ML based Anomaly Detection at the LHC

Oz Amram

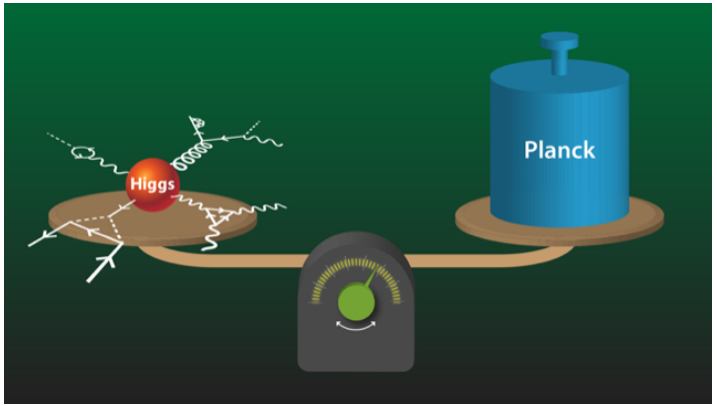
JOHNS HOPKINS UNIVERSITY

U Chicago
Rising Stars Symposium
September 2021

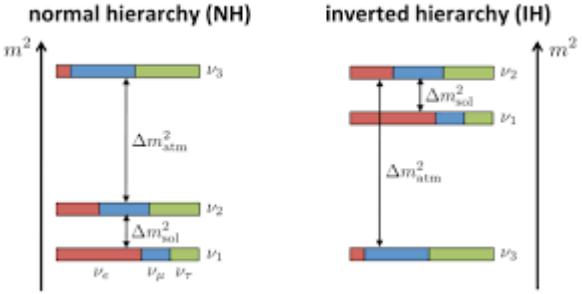
Lots of Questions

What is Dark Matter?

Why is the Higgs so light?



What is the origin of Neutrino Mass?



Baryogenesis?

Grand Unification?

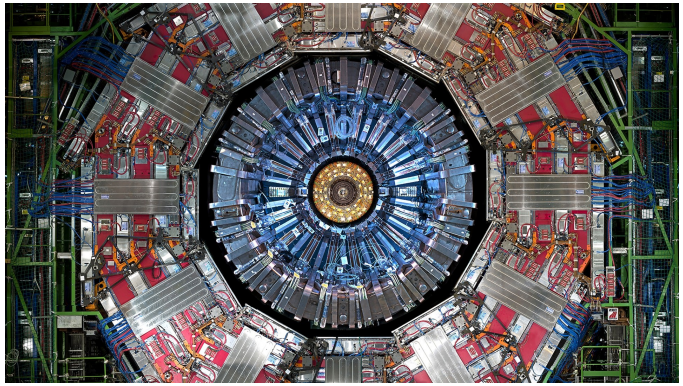
And many more…

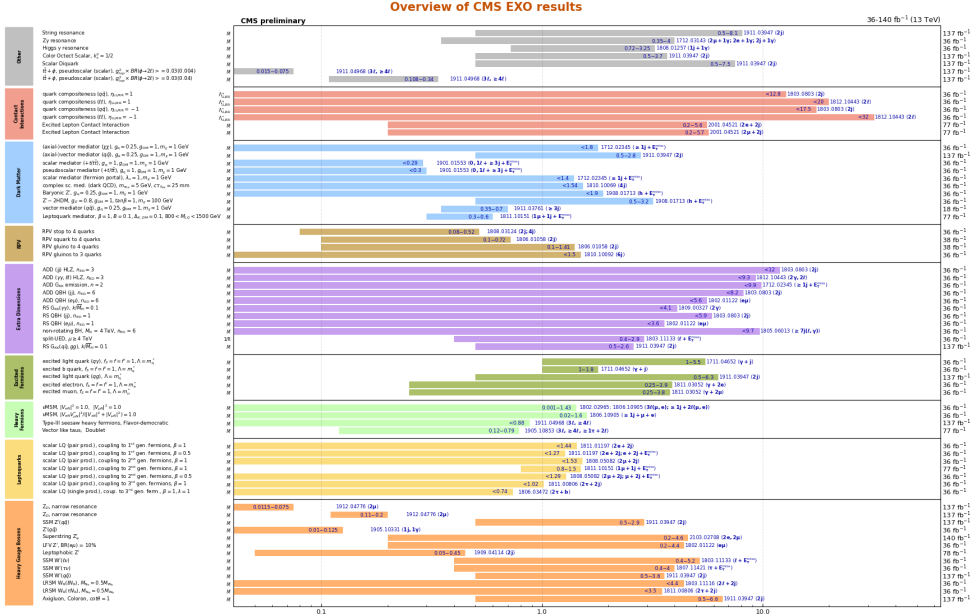
g-2?

Flavor Anomalies?

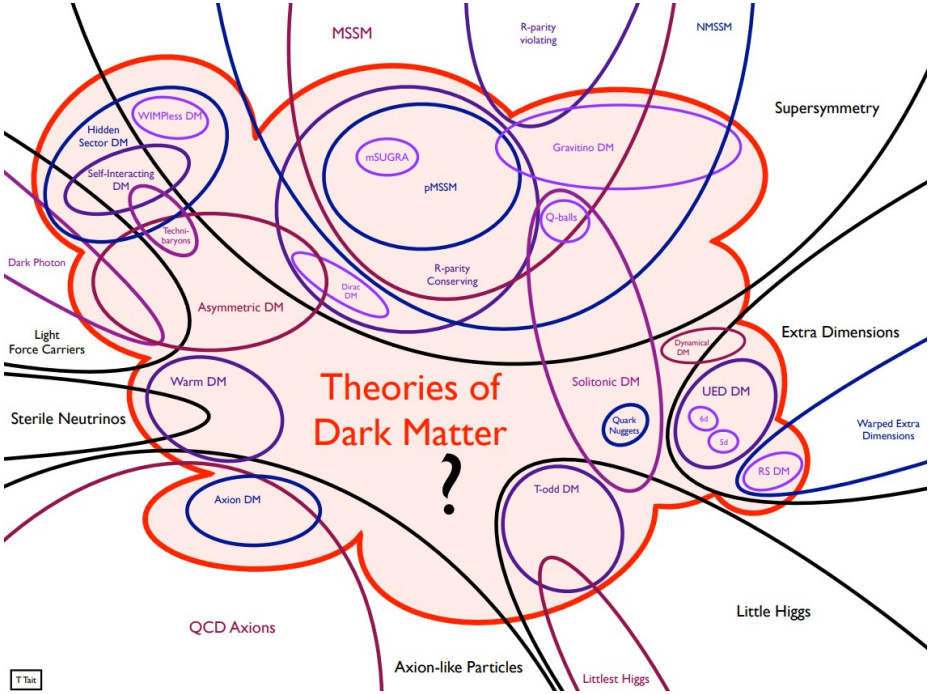
Hierarchy Problem Image from : https://physics.aps.org/articles/v13/174

No clear answers from the LHC yet

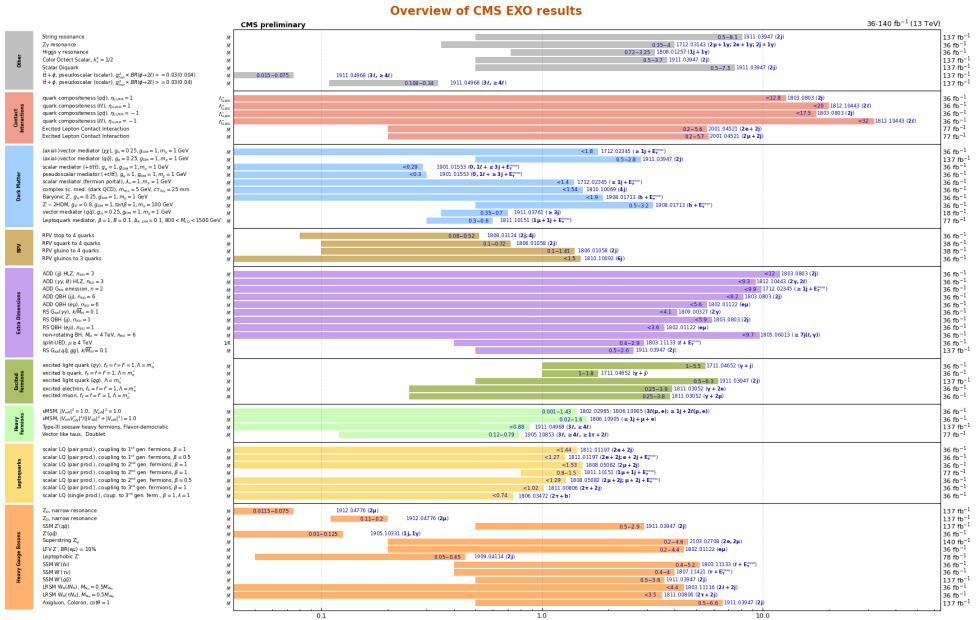




But...


