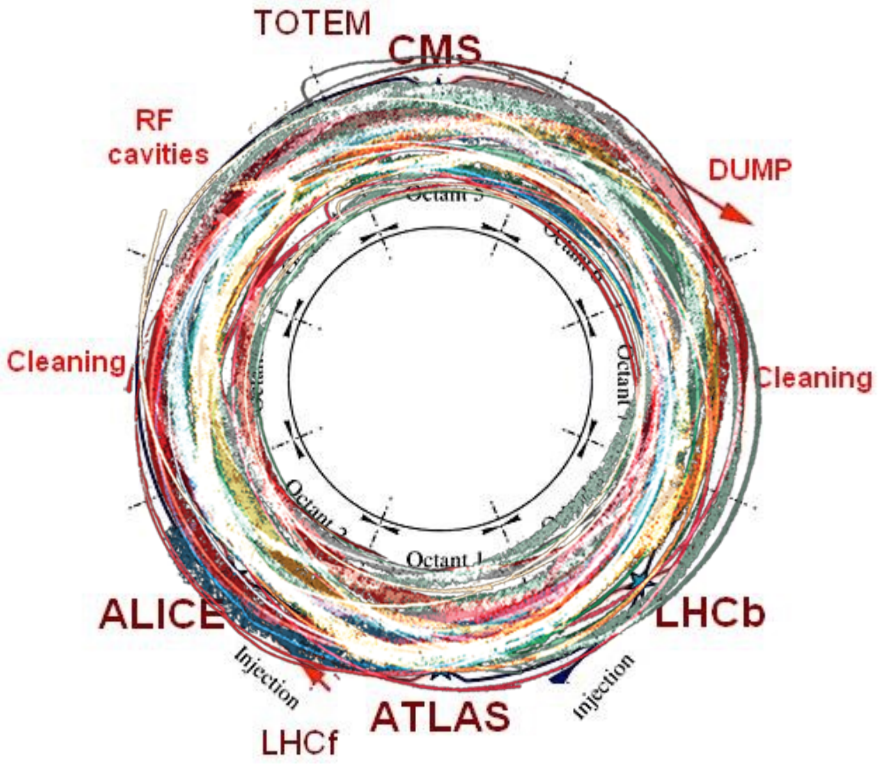


Deep Learning for the LHC physics

Myeonghun Park

서울과학기술대학교
1910
SEOUL NATIONAL UNIV. OF SCIENCE & TECHNOLOGY

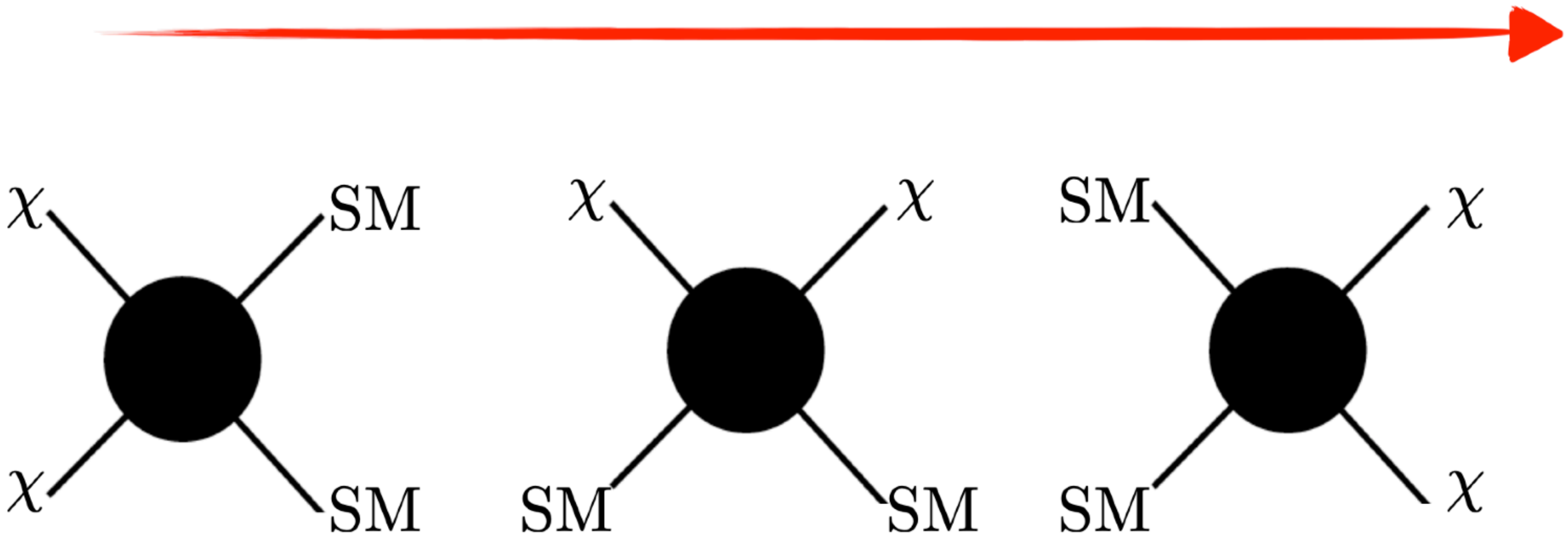
Deep Learning: arXiv:1904.08549 (JHEP 2019)

IBS-MultiDark-IPPP 2019

Experimental confirmation of out theoretical expectations

- How a **theory** is beautiful, as a physicist we need to confirm our theory with **experiments**.

“active”



Dark matter Indirect searches

Dark matter Direct searches

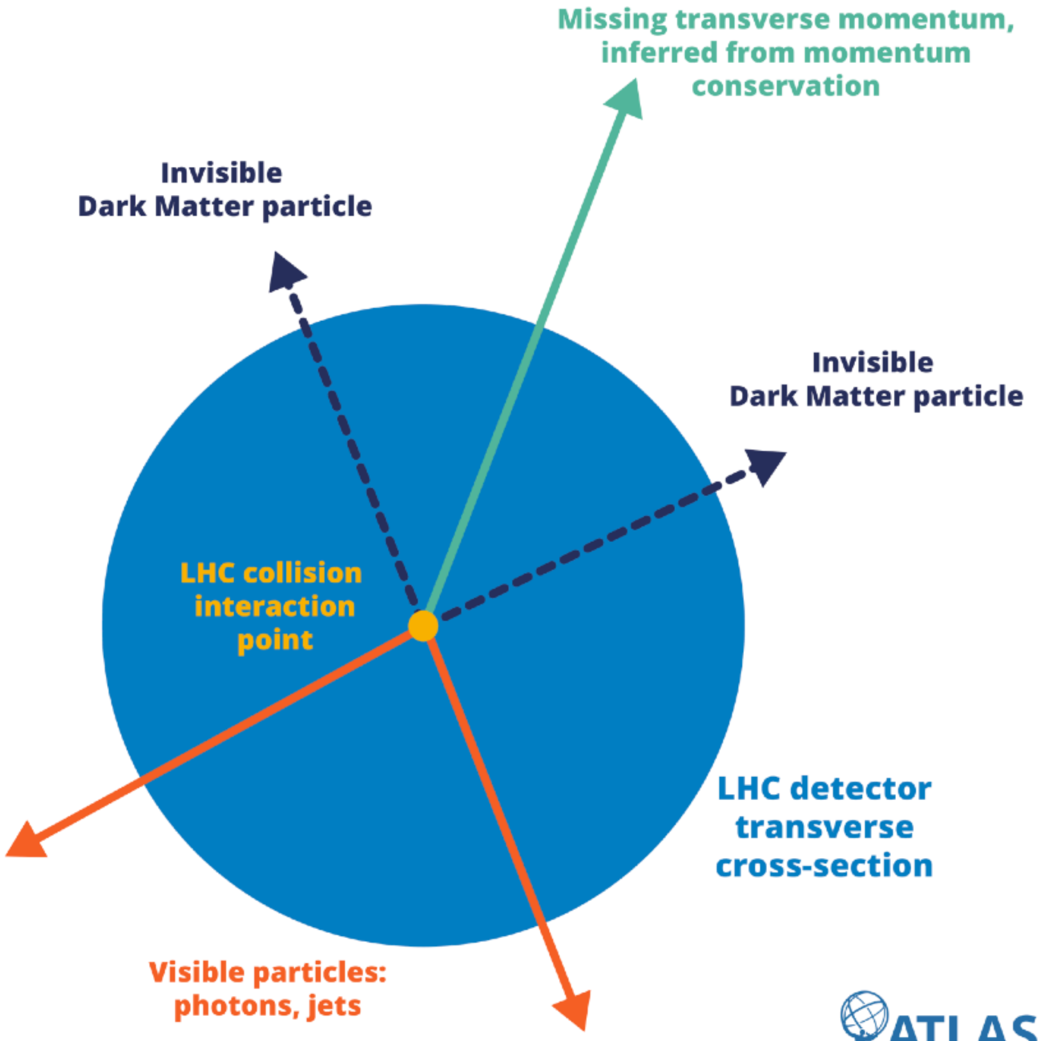
Collider

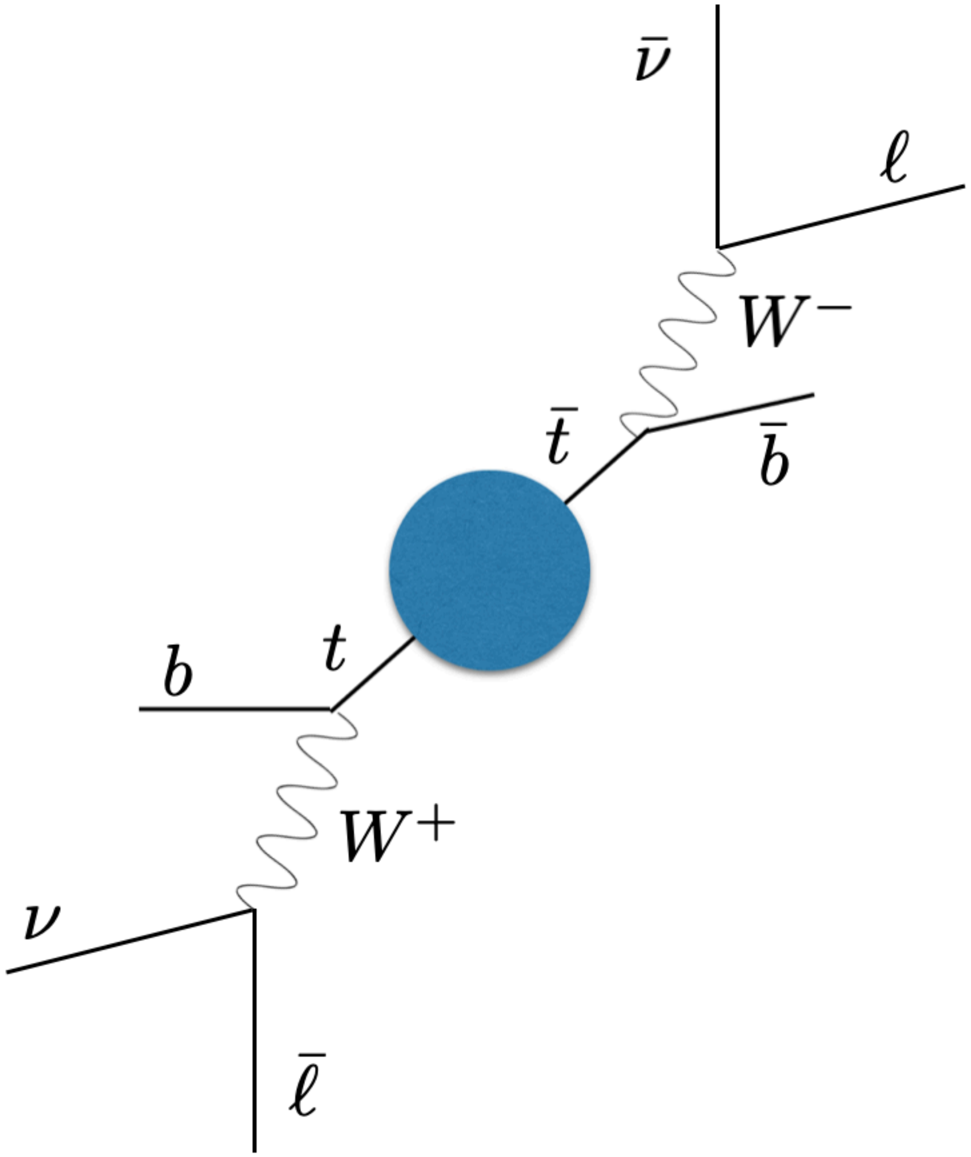
- How can we maximize the chance of the LHC ?

- LHC provides complicated data in an unprecedented way.
  - **Huge** QCD / Standard Model backgrounds.

- "Invisible" dark matter provides missing transverse energy

- Neutrinos from $t\bar{t}$ also provides "similar signature"


Missing transverse momentum, inferred from momentum conservation
Invisible Dark Matter particle
Invisible Dark Matter particle
LHC collision interaction point
LHC detector transverse cross-section
Visible particles: photons, jets
ATLAS EXPERIMENT



- LHC provides complicated data in an unprecedented way.
  - **Huge** QCD / Standard Model backgrounds.

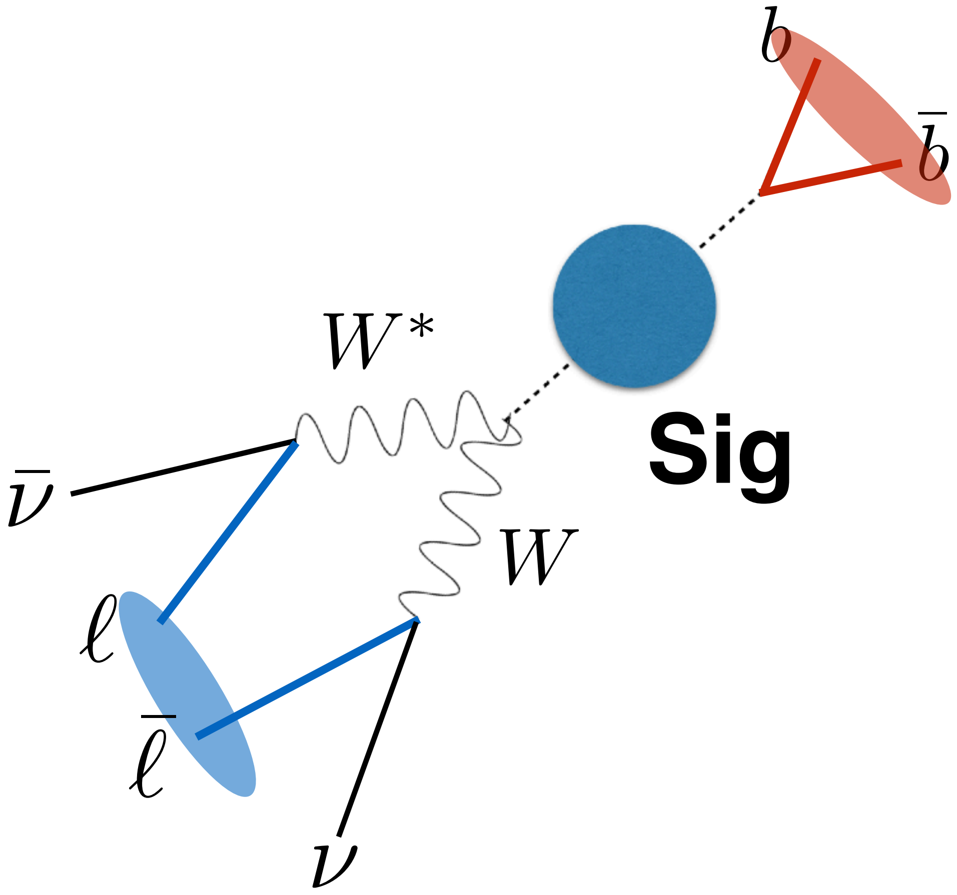
  : **Efficient way to reduce "unwanted" backgrounds with helps from data science (Deep Learning: DL)**

- Neutrinos from $t\bar{t}$ also provides "similar signature"

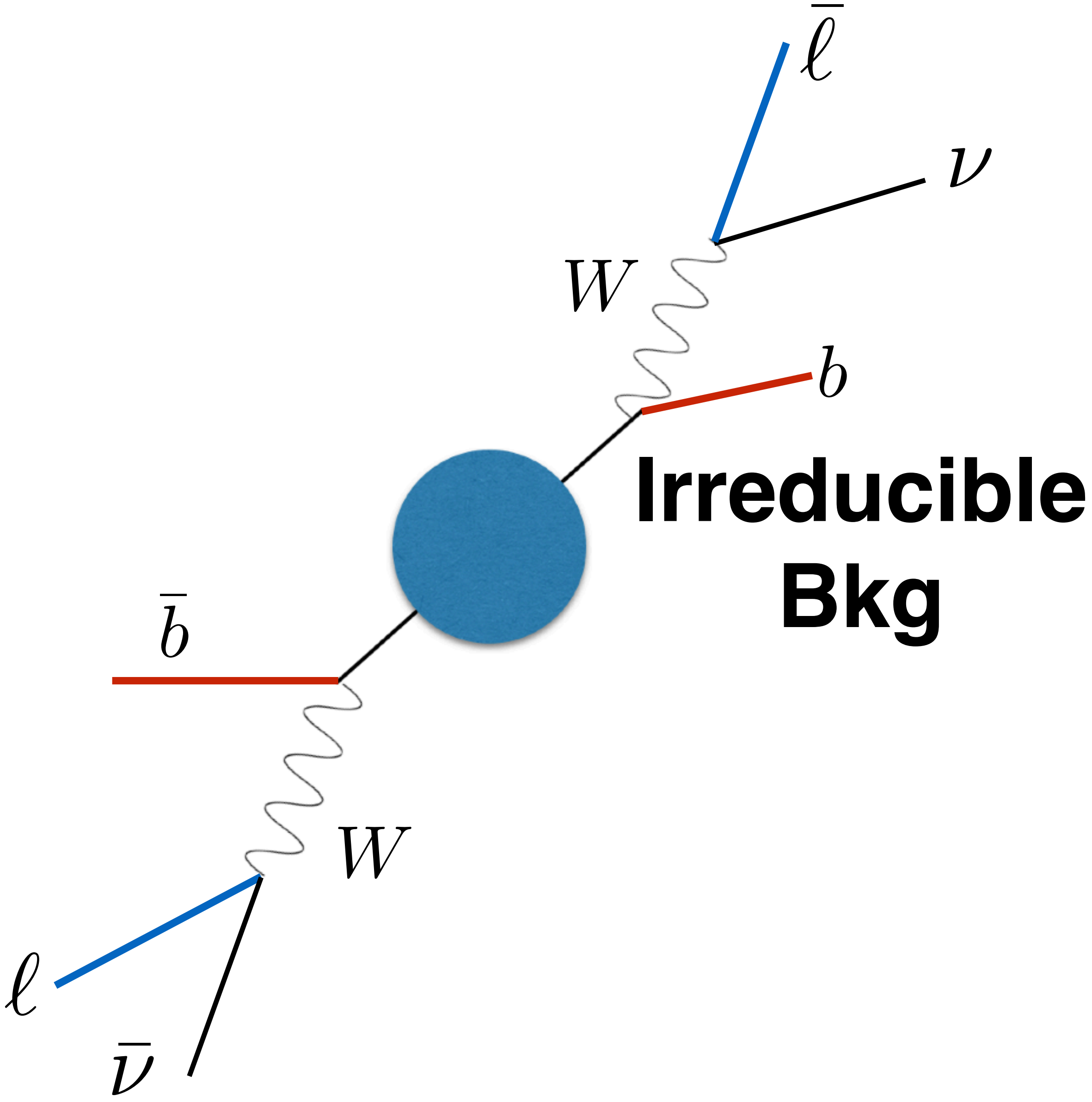
  : To understand DL performance, I will take one example from my recent works. (HH)

