Utilizing kinematics and QCD with Machine learning @ collider

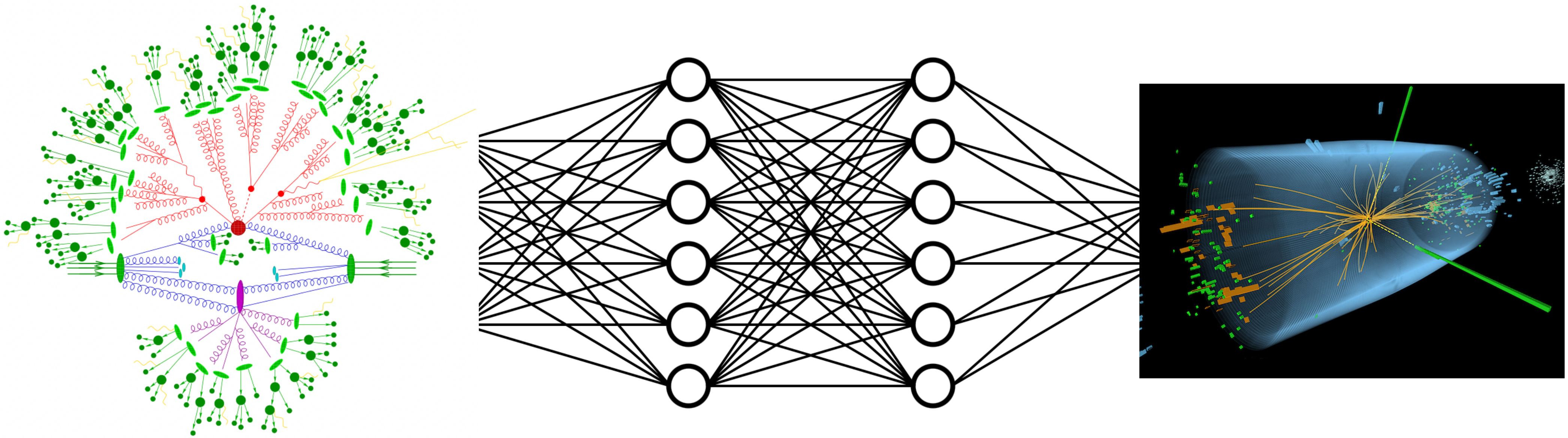
Myeonghun Park

(Seoultech)

Based on
- Overview part: Roberto F, Doojin K, KC K, Konstantin M, Prasanth S, MP (arXiv:2206.13431, accepted in **RMP**)
- Kinematics (Global) part: Doojin K, KC Kong, Konstantin M, Prasanth S MP (arXiv:2105.10126, PRD)
- QCD (Local) part: Ahmed H, MP (arXiv:2209.03898)

PPC workshop 2023

Enhancing signatures over BKG



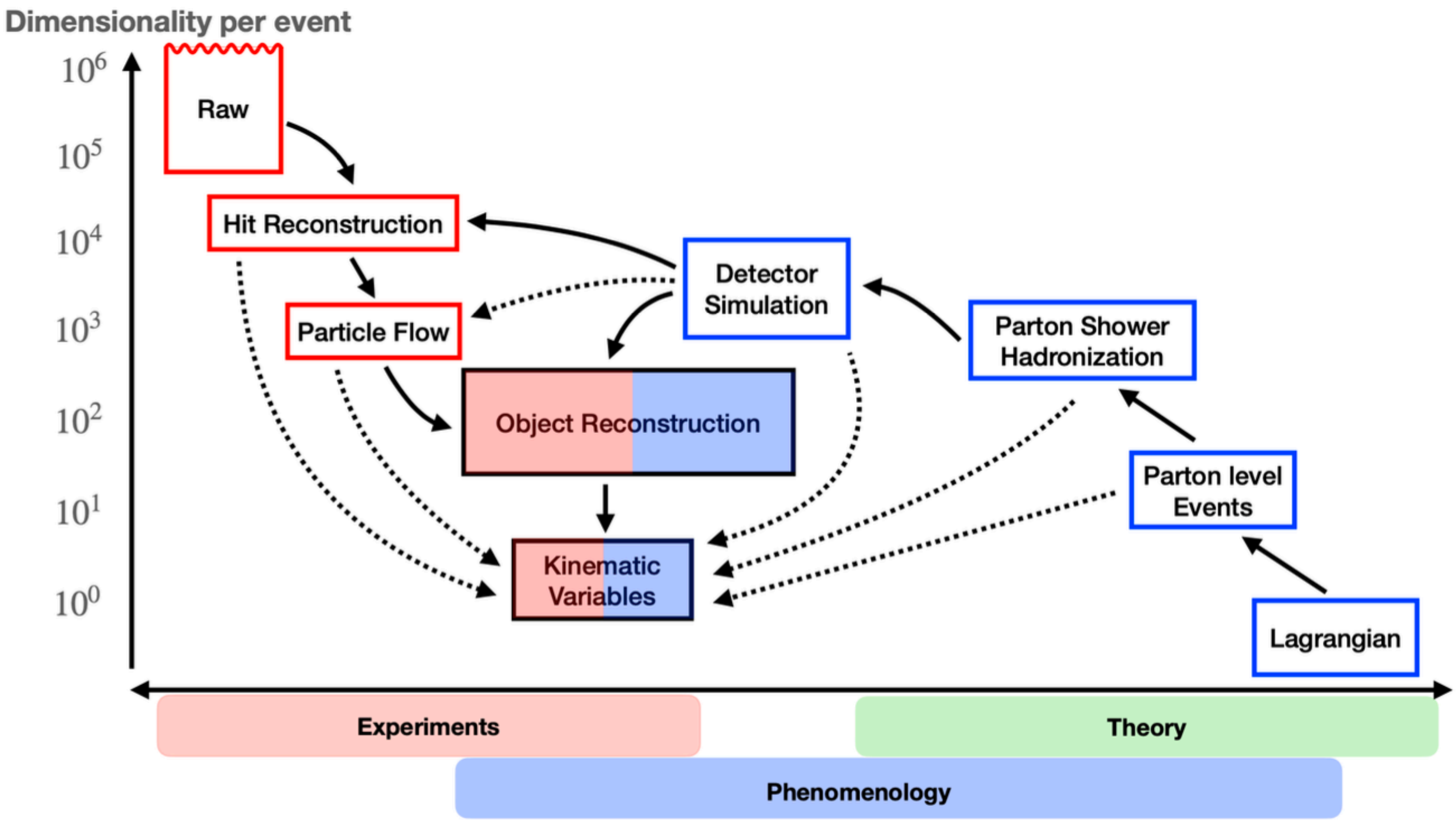
- With our elaborated **theoretical model**,

  1) Get **expectations** from **MC simulations**

  2) Get **data** from **experiments** (e.g. the LHC)

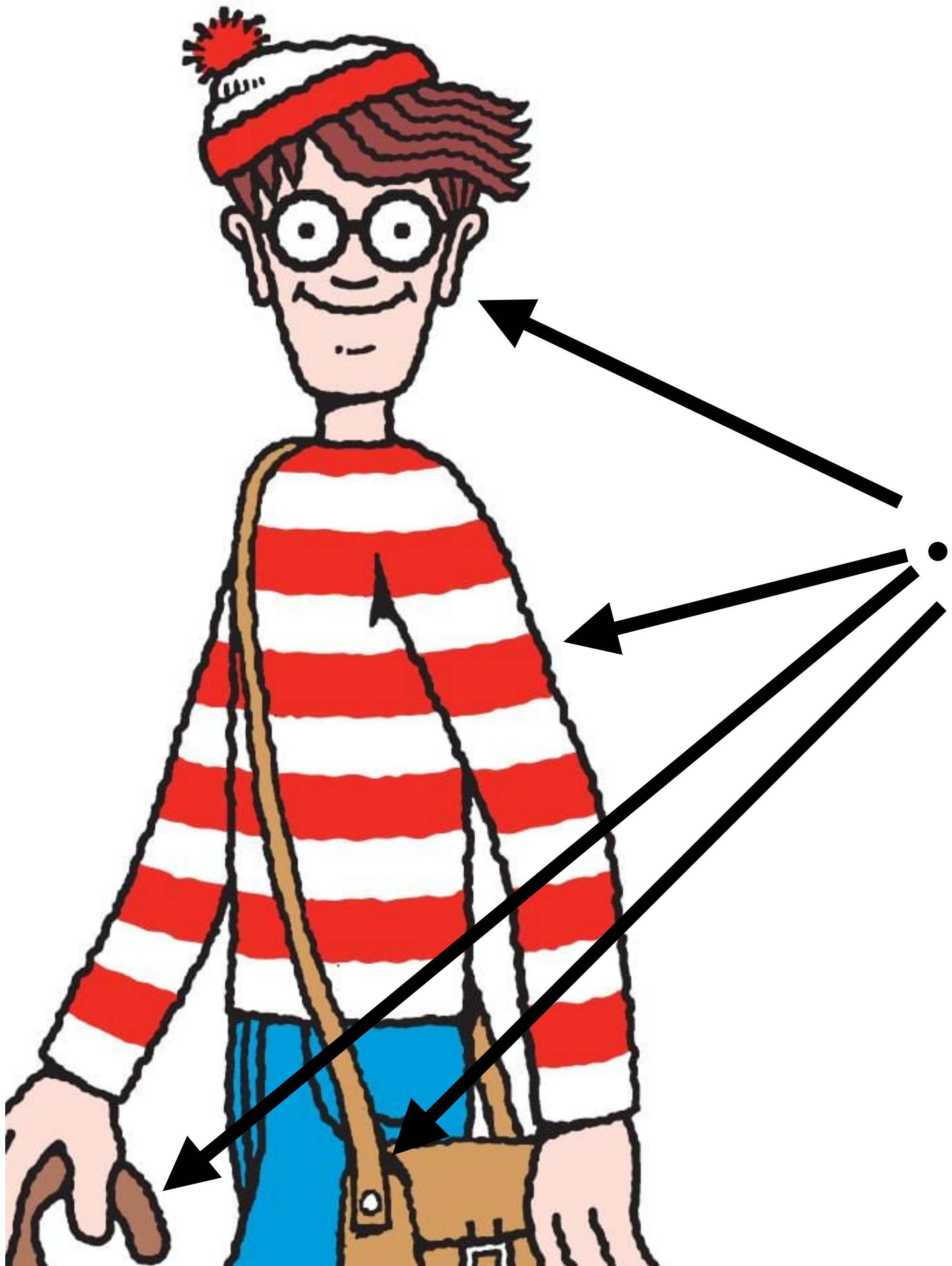
  3) Compare our expectation to data with sophisticated computer **algorithms (ML: machine learning)**

All about the dimensional reduction



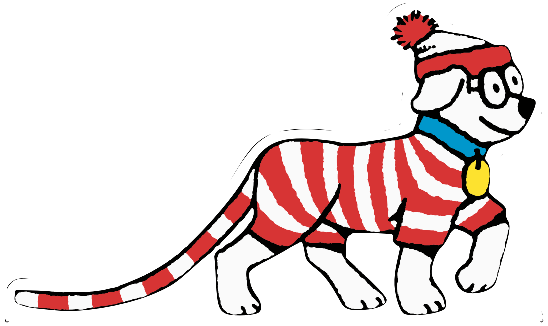
- **Kinematic variables** (features in phase-space & event-topology)
- Activities depending on a **specific Gauge** (parton shower, hadronization)

Features selection



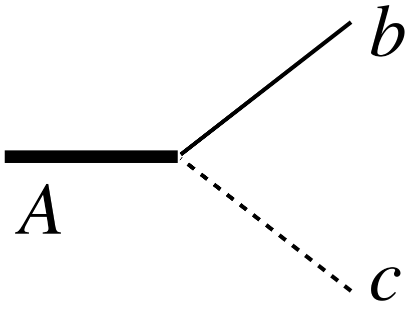
- Focusing on features of "**signal**"

Situation@Collider (maybe LHC?)