MSSM
R-parity violating
NMSSM
Supersymmetry
WIMPless DM
Hidden Sector DM
Self-Interacting DM
mSUGRA
pMSSM
Gravitino DM
Q-balls
Technicolor baryons
Dark Photon
R-parity Conserving
Dirac DM
Asymmetric DM
Light Force Carriers
Dynamical DM
Extra Dimensions
Warm DM
Theories of Dark Matter
?
Solitonic DM
UED DM
6d
5d
Quark Nuggets
Warped Extra Dimensions
Sterile Neutrinos
RS DM
Axion DM
T-odd DM
Little Higgs
QCD Axions
Axion-like Particles
Littlest Higgs
T Tait

>


Overview of CMS EXO results
CMS preliminary
36-140 fb^{-1} (13 TeV)
Other
Contact Interactions
Dark Matter
SUSY
Extra Dimensions
Excited Fermions
Heavy Fermions
Leptoquarks
Heavy Gauge Bosons
mass scale [TeV]
Moriond 2021
Selection of observed exclusion limits at 95% C.L. (theory uncertainties are not included).

What if we aren't looking in the right places ?!

The Challenge

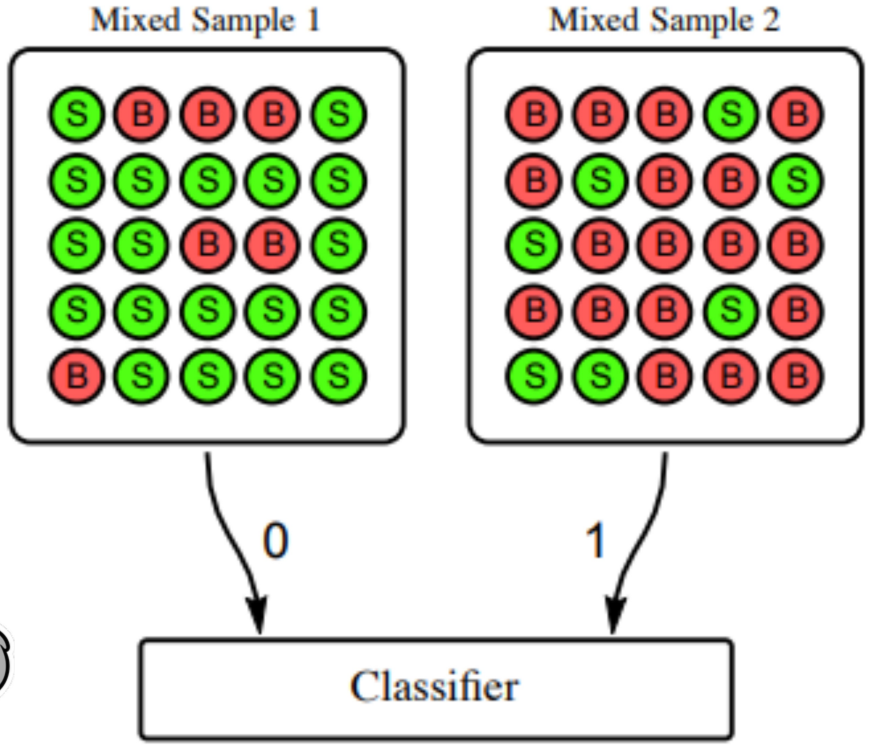
- How can we design searches with minimal assumptions but still have powerful sensitivity?

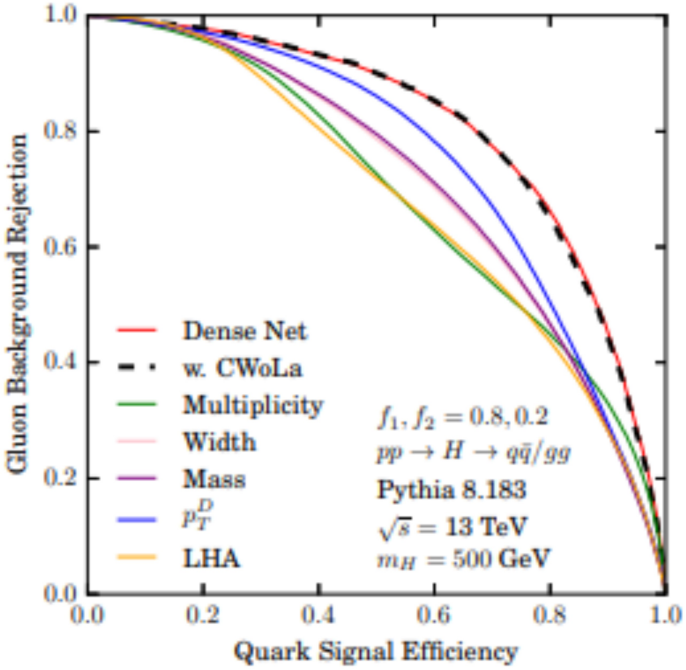
- New ideas in ML are enabling totally new search strategies!

Key Idea: Train directly on data!

