Systematically Searching for New Physics at the LHC

Daniel Whiteson
UC Irvine

Nature Guiding Theory, FNAL Aug 2014

Nature Guiding Theory

Two approaches:

Nature Guiding Theory

Two approaches:

(1) "Nature" = simplicity, aesthetics

$\Rightarrow$ need more TeV-Scale ideas for solution to outstanding theory issues

Nature Guiding Theory

<u>Two approaches:</u>

~~(1) "Nature" = simplicity, aesthetics~~

~~⇒ need more TeV-Scale ideas for solution to outstanding theory issues~~

(2) "Nature" = experiments, reality

⇒ need to find a BSM clue in LHC data

Outline

I. Strategy for unanticipated new physics

II. Deep networks for NP searches

Searching for new physics

Specific

Model

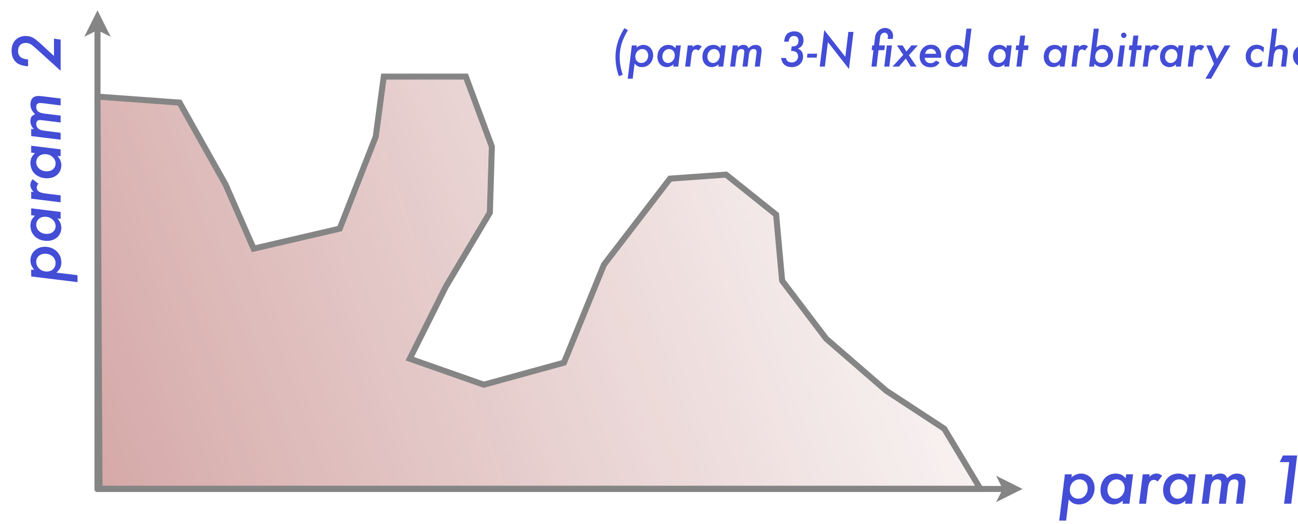
Search strategy

General

Traditional approach

Bet on a specific theory

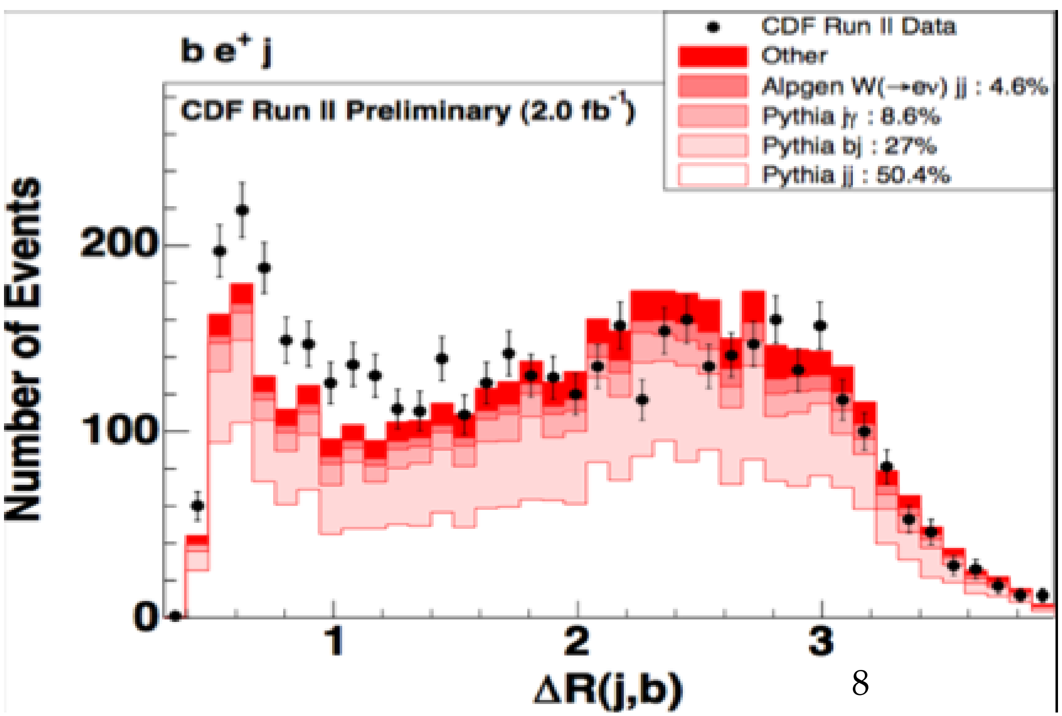
Optimize analysis to squeeze out maximal sensitivity to new physics.



Model independent search

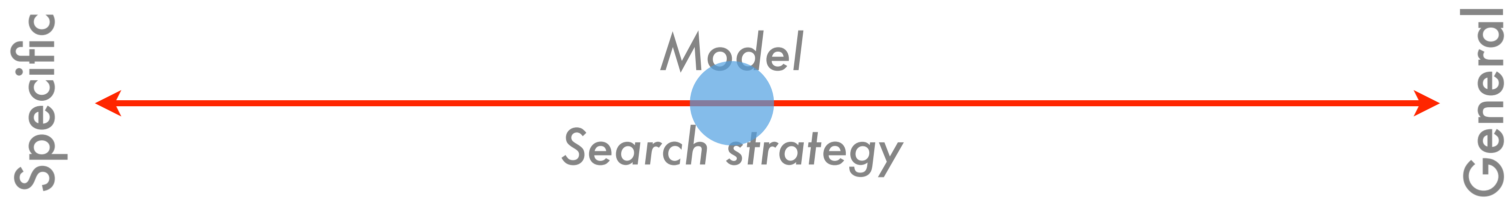
Discard the model

compare data to standard model



"Never listen to theorists.
Just go look for it."
–A. Pierce, 2010

Compromise



Admit the need for a model

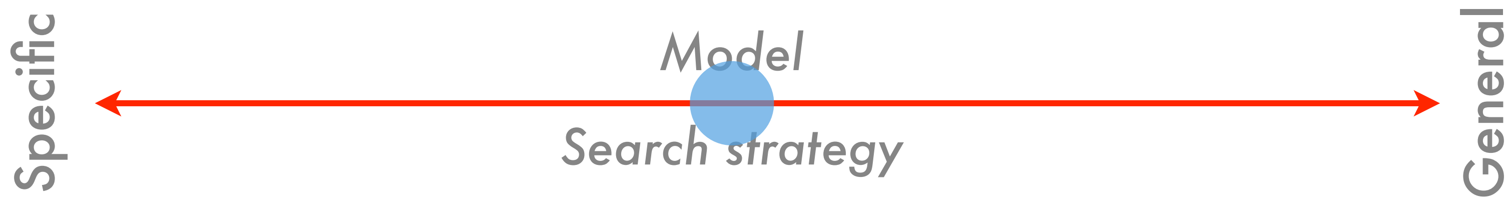
New signal requires a coherent physical explanation, even trivial or effective

Generalize your model

Construct simple models that describe classes of new physics which can be discovered at the LHC.

What are we good at discovering?

Compromise



Admit the need for a model

New signal requires a coherent physical explanation,
even trivial or effective

Generalize your model

Construct simple models that describe classes of new physics which can be discovered at the LHC.

What are we good at discovering? Resonances!

Is this being done?

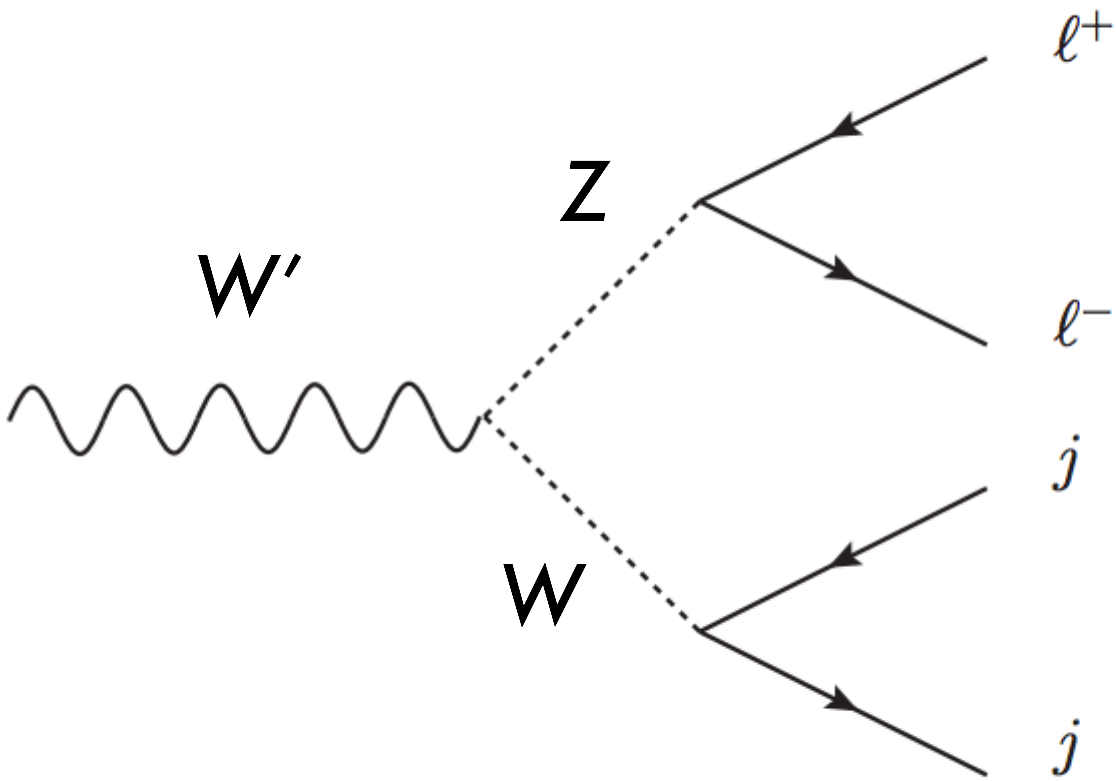
ℓ^+

ℓ^-

j

j

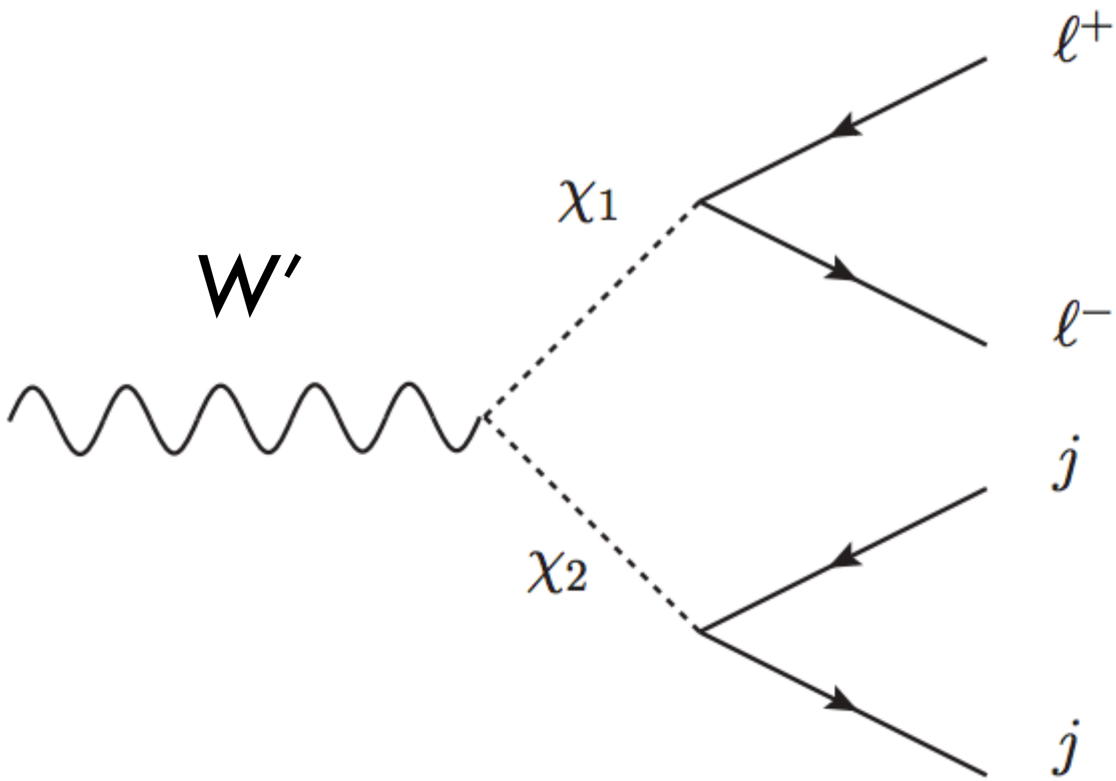
Is this being done?



