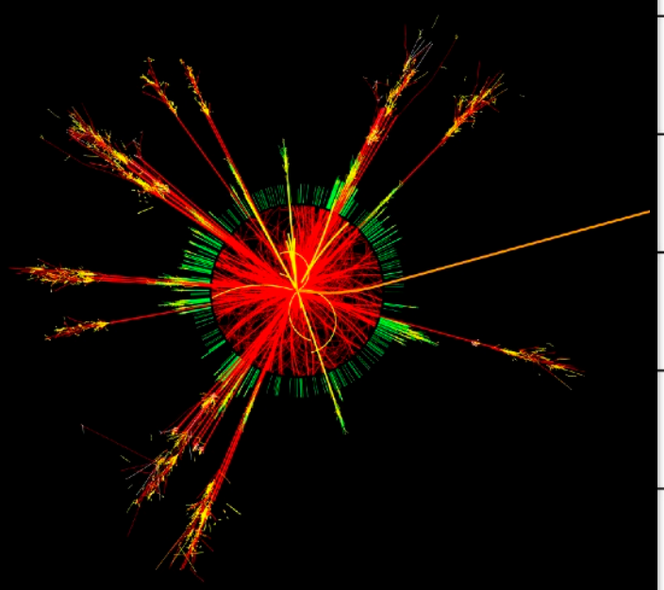
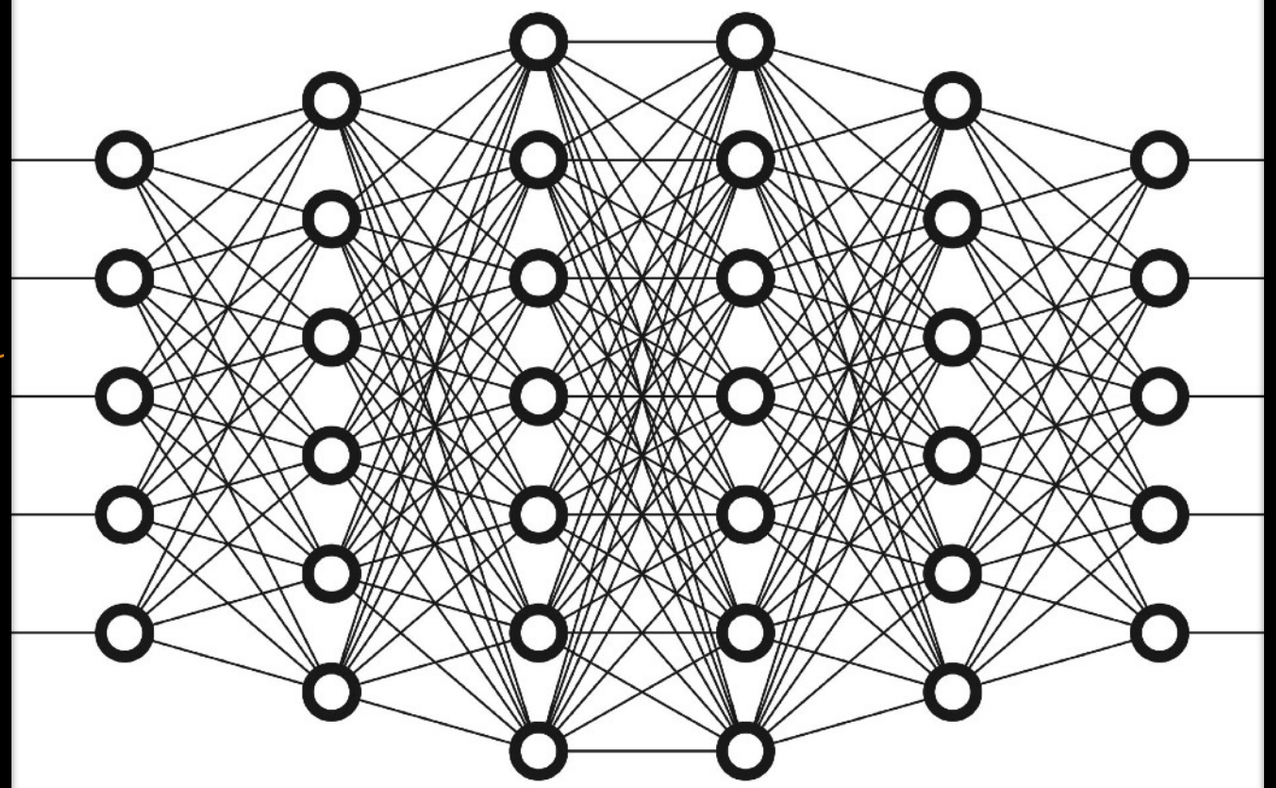
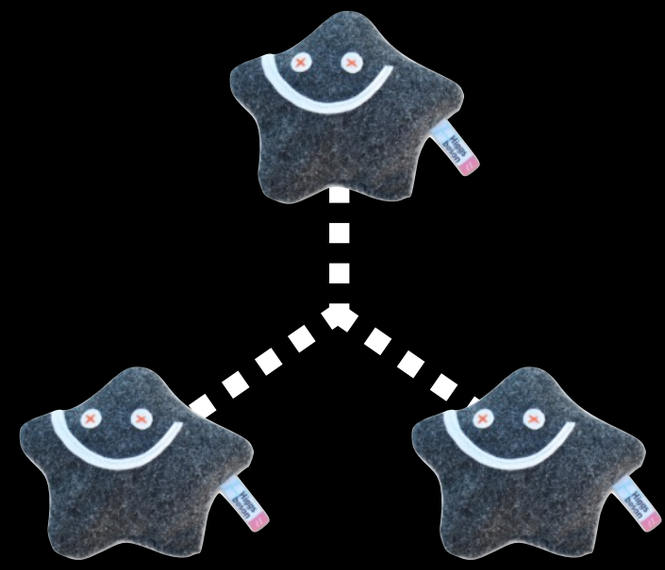
Higgs precision with Deep Learning







Myeonghun Park

(**Seoultech**)

with Minho Kim, Jeonghan Kim

K.C. Kong, Konstantin T. Matchev

Phys. Rev. Lett 122 (2019) 091801

JHEP 1909 (2019) 047

arXiv:1912.XXXXX ? maybe 2001.YYYY

IBS-PNU Joint Workshop on Physics beyond the Standard Model

Plz..... Any news?

Desperate... ?

- LHC provides complicated data in an unprecedented way.
  - **Deepen understandings about** QCD / Standard Model.

  Though these are "just" backgrounds to someone who is pursuing a new physics...

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

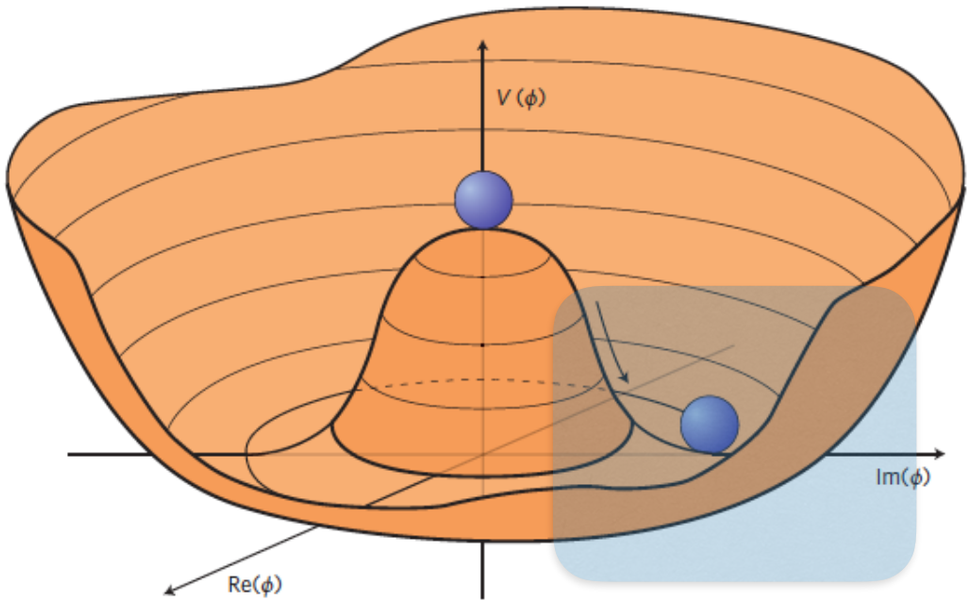
- As our preferred BSM models have been passed away...
  If we want to have **some clues from BIG data, unsupervised ML (anomaly detection) would be the way to go.**

Precision ?

- In the Standard Model, parameters related to Higgs would be the last hope to check.

- **7 years ago**, we started to understand EWSB (from LHC)

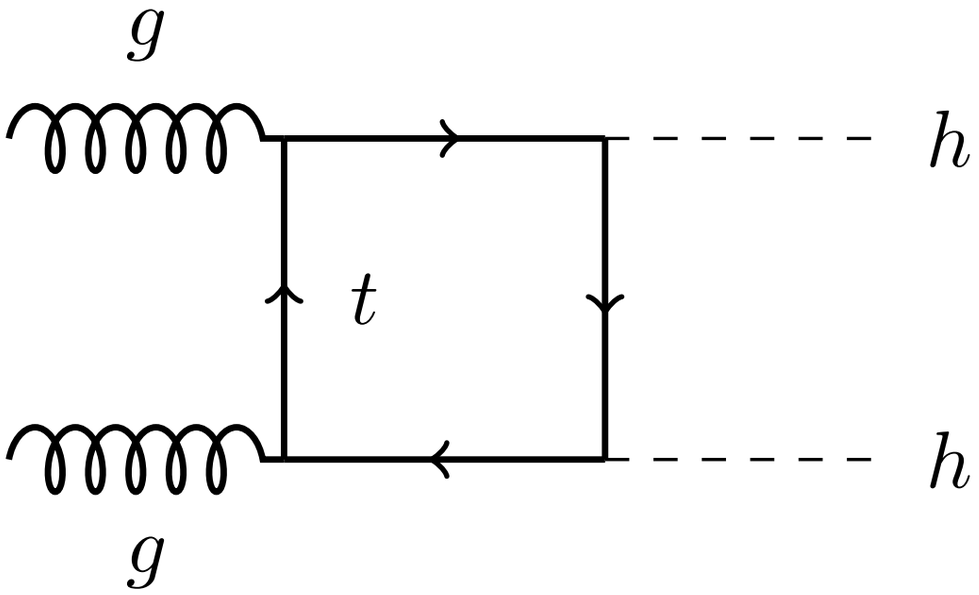
- To complete our understanding, we go further to examine the shape of the Higgs potential with data.

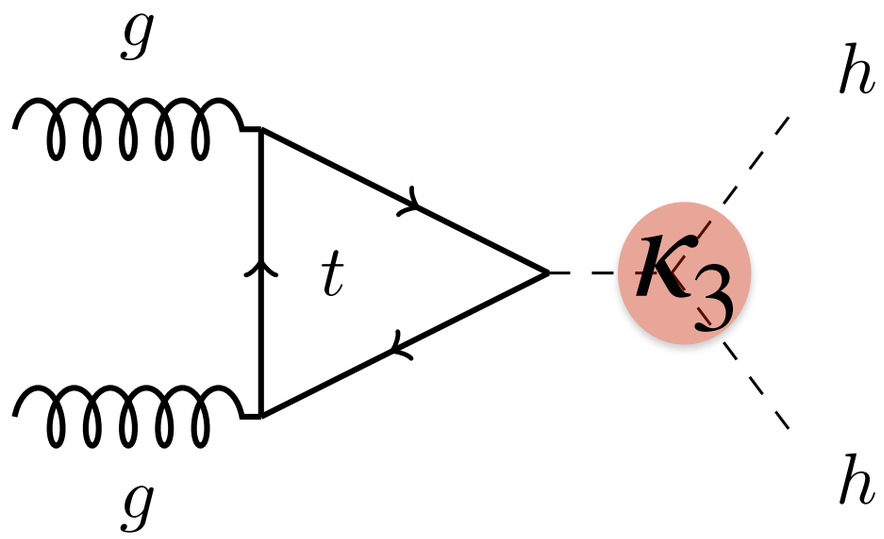


$$V_h = \frac{m_h^2}{2}h^2 + \kappa_3\frac{m_h^2}{2v}h^3 + \kappa_4\frac{m_h^2}{8v^2}h^4$$

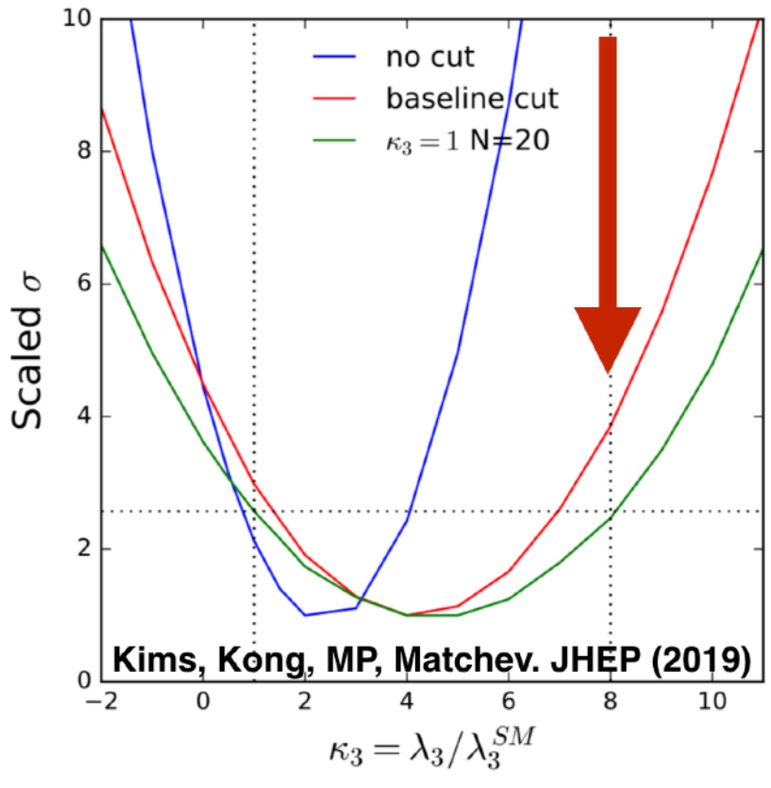
- This example is "**supervised**" Machine Learning with **feature** variables

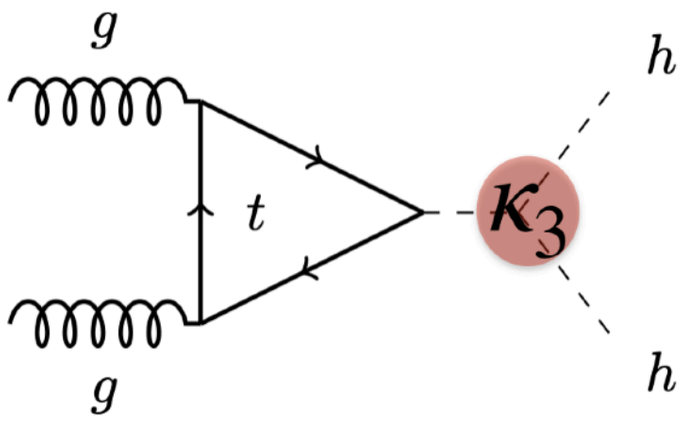
H-H-H. Higgs triple coupling





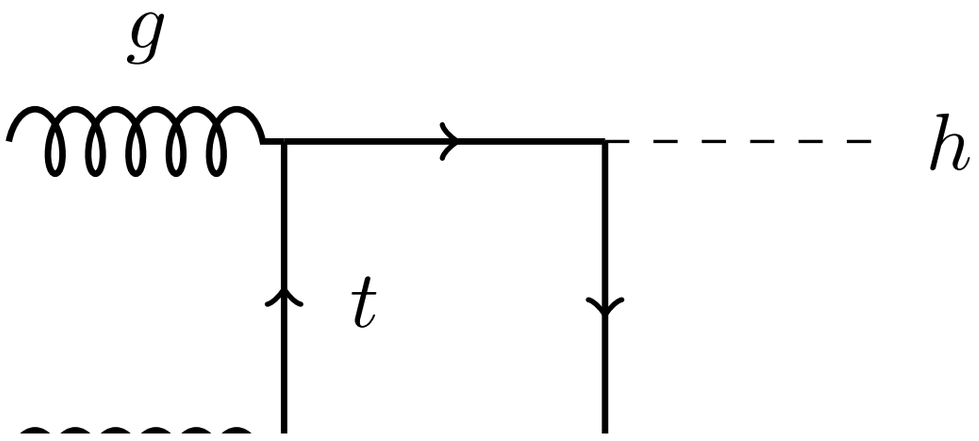
$$\sigma_{gg\to hh}(\hat{s}) = \frac{\alpha_s^2}{2^{15} v^4 \pi^2 \hat{s}^2} \int d\hat{t} (|F_1|^2 + |F_2|^2) \approx c_\triangle \, \kappa_3^2 + c_{\triangle,\square} \, \kappa_3 + c_\square$$

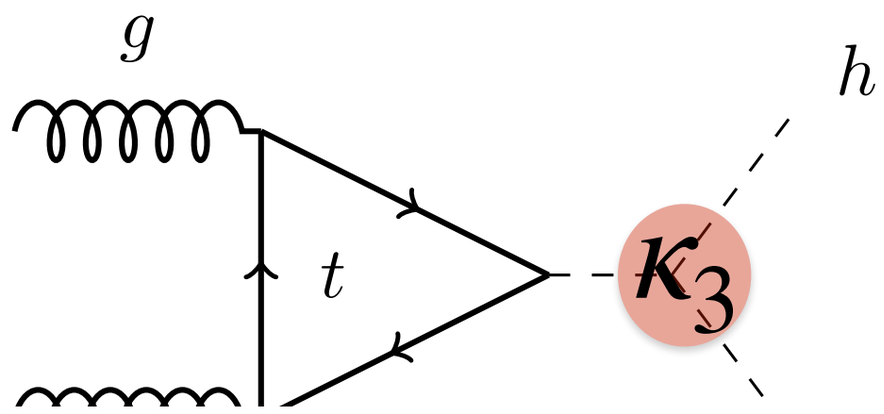




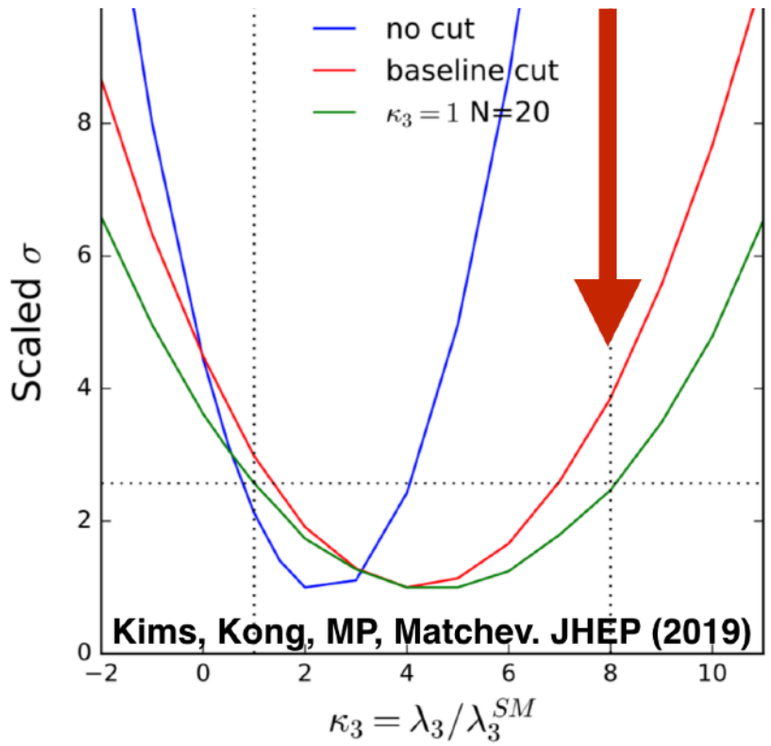
$$-\frac{m_Q^2}{\hat{s}} \left[\log\left(\frac{m_Q^2}{\hat{s}}\right) + i\pi \right]^2 + \mathcal{O}\left(\frac{m_Q^2}{\hat{s}}\right)$$

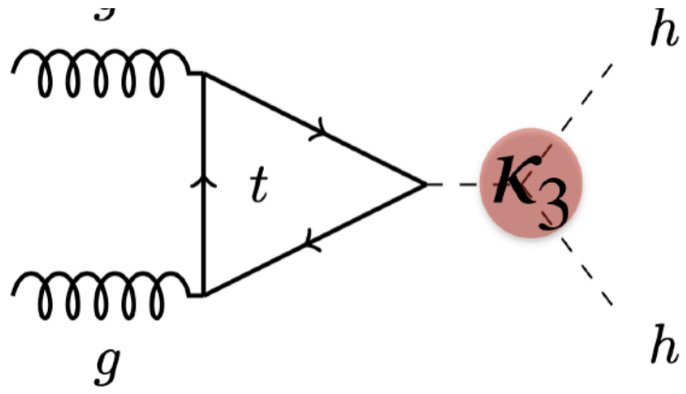
H-H-H. Higgs triple coupling





- **Problem 1**: Triangle diagram is sensitive at lower-energy (**softs**) bins where backgrounds dominate...



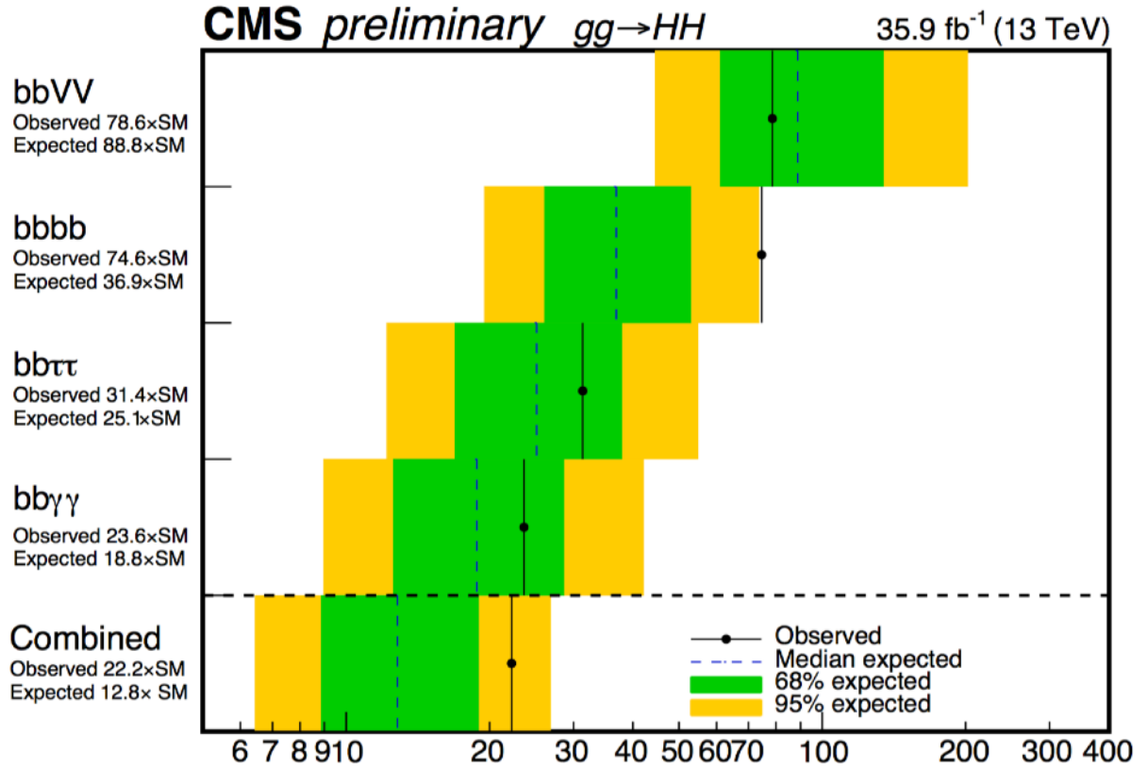


$$-\frac{m_Q^2}{\hat{s}}\left[\log\left(\frac{m_Q^2}{\hat{s}}\right)+i\pi\right]^2+\mathcal{O}\left(\frac{m_Q^2}{\hat{s}}\right)$$

$h \to XX$

$XX \leftarrow h$

	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%



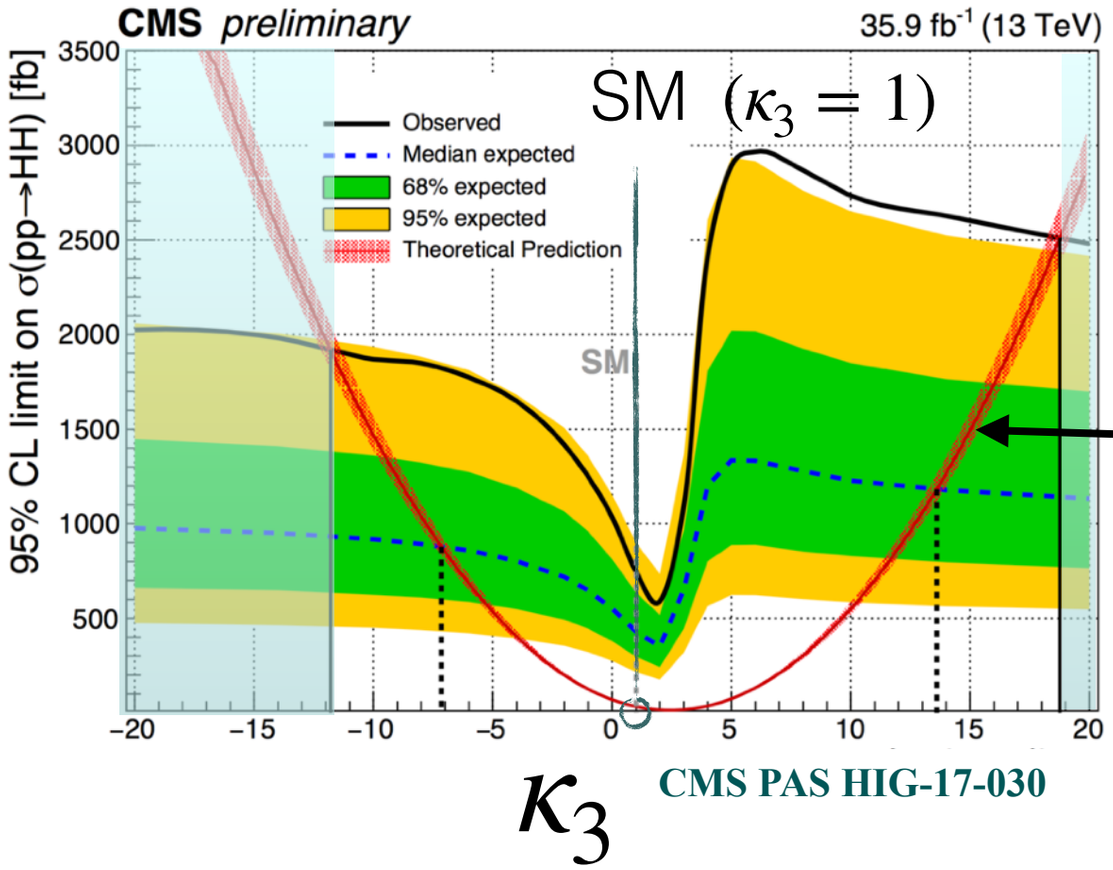
$$\sigma(hh)^{NNLO}_{SM} \simeq 40.7 \text{ fb}$$

1902.00134	Statistical-only		Statistical + Systematic	
	ATLAS	CMS	ATLAS	CMS
$HH \to b\bar{b}b\bar{b}$	1.4	1.2	0.61	0.95
$HH \to b\bar{b}\tau\tau$	2.5	1.6	2.1	1.4
$HH \to b\bar{b}\gamma\gamma$	2.1	1.8	2.0	1.8
$HH \to b\bar{b}VV(ll\nu\nu)$	-	0.59	-	0.56
$HH \to b\bar{b}ZZ(4l)$	-	0.37	-	0.37
combined	3.5	2.8	3.0	2.6
	Combined 4.5		Combined 4.0	

4σ expected for ATLAS+CMS!

- These measurements are challenged by a low σ (*hh*) and small branching ratios (*BR*).
- **No single channel** is expected to reach 5 sigma at HL-LHC.
- The combination of different channels is crucial. **bbWW has good potential for further improvement.**

Experimental status on κ_3 at LHC 13 TeV



CMS PAS HIG-17-030

- Allowed region of κ_3

$$-11.8 < \kappa_3 < 18.8$$

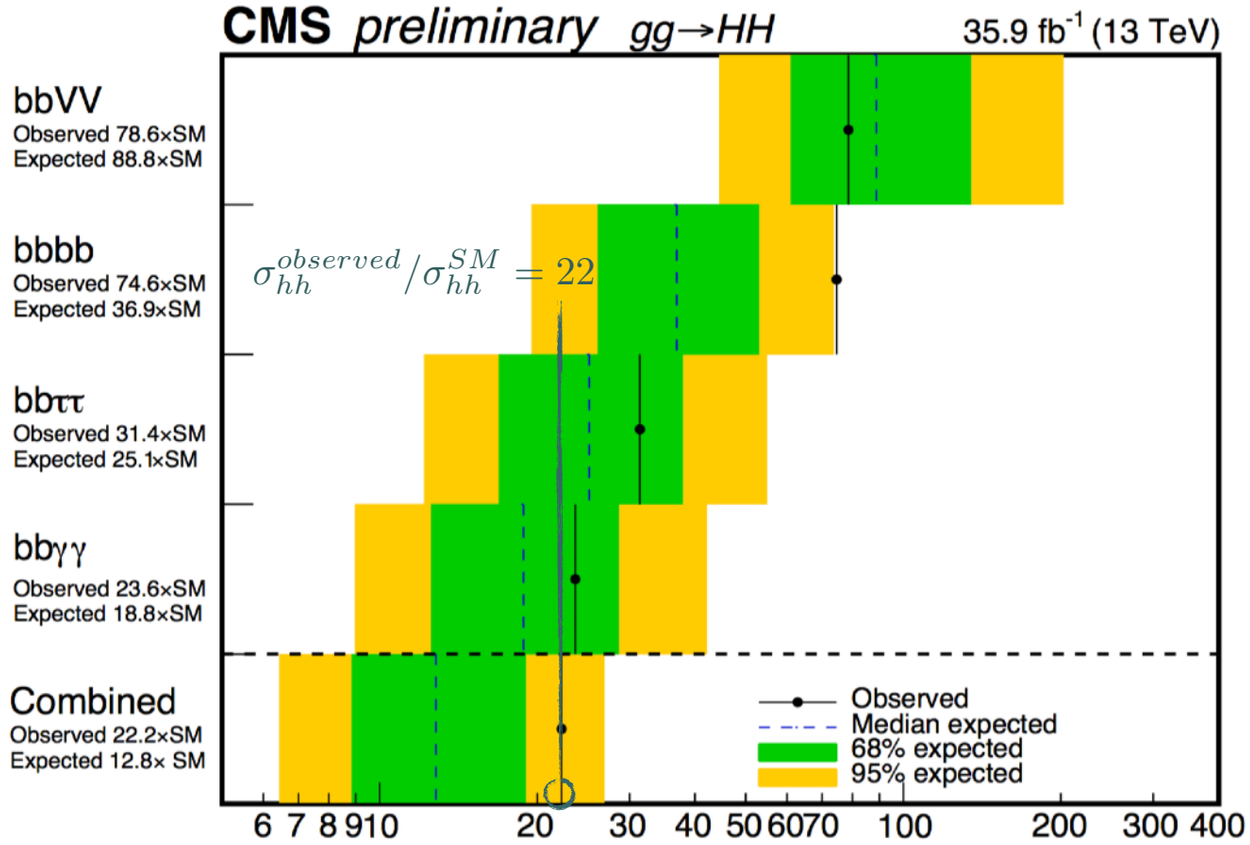
(all channels combined)

$$c_{\triangle}\,\kappa_3^2 + c_{\triangle,\square}\,\kappa_3 + c_{\square}$$

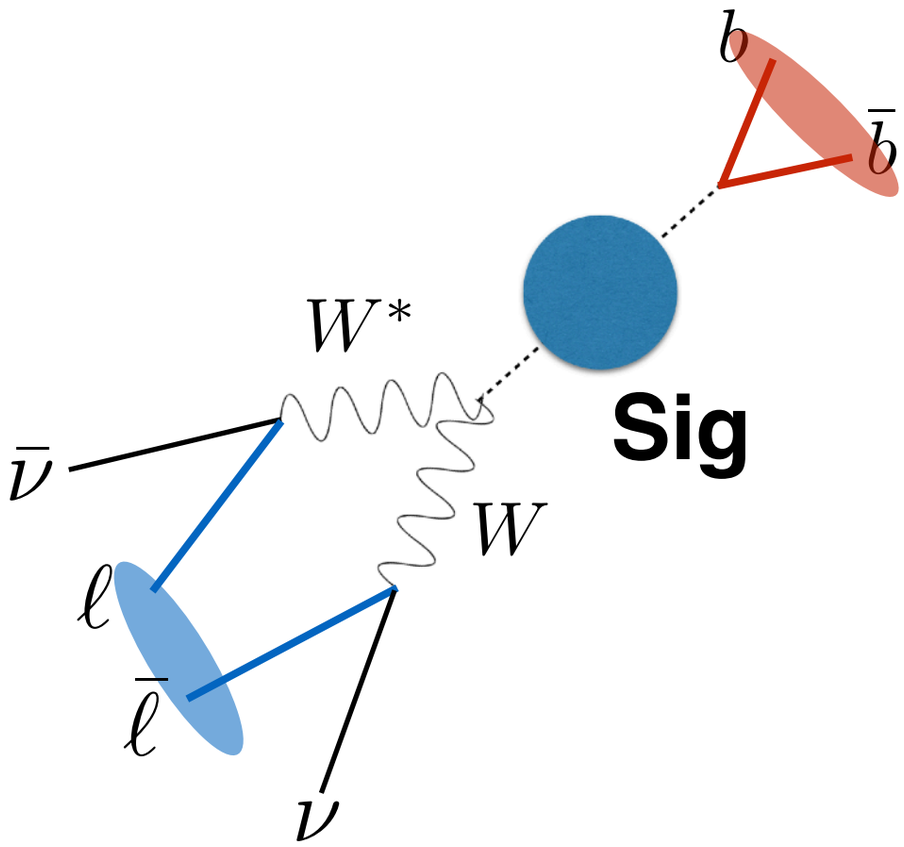
- Allowed range of hh cross sections.