Classification Without Labels

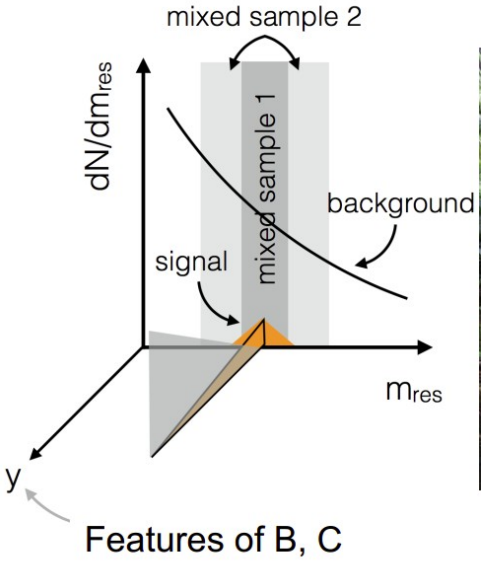
Metodiev, Nachman & Thaler 1708.02949

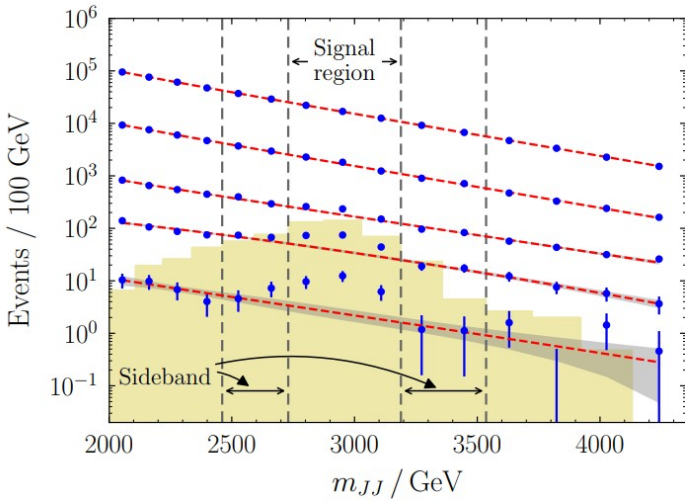




CWoLa Hunting

Colins, Howe & Nachman 1902.02634

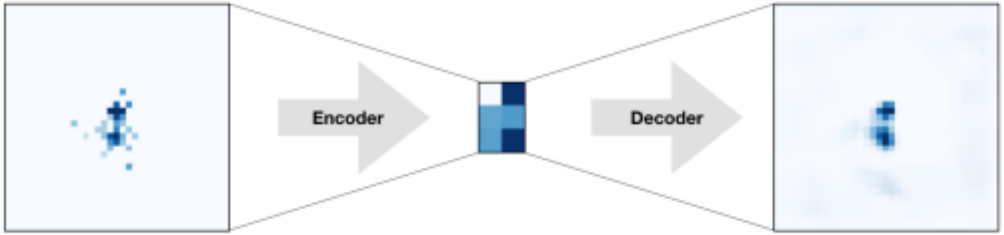


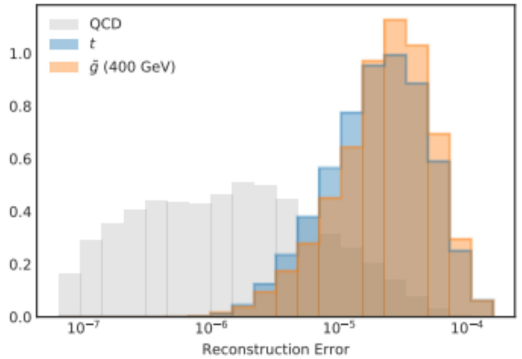


- Signal region = dijet mass window
- Train a classifier on signal region vs. others
- Select events & bump hunt

Bump-hunt graphic from here

Anomaly Detection : Autoencoders





- Train a network to compress and decompress the data
- Can train directly on data, no labels needed
- Anomalous events should have a higher reconstruction loss

Heimel, Kasieczka, Plehn & Thompson 1808.08992
Farina, Nakai & Shih 1808.08979
+ Others

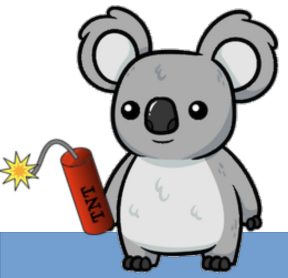
Drawbacks

- CWoLa Hunting
  - Worry about sculpting QCD dijet mass distribution
  - Apply to non-resonant signals?

- Autoencoders
  - Only ‘learns’ what QCD looks like
  - Room for improvement as a Sig vs. Bkg classifier

Tag N' Train (TNT)

- A method of training improved classifiers on data
- Assumptions:
  - Signal has **2 interesting objects** in it
  - One has a **starting classifier** for each object
  - Signal-like features in background events are uncorrelated between the 2 objects

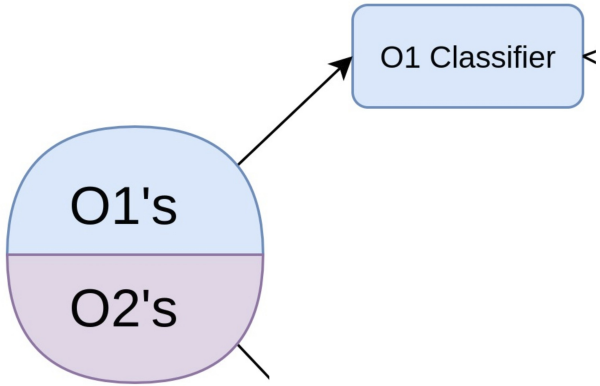


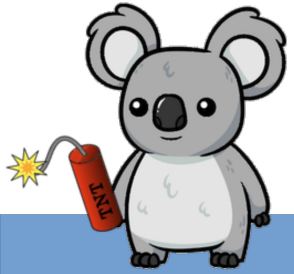
Tag with a weak classifier **N' Train** a better one!

