots

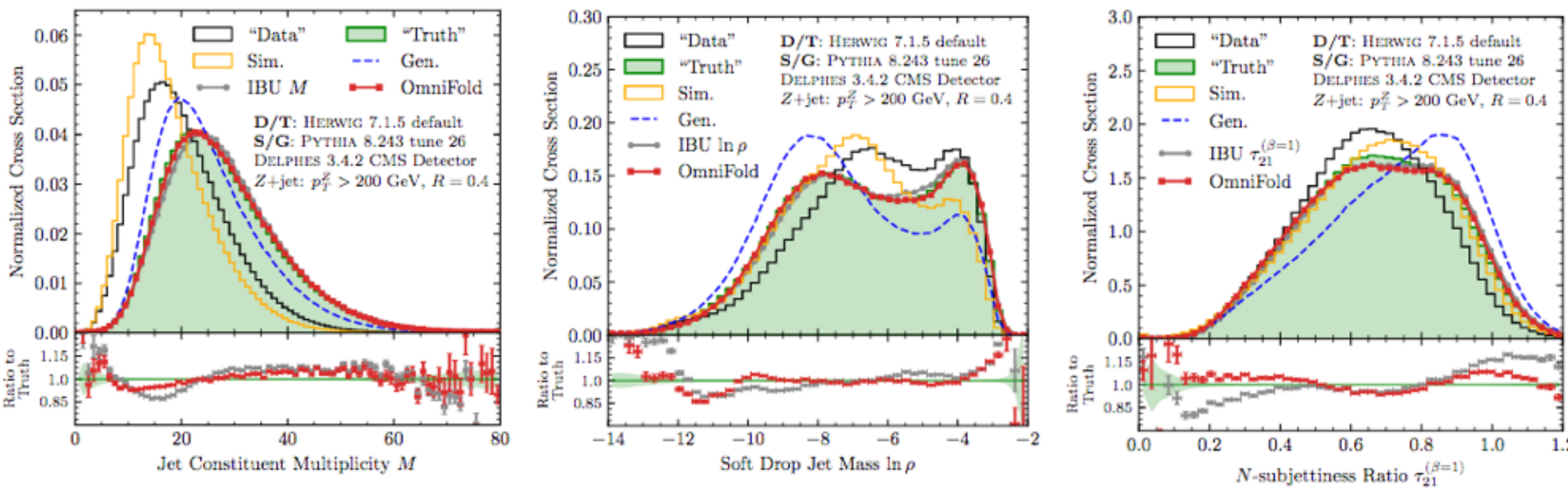
C. Shimmin and B. Nachman

Machine Learning Inference and Interp.

It is now possible to unfold the "full phase space" i.e. all of the 4-vectors and particle types in a single event !!

Single OmniFold instantiation vs. individual applications of IBU

(IBU = iterative Bayesian unfolding, standard in ATLAS)