- This example is "supervised" Machine Learning

- $pp \rightarrow HH \rightarrow b\bar{b}, \ell\bar{\ell}, \nu\bar{\nu}$



Sig



Irreducible Bkg

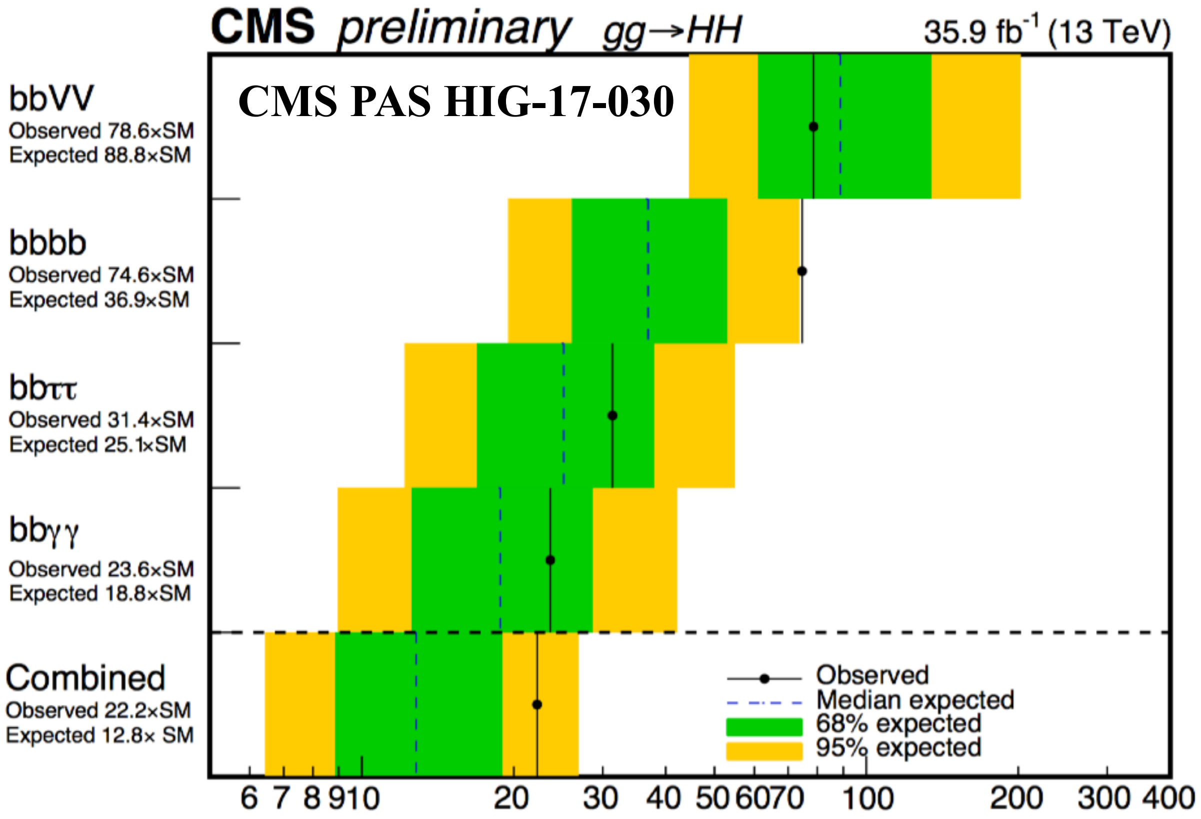
LHC Run 2 result

- Current status from various channels

$h \to XX$

$XX \leftarrow h$

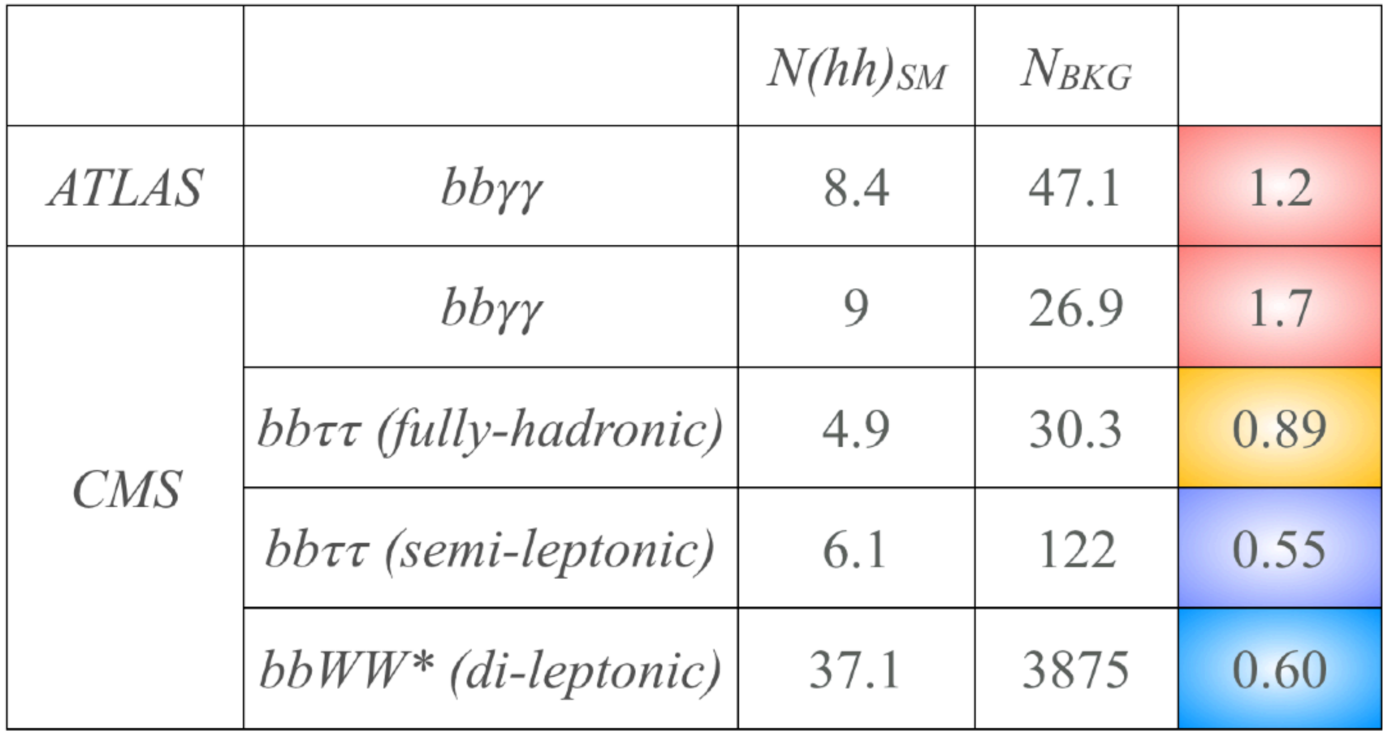
	bb	WW^*	$\tau\tau$	ZZ^*	$\gamma\gamma$
bb	33%				
WW^*	25%	4.6%			
$\tau\tau$	7.3%	2.7%	0.39%		
ZZ^*	3.1%	1.1%	0.33%	0.069%	
$\gamma\gamma$	0.26%	0.1%	0.028%	0.012%	0.0005%



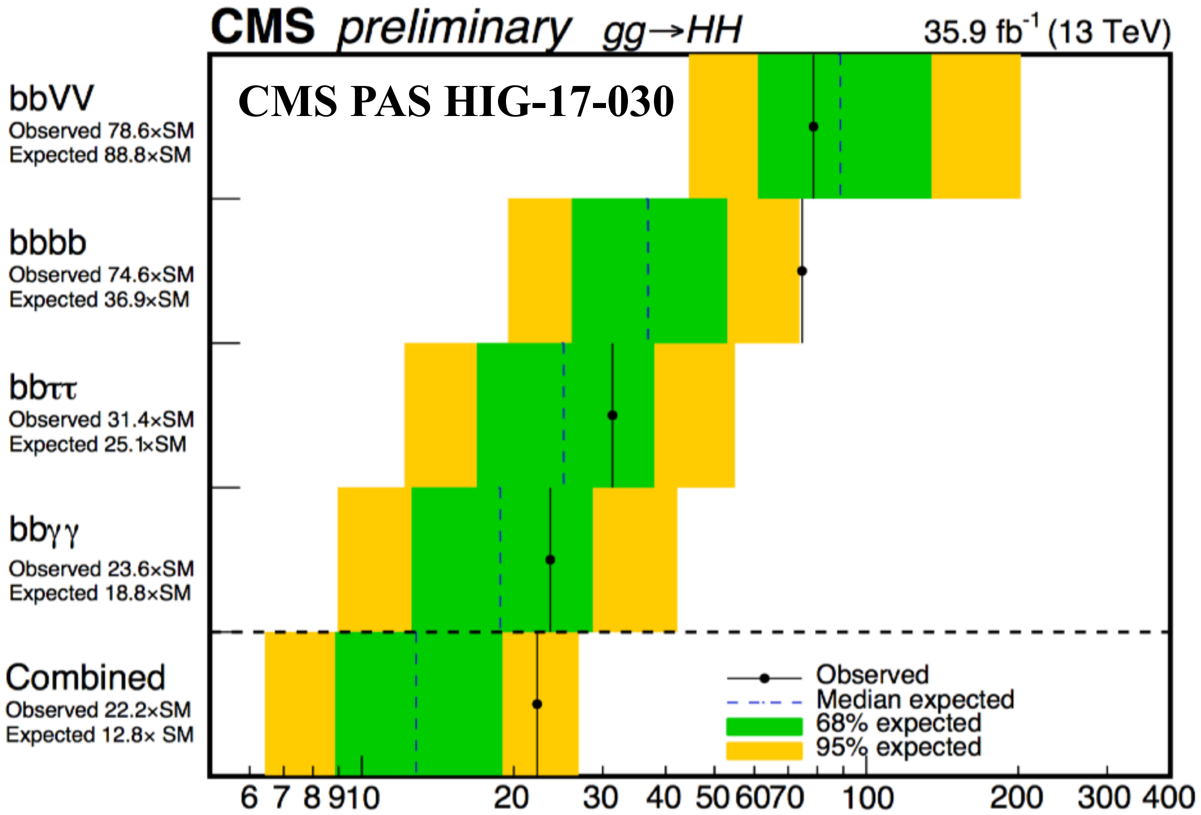
- The driven channel is the "compromised" clean channel.

LHC Run 2 result

- Current status from various channels



		$N(hh)_{SM}$	N_{BKG}	
ATLAS	$bb\gamma\gamma$	8.4	47.1	1.2
CMS	$bb\gamma\gamma$	9	26.9	1.7
	$bb\tau\tau$ *(fully-hadronic)*	4.9	30.3	0.89
	$bb\tau\tau$ *(semi-leptonic)*	6.1	122	0.55
	$bbWW^*$ *(di-leptonic)*	37.1	3875	0.60



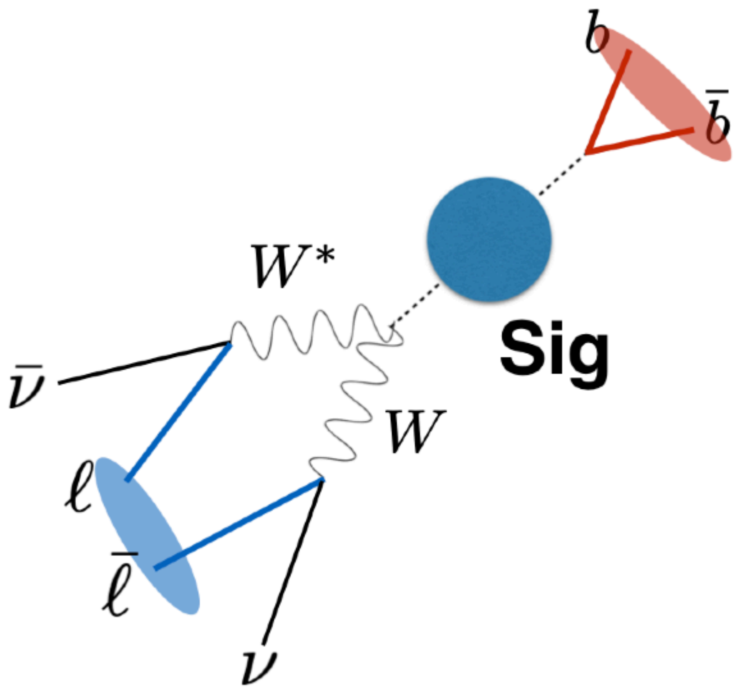
- **Why** is $bbVV$ **bad** ? **how** can one **improve** ?

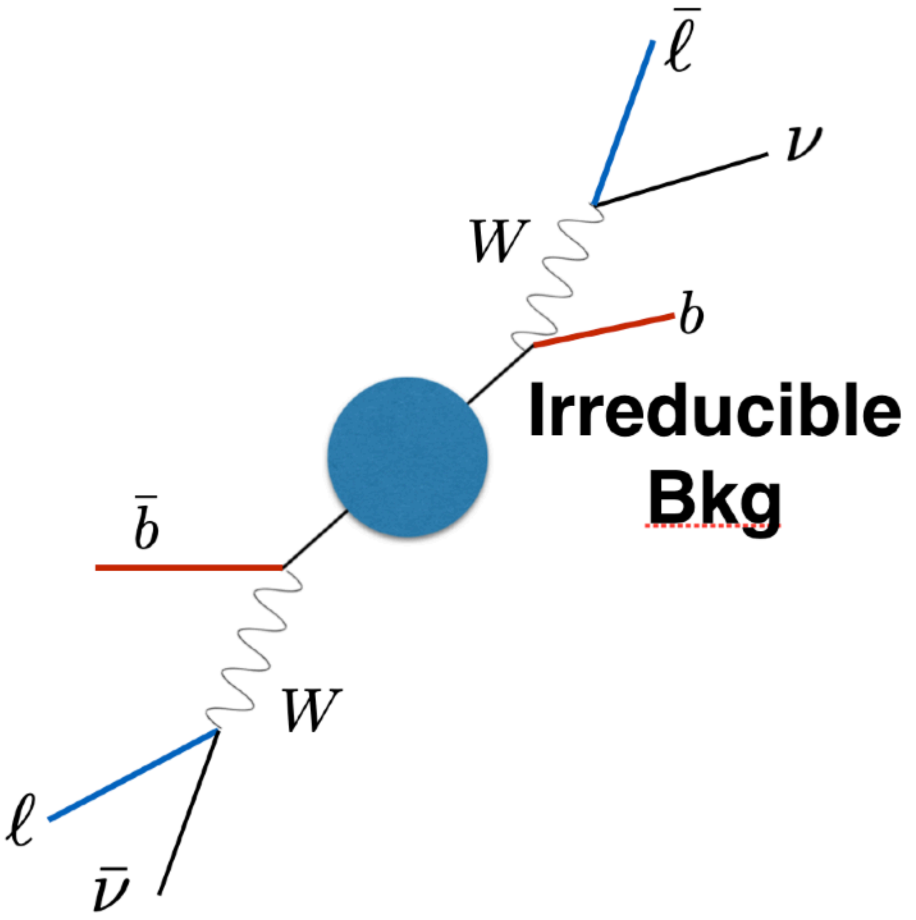
- **Why** is $bbVV$ **bad** ? LHC is the **Top-factory**

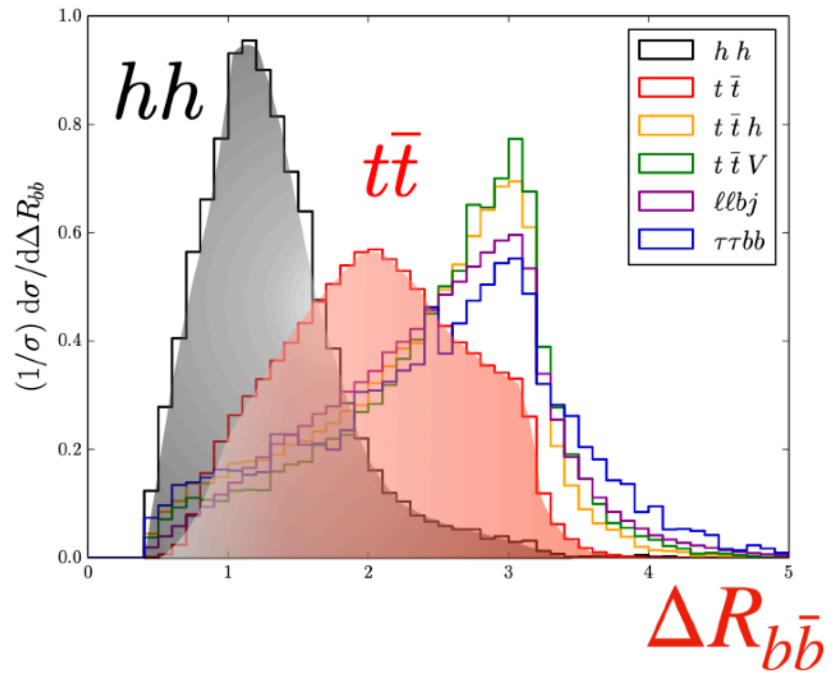
$$\left.\frac{\sigma(pp \to hh \to b\bar{b}VV^*)}{\sigma(pp \to t\bar{t} \to b\bar{b}VV)}\right|_{13\mathrm{TeV}} \simeq \frac{31\mathrm{fb}*(25\%)}{215\mathrm{pb}} \simeq \mathcal{O}(10^{-5})$$

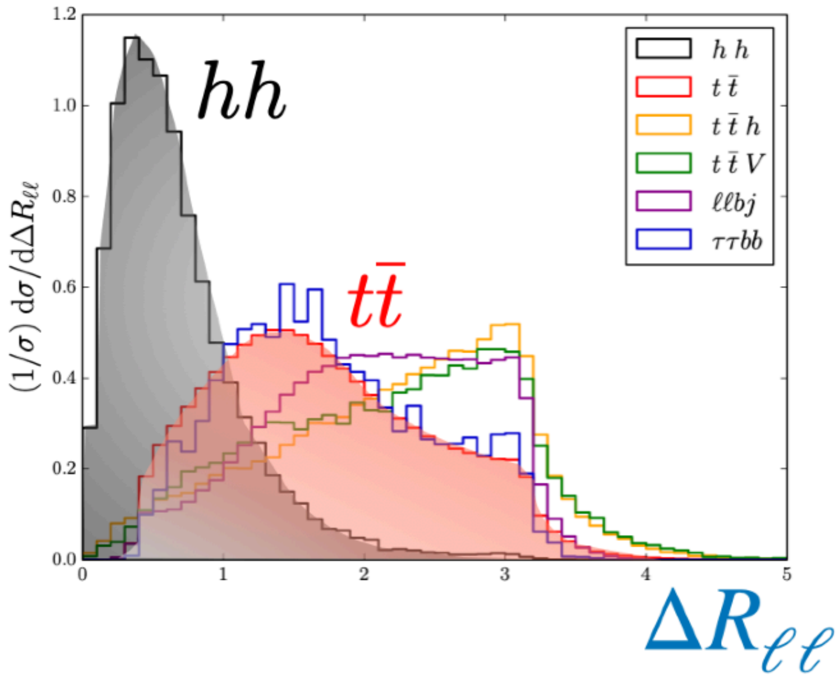
Conventional method to design cuts

- From patterns of signal events



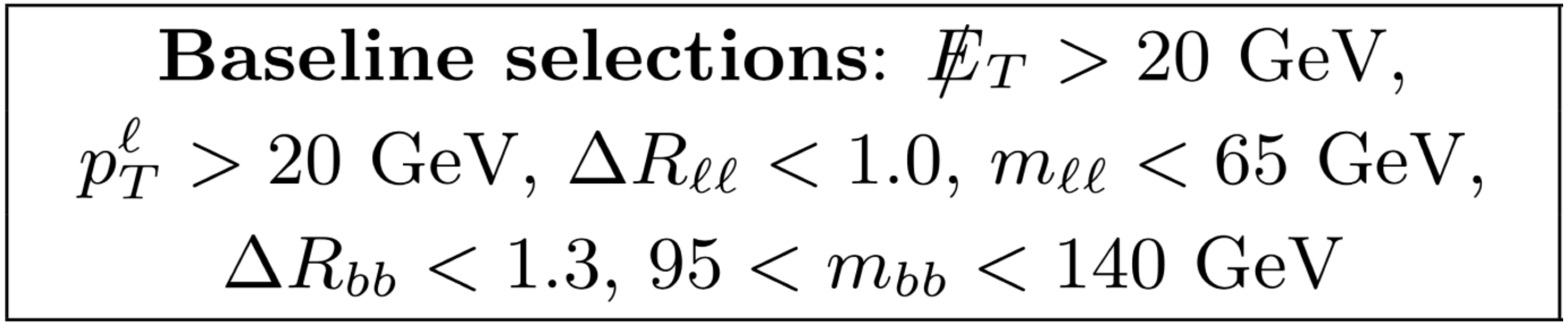




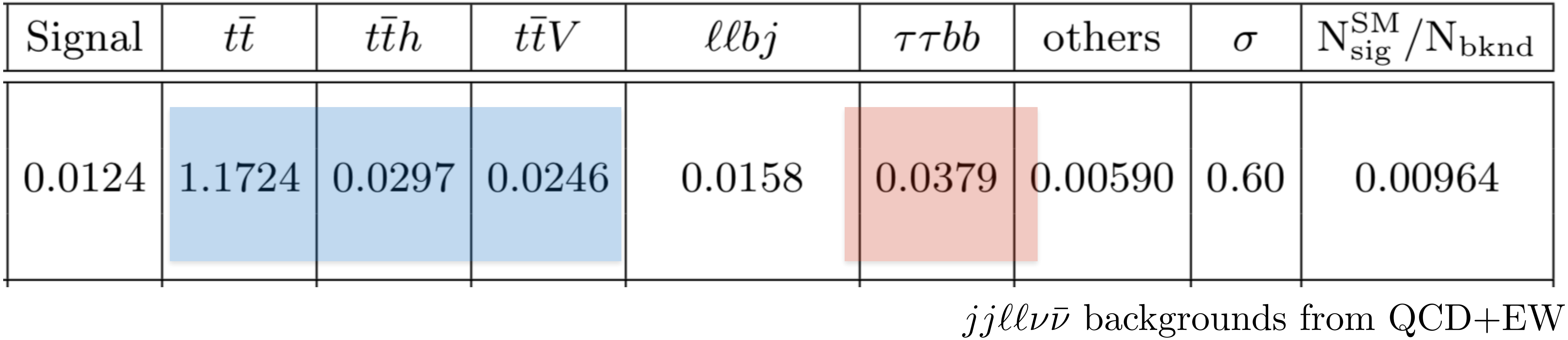


Applying **featured variables** - traditional **ABCD** method

- Applying "**low-level** " kinematic cuts based on event-topology



Baseline selections: $\not{E}_T > 20$ GeV, $p_T^\ell > 20$ GeV, $\Delta R_{\ell\ell} < 1.0$, $m_{\ell\ell} < 65$ GeV, $\Delta R_{bb} < 1.3$, $95 < m_{bb} < 140$ GeV



Signal	$t\bar{t}$	$t\bar{t}h$	$t\bar{t}V$	$\ell\ell bj$	$\tau\tau bb$	others	σ	$\mathrm{N}_{\mathrm{sig}}^{\mathrm{SM}}/\mathrm{N}_{\mathrm{bknd}}$
0.0124	1.1724	0.0297	0.0246	0.0158	0.0379	0.00590	0.60	0.00964

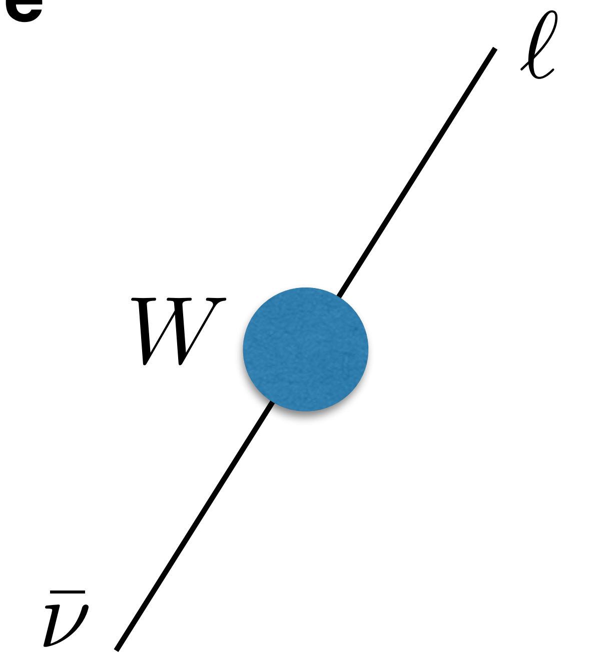
$jj\ell\ell\nu\bar{\nu}$ backgrounds from QCD+EW

- We may apply the advanced statistical tools to see correlations among "low-level" kinematic variables.