Signal	**Reducible backgrounds 1**	**Reducible backgrounds ?**	**Irreducible backgrounds**
			
$H \to b\bar{b}$	$W^{-}W^{+} \to \mu\bar{\nu}\bar{\mu}\nu$	$Z \to b\bar{b}$	$G \to b\bar{b}$
	• Different reconstructed particles • Different phase-space	• Same final sates, but different **weight** on the phase-space Well-localized	• Same final sates, but different intermediate. Huge

Extracting High-level features of a new physics

- **Kinematic variables** to utilize a **different phase-space** structures (signal, v.s. backgrounds)



(invisible)

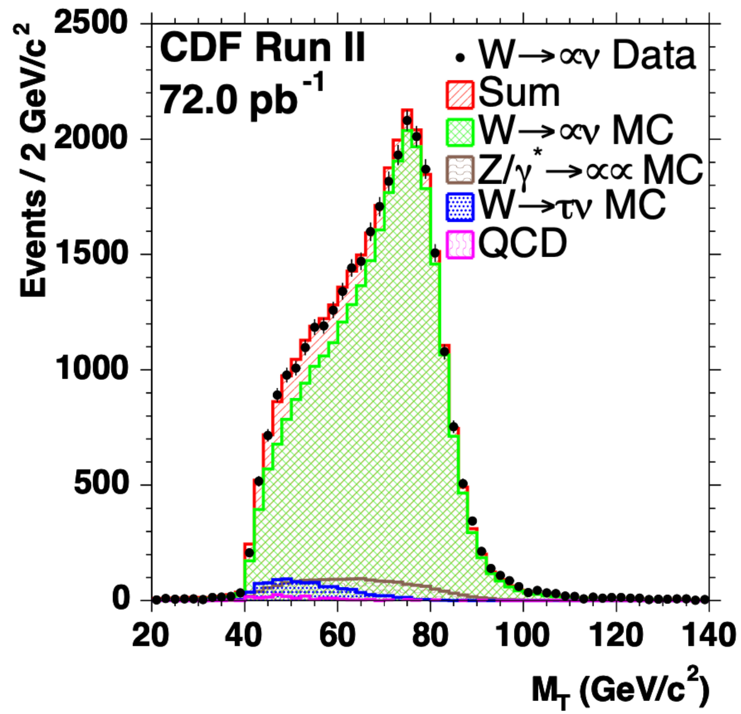
$$\theta = \{m_A\} \longrightarrow X = \{p_b^\mu, p_c^T\}$$

$$\dim(X) = 6 \to \dim(V) = 1$$

- **A human-engineered feature variable**, M_T which estimates M_A with an endpoint of its distribution

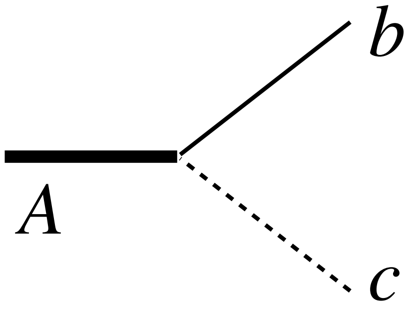
  (highly singular behavior due to its Jacobian peak)

Transverse Mass variable



Extracting High-level features of a new physics

- **Kinematic variables** to utilize a **different phase-space** structures (signal, v.s. backgrounds)

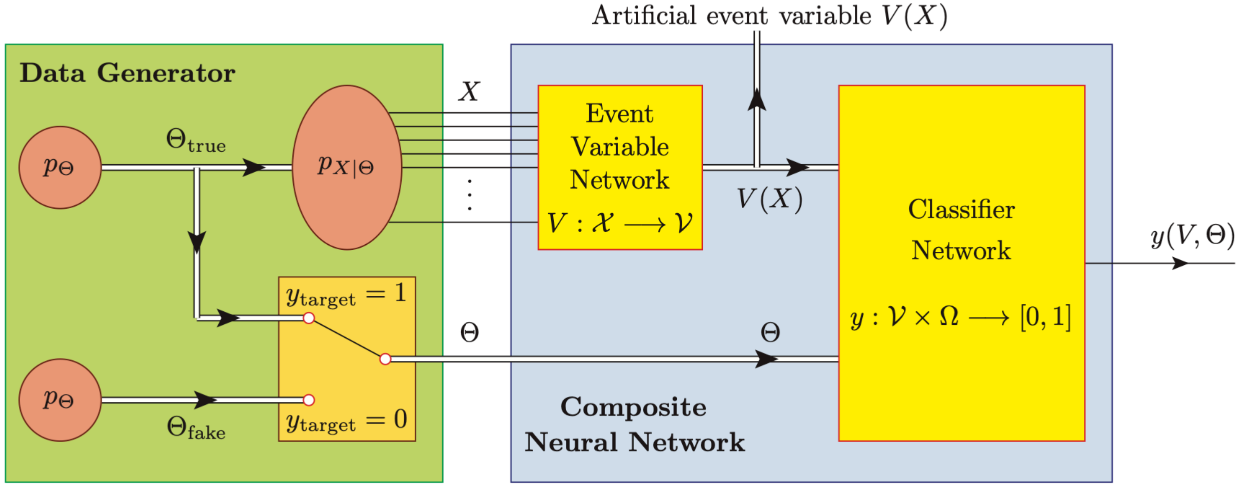


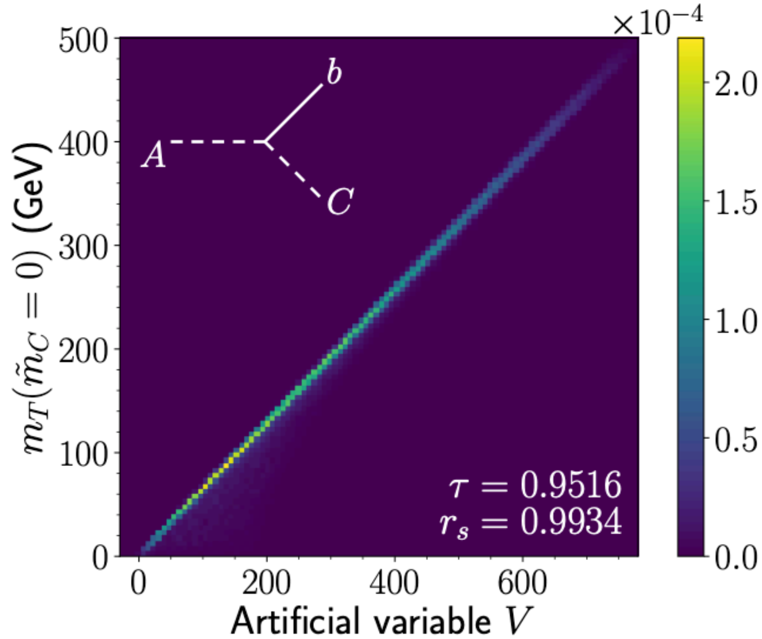
(invisible)

$$\theta = \{m_A\} \longrightarrow X = \{p_b^\mu, p_c^T\}$$

$$\dim(X) = 6 \rightarrow \dim(V) = 1$$

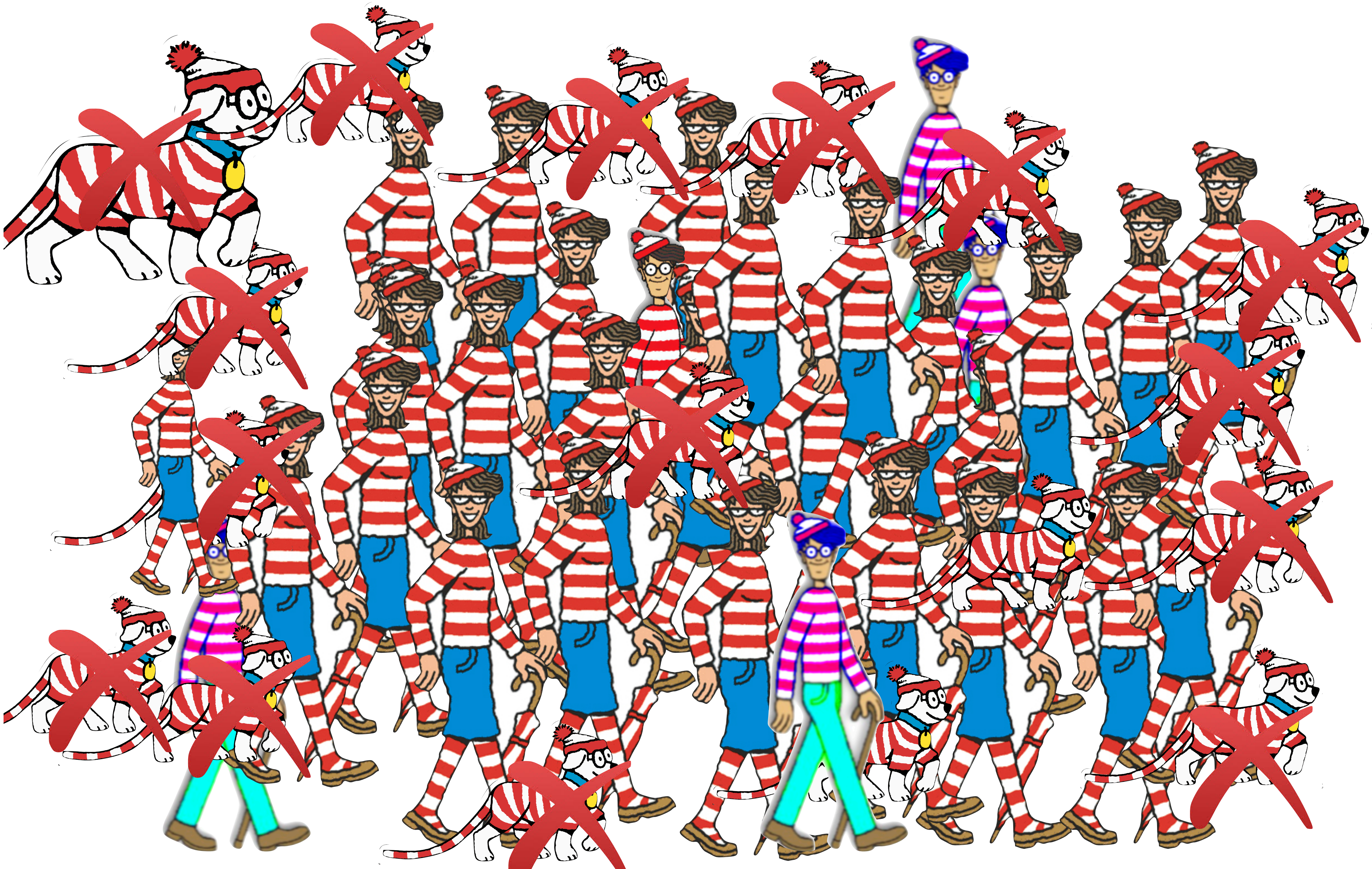
A Neural-Network can design an event-variable
(by enforcing information-bottleneck to NN)



