Jet Substructure

Daniel Whiteson, UCI

Collaborators:
Pierre Baldi, Peter Sadowski UCI CS
Taylor Faucett, Edison Weik, Kevin Bauer, Chase Shimmin, UCI Physics
Andreas Soegaard, Edinburgh Physics
Jesse Thaler, MIT Physics

Our HEP-ML work

1402.4735: DL for event selection

1410.3469: DL for Higgs tau tau

1601.07913: Parameterized NN

1603.09349: DL jet substructure

1607.08633: DL jet flavor tagging

1703.03507: Adv, parametrized DL

Next : Interpreting DL jetsub. solutions

Outline

I. Jet substructure classification

II. Decorrelated jet substructure

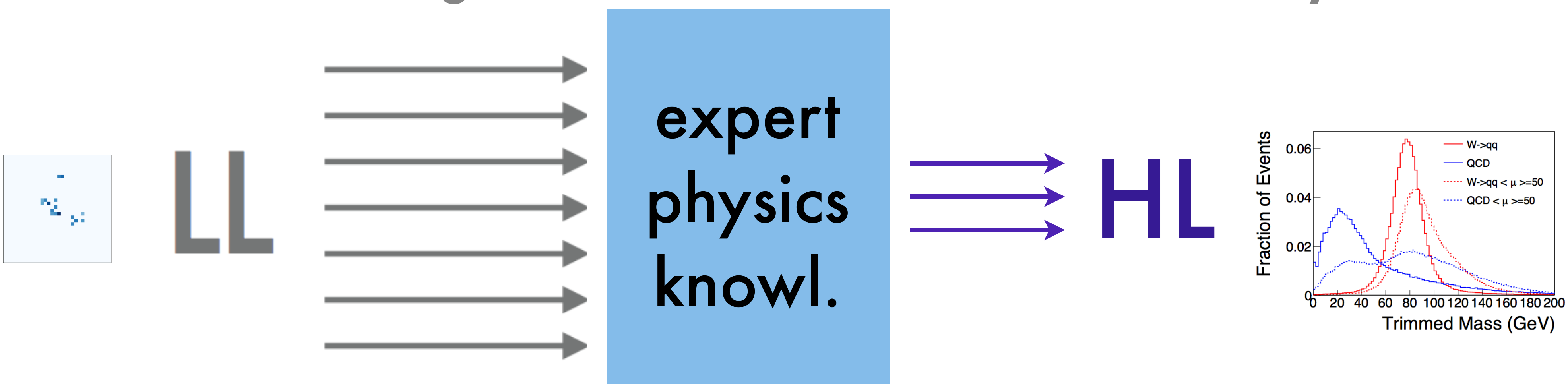
III. Interpreting ML

Approach

[Use a structured dataset]

LL: lower-level, higher dimensionality

HL: higher-level, lower dimensionality



HL is a strict function of LL

If NN(LL) > NN(HL)

HL has lost information

Jets LL

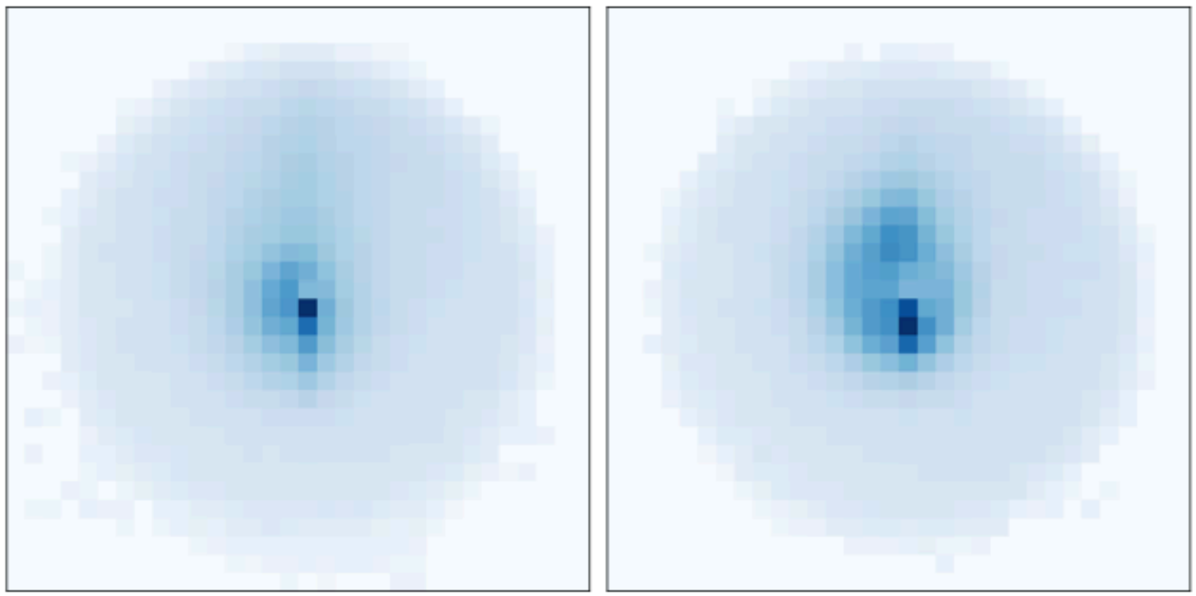
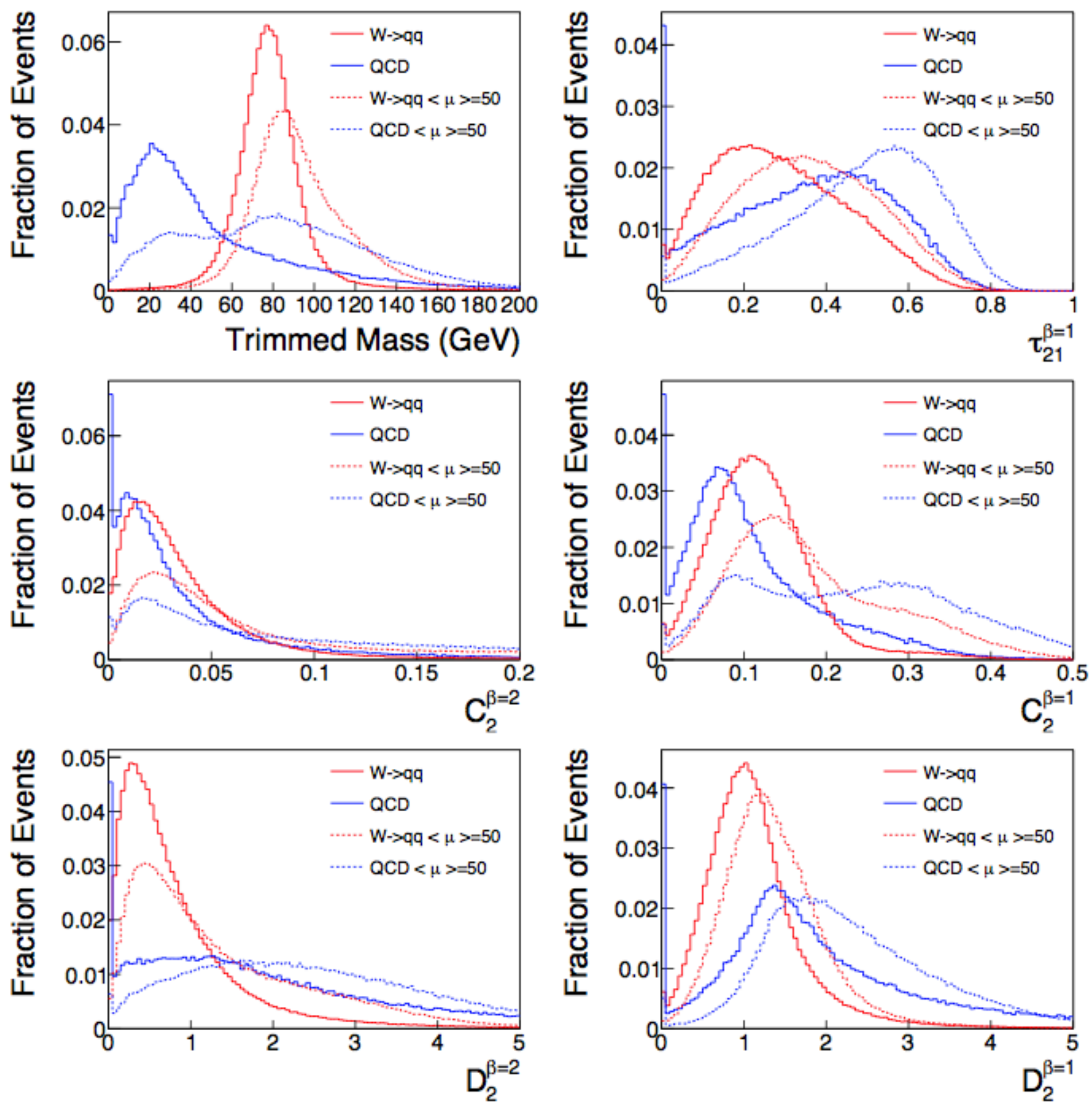