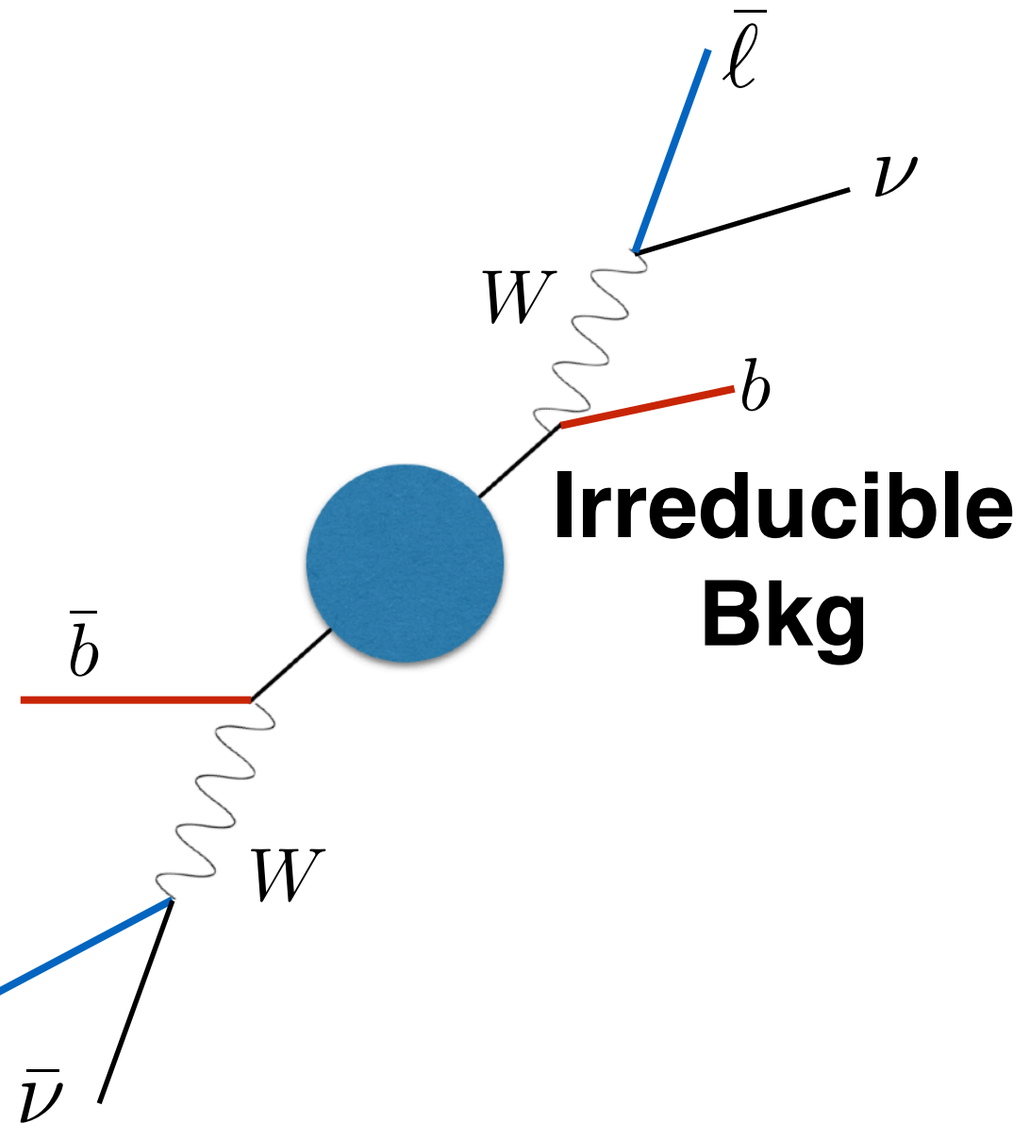
$$\sigma_{hh}/\sigma_{hh}^{\mathrm{SM}} = 22$$

- The $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\bar{\tau}$ are leading channels.



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

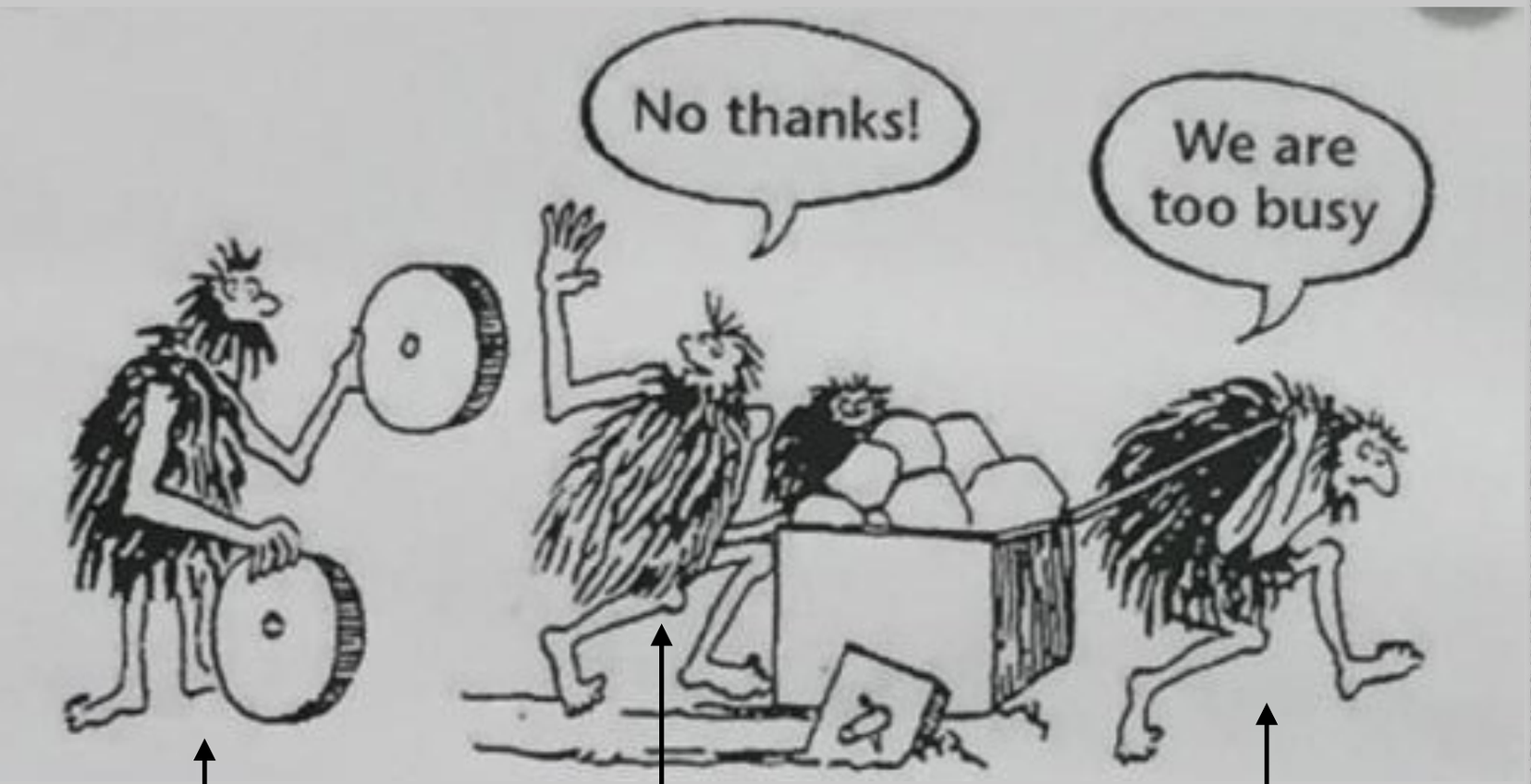




- **Problem 2**: In $bbVV$ channel, $t\bar{t}$ **backgrounds** are HUGE...

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

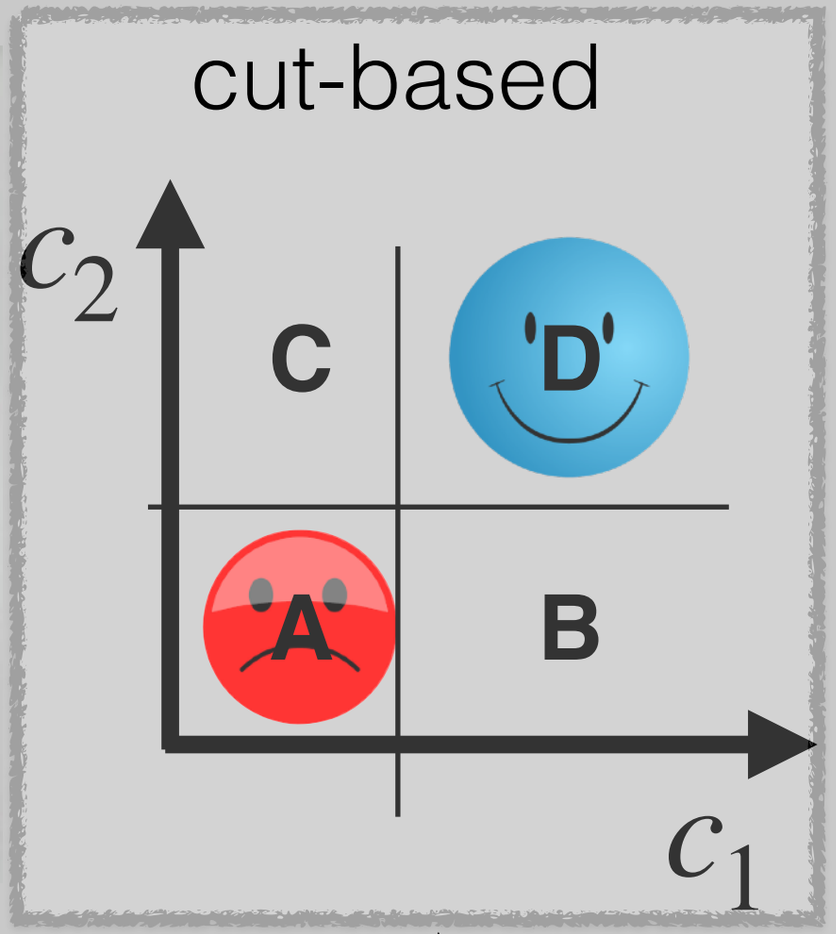
1. Applying **variables**



young generation
(Minho Kim, Jeong Han Kim)

my old friends

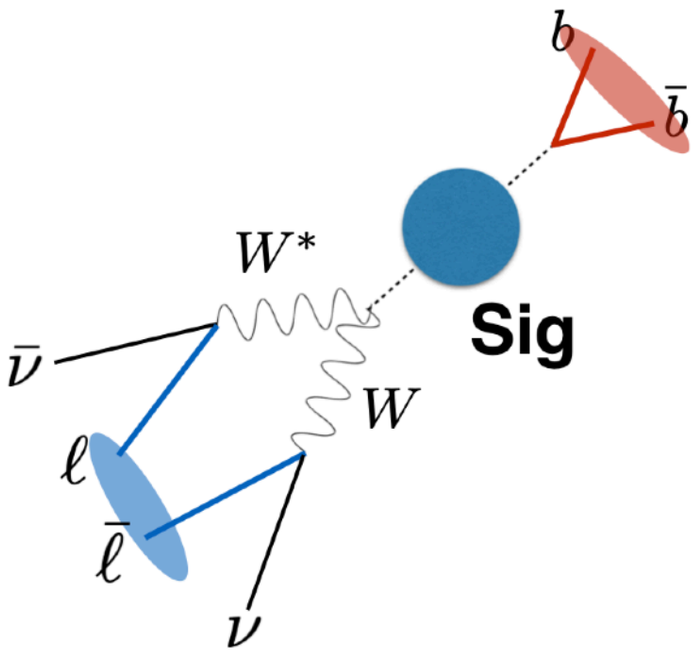
me!

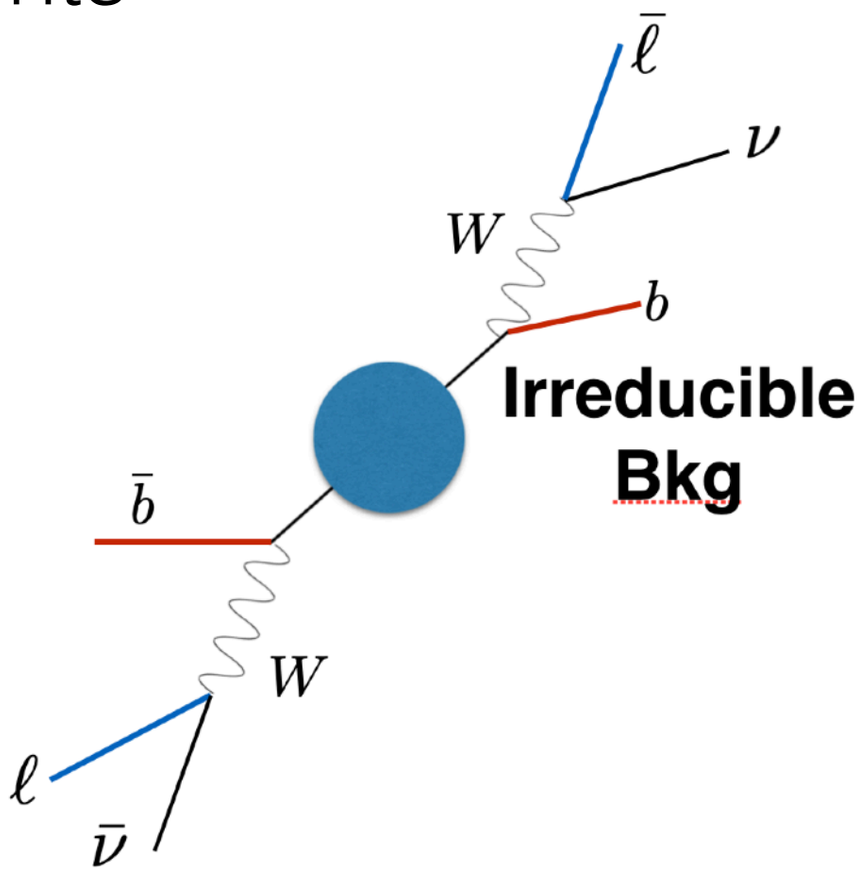


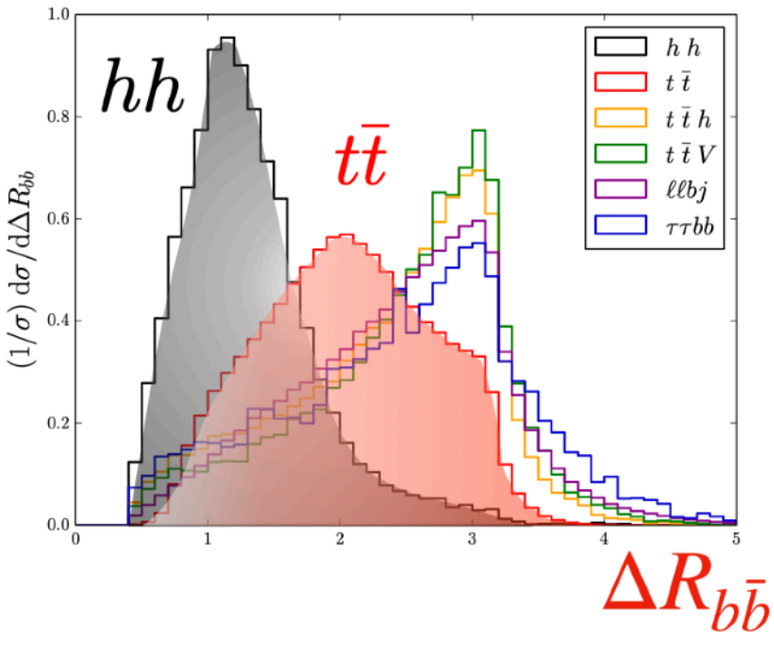
my previous favorite

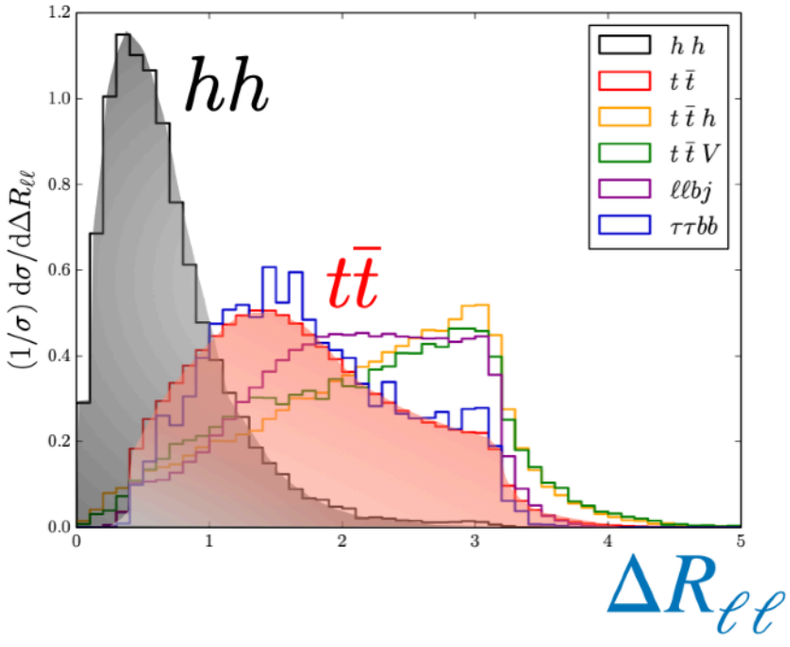
Conventional method to design cuts

- From patterns of signal events









- 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	N_{sig}^{SM}/N_{bknd}
0.0124	1.1724	0.0297	0.0246	0.0158	0.0379	0.00590	0.60	0.00964

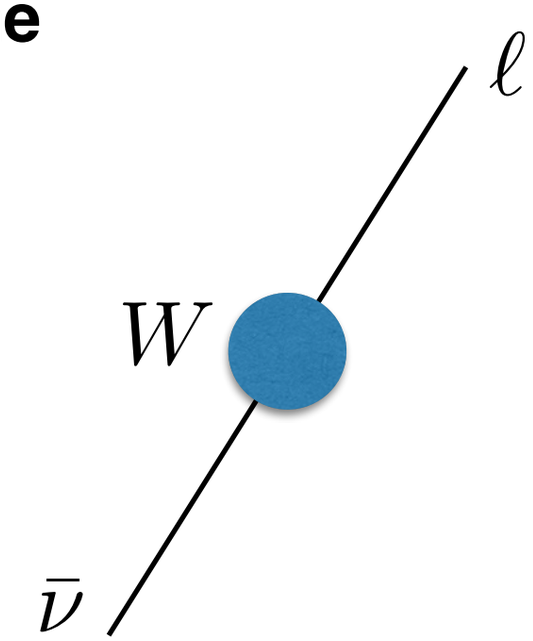
$jj\ell\ell\nu\bar{\nu}$ backgrounds from QCD+EW

- 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**

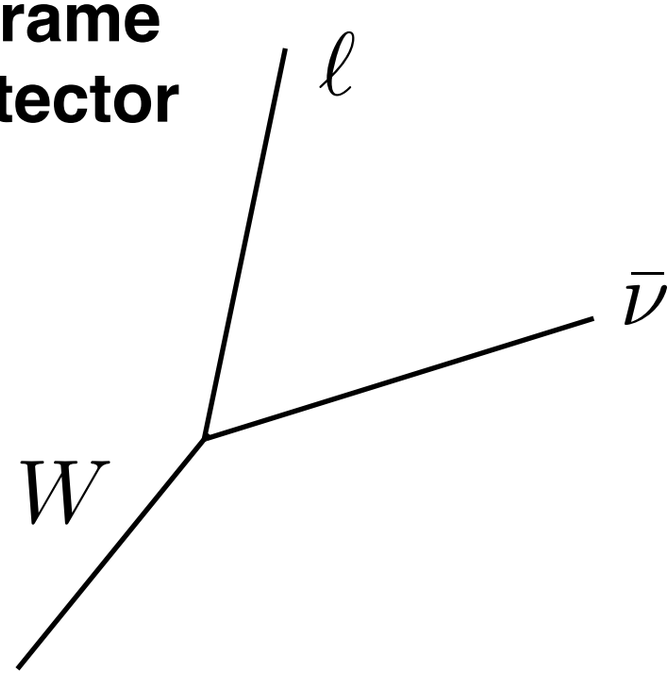
- A **low-level** variable contains various information

Rest frame of W

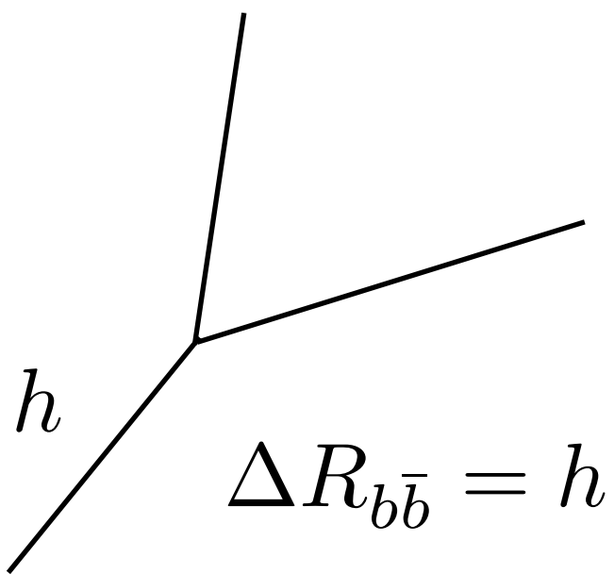


$$P_{t(\ell)} = f(M_W, M_\nu, M_\ell)$$

Lab frame of Detector



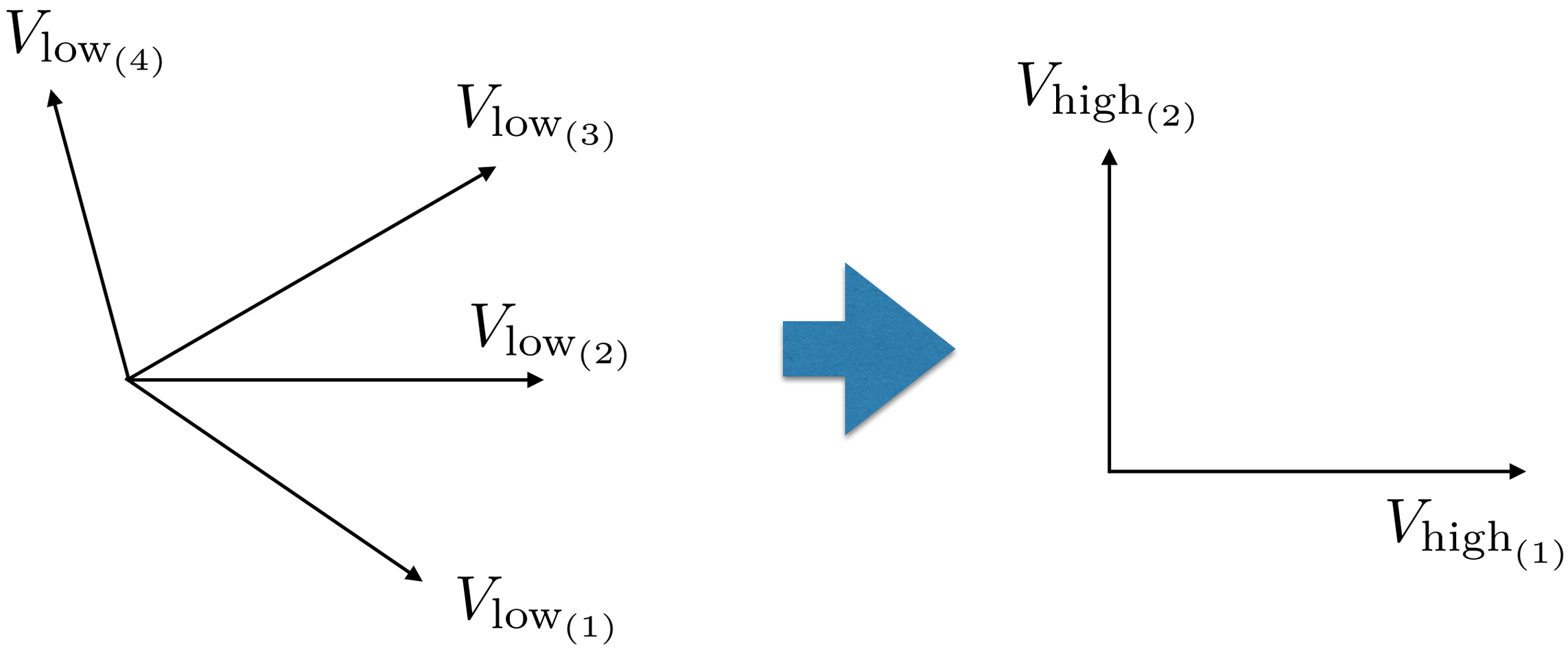
$$P_{t(\ell)} = g(M_W, M_\nu, M_\ell, \eta_W)$$



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

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

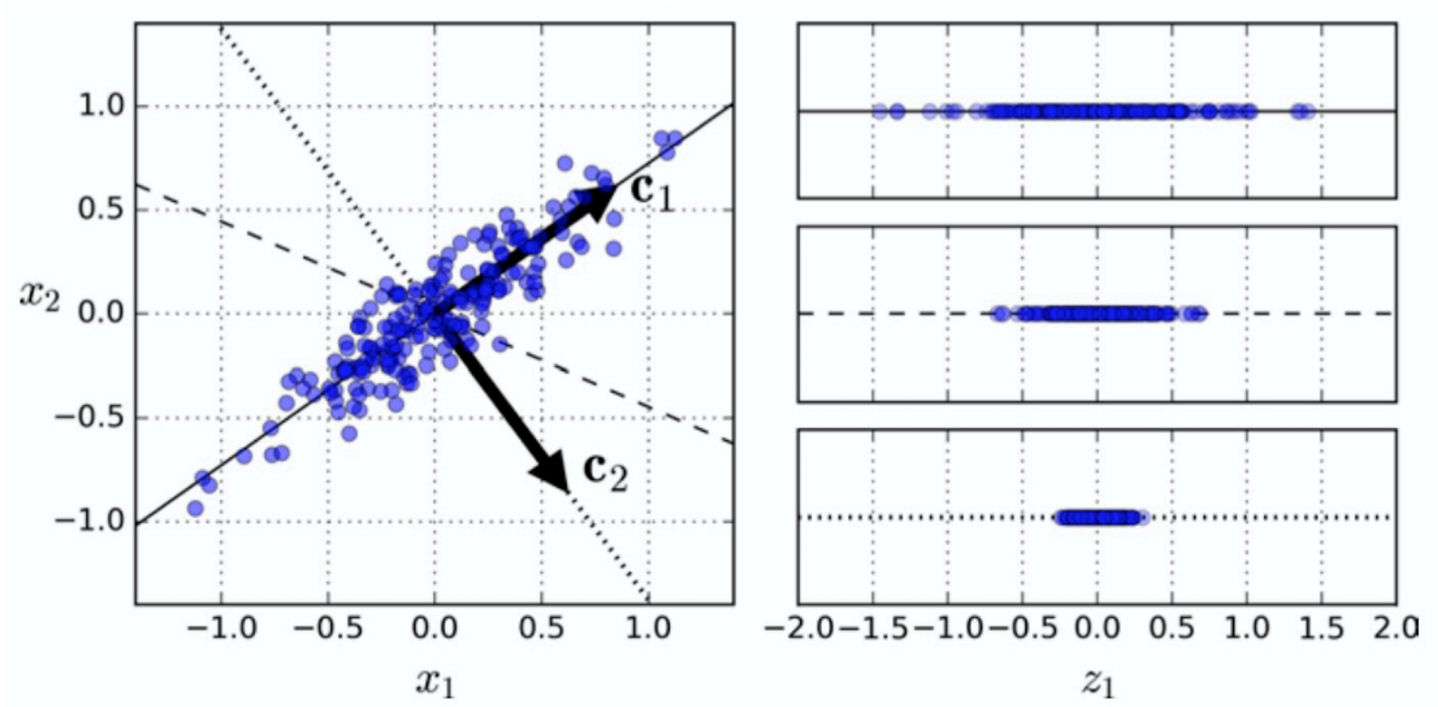
example of 2-dim



- Human can do this... and also **Computer can achieve this**..

- There have been various methods to resolve this issue.

- Principal Component Analysis (PCA)



- Manifold learning

Kilian Q. Weinberger et.al. **2005**

Maximizing correlations

VOLUME 65, NUMBER 11 PHYSICAL REVIEW LETTERS 10 SEPTEMBER 1990

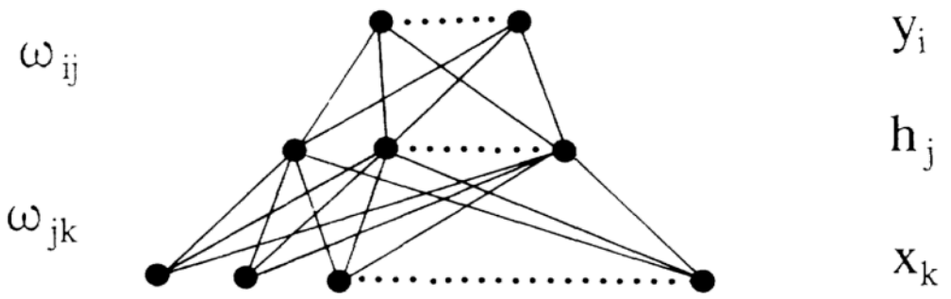
Finding Gluon Jets with a Neural Trigger

Leif Lönnblad,[a] Carsten Peterson,[b] and Thorsteinn Rögnvaldsson[c]
Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden
(Received 6 April 1990)

Using a neural-network classifier we are able to separate gluon from quark jets originating from Monte Carlo–generated e^+e^- events with 85%–90% accuracy.

PACS numbers: 13.87.Fh, 12.38.Qk, 13.65.+i

The hidden nodes have the task of correlating and building up an "internal representation" of the patterns to be learned. Training the network corresponds to changing the weights ω_{ij} such that a given input parameter $\mathbf{x}^{(p)}$ gives rise to an output (feature) value $\mathbf{y}^{(p)}$ that equals the desired output or target value $\mathbf{t}^{(p)}$. A frequently used procedure for accomplishing this is the *back-propagation* learning rule[5] where the error function



For each of the different data sets we use two different approaches of presenting the jets to the network. One is to show only the four-momenta $(\mathbf{p}_k, E_k)$ of the four leading particles in the jet. In this way we do not reveal too

[5]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition,* edited by D. E. Rumelhart and J. L. McClelland (MIT Press, Cambridge, 1986), Vol. 1.

2. Hypothesis test

For ttbar Background: A mass variable

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

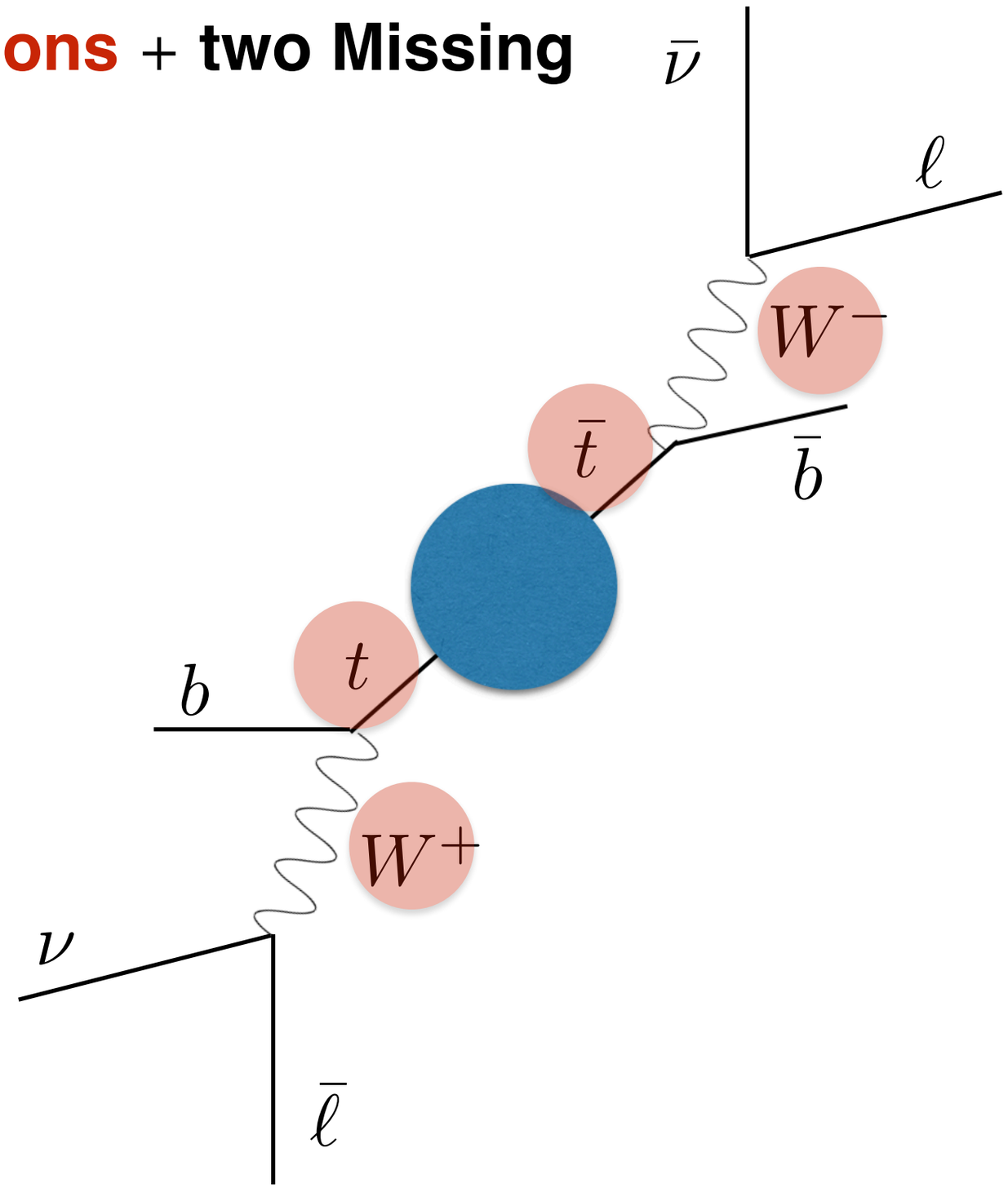
- We have **four mass-shell conditions** + **two Missing Transverse Energy conditions**