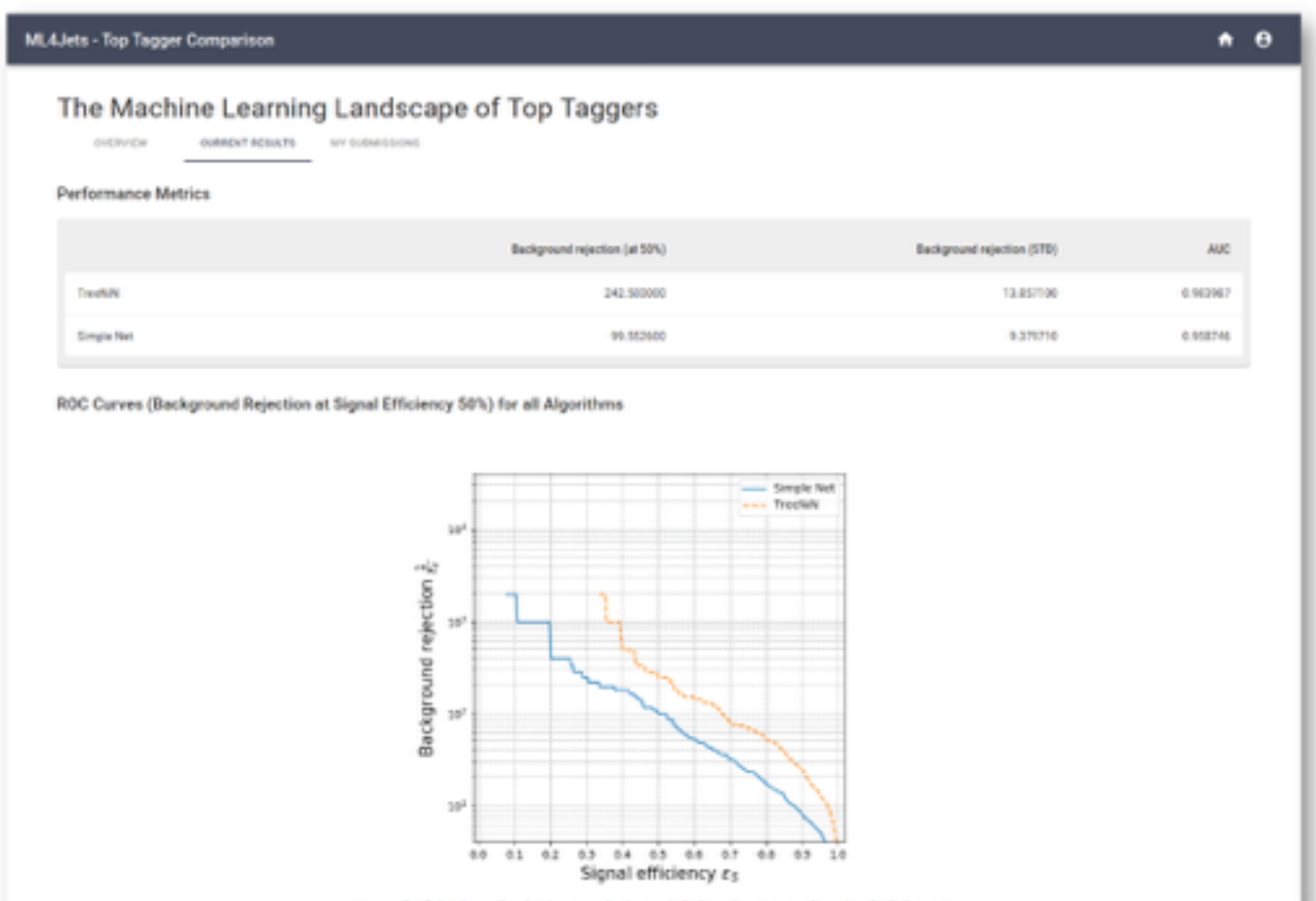


IRC unsafe | **IRC safe** | **Sudakov safe**

P. Komiske, A. Andreassen, E. Metodiev, B. Nachman, J. Thaler

Benchmarking

Included a live demo! Could be used for future challenges as well as for analysis in the collaboration - can we update all results at once if e.g. b-tagging is improved?



Reproducible Open Benchmarks for Data Analysis Platform

Kyle Cranmer, Irina Espejo,
Sebastian Macaluso, Heiko Mueller
New York University