$m_{ll} = m_Z$

$m_{jj} = m_W$

What about this?



m_{ll} != m_Z

m_{jj} != m_W

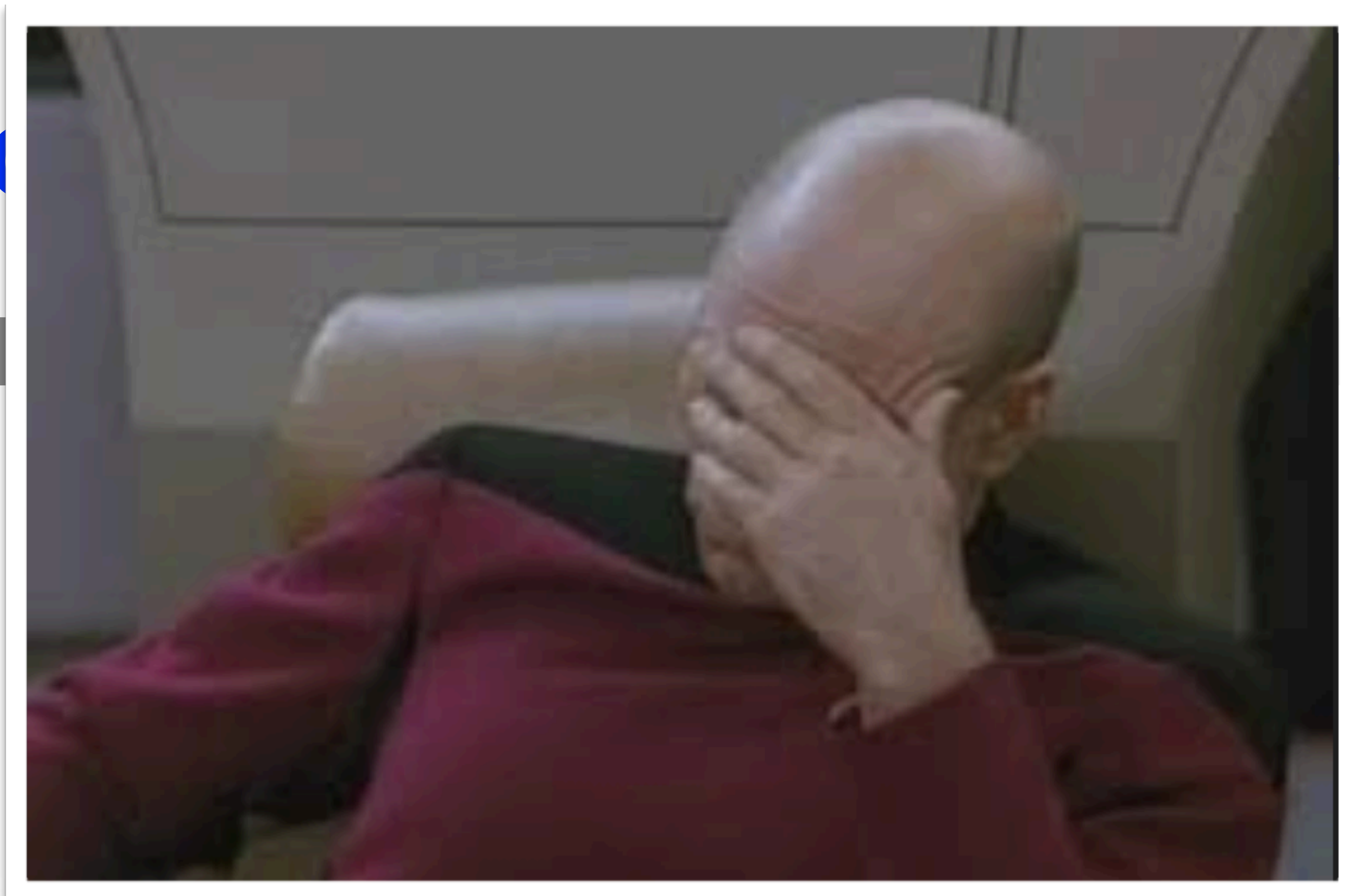
Missed resonances?

Easy-to-find resonances

may exist in our data and

nobody has looked!

Missed resonances?



Topological models

UC Irvine Physics 247
Final project
arXiv: 1401.1462, PRD

FERMILAB-PUB-13-529-T

Systematically Searching for New Resonances at the Energy Frontier using Topological Models

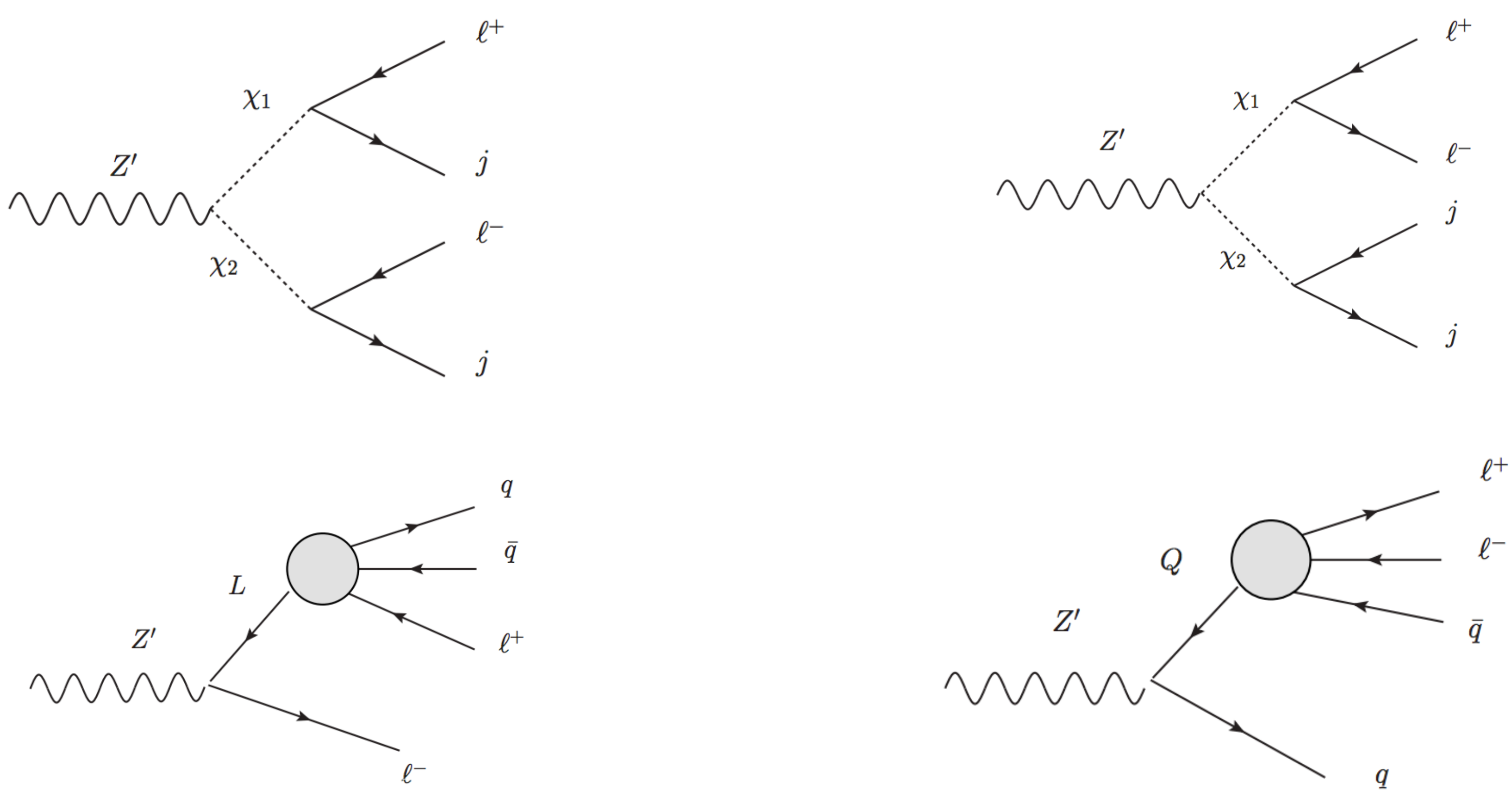
Mohammad Abdullah,[1] Eric Albin,[1] Anthony DiFranzo,[1] Meghan Frate,[1] Craig Pitcher,[1] Chase Shimmin,[1] Suneet Upadhyay,[1] James Walker,[1] Pierce Weatherly,[1] Patrick J. Fox,[2] and Daniel Whiteson[1]

[1]*Department of Physics and Astronomy, University of California, Irvine, CA 92697*
[2]*Fermi National Accelerator Laboratory, Batavia, IL 60615*

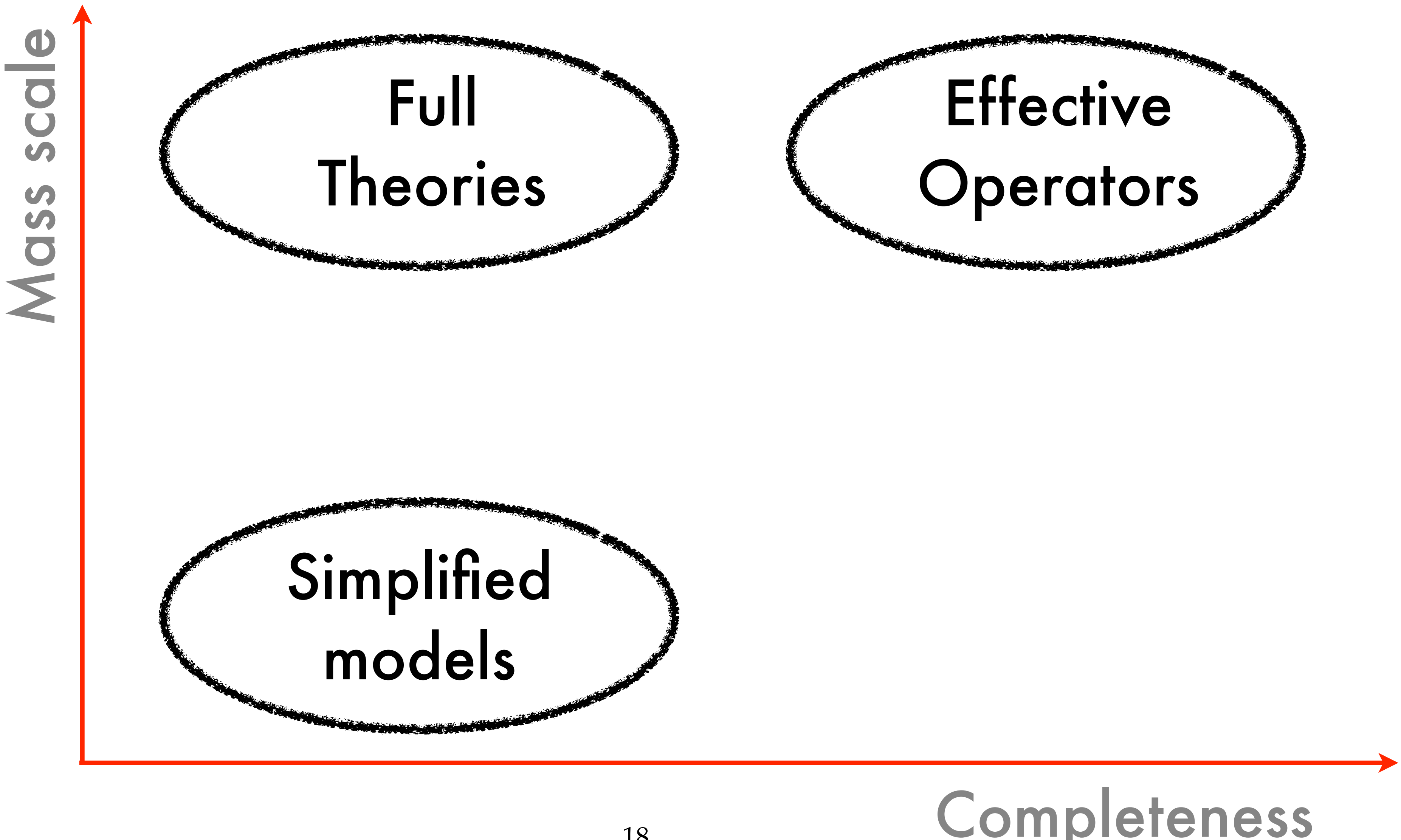
We propose a new strategy to systematically search for new physics processes in particle collisions at the energy frontier. An examination of all possible topologies which give identifiable resonant features in a specific final state leads to a tractable number of 'topological models' per final state and gives specific guidance for their discovery. Using one specific final state, $\ell\ell jj$, as an example, we find that the number of possibilities is reasonable and reveals simple, but as-yet-unexplored, topologies which contain significant discovery potential. We propose analysis techniques and estimate the sensitivity for pp collisions with $\sqrt{s} = 14$ TeV and $\mathcal{L} = 300$ fb^{-1}.