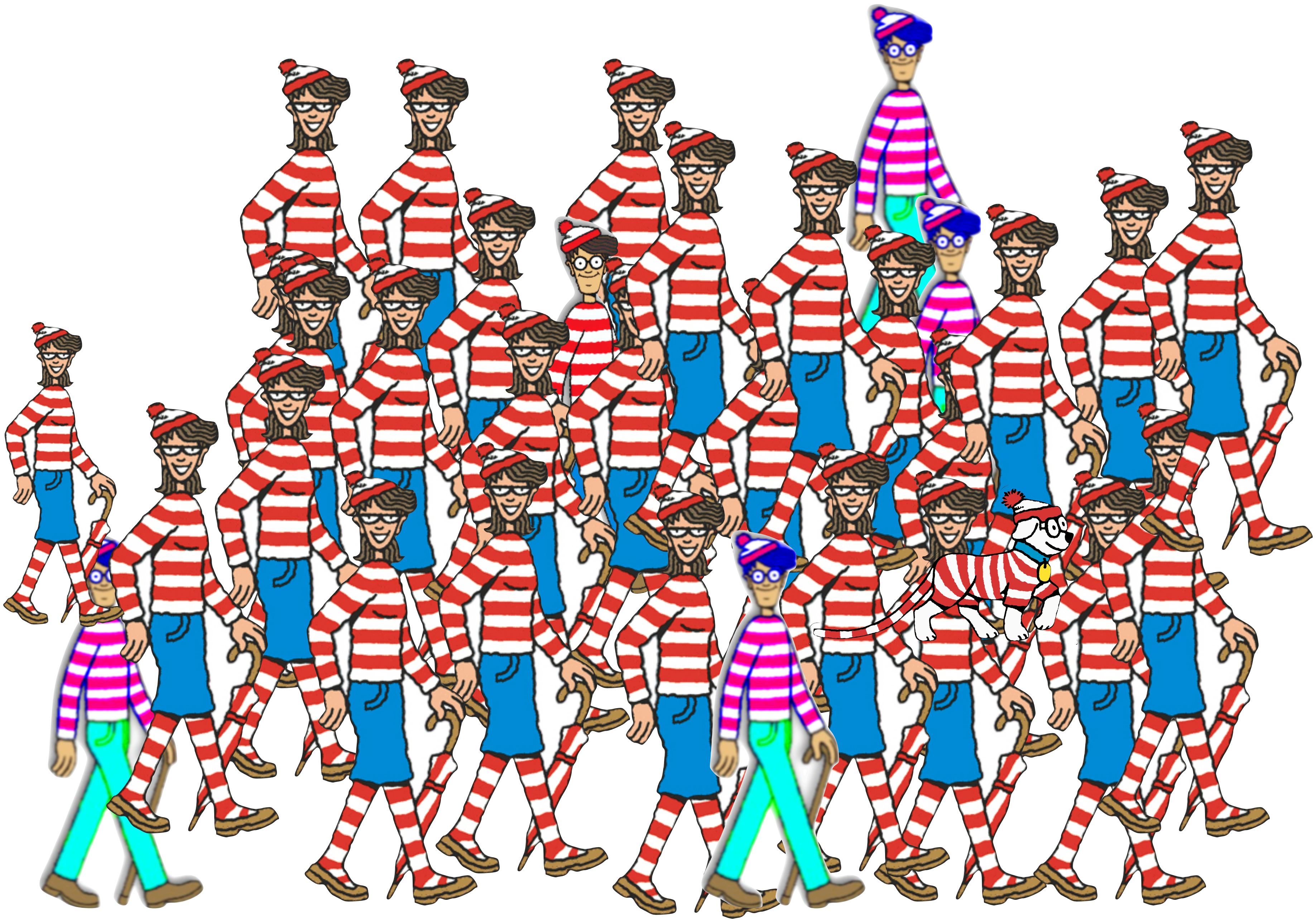


Doojin Kim, KC Kong, Konstantin Matchev, Prasanth, **MP**. (2023)

- "High-level" kinematic variables to remove **easy reducible backgrounds**



- "High-level" kinematic variables to remove **easy reducible backgrounds**



With low-level information

- Low-level information:
  - simple variables like p_T, η, $\Delta\phi$, $\cos\theta_{jj}$, and any basic kinematic variables including m_{jj} , ...
  - **Four-vector** of reconstructed particles.

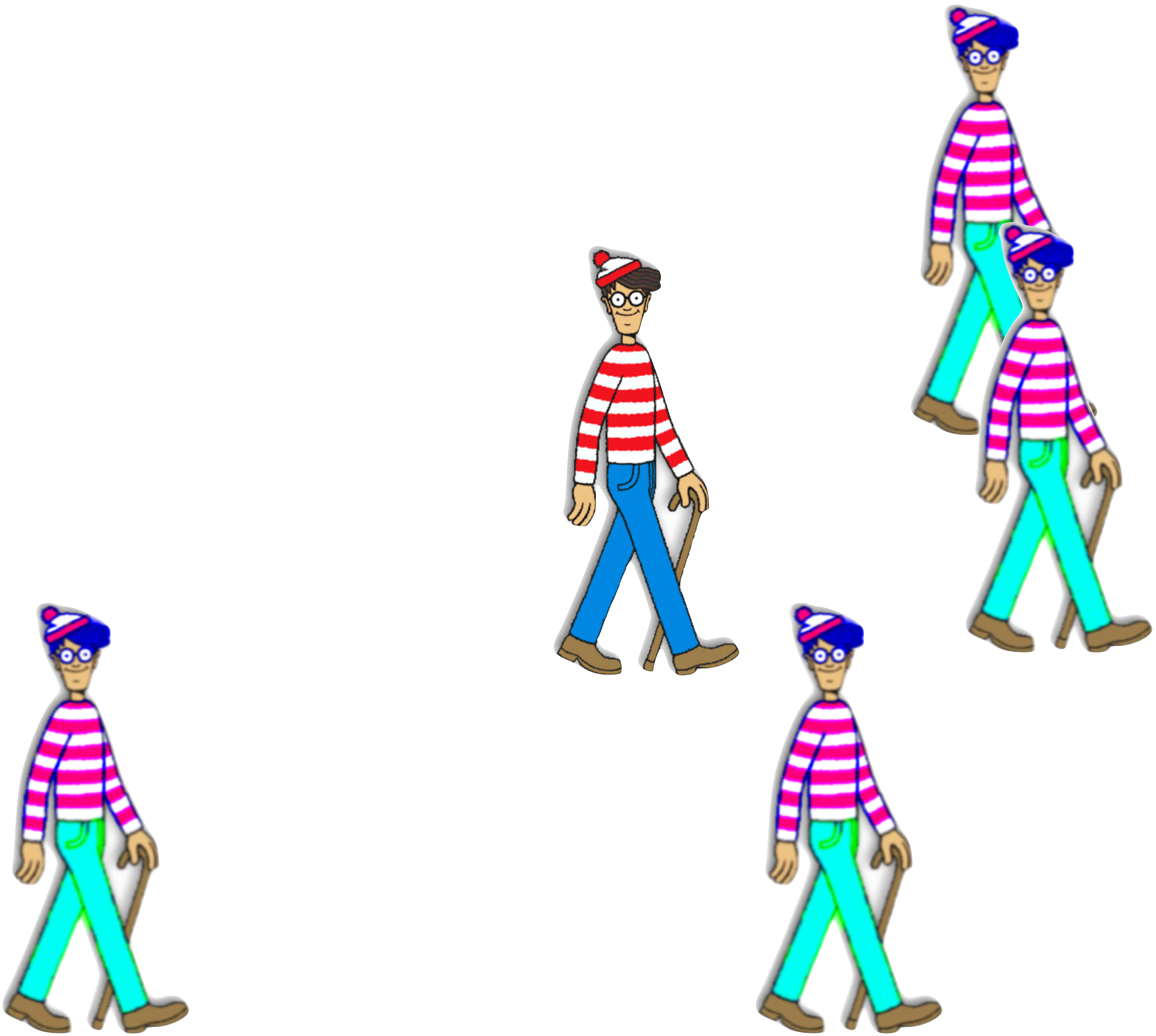
- Provide a "**freedom**" to a **Neural Nets (expensive GPU machine)** so that it can design some decision criteria to suppress backgrounds.

  - Freedom : More data and more complicated and deep design of NN

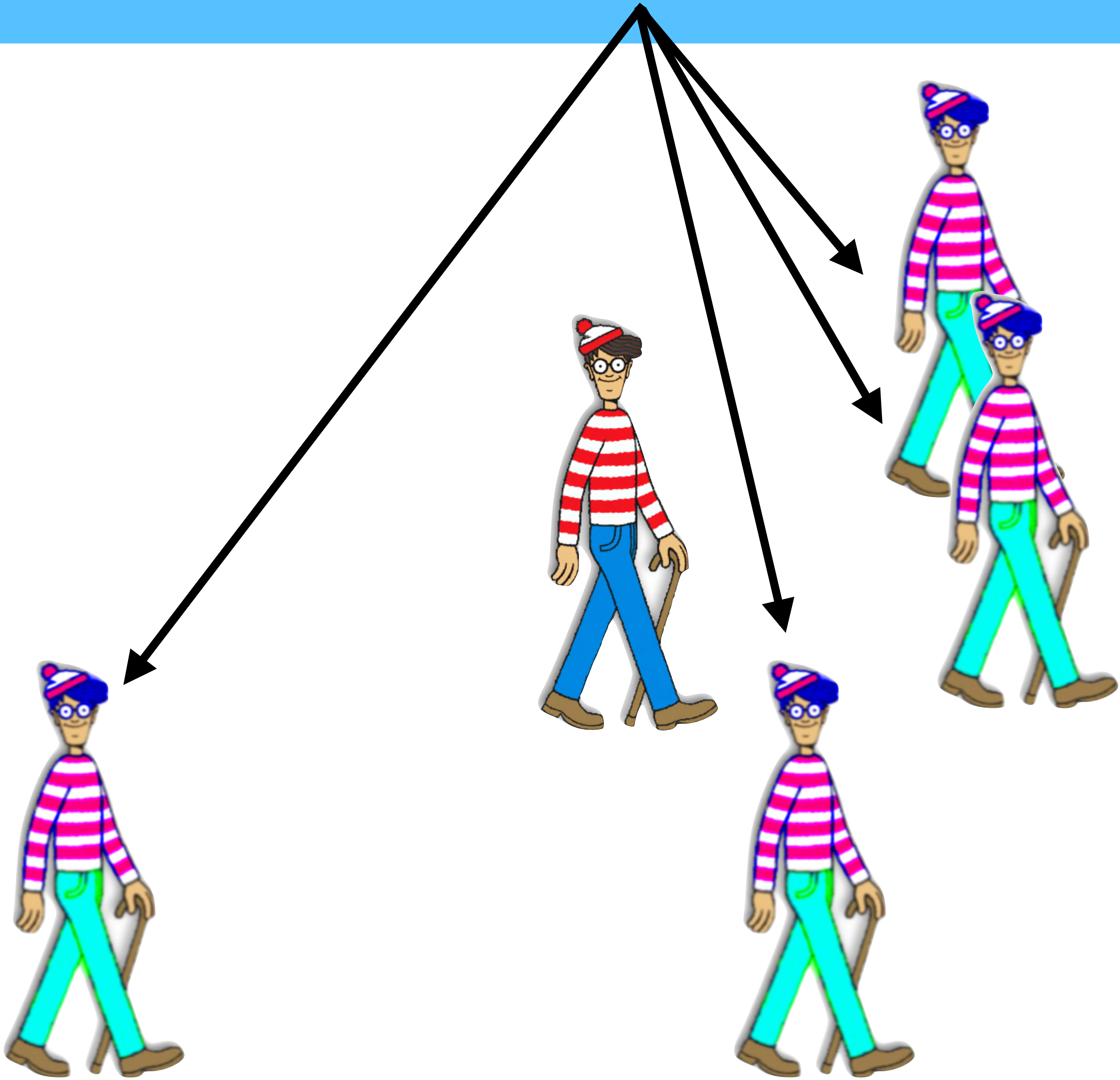
- We can borrow any fancy Neural Networks, including Recurrent Neural net, Graph Neural net, Teacher-Student net, and so on... (So we need to give eyes on the recent ML developments, and enjoy shopping...)