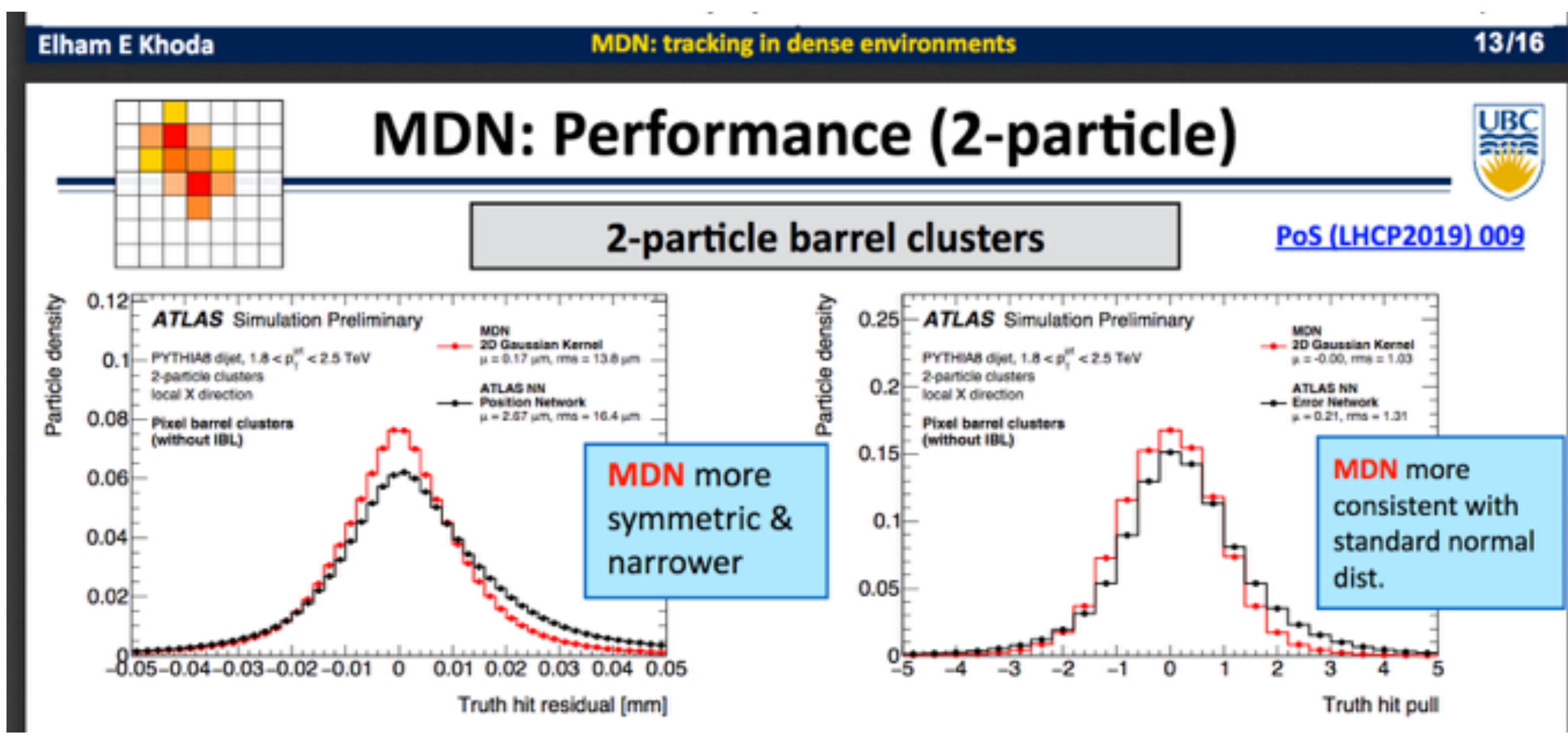
Shih-Chieh Hsu, Aaron Maritz,
Ajay Rawat, Cha Suaysom
University of Washington

(click this to see the slides)

Applications and experimental methods

- Vertexing with graph networks
  - Set to graph networks
- Mixture density networks for pixel clusters in dense environments

(Gaussian mixture model where mixture coefficients are NNs)