Topological models

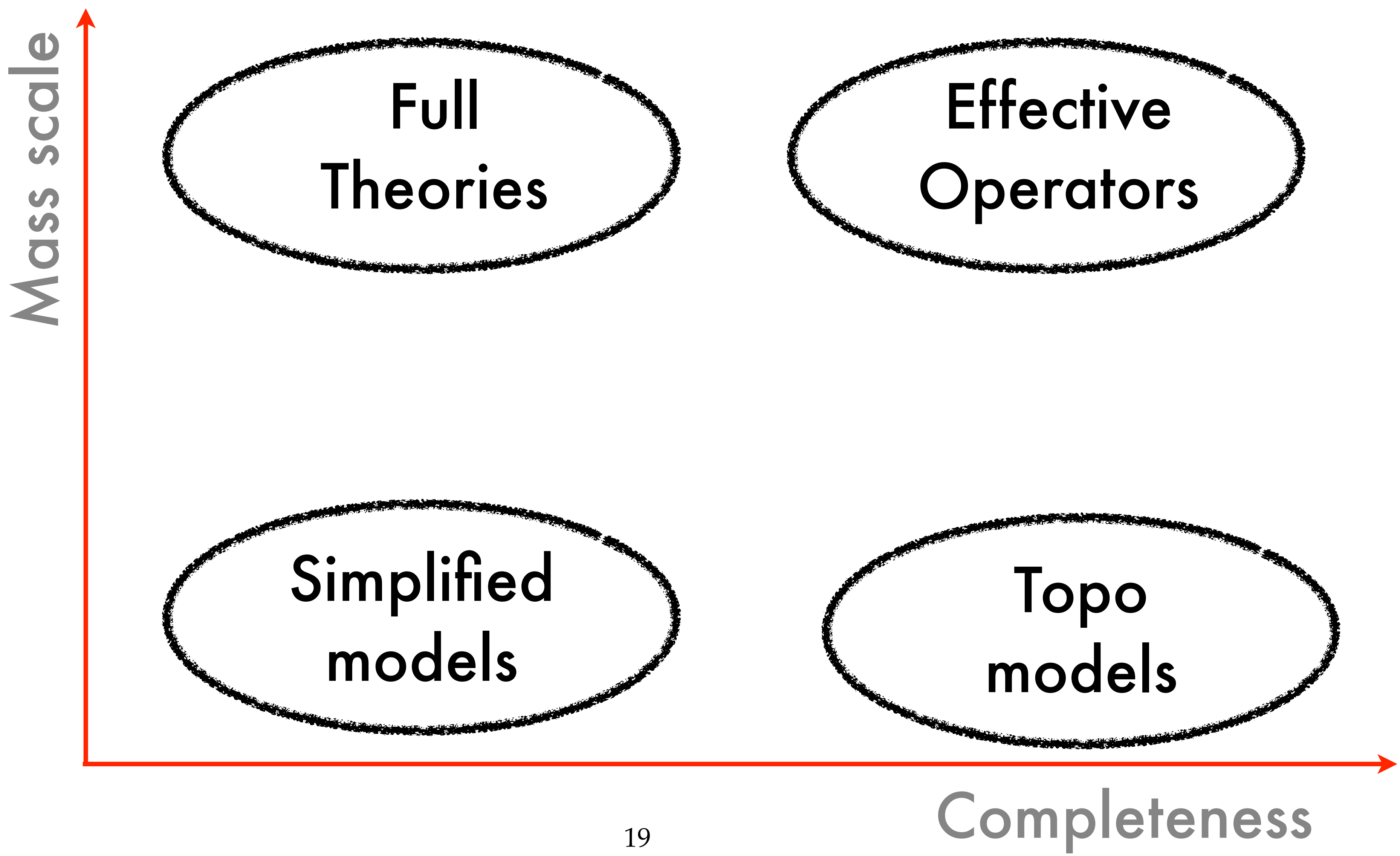
For a given final state (eg lljj) construct all models with resonances. Then look for them!



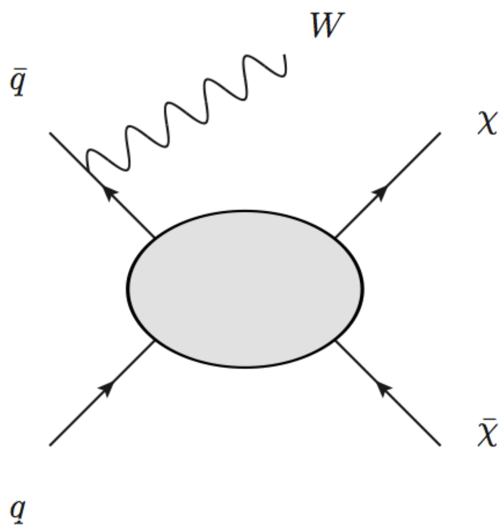
Connections to EFT, Simp. Models

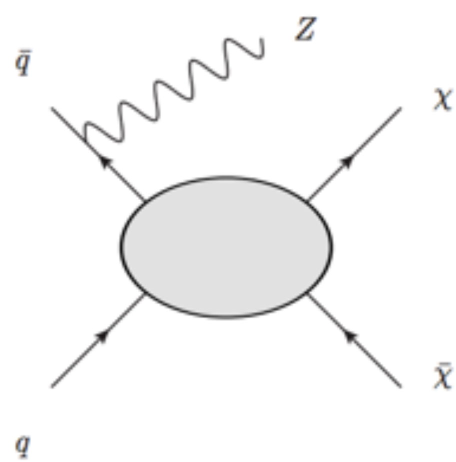


Connections to EFT, Simp. Models



Mono-Z'

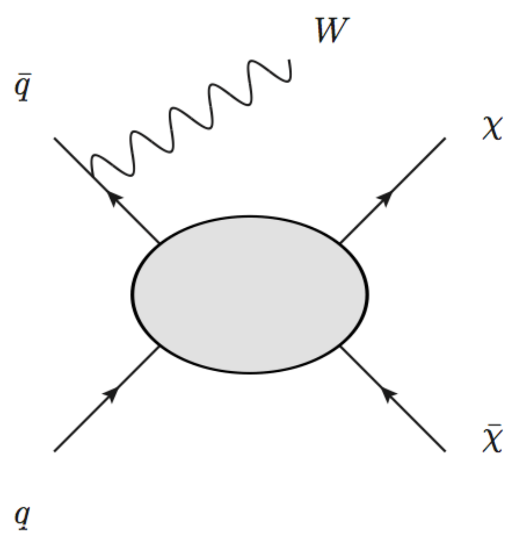


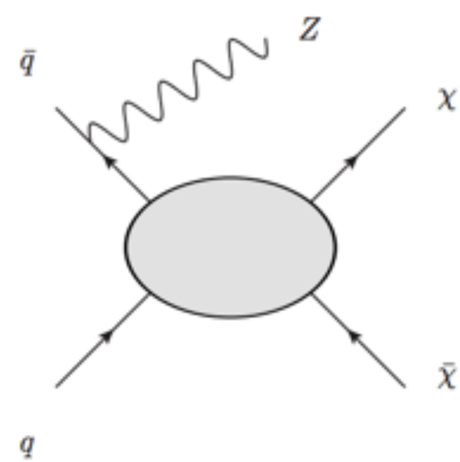


$m_{jj} = m_W$ or m_Z

$m_{ll} = m_Z$

Mono-Z'





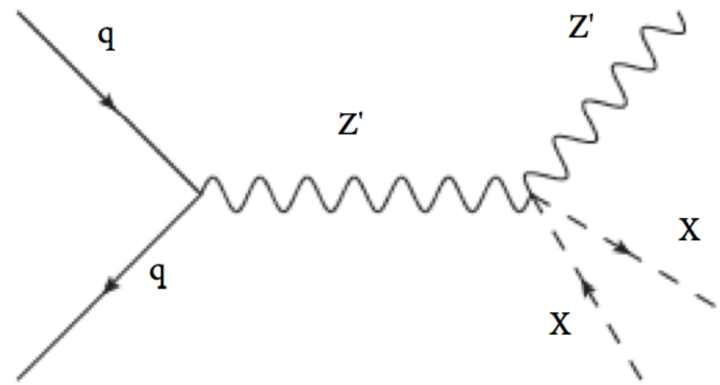
$m_{jj} = m_W$ or m_Z $\quad$ $m_{ll} = m_Z$

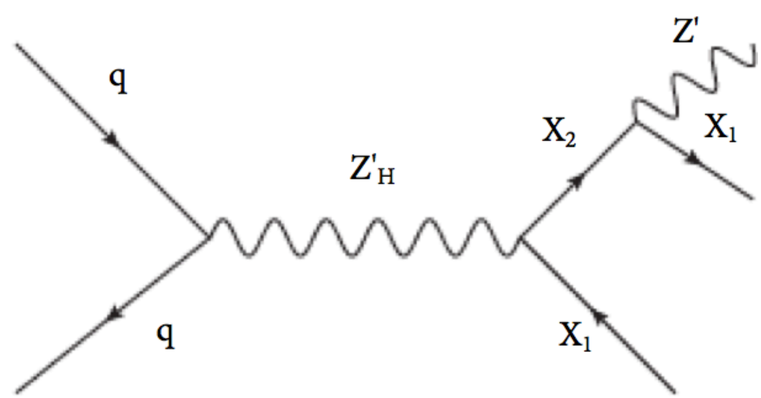
What about other values?

Mono-....