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

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

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

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

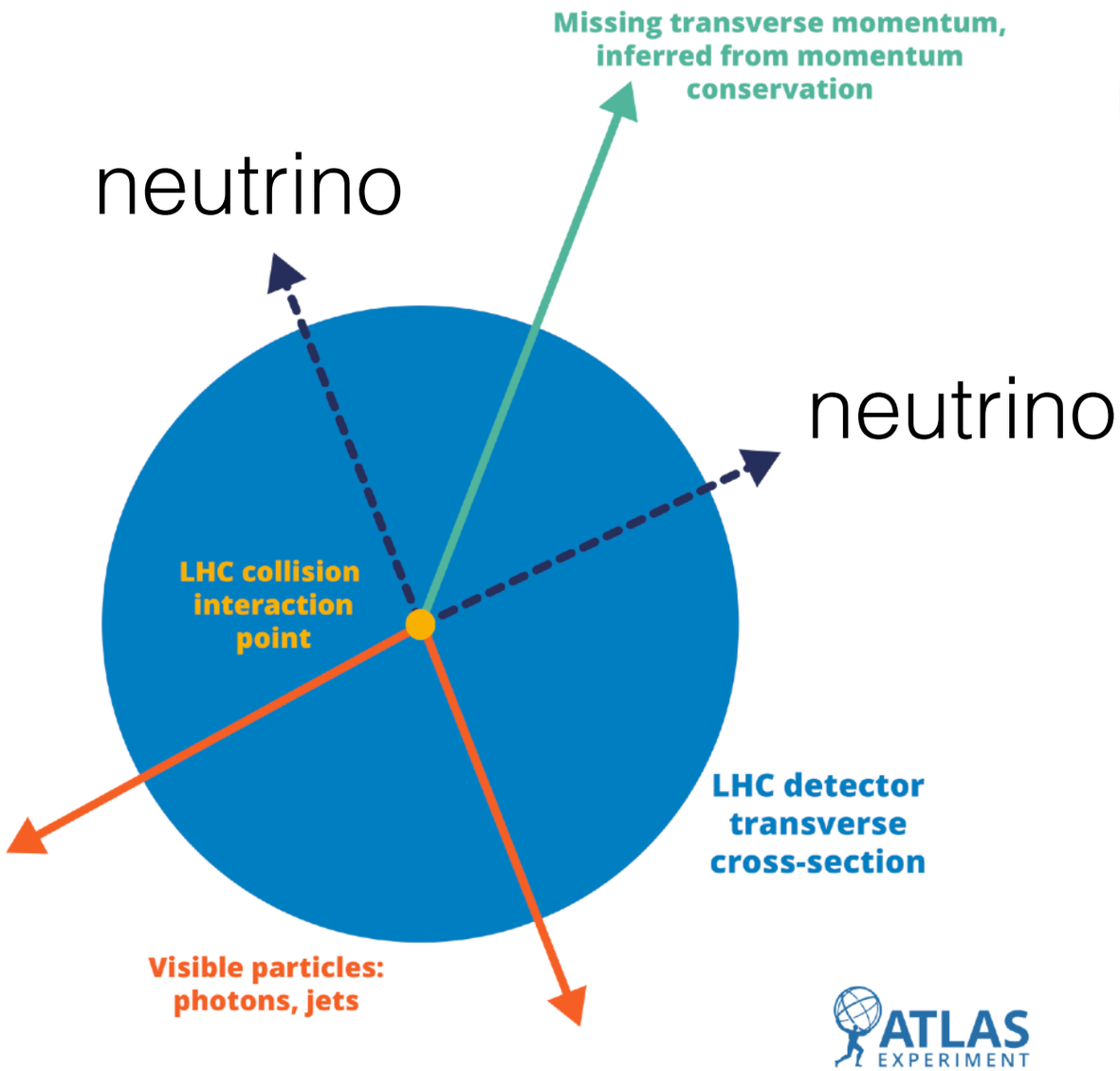


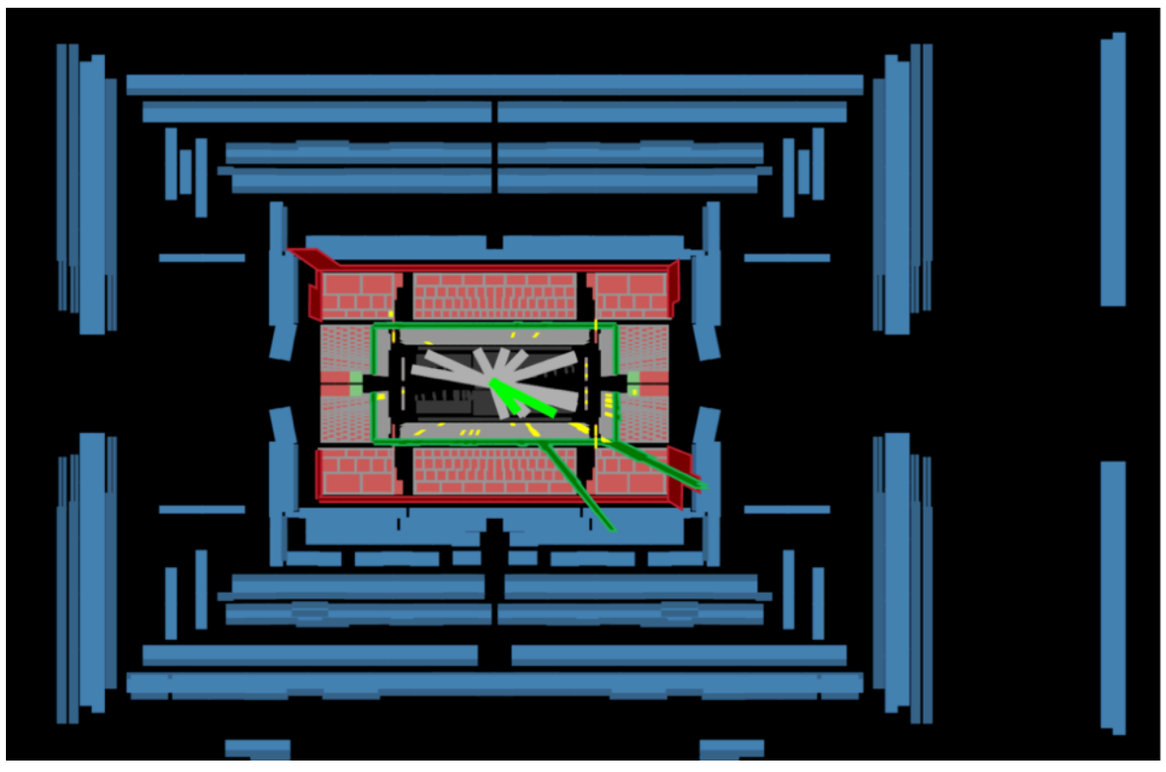
For ttbar Background: A mass variable

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

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

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

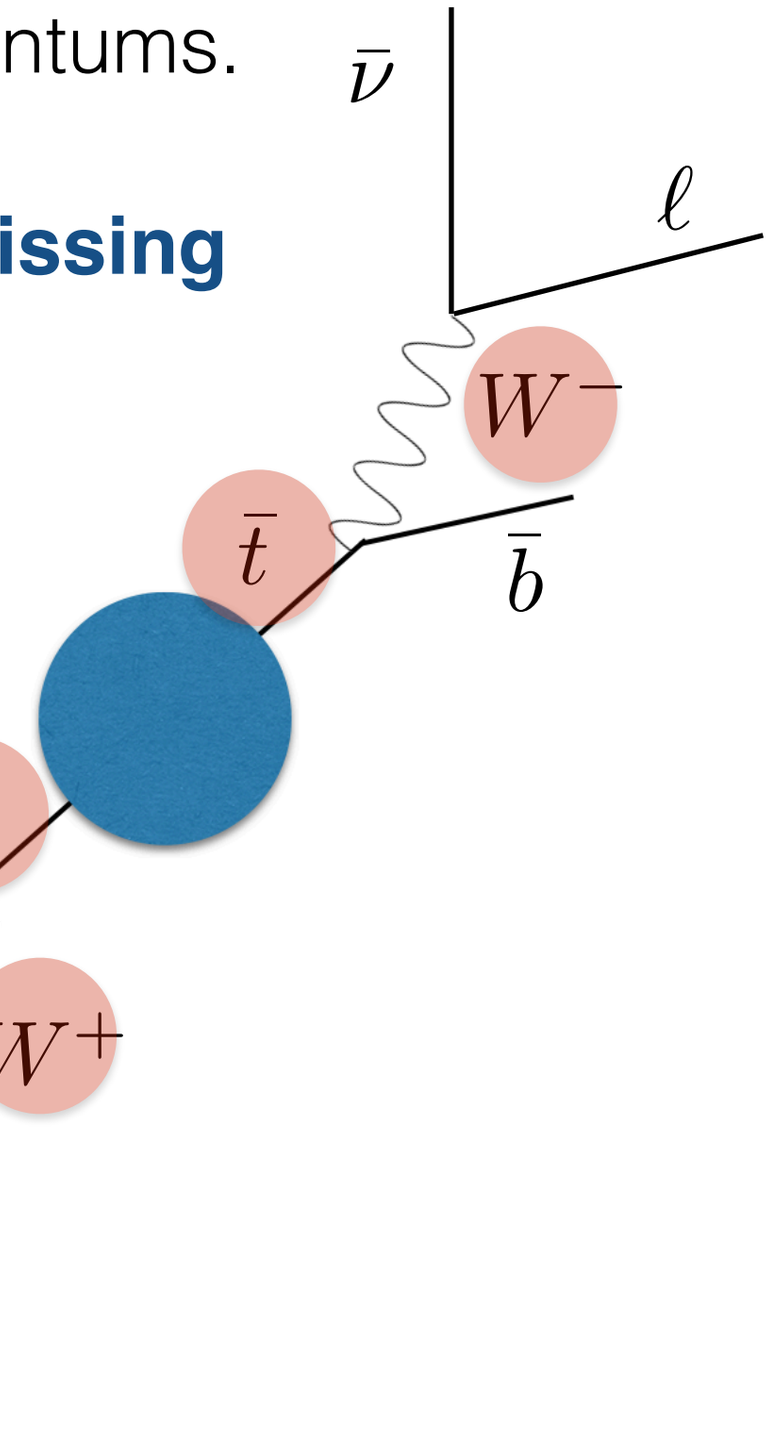




For ttbar Background: A mass variable

- 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

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



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

For HH signal events: Utilizing Mass information

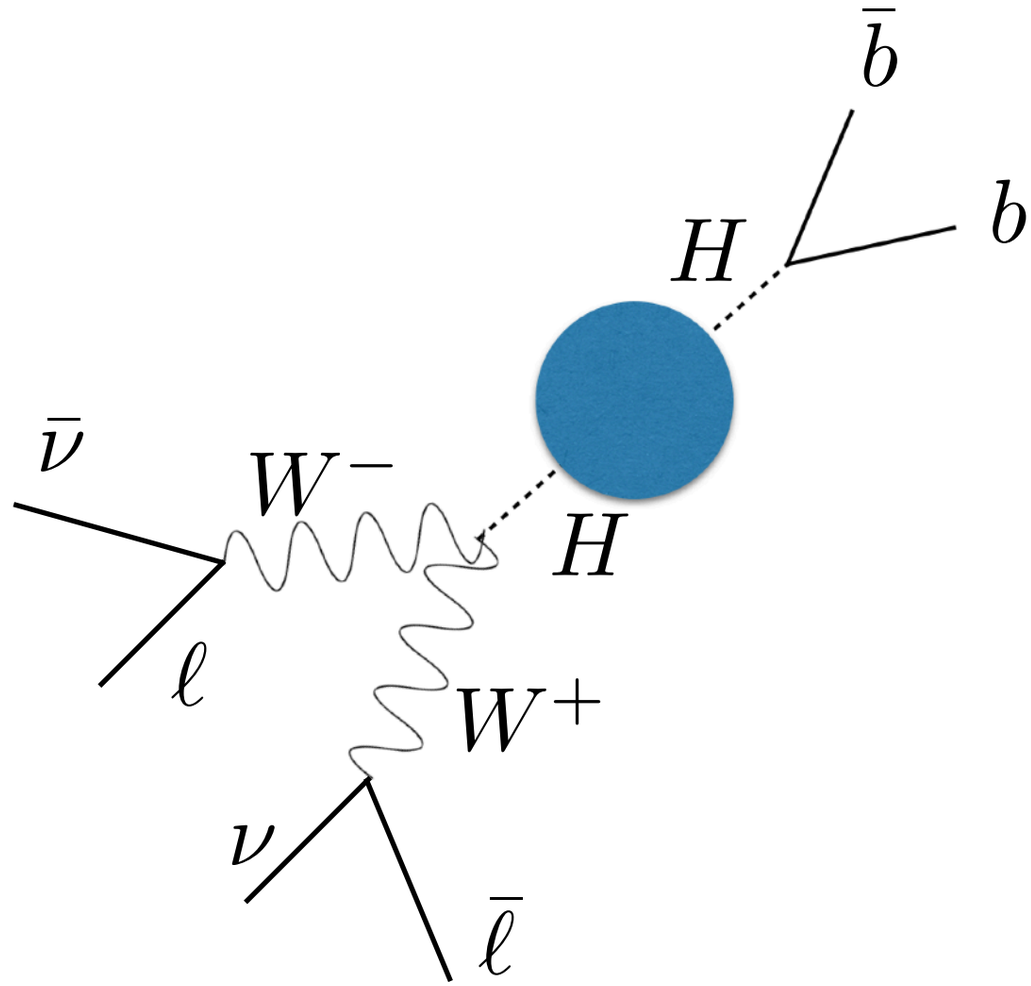
- We have **six unknowns** for two neutrino momentums.

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

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

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

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



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

HH

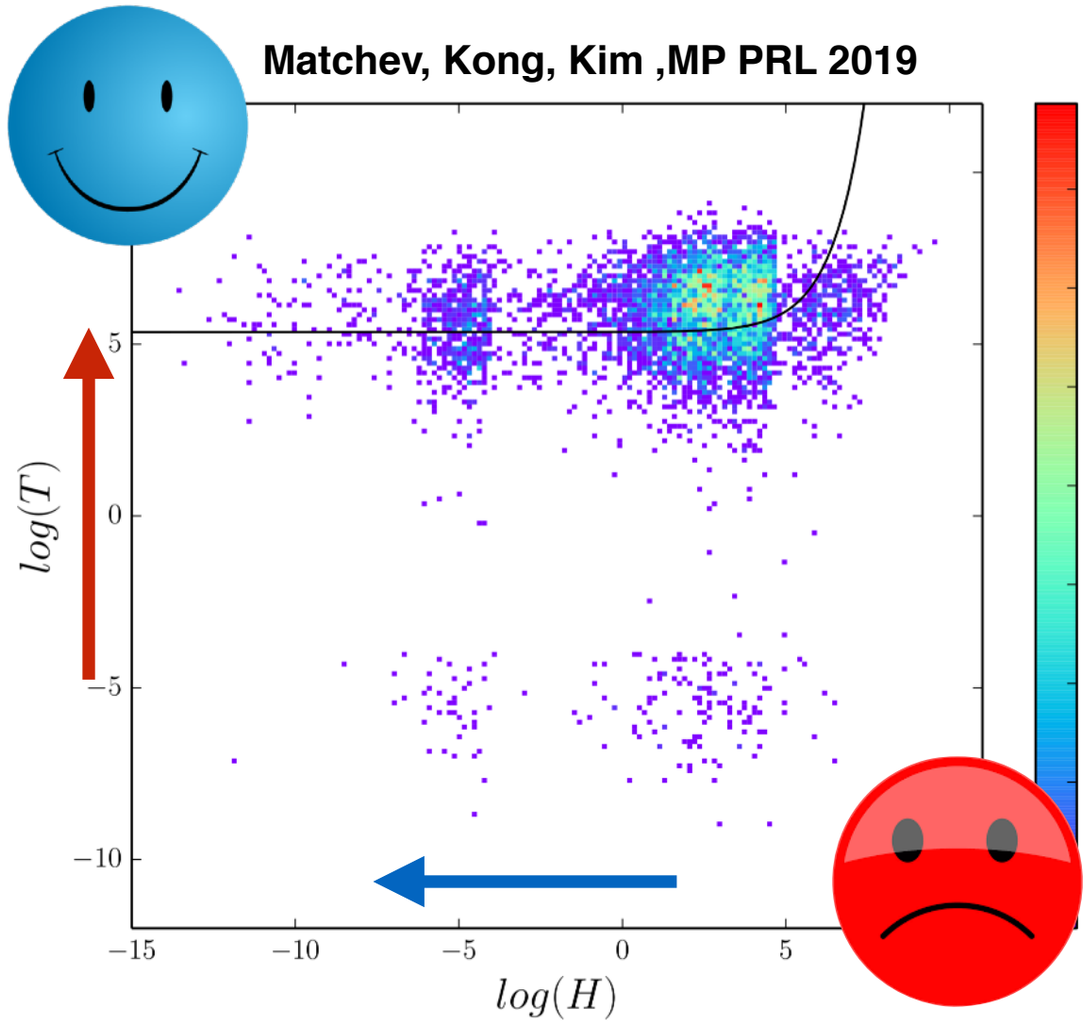
Small H (Higgs-ness)
compatible with a **Higgs-topology**

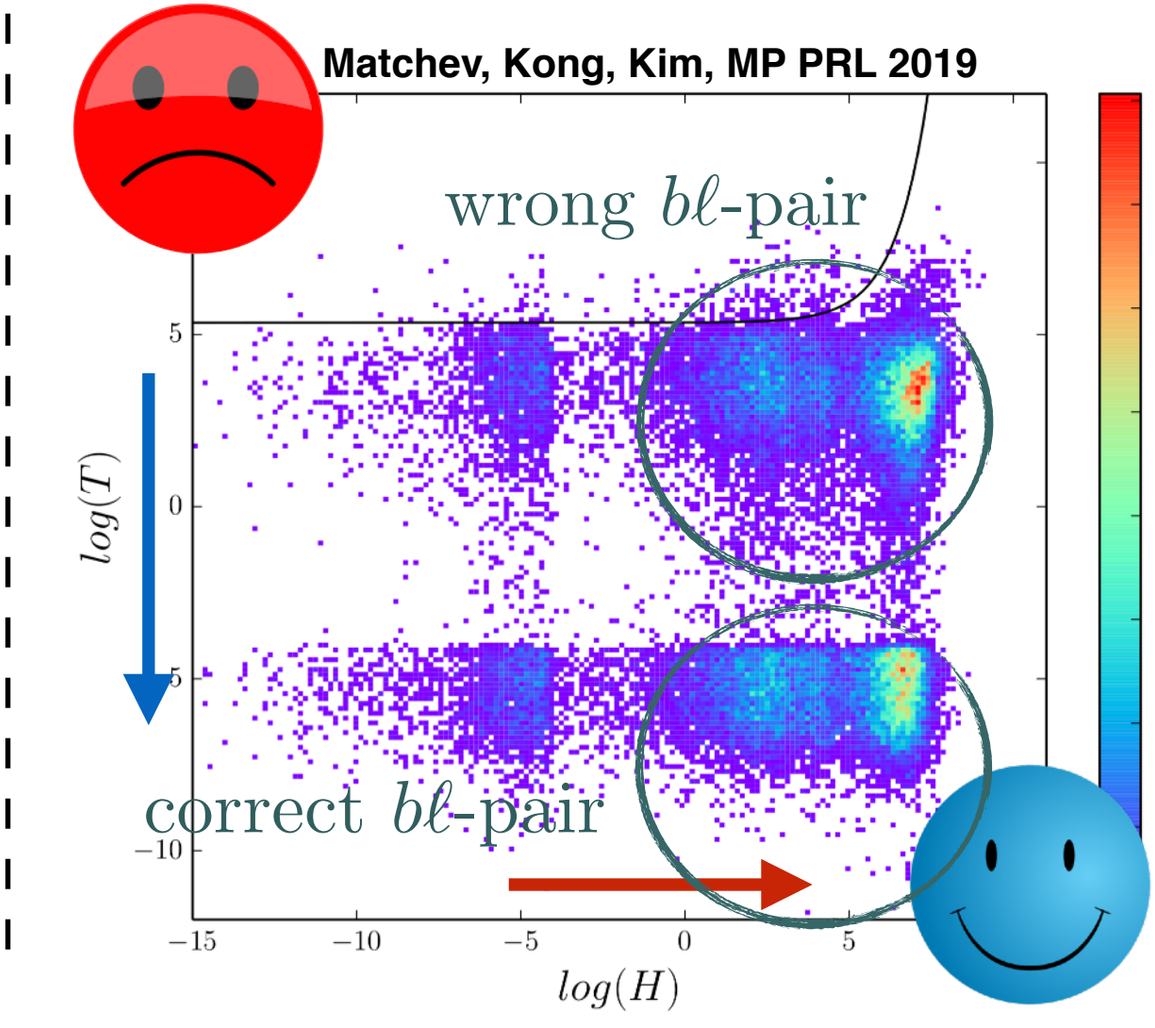
Large χ_{ij} (Top-ness)
NOT compatible with a $t\bar{t}$ -**topology**

$t\bar{t}$

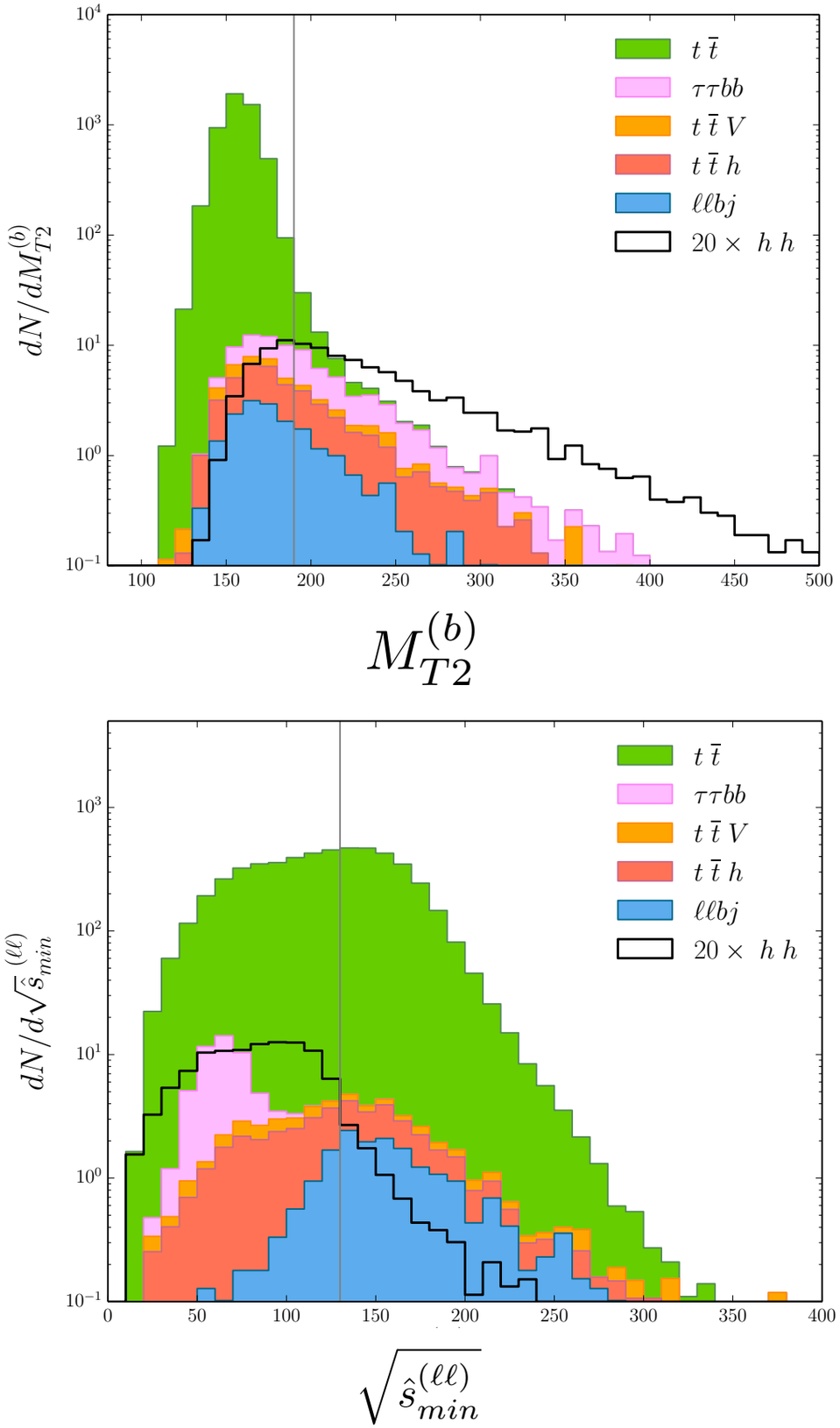
Large H (Higgs-ness)
NOT compatible with a **Higgs-topology**

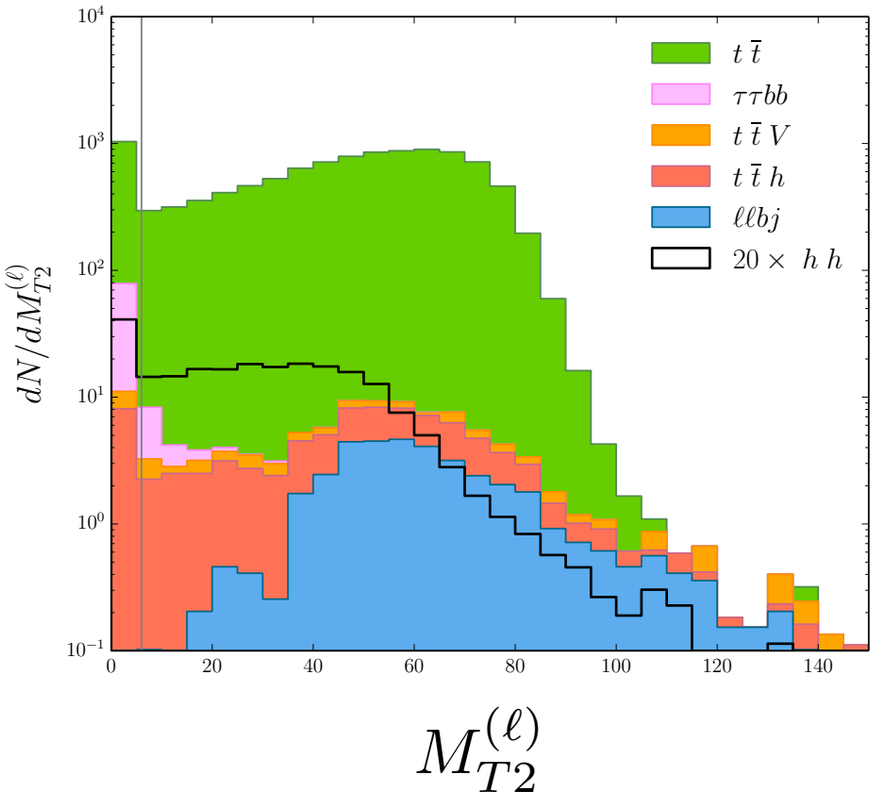
Small χ_{ij} (Top-ness)
compatible with a $t\bar{t}$ -**topology**





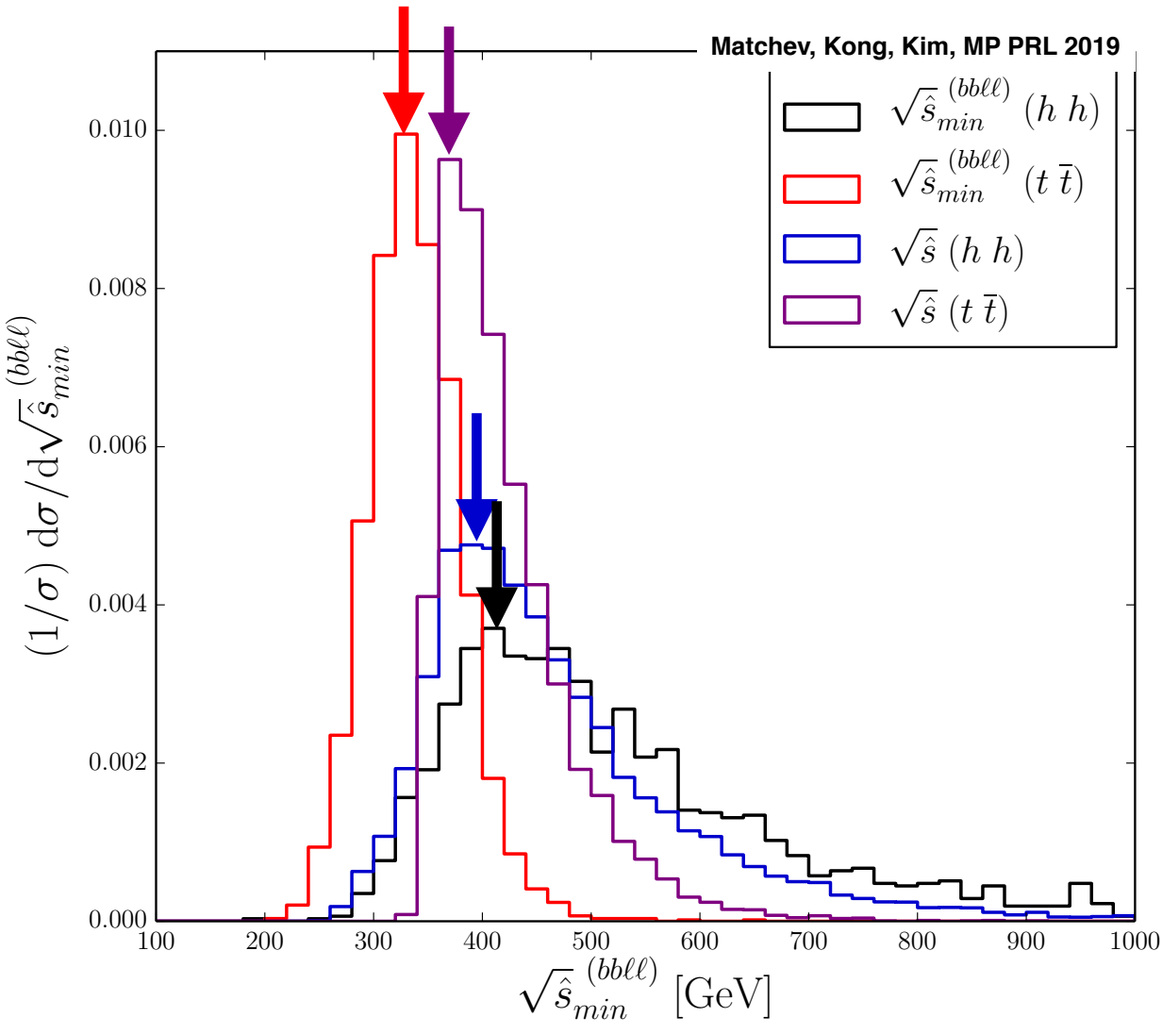
Conventional "**super**-cuts" (feature variables)



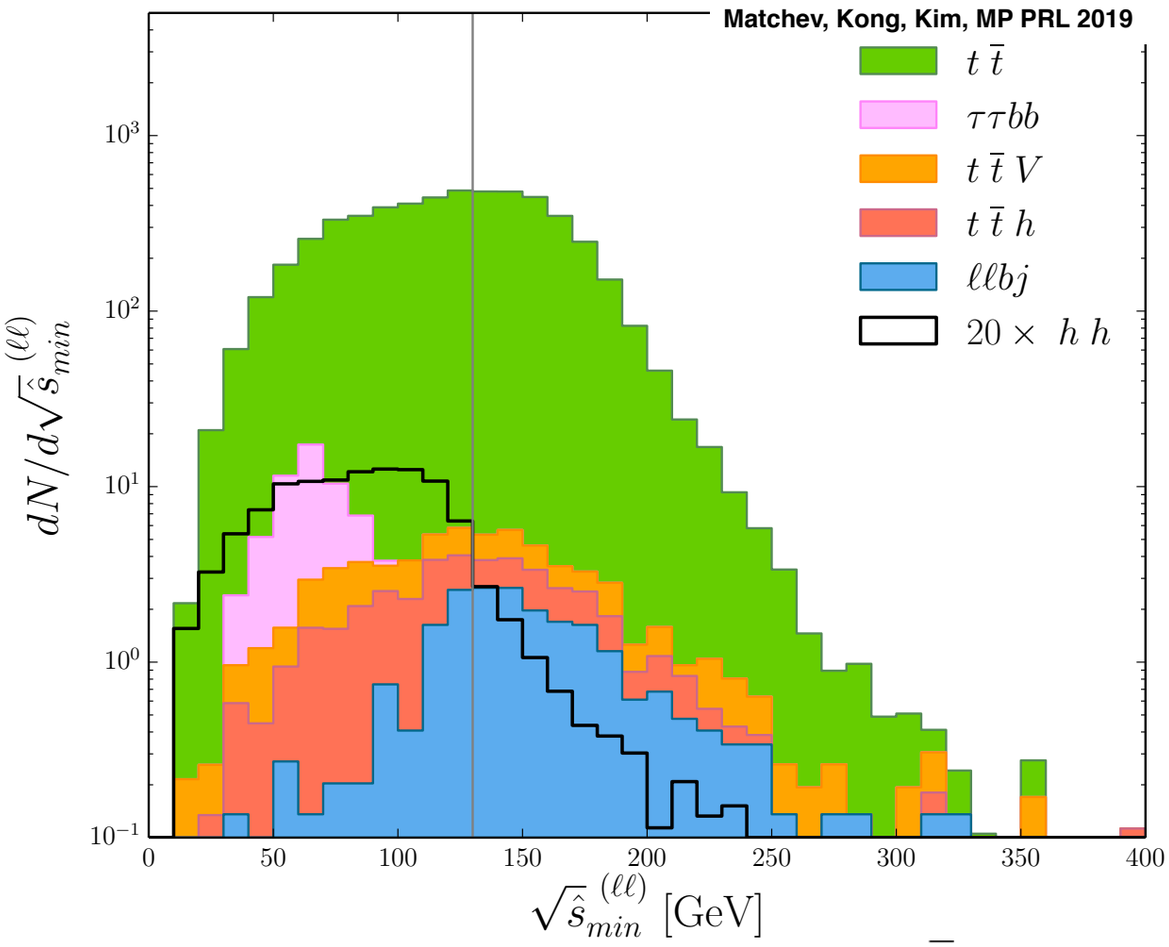


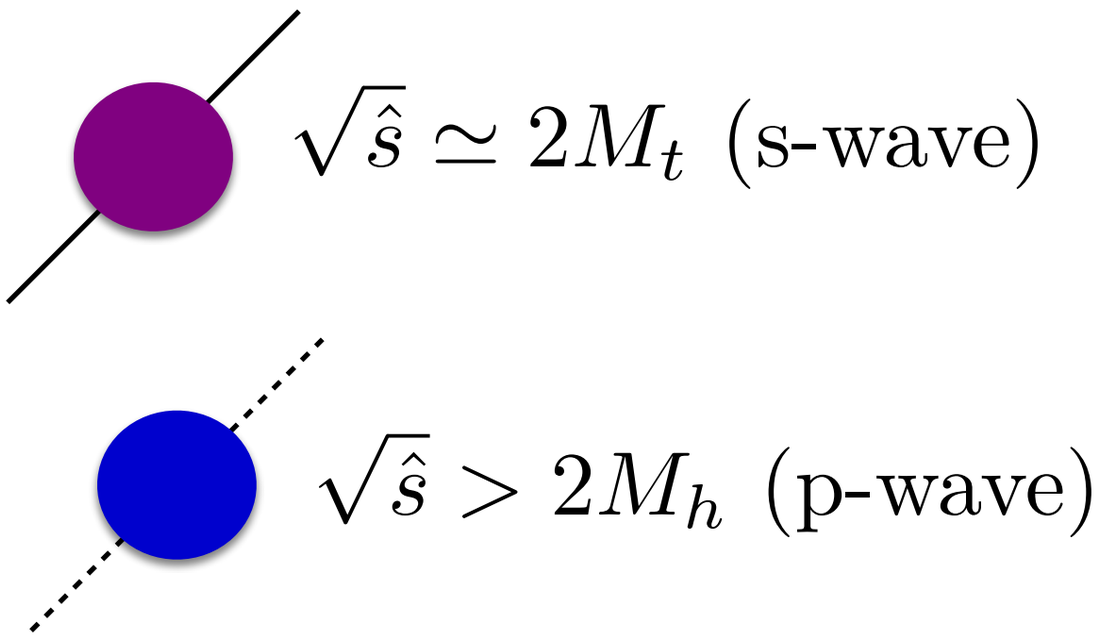
- $M_{T2}^{(b)}$ is designed to specifically kill $t\bar{t}$.
- $M_{T2}^{(\ell)}$ is designed to specifically kill $\tau\tau b\bar{b}$.
- $\sqrt{\hat{s}_{min}^{(\ell\ell)}}$ can be used to suppress $t\bar{t}$ further.