OA & Suarez 2002.12376

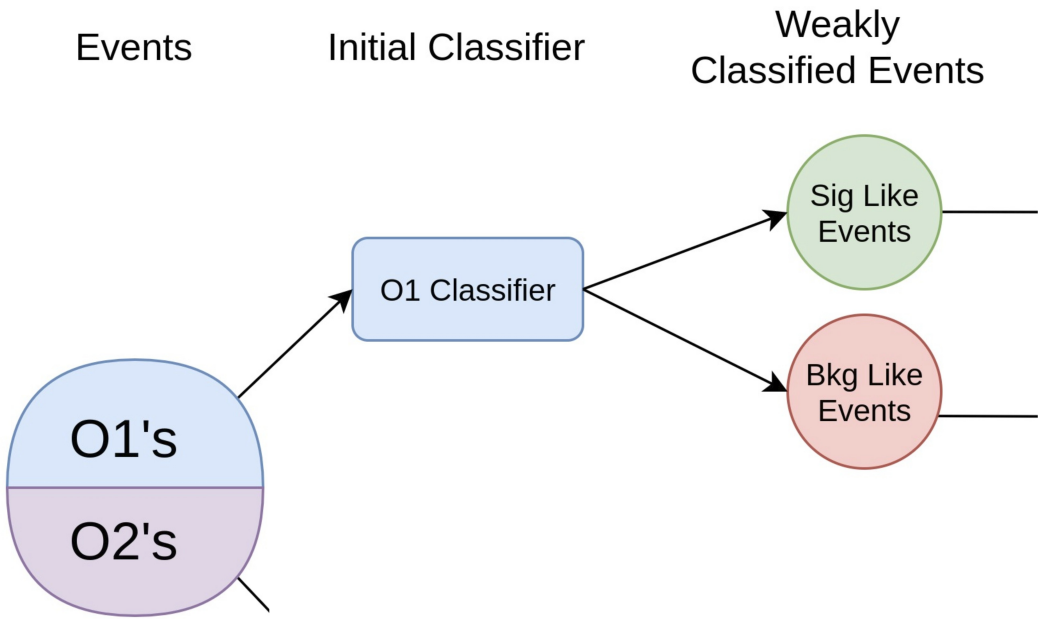
Tag N’ Train

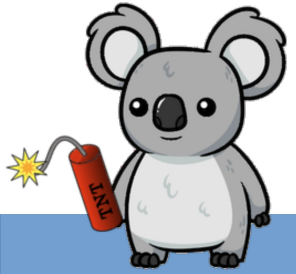
Events Initial Classifier



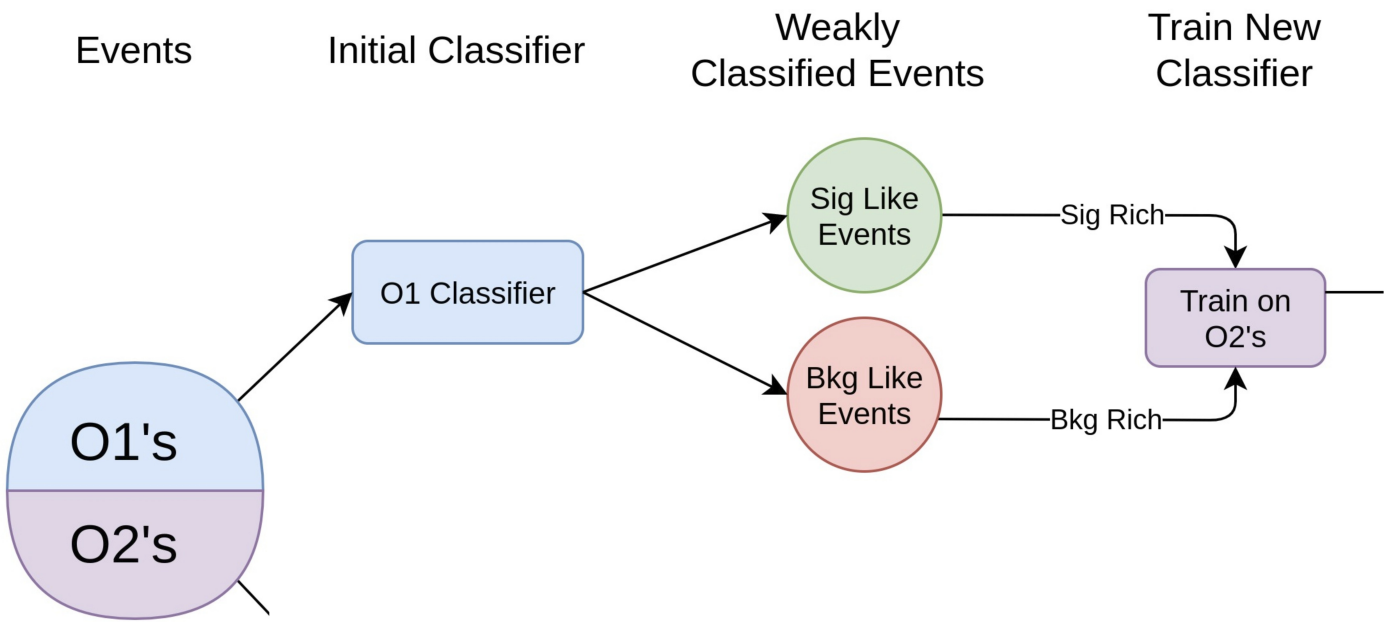


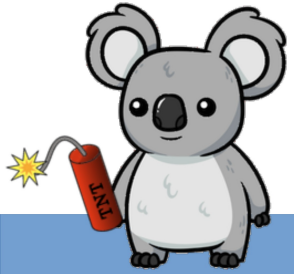
Tag N' Train


Events
Initial Classifier
Weakly Classified Events
O1's
O2's
O1 Classifier
Sig Like Events
Bkg Like Events

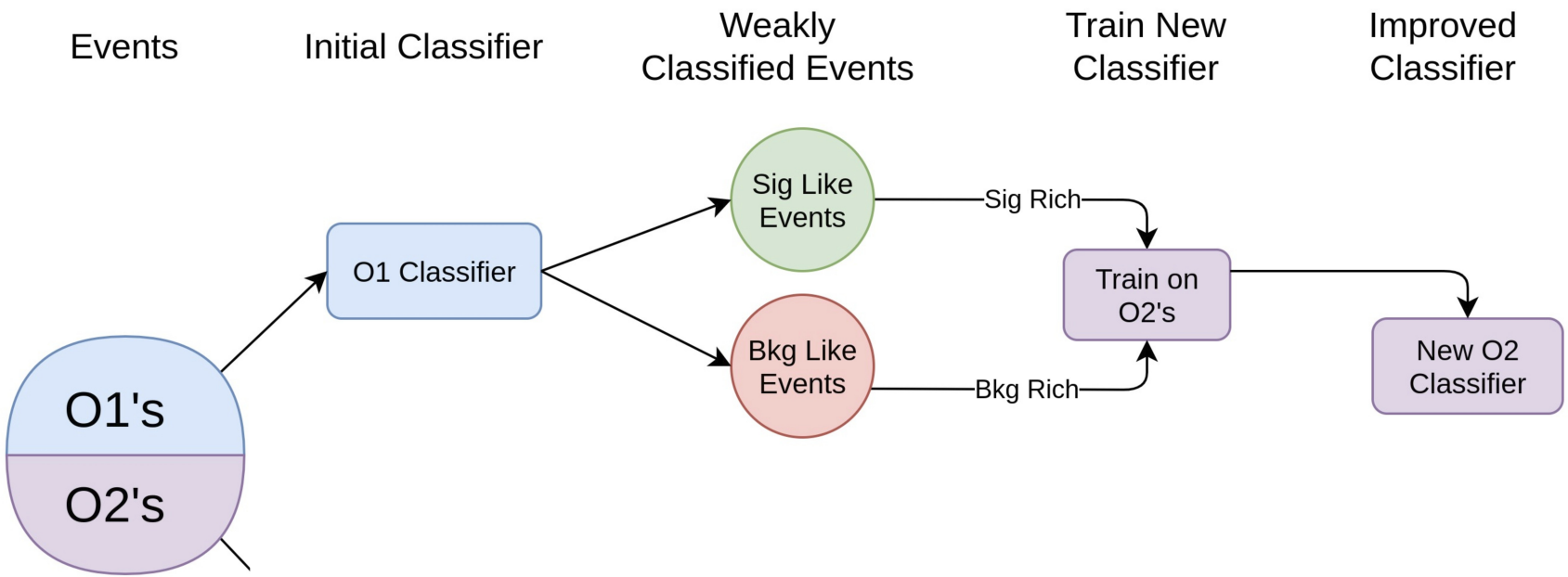

TNT

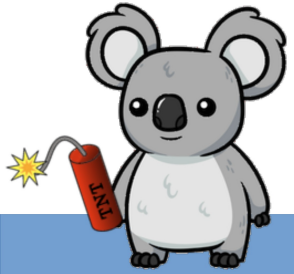
Tag N' Train


Events
Initial Classifier
Weakly Classified Events
Train New Classifier
O1's
O2's
O1 Classifier
Sig Like Events
Bkg Like Events
Sig Rich
Bkg Rich
Train on O2's

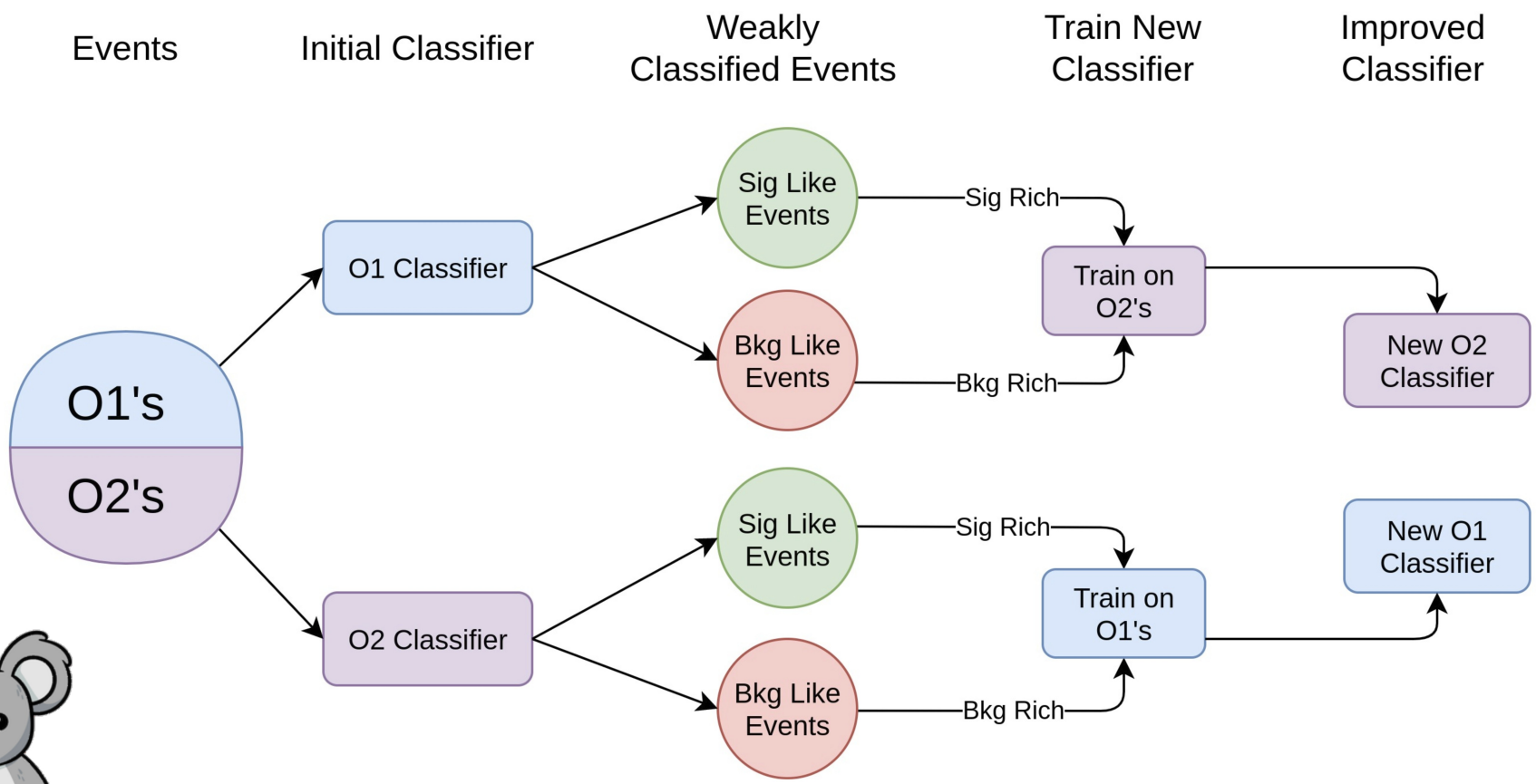

TNT

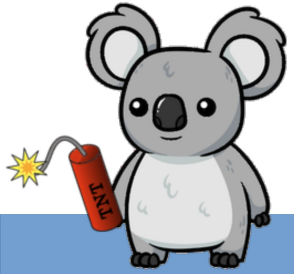
Tag N' Train


Events
Initial Classifier
Weakly Classified Events
Train New Classifier
Improved Classifier
O1's
O2's
O1 Classifier
Sig Like Events
Bkg Like Events
Sig Rich
Bkg Rich
Train on O2's
New O2 Classifier