<u>Signature</u>

Heavy resonance
+ MET





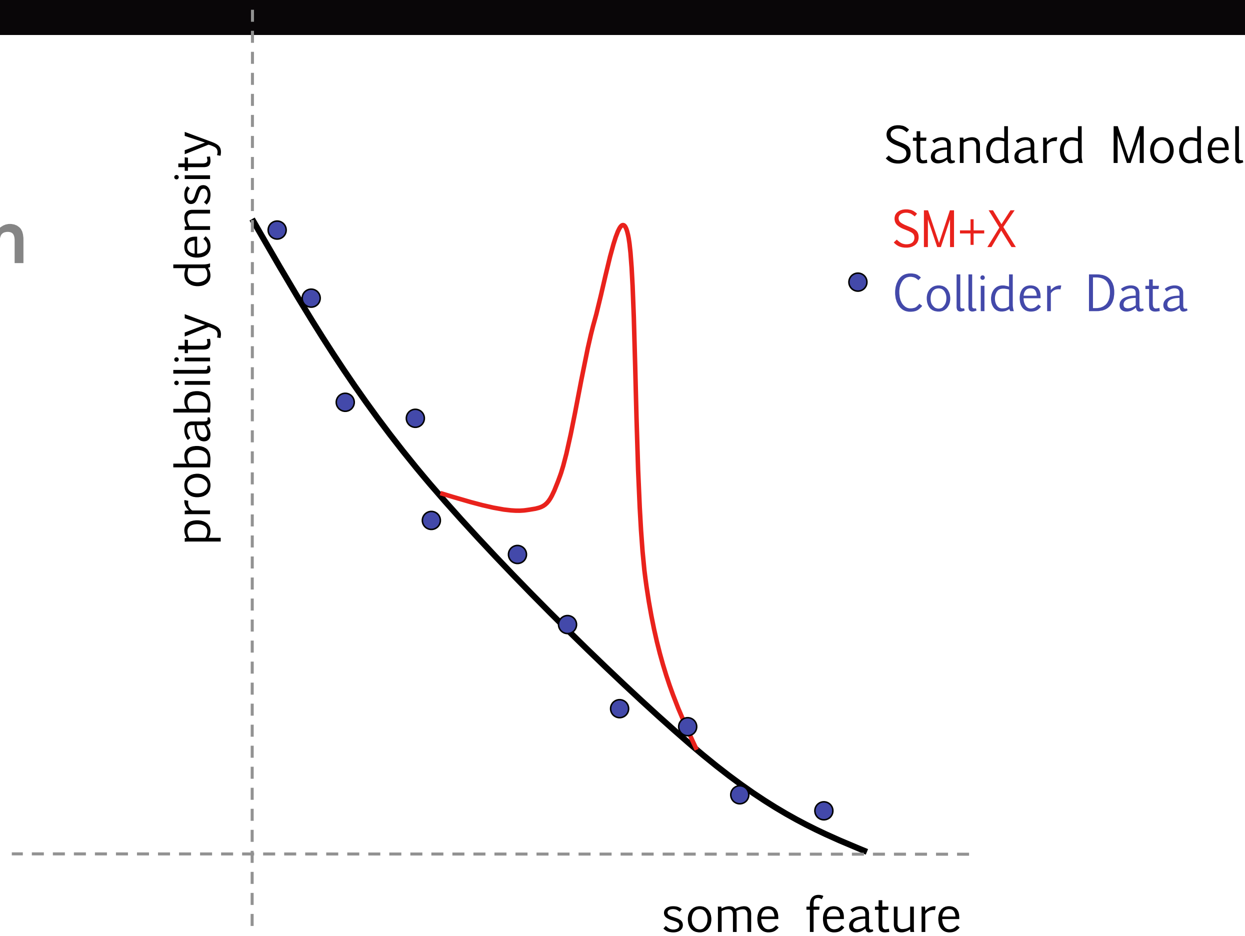
Outline

I. Strategy for unanticipated new physics

II. Deep networks for NP searches

How to find NP

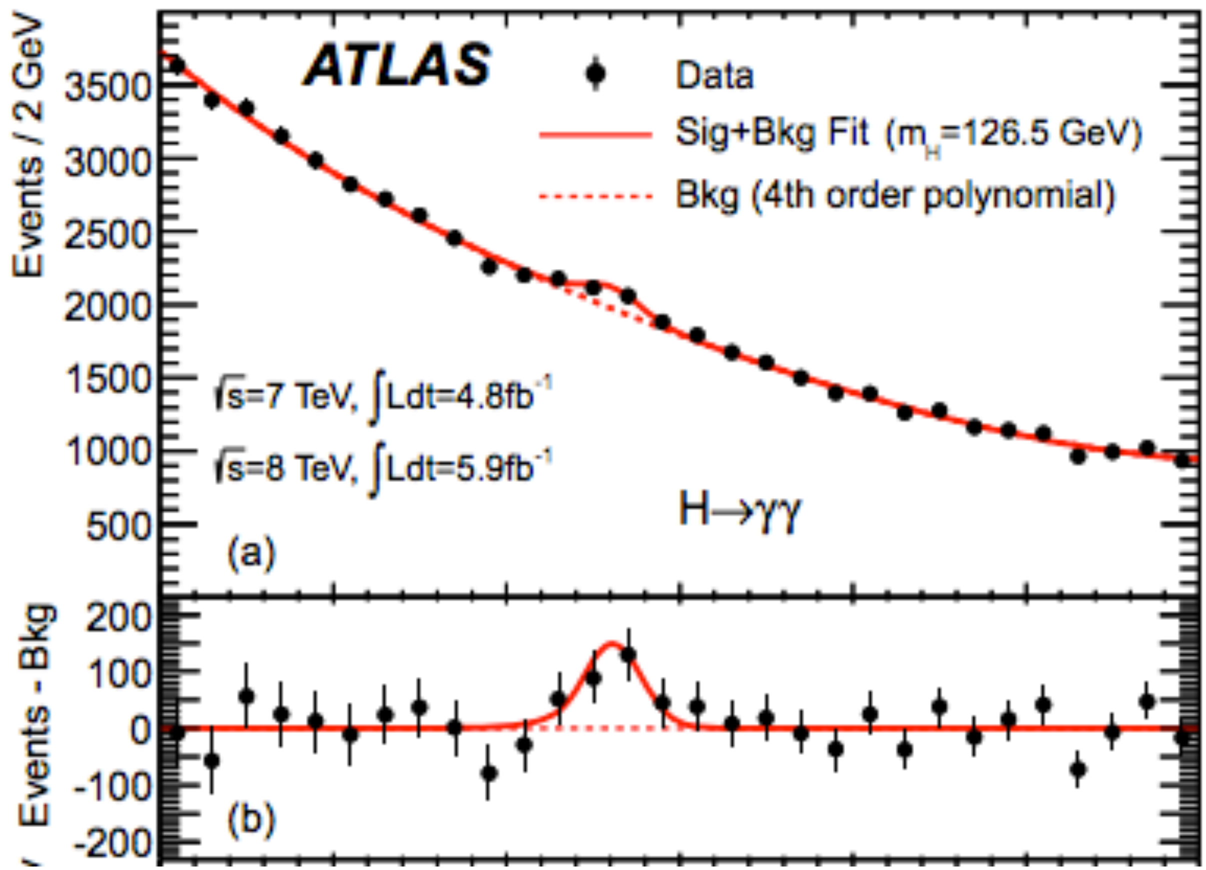
Isolate some feature in which two theories SM, SM+X can be best distinguished.



The data can tell us which hypothesis is preferred via a likelihood ratio:

$$\frac{L_{SM+X}}{L_{SM}} \qquad \frac{P(\text{data} \mid SM+X)}{P(\text{data} \mid SM)}$$

e.g.



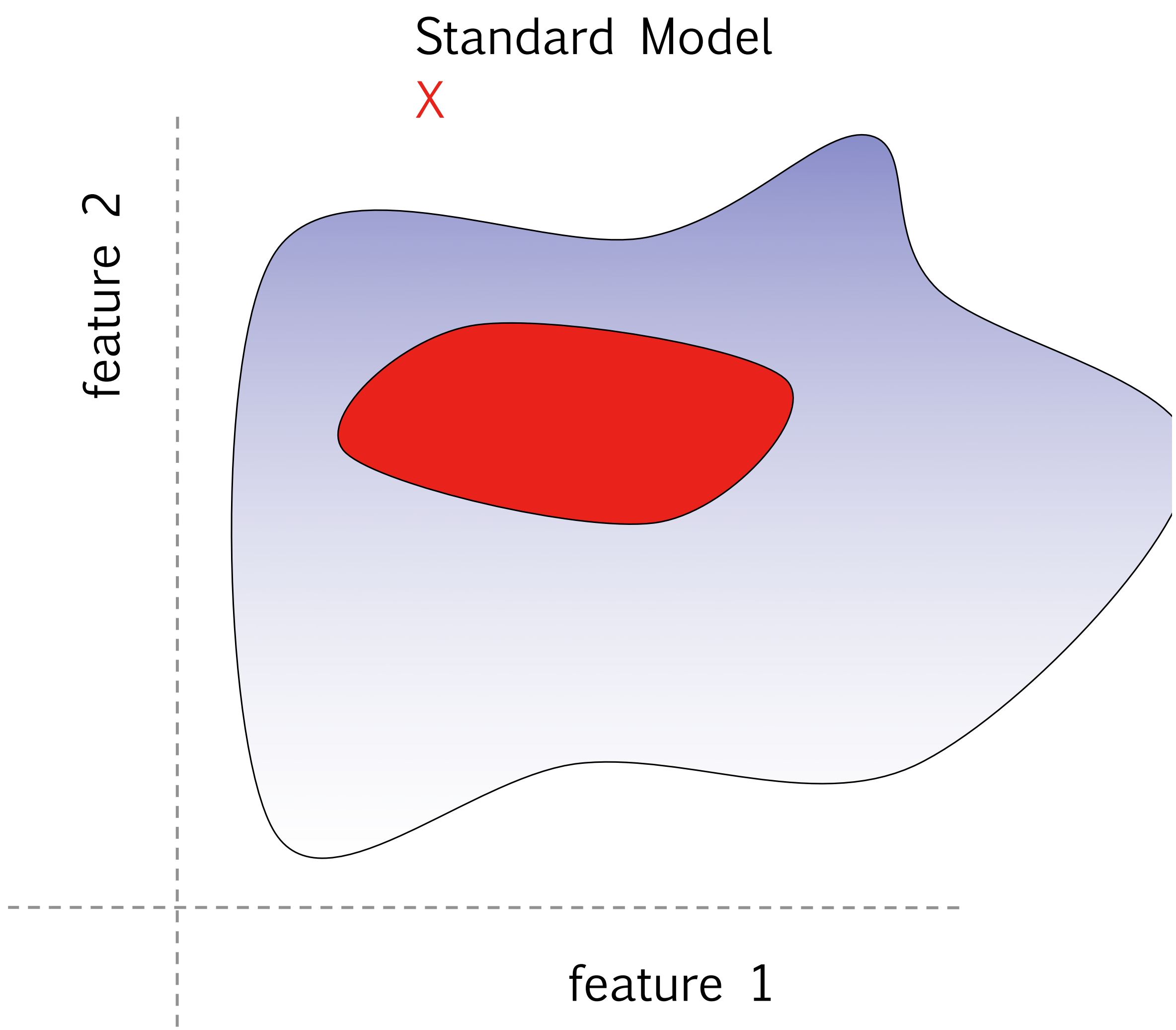
ATLAS
Data
Sig+Bkg Fit (m_H=126.5 GeV)
Bkg (4th order polynomial)
$\sqrt{s}$=7 TeV, $\int Ldt$=4.8fb^{-1}
$\sqrt{s}$=8 TeV, $\int Ldt$=5.9fb^{-1}
H→γγ
(a)
(b)
Events / 2 GeV
Events - Bkg
3500
3000
2500
2000
1500
1000
500
200
100
0
-100
-200

But...

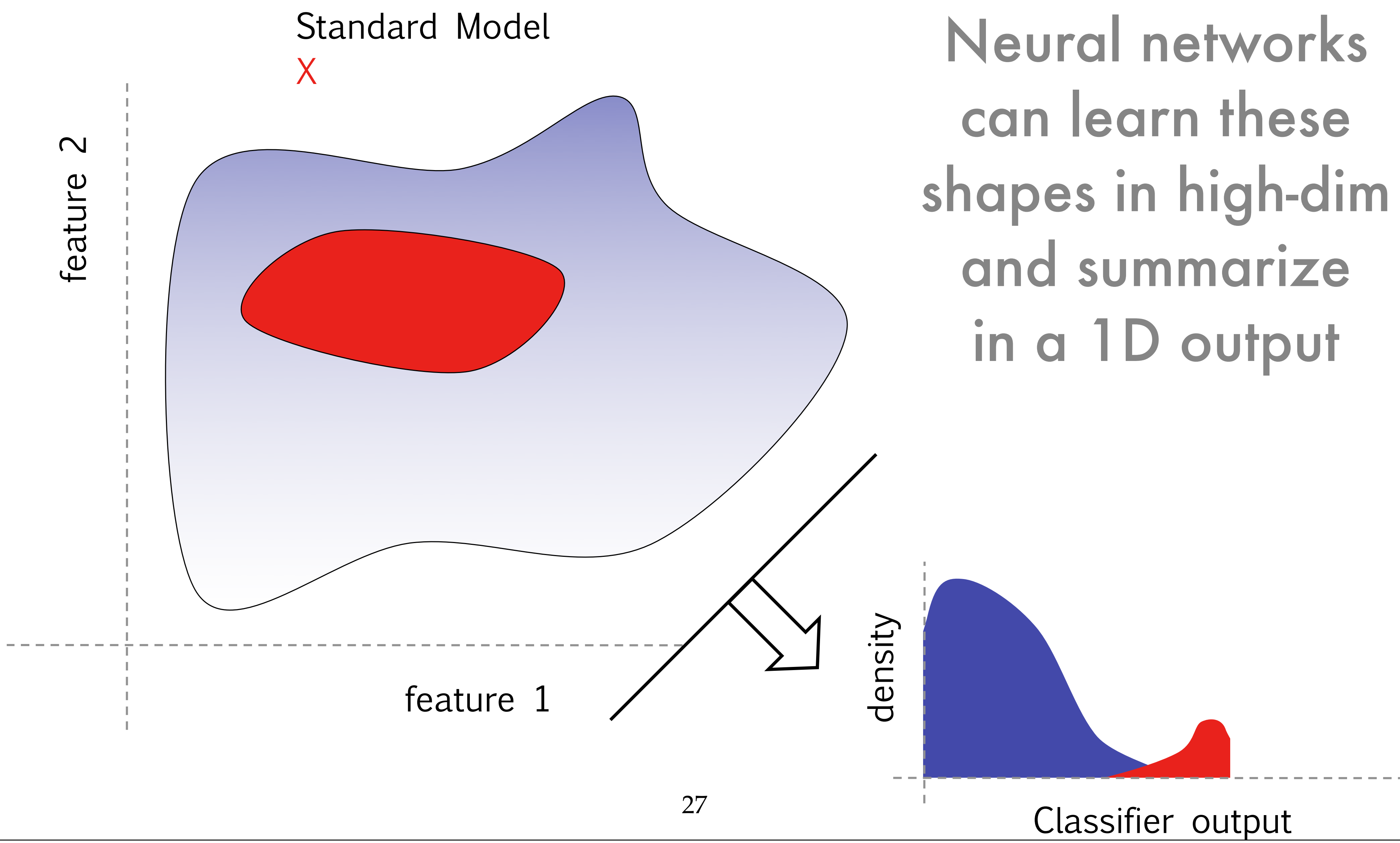
Reality is more complicated.

The full space can be very high dimensional.

Calculating likelihood in d-dimensional space requires $\sim 100^d$ MC events.