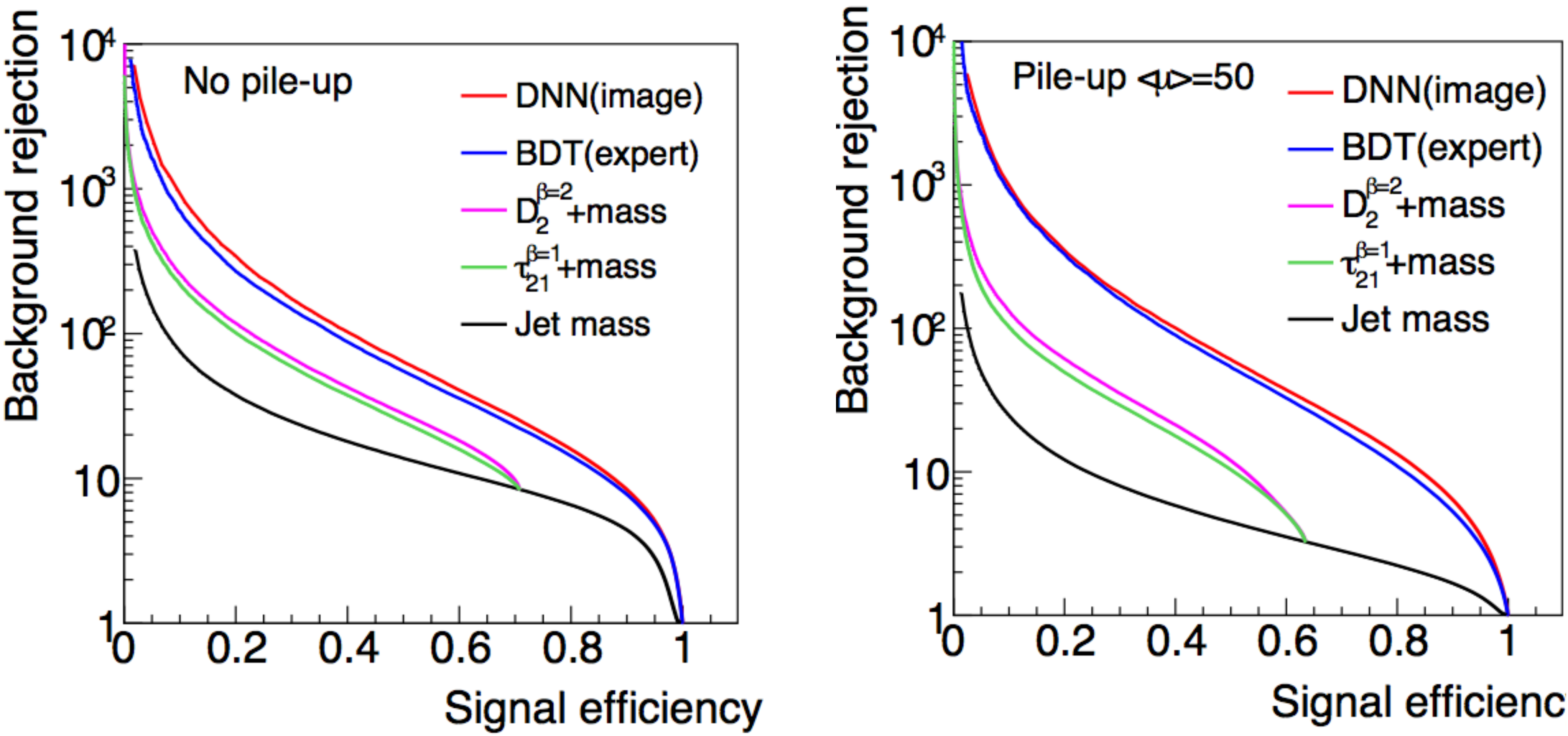
FIG. 3: Average of 100,000 jet images from class 1 (single QCD jet from q or g) on the left, and class 2 (two overlapping jets from $W \to qq'$) on the right, after preprocessing.

Jets HL



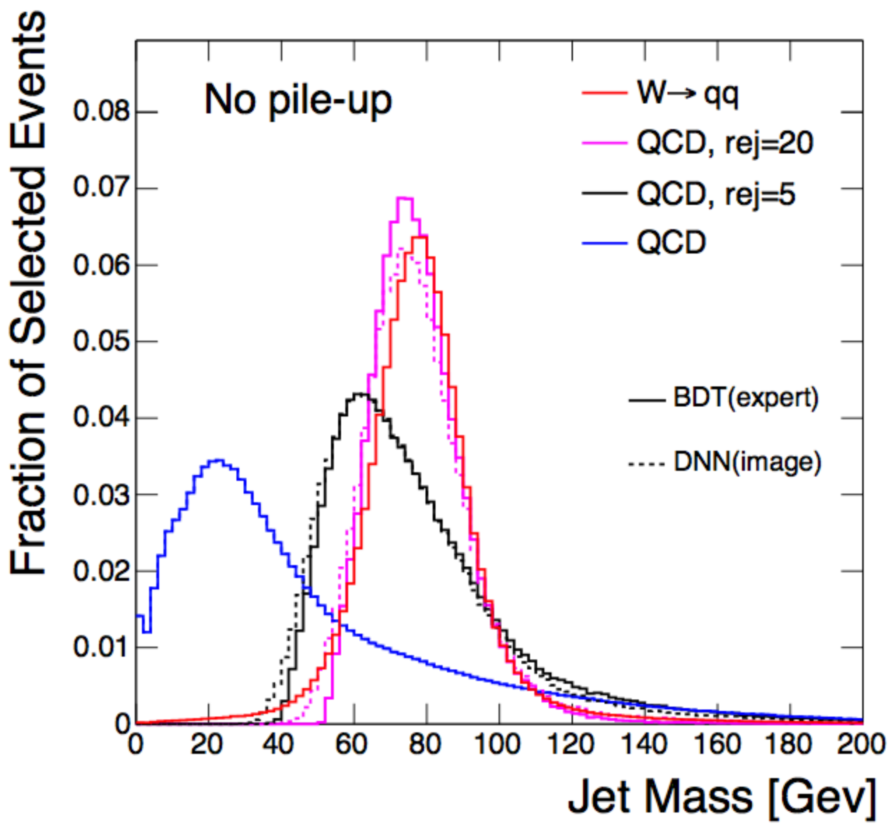
W->qq
QCD
W->qq < μ >=50
QCD < μ >=50
Fraction of Events
Trimmed Mass (GeV)
0 20 40 60 80 100 120 140 160 180 200
0 0.02 0.04 0.06
$\tau_{21}^{\beta=1}$
0 0.2 0.4 0.6 0.8 1
0 0.01 0.02 0.03 0.04
$C_2^{\beta=2}$
0 0.05 0.1 0.15 0.2
0 0.02 0.04 0.06
$C_2^{\beta=1}$
0 0.1 0.2 0.3 0.4 0.5
0 0.01 0.02 0.03 0.04
$D_2^{\beta=2}$
0 1 2 3 4 5
0 0.01 0.02 0.03 0.04 0.05
$D_2^{\beta=1}$
0 1 2 3 4 5
0 0.01 0.02 0.03 0.04

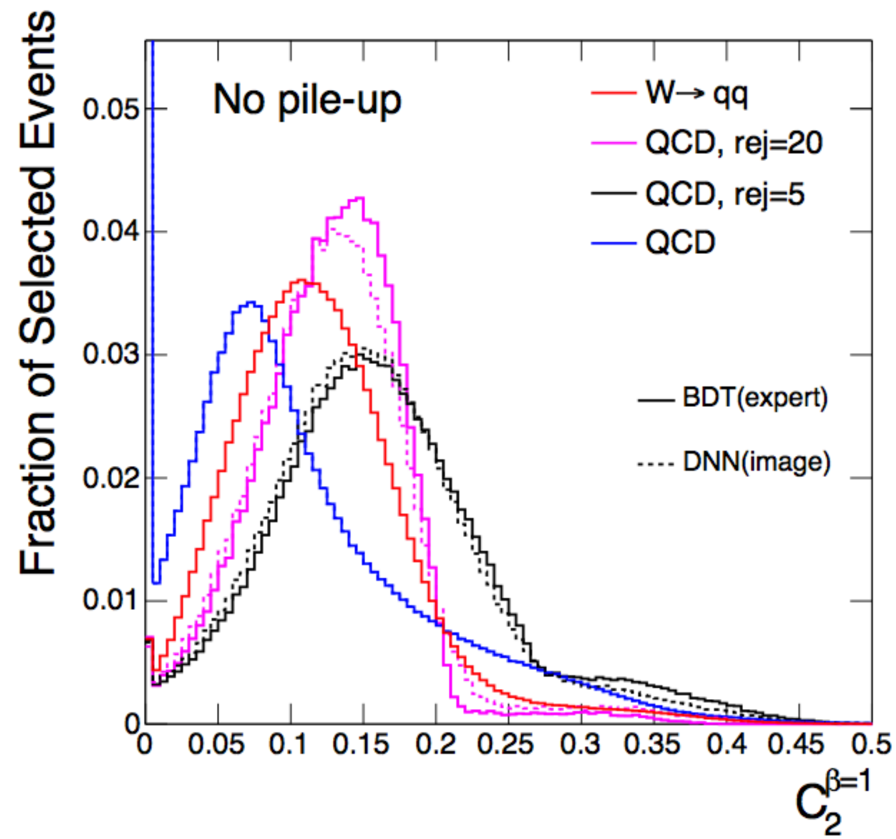
Performance



LL slightly better than HL

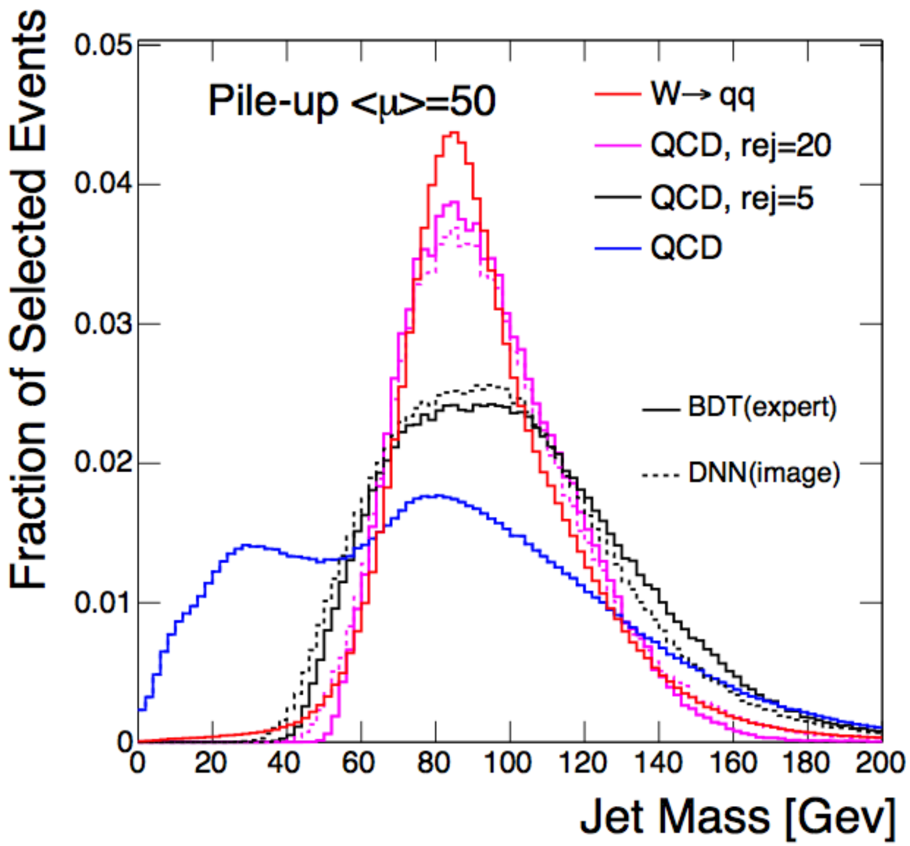
How? (no pileup)