- But the **efficiency based on "low-level cuts" is NOT GOOD**

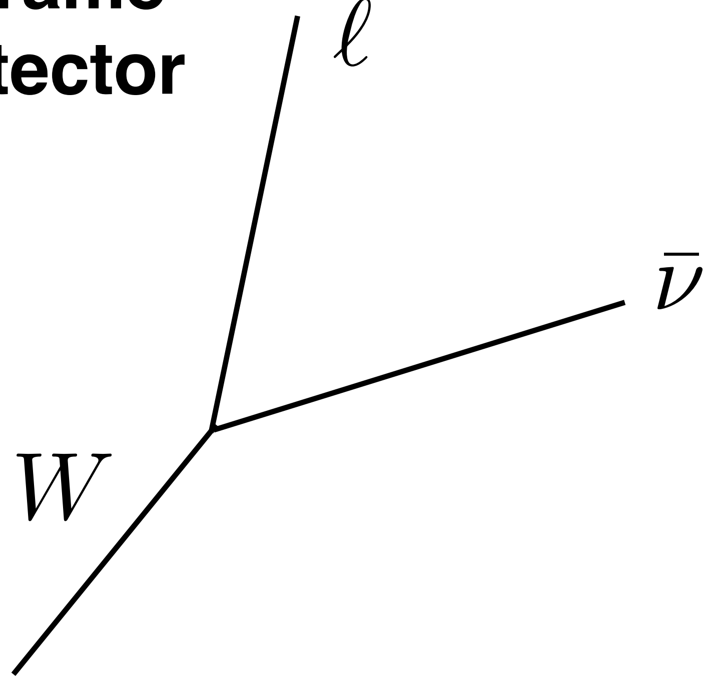
- A **low-level** variable contains various information

Rest frame of W

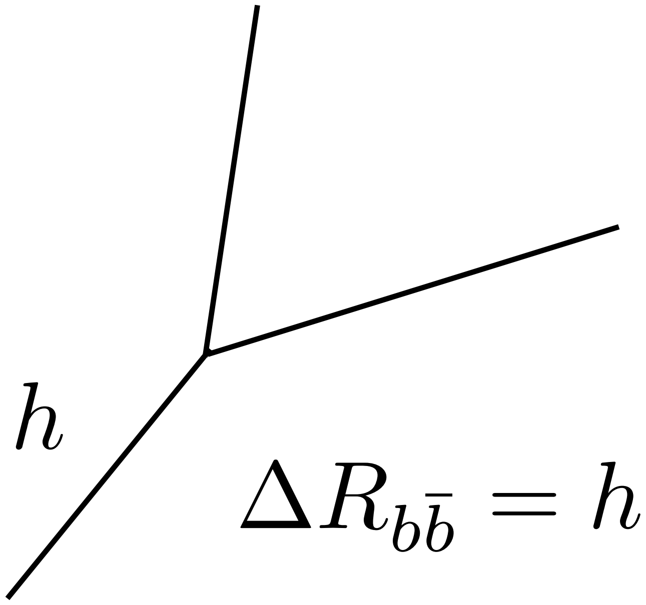


$$P_{t(\ell)} = f(M_W, M_\nu, M_\ell)$$

Lab frame of Detector

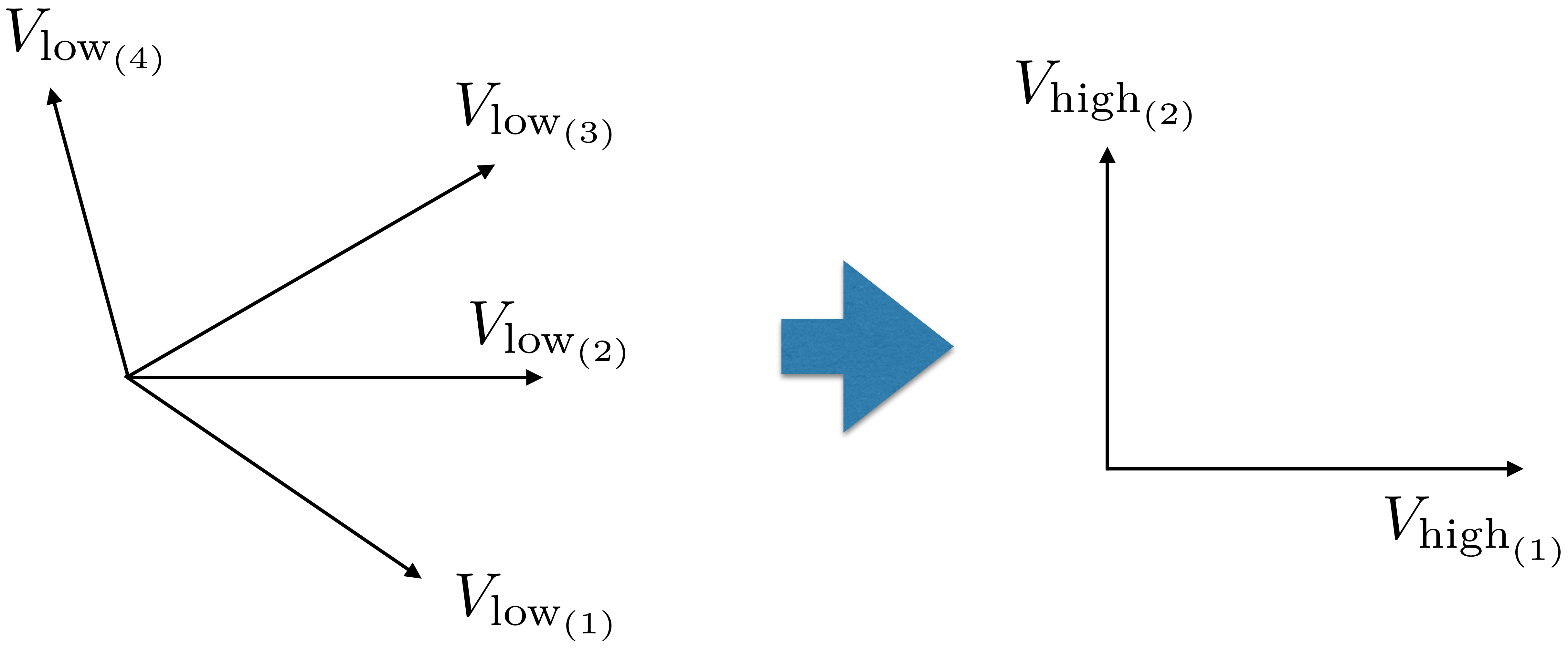


$$P_{t(\ell)} = g(M_W, M_\nu, M_\ell, \eta_W)$$



$$\Delta R_{b\bar{b}} = h(M_h, M_b, \eta_h)\ ,\ \eta_h = h'(\sqrt{\hat{s}}, M_h)$$

- We need to "**reduce dimensions**" by finding "**mutually orthogonal variables**" to maximize sensitivity.

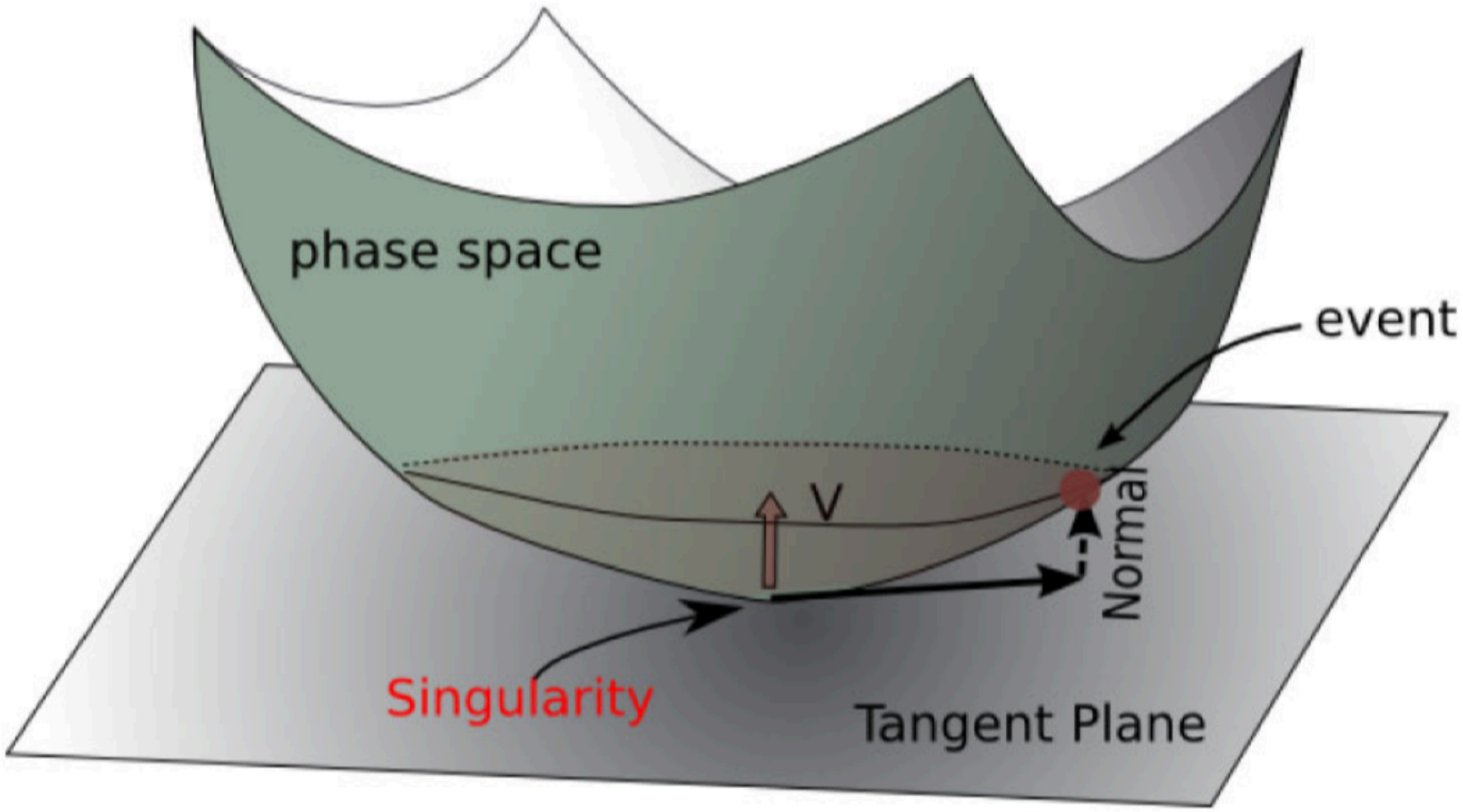


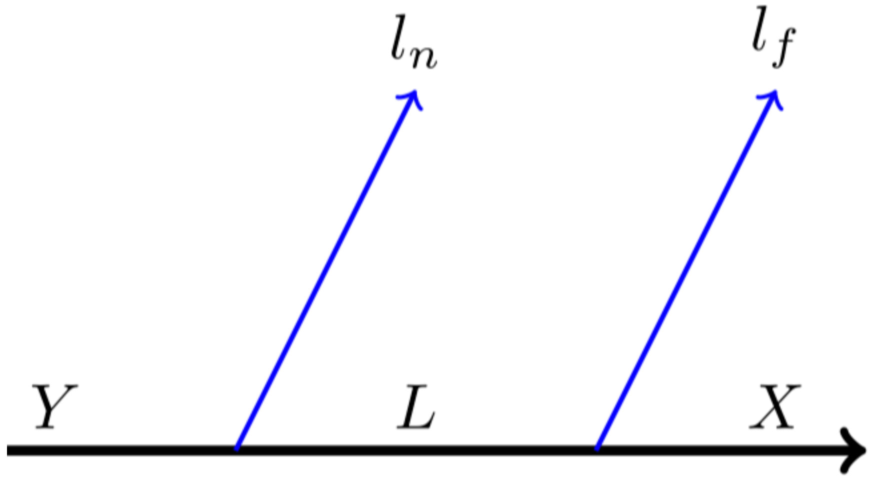
example of 2-dim

- Considering "**featured (high-level)**" kinematic cuts based on event-topology

- What are the "**featured**" kinematic variables?

- **Represent** Phase-space / Physics very well (**singular behavior**)





invariant mass

Ian-Woo Kim, 2010 PRL

For ttbar Background: A mass variable

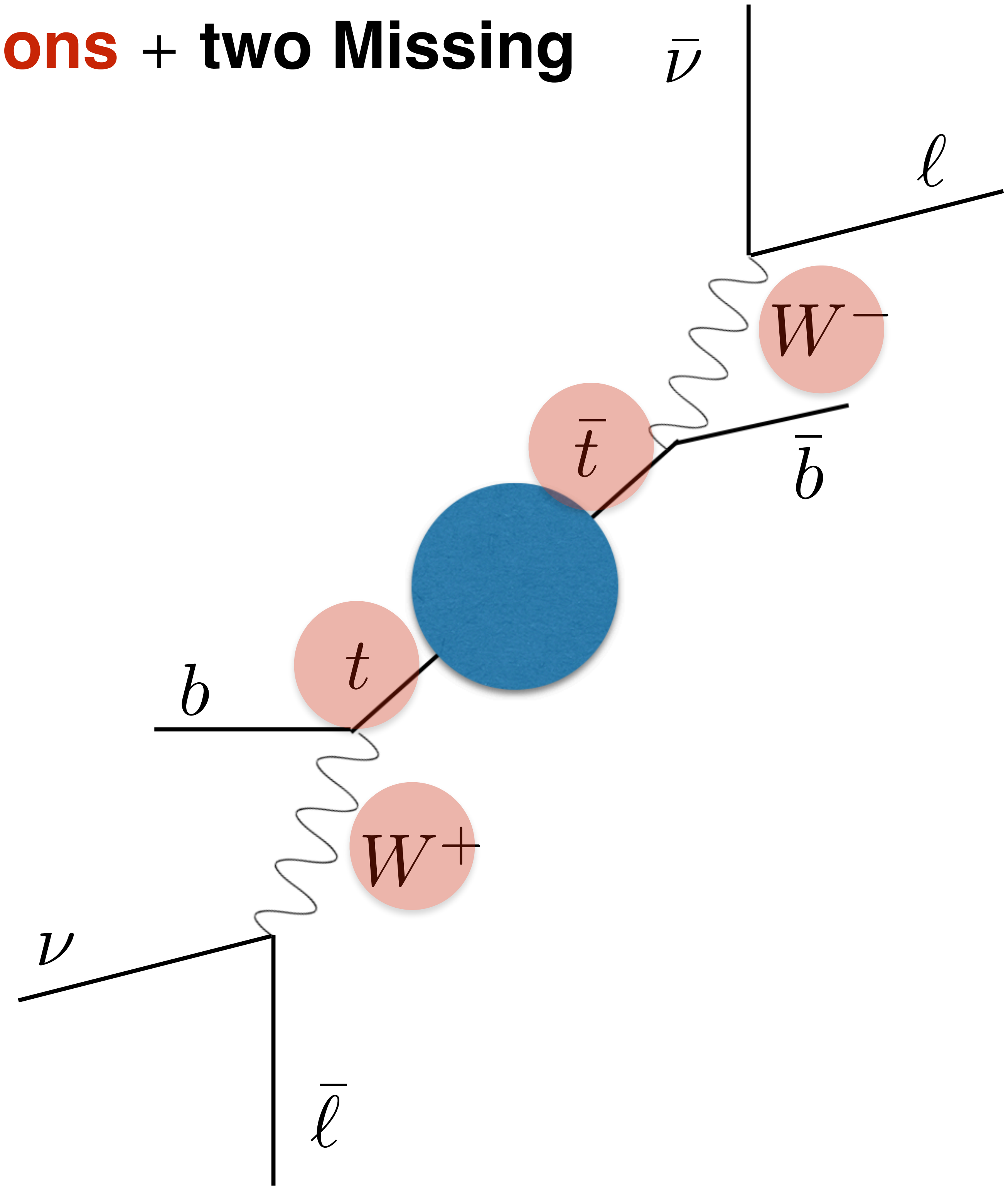
- We have **six unknowns** for two neutrino momentums.
- We have **four mass-shell conditions** + **two Missing Transverse Energy conditions**

$$(p_{\bar{\nu}} + p_{\ell})^2 = m_W^2$$

$$(p_{\bar{\nu}} + p_{\ell} + p_{\bar{b}})^2 = m_{\bar{t}}^2$$

$$(p_{\nu} + p_{\bar{\ell}})^2 = m_W^2$$

$$(p_{\nu} + p_{\bar{\ell}} + p_b)^2 = m_t^2$$

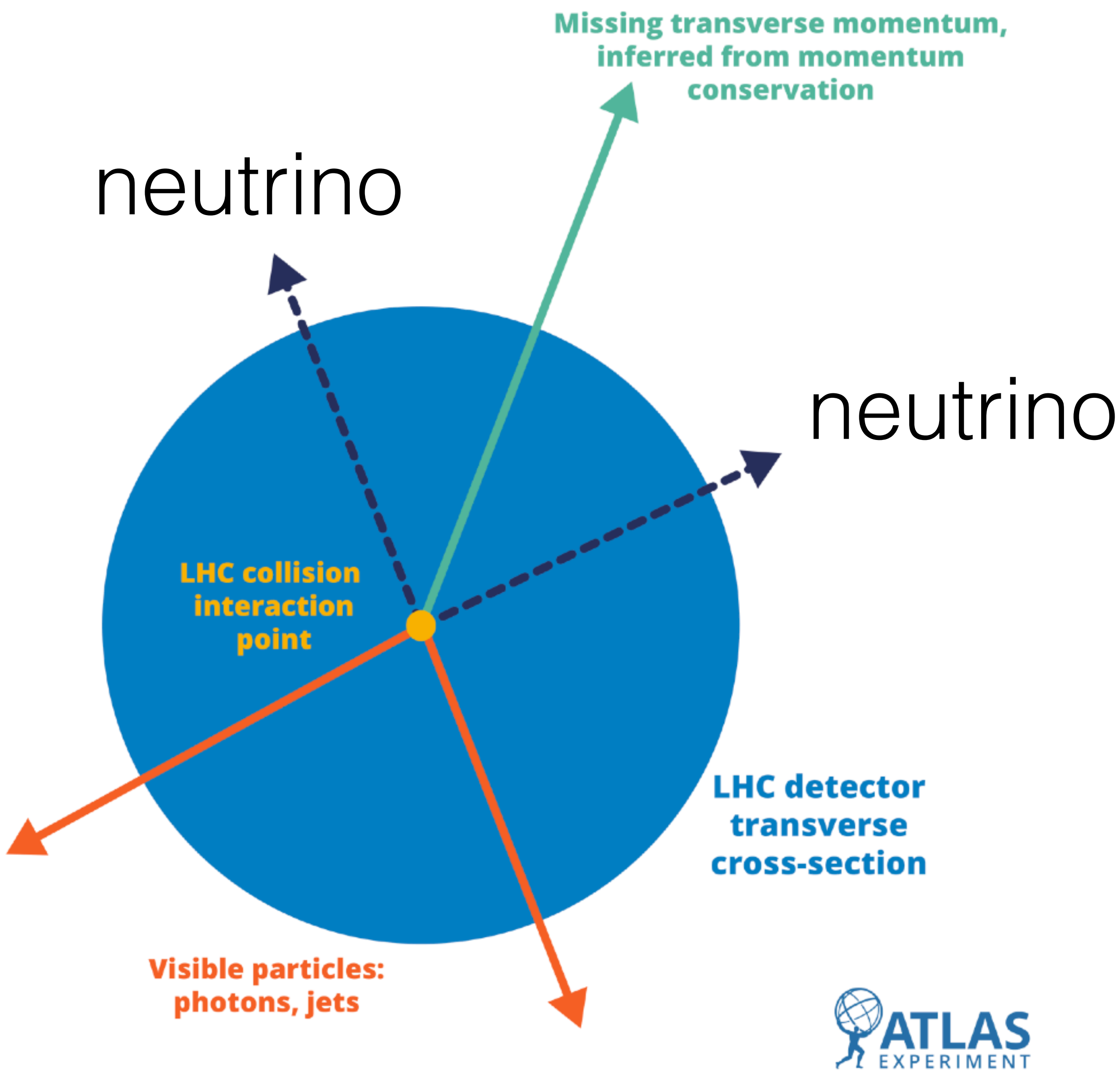


For ttbar Background: A mass variable

- We have **six unknowns** for two neutrino momentums.