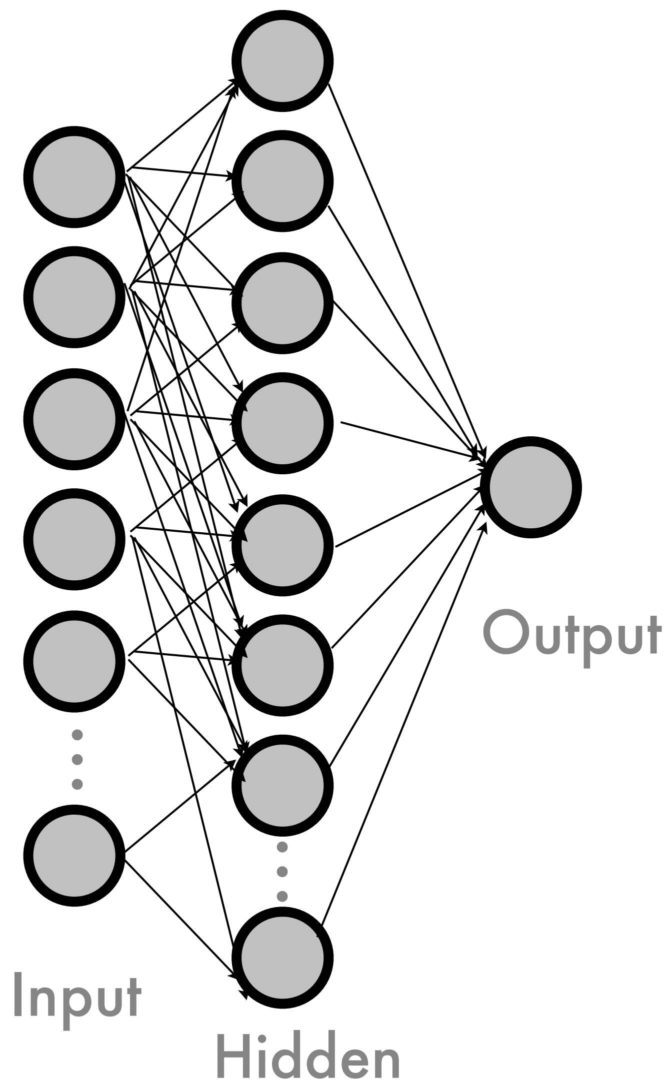
ML tools



Neural networks can learn these shapes in high-dim and summarize in a 1D output

Neural Networks

Essentially a functional fit with many parameters



Function

Each neuron's output is a function of the weighted sum of inputs.

Goal

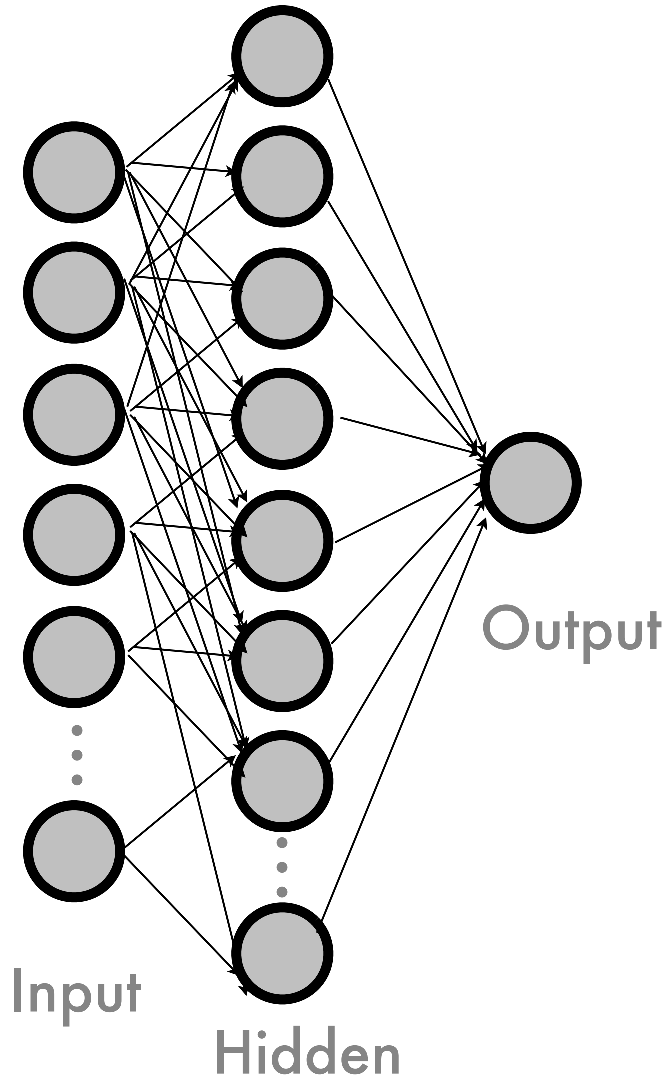
find set of weights which give most useful function

Learning

give examples, back-propagate error to adjust weights

Neural Networks

Essentially a functional fit with many parameters



Problem:
Networks with > 1 layer are very difficult to train.

Consequence:
Networks are not good at learning non-linear functions.
(like invariant masses!)

In short:
Can't just throw 4-vectors at NN.

Search for Input

ATLAS-CONF-2013-108

Can't just use 4v

Can't give it too many inputs

Painstaking search through input feature space.

Variable	VBF			Boosted		
	$\tau_{\rm lep}\tau_{\rm lep}$	$\tau_{\rm lep}\tau_{\rm had}$	$\tau_{\rm had}\tau_{\rm had}$	$\tau_{\rm lep}\tau_{\rm lep}$	$\tau_{\rm lep}\tau_{\rm had}$	$\tau_{\rm had}\tau_{\rm had}$
$m_{\tau\tau}^{\rm MMC}$	•	•	•	•	•	•
$\Delta R(\tau,\tau)$	•	•	•		•	•
$\Delta\eta(j_1,j_2)$	•	•	•			
m_{j_1,j_2}	•	•	•			
$\eta_{j_1}\times\eta_{j_2}$		•	•			
$p_{\rm T}^{\rm Total}$		•	•			
sum $p_{\rm T}$					•	•
$p_{\rm T}(\tau_1)/p_{\rm T}(\tau_2)$					•	•
$E_{\rm T}^{\rm miss}\phi$ centrality		•	•	•	•	•
$x_{\tau 1}$ and $x_{\tau 2}$						•
$m_{\tau\tau,j_1}$				•		
m_{ℓ_1,ℓ_2}				•		
$\Delta\phi_{\ell_1,\ell_2}$				•		
sphericity				•		
$p_{\rm T}^{\ell_1}$				•		
$p_{\rm T}^{j_1}$				•		
$E_{\rm T}^{\rm miss}/p_{\rm T}^{\ell_2}$				•		
$m_{\rm T}$		•			•	
$\min(\Delta\eta_{\ell_1\ell_2,\rm jets})$	•					
j_3 η centrality	•					
$\ell_1\times\ell_2$ η centrality	•					
ℓ η centrality		•				
$\tau_{1,2}$ η centrality			•			

Table 3: Discriminating variables used for each channel and category. The filled circles identify which variables are used in each decay mode. Note that variables such as $\Delta R(\tau,\tau)$ are defined either between the two leptons, between the lepton and $\tau_{\rm had}$, or between the two $\tau_{\rm had}$ candidates, depending on the decay mode.

Search for Input

ATLAS-CONF-2013-108

Can't just use 4v

Can't give it t[...]
many inp[...]

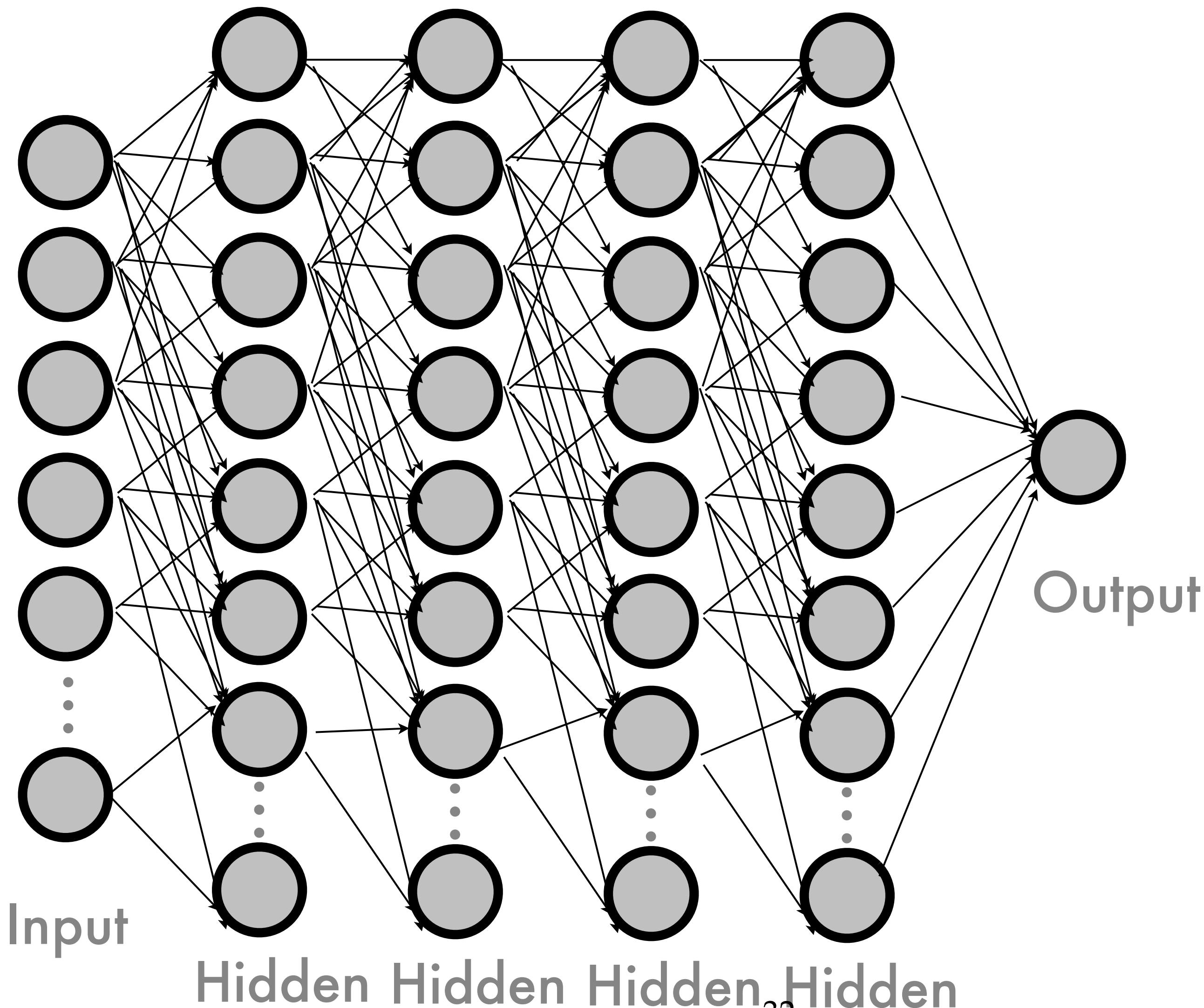
Painstaking[...]
through inp[...]
feature space.

Also true for BDTs, SVNs, etc

Variable	VBF			Boosted		
	$\tau_{lep}\tau_{lep}$			$\tau_{lep}\tau_{lep}$	$\tau_{lep}\tau_{had}$	$\tau_{had}\tau_{had}$
$m_{\tau\tau}^{MMC}$				•	•	•
$\Delta R(\ldots$					•	•
					•	•
					•	•
					•	•
						•
				•		
				•		
				•		
				•		
$p_T^{j_1}$				•		
$E_T^{miss}/p_T^{\ell_2}$				•		
m_T		•			•	
$\min(\Delta\eta_{\ell_1\ell_2,jets})$	•					
j_3 η centrality	•					
$\ell_1 \times \ell_2$ η centrality	•					
ℓ η centrality		•				
$\tau_{1,2}$ η centrality			•			

Table 3: Discriminating variables used for each channel and category. The filled circles identify which variables are used in each decay mode. Note that variables such as $\Delta R(\tau,\tau)$ are defined either between the two leptons, between the lepton and τ_{had}, or between the two τ_{had} candidates, depending on the decay mode.

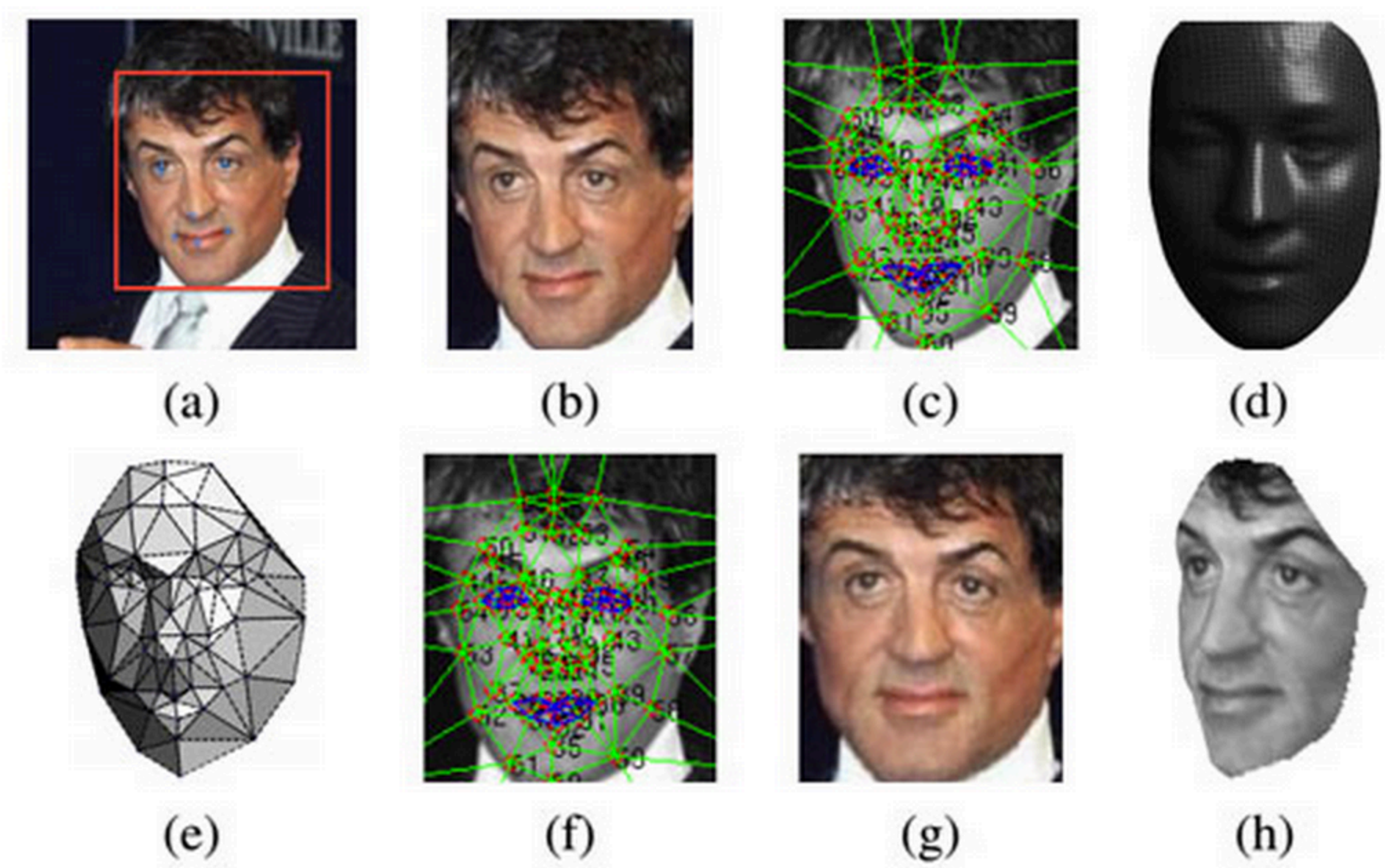
Deep networks



New tools let us train deep networks.