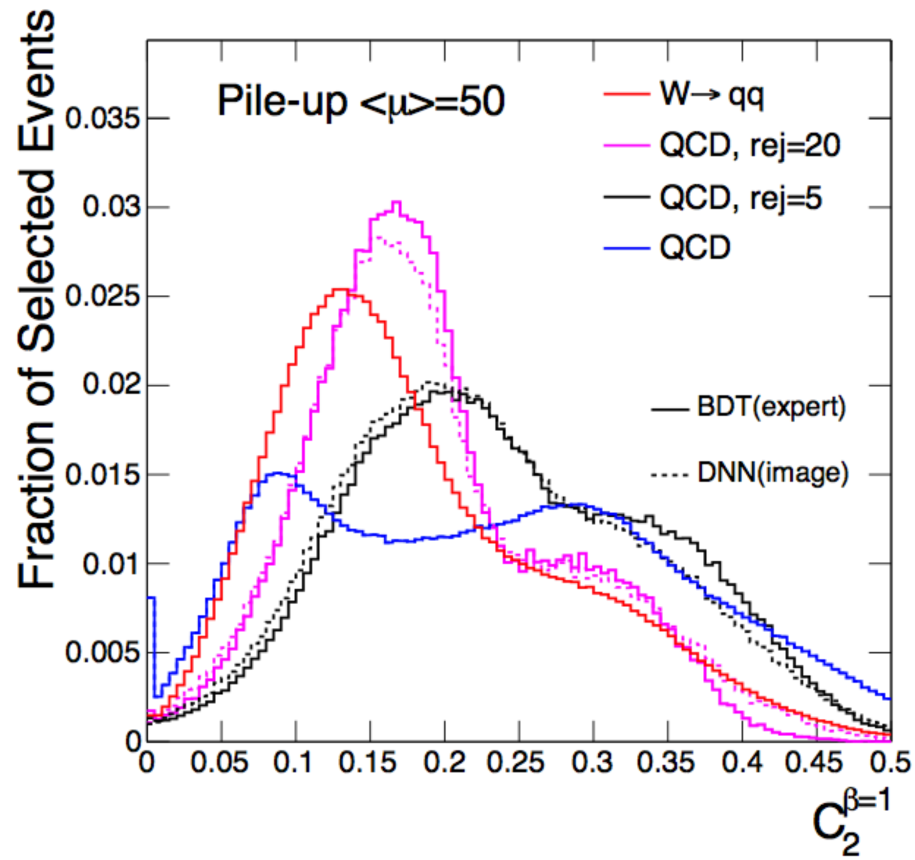




NN(LL) uses similar features to HL

How? (with pileup)





NN(LL) uses similar features to HL

Conclusions (I)