- Deep NN with "low-level" kinematics to remove **somewhat complicated backgrounds**

- Deep NN with "low-level" kinematics to remove **somewhat complicated backgrounds**



So leftovers are the imposters with different **color**



Orthogonal information to the Kinematics

- Differences in kinematics are from **"high P_T" region, i.e. reconstructed (reco) level**

  - Telling us about the structure of "Feynman-diagram"
    (Event-topology, Mass spectrum)

- We can further utilize $|\mathscr{M}|^2$ differences (Density bounded by phase-space)
  - e.g.) Decaying angle of the Higgs

- **Differences in radiation patterns of a Gauge charge** are coming from **"soft P_T" region**

  - eg) Telling us about the state under a gauge group, $SU(3)_C$

More than Kinematics difference

- In many cases, the **soft QCD radiation patterns** from signals are different from Backgrounds. (e.g. : rapidity gap)

Jason Gallicchio, Matthew D. Schwartz 2010

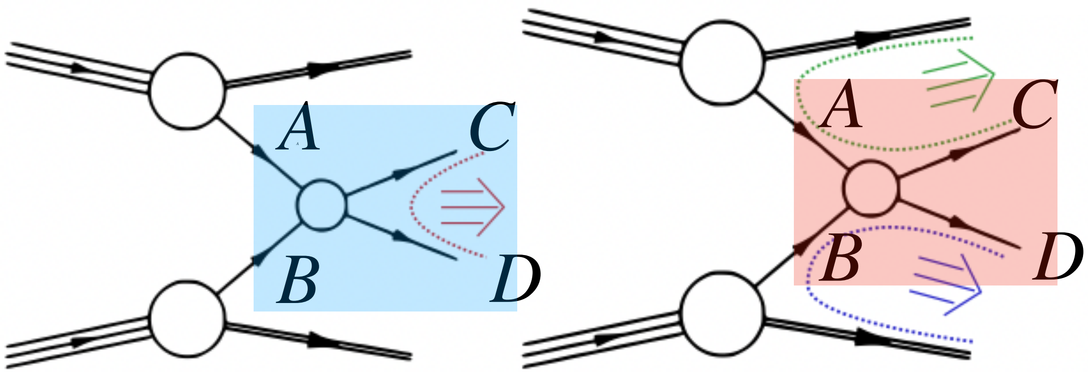
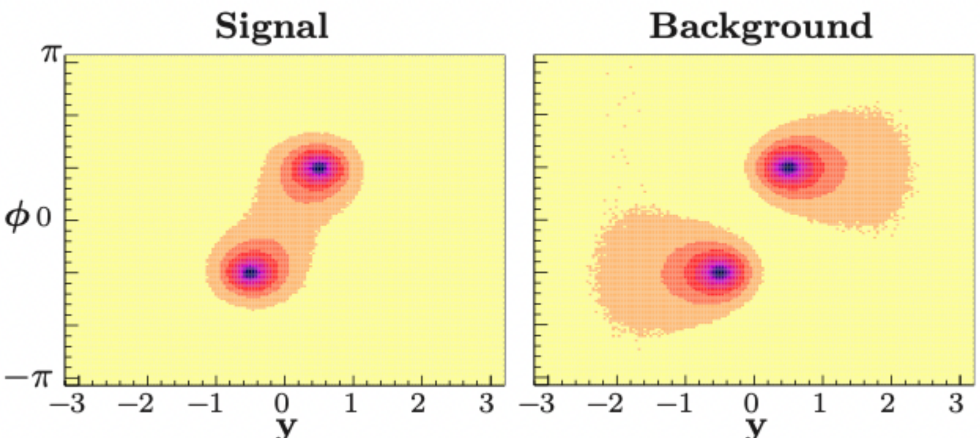


FIG. 1: Possible color connections for signal ($pp \to H \to b\bar{b}$) and for background ($pp \to g \to b\bar{b}$).



$$gg \to h \to b\bar{b}$$

$$\mathrm{Tr}[T^A T^B]\mathrm{Tr}[T^C T^D]$$

V.S.

$$gg \to b\bar{b}$$

$$\mathrm{Tr}[T^A T^C]\mathrm{Tr}[T^B T^D]$$

$$\mathrm{Tr}[T^A T^D]\mathrm{Tr}[T^B T^C]$$

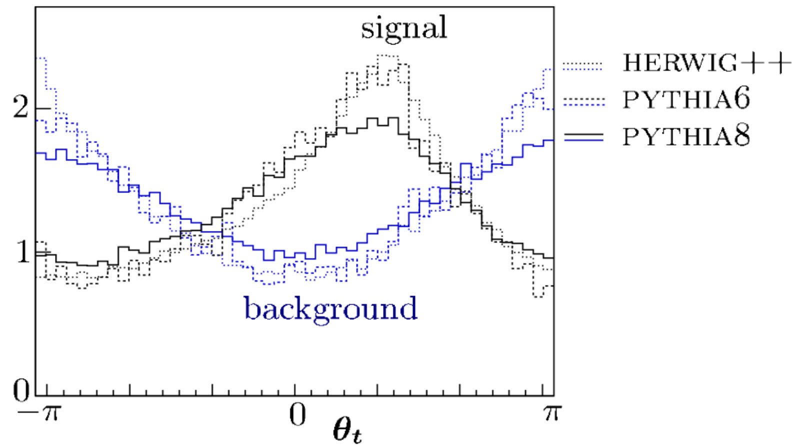
Utilizing QCD information

- One can design a QCD variable, for example a pull-vector

$$\vec{t} \equiv \sum \frac{p_T^i |r_i|}{p_T^{\text{jet}}} \vec{r}_i$$

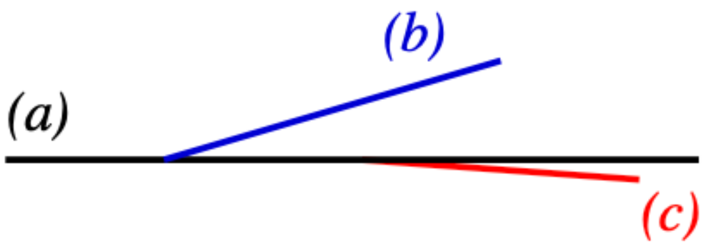
(Jason Gallicchio, Matthew D. Schwartz 2010)

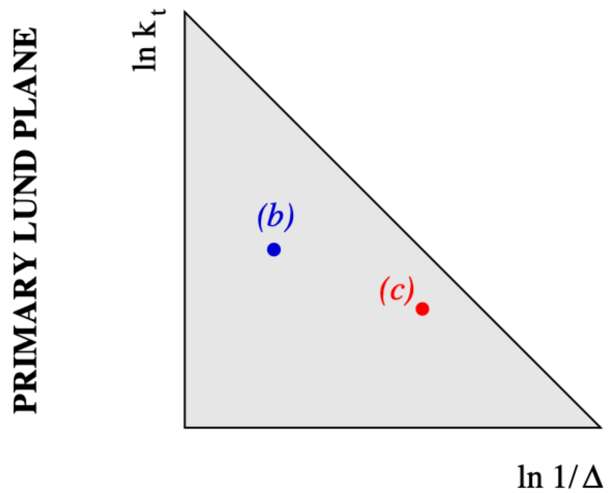
provides an one-dimensional feature



- Or one can get two-dimensional features,

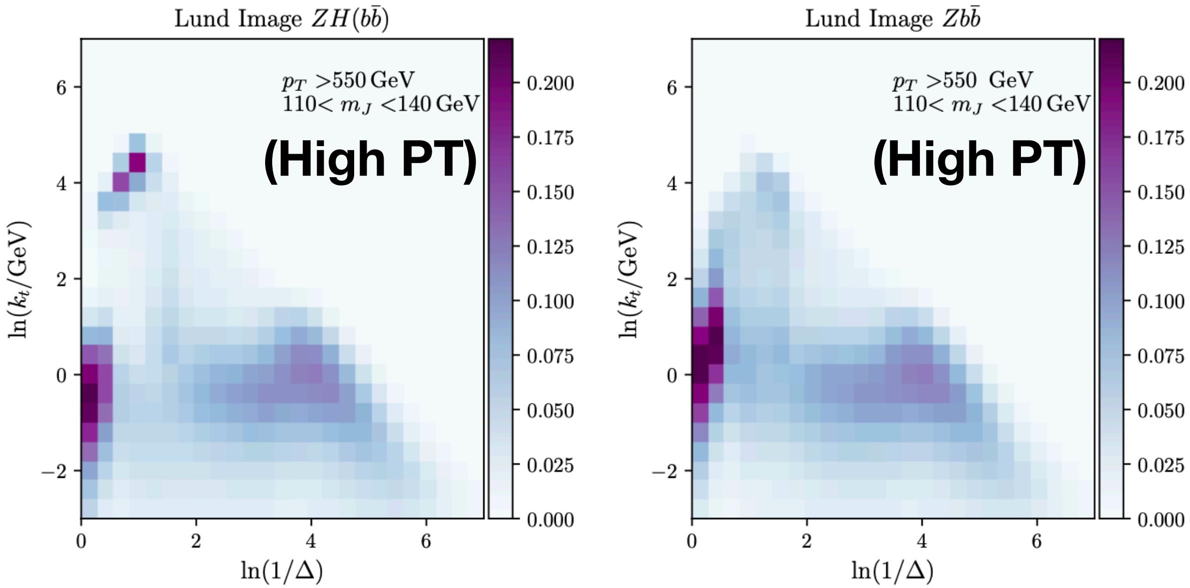
(Frederic A. Dreyer, Gavin P. Salam, Gregory Soyez 2018)



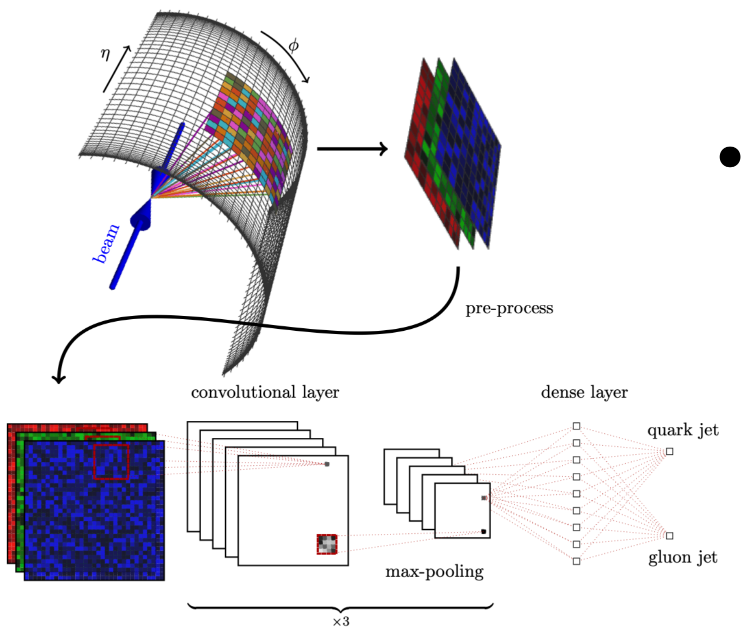


Fully utilizing QCD information?

- One needs to understand differences in **"image"**



(Charanjit K. Khosa, Simone Marzani, 2021)

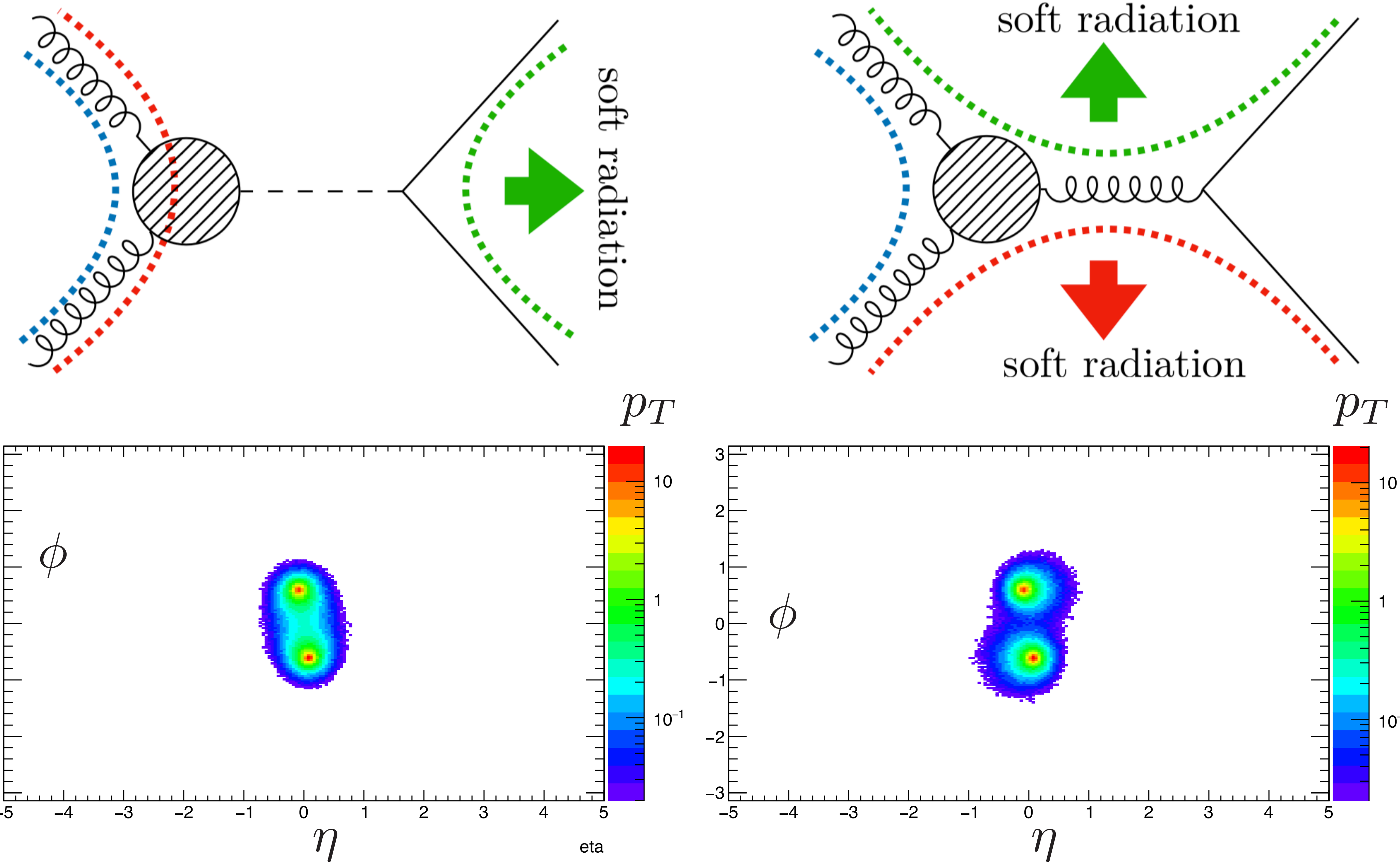


- A **neural network** (designed to understand a **picture**) can tell differences in QCD