- We have **four mass-shell conditions** + **two Missing Transverse Energy conditions**

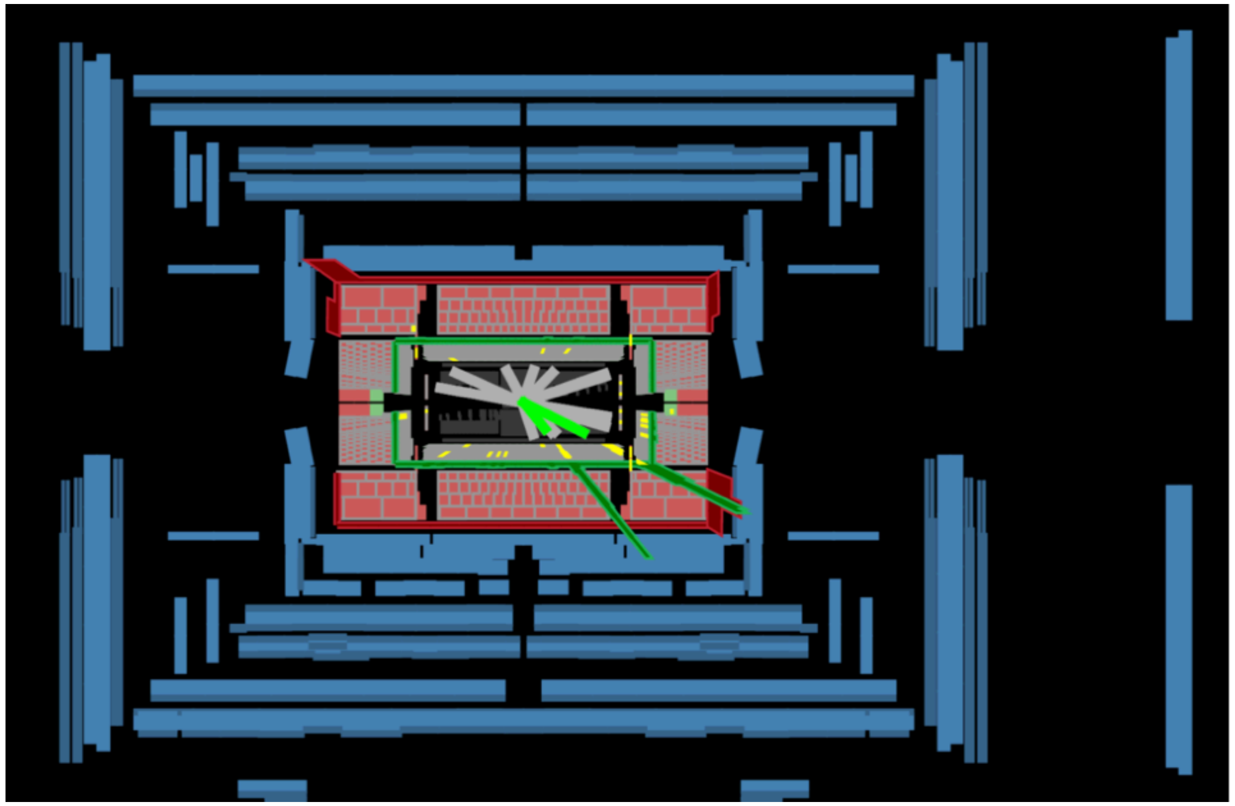
$$\left(\sum_{\text{visible particles}} \vec{P}_T\right) + \left(\sum_{\text{neutrinos}} \vec{P}_T\right) = 0$$



neutrino

neutrino

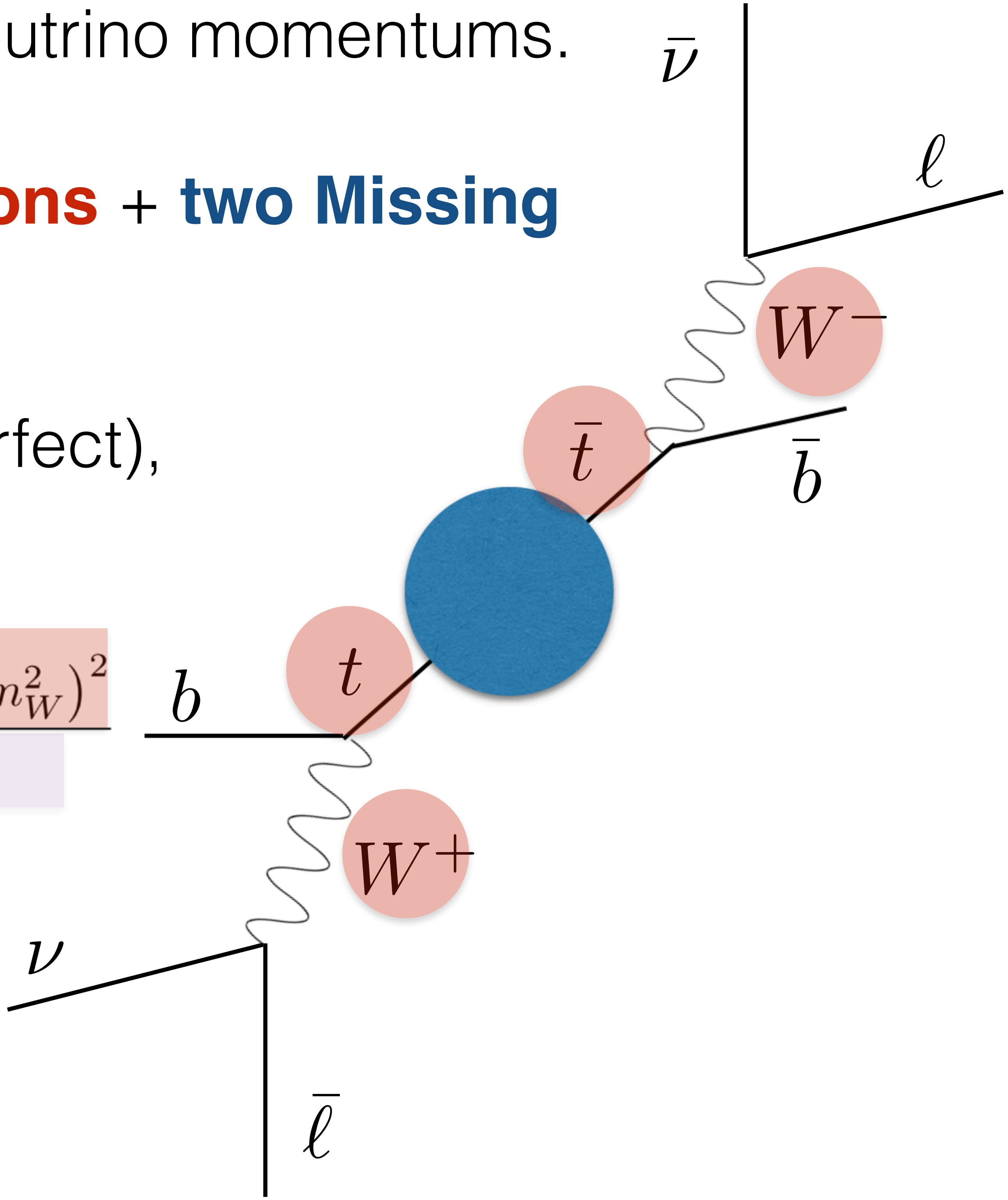
Missing transverse momentum, inferred from momentum conservation
LHC collision interaction point
LHC detector transverse cross-section
Visible particles: photons, jets
ATLAS EXPERIMENT



For ttbar Background: A mass variable

- We have **six unknowns** for two neutrino momentums.
- We have **four mass-shell conditions** + **two Missing Transverse Energy conditions**
- In a reality (as a detector is not perfect), we allow some "**smearing**" effects

$$\chi^2_{ij} \equiv \min_{\vec{\not p}_T = \vec p_{\nu T} + \vec p_{\bar\nu T}} \left[\frac{\left(m^2_{b_i \ell^+ \nu} - m_t^2\right)^2}{\sigma_t^4} + \frac{\left(m^2_{\ell^+\nu} - m_W^2\right)^2}{\sigma_W^4} + \frac{\left(m^2_{b_j \ell^- \bar\nu} - m_t^2\right)^2}{\sigma_t^4} + \frac{\left(m^2_{\ell^-\bar\nu} - m_W^2\right)^2}{\sigma_W^4} \right]$$



Small χ_{ij} (Top-ness) = compatible with a **ttbar event topology**

For HH signal events: Utilizing Mass information

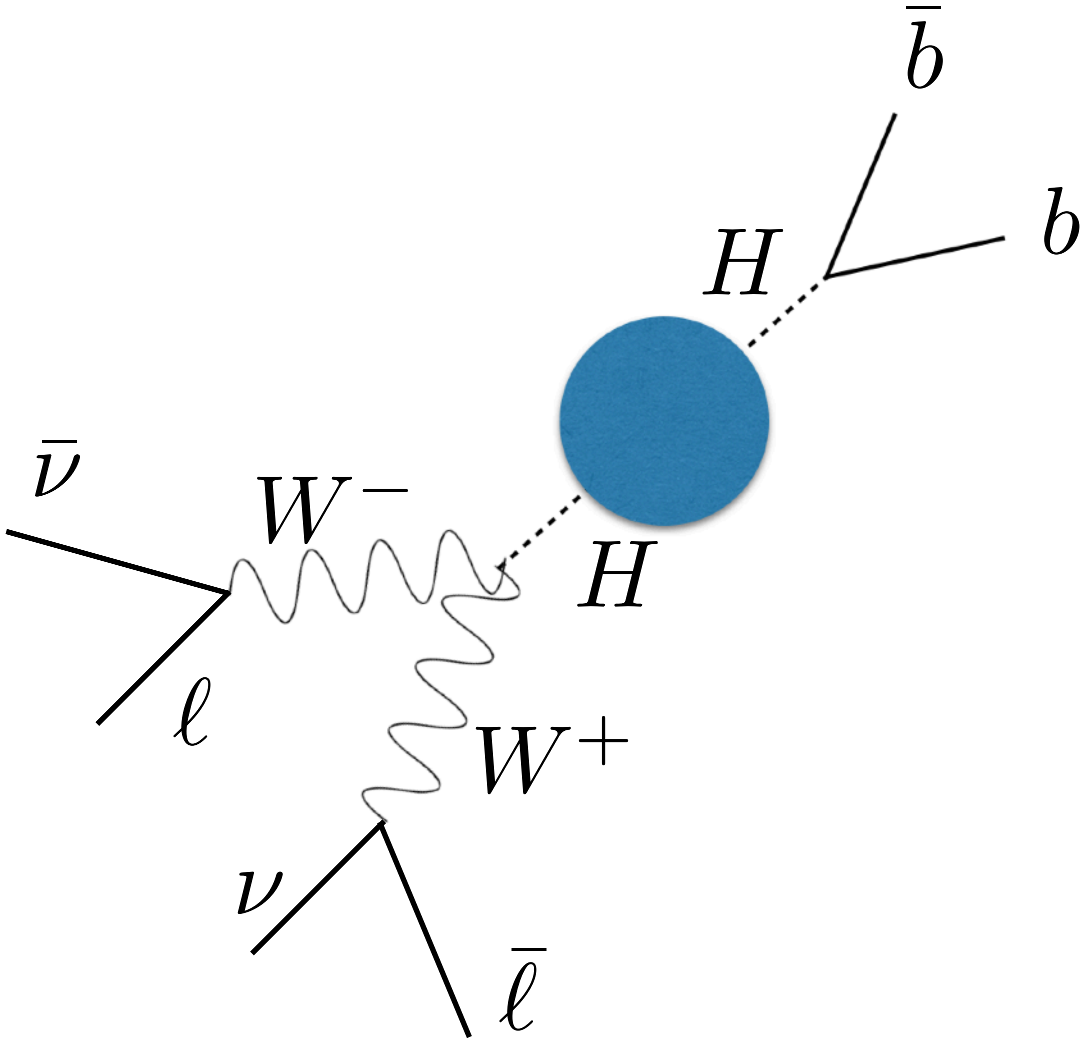
- We have **six unknowns** for two neutrino momentums.

- We have **two mass-shell conditions** + **two "mass" constraints** + **two Missing Transverse Energy conditions**

$$H \equiv \min \Bigg[\frac{\left(m^2_{\ell^+\ell^-\nu\bar{\nu}} - m^2_h\right)^2}{\sigma^4_{h_\ell}} + \frac{\left(m^2_{\nu\bar{\nu}} - m^2_{\nu\bar{\nu},peak}\right)^2}{\sigma^4_\nu}$$

$$+min\left(\frac{\left(m^2_{\ell^+\nu} - m^2_W\right)^2}{\sigma^4_W} + \frac{\left(m^2_{\ell^-\bar{\nu}} - m^2_{W^*,peak}\right)^2}{\sigma^4_{W_*}} \right.$$

$$\left. \frac{\left(m^2_{\ell^-\bar{\nu}} - m^2_W\right)^2}{\sigma^4_W} + \frac{\left(m^2_{\ell^+\nu} - m^2_{W^*,peak}\right)^2}{\sigma^4_{W_*}} \right) \Bigg]$$

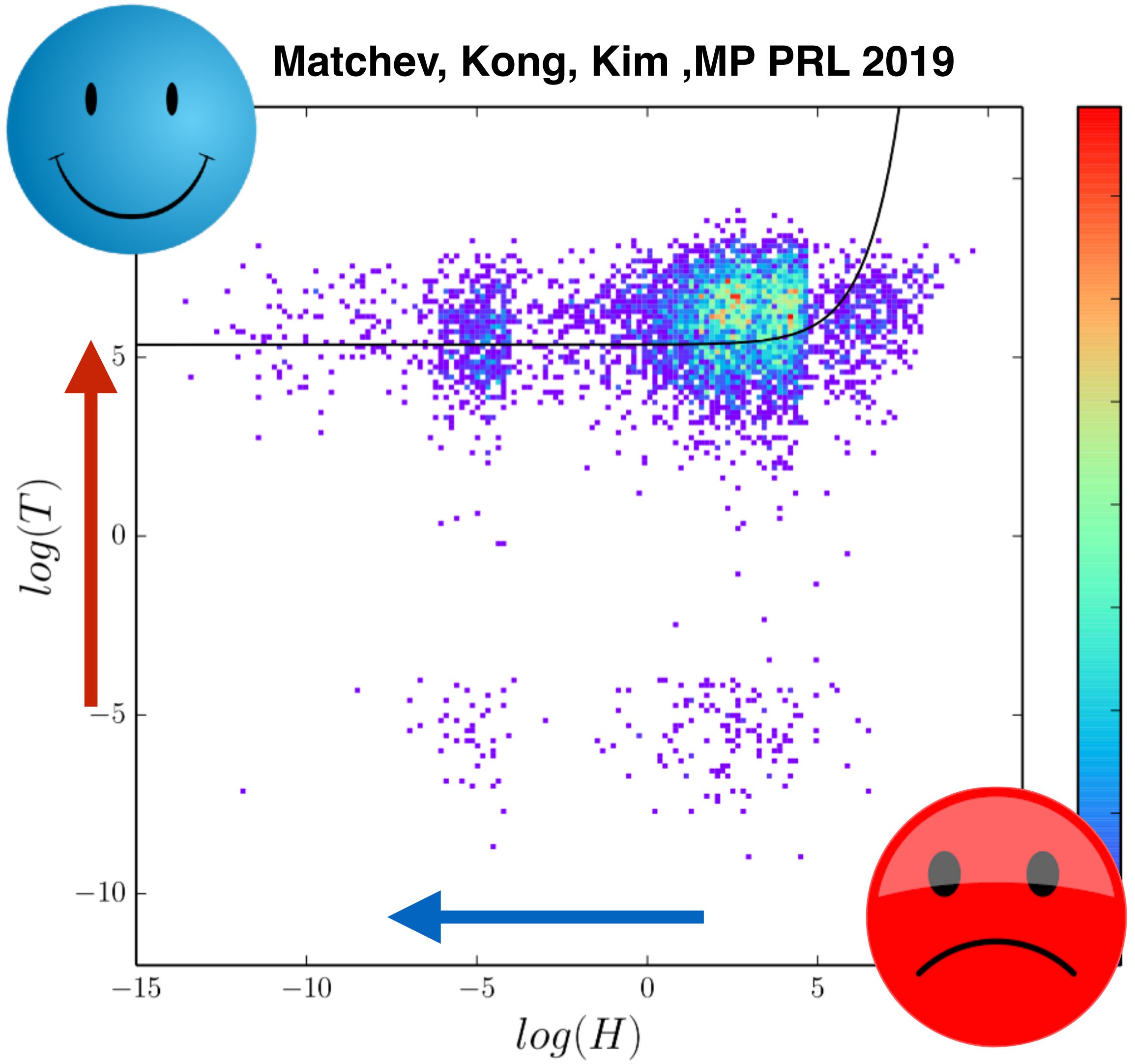


Small H (Higgs-ness) = compatible with a **Higgs event topology**

HH

Small H (Higgs-ness)
compatible with a **Higgs-topology**

Large χ_{ij} (Top-ness)
NOT compatible with a $t\bar{t}$ -**topology**

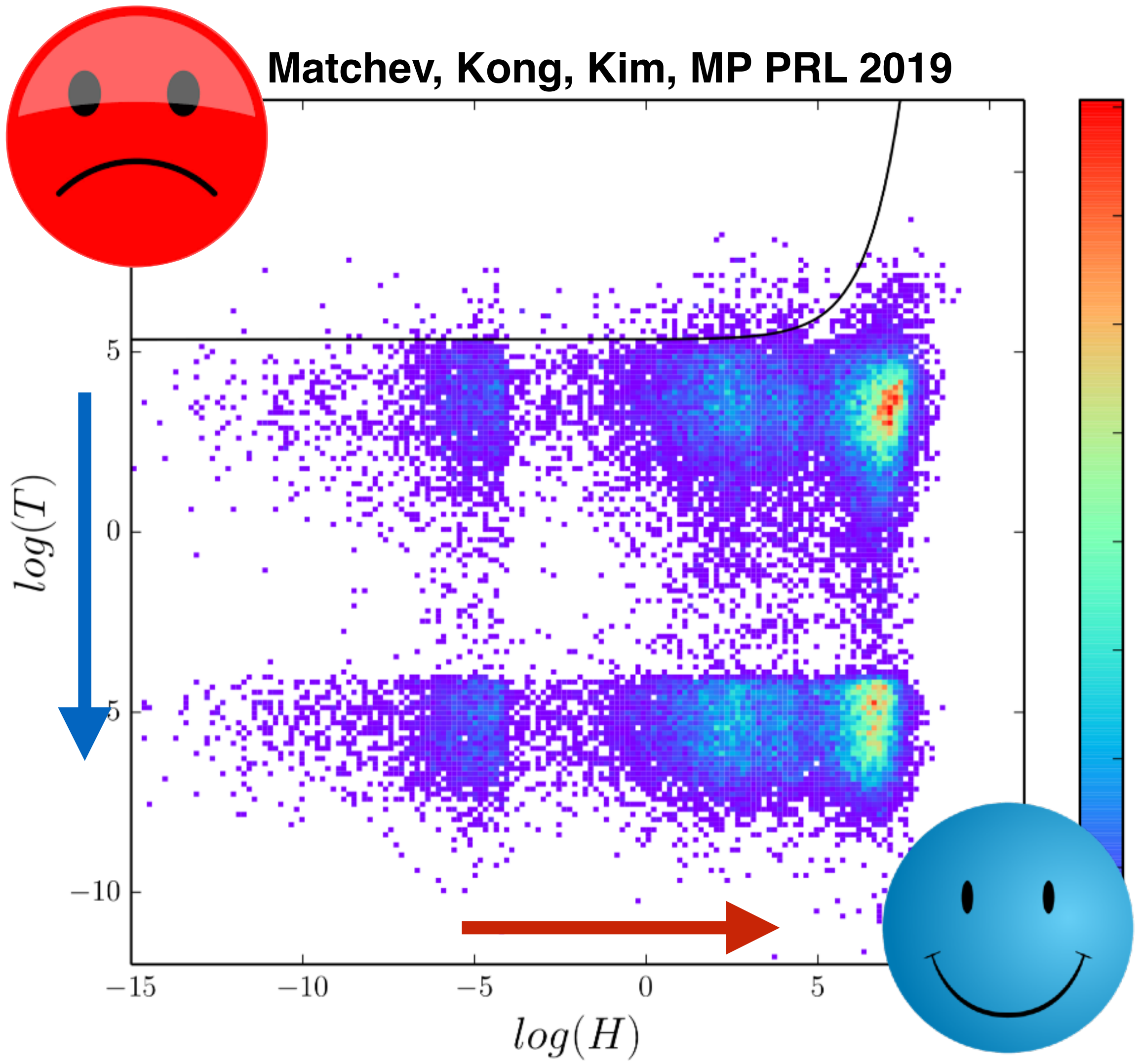


Matchev, Kong, Kim ,MP PRL 2019

$t\bar{t}$

Large H (Higgs-ness)
NOT compatible with a **Higgs-topology**

Small χ_{ij} (Top-ness)
compatible with a $t\bar{t}$ -**topology**

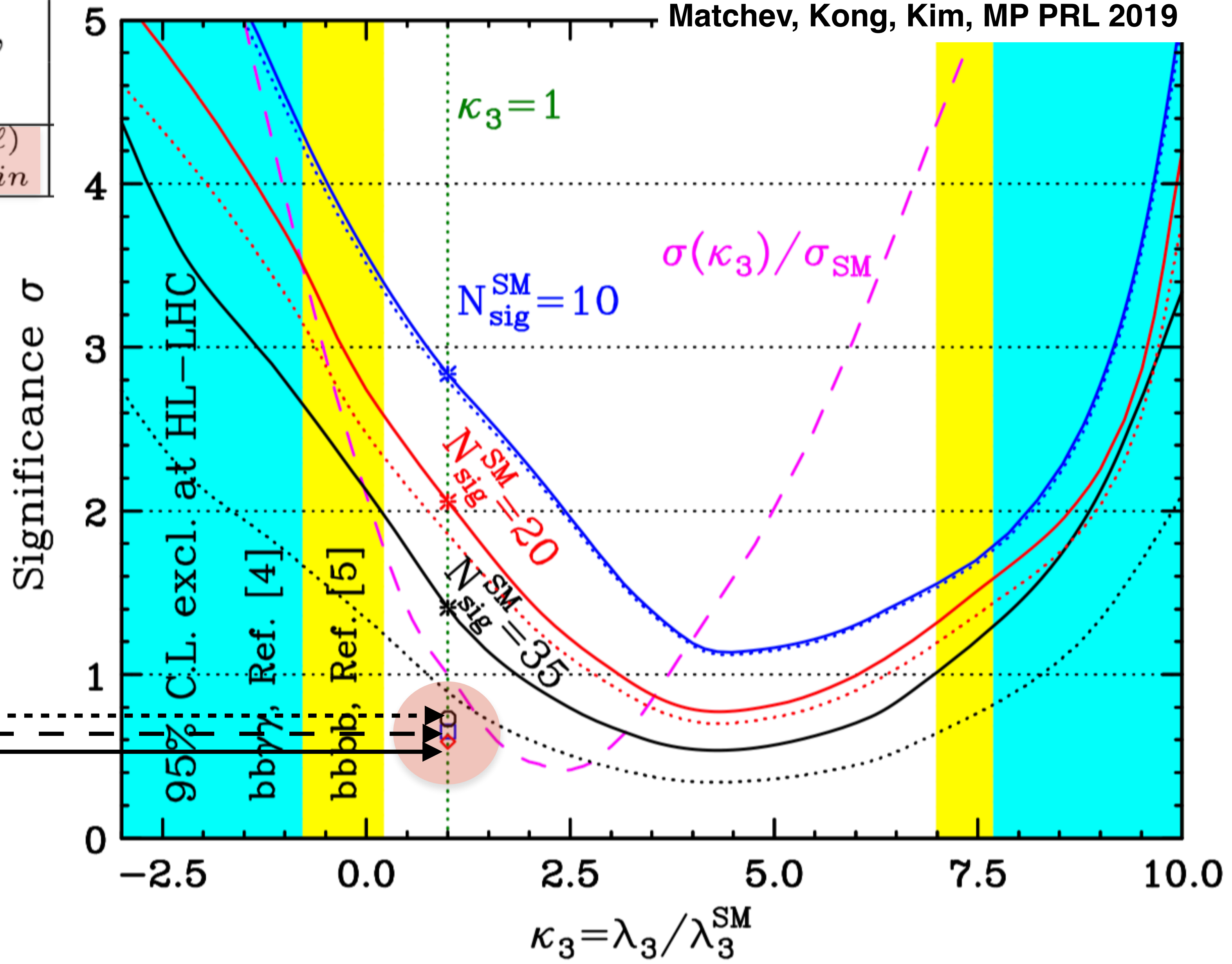


Matchev, Kong, Kim, MP PRL 2019

Baseline selections: $\not{E}_T > 20$ GeV, $p_T^\ell > 20$ GeV, $\Delta R_{\ell\ell} < 1.0$, $m_{\ell\ell} < 65$ GeV, $\Delta R_{bb} < 1.3$, $95 < m_{bb} < 140$ GeV