I. DL learned features on its own from LL information

II. HL features already capture most of the classification

III. DL still important for extension/application to other problems

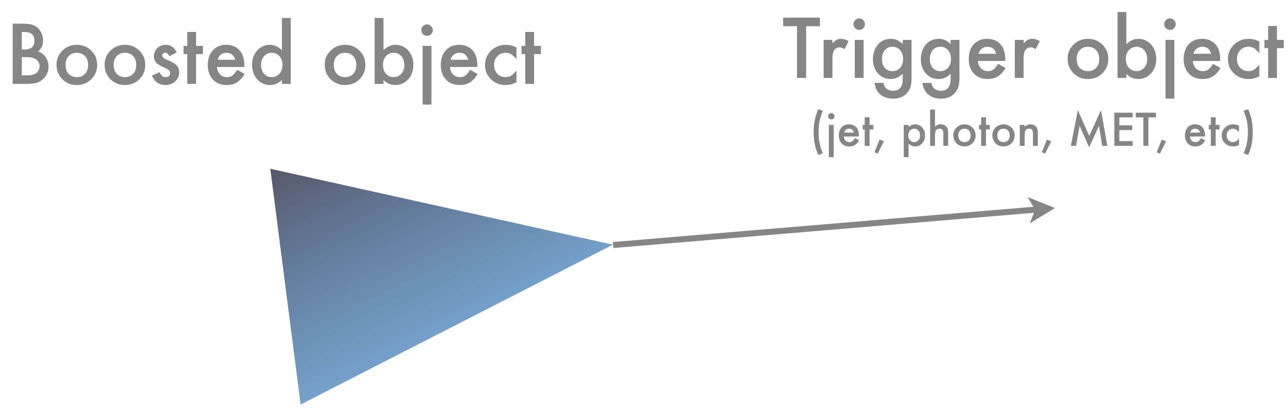
Outline

I. Jet substructure classification

II. Decorrelated jet substructure

III. Interpreting ML

The problem



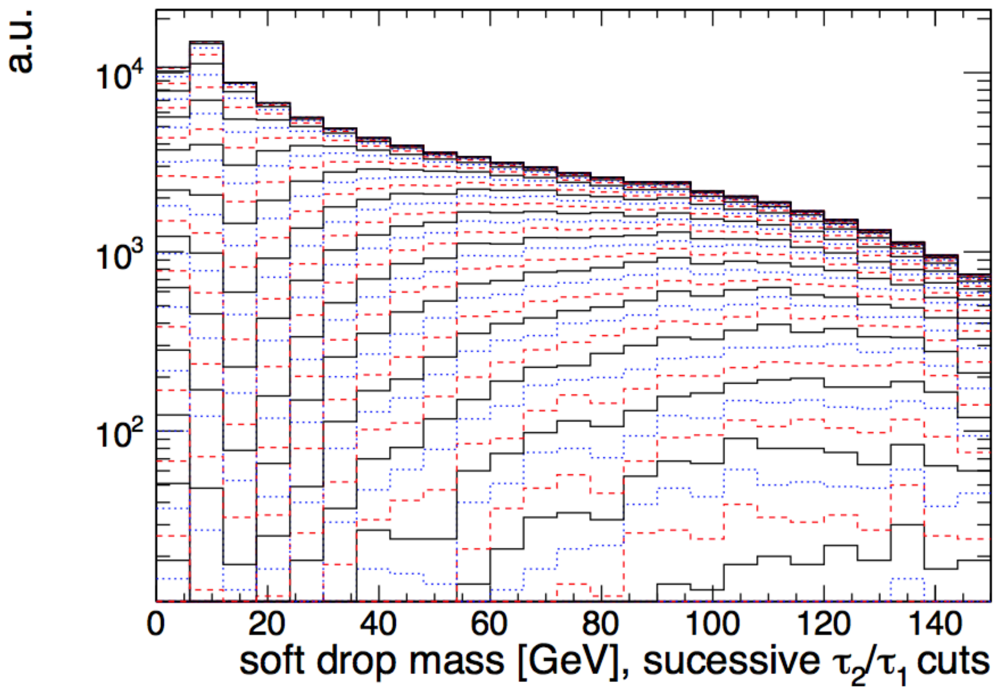
Want

Use jet substructure variables

Avoid sculpting background jet mass distribution

Smooth variation with theory mass

Background sculpting



1603.00027

Jet substructure

correlated with mass

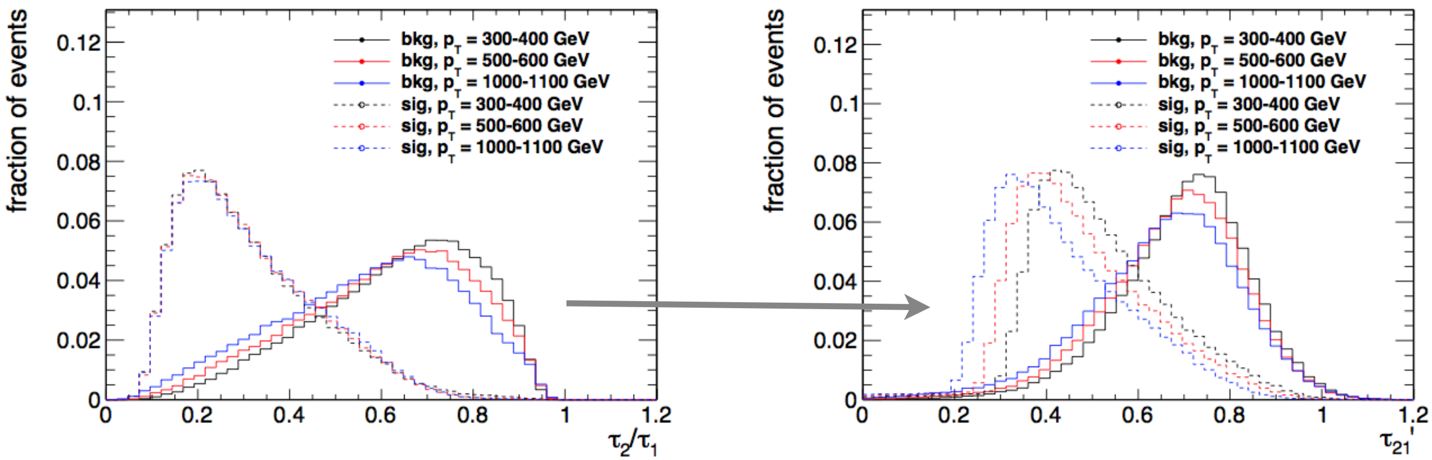
Sculpting undesirable

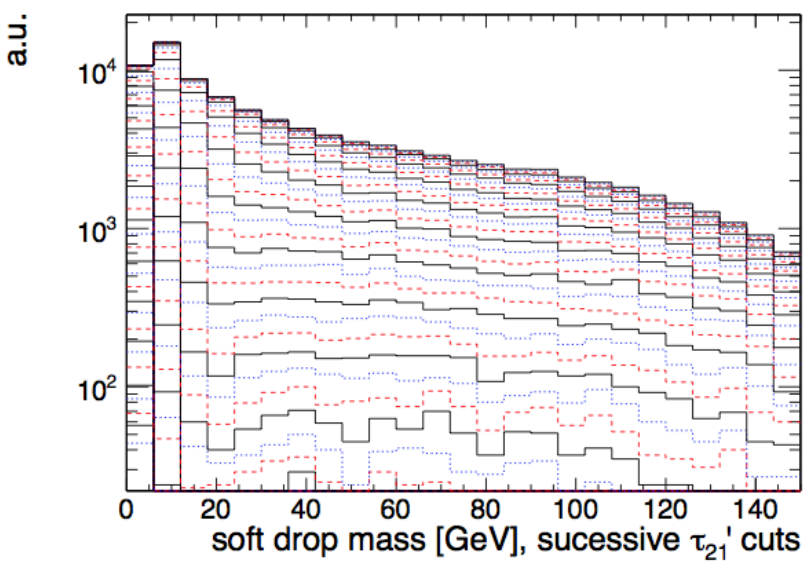
mimics signal

lost sideband constraints

lost simple behavior

One approach: DDT

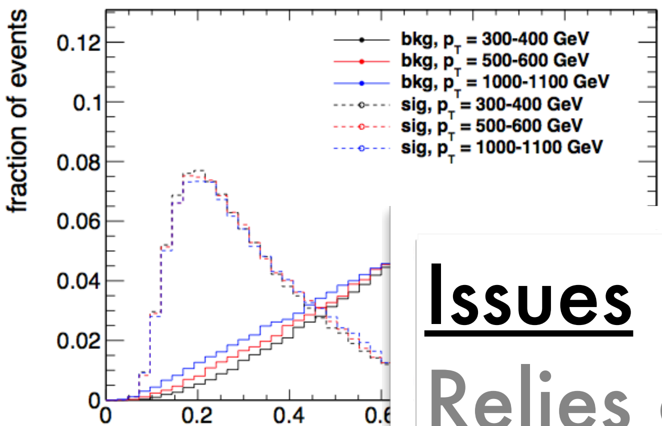


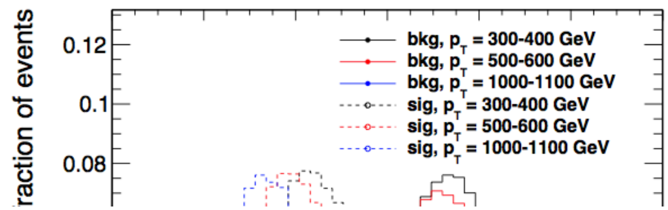


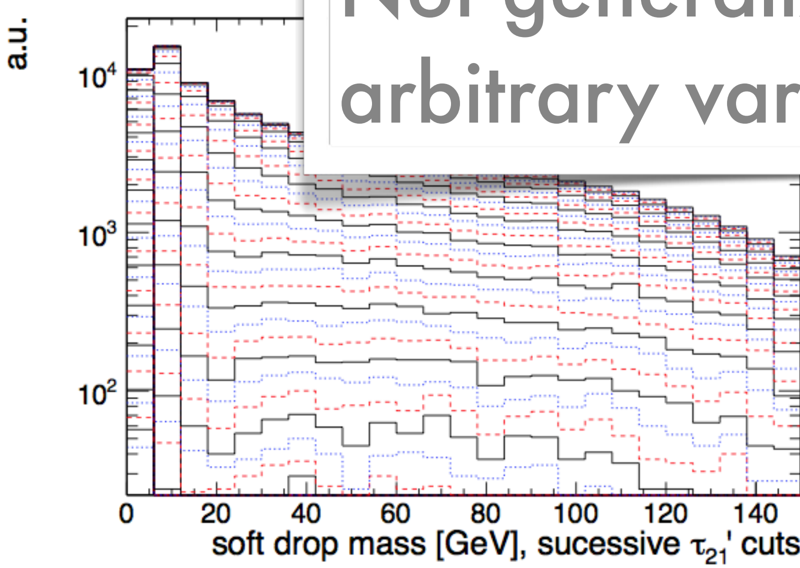
Find transformation such that variable is no longer correlated with mass

1603.00027

DDT





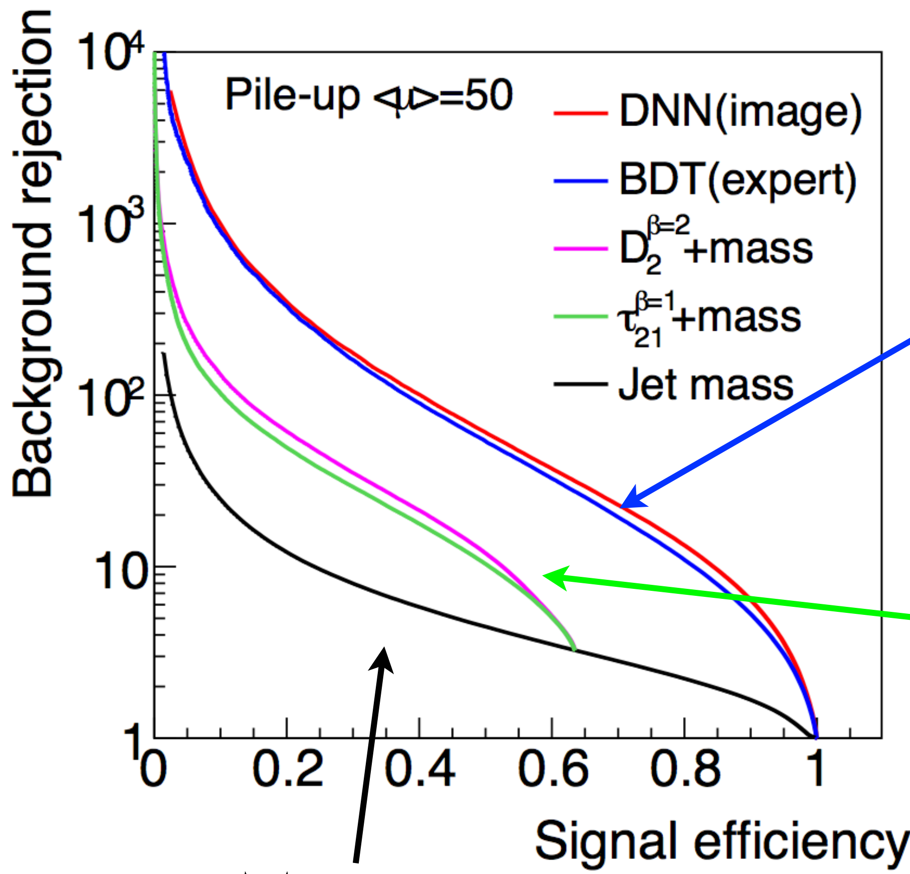


Issues

Relies on behavior of tau21.
Not generalizable to arbitrary variable.