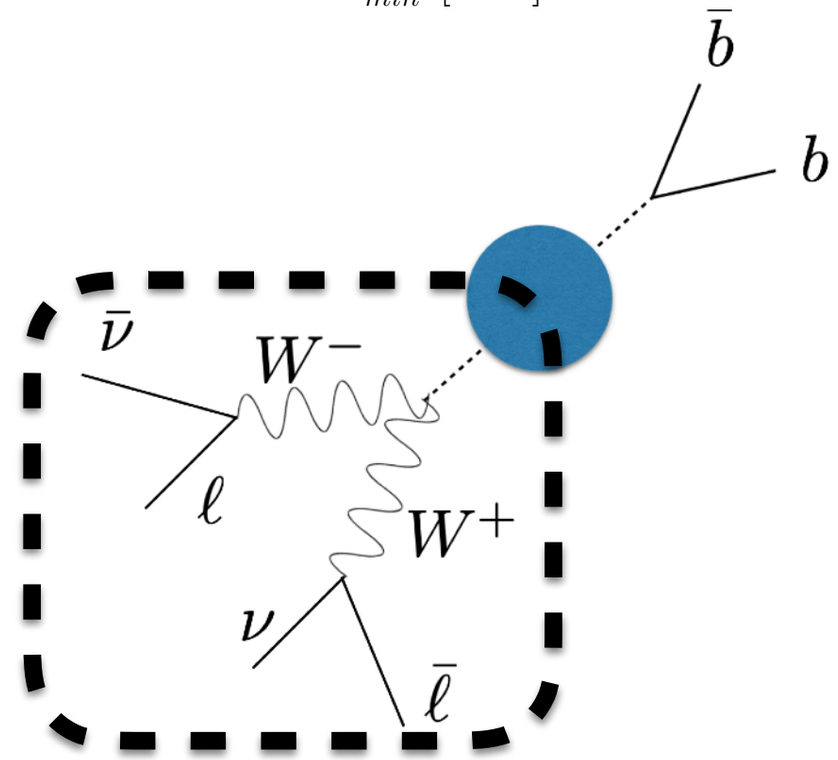
"Production Energy" variable



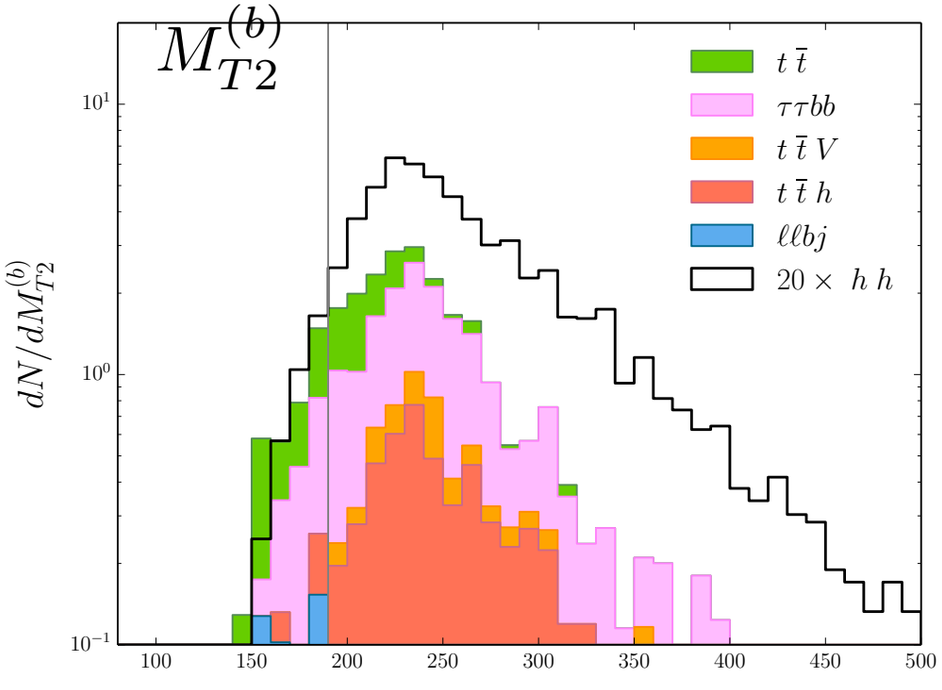
$$\sqrt{\hat{s}}(\text{Higgs decays}) \gtrapprox M_h$$

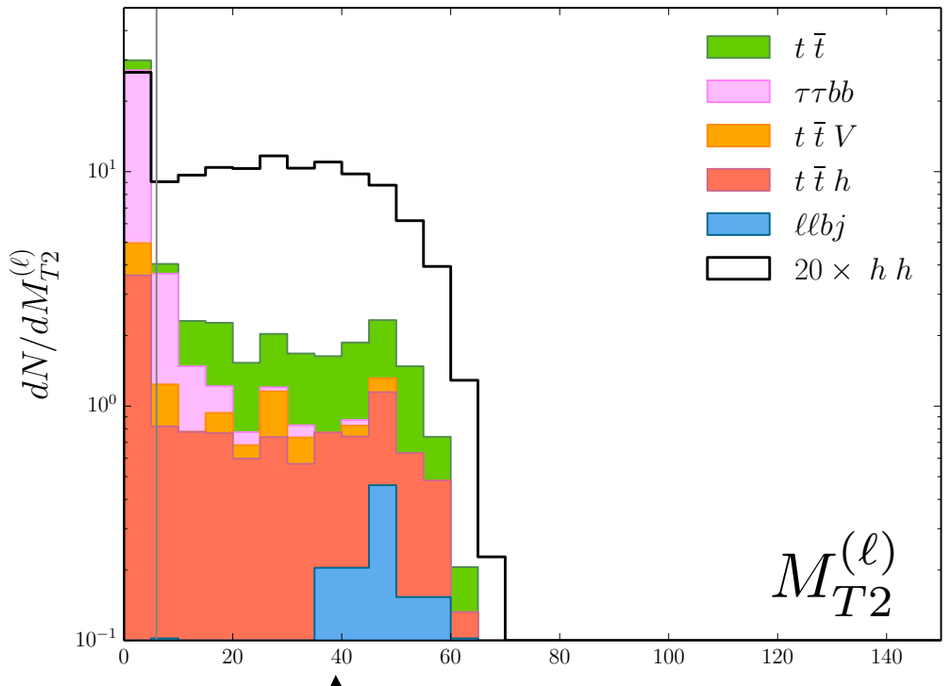


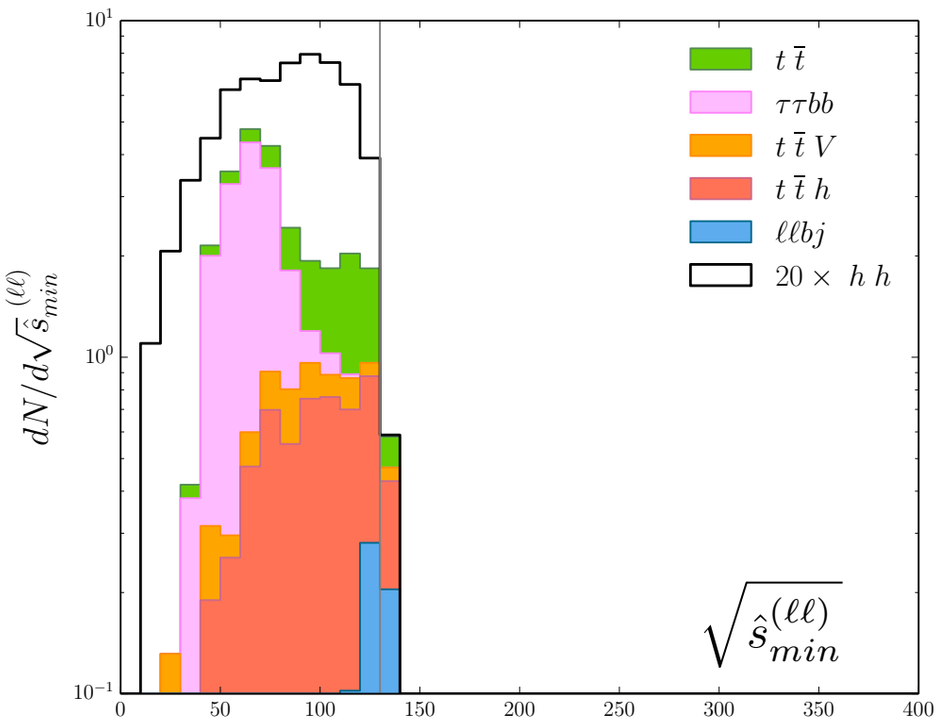




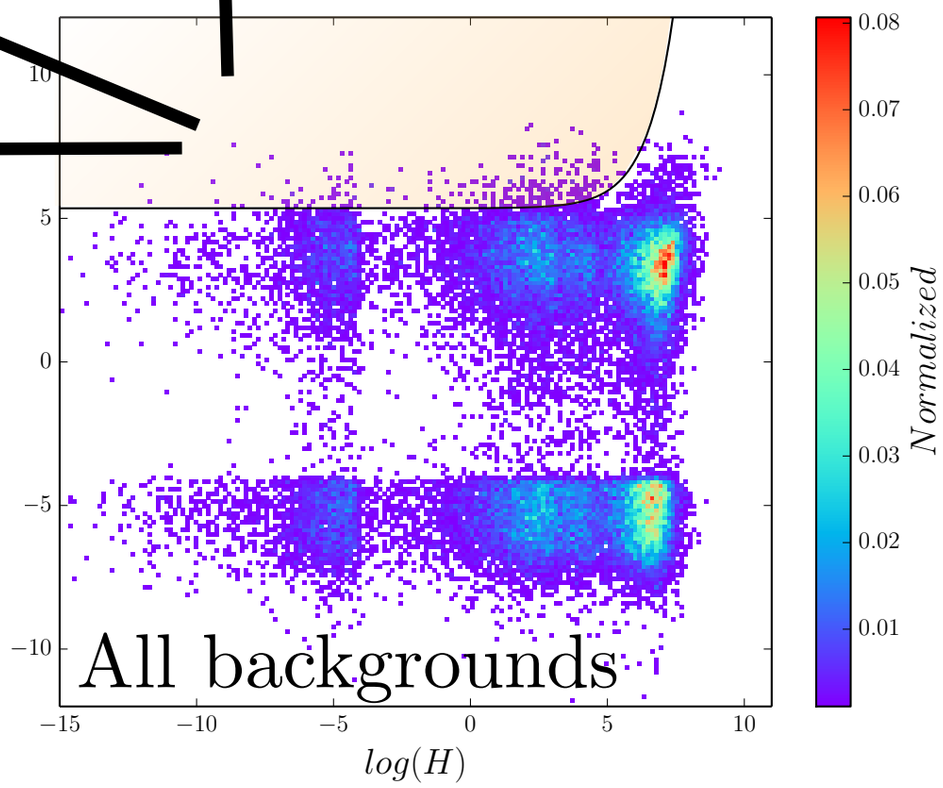
BKGs which are **compatible to signal hypothesis**



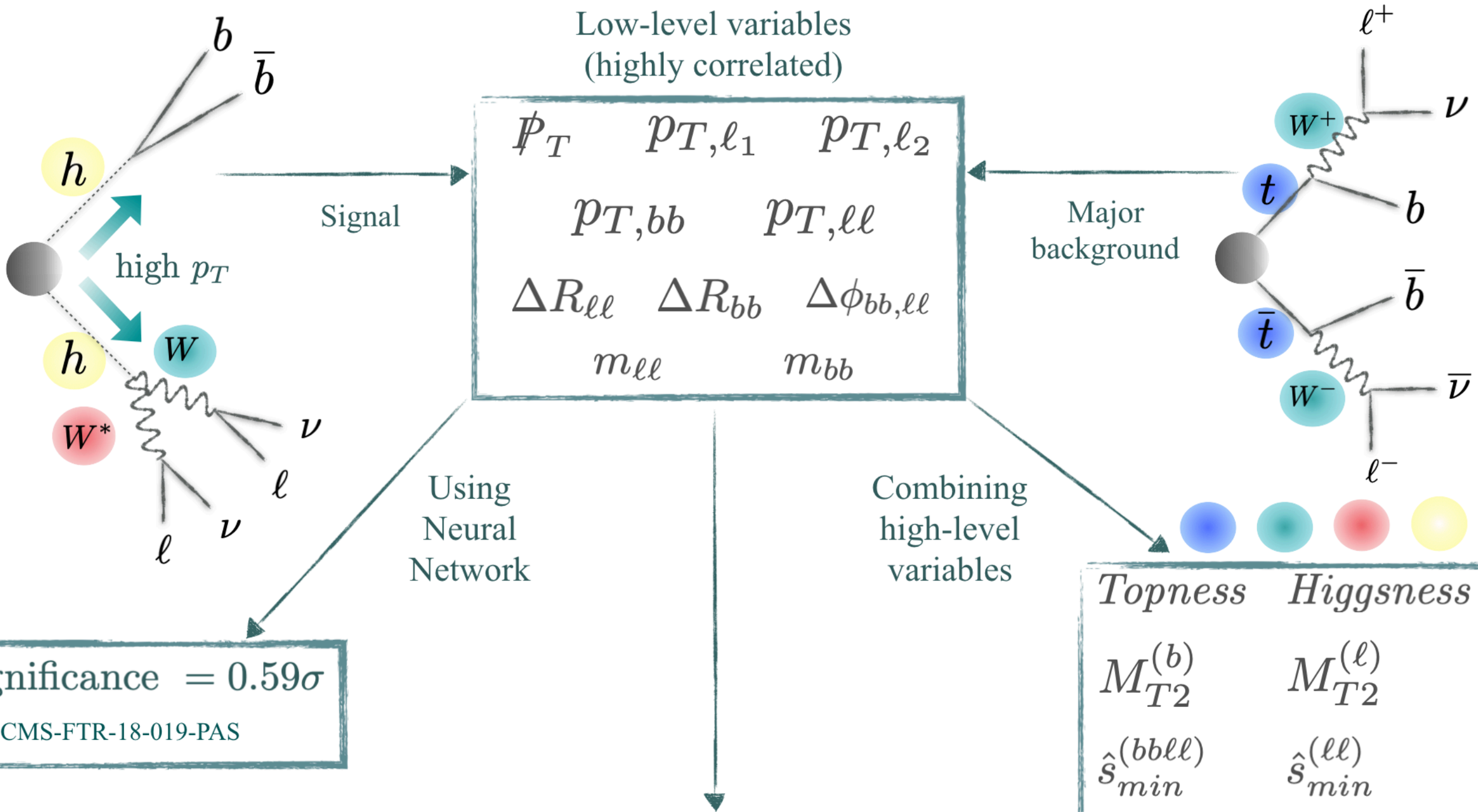




The power of "super-cuts" is **reduced**



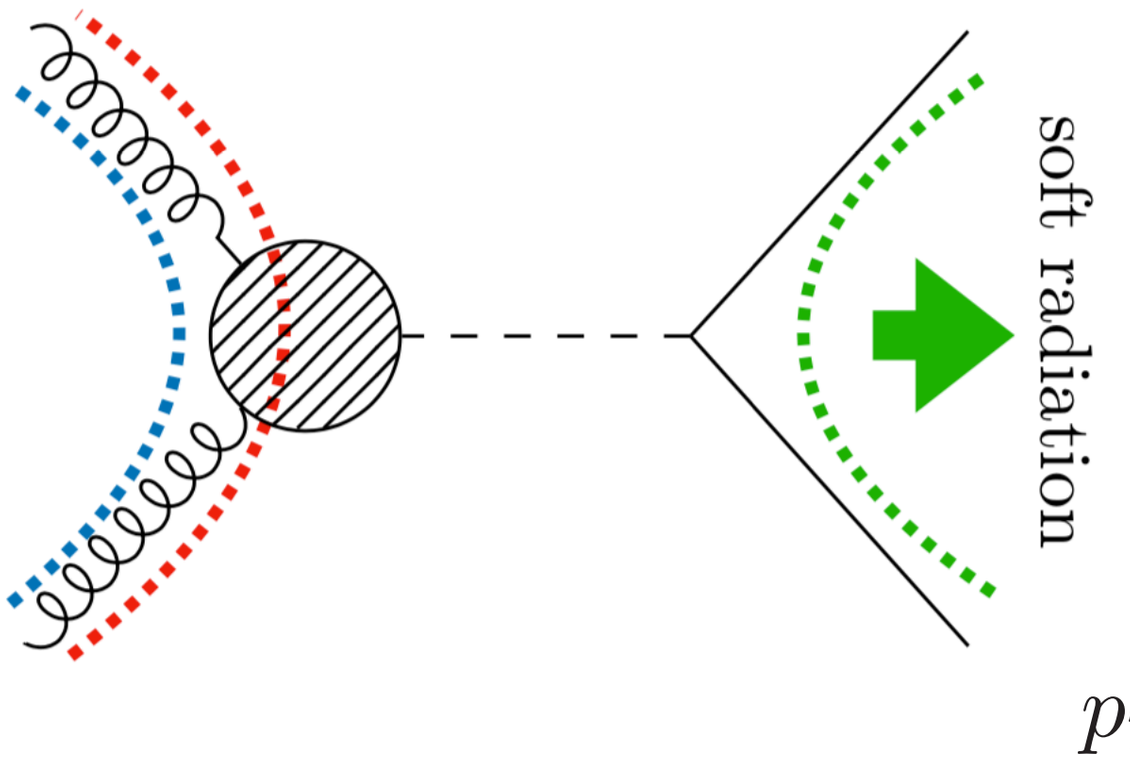
How to rescue *bbWW** ?

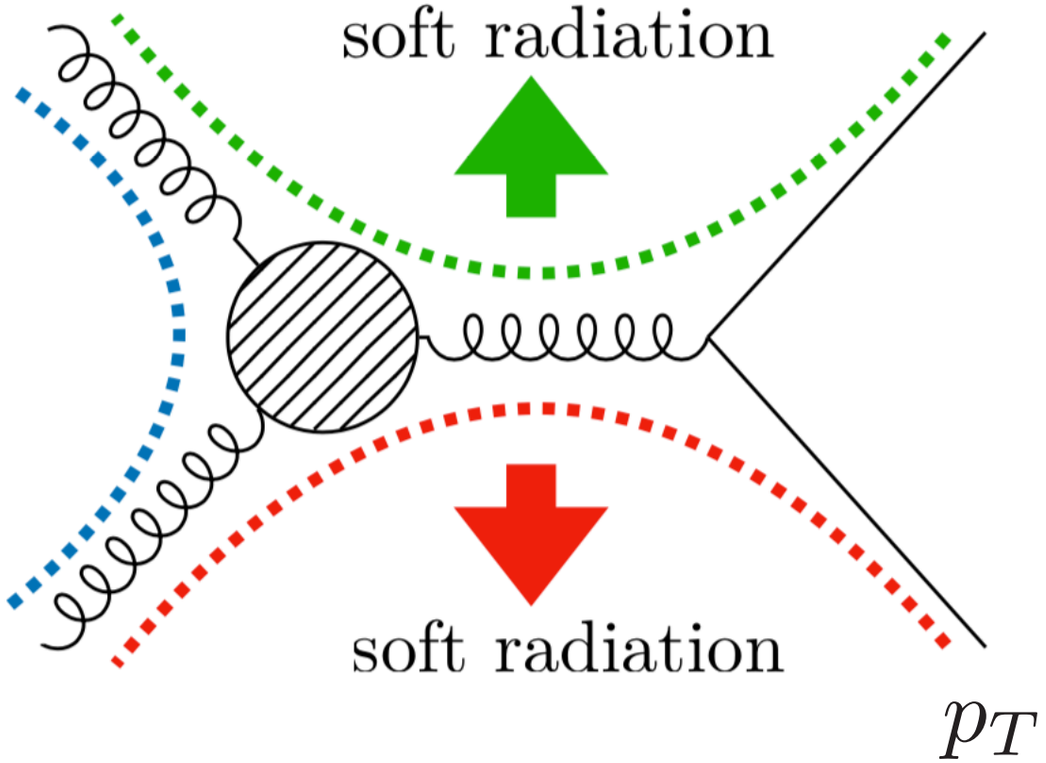


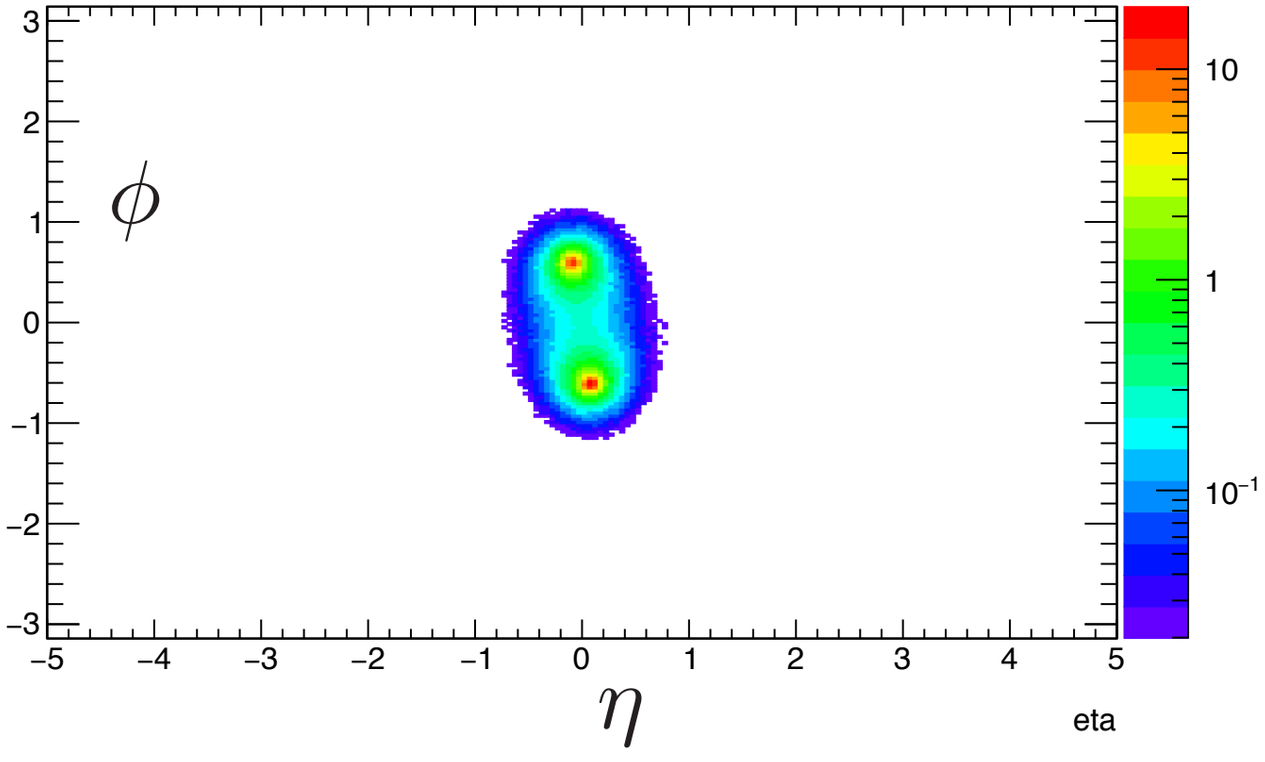
We utilize jet images, and let the machine deal with correlations.

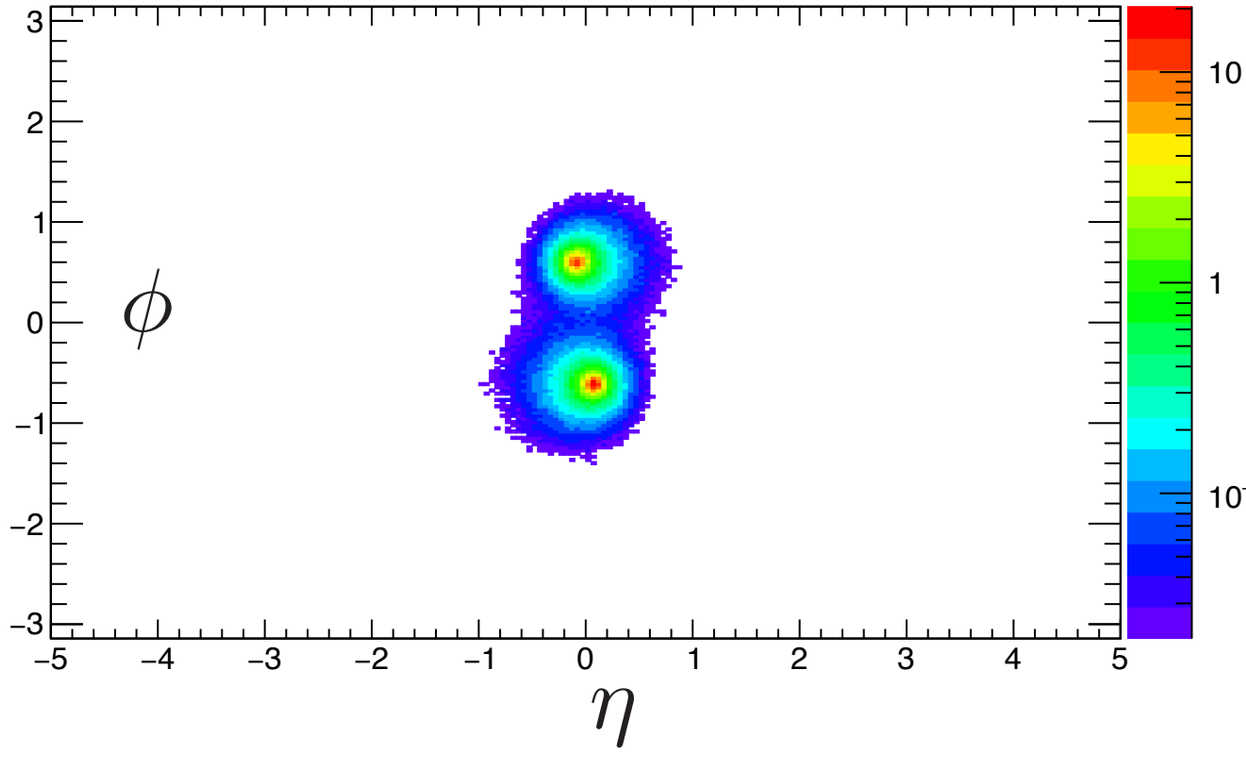
J. H. Kim, K. C. Kong, K. T. Matchev, M. Park [2018]

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



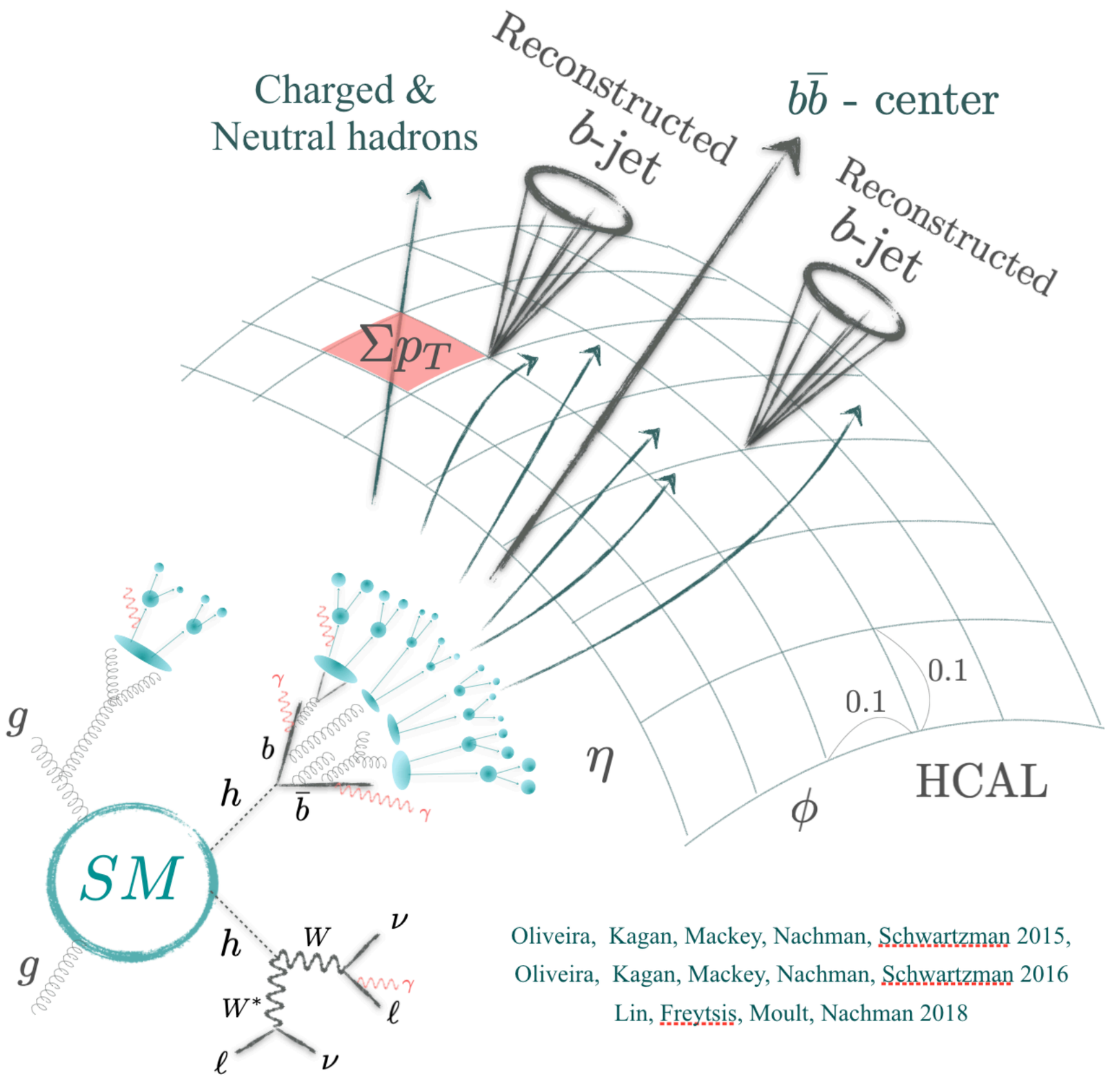






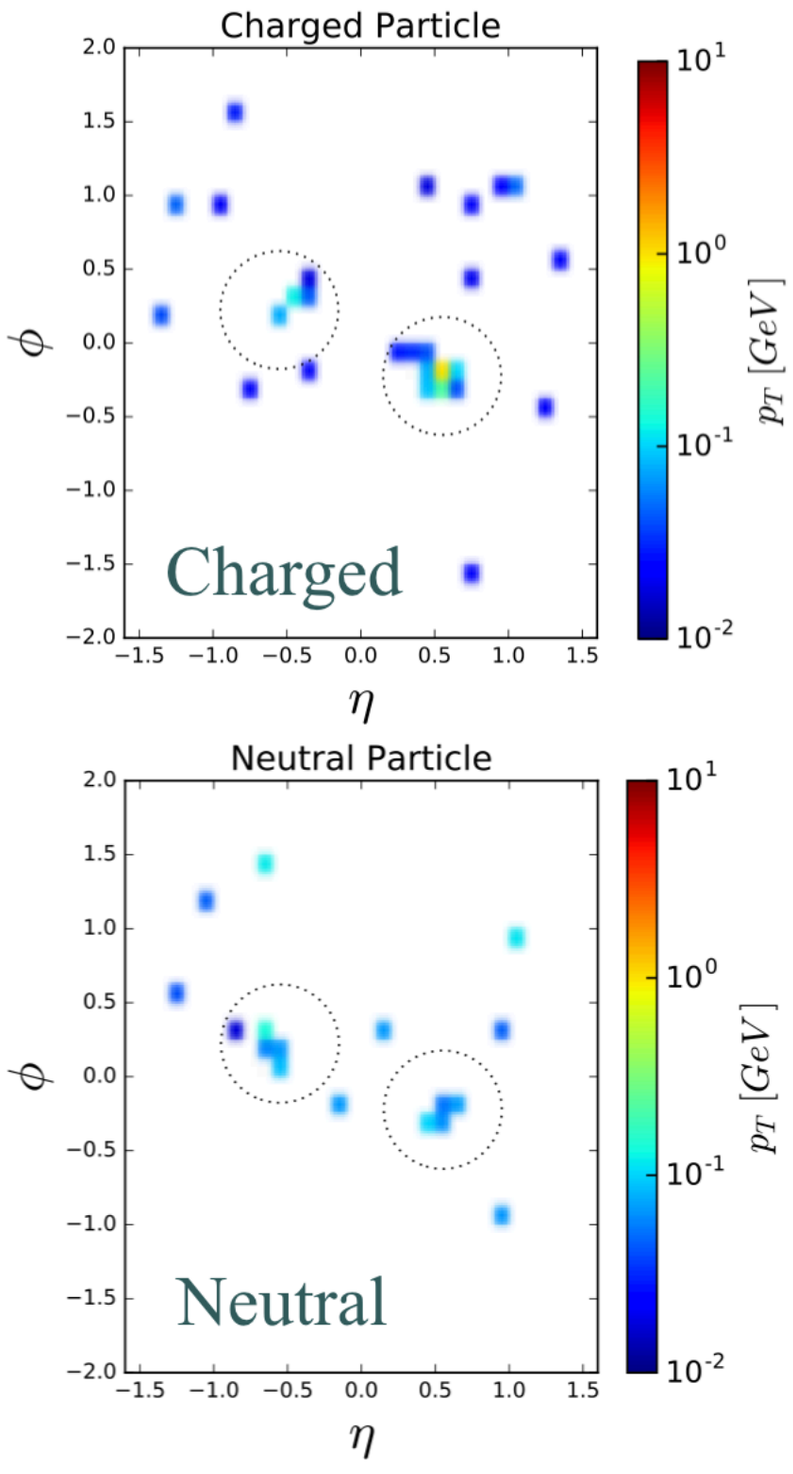
Energy deposits

Processing Hadron Images (hh)