  - Pixels are energy deposits from various sub-detectors (e.g. : tracks, e-cal, h-cal)

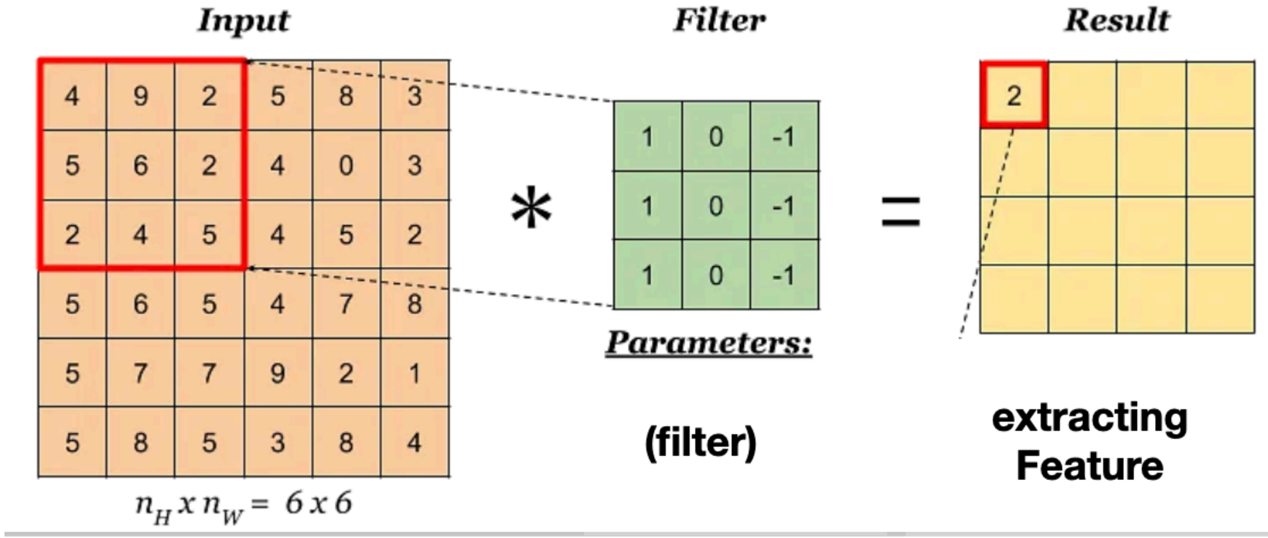
M. Schwartz et.al. arXiv:1612.01551

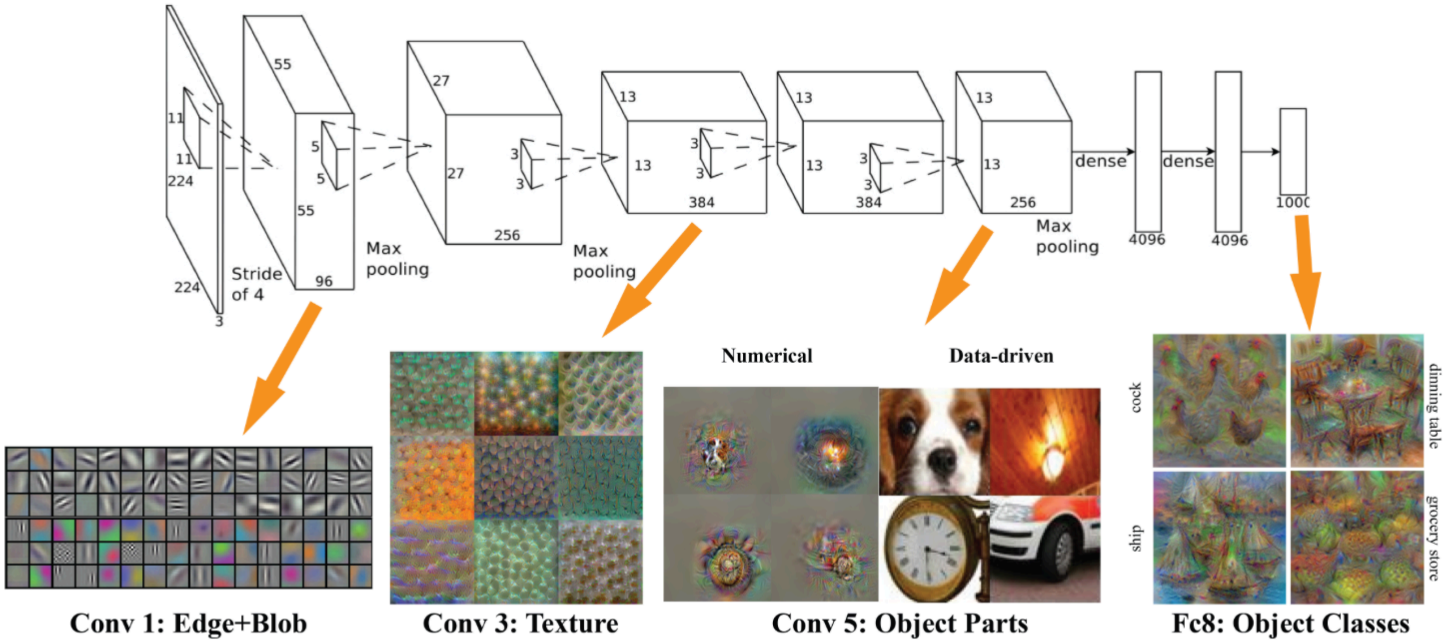
- Consider "**orthogonal**" method to kinematics; **QCD Color-flow**



Energy deposits

Conventional Image recognition
(applying series of filters)

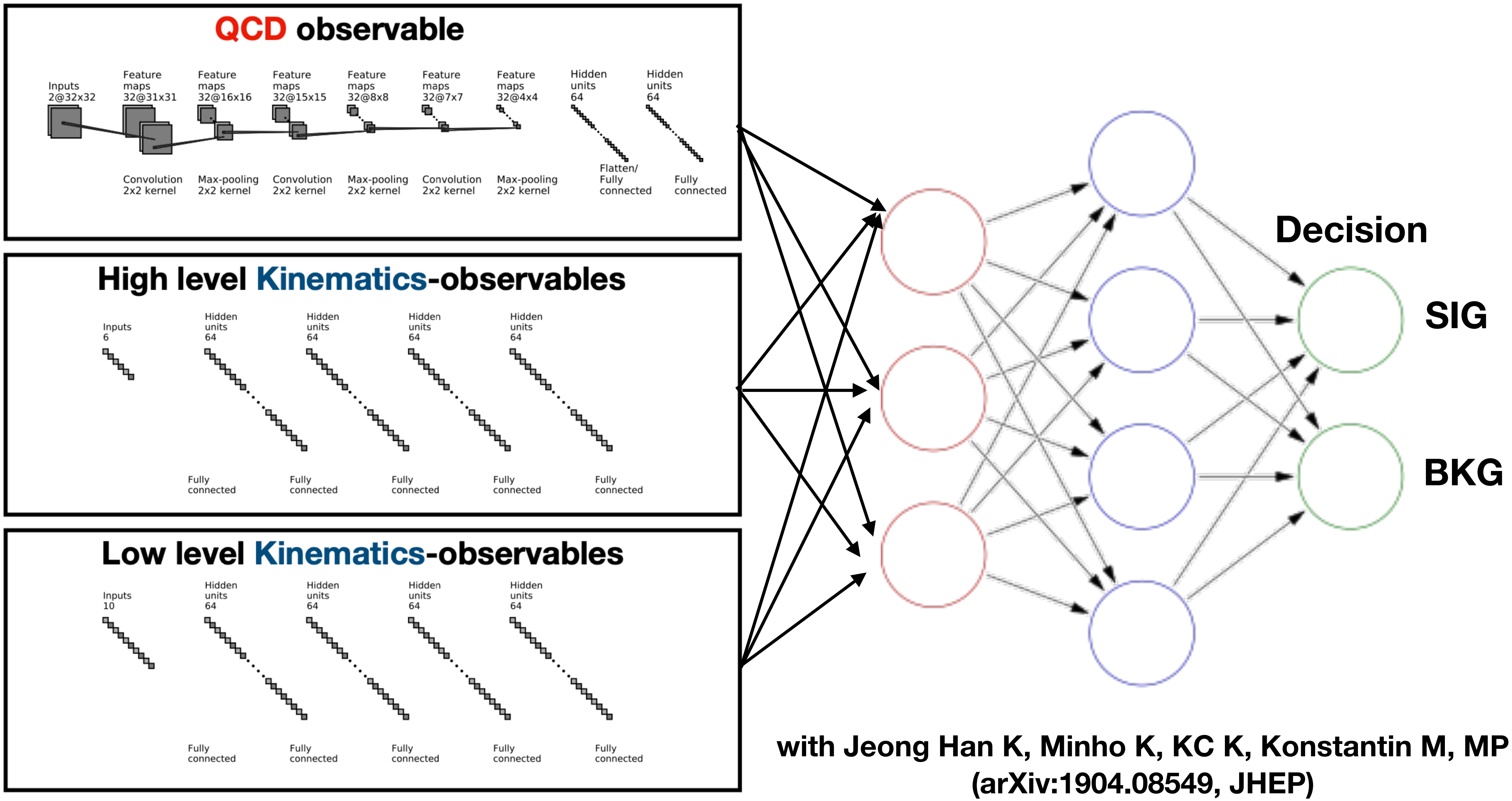




Basic structure of Image recognition using CNN

Combining pieces together

- A multi-modal learning can be done.