TNT

Tag N' Train


Events
Initial Classifier
Weakly Classified Events
Train New Classifier
Improved Classifier
O1's
O2's
O1 Classifier
O2 Classifier
Sig Like Events
Bkg Like Events
Sig Like Events
Bkg Like Events
Sig Rich
Bkg Rich
Sig Rich
Bkg Rich
Train on O2's
Train on O1's
New O2 Classifier
New O1 Classifier

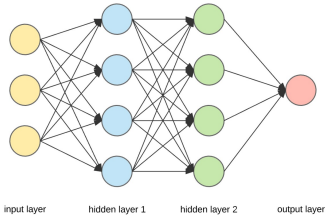

TNT

Dijet Anomaly Search

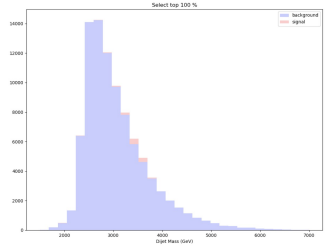
Applying TNT to a Resonance Search

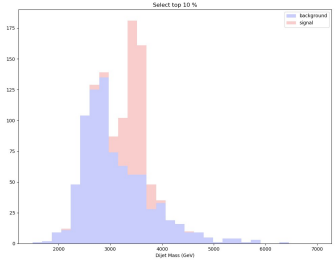
Events A

Events B

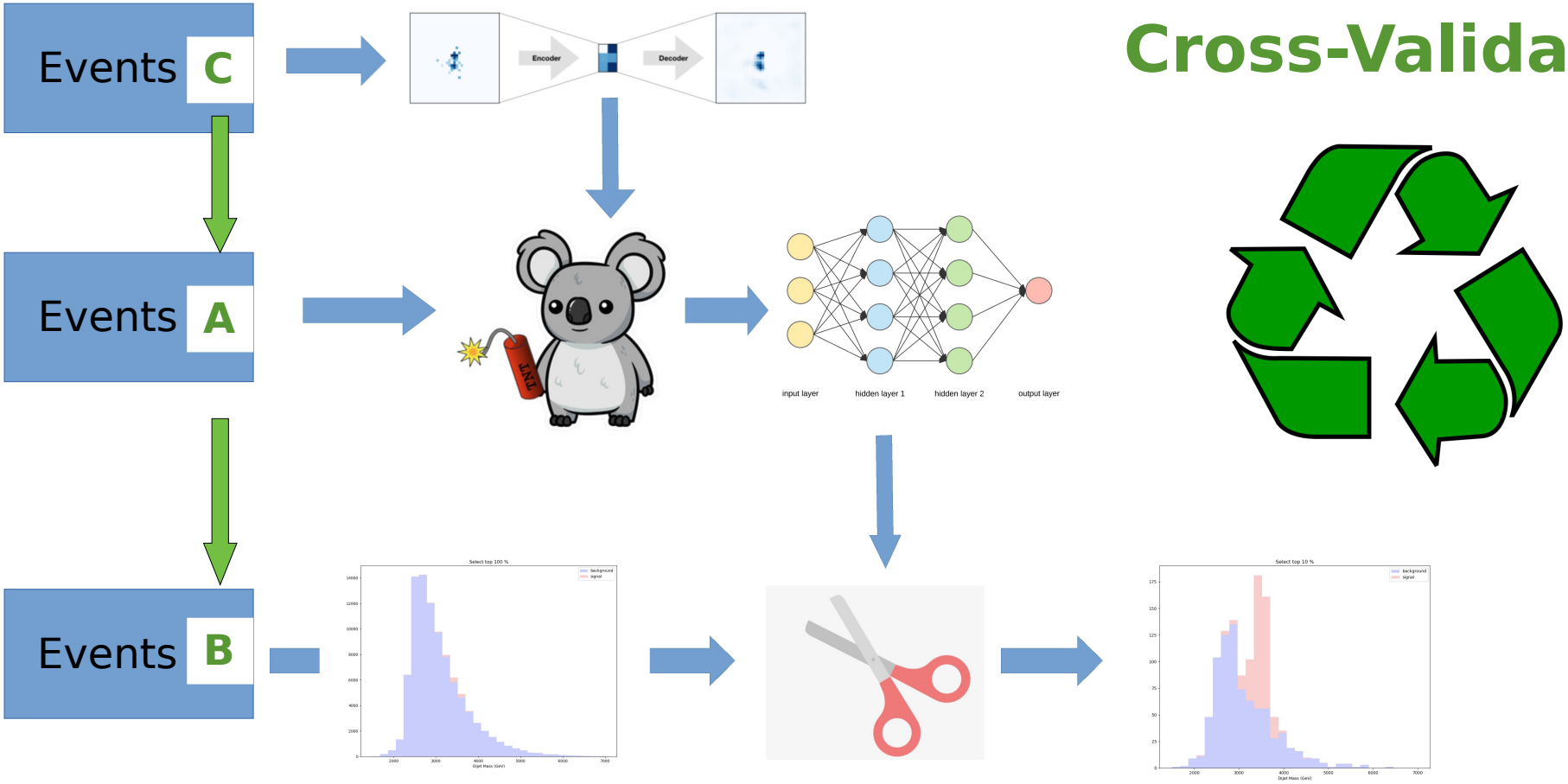


Events C





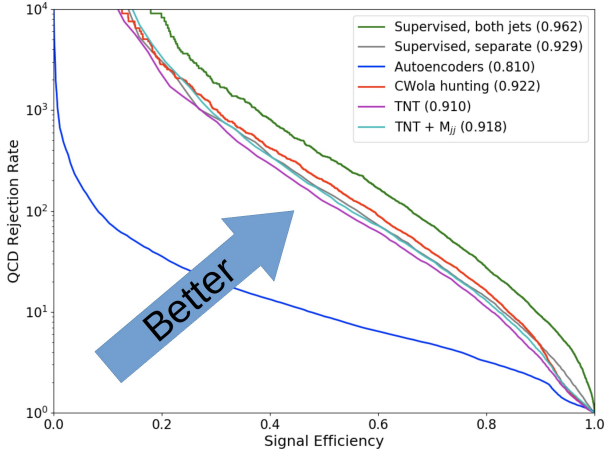
Applying TNT to a Resonance Search



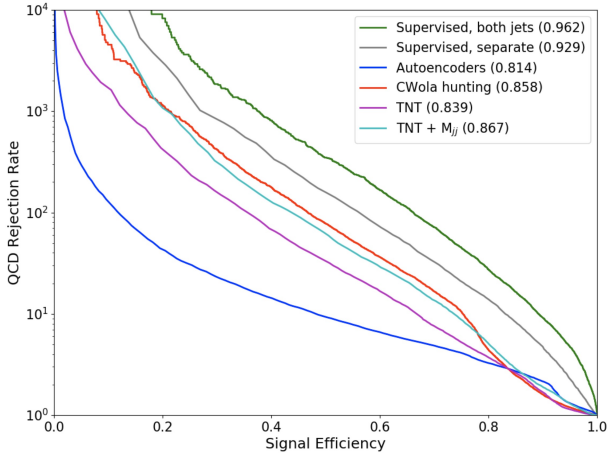
Cross-Validate!