Higgsness $\oplus$ Topness $\oplus$ $M_{T2}^{(b)} \oplus M_{T2}^{(\ell)} \oplus \sqrt{\hat{s}}_{min}^{(\ell\ell)}$

Orthogonal / singular variables

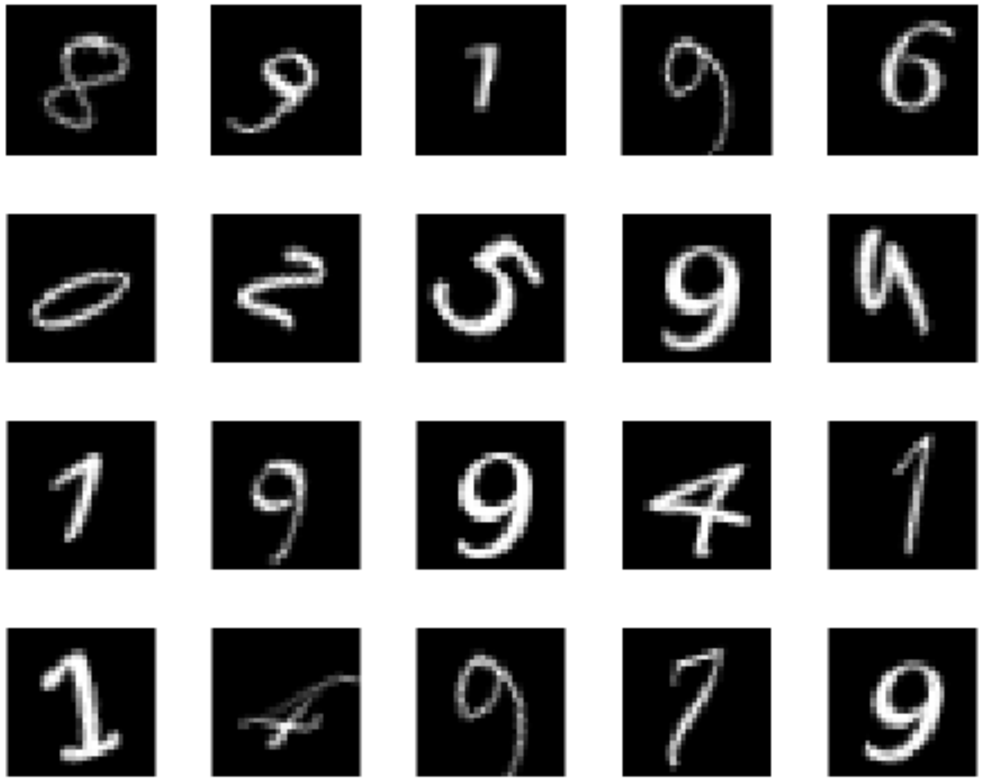

Matchev, Kong, Kim, MP PRL 2019

- CMS-PAS-FTR-15-002, Neural Network (NN) with $\left(p_T, \eta, M_{l\bar{l}}, M_{b\bar{b}}, \Delta R_{l\bar{l}}, \Delta R_{b\bar{b}}, \Delta\phi_{b\bar{b},l\bar{l}}\right)$
- CMS-PAS-HIG-16-024, BDT based on $\left(M_{l\bar{l}}, \Delta R_{l\bar{l}}, \Delta R_{jj}, \Delta\phi_{l\bar{l},jj}, p_T^{l\bar{l}}, p_T^{jj}, min(\Delta R_{j,l}), M_T\right)$
- A. Adhikary et.al (1712.05346) , BDT based on $\left(p_T^l, M_{ll}, M_{bb}, \Delta R_{ll}, \Delta R_{bb}, p_T^{bb}, p_T^{ll}, MET\right)$

- Mathematically well-designed "feature" variables are very strong even in a ABCD (old) way.

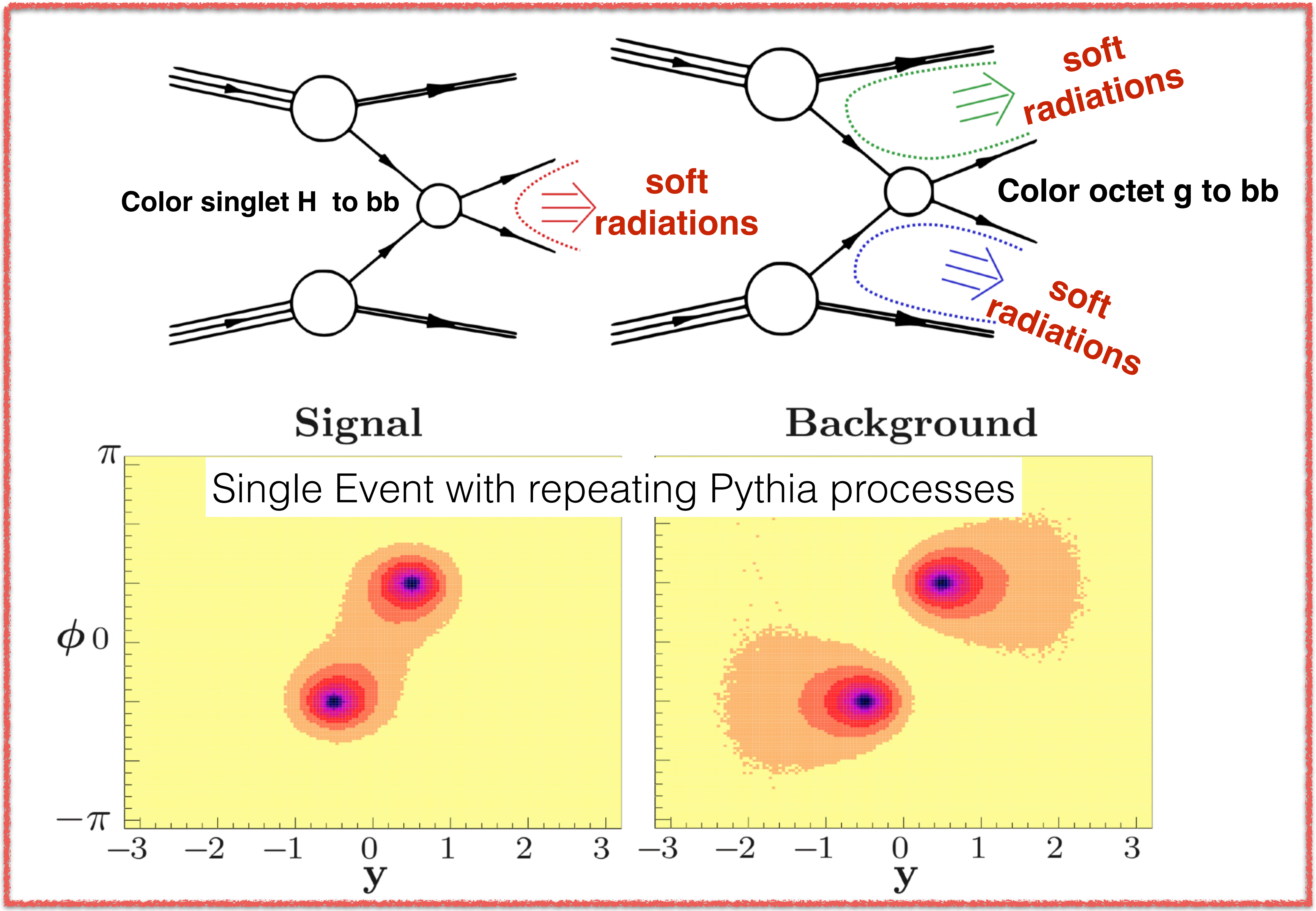
- We can utilize Deep Learning (DL) to maximize correlations among feature variables.

- We can apply DL to utilize "energy deposit patterns" directly (Pattern recognition in images)



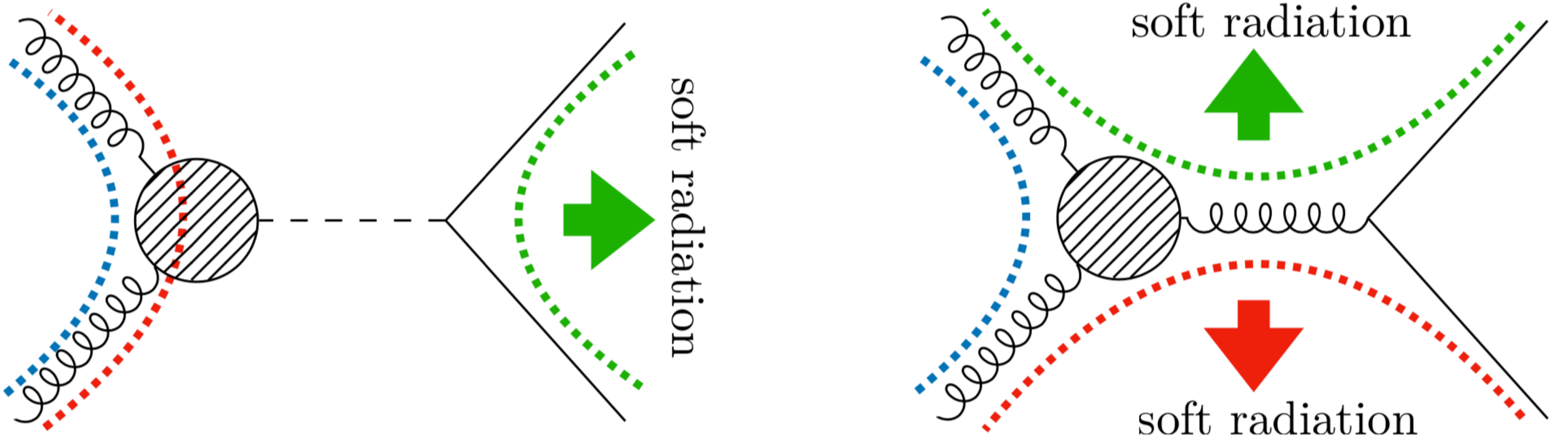
- Consider "**orthogonal**" method to kinematics; **Color-flow**

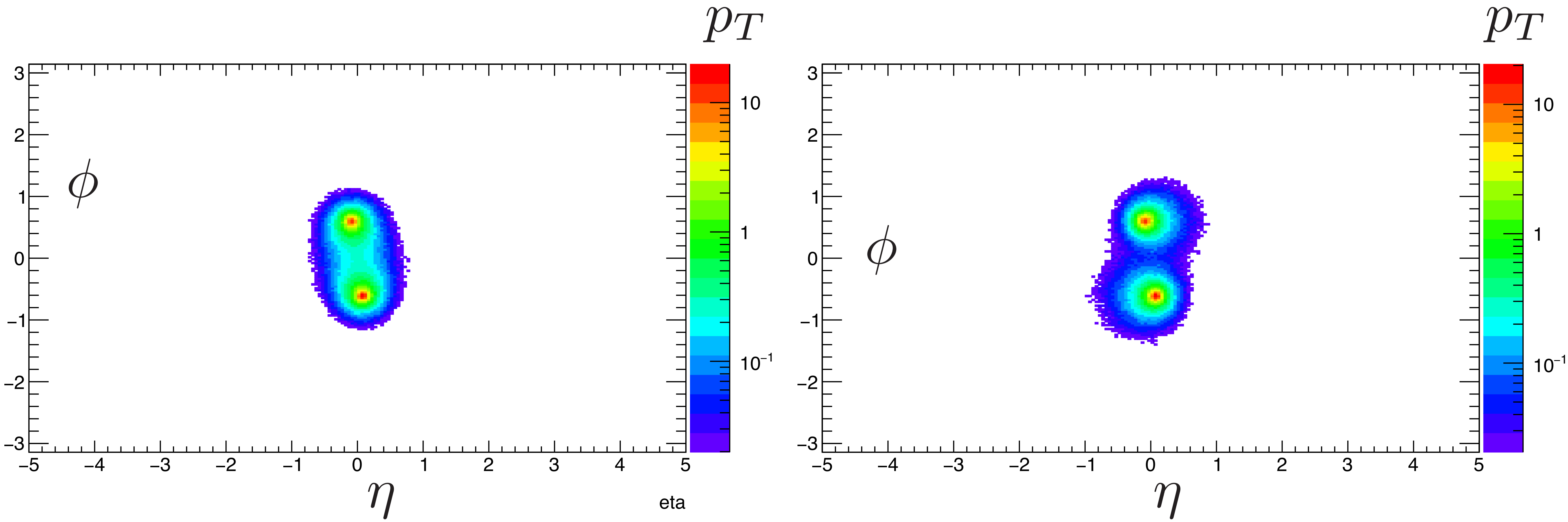
Jason Gallicchio, Matthew D. Schwartz, PRL 105:022001, 2010



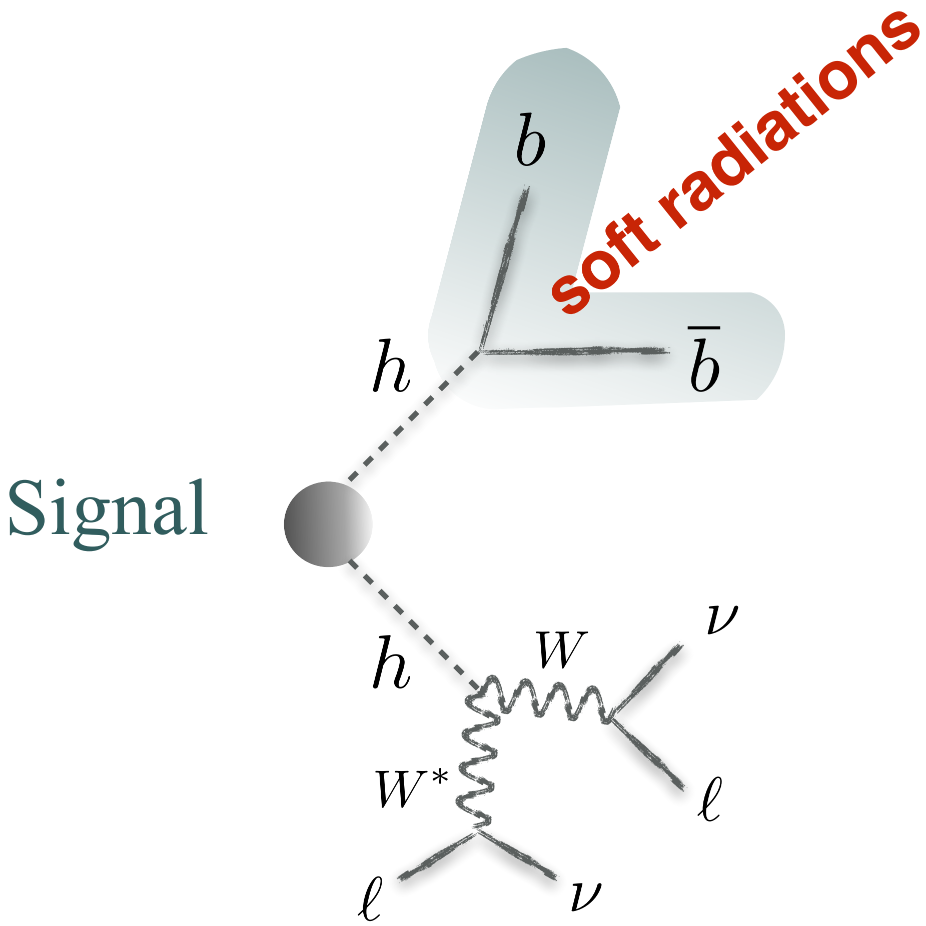
Single Event with repeating Pythia processes

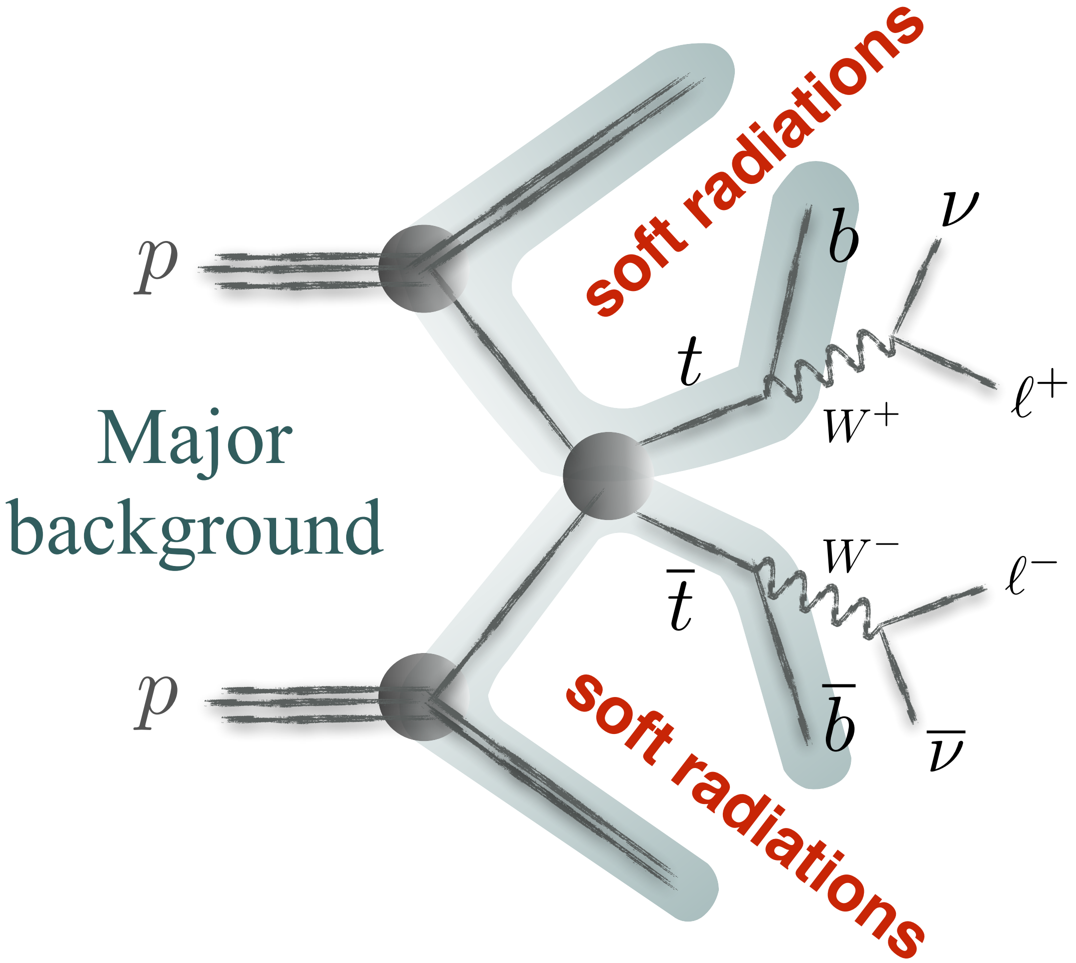
- Consider "**orthogonal**" method to kinematics; **QCD Color-flow**