ation
such that variable
is no longer
correlated with mass

Jet substructure vars

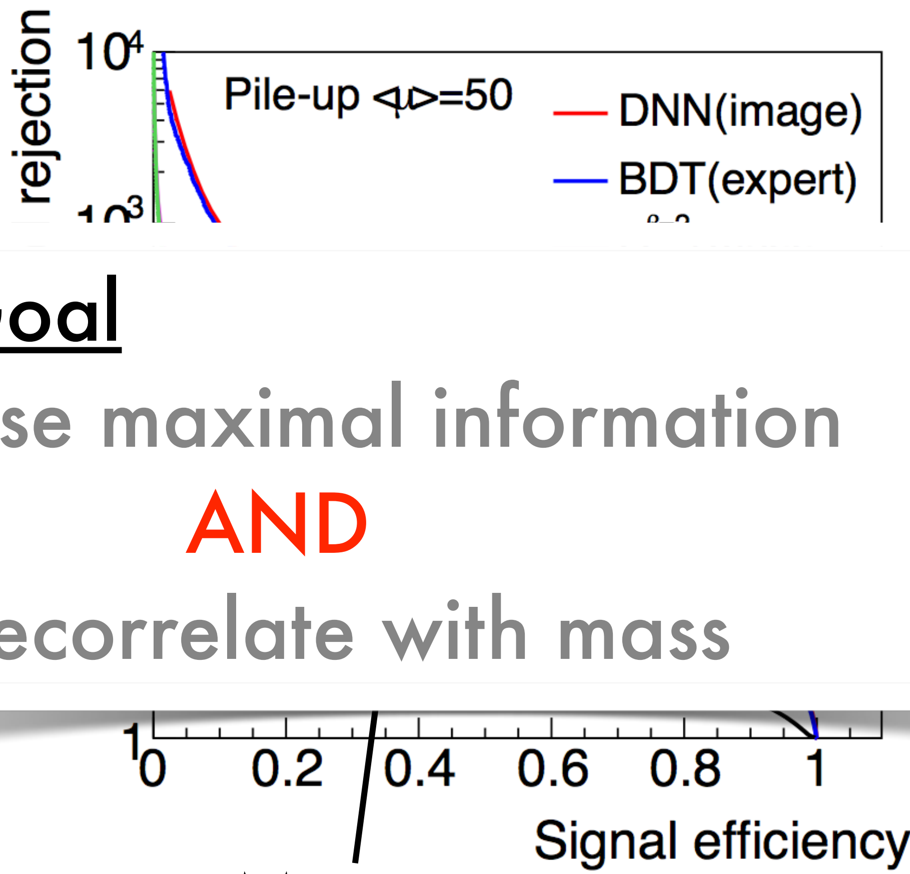


Mass + many variables
much more powerful

Mass + one variable
more powerful

Mass
powerful

Jet substructure vars



Goal

Use maximal information

AND

decorrelate with mass

Mass

powerful

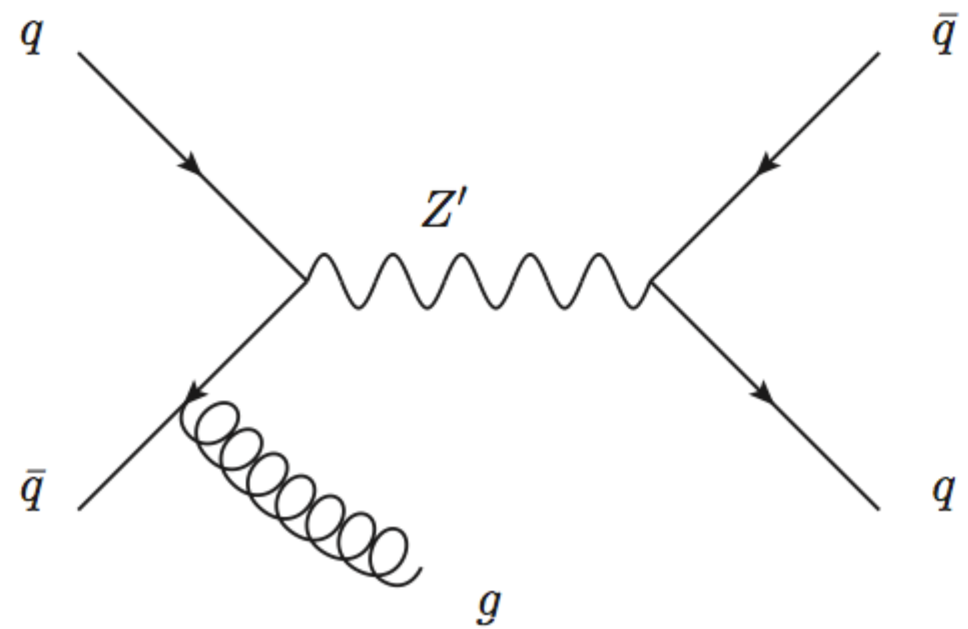
Mass + many variables

much more powerful

Mass + one variable

more powerful

Problem setup



Signal

Z′ + light jet

Background

QCD

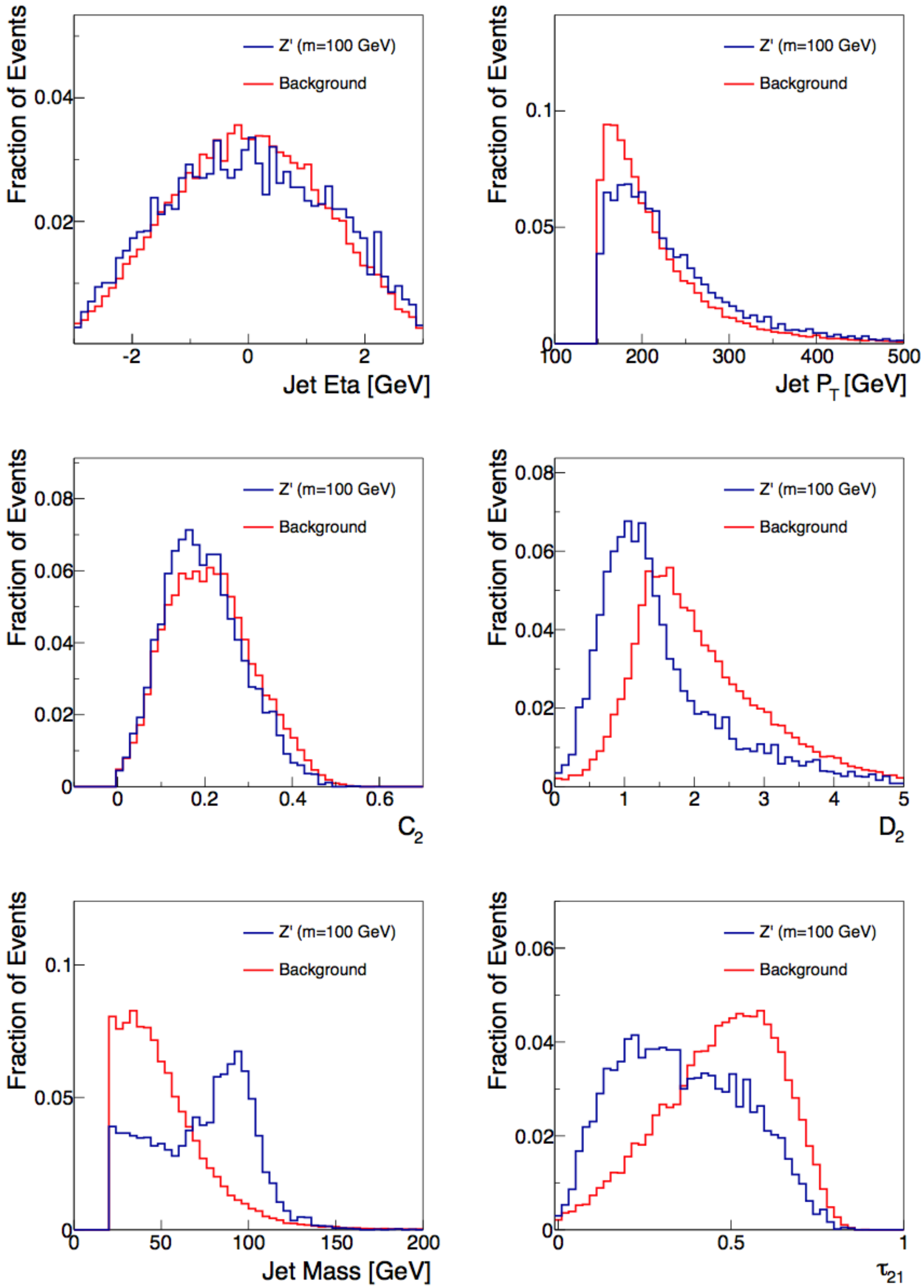
Simulation

Madgraph+pythia+delphes

Jet substructure

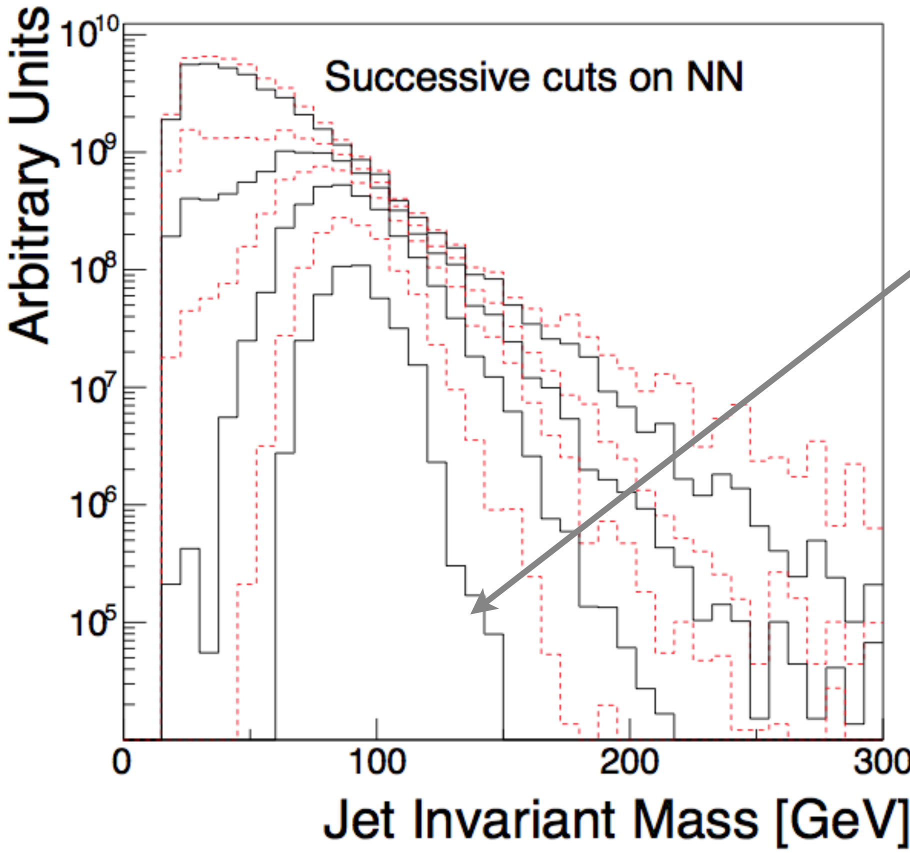
Trimmed mass, N-subj, C2, D2

Variables

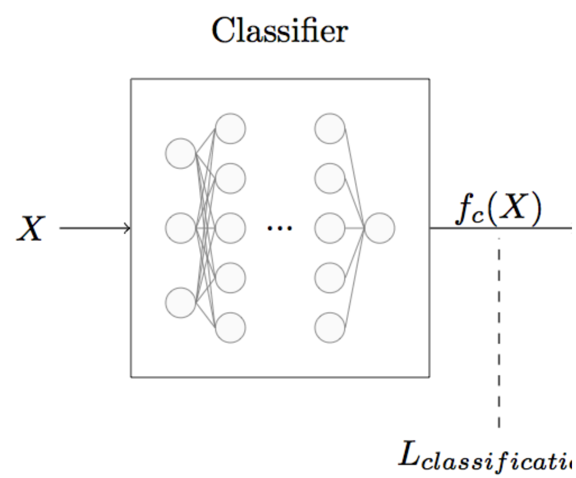

Fraction of Events
Z' (m=100 GeV)
Background
Jet Eta [GeV]
Jet P$_T$ [GeV]
C_2
D_2
Jet Mass [GeV]
τ_{21}

Mass sculpting

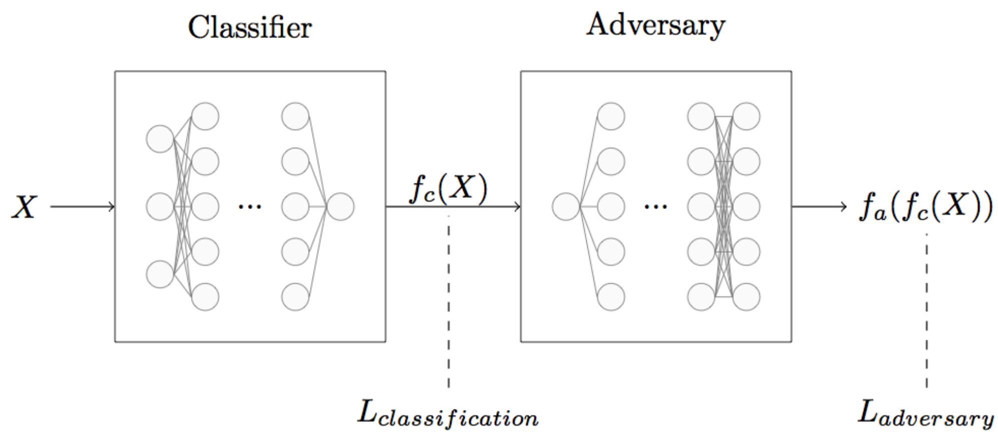
Generic Classifier



NN sculpts background
to look like signal



Adversarial NN



<u>Optimize</u>:

classification accuracy

- adversary accuracy

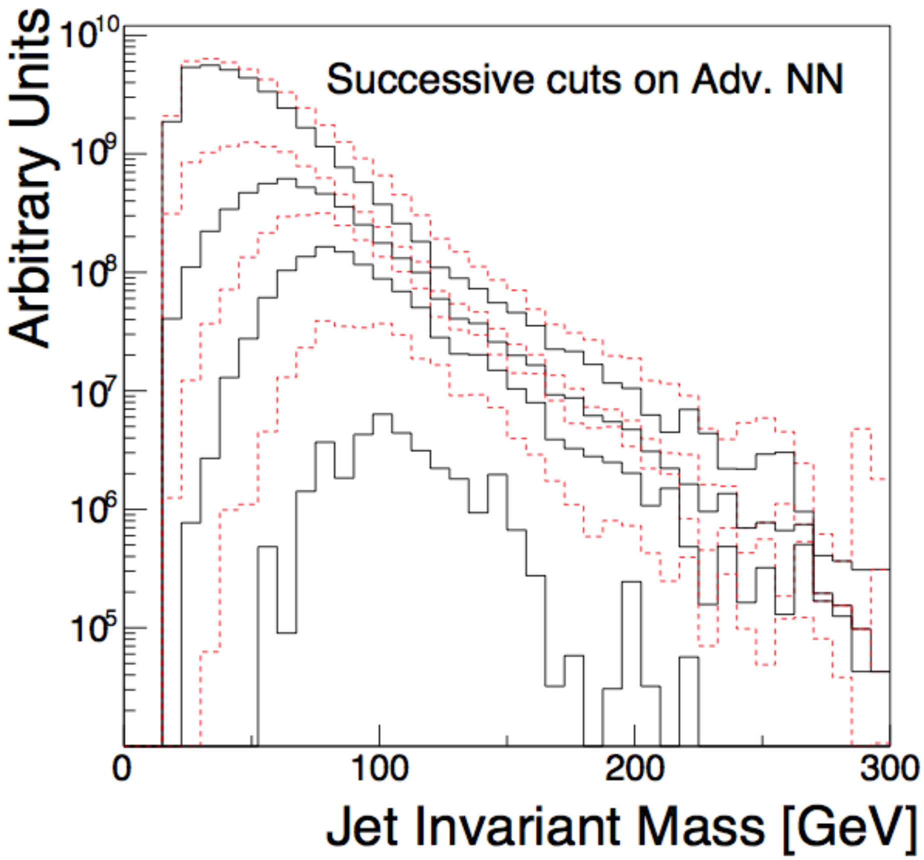
How well does classifier discriminate?