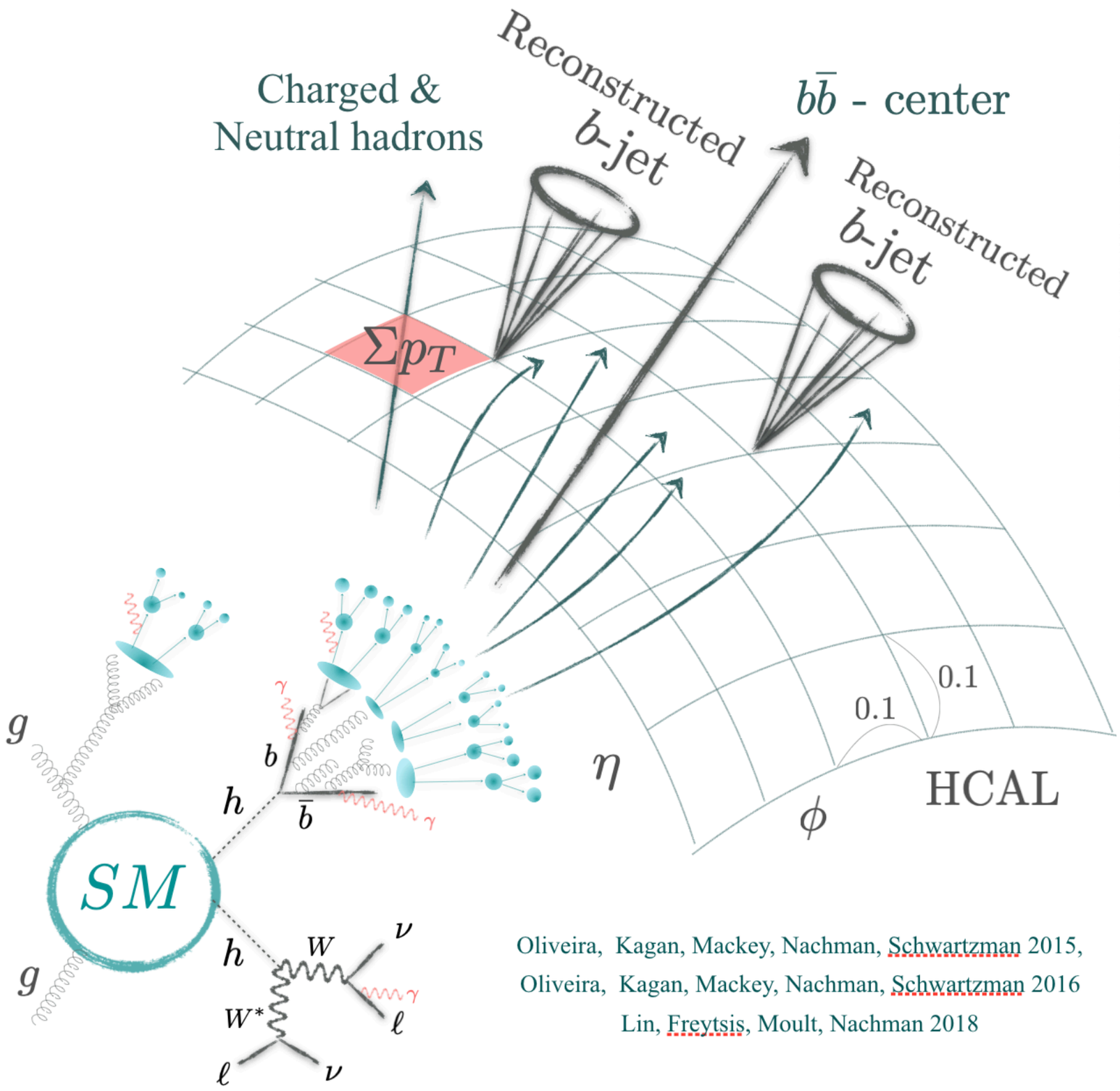
Oliveira, Kagan, Mackey, Nachman, Schwartzman 2015,
Oliveira, Kagan, Mackey, Nachman, Schwartzman 2016
Lin, Freytsis, Moult, Nachman 2018

Each event

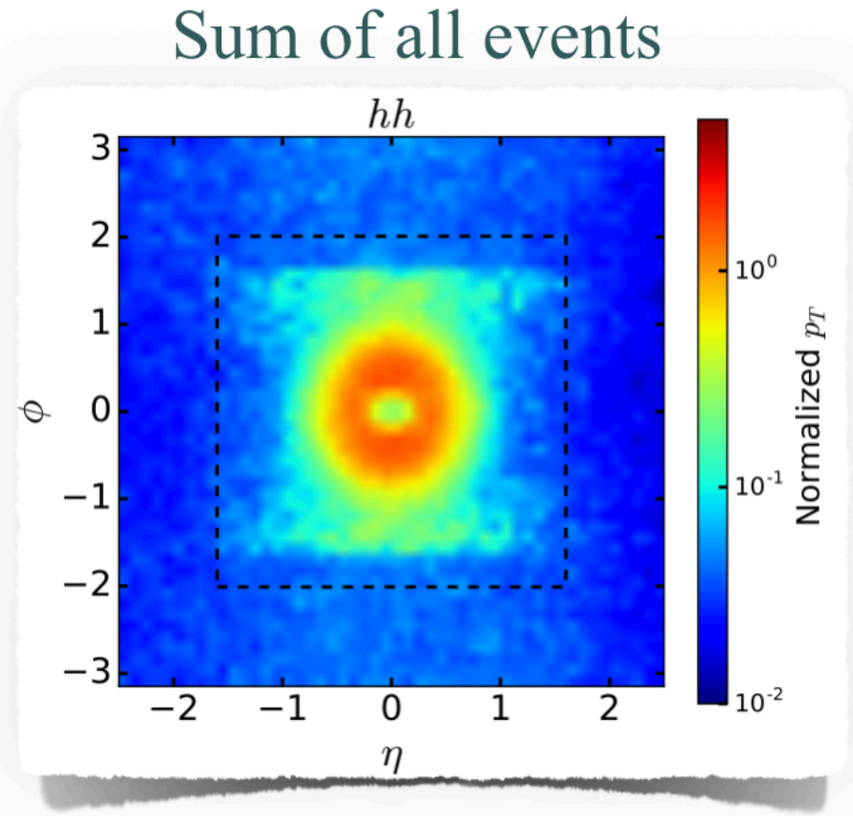


Kim, Kong, Matchev, Park JHEP 2019

Processing Hadron Images (hh)

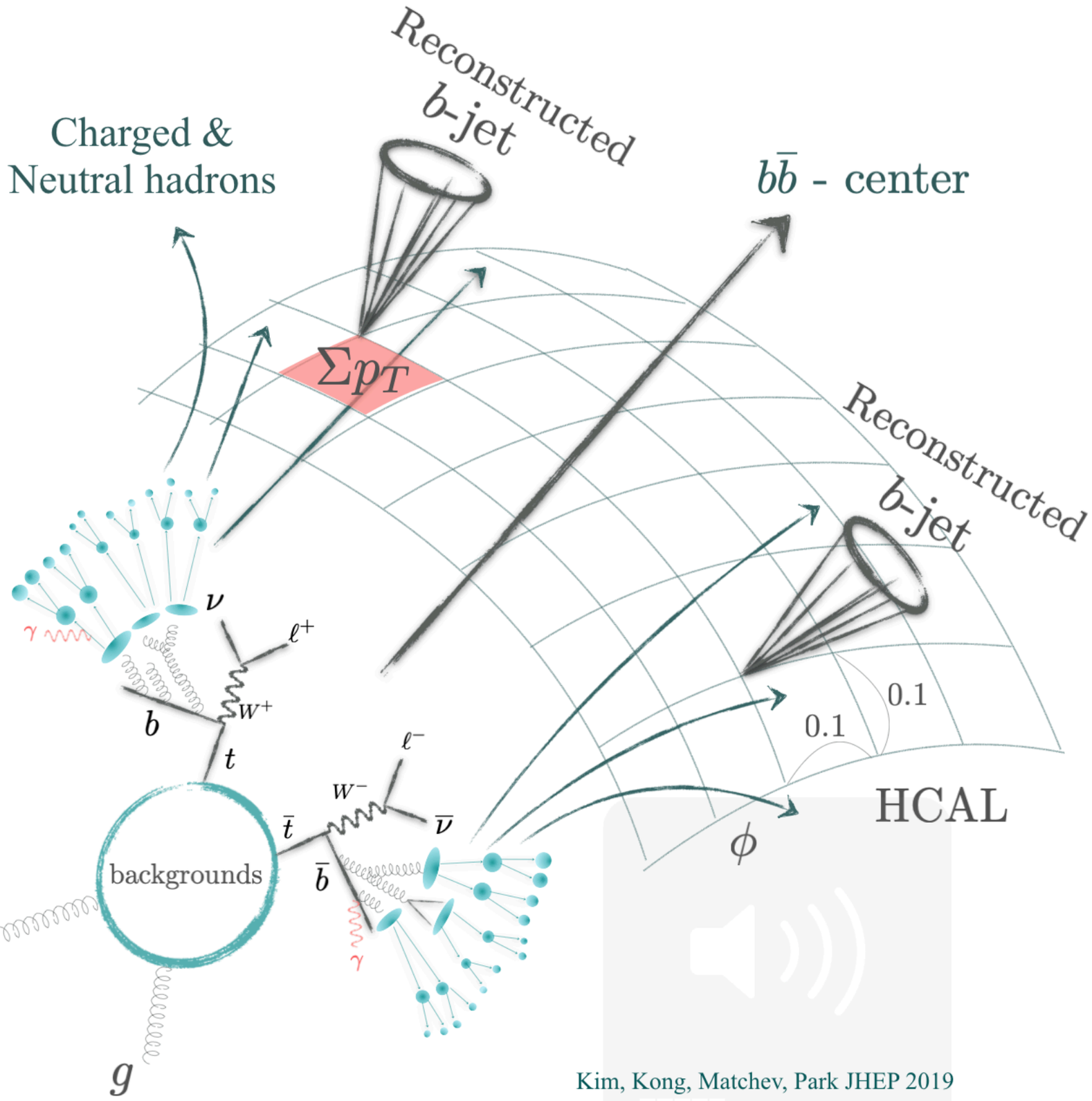


Oliveira, Kagan, Mackey, Nachman, Schwartzman 2015,
Oliveira, Kagan, Mackey, Nachman, Schwartzman 2016
Lin, Freytsis, Moult, Nachman 2018



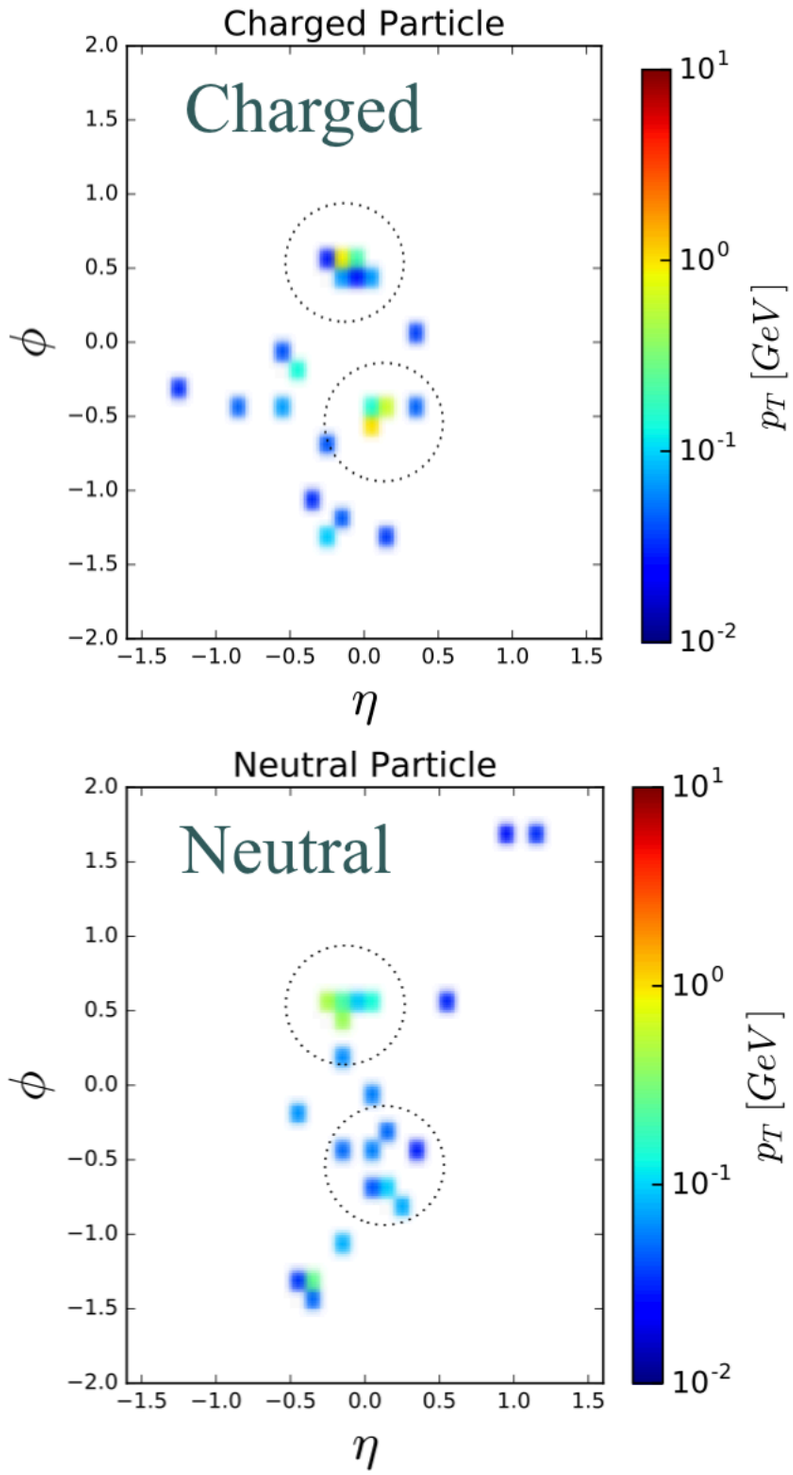
Kim, Kong, Matchev, Park JHEP 2019

Processing Hadron Images ($t\bar{t}$)

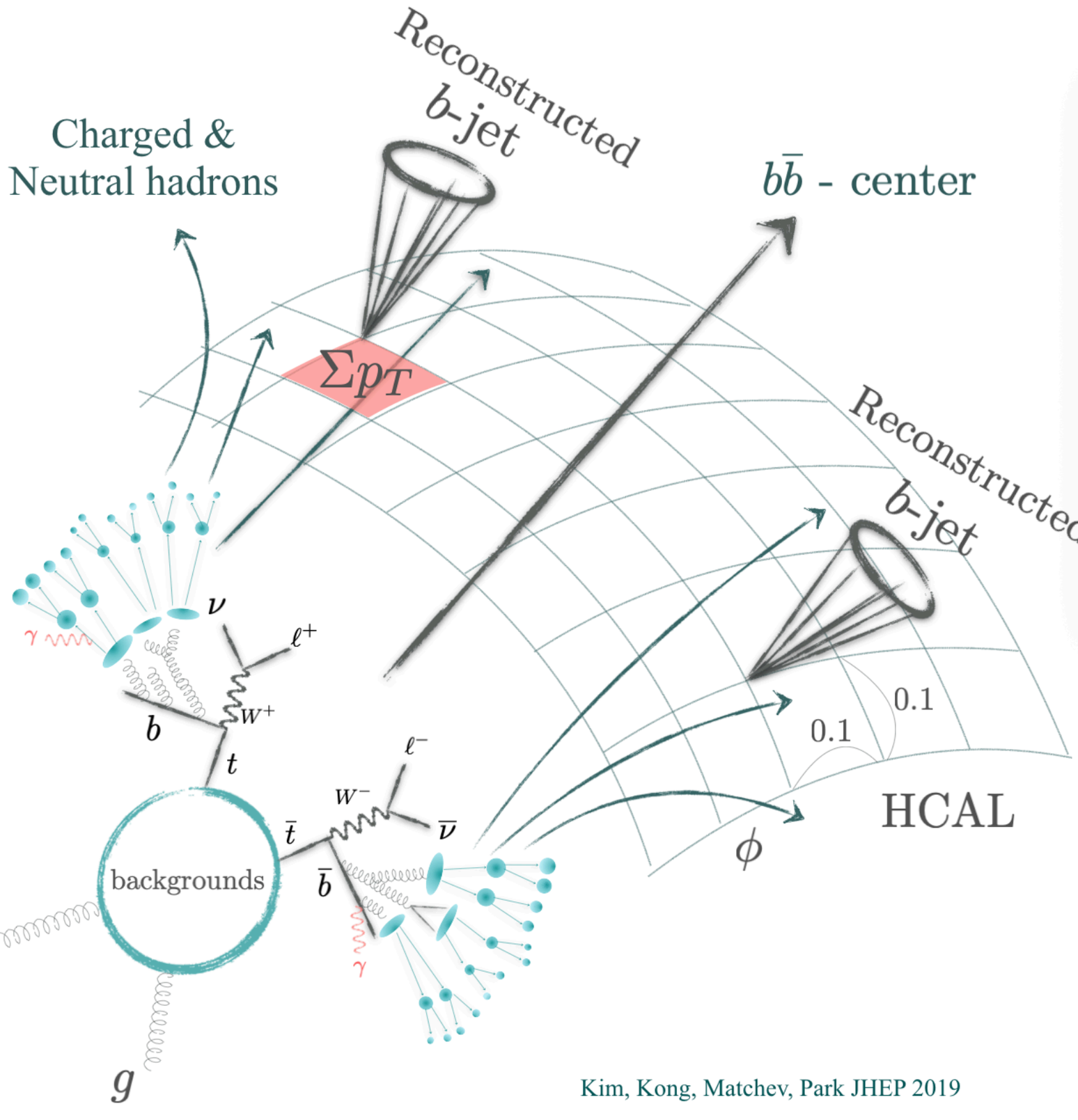


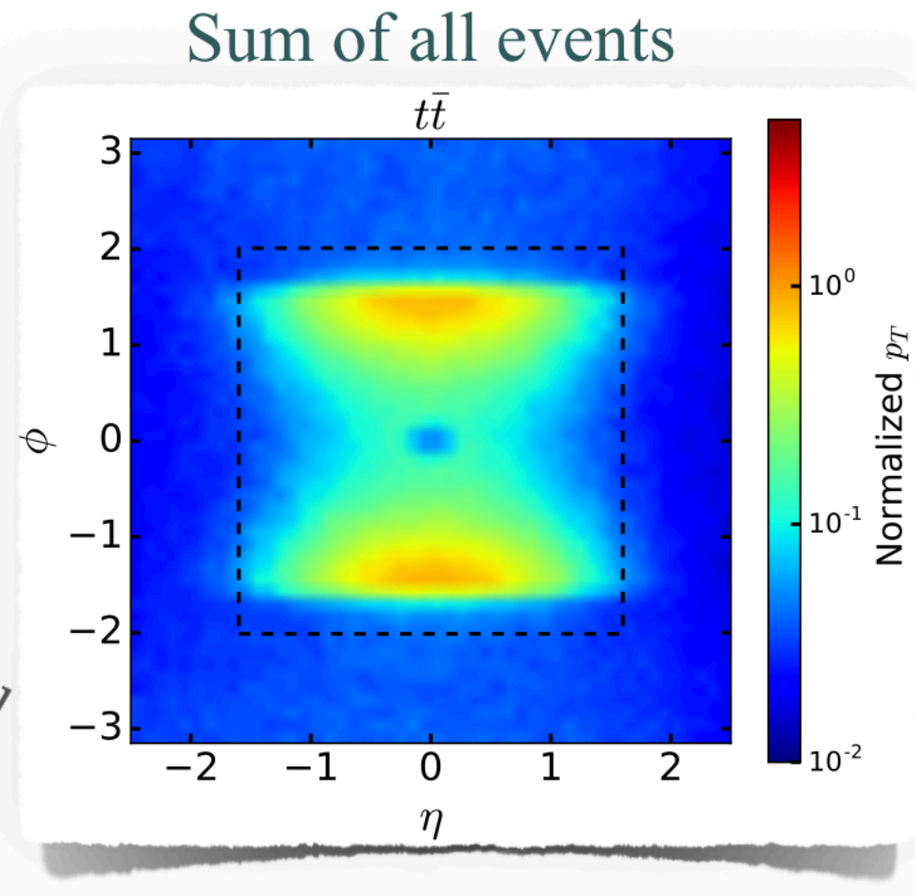
Kim, Kong, Matchev, Park JHEP 2019

Each event



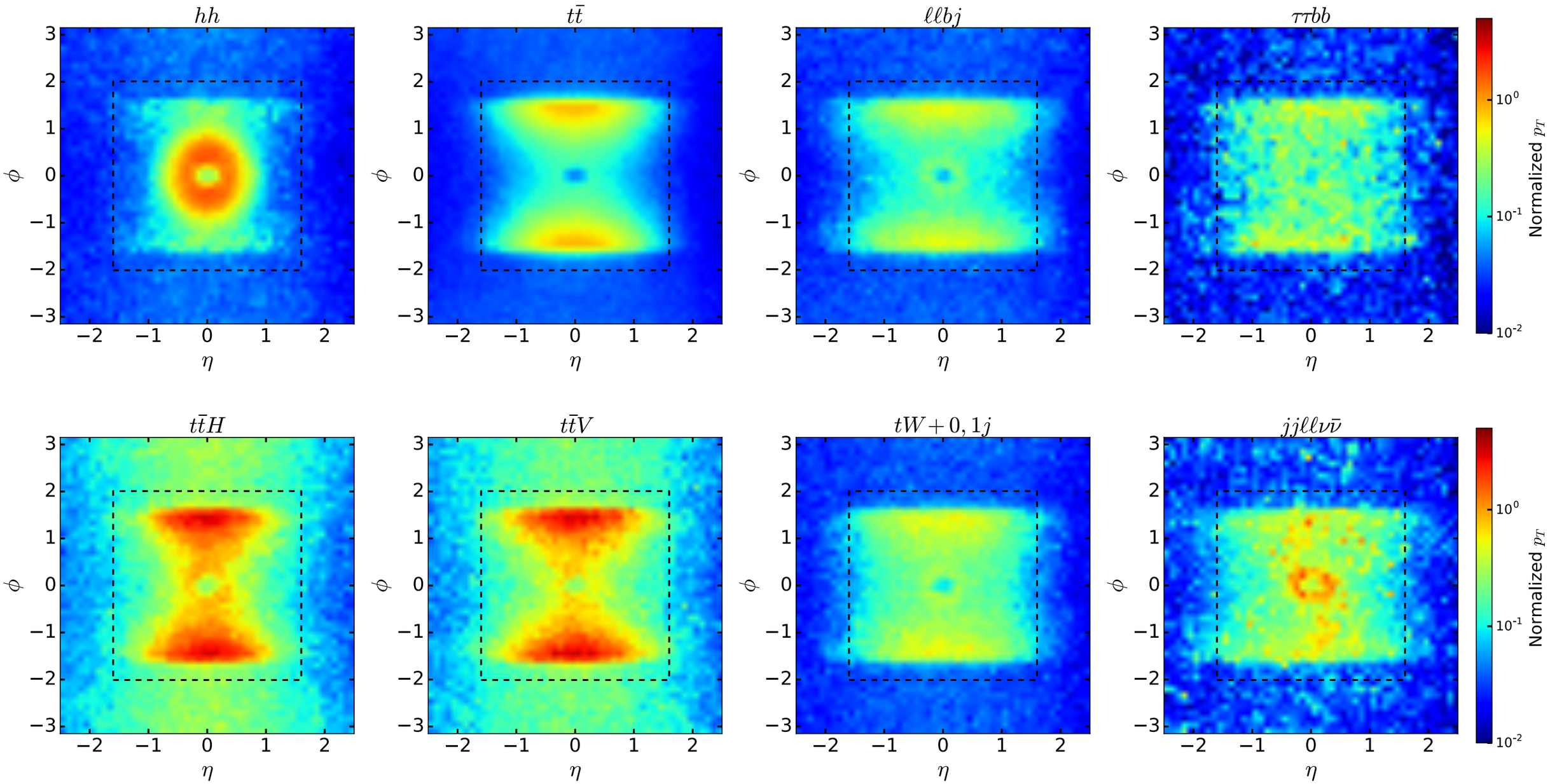
Processing Hadron Images ($t\bar{t}$)