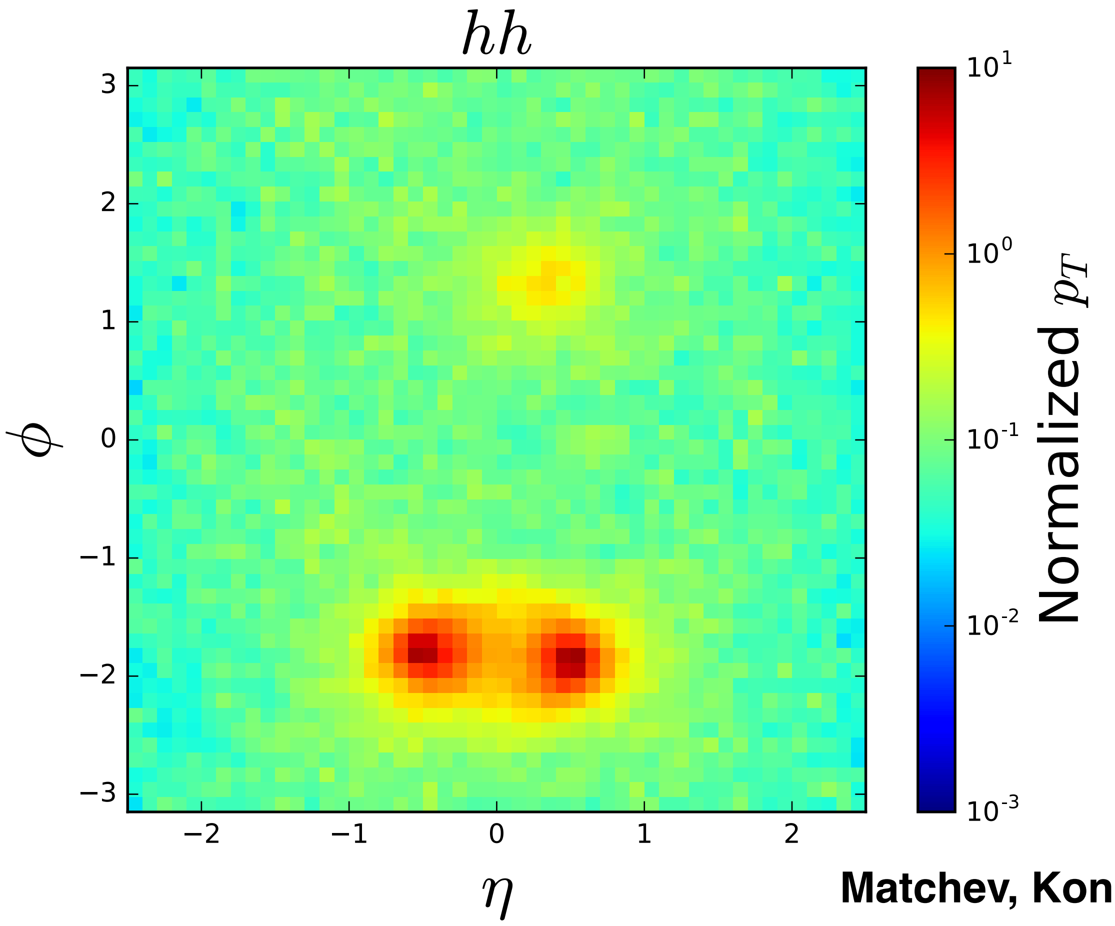


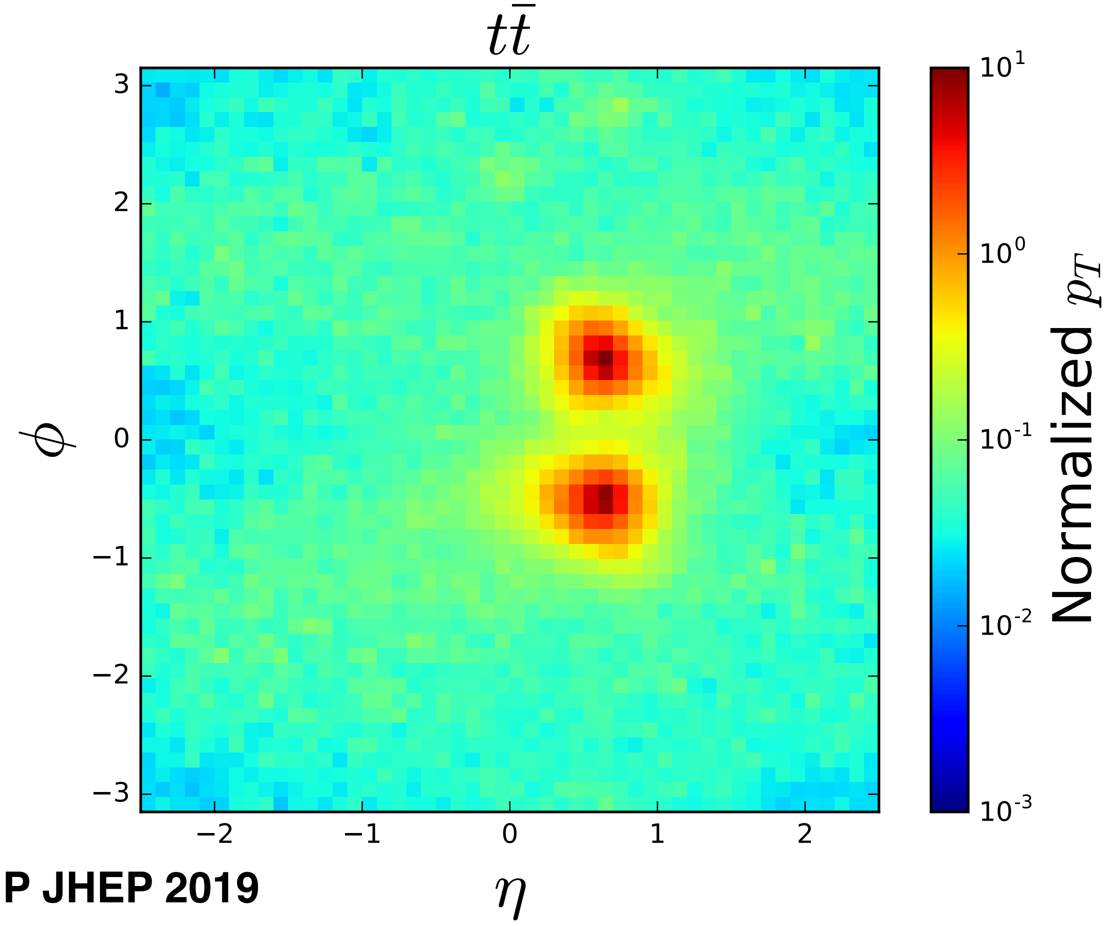


Energy deposits

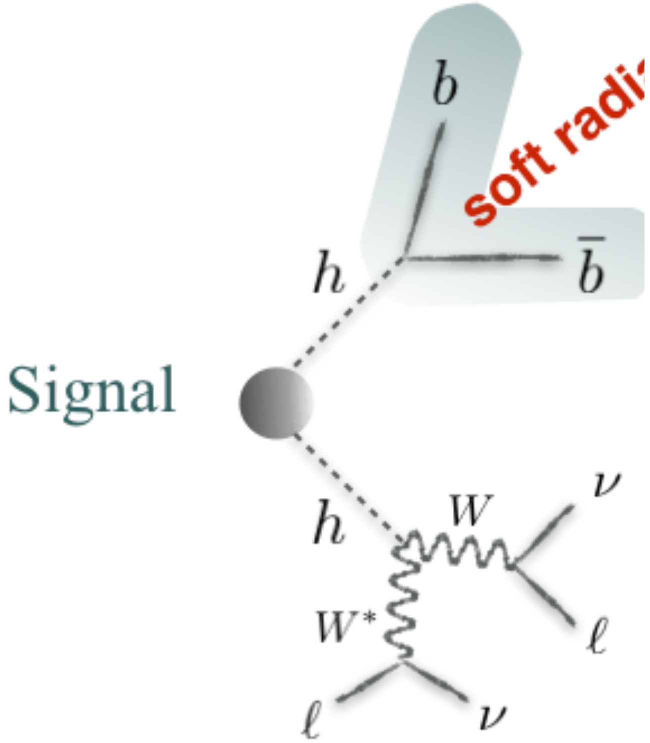


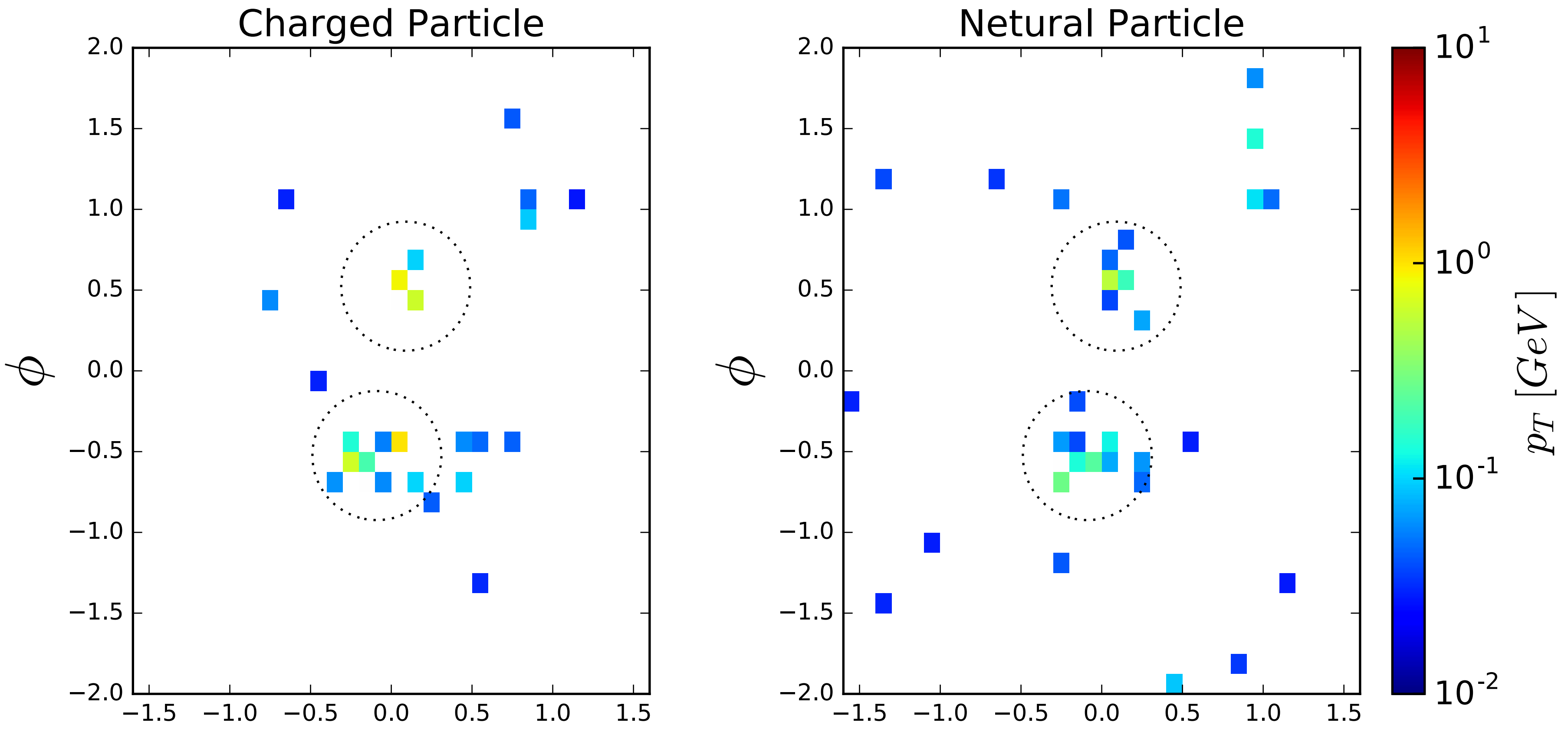




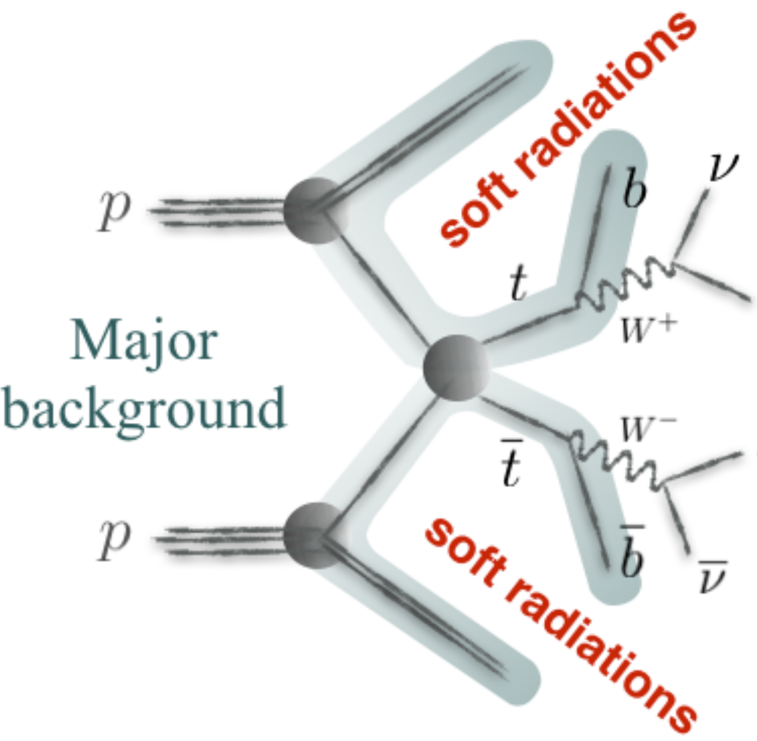


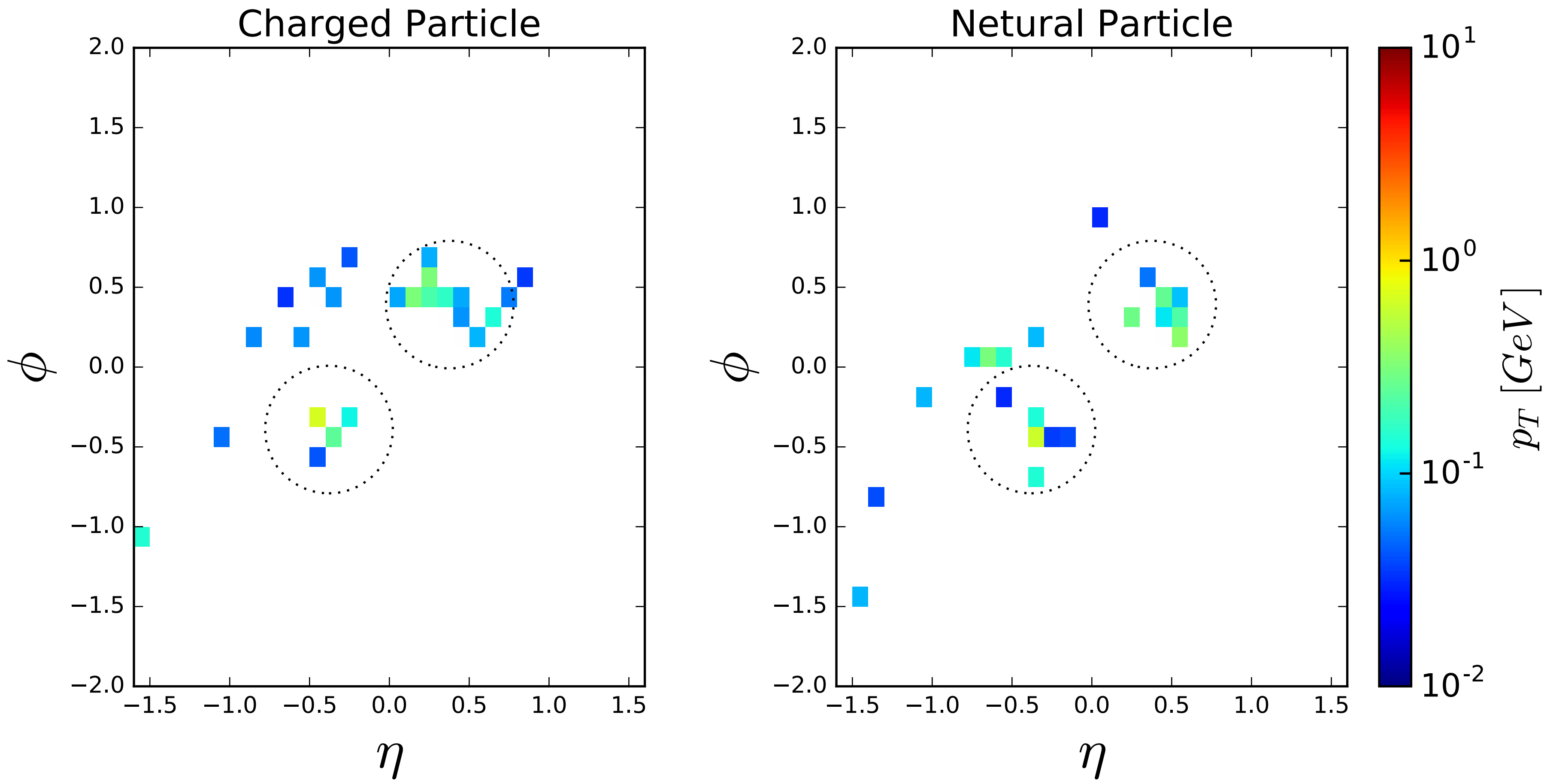
Matchev, Kong, Kims,MP JHEP 2019





Not easy to determine event by event basis

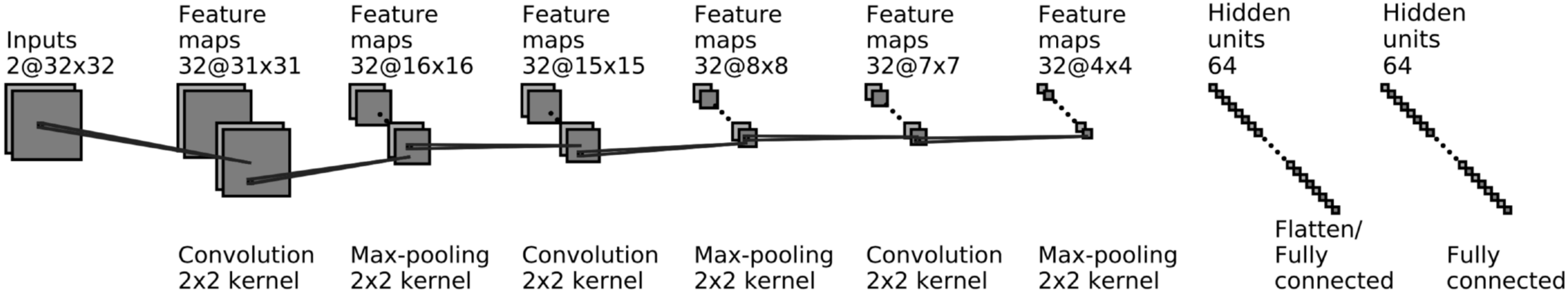


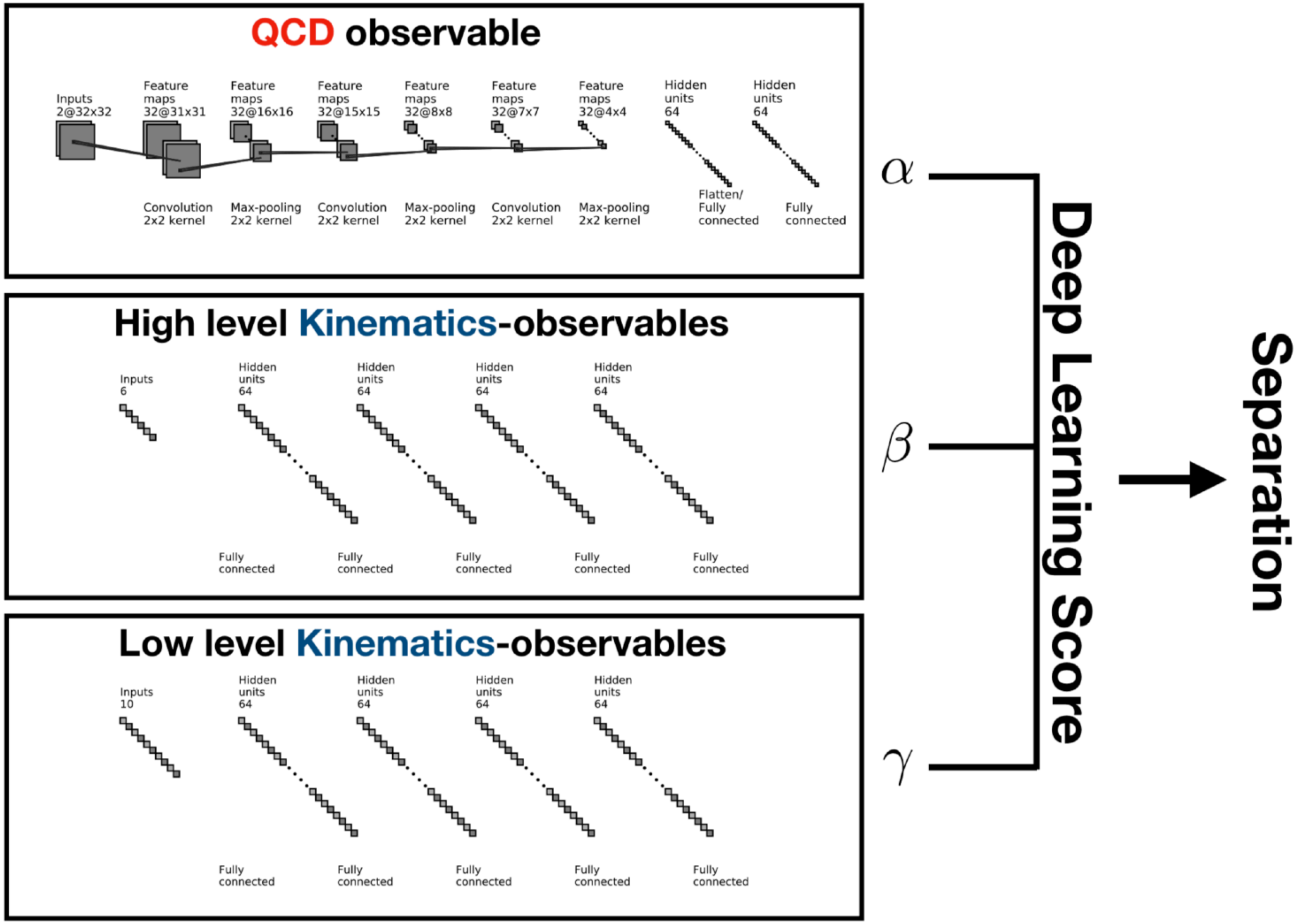


Maximizing information with Deep Neural Network

- For **jet-image**, we use 32x32 pixels for -2.5<eta<2.5, -pi<phi < pi.