Classification Performance

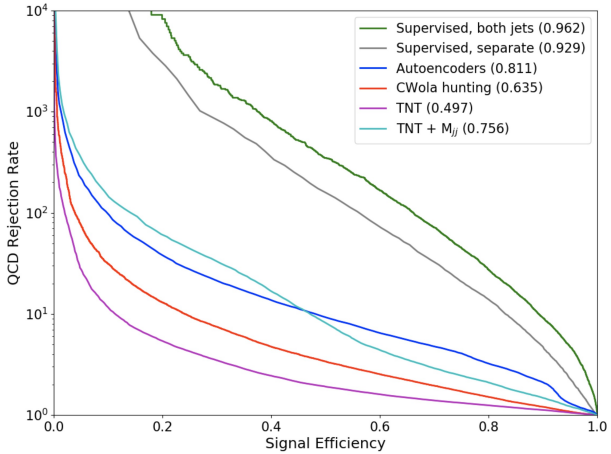
S/B = 35%



S/B = 5%



S/B = 0.5%

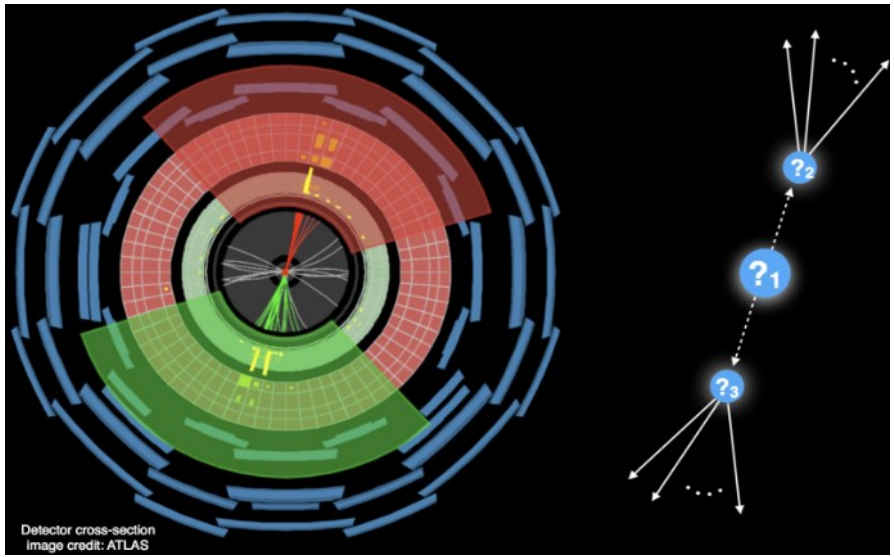


Less Signal

- CWoLa based methods approach **supervised** case when lots of signal
- **Autoencoders** performance independent of signal
- **TNT** matches **CWoLa hunting** high/medium signal, better at low signal

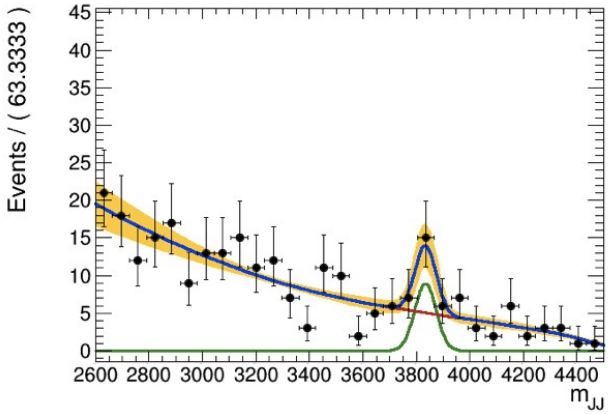
In Action: LHC Olympics 2020

- A competition to test out these new anomaly detection methods
- Blackboxes with:
  - 1M events, R=1 jet pt > 1.2 TeV trigger
  - 4 vectors of all reconstructed particles
  - Mostly background + some hidden new physics (?)



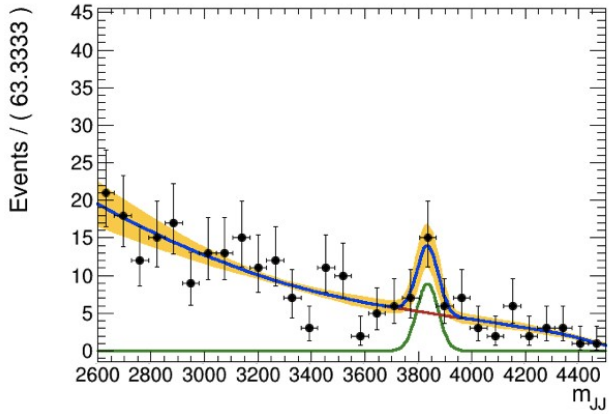
arXiv:2101.08320

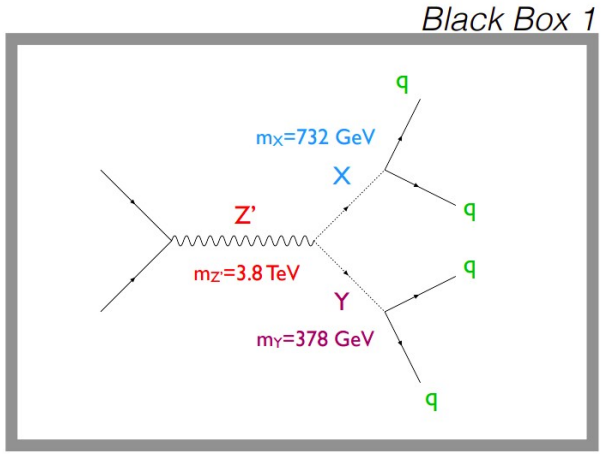
LHC Olympics Results



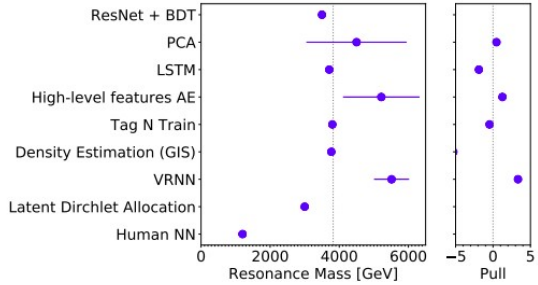
- TNT found a resonance at ~3800 GeV with 4σ evidence

LHC Olympics Results



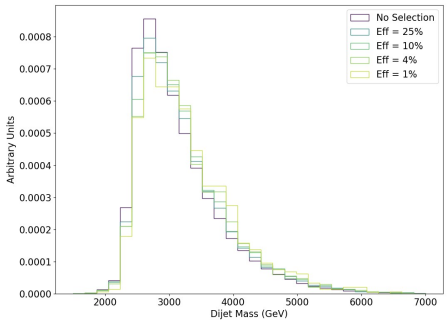


- TNT found a resonance at ~3800 GeV with 4σ evidence
- One of the few groups able to find the signal!

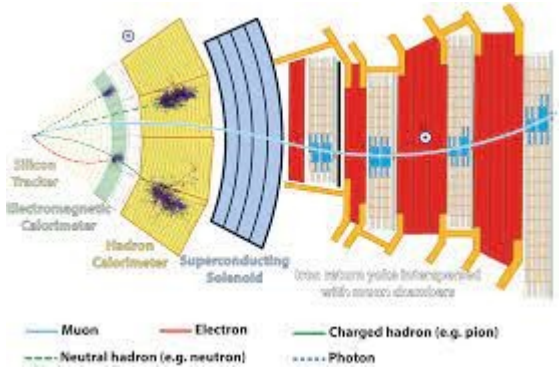


Many Challenges in Applying to Real Data!