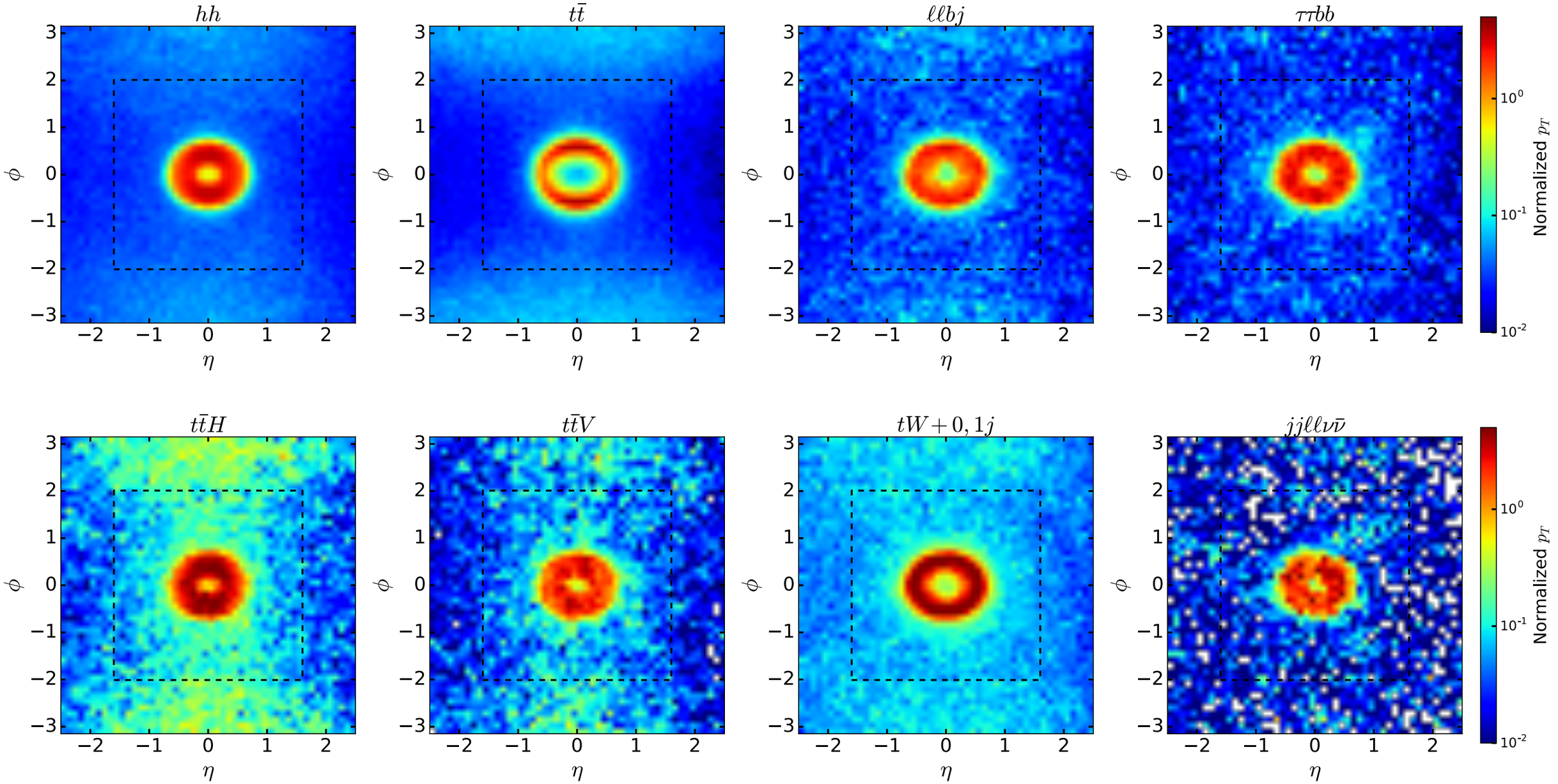


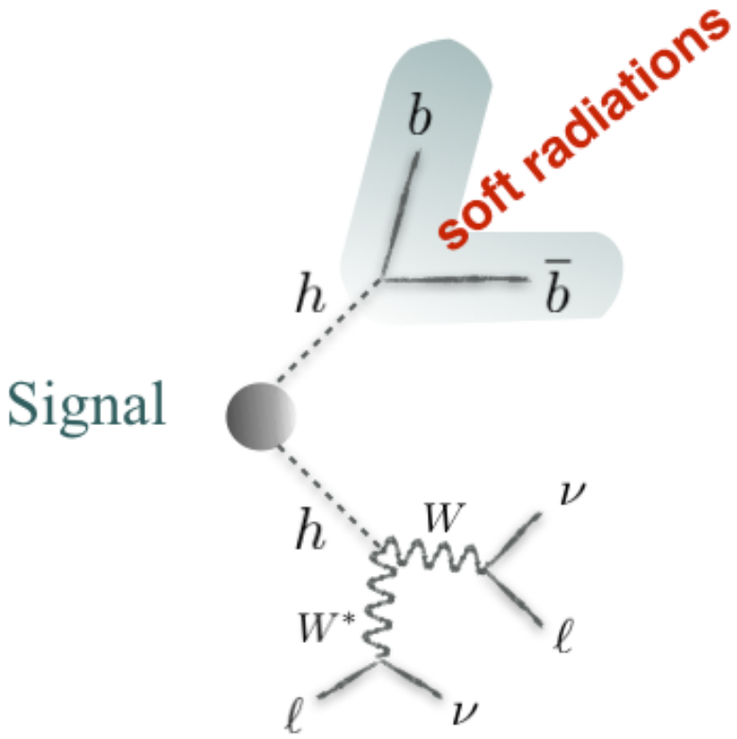
Kim, Kong, Matchev, Park JHEP 2019

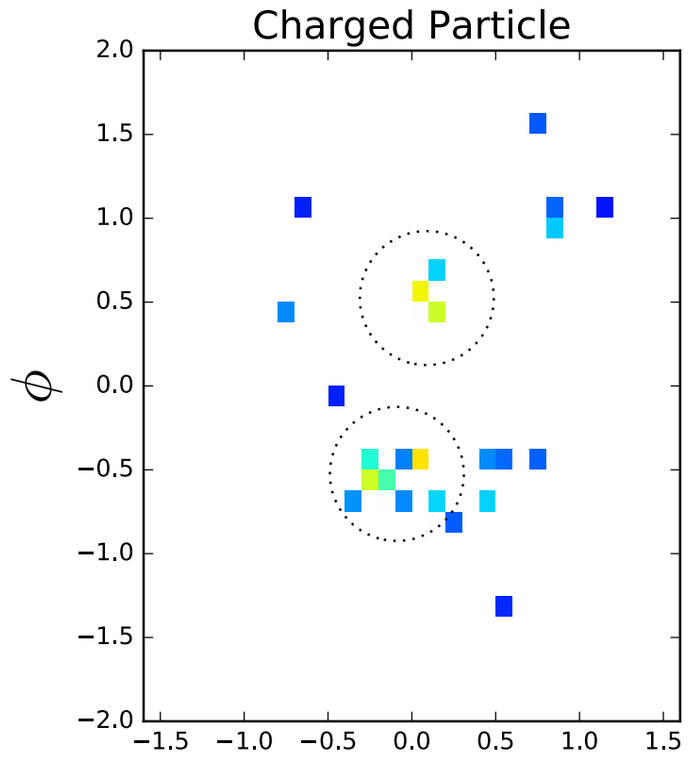
Jet images before baseline cuts

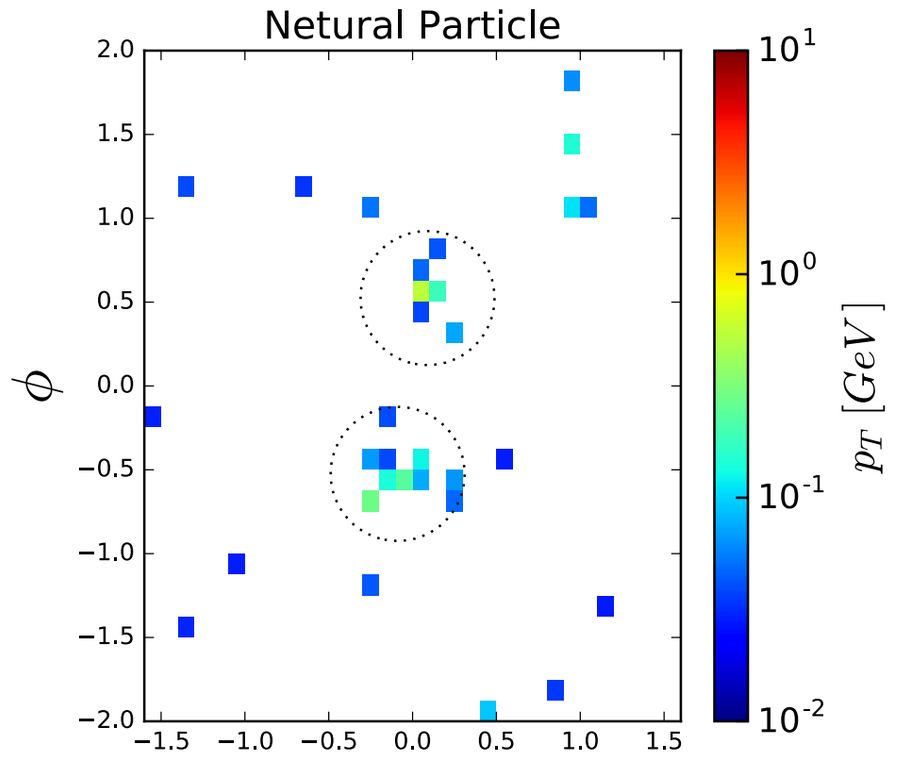


Jet images after baseline cuts

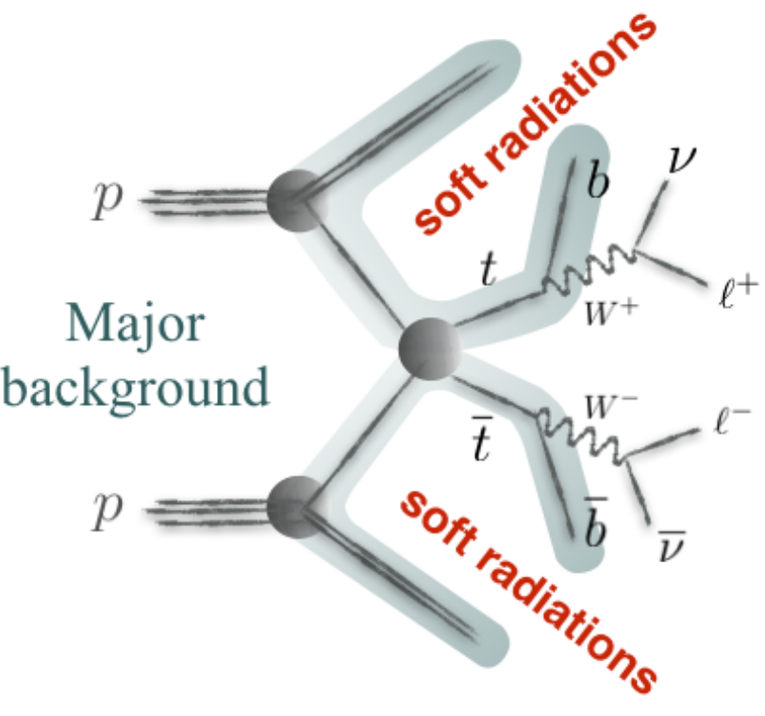


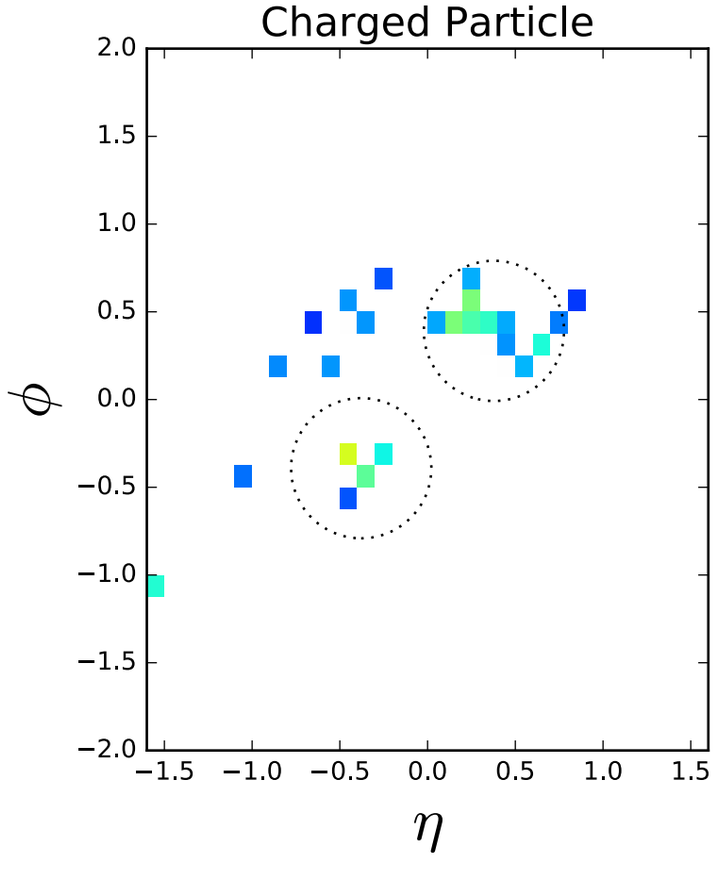


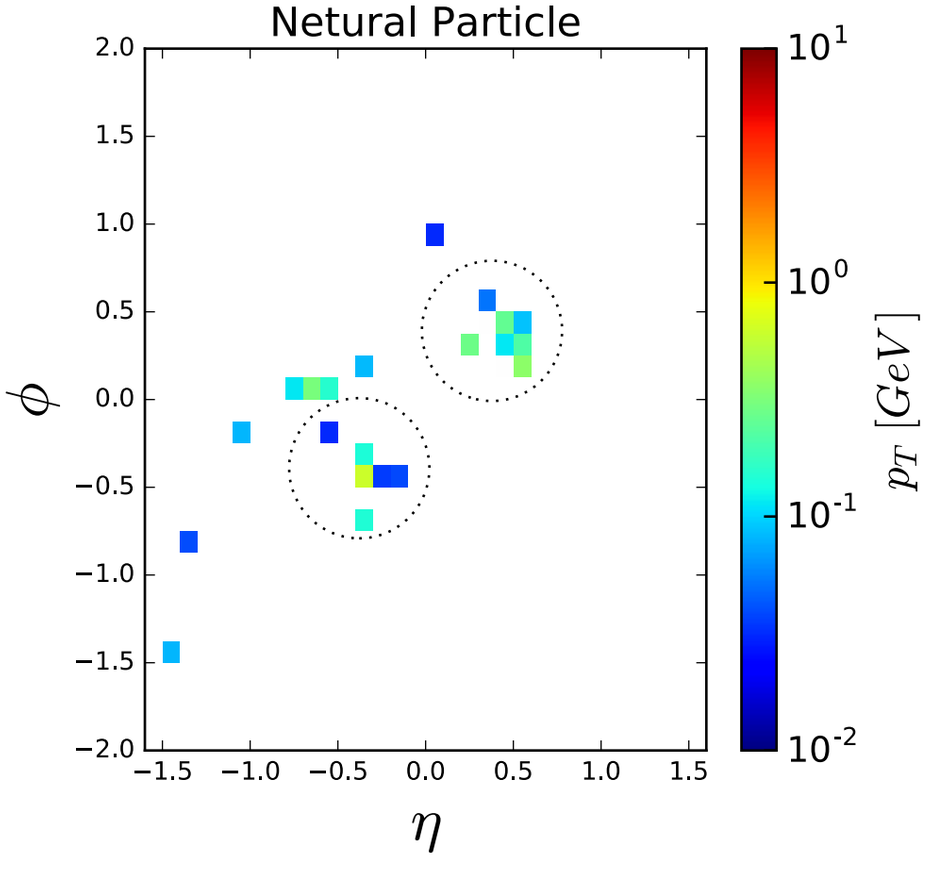




Not easy to determine event by event basis



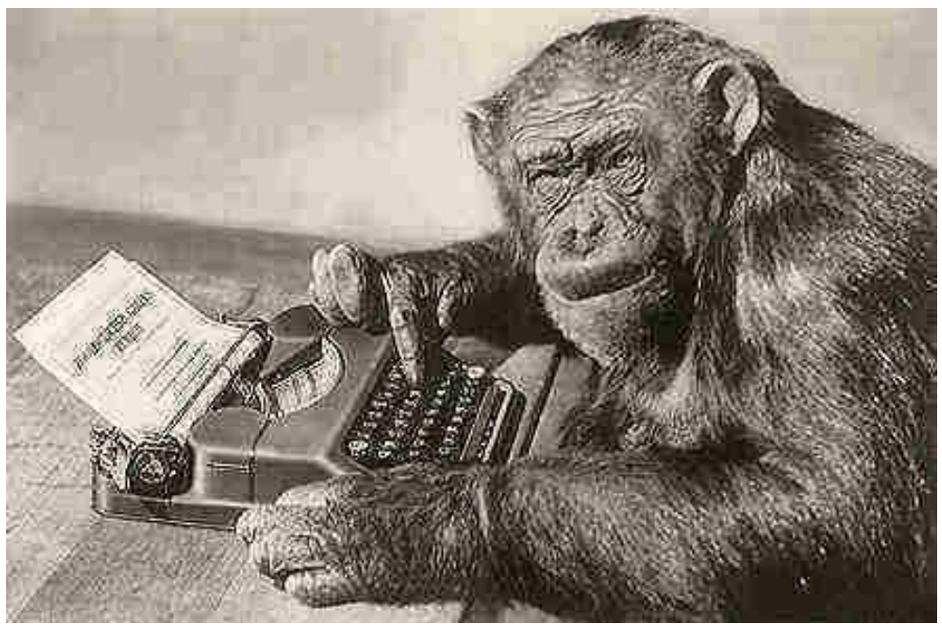




Let's pass the difficult and dirty jobs to a Machine

II

Typing Monkey with endless effort



Convolution network for image data

50×50 $|\eta| < 2.5,\ |\phi| < \pi$

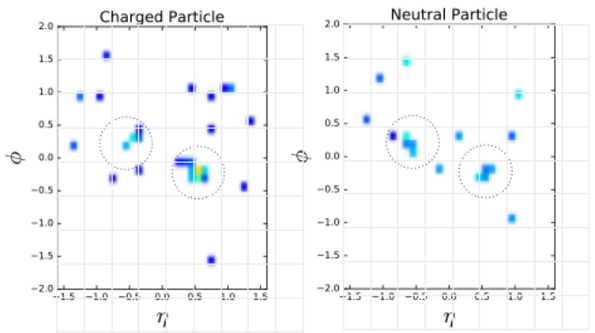
$I_{ij}/I_{\max}$

cropping with (for locality) $|\eta| < 1.6,\ \ |\phi| < 2.01$

Event 1

32×32 Charged

32×32 Neutral

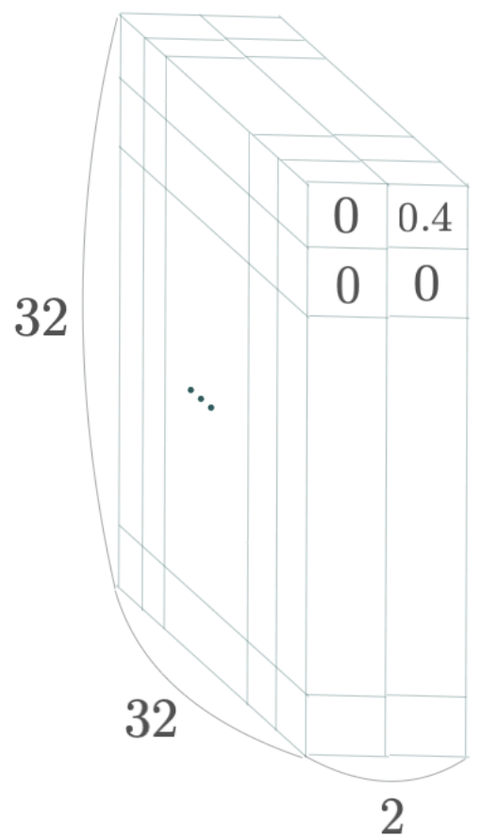


hh label

value

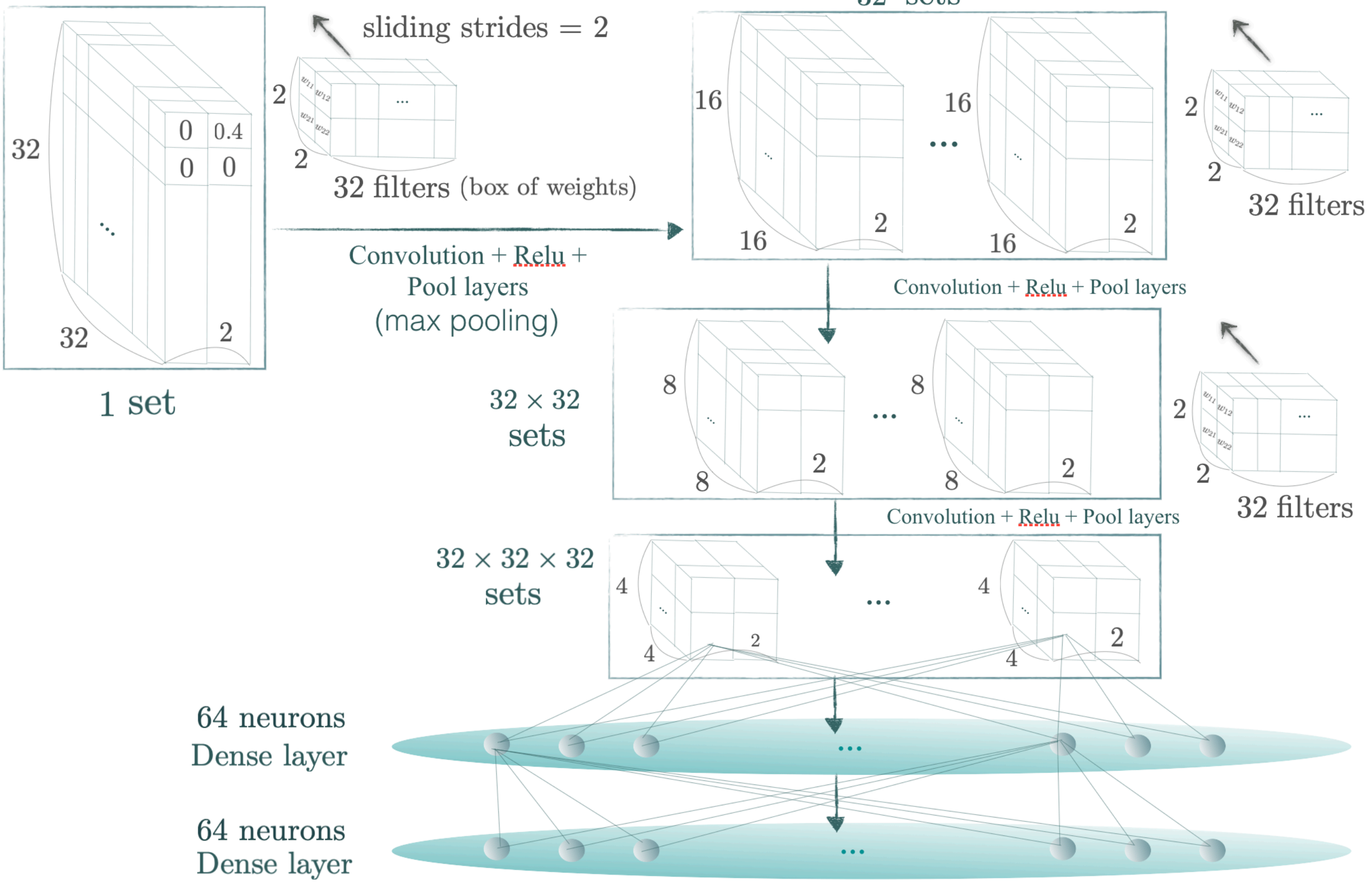
0 0 0 0.4 … 0 0 0.1

- The jet-image data is processed by a typical convolution network.
- The charged and neutral image data are shaped as 3D matrix.

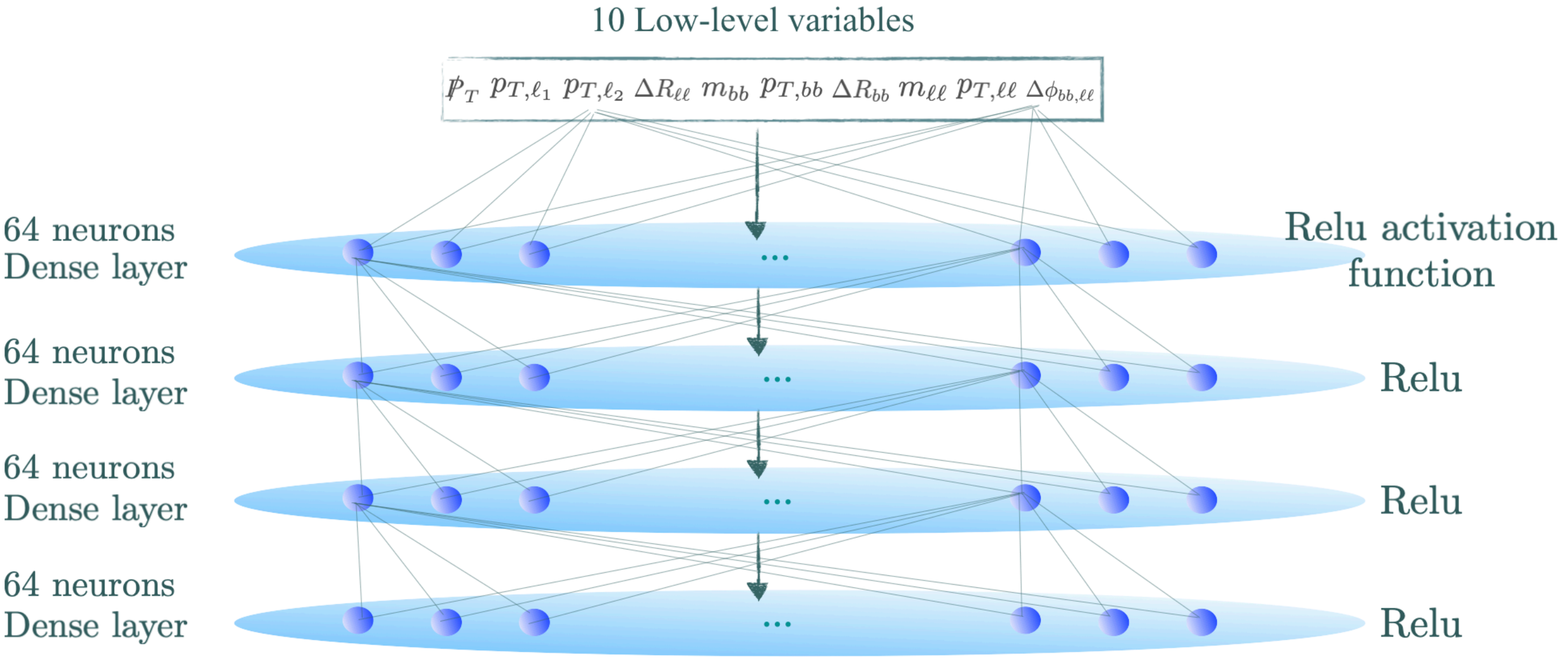


- This 3D matrix preserves the spatial relationship between pixels.

Convolution network for image data

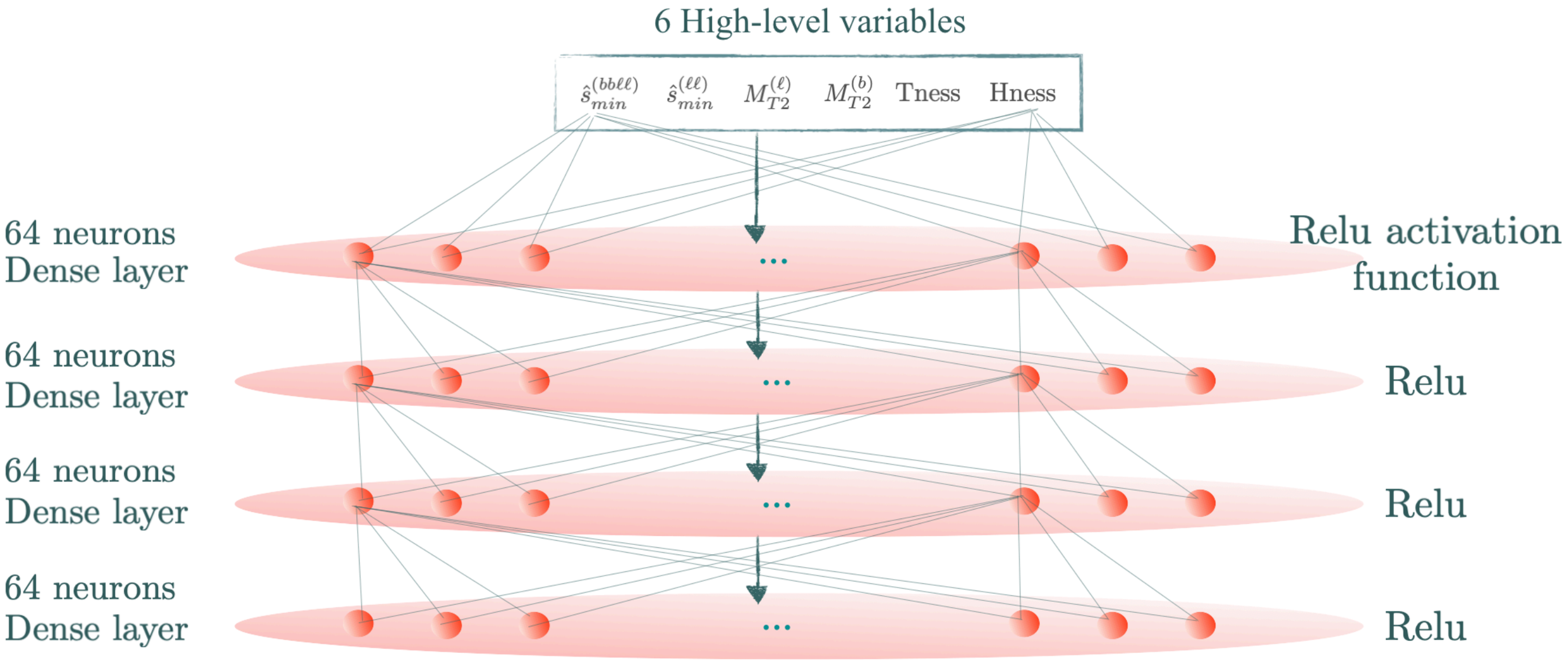


Dense neural network for low-level variables



- The 10 low-level variables are processed by four dense layers.

Dense neural network for high-level variables

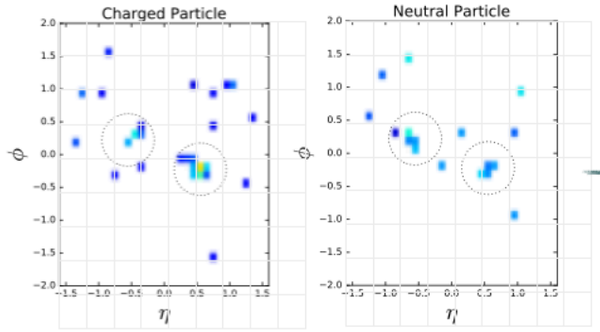


- The 6 high-level variables are processed by four dense layers.

Combining dense neural networks

32×32
Charged

32×32
Neutral



Dense layer (128)

Dense layer (1)

No Activation Function
Dense layer (3)

Output layer (2)

Signal

Background

10 Low-level variables

$\not{p}_T$ p_{T,ℓ_1} p_{T,ℓ_2} $\Delta R_{\ell\ell}$ m_{bb}

$p_{T,bb}$ ΔR_{bb} $m_{\ell\ell}$ $p_{T,\ell\ell}$ $\Delta\phi_{bb,\ell\ell}$

Dense layer (64)

Dense layer (1)

Combination

6 High-level variables

$\hat{s}^{(bb\ell\ell)}_{min}$ $\hat{s}^{(\ell\ell)}_{min}$ $M^{(\ell)}_{T2}$ $M^{(b)}_{T2}$ Tness Hness

Dense layer (64)

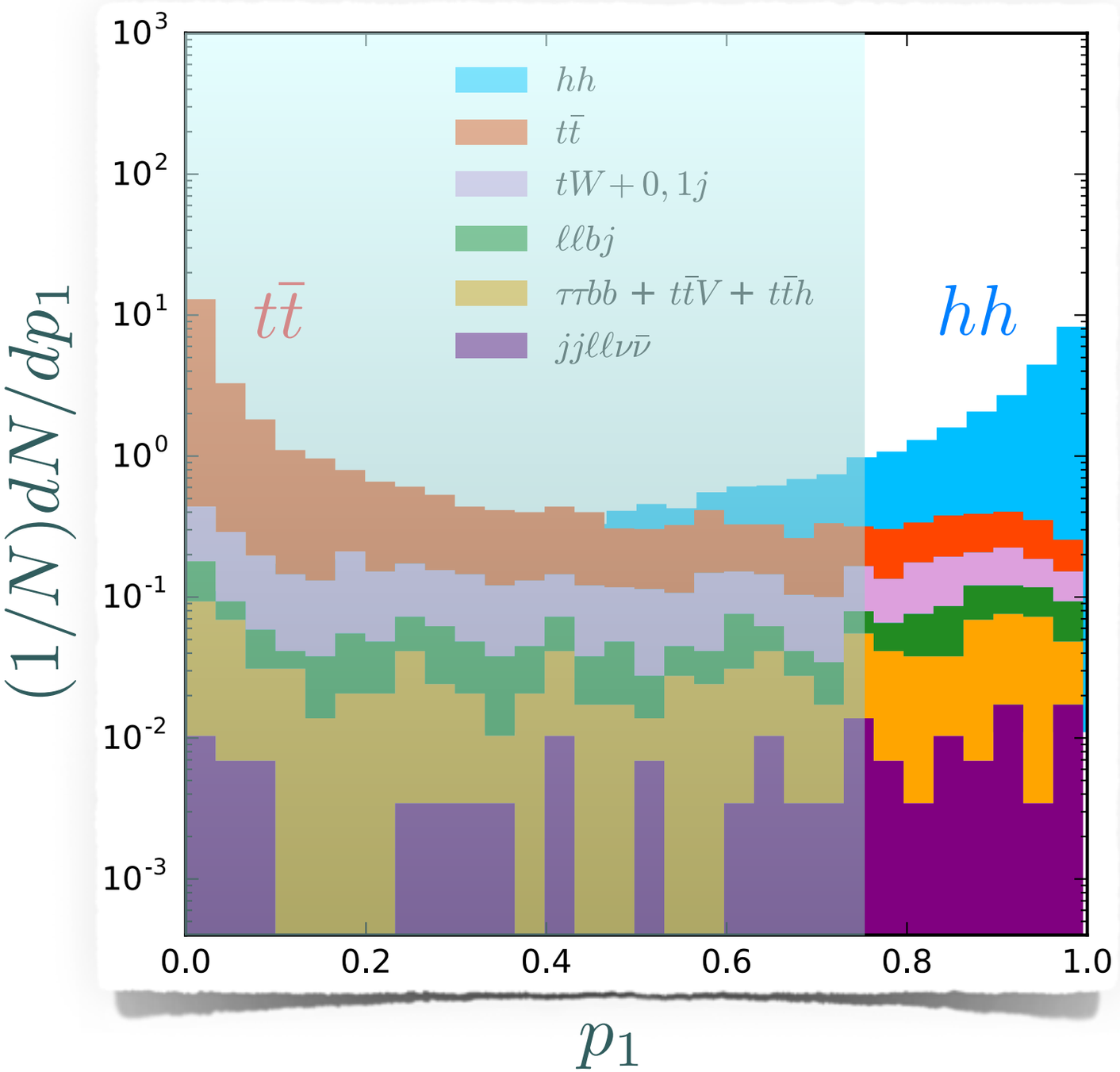
Dense layer (1)

$$H = -y_1 \log(p_1) - y_2 \log(p_2)$$

$$\text{Signal}: \ (y_1, y_2) = (1, 0)$$

$$\text{Backgrounds}: \ (y_1, y_2) = (0, 1)$$

DNN results



- Once the training is complete, we compute the probability (p_1) that a given event is classified as *hh*.
- Most of backgrounds are unlikely to be classified as *hh*.
- We place a cut on p_1 to disentangle the backgrounds.

$hh \to bbWW^*$ discovery significance

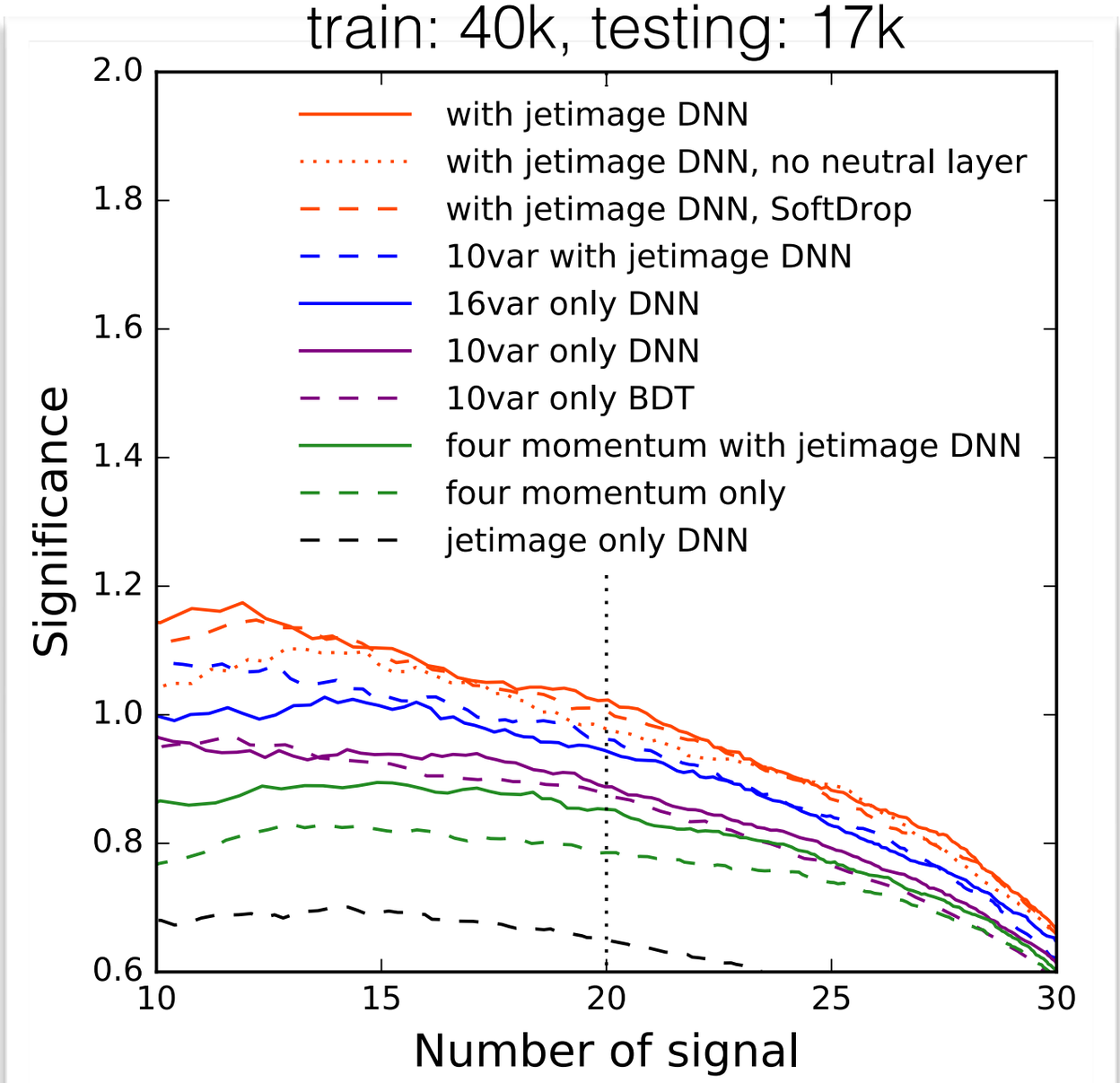
Using Delphes

$3\ \text{ab}^{-1}(14\ \text{TeV})$

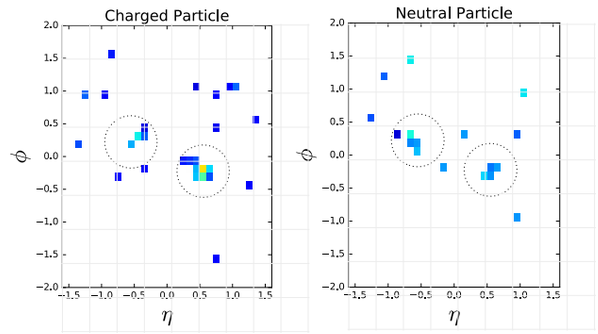
$\kappa_3 = 1$

$\text{significance} = 0.59\sigma$

CMS-FTR-18-019-PAS



- We can see a relative improvement in each layer of information.
- The DNN with jet images and high-level variables improves the final significance.



+

6 High-level variables

$\hat{s}_{min}^{(bb\ell\ell)}$ $\hat{s}_{min}^{(\ell\ell)}$ $M_{T2}^{(\ell)}$ $M_{T2}^{(b)}$ Tness Hness

+

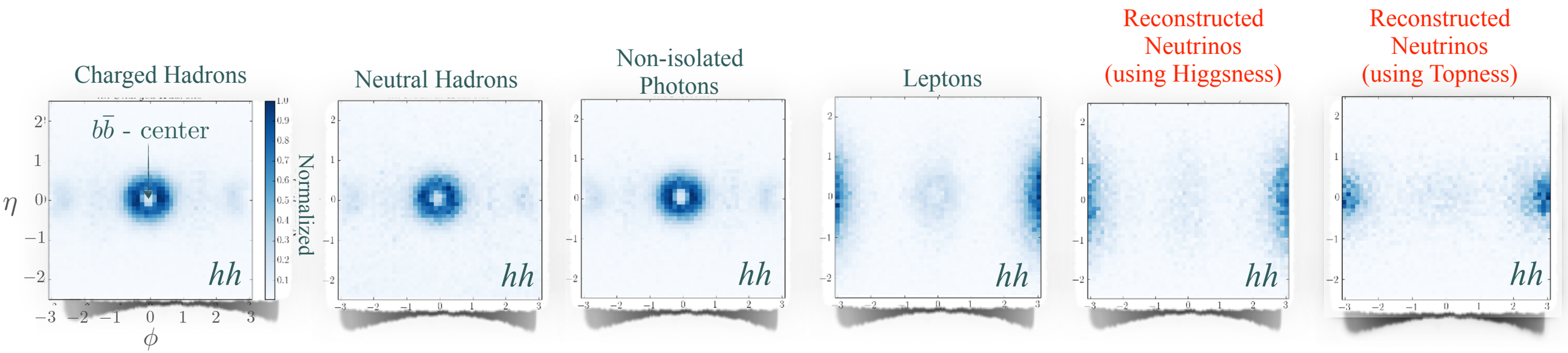
10 Low-level variables

$\not{p}_T$ p_{T,ℓ_1} p_{T,ℓ_2} $\Delta R_{\ell\ell}$ m_{bb}

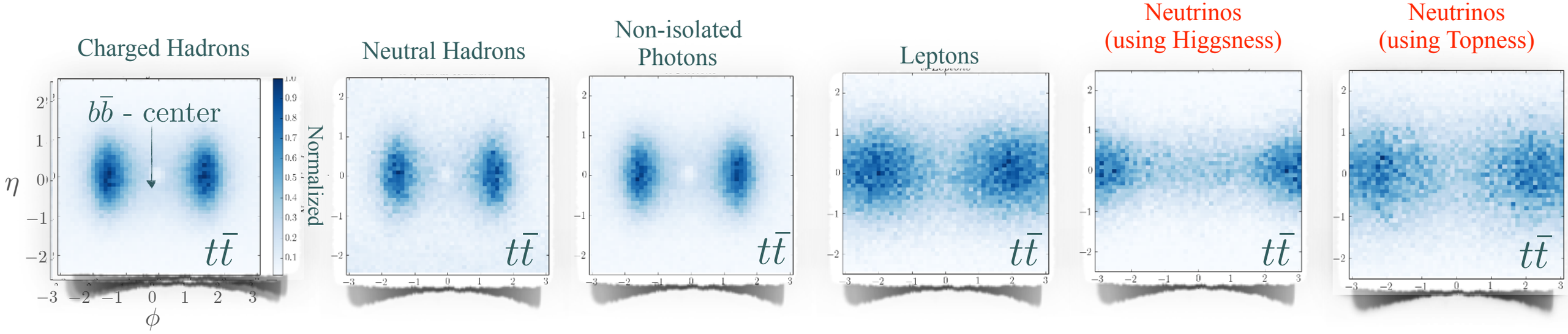
$p_{T,bb}$ ΔR_{bb} $m_{\ell\ell}$ $p_{T,\ell\ell}$ $\Delta\phi_{bb,\ell\ell}$

The Di-Higgs Photography

- Totally hadrons, lepton, photon, and neutrino images are shown for *hh*.



- Analogous images are shown for $t\bar{t}$ background.



- A sharp difference between *hh* and $t\bar{t}$.
- Construct 5 images (L+T, L+H) with 6 images data set.