  Input channels for **CNN** are divided into two with particle flow:
  - Neutral particles
  - Charged particles



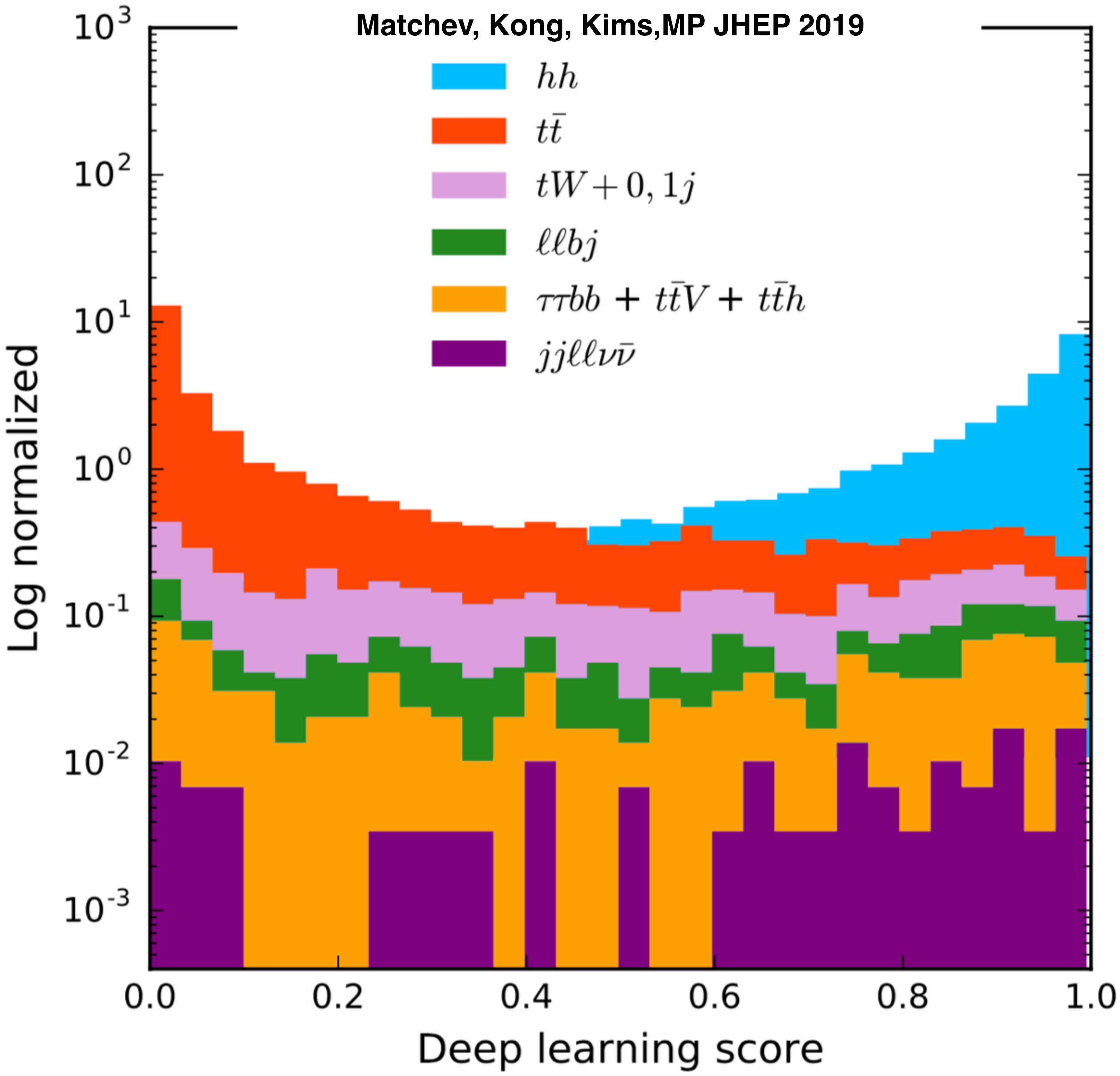


6 High Level Variables

Input data: $\sqrt{\hat{s}_{\text{min}}^{(\bar{b},b,\bar{\ell},\ell)}}$ $\sqrt{\hat{s}_{\text{min}}^{(\bar{\ell},\ell)}}$ M_{T2}^{b} M_{T2}^{ℓ} Higgsness, Topness (feature variables)

10 Low Level Variables

Input data: MET $p^t_{\bar{\ell}}$ p^t_{ℓ} $\Delta R_{(\bar{\ell},\ell)}$ $M_{(\bar{b},b)}$ $p^t_{(\bar{b},b)}$ $\Delta R_{(\bar{b},b)}$ $M_{(\bar{\ell},\ell)}$ $p^t_{(\bar{\ell},\ell)}$, $\Delta\phi_{\{(\bar{\ell},\ell),(\bar{b},b)\}}$

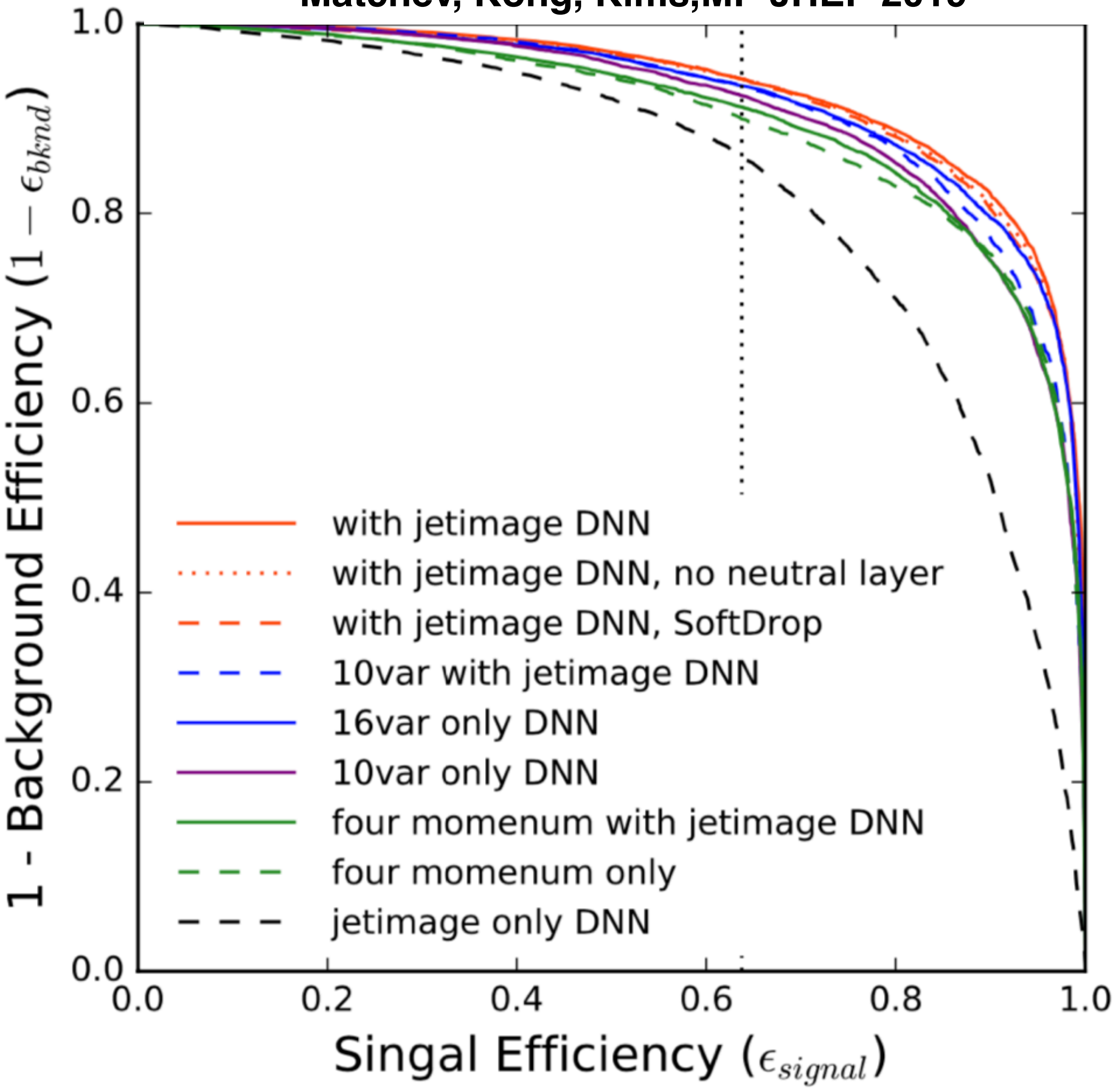


"Backgrounds as stacked Histogram"

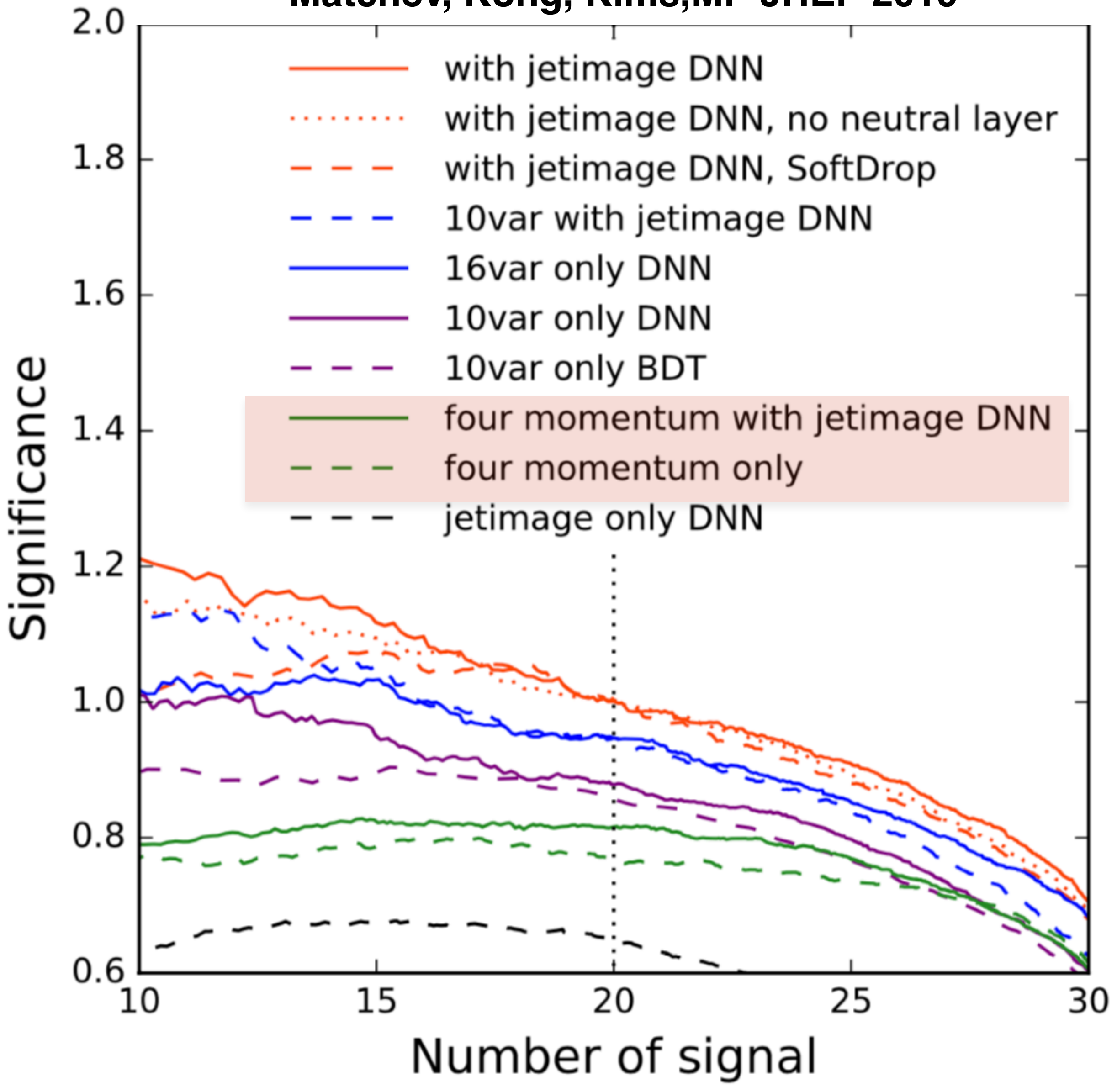
- To estimate effects from **pileup** removal (important in using QCD info),

  0. No additional processes.
  1. we apply SoftDrop to a fat-jet (R= 1.2 anti-Kt)
  2. we use "charged layer only"
  (Various pile-up removers use "longitudinal vertex information through tracking)

Matchev, Kong, Kims,MP JHEP 2019



Matchev, Kong, Kims,MP JHEP 2019



Difficulties in DL

- Feature learning (data preprocessing)

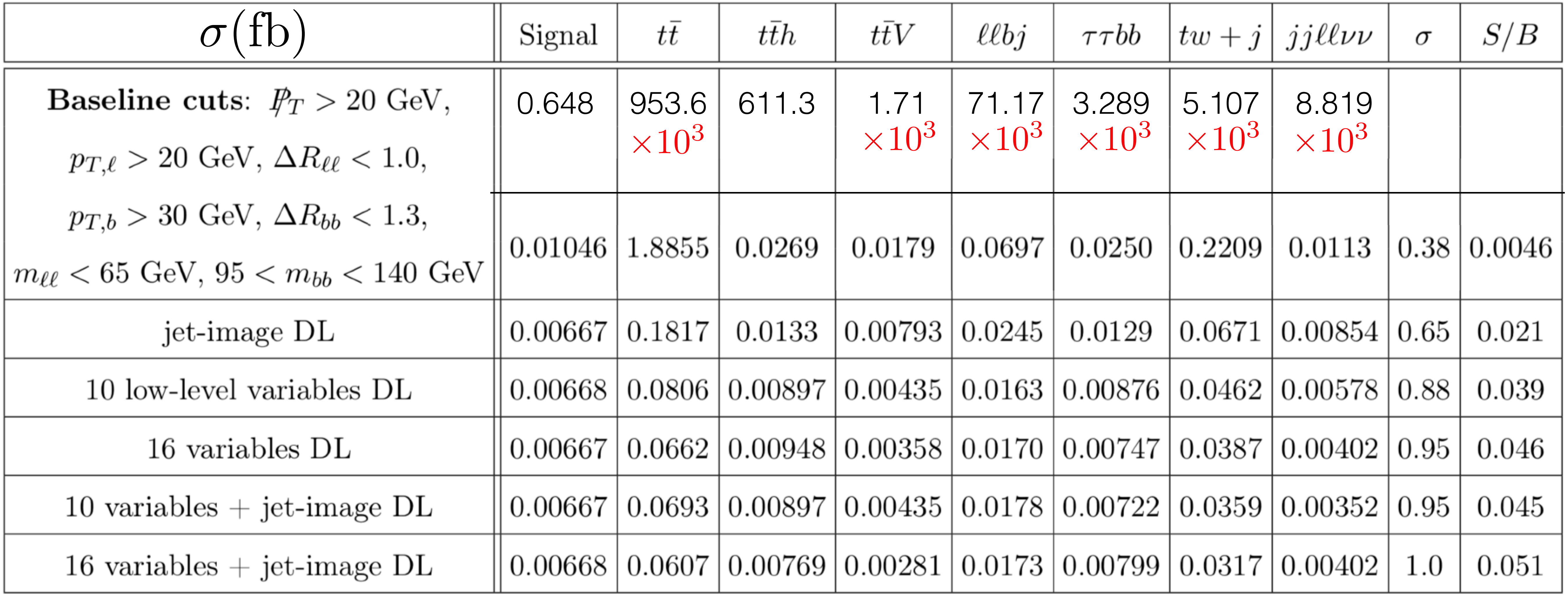
  - Is it really necessary if we have smart DL ?

- In the case of Dark matter searches,

  - For a given model, we have no idea about parameters (mass of mediator, dark matter)

  - There would be Dark matter model which we have not though about.

Difficulties in preparing MC



$\sigma(\text{fb})$	Signal	$t\bar{t}$	$t\bar{t}h$	$t\bar{t}V$	$\ell\ell bj$	$\tau\tau bb$	$tw+j$	$jj\ell\ell\nu\nu$	σ	S/B
Baseline cuts: $\not{p}_T > 20$ GeV, $p_{T,\ell} > 20$ GeV, $\Delta R_{\ell\ell} < 1.0$, $p_{T,b} > 30$ GeV, $\Delta R_{bb} < 1.3$, $m_{\ell\ell} < 65$ GeV, $95 < m_{bb} < 140$ GeV	0.648	953.6 $\times 10^3$	611.3	1.71 $\times 10^3$	71.17 $\times 10^3$	3.289 $\times 10^3$	5.107 $\times 10^3$	8.819 $\times 10^3$		
	0.01046	1.8855	0.0269	0.0179	0.0697	0.0250	0.2209	0.0113	0.38	0.0046
jet-image DL	0.00667	0.1817	0.0133	0.00793	0.0245	0.0129	0.0671	0.00854	0.65	0.021
10 low-level variables DL	0.00668	0.0806	0.00897	0.00435	0.0163	0.00876	0.0462	0.00578	0.88	0.039
16 variables DL	0.00667	0.0662	0.00948	0.00358	0.0170	0.00747	0.0387	0.00402	0.95	0.046
10 variables + jet-image DL	0.00667	0.0693	0.00897	0.00435	0.0178	0.00722	0.0359	0.00352	0.95	0.045
16 variables + jet-image DL	0.00668	0.0607	0.00769	0.00281	0.0173	0.00799	0.0317	0.00402	1.0	0.051

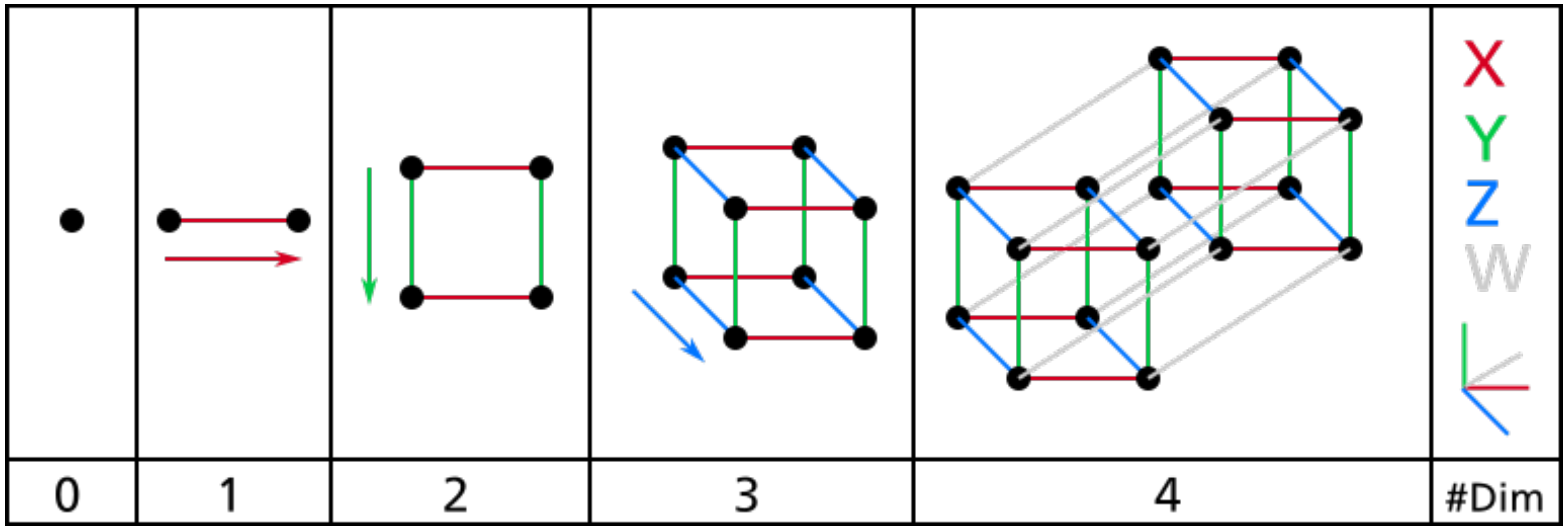
- To generate backgrounds properly, we need to make HUGE Monte Carlo samples
- Preparing "Good enough" MC samples for testing is **NOT EASY.**
- Thus, we should find very good features (High-level observables)

Curse of dimensionality problem

(1957)

n-dim cubic with length 1

average distance between two points



0.3332 0.5211 0.6616 0.7776 $\leq \sqrt{n/6}$

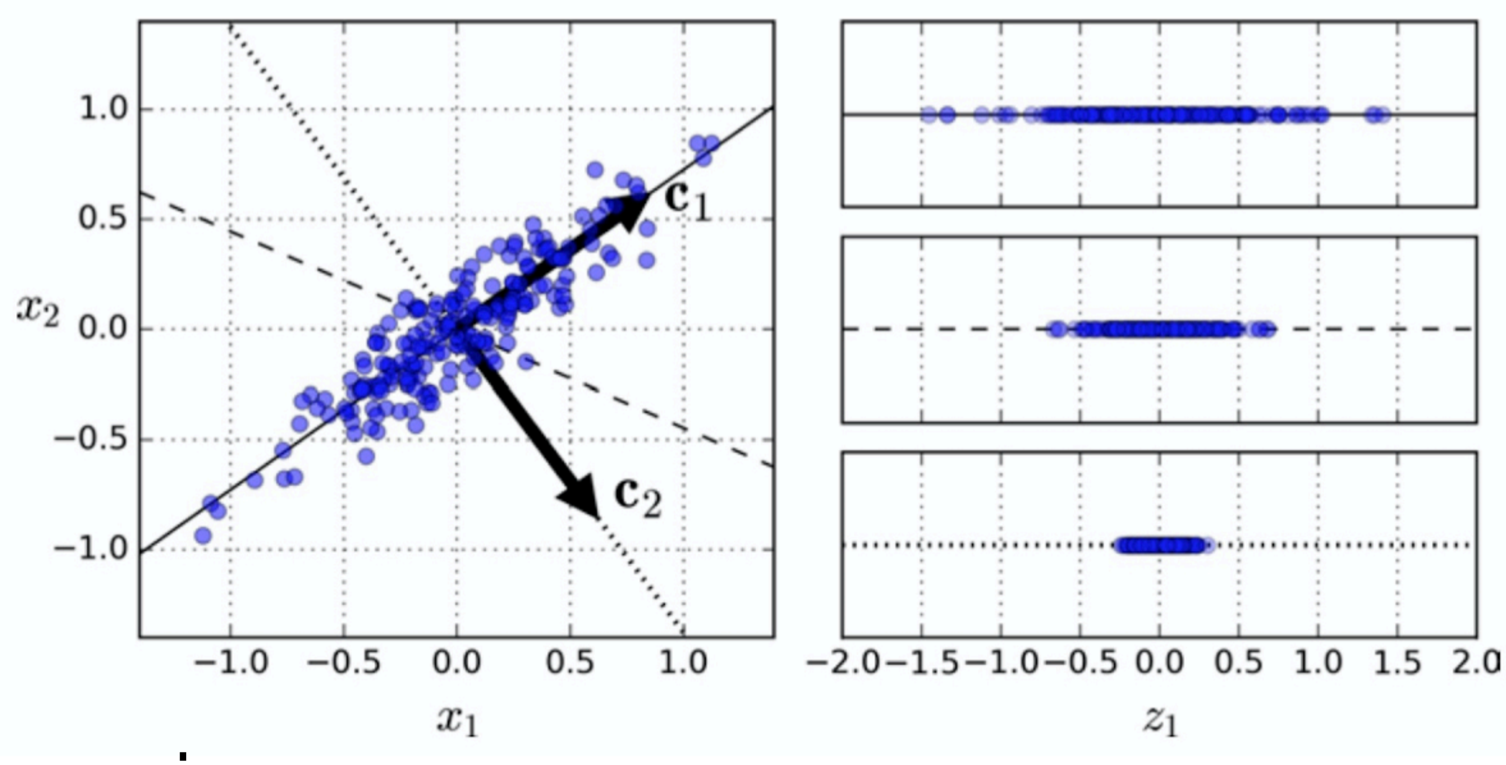
- If we increase density by a conventional grid method, number of points for $n-$dimension is proportional to d^n where d-distance in one dimension.

$$P||E_{\text{training}}(f_{\text{estimator}}) - E_{\text{test}}(f_{\text{estimator}})| > \epsilon|| \leq N_{\text{hypothesis}} e^{-2\epsilon^2 N_{\text{training}}}$$

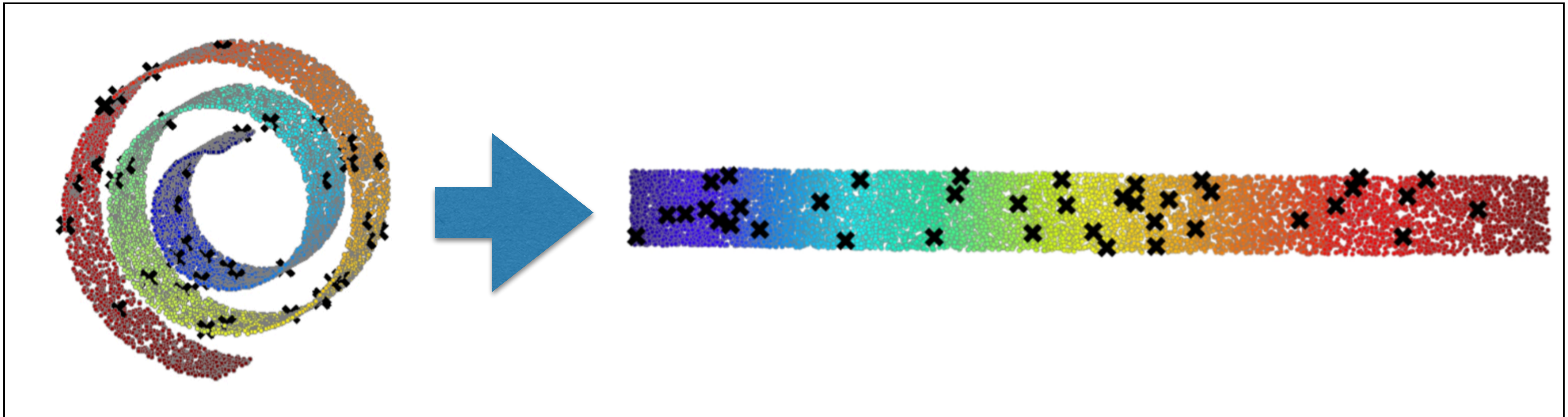
-"Hoeffding inequality" (from a textbook)

- There have been various methods to resolve this issue.