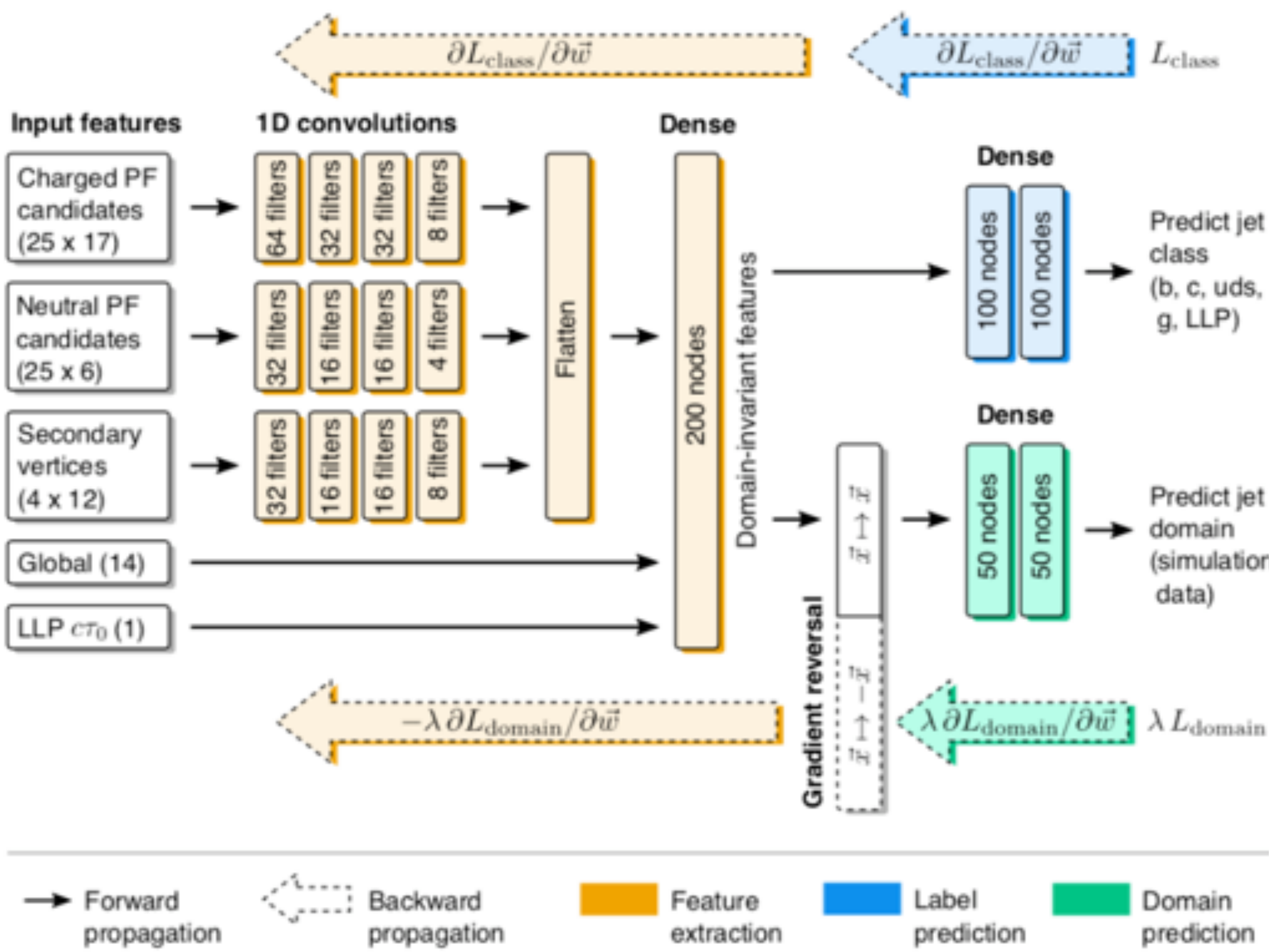


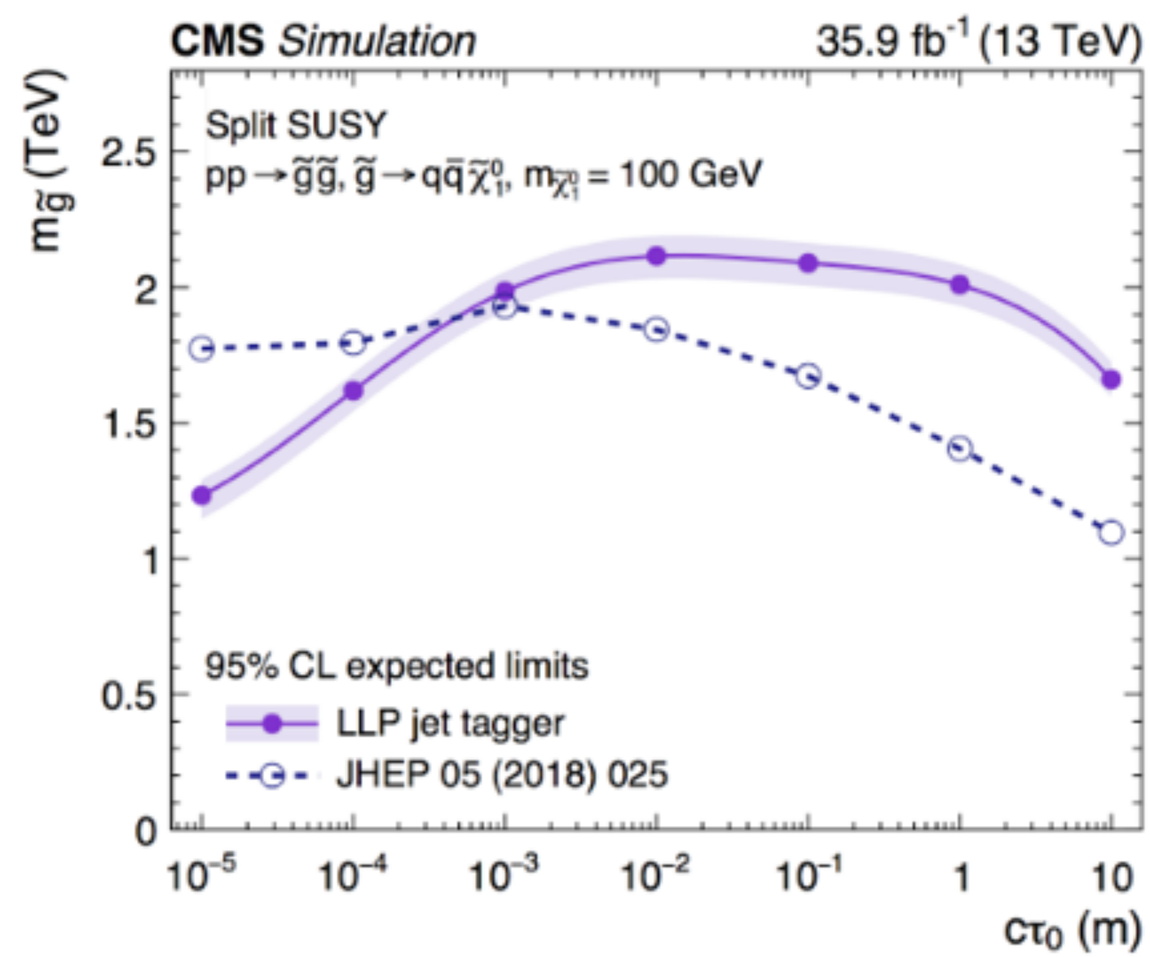
- Displaced jet tagging (next slide)