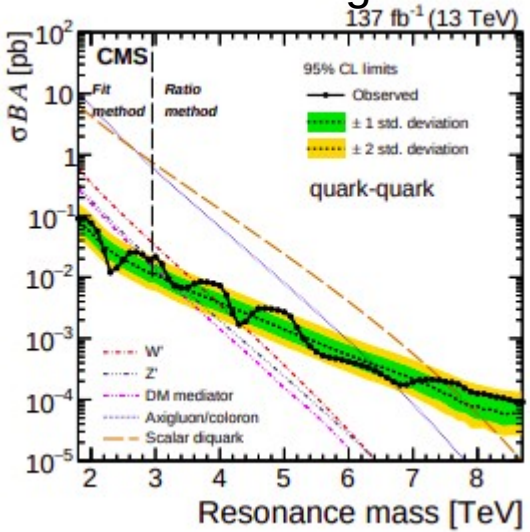
Ensure no mass sculpting



Don't "discover" a detector glitch!

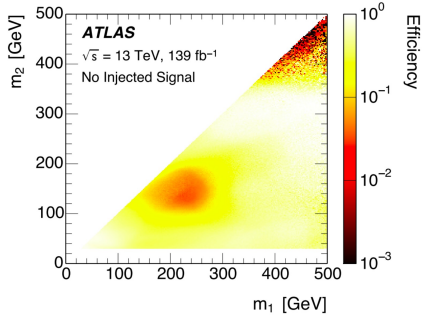


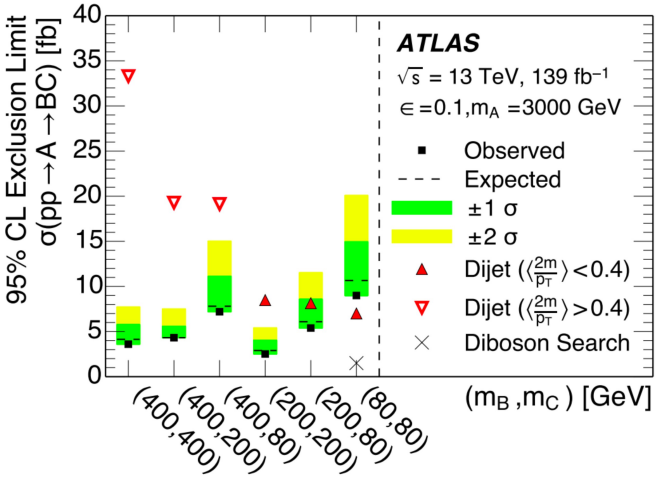
Limit setting ?!



Results on Data

ATLAS: 2005.02983



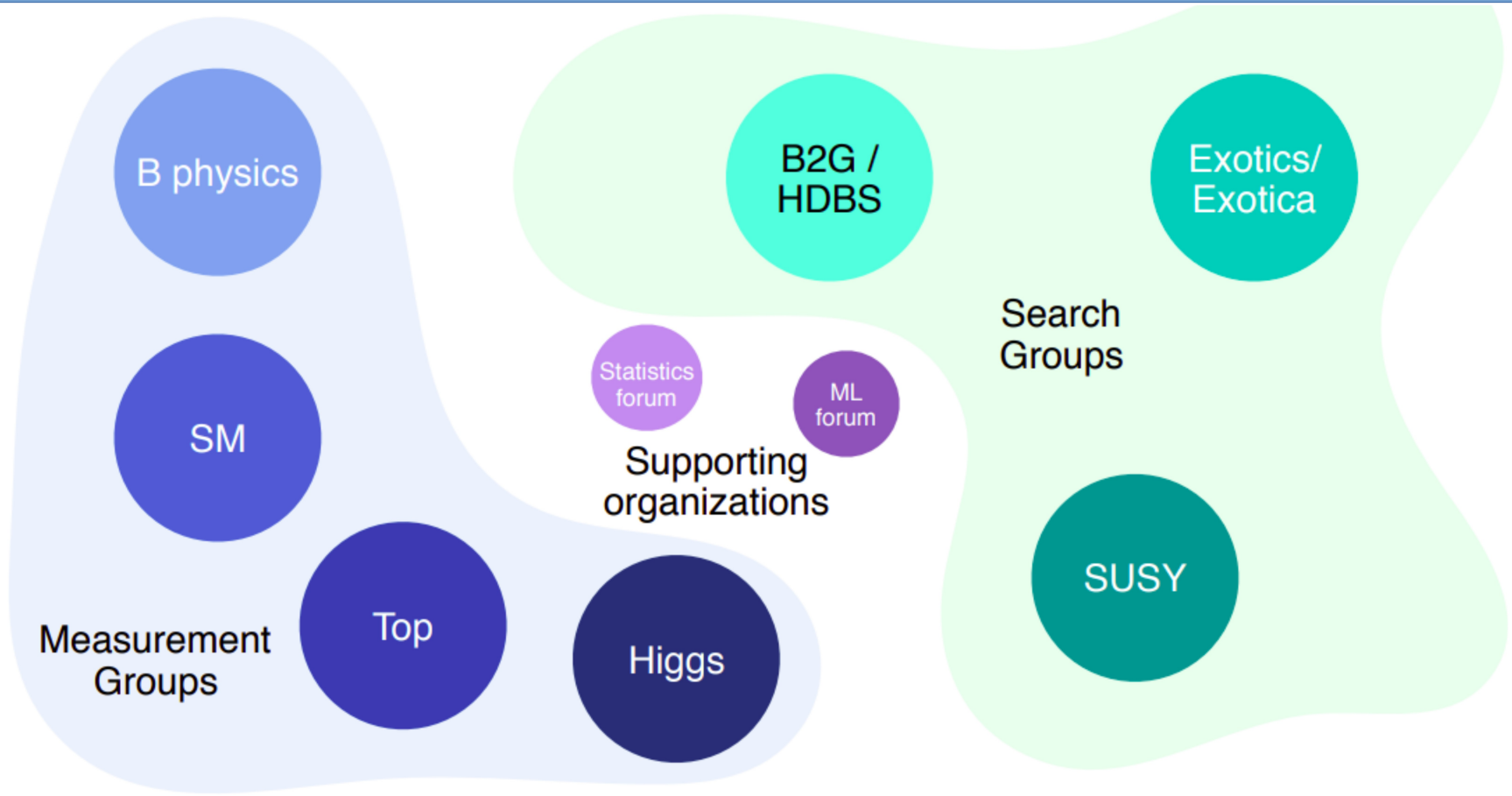


CMS

Just the Beginning...