- Principal Component Analysis (PCA)

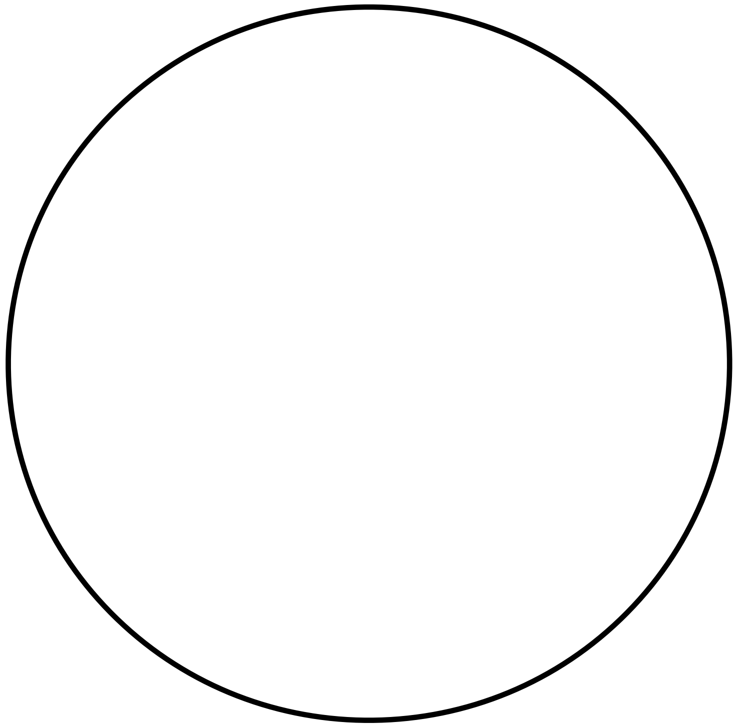


- Manifold learning

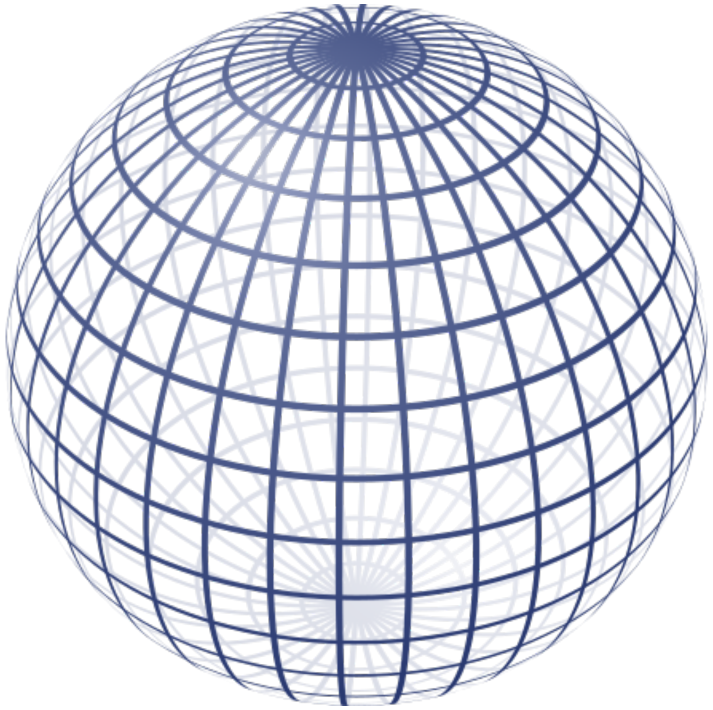


Kilian Q. Weinberger et.al. 2005

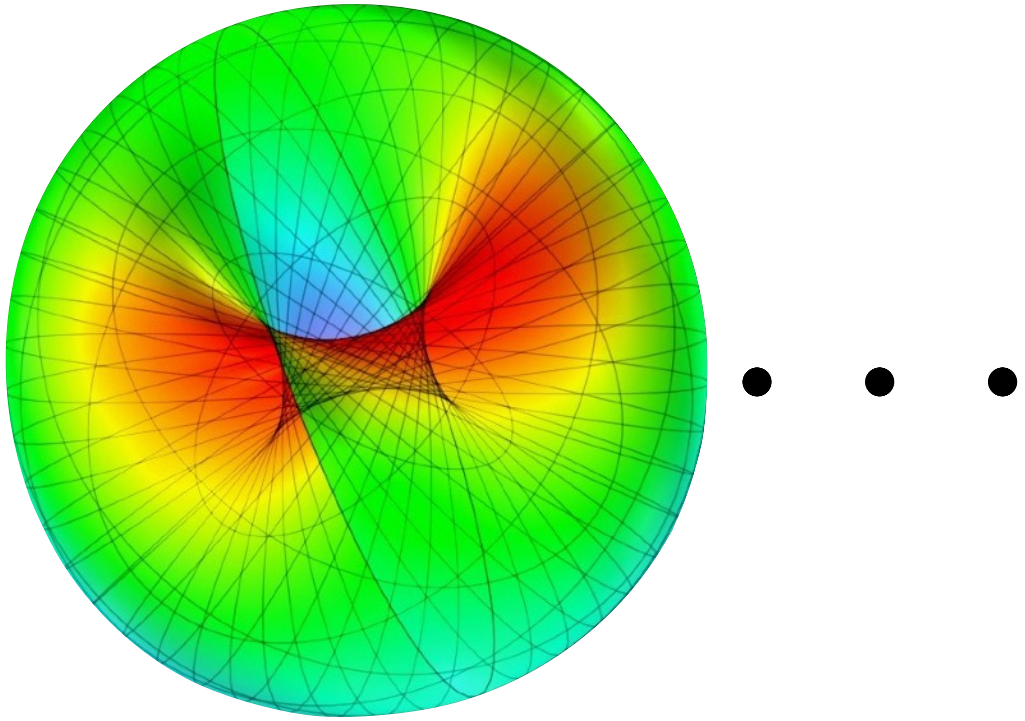
- But, **feature variable** in our hands (HEP) are not simple function of raw data. The transformation is **highly non-linear.**
  - we try to find a good DL architecture... **also** ...
  - we design feature variables for the input of DL



S^1

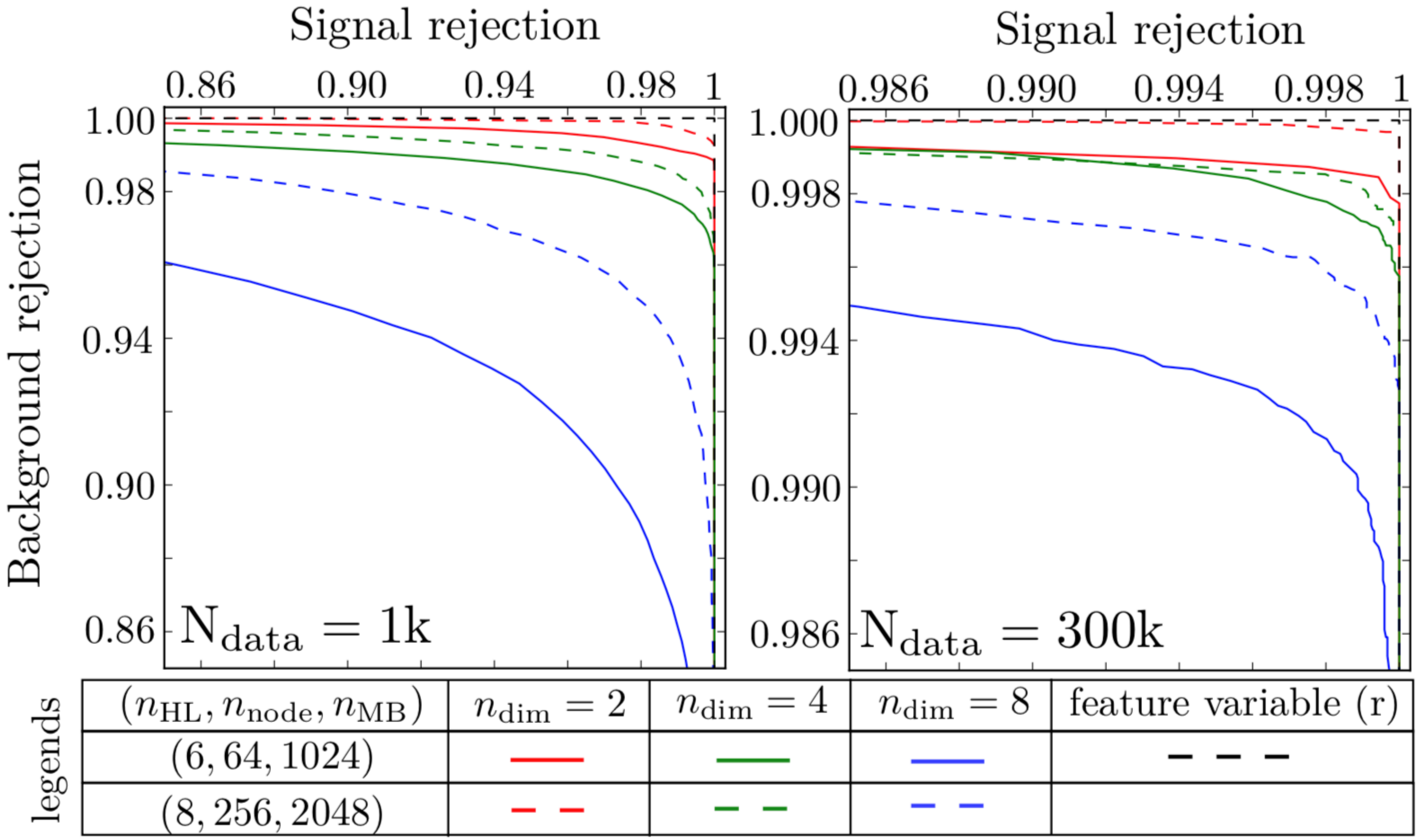


S^2



Projection of S^3 into R^3

- Inside n-sphere: Signal events
  Outside n-sphere: Backgrounds.
- Featured variable: Radius.
  Raw observable: coordinate

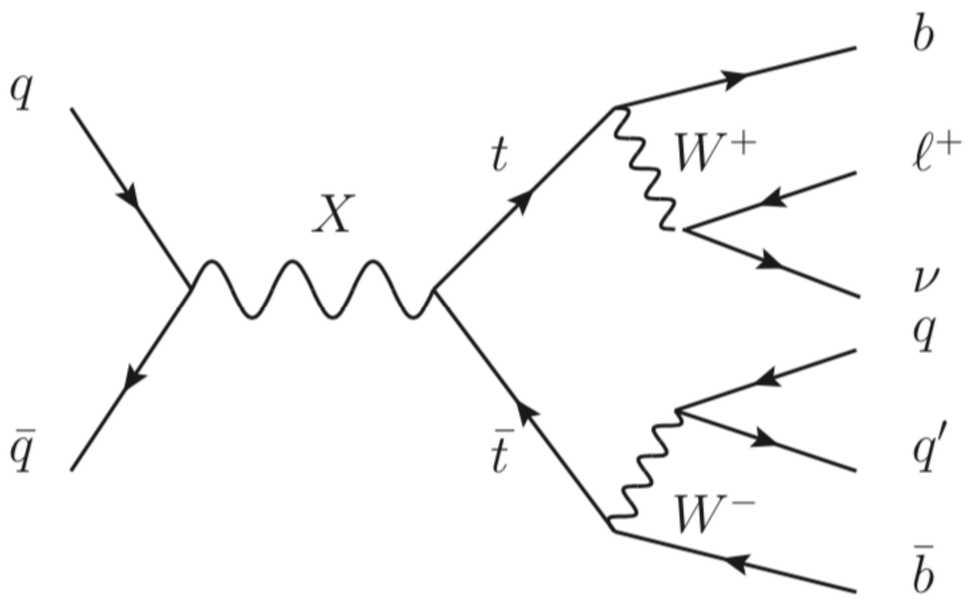


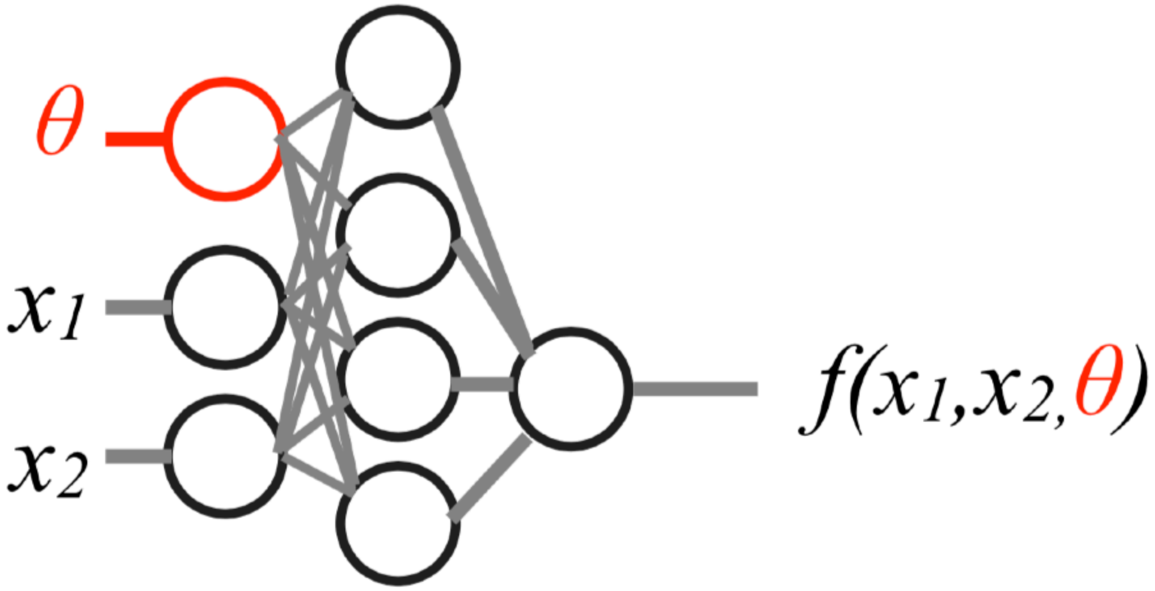
$(n_{\mathrm{HL}}, n_{\mathrm{node}}, n_{\mathrm{MB}})$	$n_{\mathrm{dim}} = 2$	$n_{\mathrm{dim}} = 4$	$n_{\mathrm{dim}} = 8$	feature variable (r)
$(6, 64, 1024)$	―	―	―	– – –
$(8, 256, 2048)$	– –	– –	– –	

- Featured variables are effective in learning with few data.

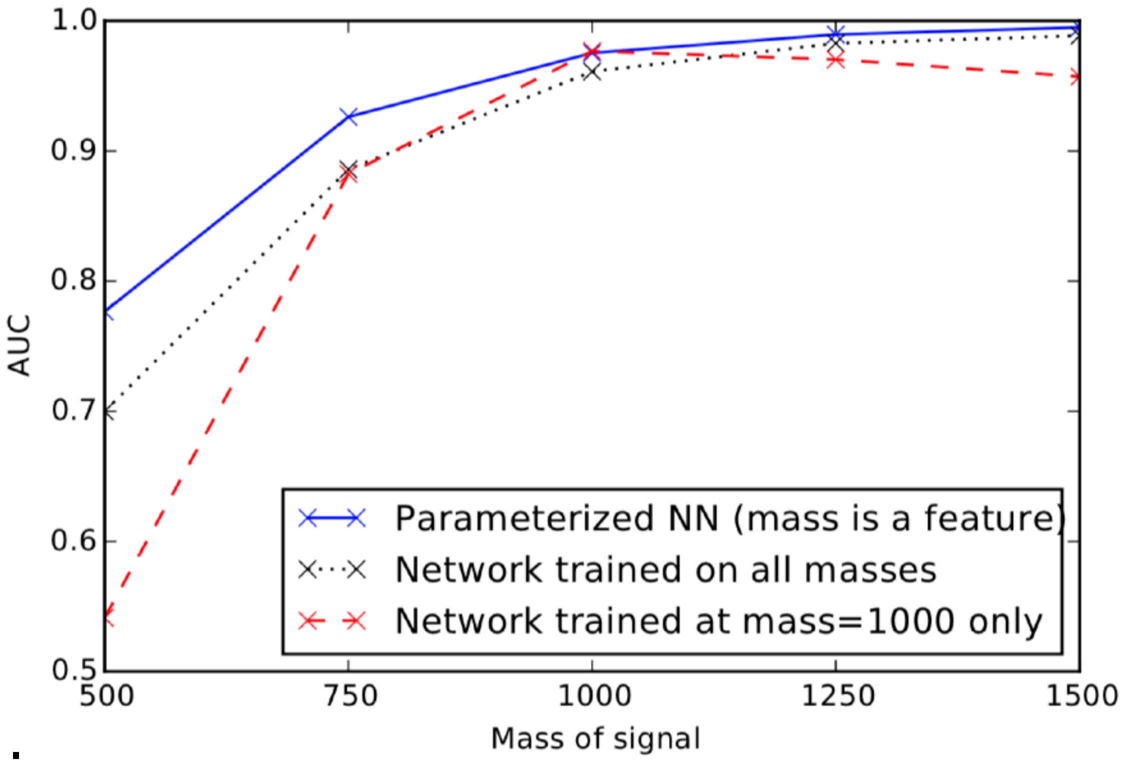
- Featured variable: Radius.
  Raw observable: coordinate

Out of parameter problem



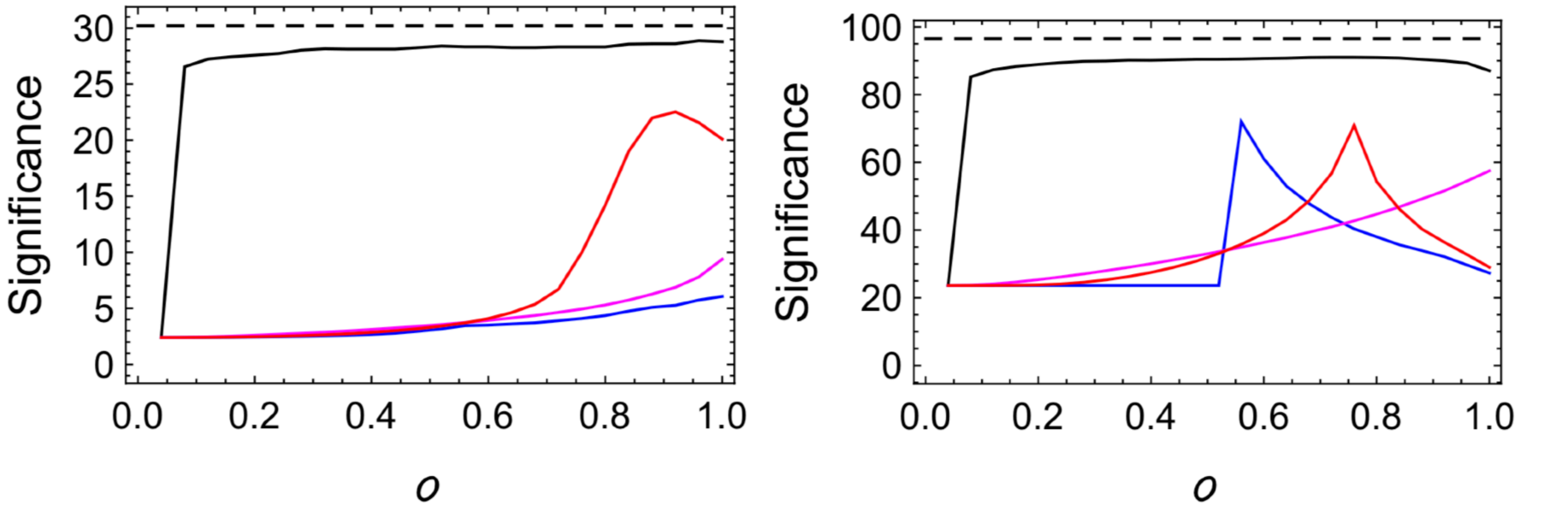


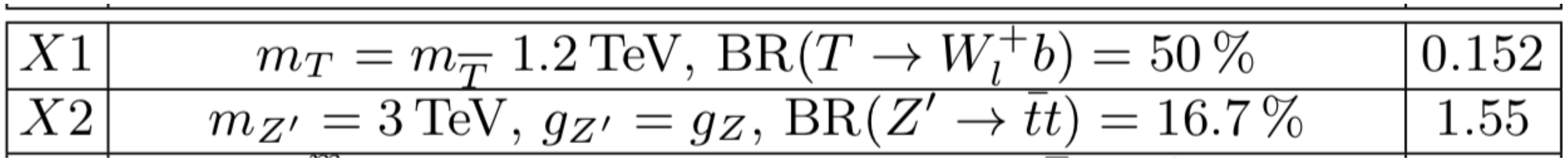
- Parameterized NN:
  A single neural network with the true mass, as an input feature.



arXiv:1601.07913 by Pierre Balda et.al.

Out of Model ambiguity problem





$X1$	$m_T = m_{\overline{T}}$ 1.2 TeV, BR$(T \to W_l^+ b) = 50\,\%$	0.152
$X2$	$m_{Z'} = 3$ TeV, $g_{Z'} = g_Z$, BR$(Z' \to t\bar{t}) = 16.7\,\%$	1.55

- Based on autoencoder, find an anomaly away from Standard Model (backgrounds) expectations.

arXiv:1807.10261 by Tao Liu et.al.

Conclusions

- Various DL algorithms can enhance searches at the LHC

- When we are targeting a specific NP scenario,
  we can maximize a sensitivity by aggressively utilizing "**feature variables**" through DL.
  - reducing the issue of Dimensionality.

- When we don't have any preferred parameter in a given model, still DL would provide the best performance.

- When we don't have any MODEL in our mind, DL can provide a "good" results via an anomaly detection....
  - We need to check what kind of models we have missed so far !