Kim, Kong, Matchev, Park JHEP 2019

Kim, Kong, Matchev, Park, preliminary

Hypothesis test

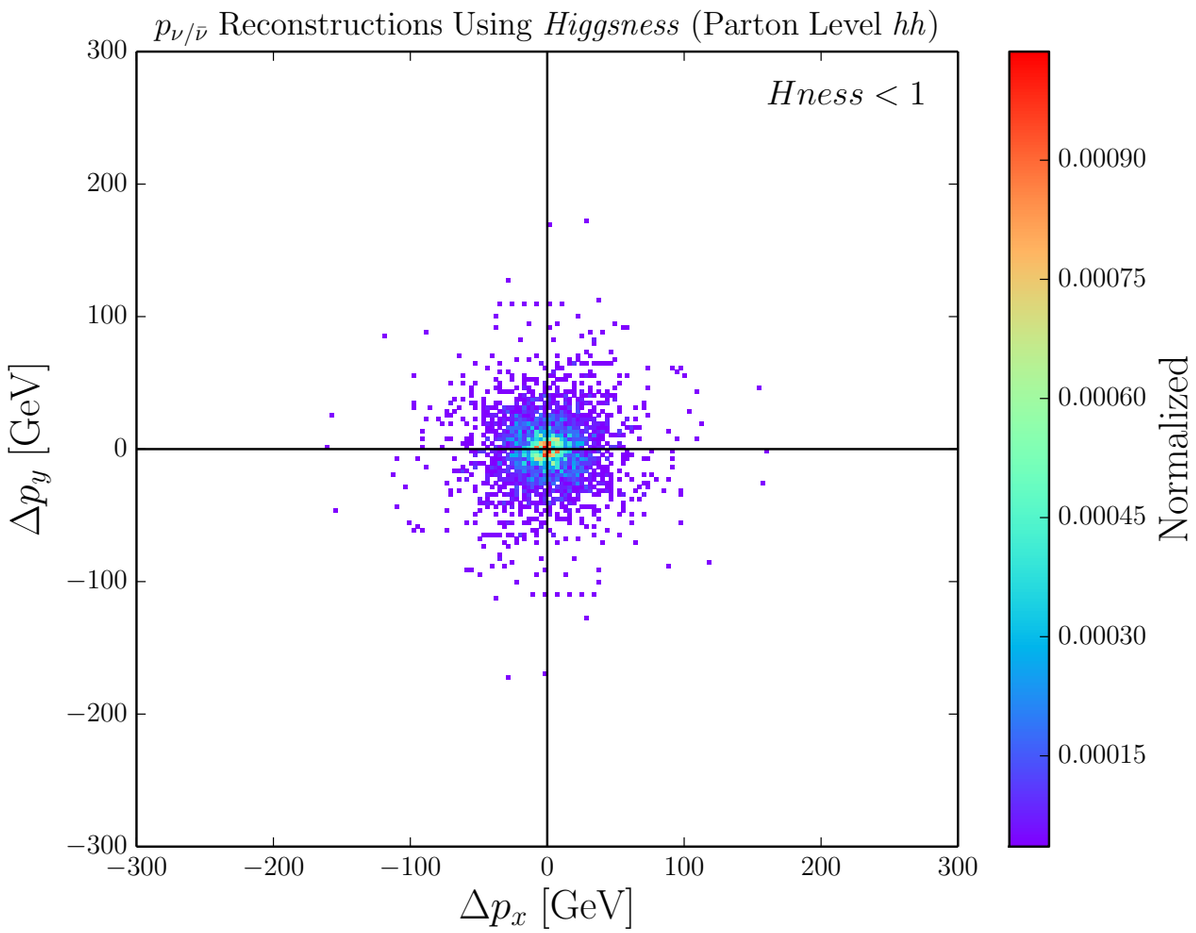
- With trying to "**reconstruct**" 4-momentum of **neutrino** under
  - signal hypothesis

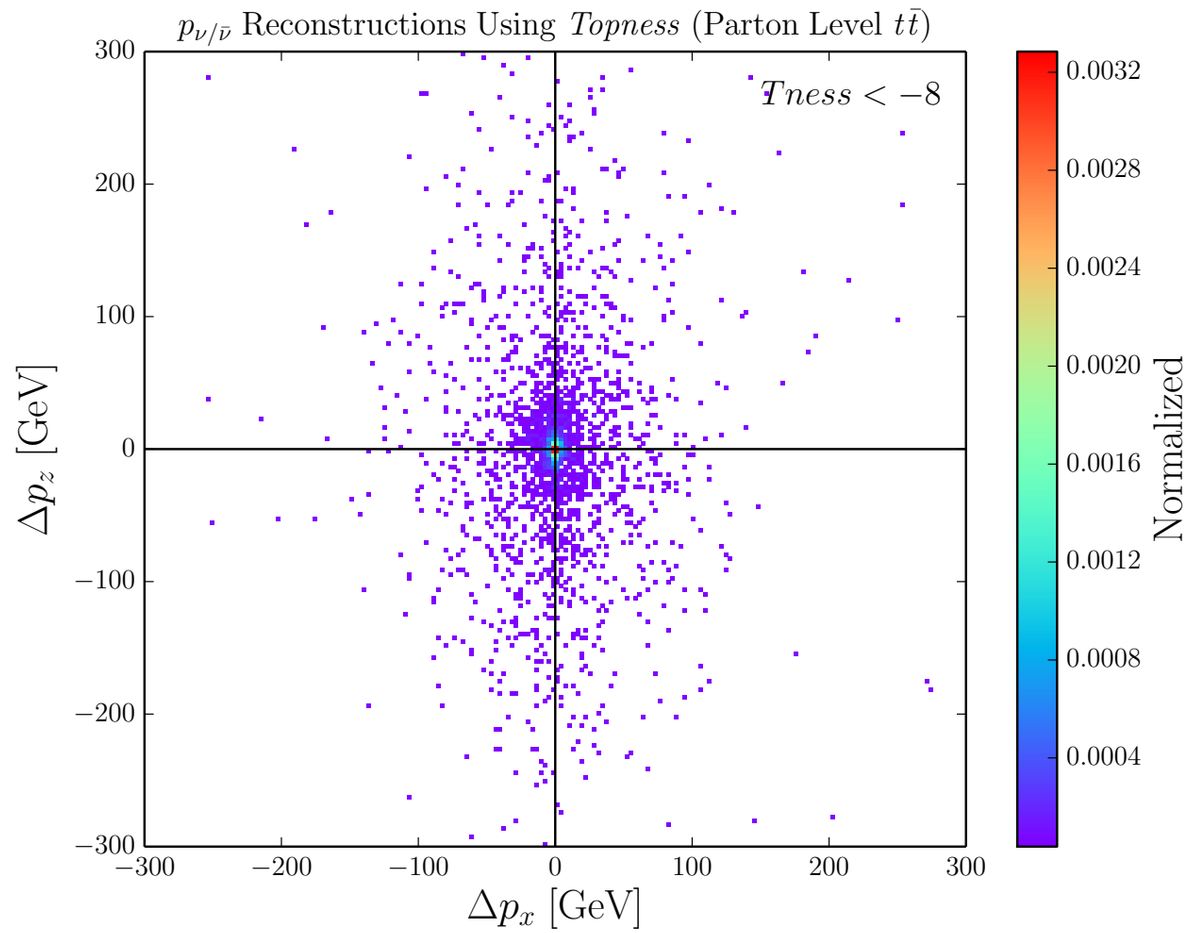
$$H \equiv \min\left[\frac{\left(m^2_{\ell^+\ell^-\nu\bar{\nu}} - m^2_h\right)^2}{\sigma^4_{h_\ell}} + \frac{\left(m^2_{\nu\bar{\nu}} - m^2_{\nu\bar{\nu},peak}\right)^2}{\sigma^4_\nu}\right.$$
$$+min\left(\frac{\left(m^2_{\ell^+\nu} - m^2_W\right)^2}{\sigma^4_W} + \frac{\left(m^2_{\ell^-\bar{\nu}} -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

  - background hypothesis

$$\chi^2_{ij} \equiv \min_{\vec{\not{p}}_T = \vec{p}_{\nu T} + \vec{p}_{\bar{\nu} T}}\left[\frac{\left(m^2_{b_i\ell^+\nu} - m^2_t\right)^2}{\sigma^4_t} + \frac{\left(m^2_{\ell^+\nu} - m^2_W\right)^2}{\sigma^4_W}\right.$$
$$\left.+\frac{\left(m^2_{b_j\ell^-\bar{\nu}} - m^2_t\right)^2}{\sigma^4_t} + \frac{\left(m^2_{\ell^-\bar{\nu}} - m^2_W\right)^2}{\sigma^4_W}\right]$$

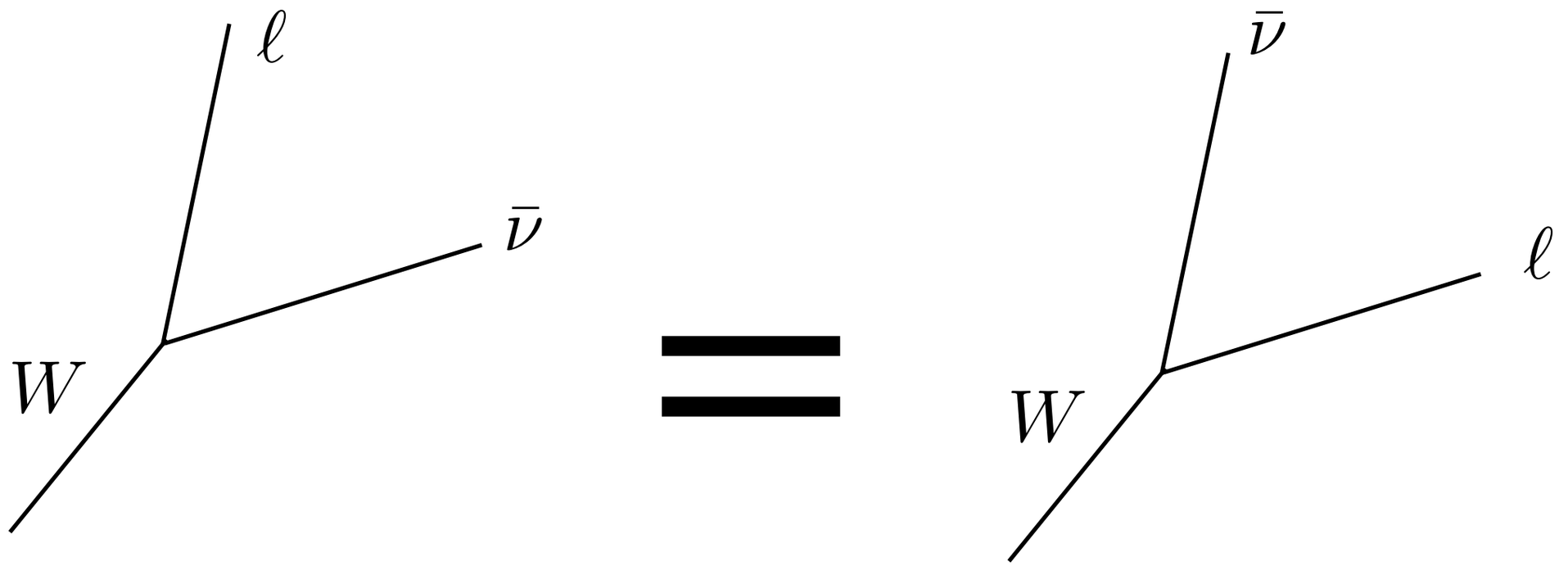
Neutrino momenta (parton-level)





- Just like MT2, Higgsness and Topness provide momentum of neutrinos.
- They can be used to study other quantities.

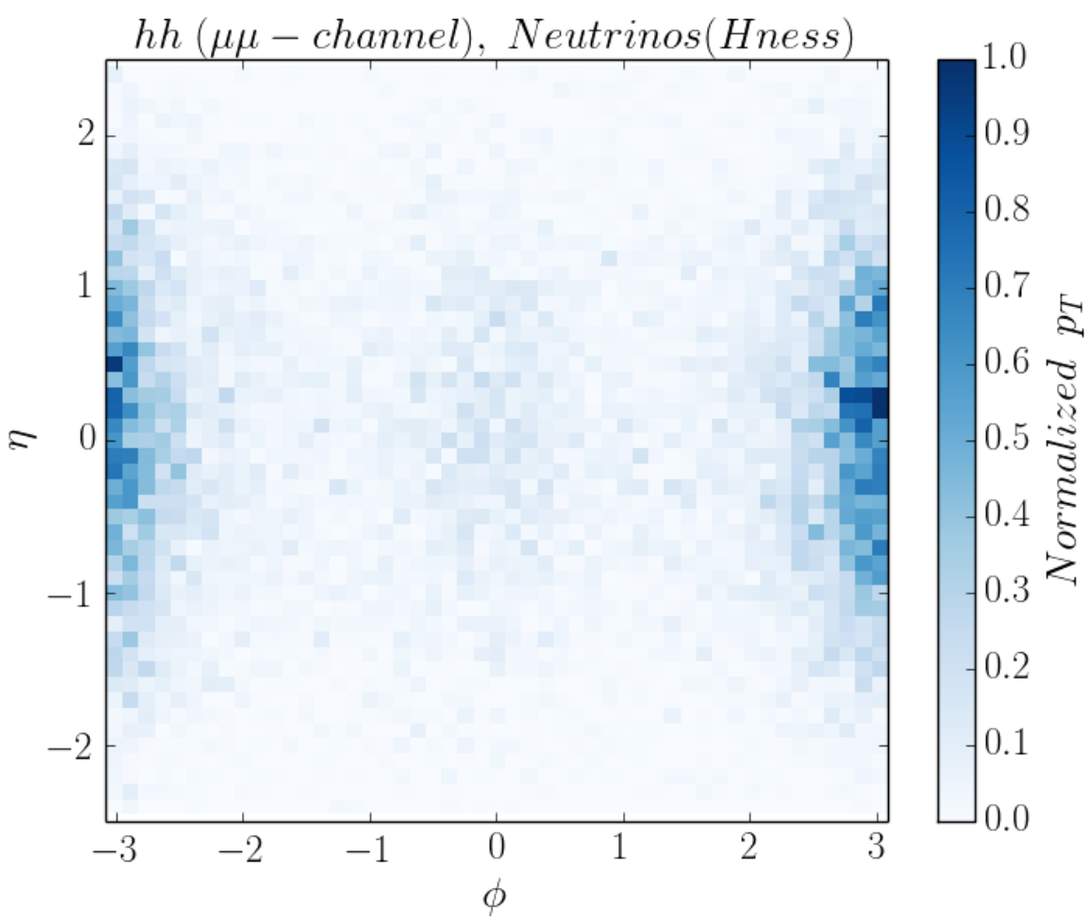
Symmetry of $\ell \leftrightarrow \nu$ momentum



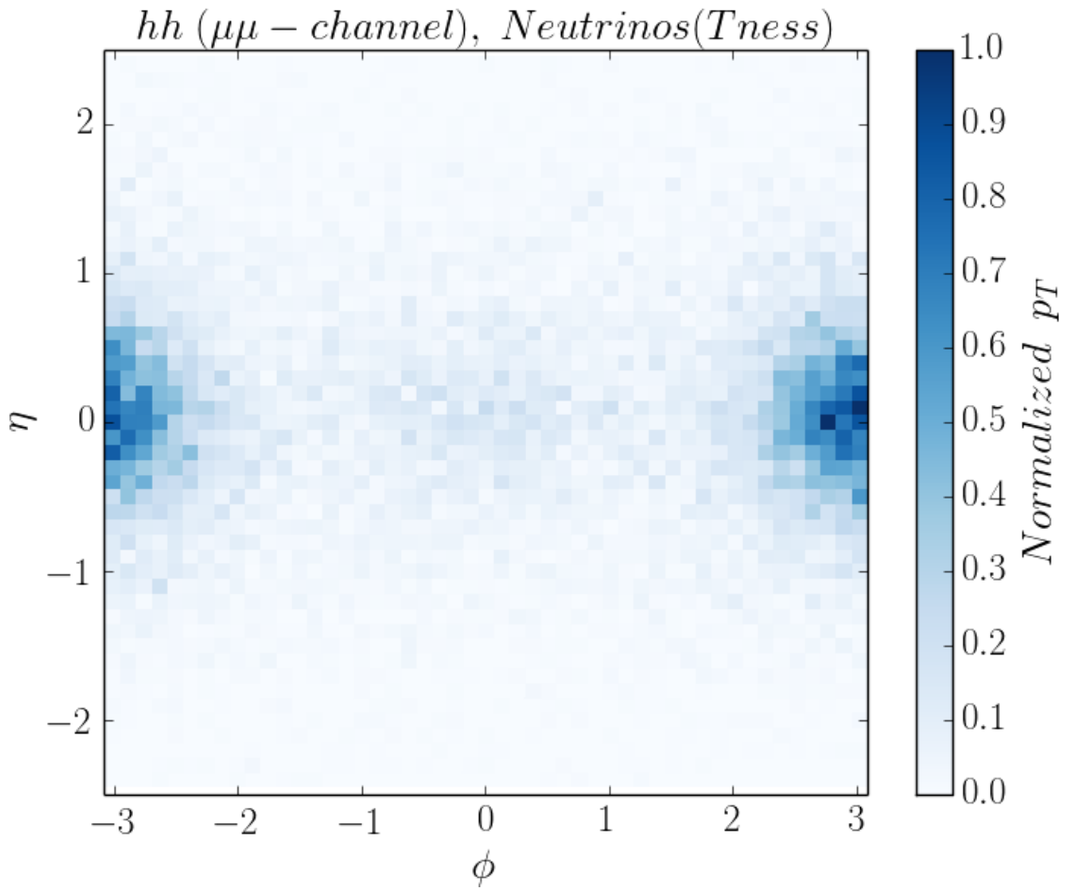
If we have this event, this event is also possible!

Hypothesis test: Signal

- With trying to "reconstruct" 4-momentum of neutrino under
  - signal hypothesis



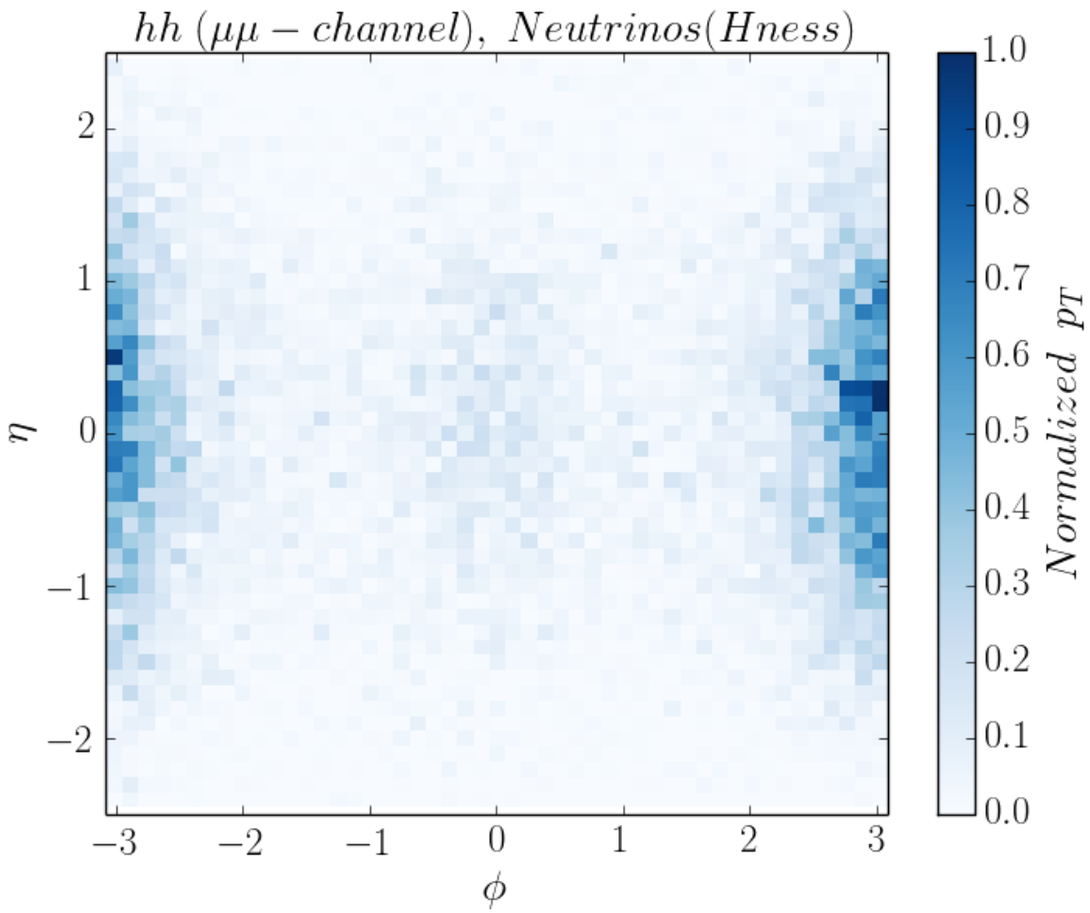
  - background hypothesis



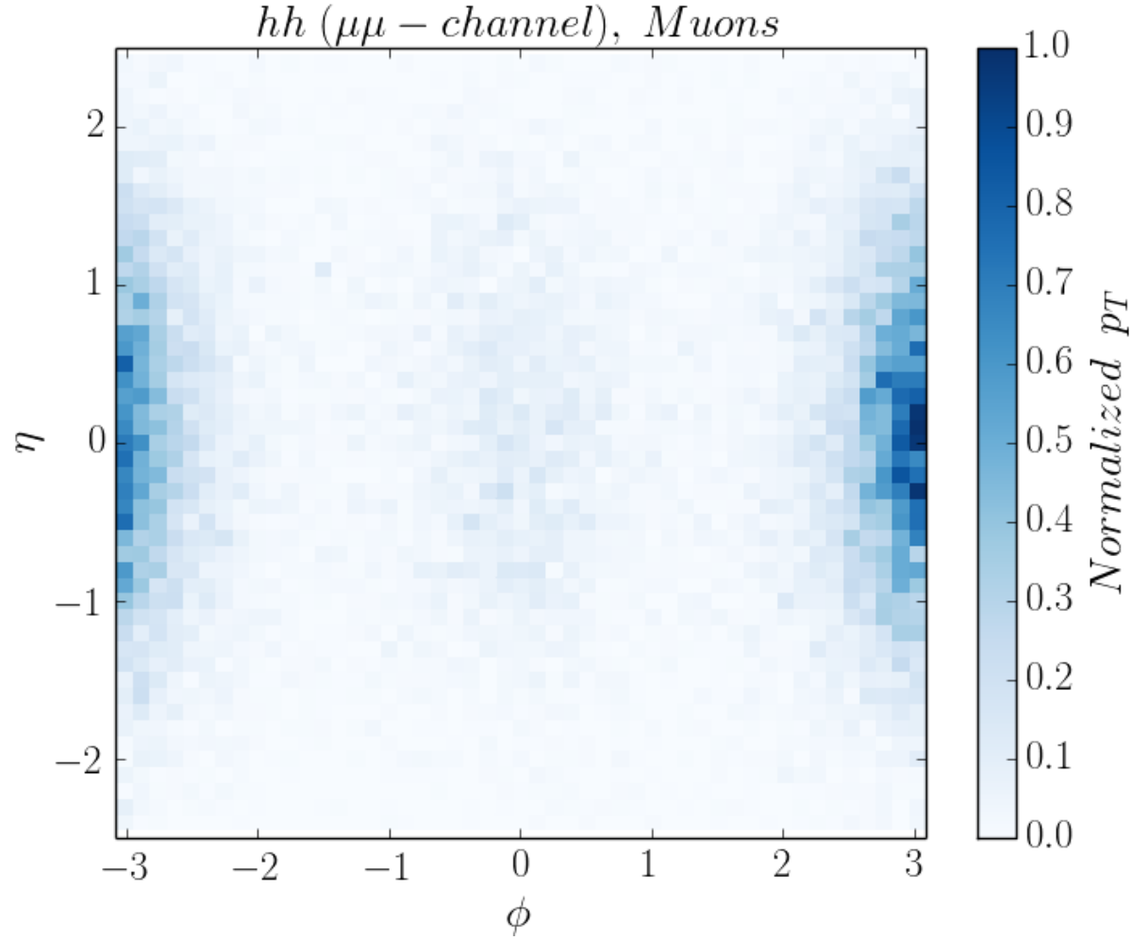
Hypothesis test: Signal

- With trying to "reconstruct" 4-momentum of neutrino under
  - correct hypothesis

(reconstructed neutrinos)

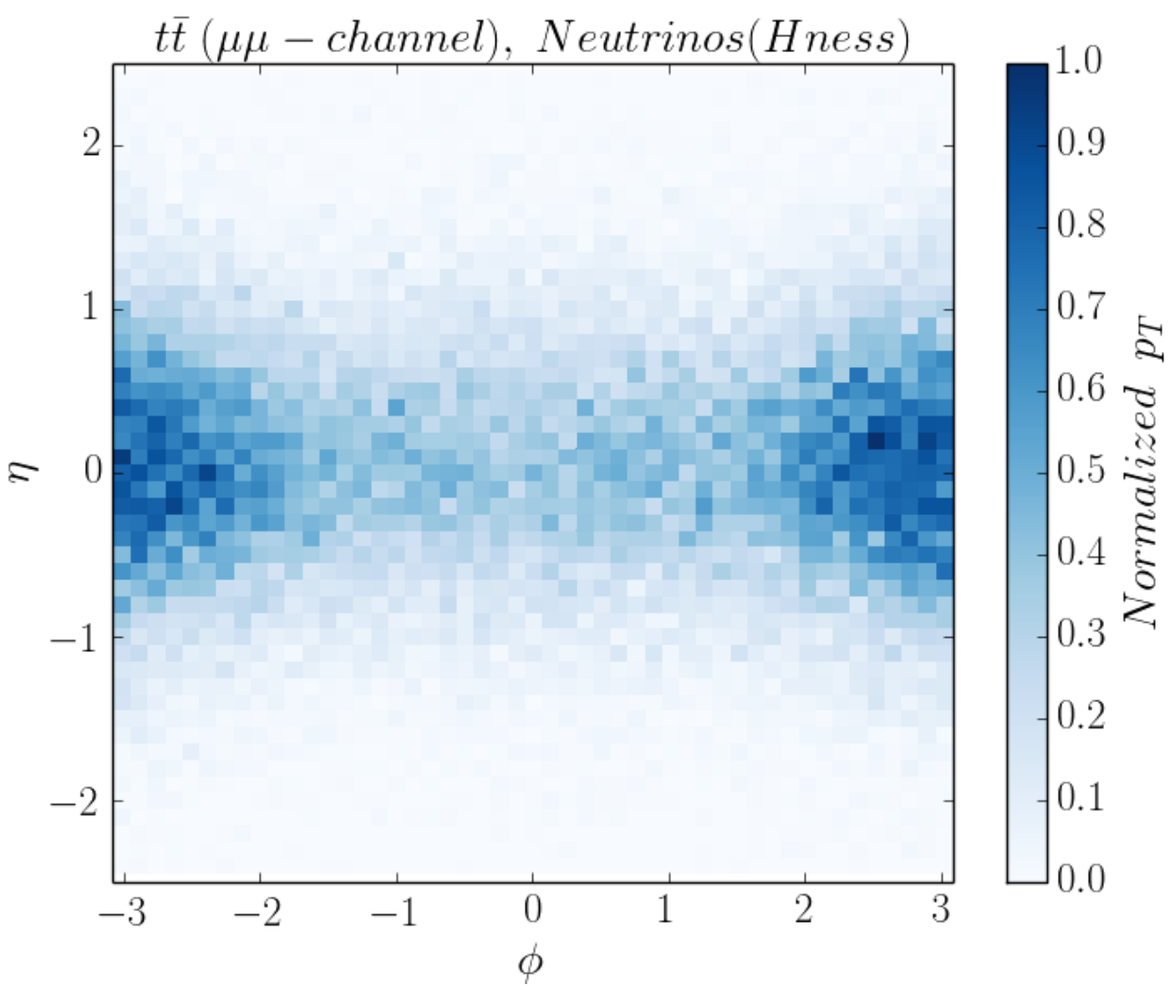


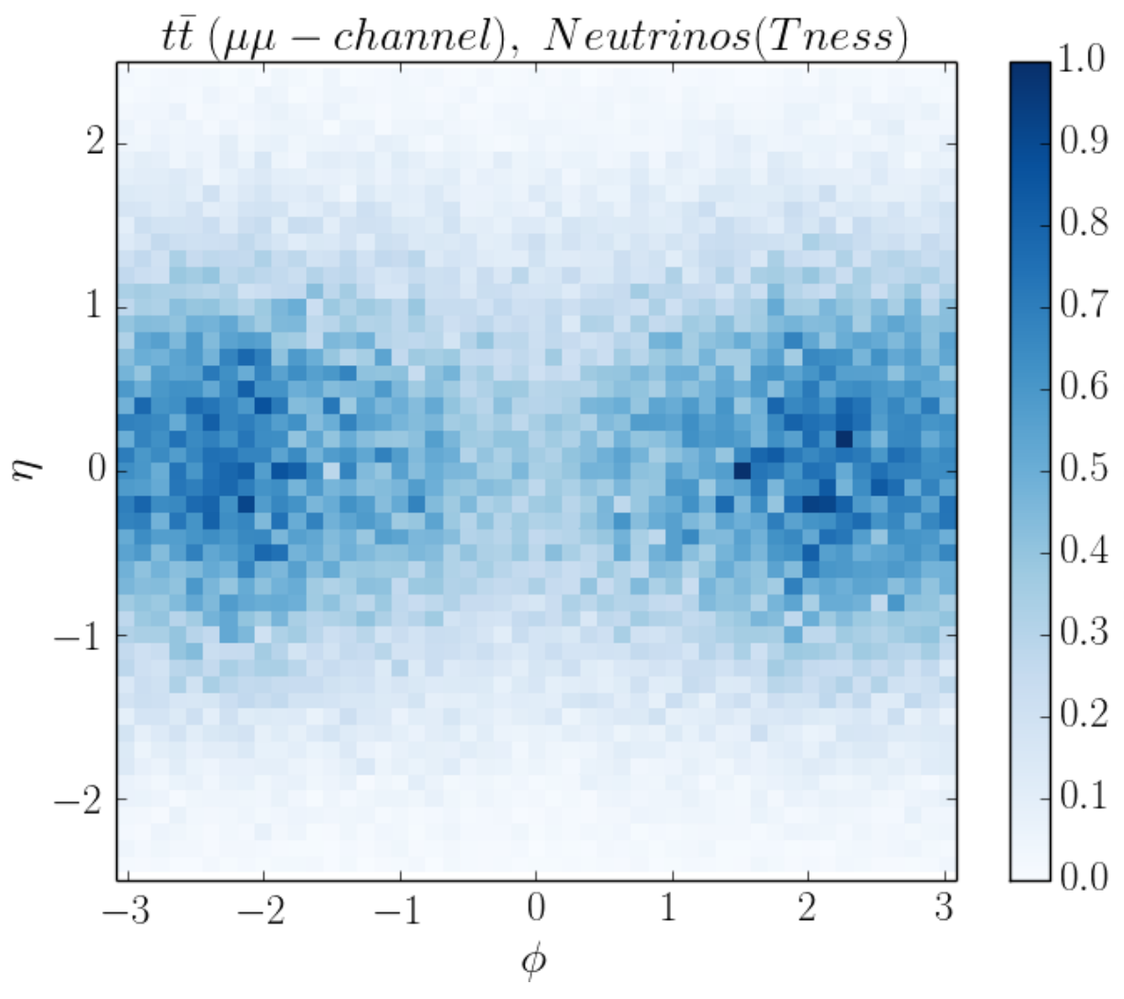
- **information from leptons**
  (neutrino and leptons from same W)



Hypothesis test: Background

- With trying to "reconstruct" 4-momentum of neutrino under
  - signal hypothesis
  - background hypothesis

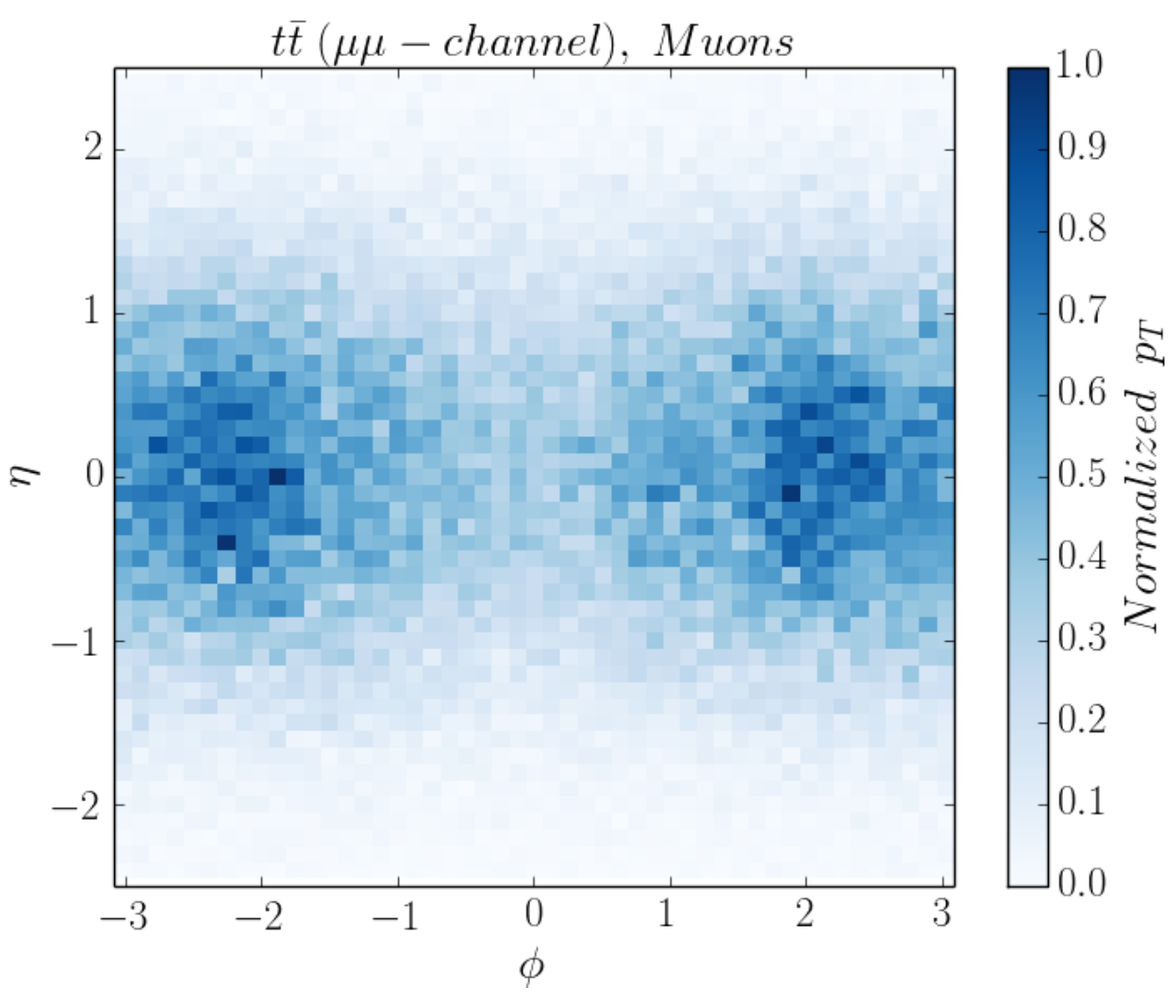




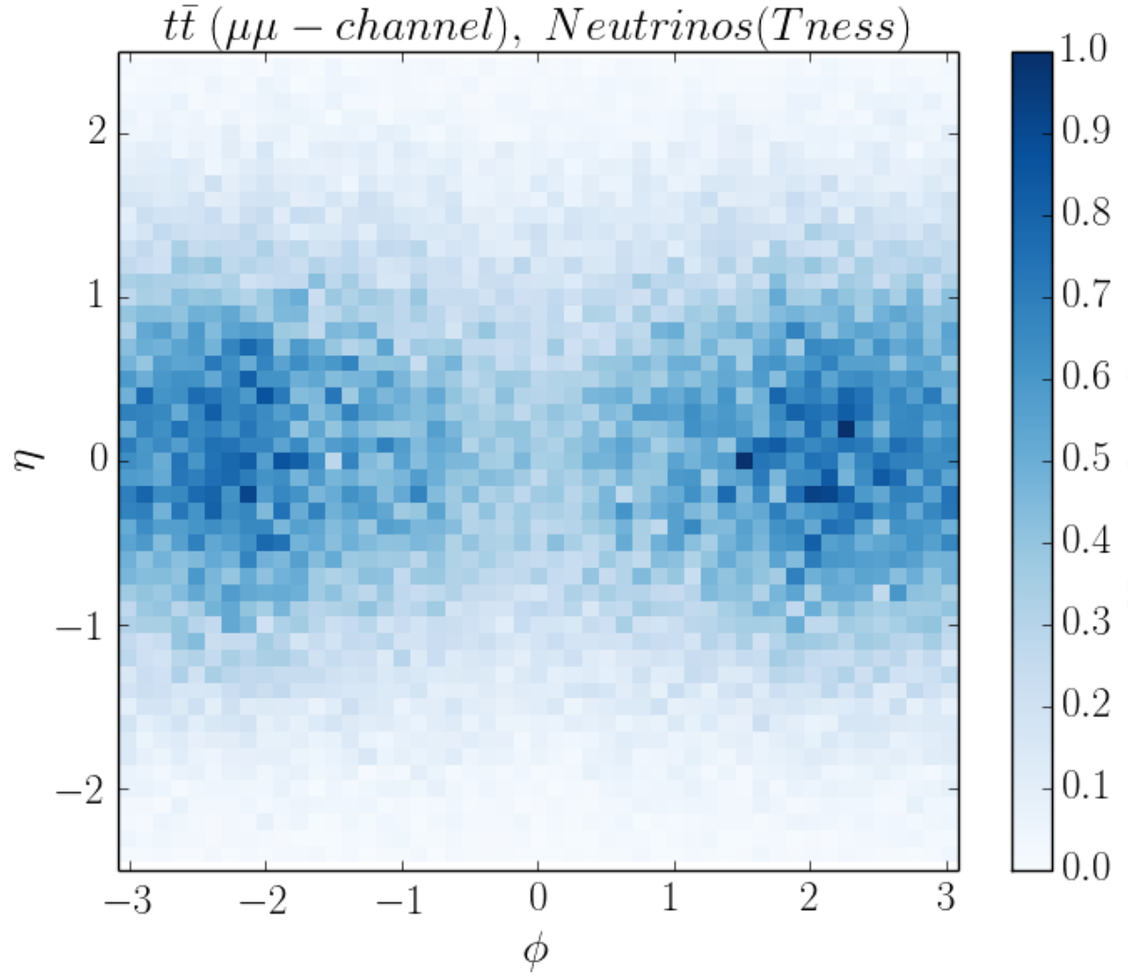
Hypothesis test: Background

- With trying to "reconstruct" 4-momentum of neutrino under

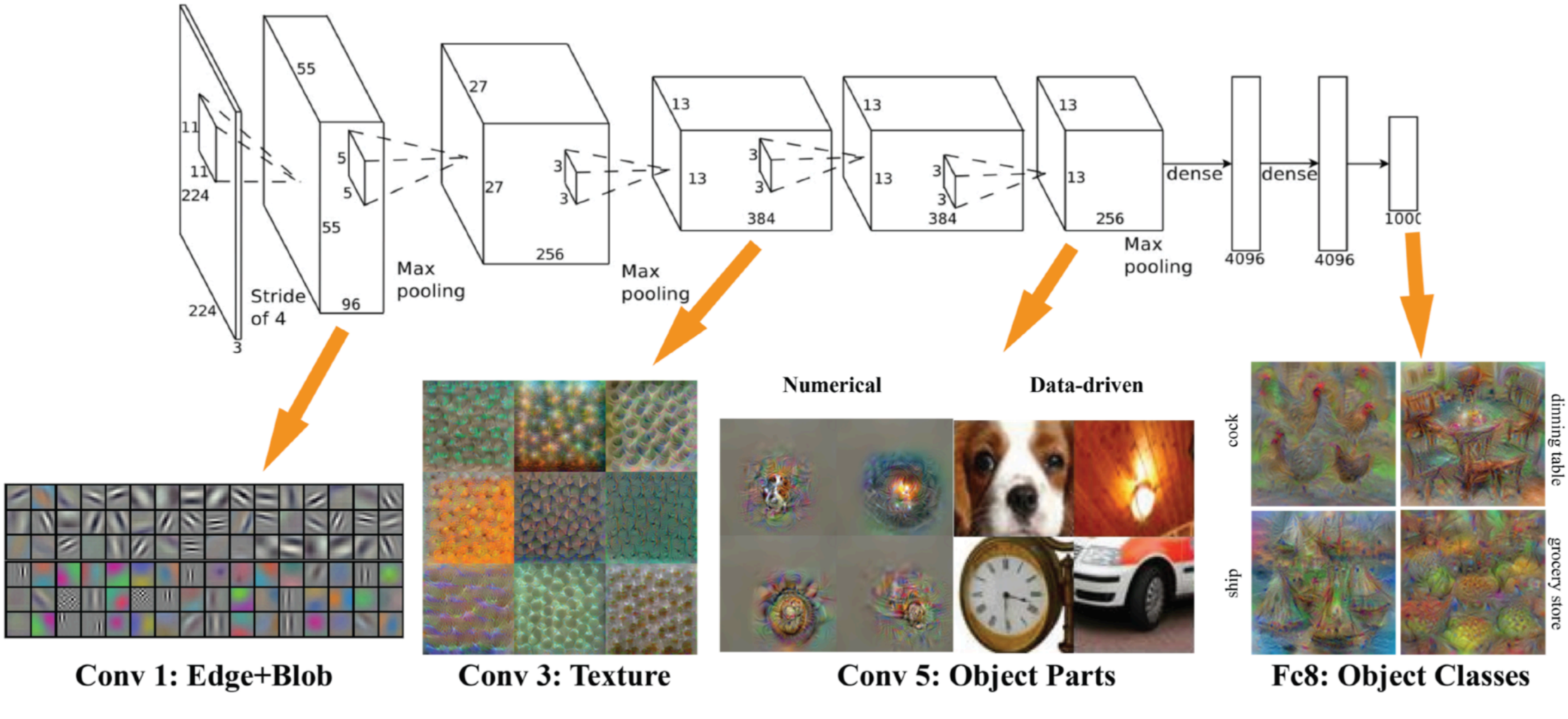
- **information from leptons**
  (neutrino and leptons from same W)