How well do they work?

Real world applications



Head turn: DeepFace uses a 3-D model to rotate faces, virtually, so that they face the camera. Image (a) shows the original image, and (g) shows the final, corrected version.

Paper

Deep Learning in High-Energy Physics: Improving the Search for Exotic Particles

P. Baldi,[1] P. Sadowski,[1] and D. Whiteson[2]

[1] *Dept. of Computer Science, UC Irvine, Irvine, CA 92617*

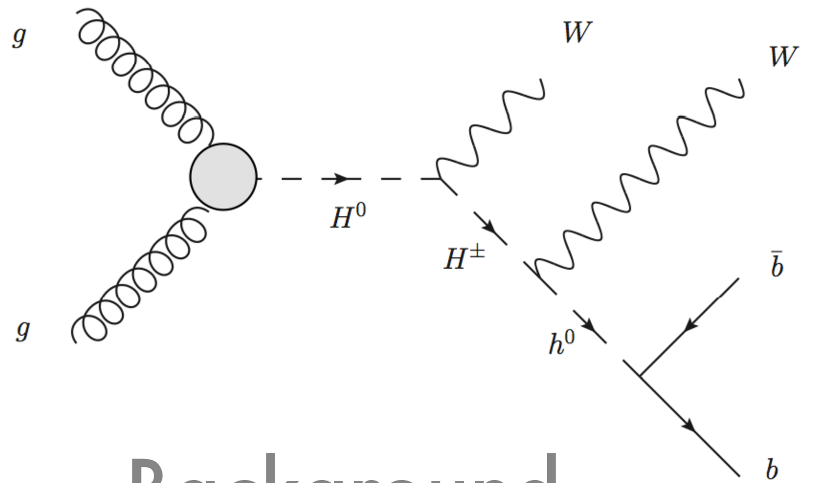
[2] *Dept. of Physics and Astronomy, UC Irvine, Irvine, CA 92617*

arXiv: 1402.4735

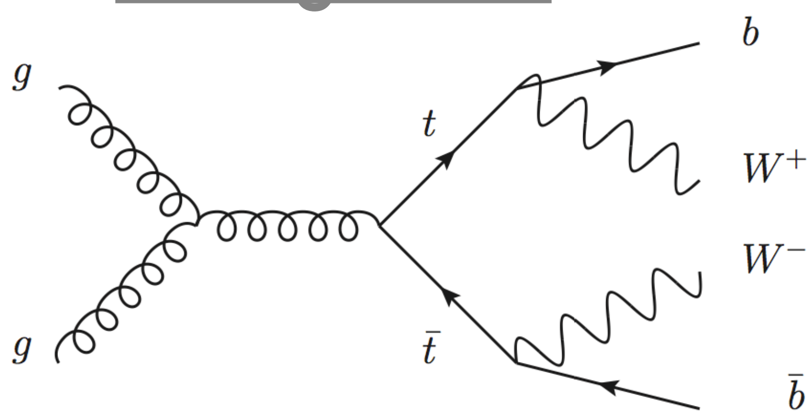
Accepted in *Nature Comm.*

Benchmark problem

Signal



Background



Can deep networks automatically discover useful variables?

4-vector inputs

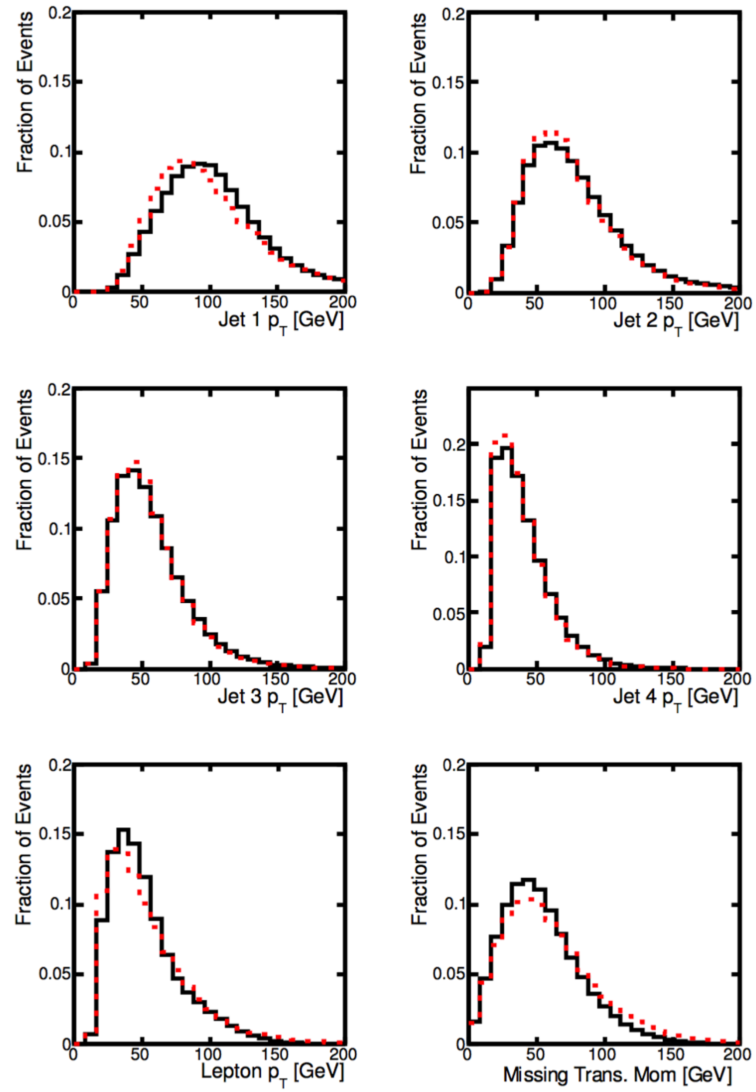
<u>21 Low-level vars</u>

jet+lepton mom. (3x5)

missing ET (2)

jet btags (4)

Not much separation visible in 1D projections

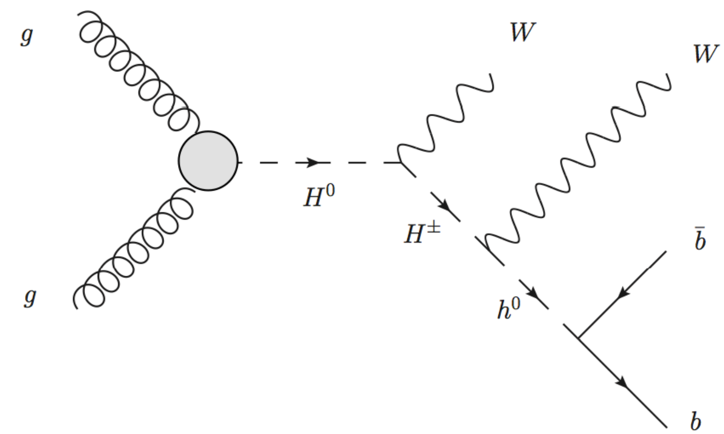


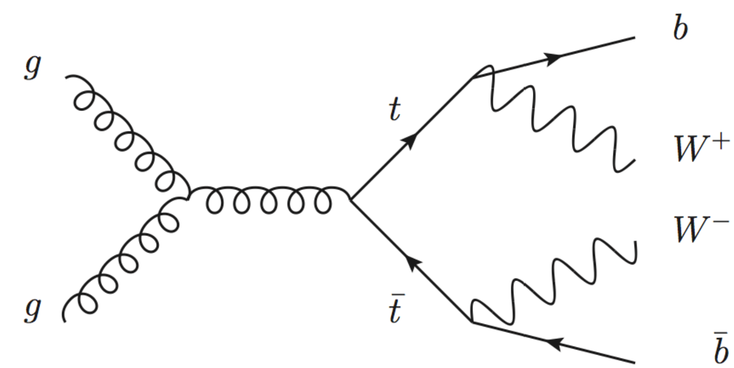
4-vector inputs

<u>7 High-level vars</u>

m(WWbb)
m(Wbb)
m(bb)

m(bjj)
m(jj)
m(lv)
m(blv)





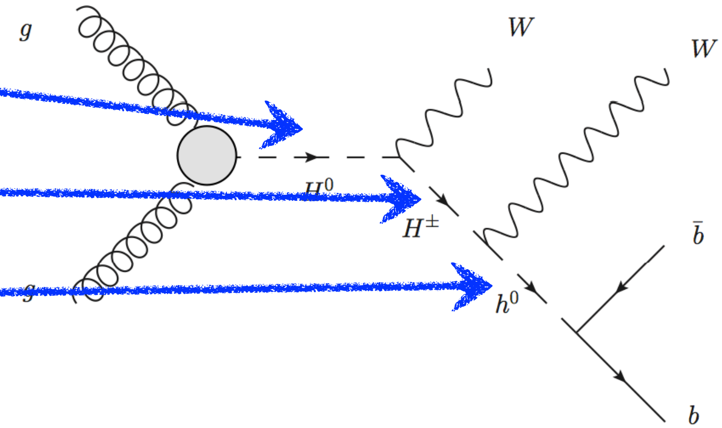
4-vector inputs

<u>7 High-level vars</u>

m(WWbb)

m(Wbb)

m(bb)

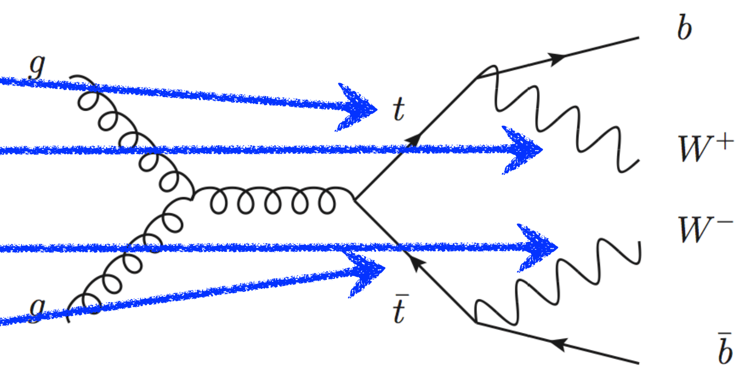


m(bjj)

m(jj)

m(lv)

m(blv)