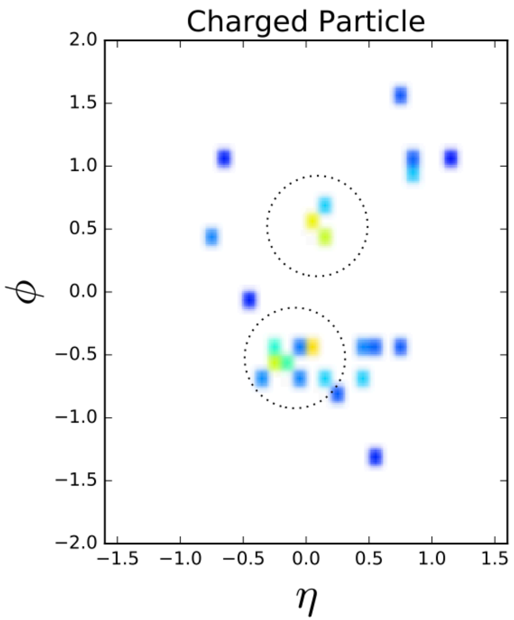


with Jeong Han K, Minho K, KC K, Konstantin M, MP (arXiv:1904.08549, JHEP)

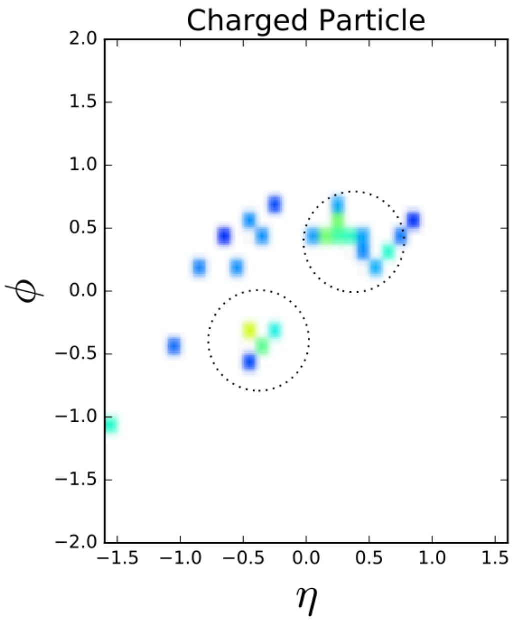
So, are we OK?

The problem of conventional ML (1)

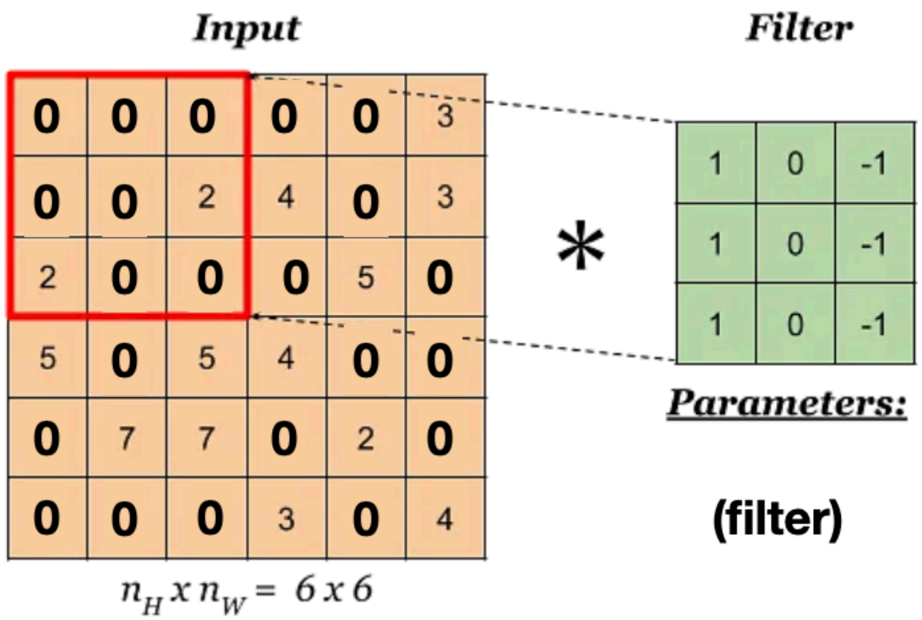
- The direct use of a neural network **(designed for commercial image)** is **not suitable (= not efficient) to our physics cases.**
- The "image" from our LHC data is very **sparse**



$h \to b\bar{b}$



$b\bar{b}$
in Backgrounds

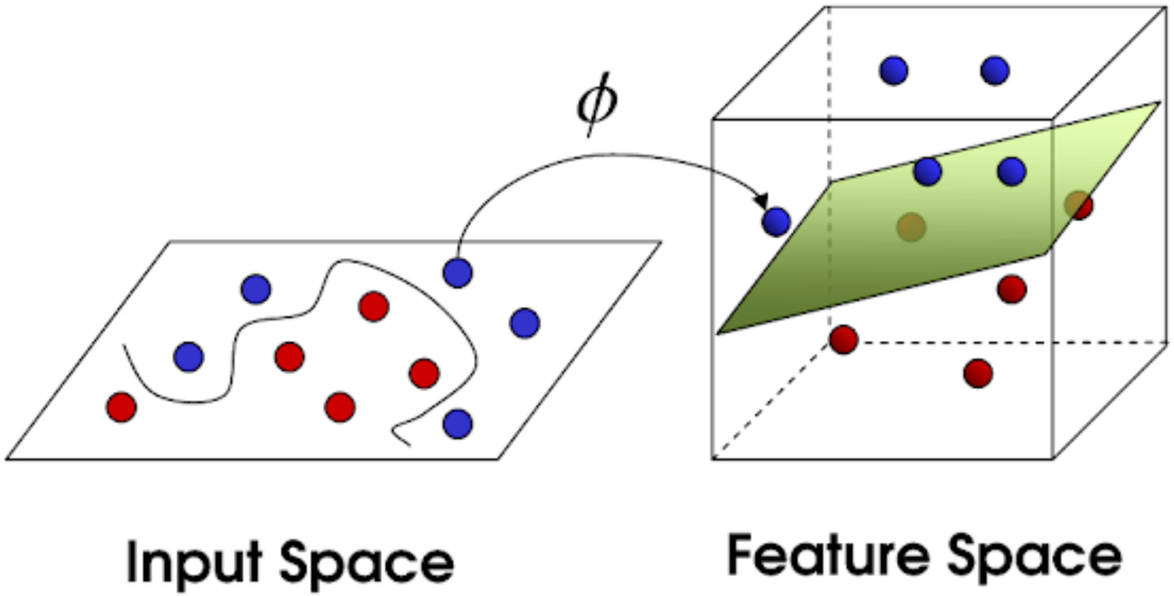


Mostly empty image (value = 0)
Any meaning to apply a "filter"?

One solution: Kernel Method

- We can provide a good kernel to separate data **efficiently,** namely with a few and sparse "image" data by making **"linearly" separable**

"OLD" Kernel Method

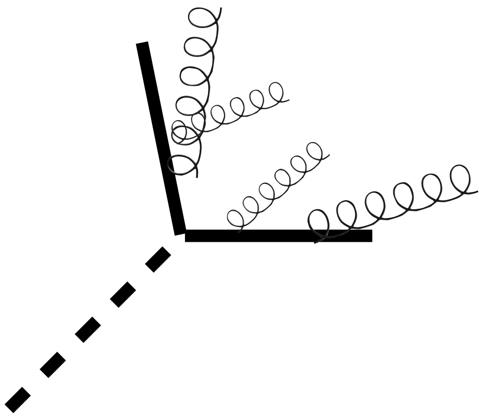


- Designing a kernel requires a domain knowledge (based on our expertise)
  - This means "**old**"
  : **Conventional ML : end - to - end (**Blackbox): No human intervention

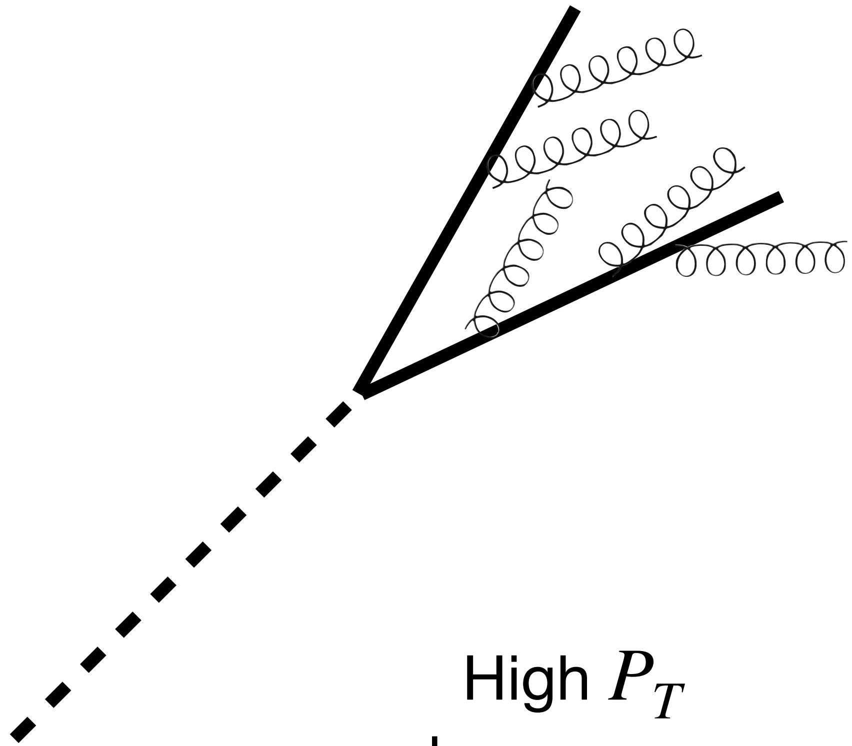
The problem in data (2)

- The soft patterns are not fully detangled with kinematics.