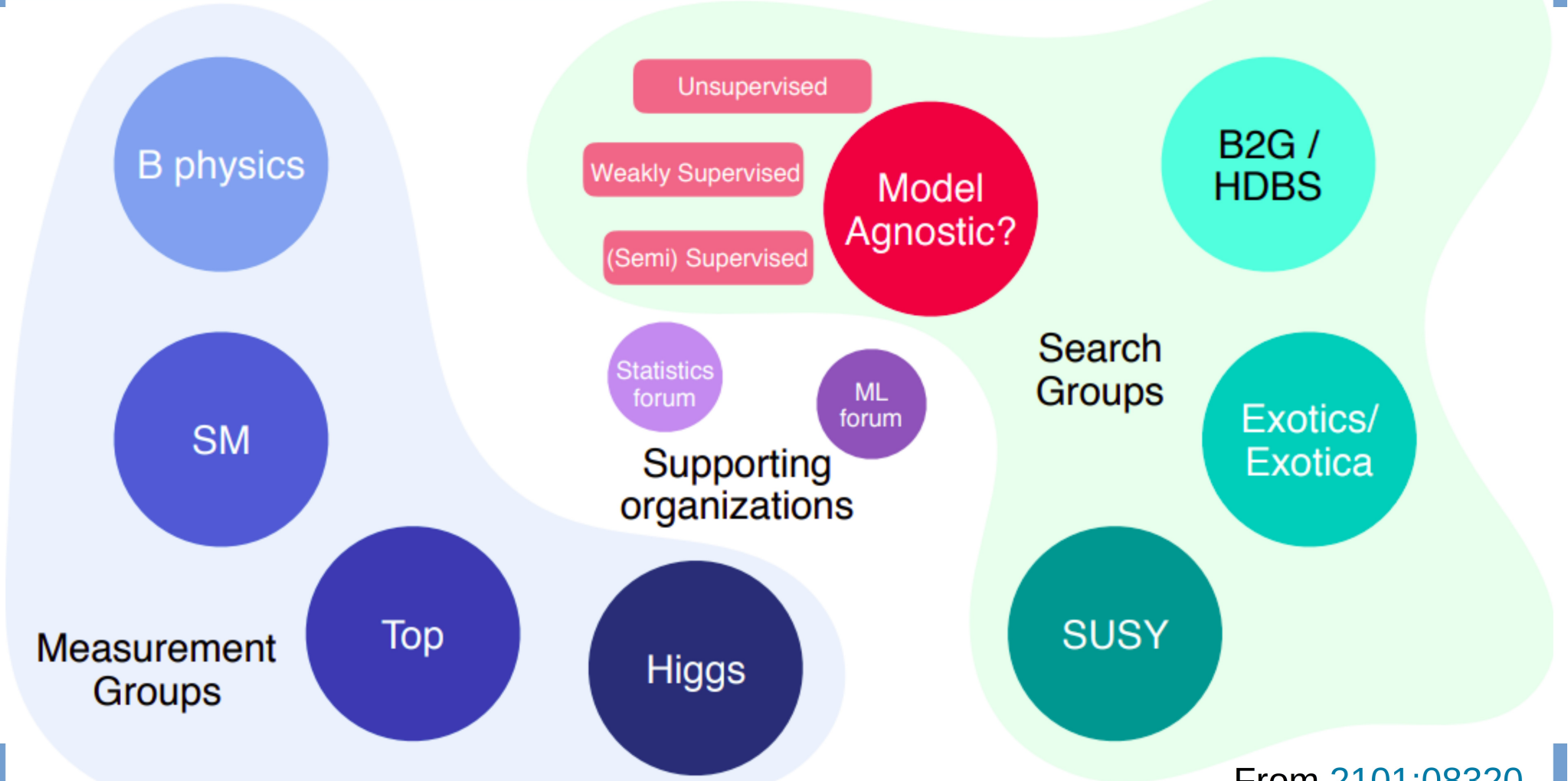
- New techniques!
  - New ideas innovations from the ML side
  - Hybrid approaches with traditional searches
- New Searches!
  - Other anomalies besides jets with substructure
  - Non-resonant searches
- Do it fast!
  - Incorporate these ideas into triggers
  - Recently announced Anomaly Detection at 40 MHz challenge!

Current LHC Analysis Group Organization



From 2101:08320

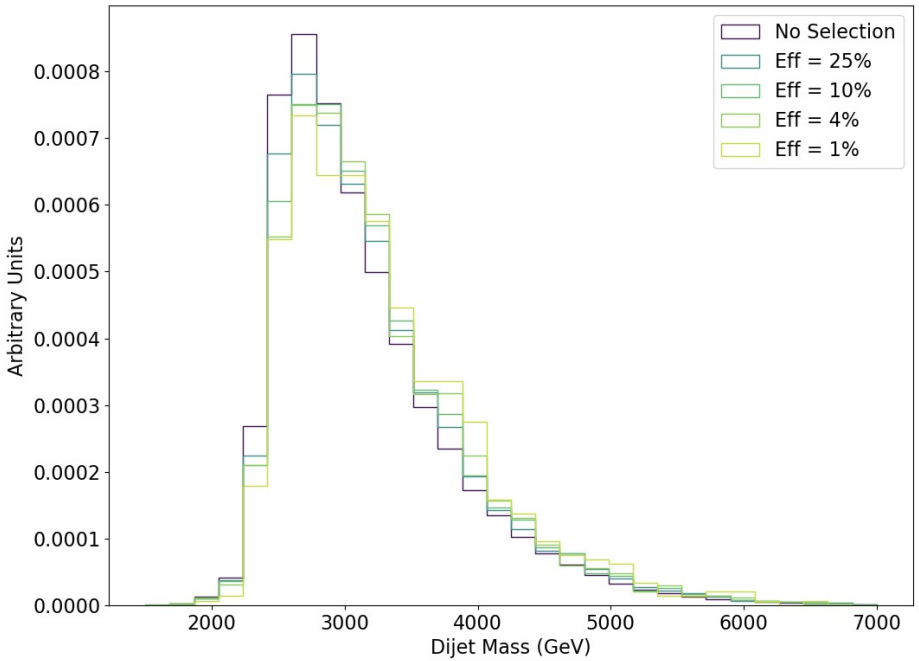
In 10 Years?



From 2101:08320

Backup

Dijet Mass Sculpting



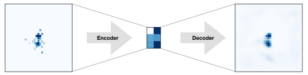
- No sculpting of dijet mass!
- Decorrelation methods also possible with TNT
  - p_T reweighting tried, found no difference

TNT Technical Details

- 2 objects: heavy jet and light jet in event
- TNT Classifiers and autoencoders are CNN's based on jet images
- Top 20% 'sig-like', bottom 40% 'background-like'
  - Optional: require signal events in dijet mass window
- Combine 2 classifiers into 1
  - Require both jet's scores be in top X% of scores

Trade Offs*

(V)AE's



+ Performance indep. of amount of signal

+ Minimal assumptions

– Inherently 'anti-QCD' rather than a 'pro-signal'

CWoLa Hunting

+ Great performance for large to medium signals

+ Can do full-event classification

– Assumption: resonant signal

– Must fully decorrelate features with M_{jj}

TNT

+ Great performance for medium/large signals and maintains performance for smaller signals

+ Mass sculpting mitigation possible

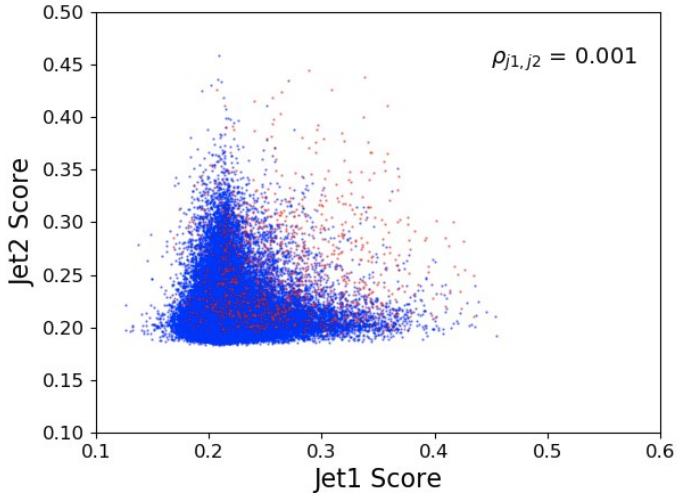
– Requires a starting classifier

– Assumption: Signal has 2 interesting objects

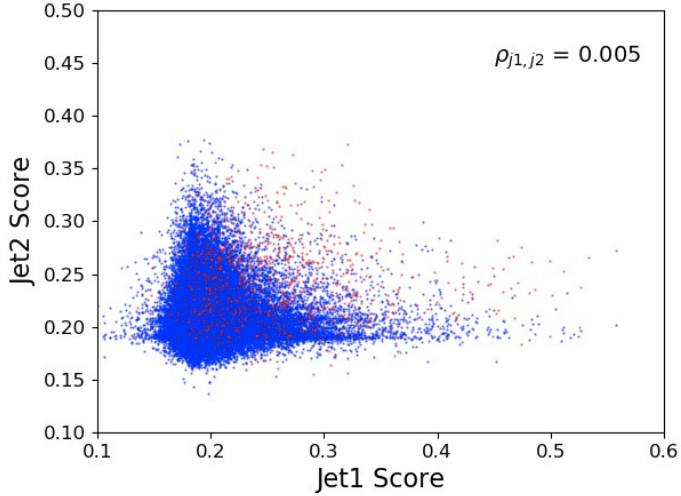
*Of course there are other interesting techniques with different trade offs too

Assumption: Correlations

"Pure" CWoLa



Tag N' Train



- Key assumption: Anomalous features of background events are uncorrelated
- Empirically (?) seems to hold