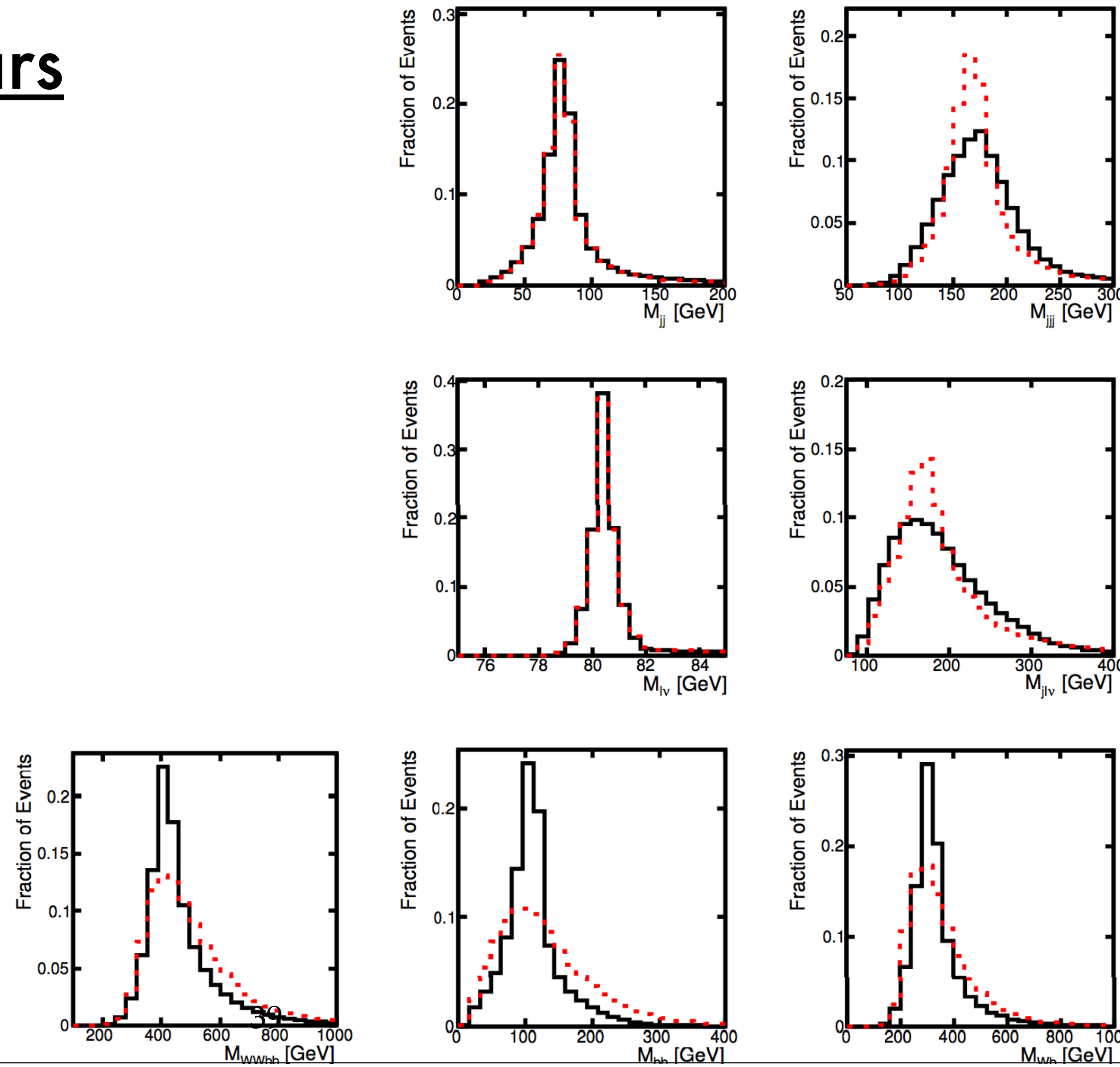


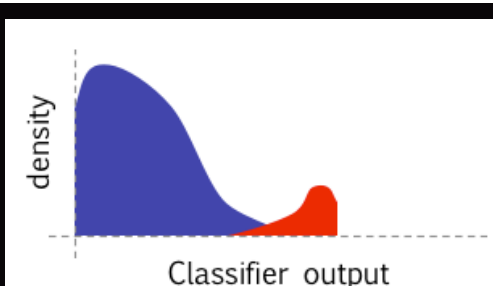
4-vector inputs

7 High-level vars

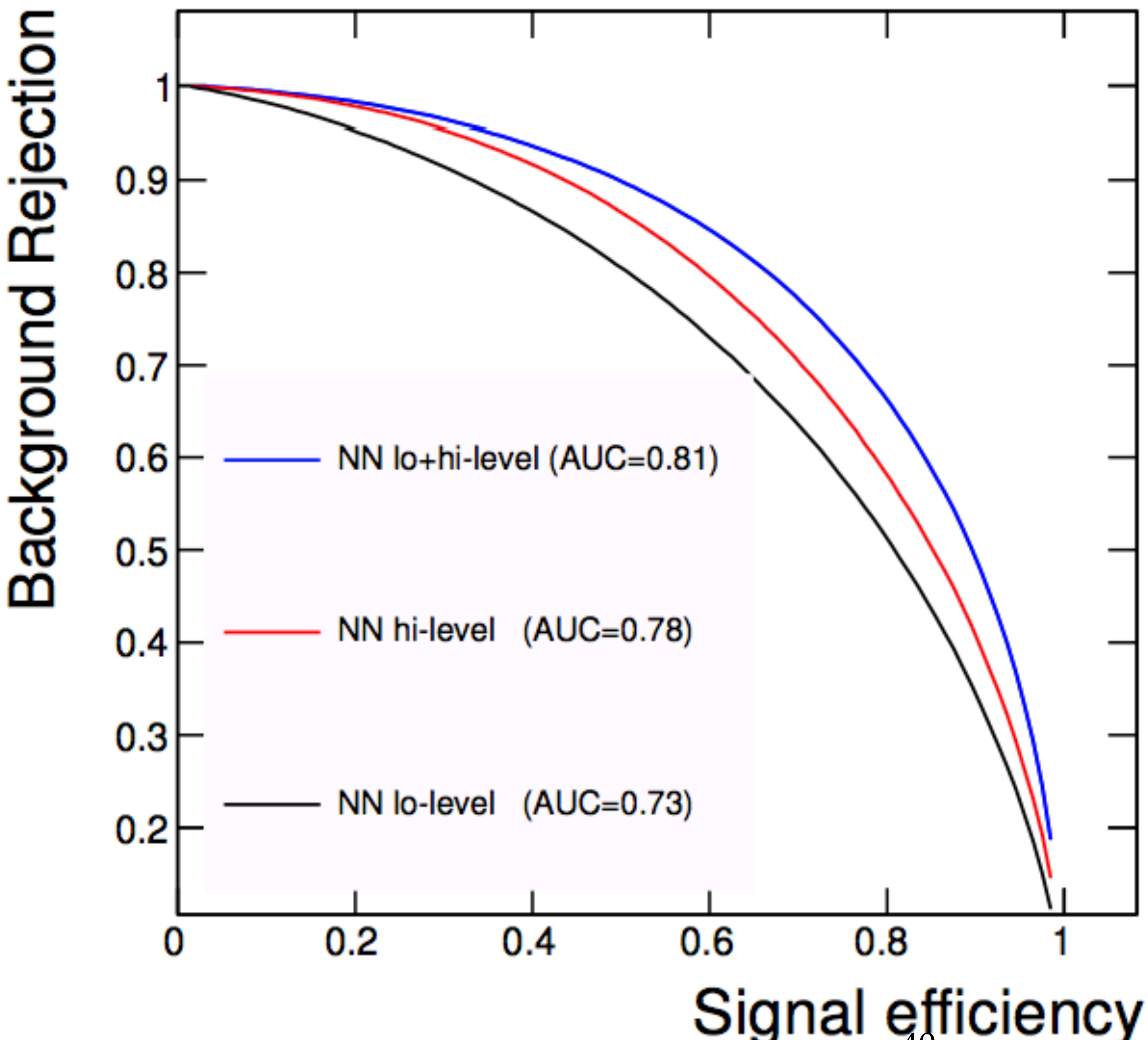
m(WWbb)
m(Wbb)
m(bb)

m(bjj)
m(jj)
m(lv)
m(blv)





Standard NNs



Results

Adding hi-level boosts performance

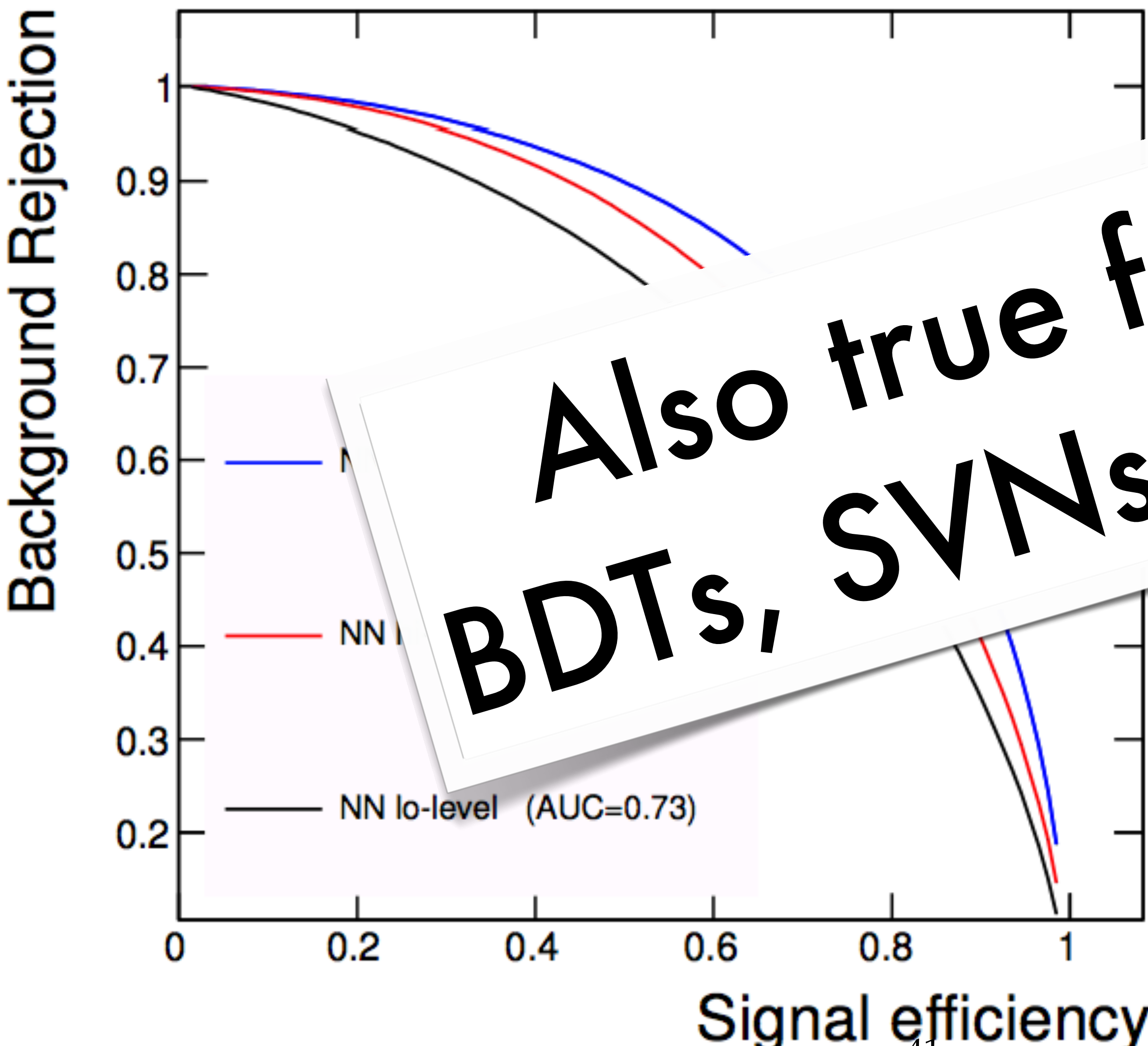
Better: lo+hi-level.

Conclude:

NN can't find hi-level vars.

Hi-level vars do not have all info

Standard NNs



Also true for BDTs, SVNs, etc

Results

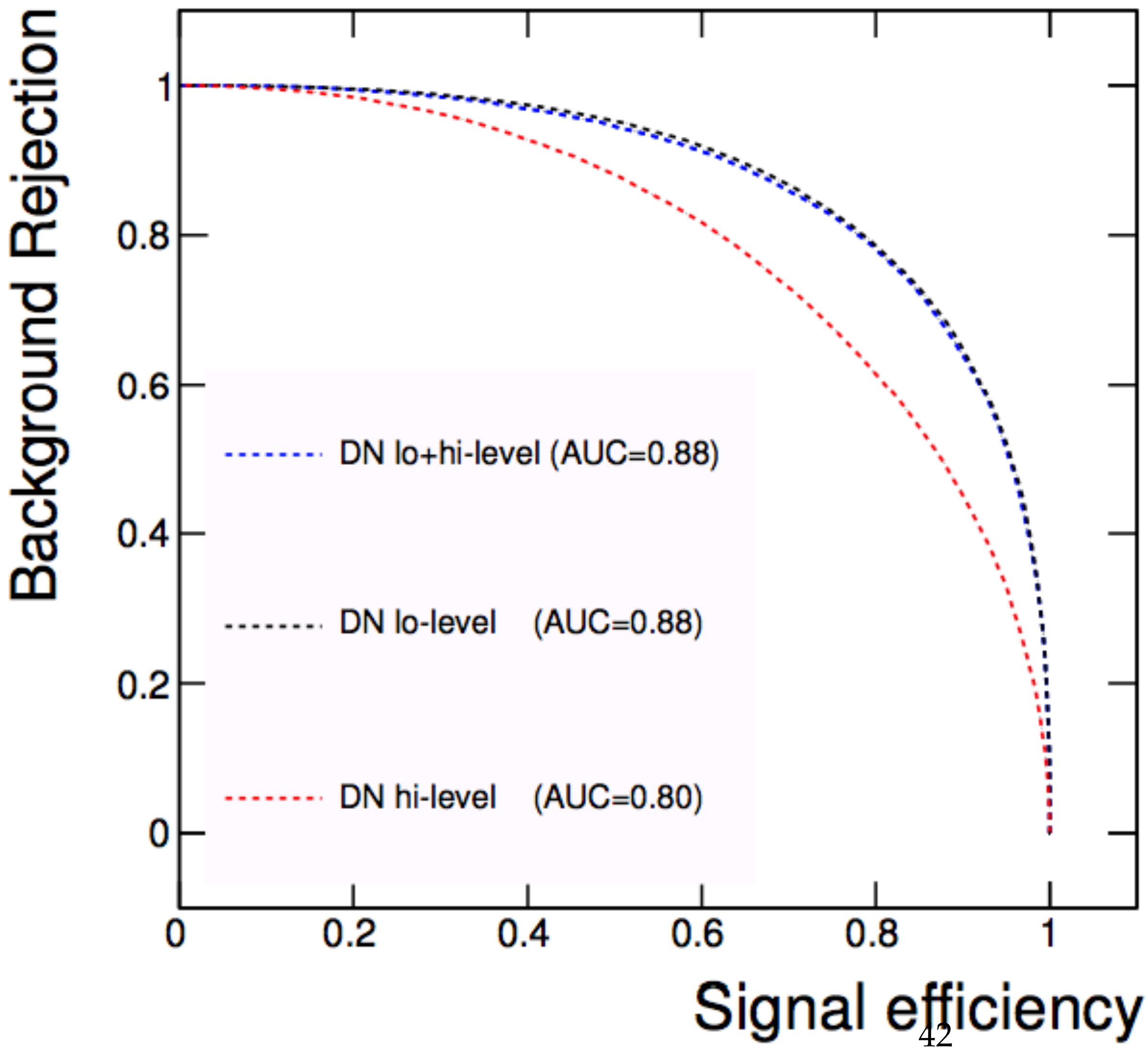
...g hi-level
...s performance
...lo+hi-level.