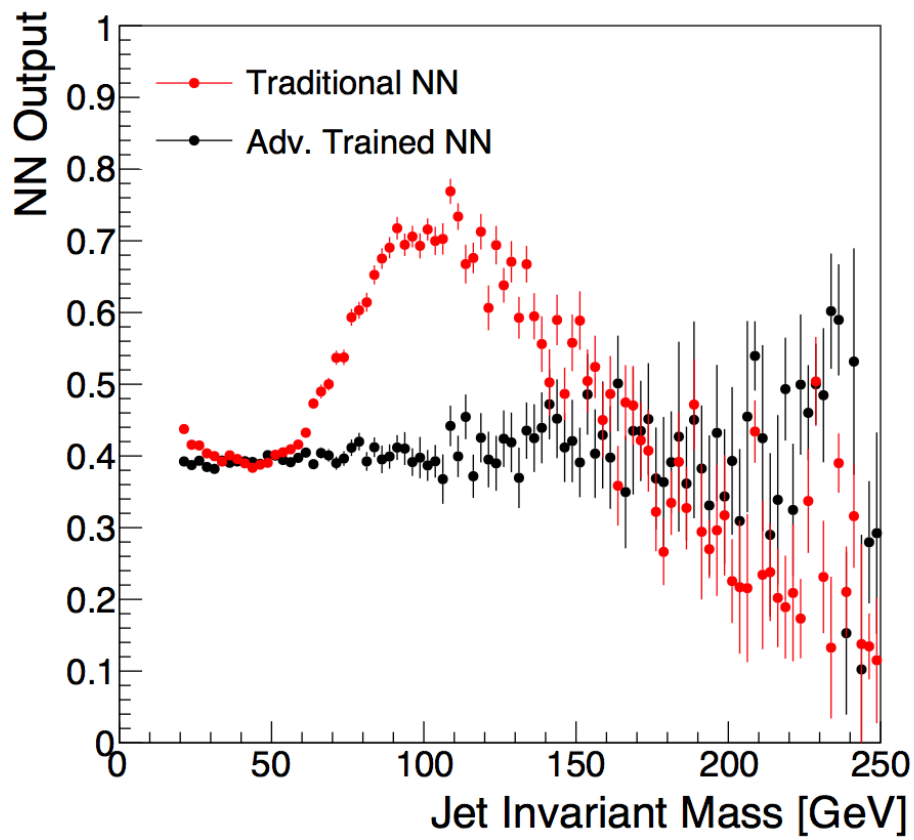
Can adversary guess the jet mass from the classifier output?

$$L_{\text{system}} = L_{\text{classification}} - \lambda L_{\text{adversary}} \ .$$

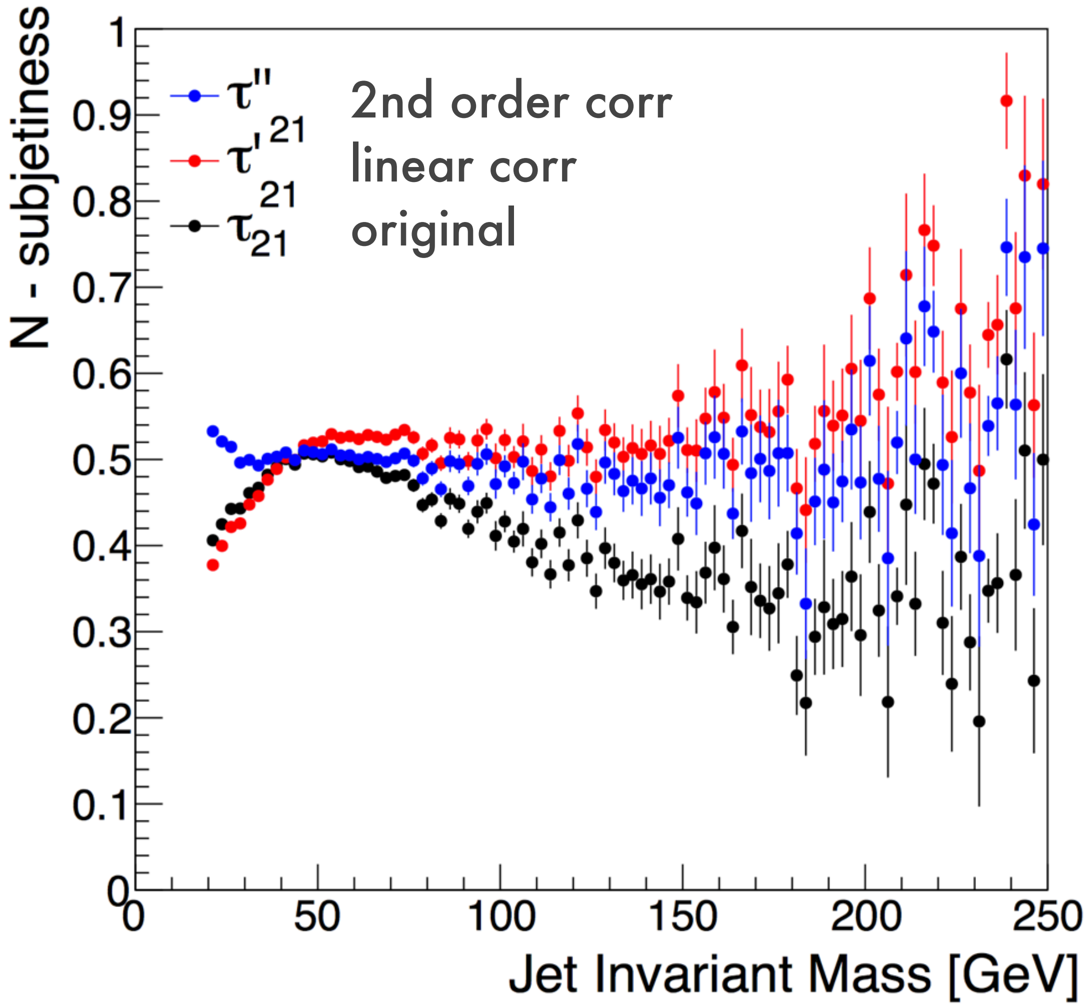
Adv. NN

Classifier is less dependent on jet mass





DDT

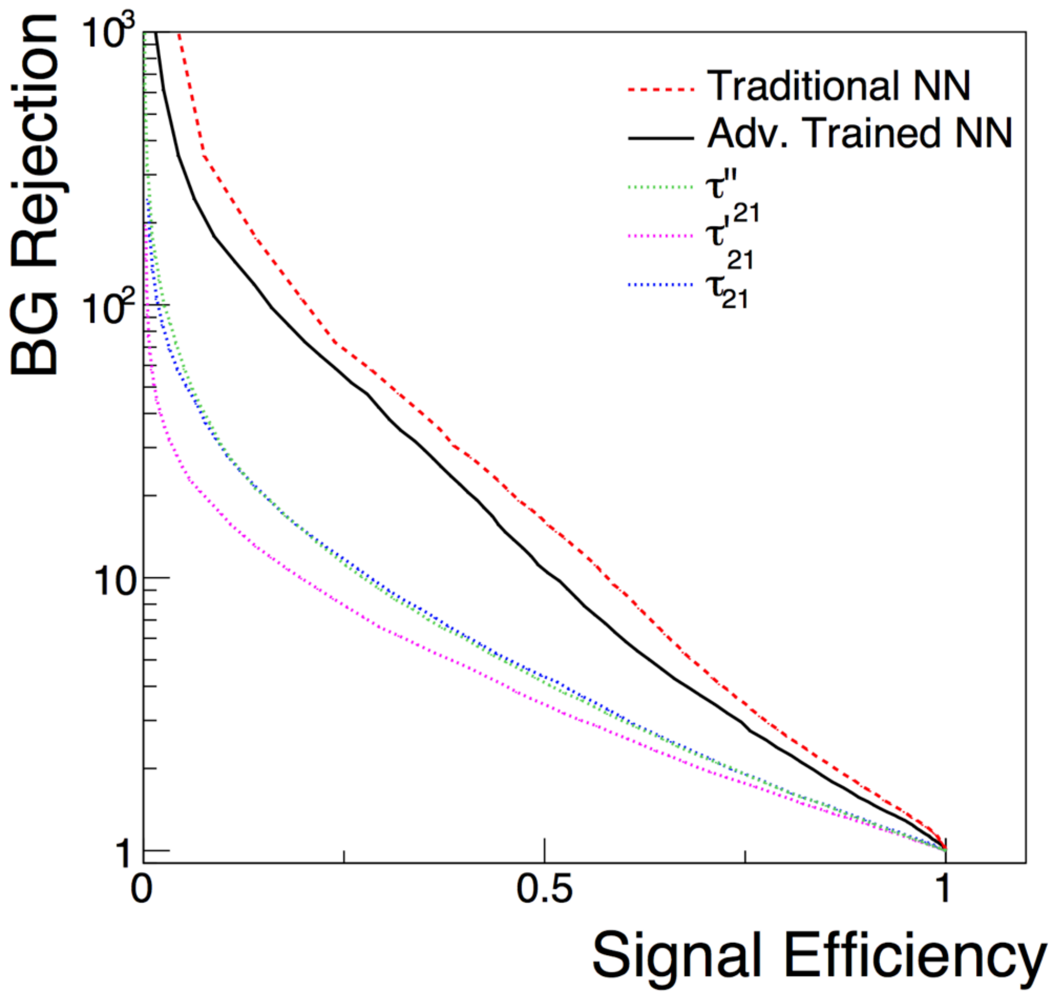


DDT

Don't see linear behavior with rho

Try similar method to reduce mass dependence for a single variable

But at what cost?



Traditional classifier is better at S/B discrimination.

That's what it is optimized for!