$$m_{jj} \propto p_{T_1} p_{T_2} \Delta R$$



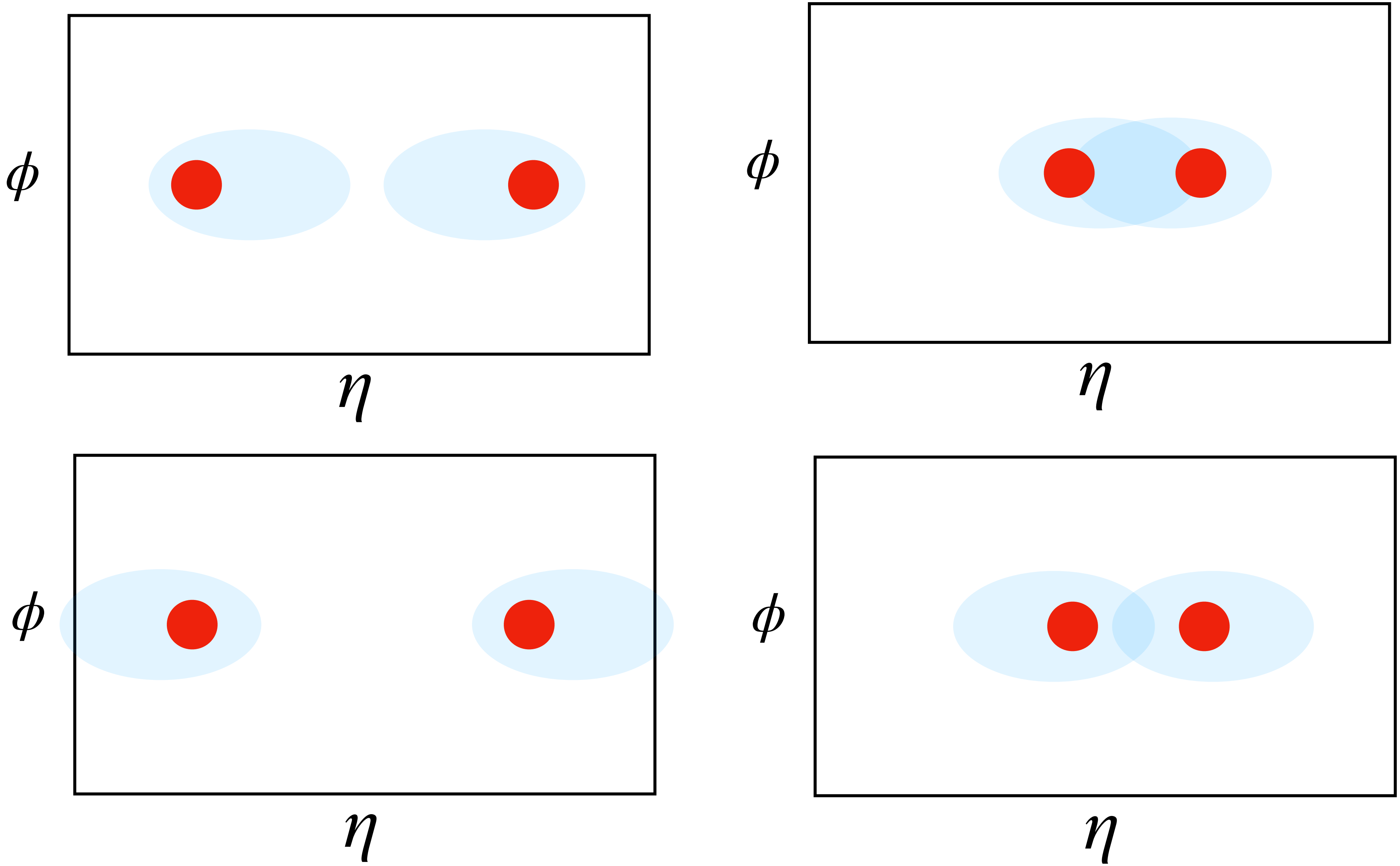
Low P_T
phase-space



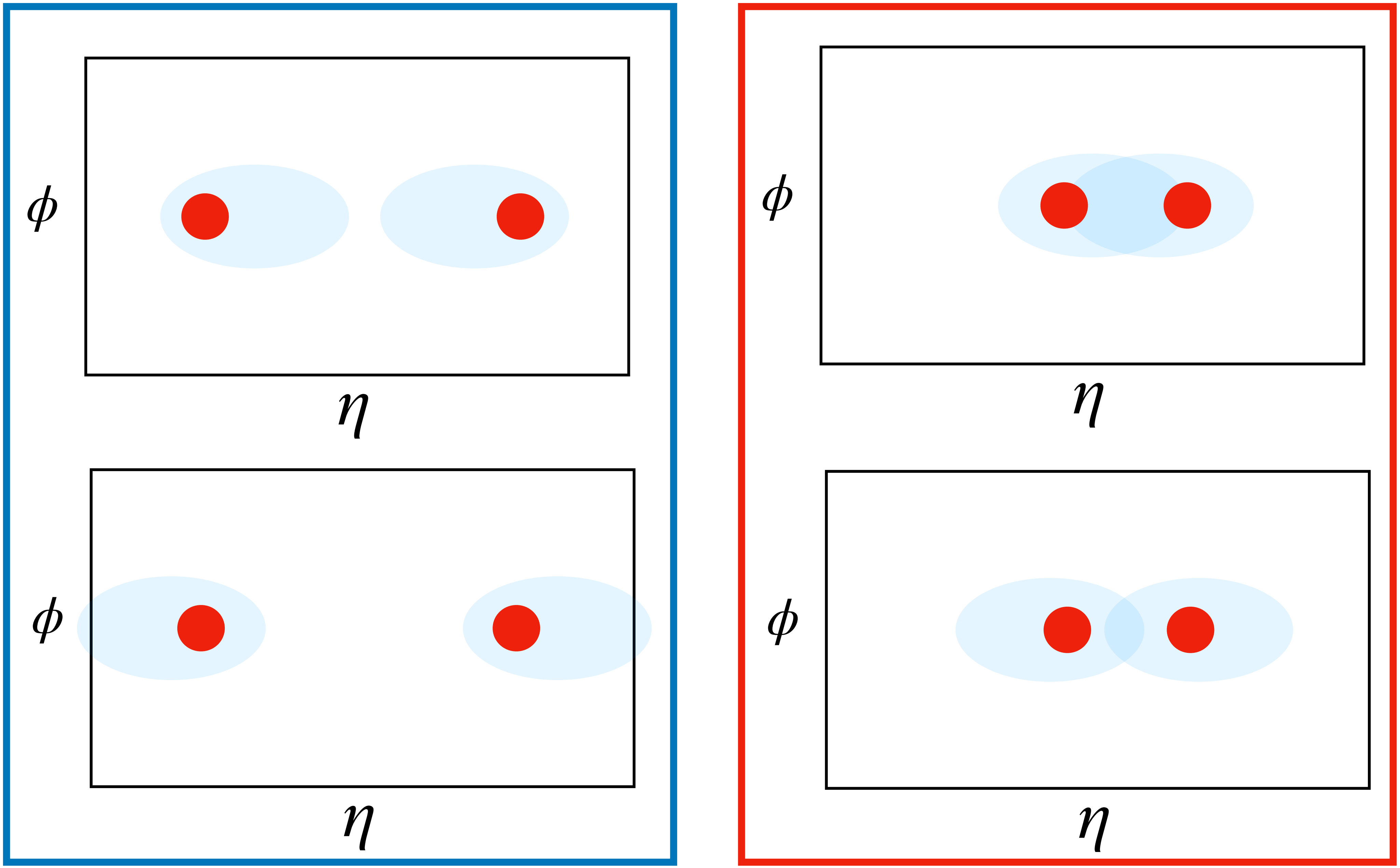
High P_T
phase-space

The problem in data (2)

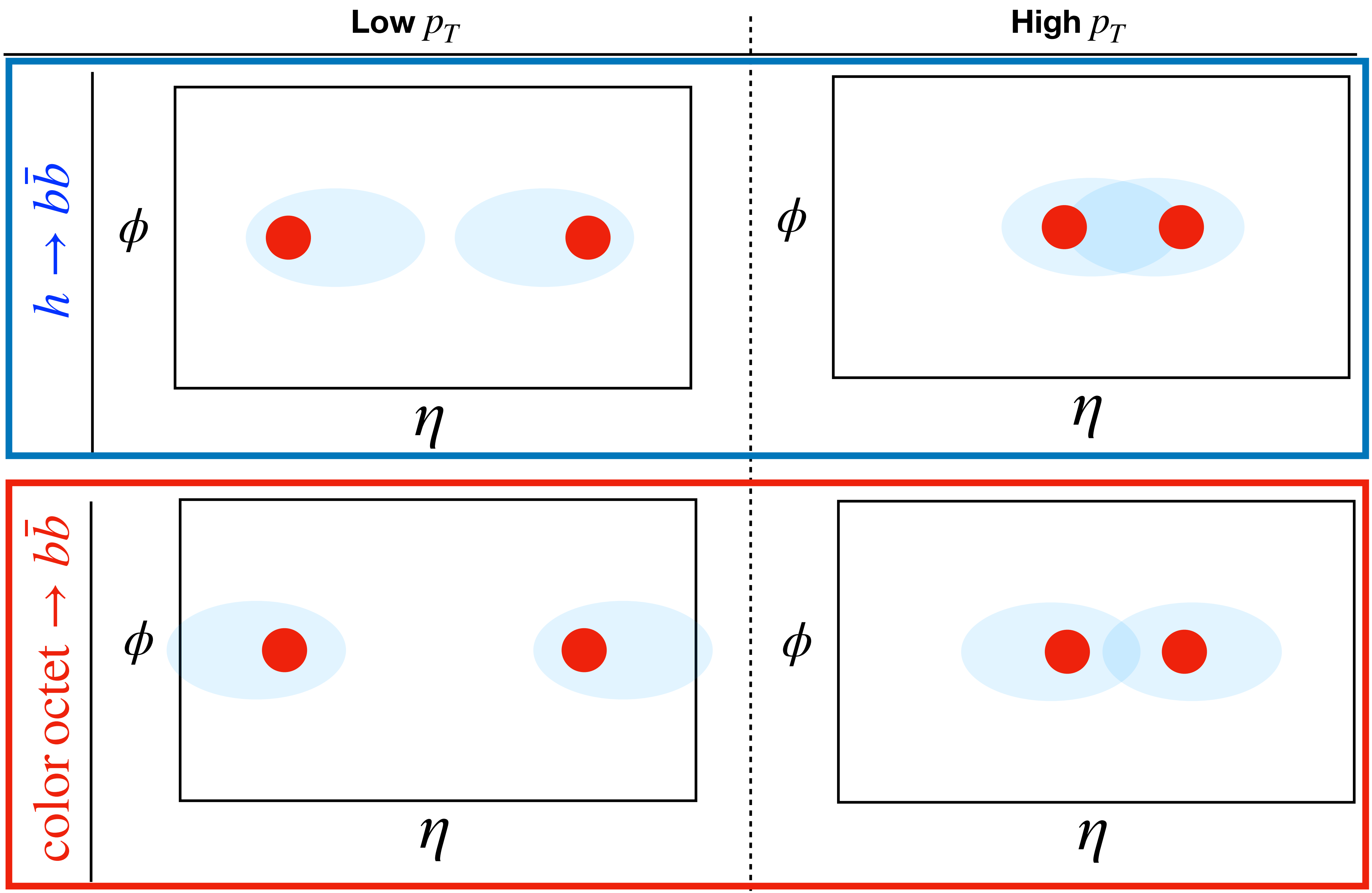
- Let's make two categories : Divide below into two categories

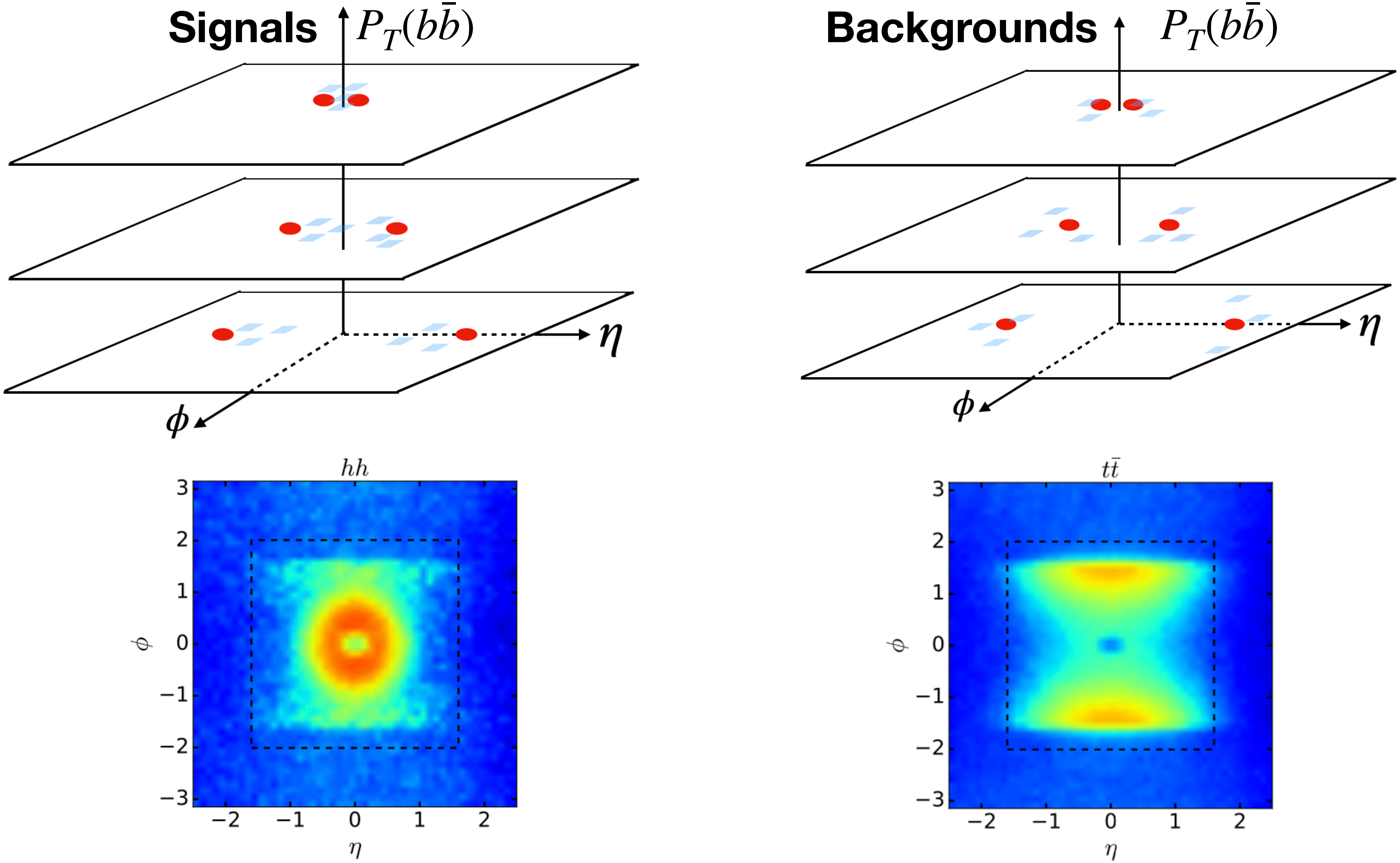


- **A quick trial: Attentions are on hot cores**



- Due to the **softness of radiations**, everyone (even ML) can easily gets focused on **hot cores ($b/\bar{b}$)**. **To train, we need more more more data**