...nclude:

NN can't find hi-level vars.

Hi-level vars do not have all info

Deep Networks



Results

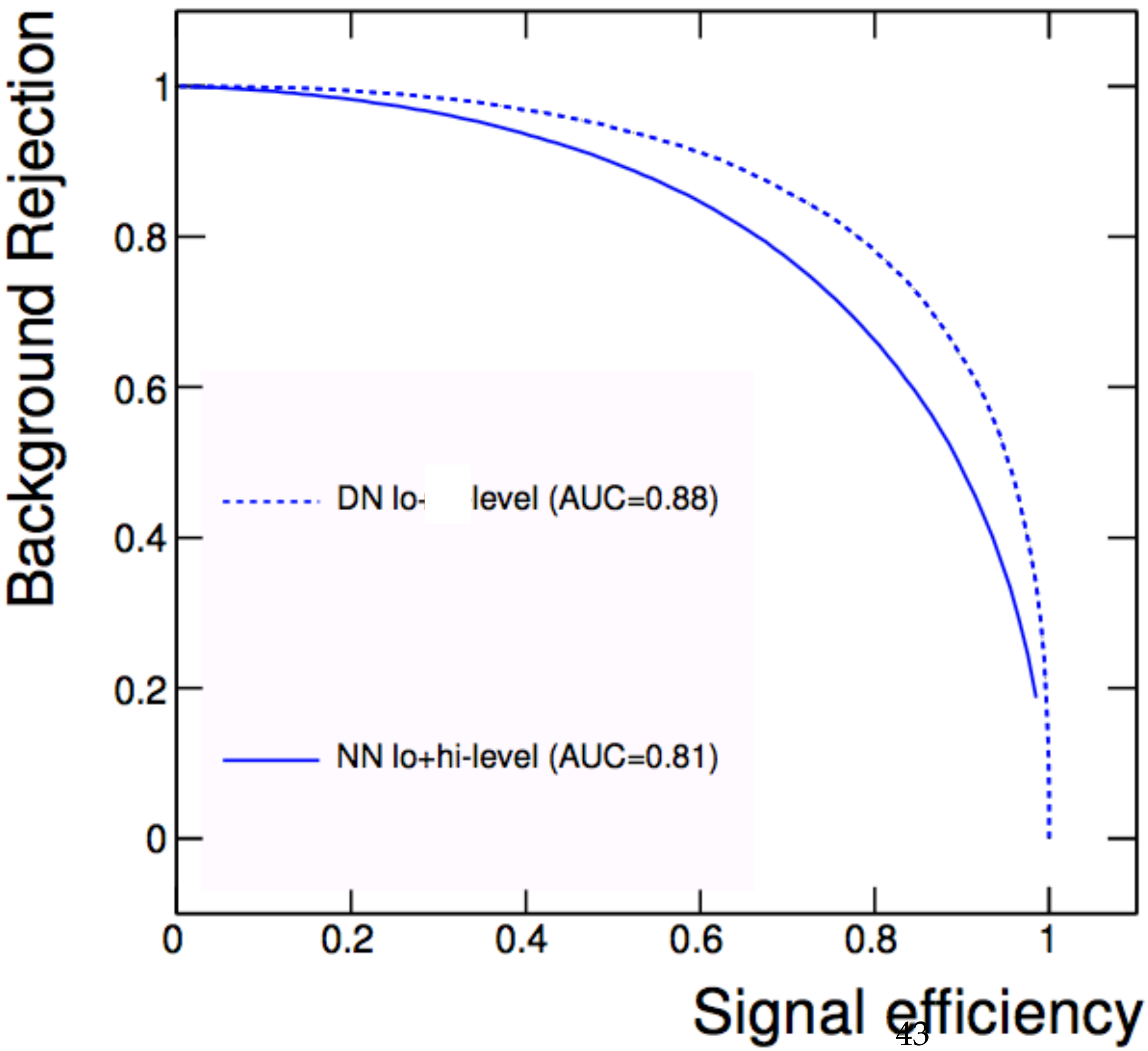
Lo+hi = lo.

Conclude:

DN can find hi-level vars.

Hi-level vars do not have all info are unnecessary

Deep Networks



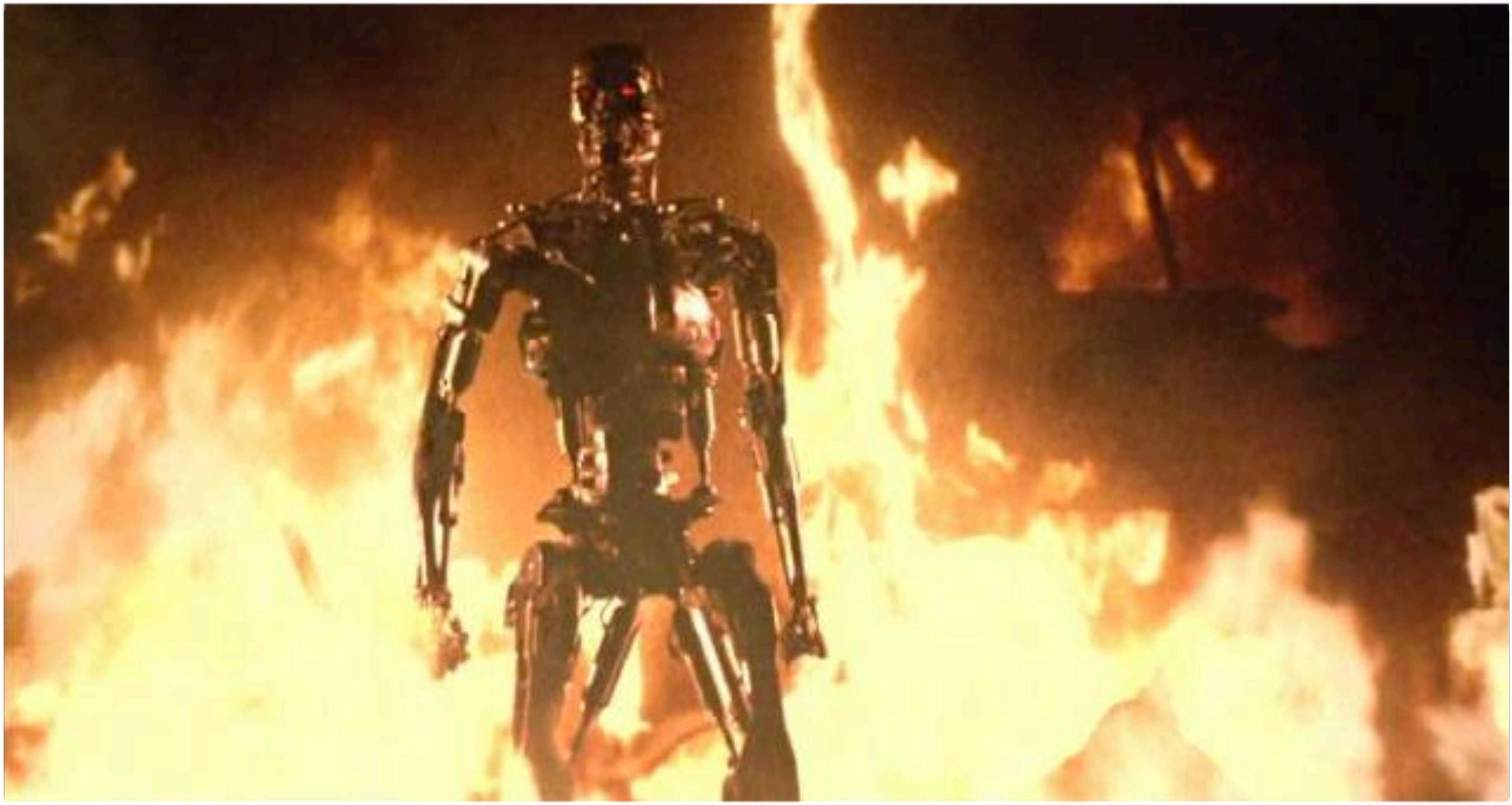
Results

DN > NN

Conclude:

DN does better than human assisted NN

The AIs win



Results

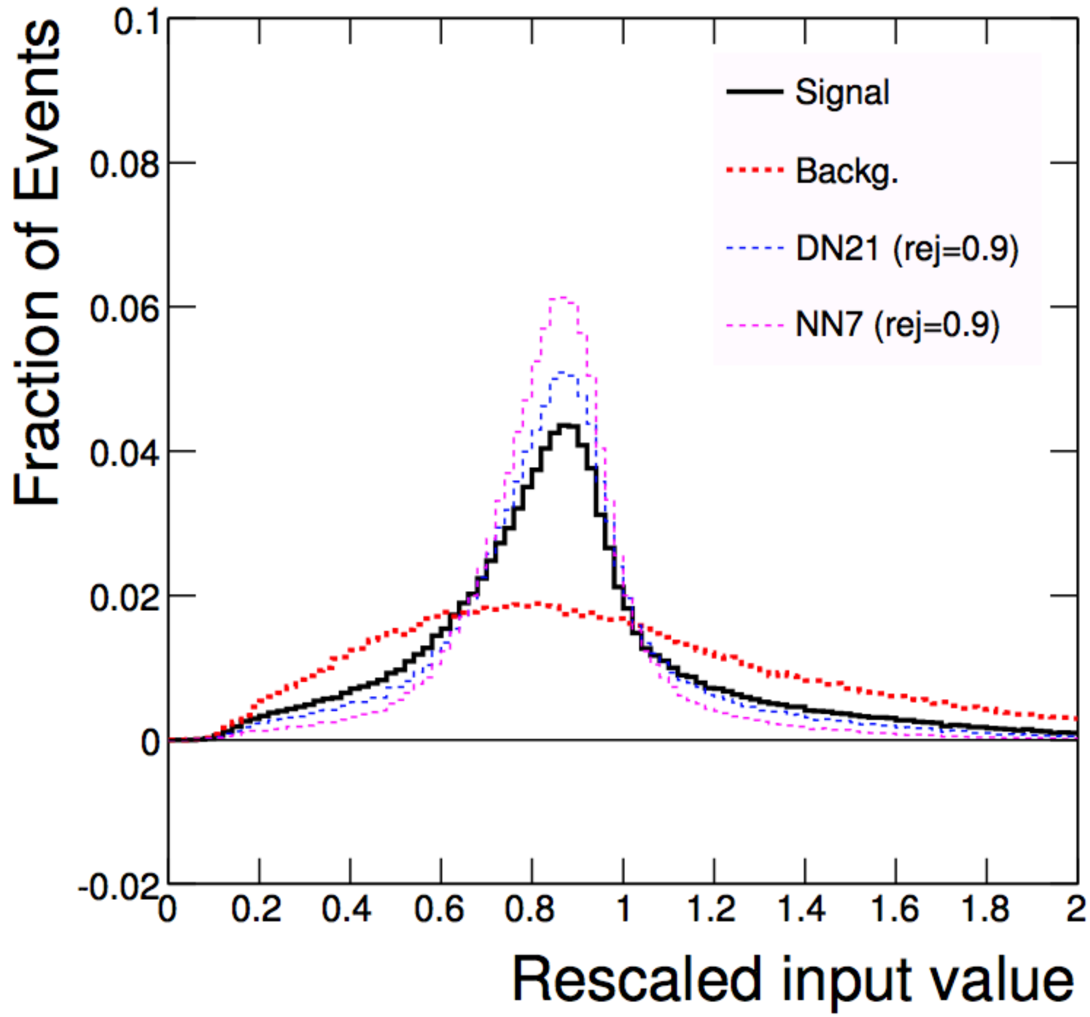
Identified example benchmark where traditional NNs fail to discover all discrimination power.

Adding human insight helps traditional NNs.

Deep networks succeed **without human insight.** Outperform human-boosted traditional NNs.

Why?

DN not as reliant on signal features. Cuts into background space.



Fin