Real Goal

More to the point

Fit jet mass histogram

Measure discovery significance

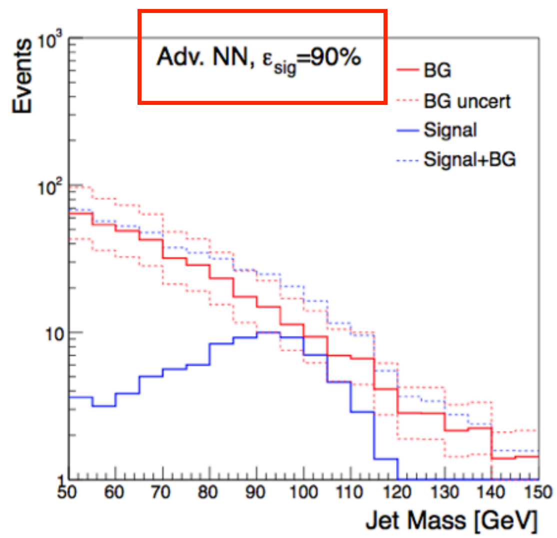
N_{bg} = 1000

N_{sig} = 100

Bg rate uncertainty

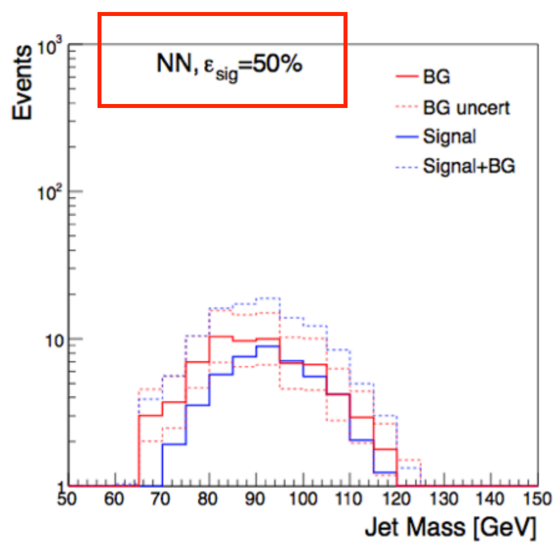
5% or 50%

(shape uncertainty very dependent on specific unc)



No bg sculpting

So sidebands can constrain bg rate



Bg sculpting

No sidebands to constrain bg rate

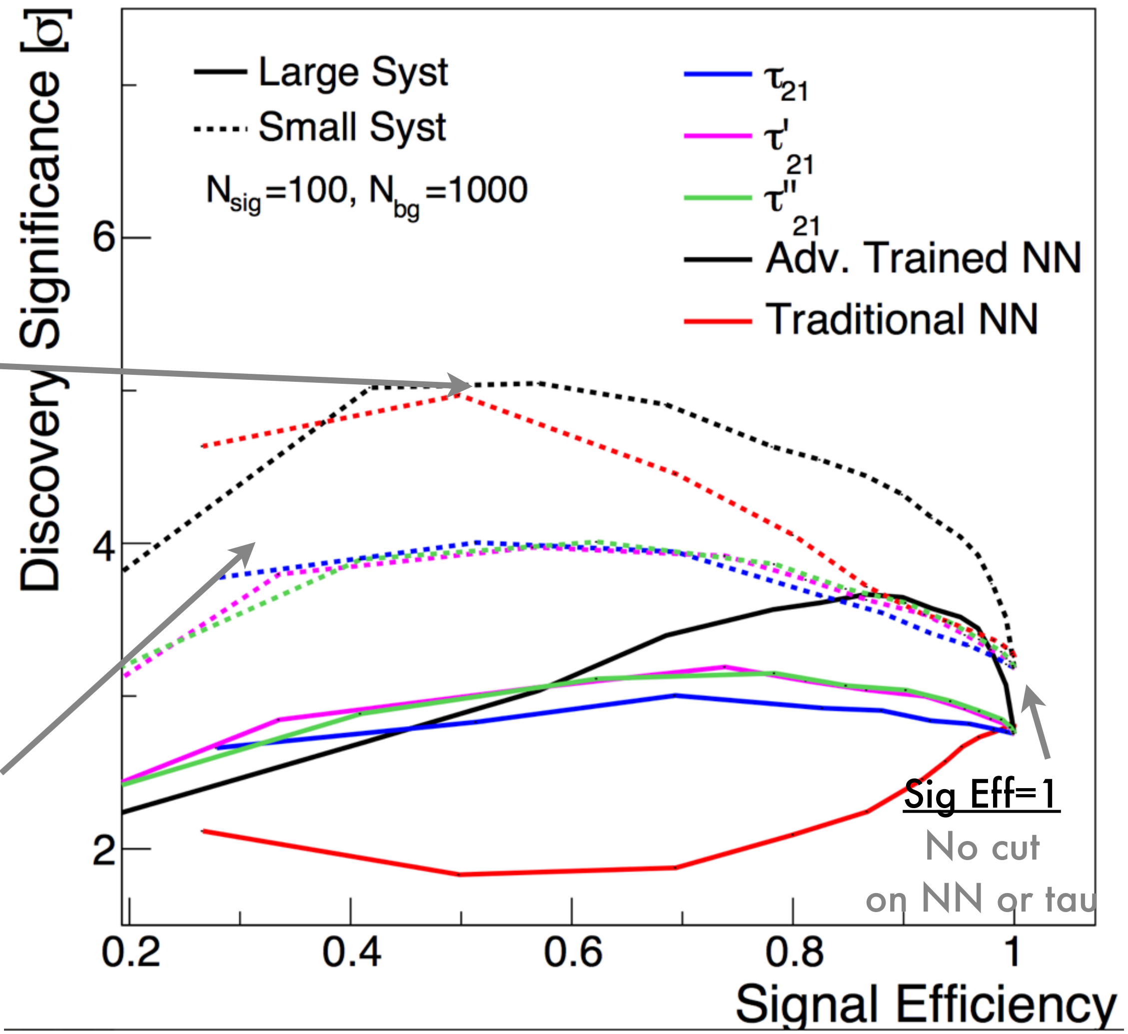
Real Goal

<u>**Small systematics**</u>

Background shape and rate ~known

Classifier improves significance

Classifier has higher max significance



Low signal eff

Strict cut on NN or tau

Sig Eff=1

No cut on NN or tau

Real Goal

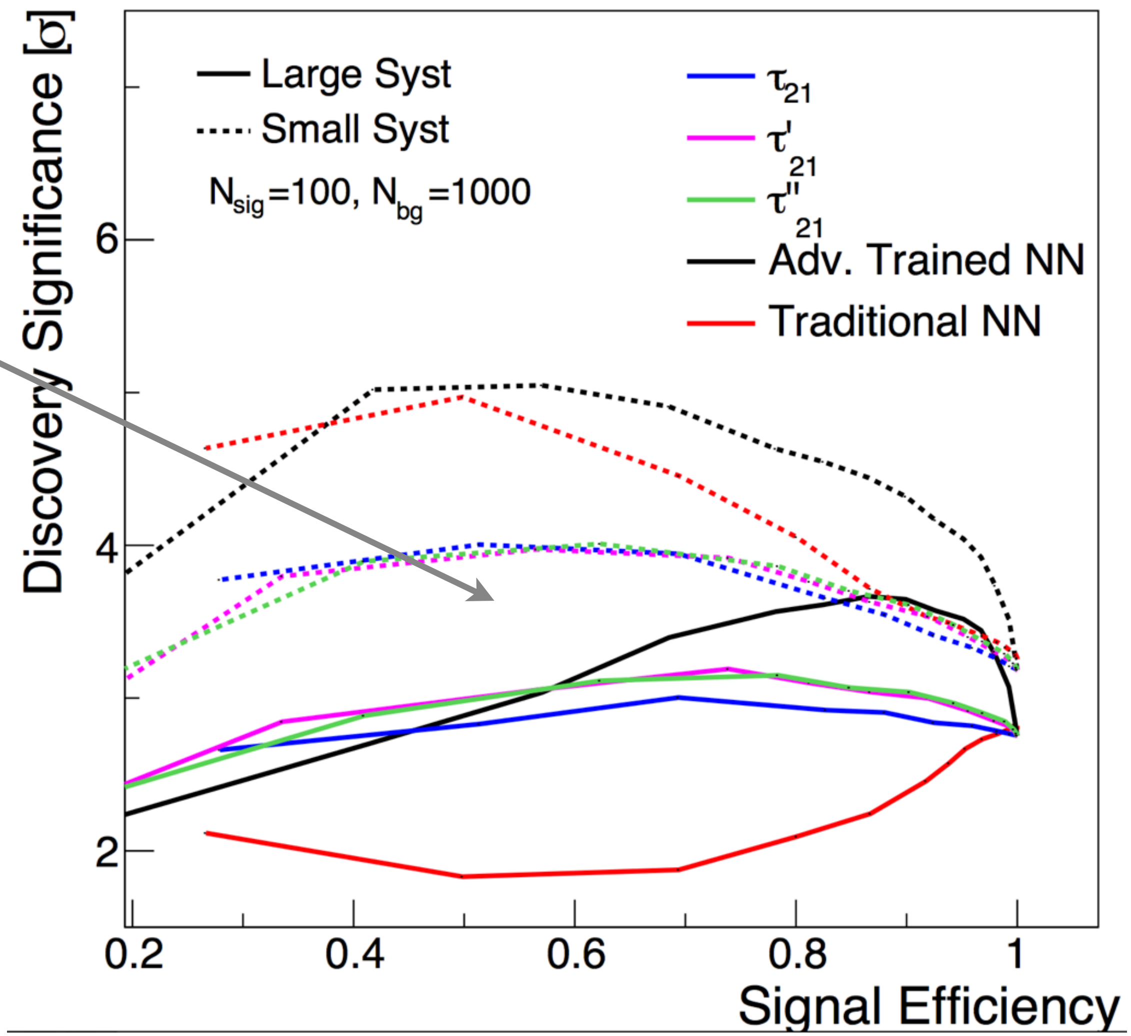
<u>**Large systematics**</u>

Background shape known, rate uncert.

Classifier sculpts bg like signal
- S and B are ~identical
- no sidebands to constrain B
- cutting on classifier worsens signif.

Adversarial network
- maintains bg shape
- keeps sidebands to constrain

Adv NN has higher max significance.