Applications and experimental methods

Displaced jet tagging in CMS - very sophisticated!

high dimensional + decorelation





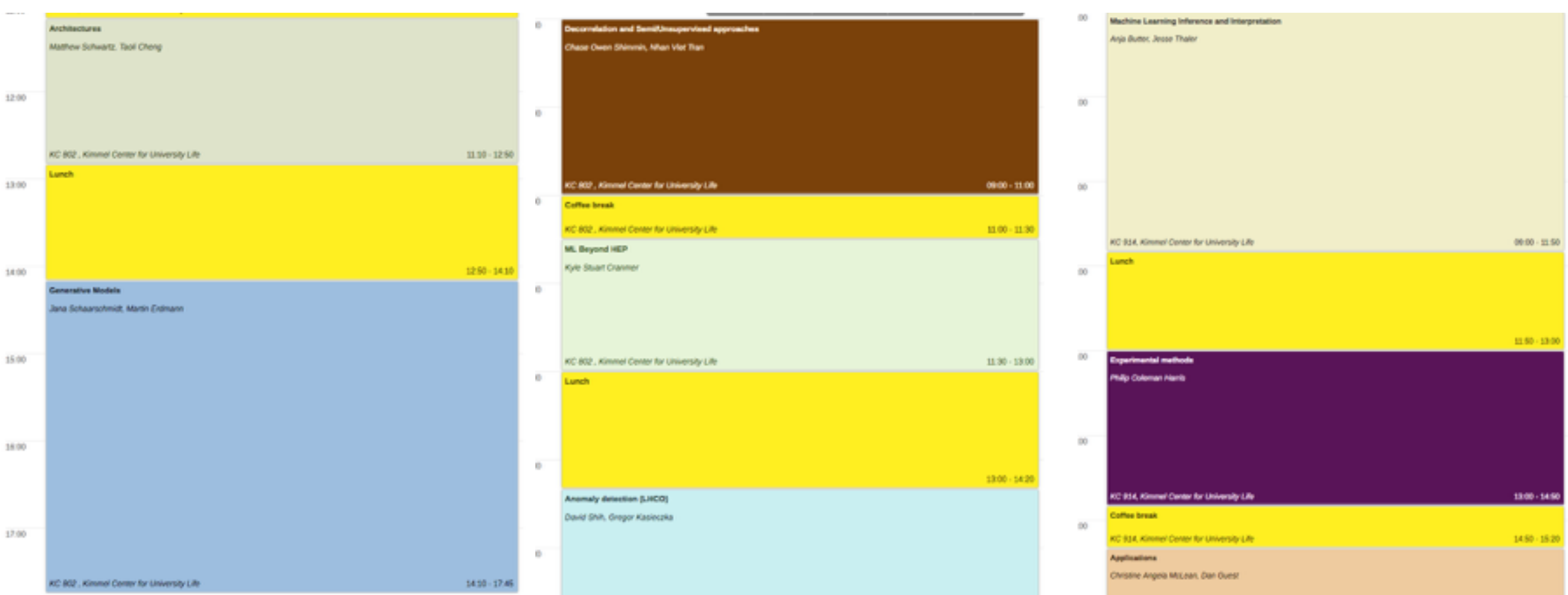
no idea how to do signal uncertainties
… and they don't either

R. Bainbridge, for CMS

Summary

There was a packed agenda with three very full days of interesting talks!

I won’t review every talk … please take a look at the slides for all the contributions and for details. These slides are some personal highlights.



Outlook

ML4Jets 20/21

Philosophenweg Heidelberg, week of Dec 14, 2020

Questions?