- correct hypothesis
  (reconstructed neutrinos)

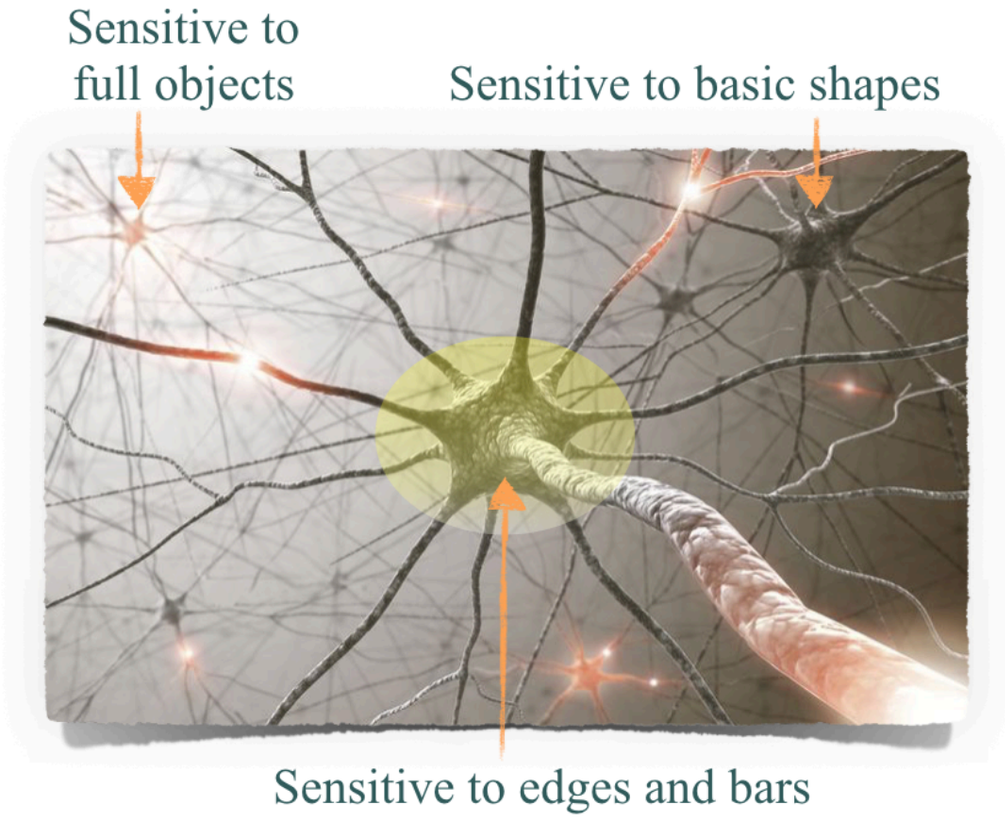


3. Deeper and Deeper

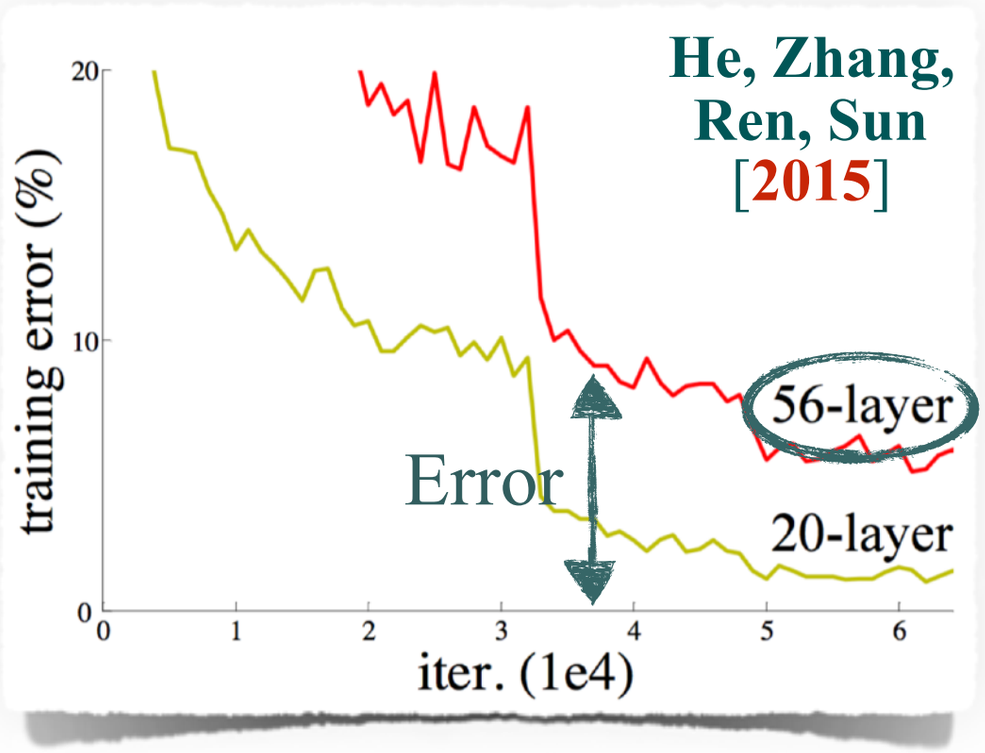


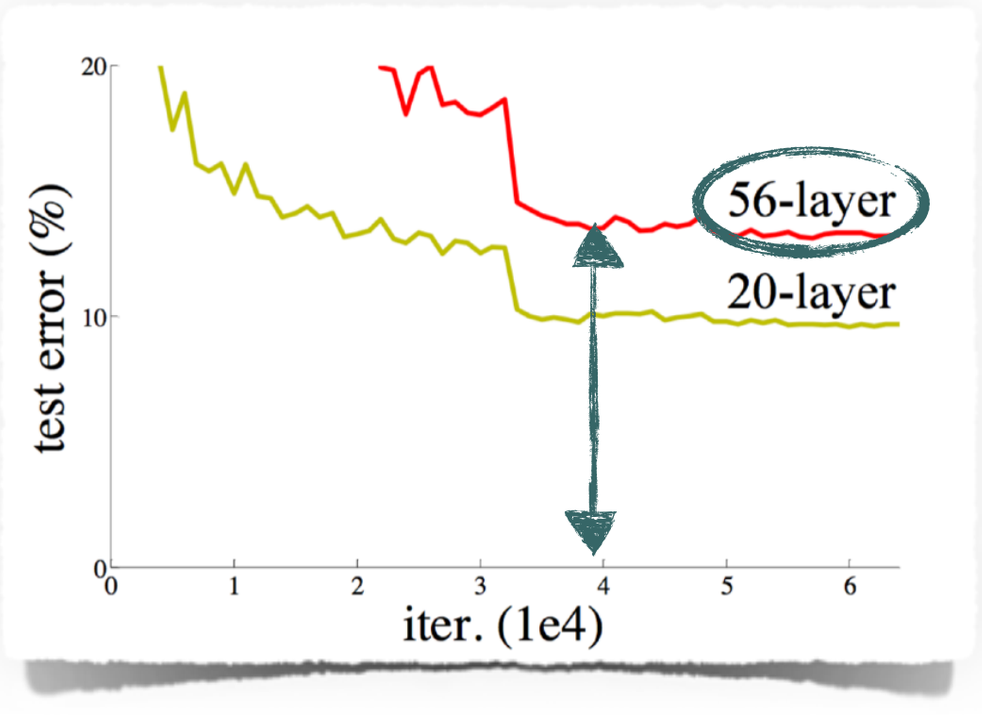
Good for shallow NN for simple correlations

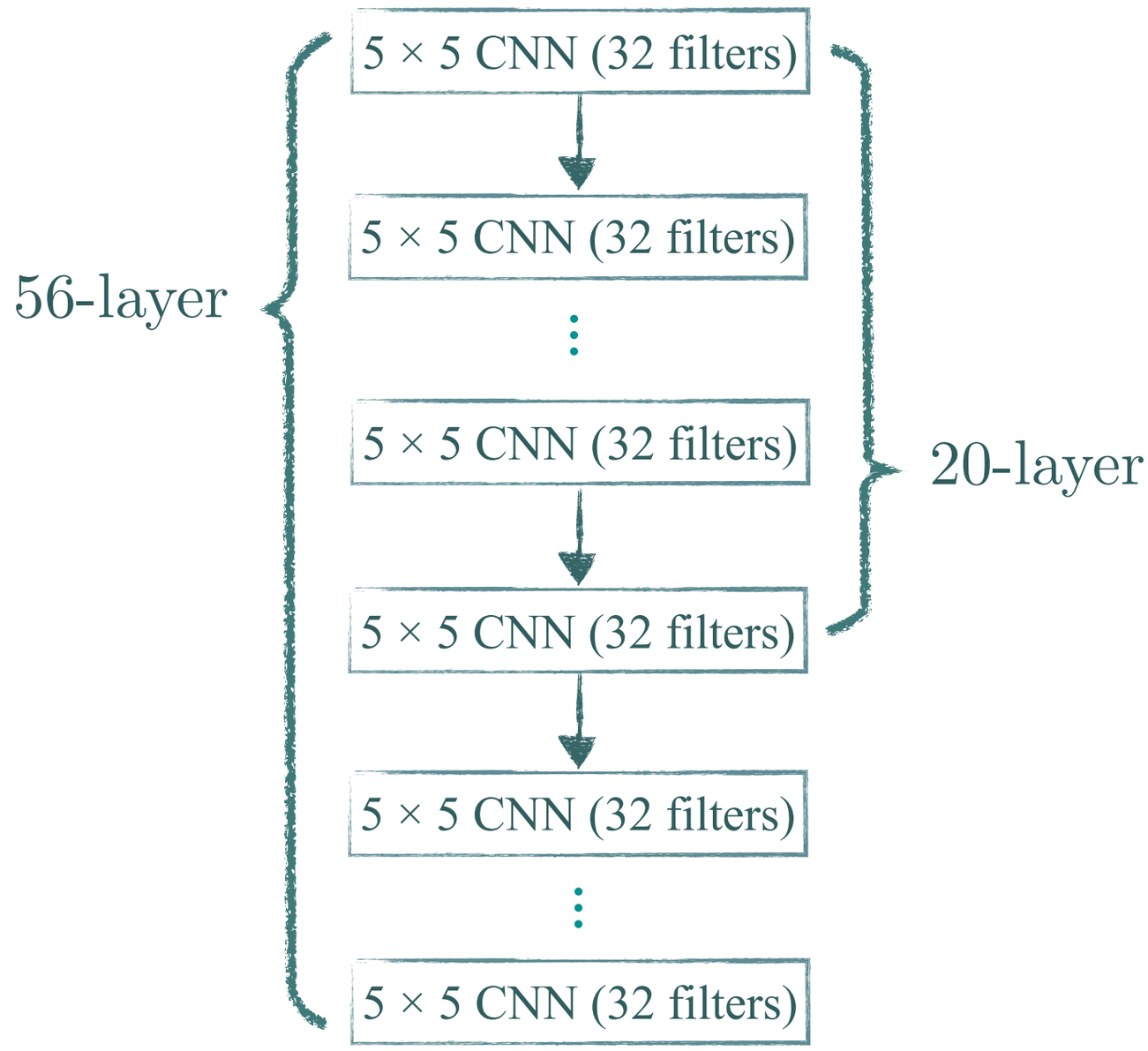
- To make DL understand the object properly, we need to increase layers...



Vanishing gradient



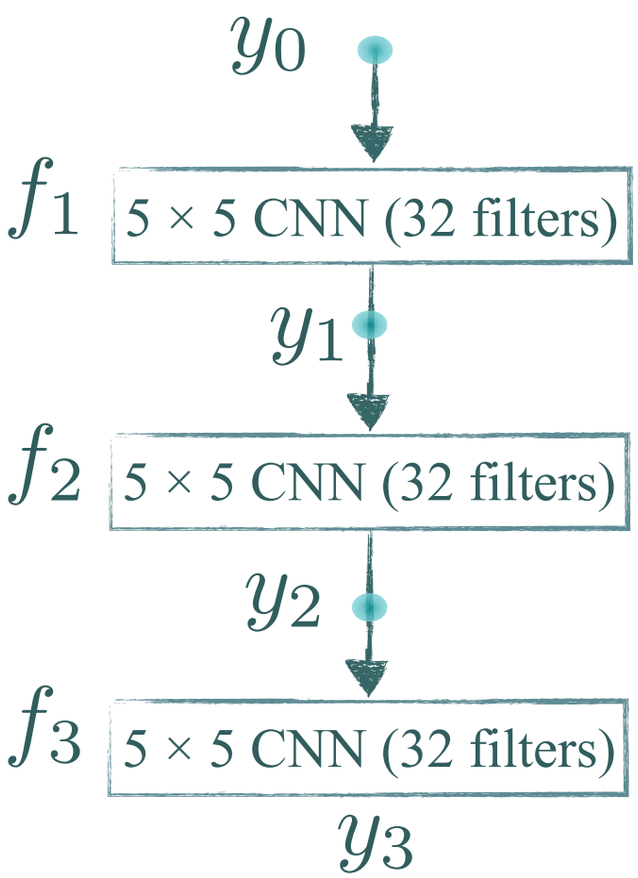




- However, as the network goes deeper, its performance gets saturated or even starts degrading rapidly.

A Residual Neural Network (ResNet)

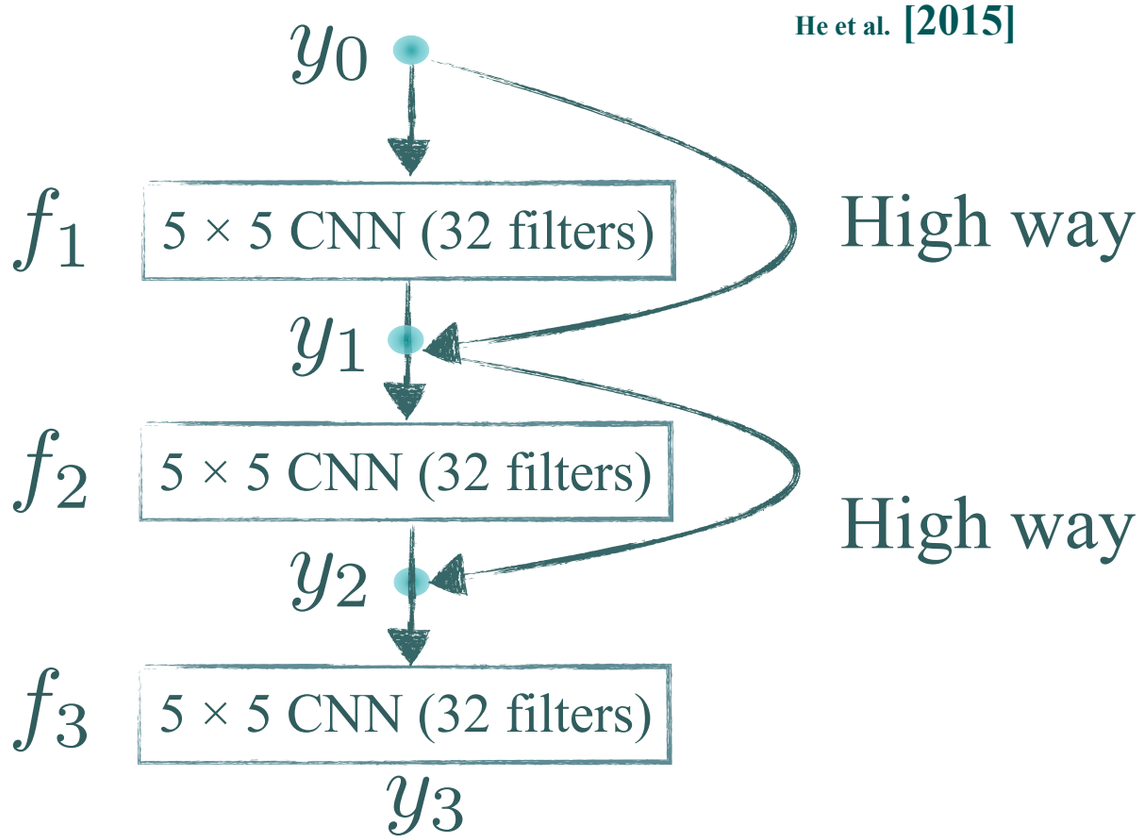
Classic CNN



$$y_3 = f_3(f_2(f_1(y_0)))$$

Residual Neural Network (ResNet)

He et al. [2015]

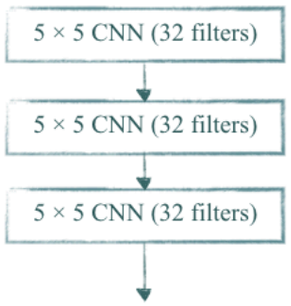


$$y_3 = \underbrace{y_0}_{\text{original}} + \underbrace{f_1(y_0)}_{\text{angle1}} + \underbrace{f_2(y_0 + f_1(y_0))}_{\text{angle2}} + \underbrace{f_3(y_0 + f_1(y_0) + f_2(y_0 + f_1(y_0)))}_{\text{angle3}}$$

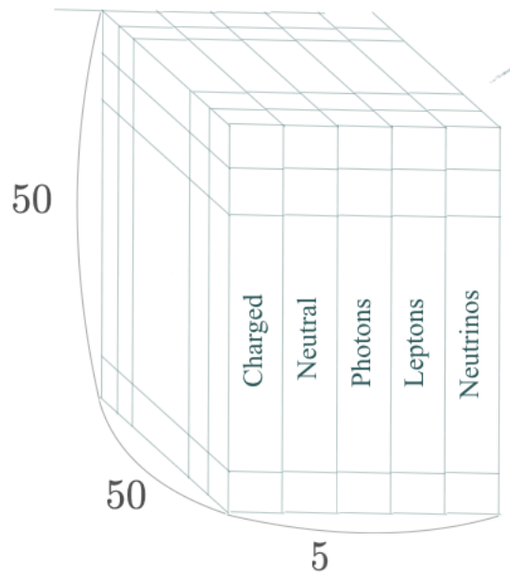
- The ResNet was proposed to go much deeper.
- The core idea of ResNet is introducing high ways skipping the nodes.
- ResNet has a variety of angles to detect full images, as compared to the classic CNN.

Our ResNet

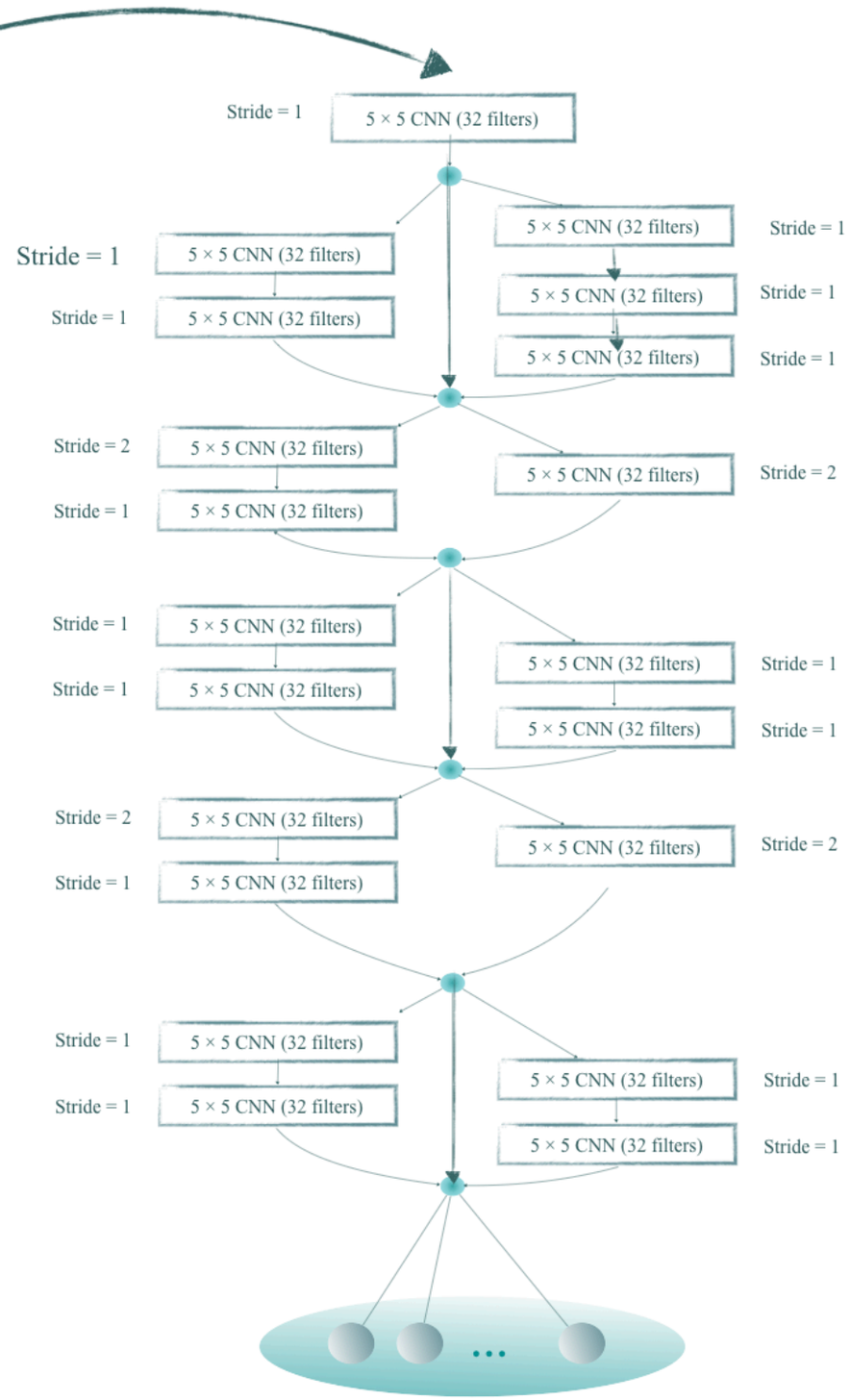
Classic CNN



- As the network becomes deeper, the performance of CNN gets saturated or even starts degrading rapidly.
- The classic CNN starts to degrade after 3 layers, with our 5-image data.
- Absolutely not suitable.



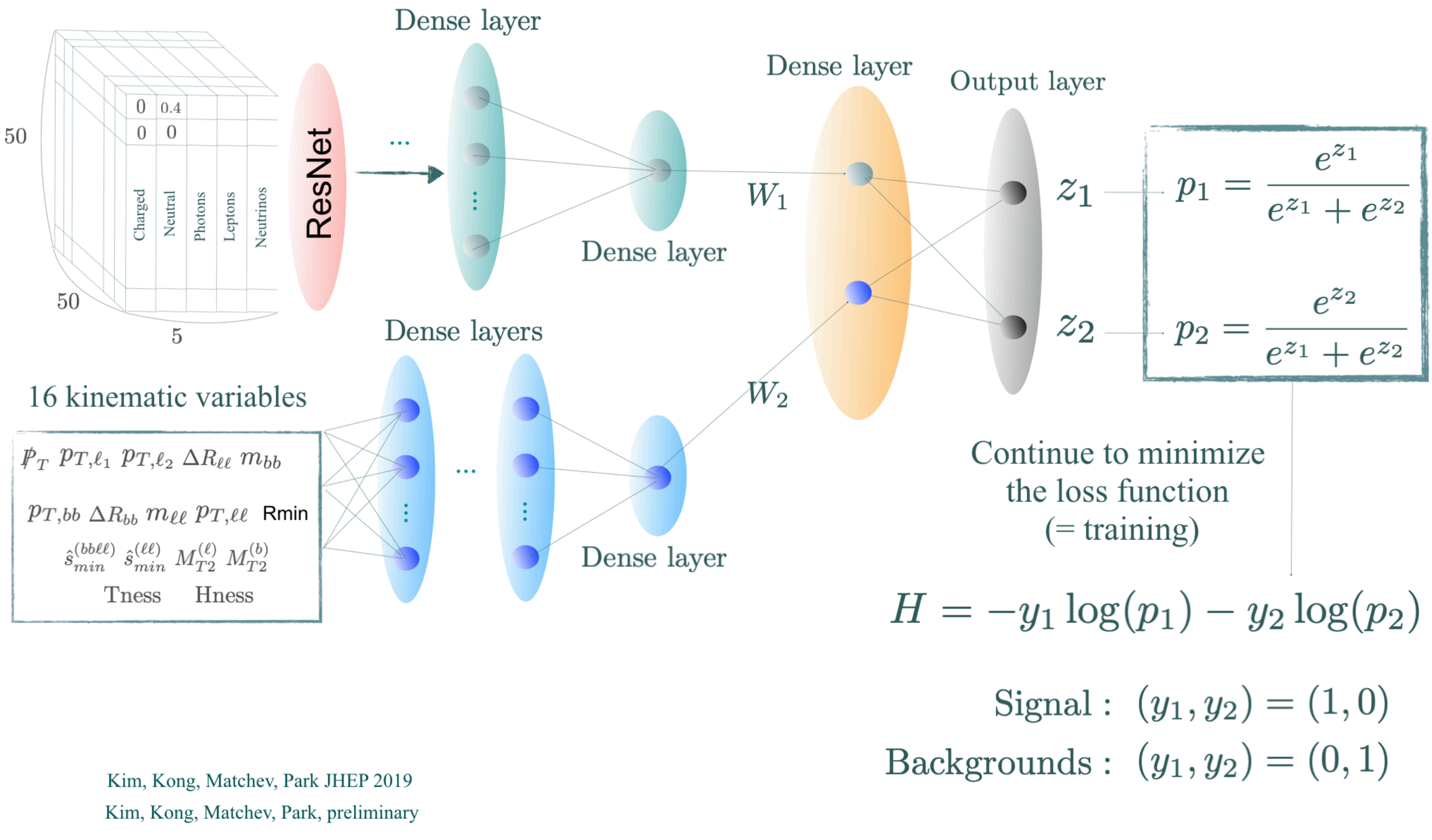
- Motivated by ResNet, we could design a much deeper network.
- The machine will be able to resolve much detailed features of 5-image data.
- This ensemble-like topology is much resilient against the change in network hyper-parameter.



Kim, Kong, Matchev, Park JHEP 2019

Kim, Kong, Matchev, Park, preliminary

Combining image data and high level kinematic variables in dense neural networks



Continue to minimize the loss function (= training)

$$H = -y_1 \log(p_1) - y_2 \log(p_2)$$

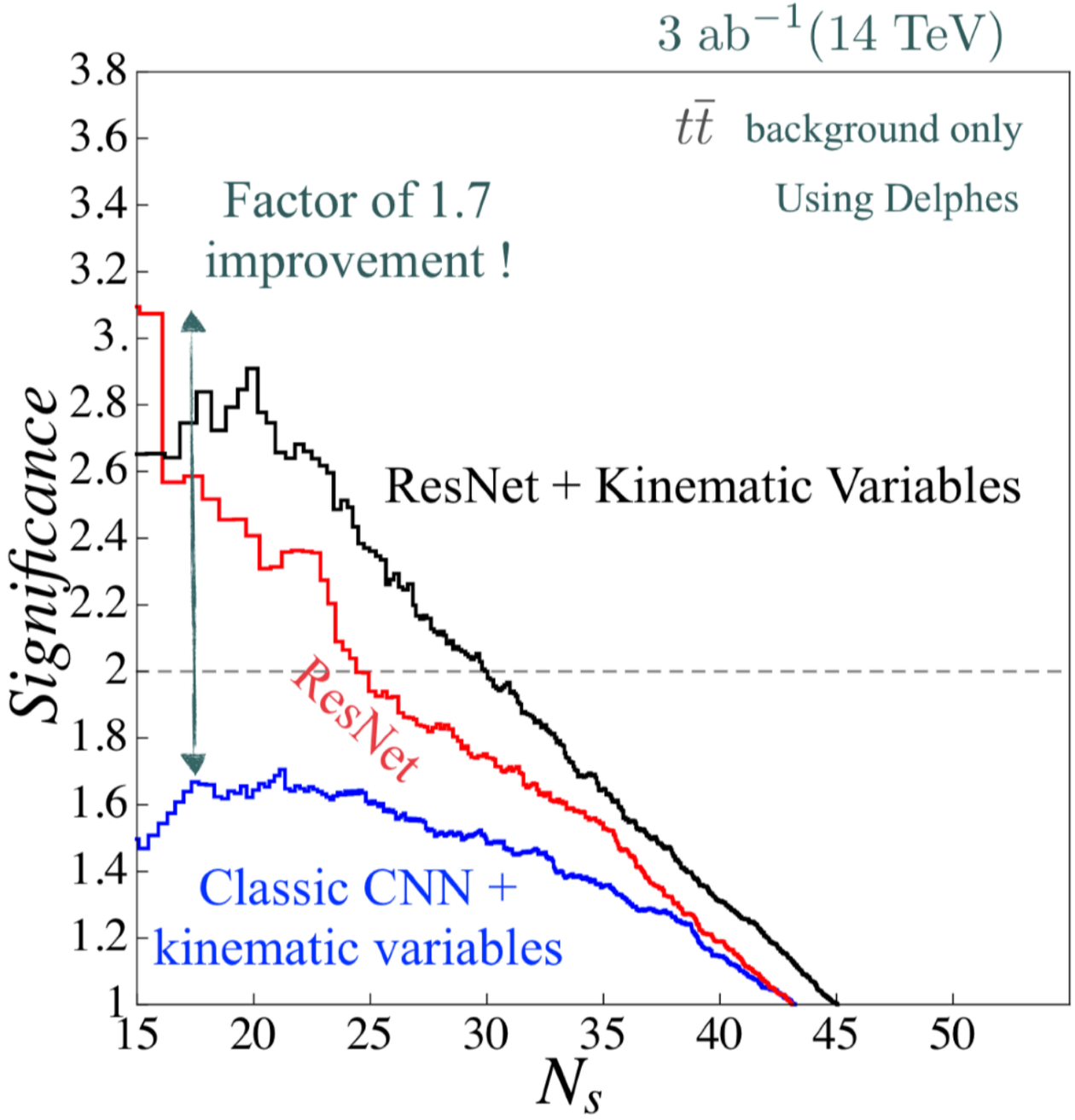
$$\text{Signal}: \ (y_1, y_2) = (1, 0)$$

$$\text{Backgrounds}: \ (y_1, y_2) = (0, 1)$$

Kim, Kong, Matchev, Park JHEP 2019

Kim, Kong, Matchev, Park, preliminary

Conclusions



- Various DL algorithms can **maximize** the reach of the Energy Frontier.
- DL, one of **Computer Frontier**, **the level of industries is way higher** than HEP applications. **We need young power!**
- We need to understand the **types of data** (**physics**) to get maximum performance with given DL architectures.