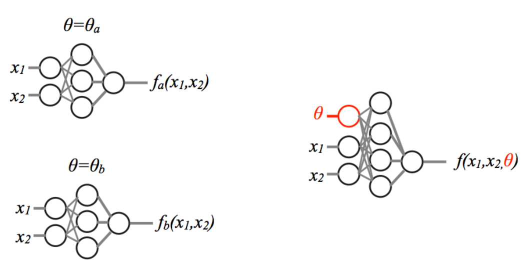


Parameterized NN

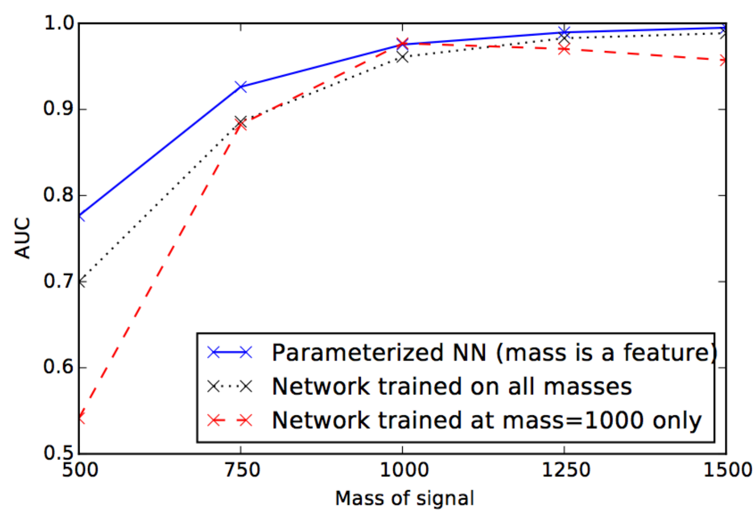
Use a NN parameterized with theory mass

Allows optimal combination of jet vars
(not just one variable)

Smoothly varies with theory mass
(allows interpolation)

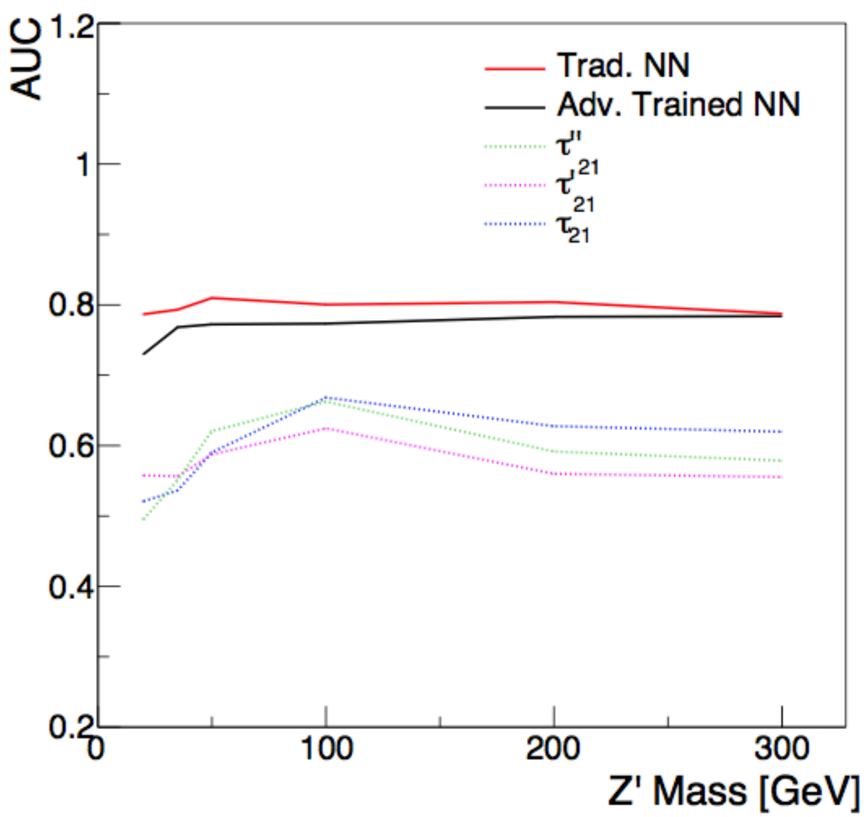


1601.07913

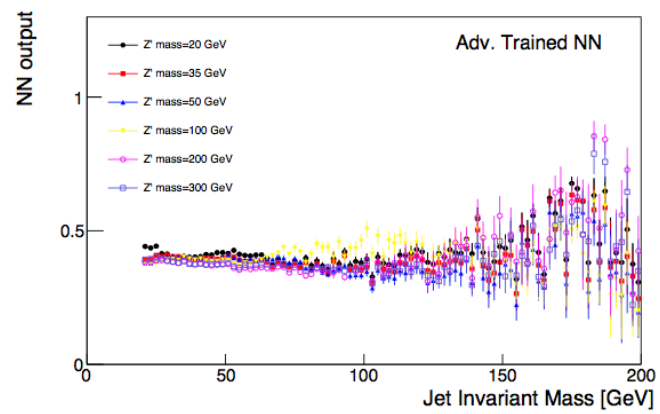


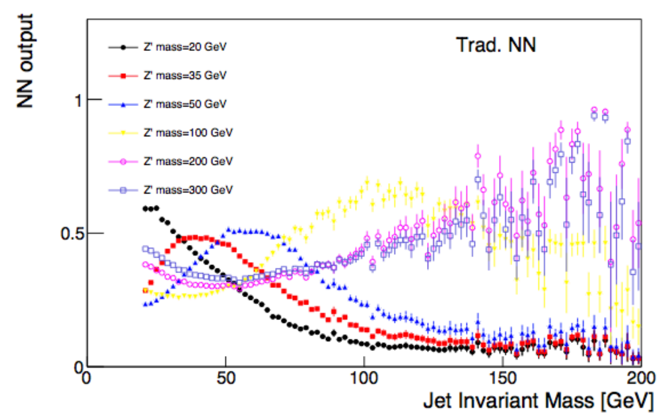
Results

At any point trad NN has better classification (but not our goal)



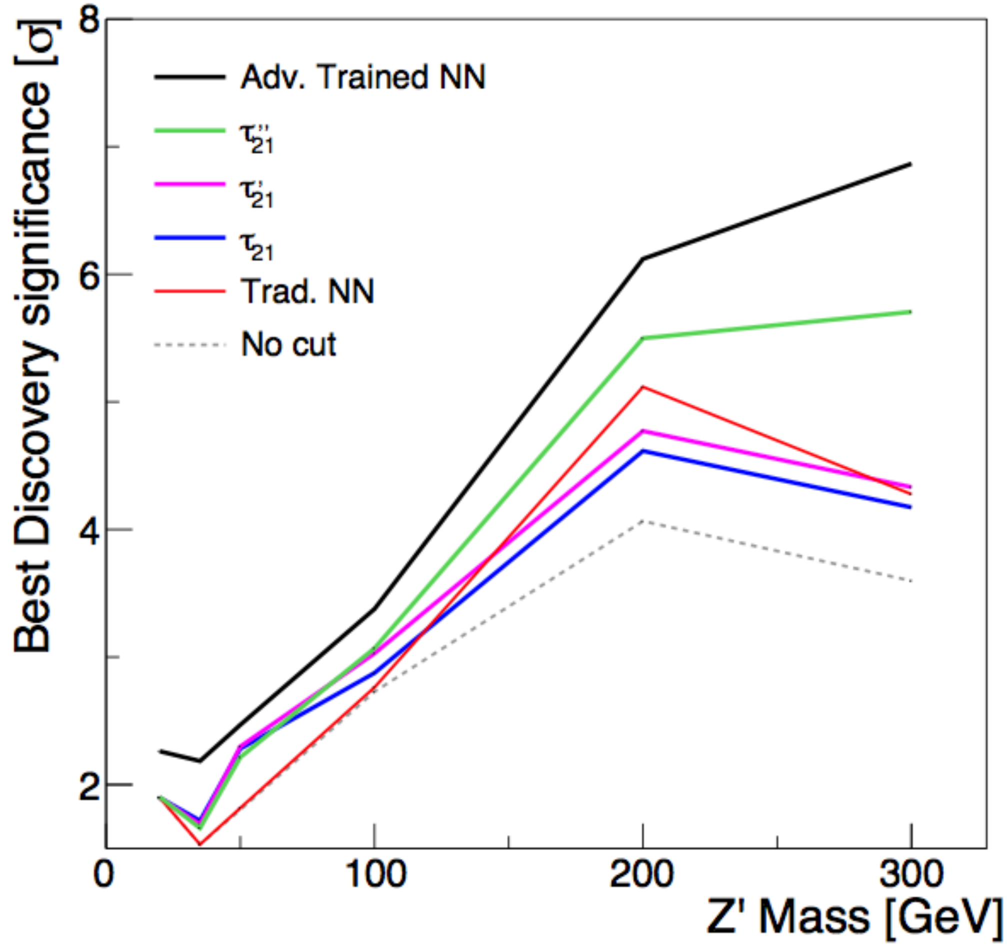
Param adv network remains decorrelated





Results

Adv. param gives best performance over entire space



Conclusions (II)

<u>Adversarial parameterized NN</u>

Constrains NN to maintain bg shape

More robust against systematics

Smoothly interpolatable

Better use of all jet substructure variables

Outline

I. Jet substructure classification

II. Decorrelated jet substructure

III. Interpreting ML

We prefer HL

If HL data includes all necessary information...

- It is easier to understand
- Its modeling can be verified
- Uncertainties can be sensibly defined
- It is more compact and efficient

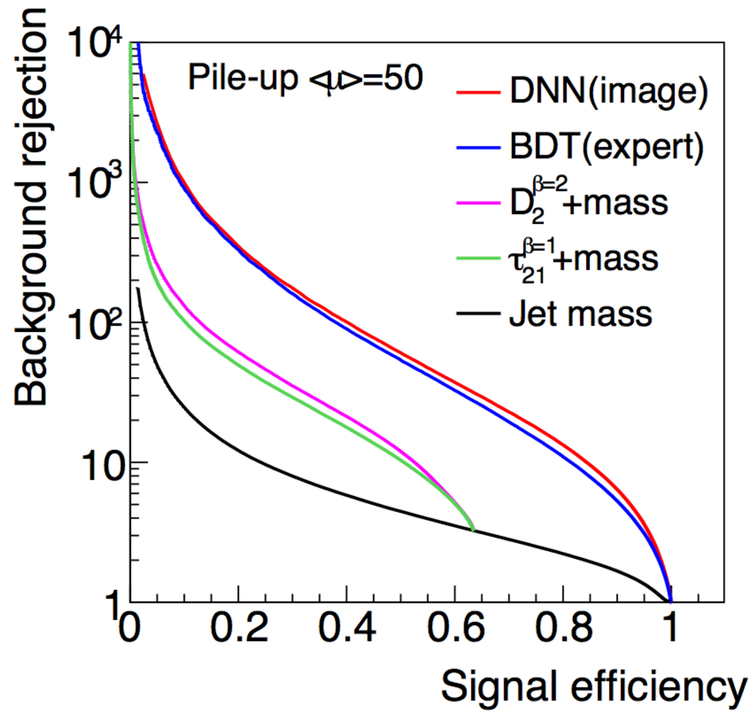
- LL -> HL is physics, so we like it.

Our question

How has the DNN found its solution?
What can we learn from it?

<u>Residual knowledge:</u>
Is there a **new** HL variable?
Can it reveal physics?

<u>Translating complete solutions:</u>
What is the **structure** of its solution?
Has it just rediscovered and
optimized the existing HL vars?



How?

<u>I. Define space of possible human solutions</u>

- provides context for NN solution
- defines problem
- does NN live in this space?
- Can it be compactly represented?
- Yes or No are both interesting!

<u>II. Define mapping metric</u>

- how do you compare two solutions?
- can’t use functional identity or linear correlation

How?

I. Define space of possible human solutions

- provides context for NN soluti
- defines problem
- does NN live in t
- Can it b
- Ye ng!

Results soon!

II. Defi pping metric

- how do you compare two solutions?
- can't use functional identity or linear correlation

Conclusions

<u>Jet substructure is theoretically mature</u>

- existing HL functions work well

<u>DL can rediscover existing ideas</u>

- generalize them
- decorrelate them

<u>DL might be able to extract new insights</u>

- mapping back to human ideas