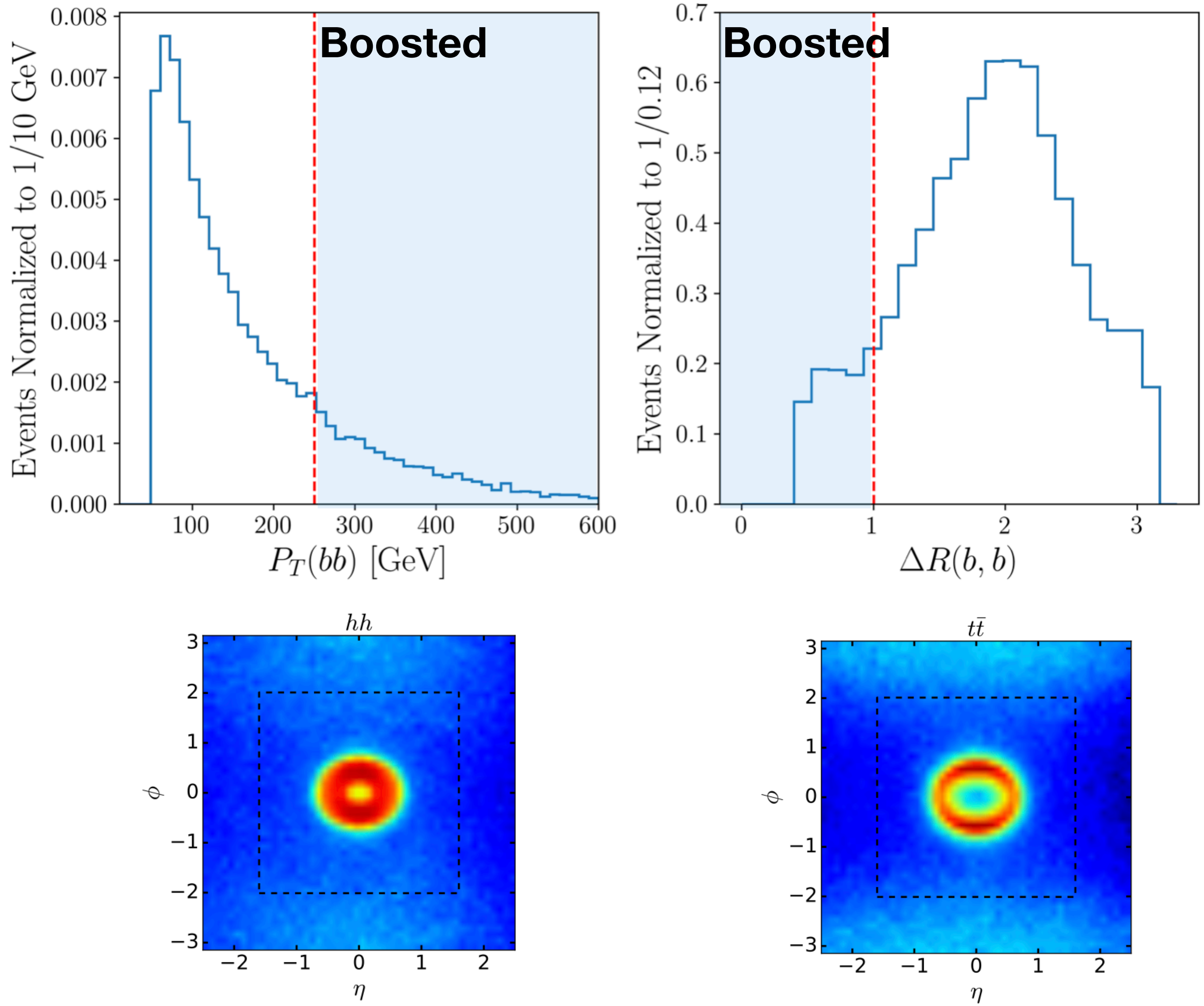




- Conventional Machine Learning can not focus on soft-patterns, **rather on different kinematics.**

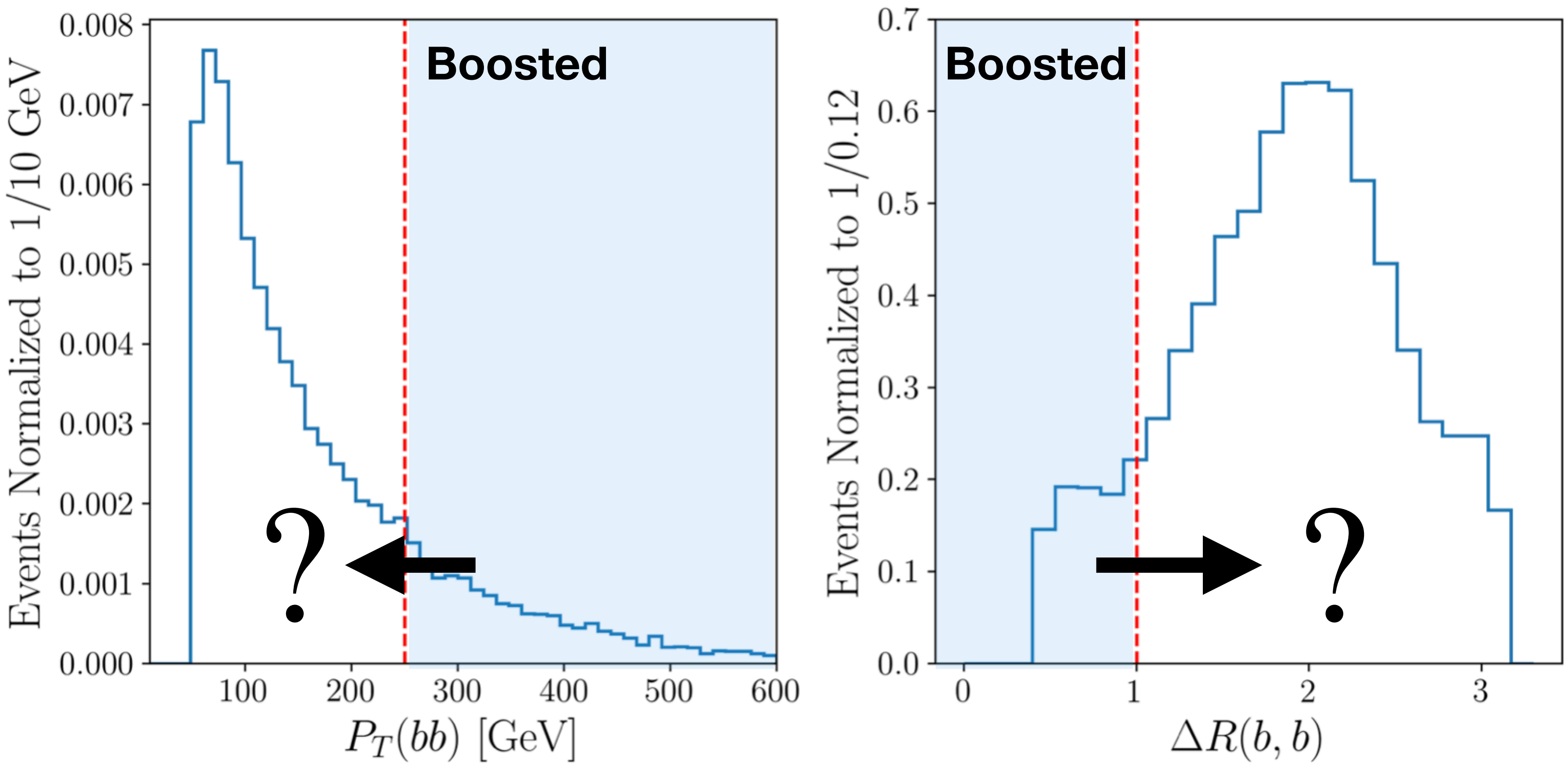
- **It requires "BIG" data** to pay attention to **soft patterns.**

One solution: focusing on small region



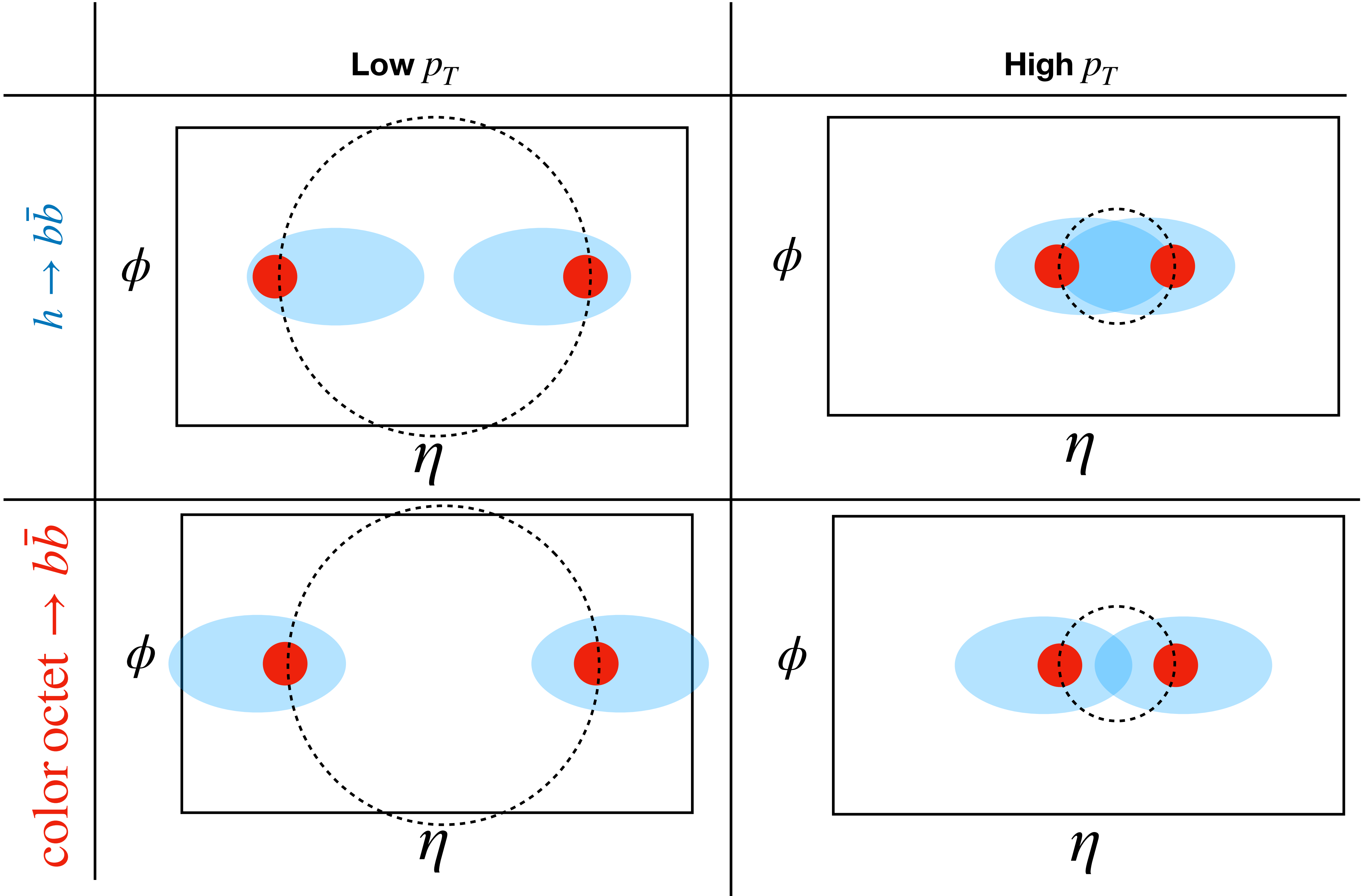
- In a boosted region, the dependency on "kinematics" becomes mild

Want to use "Full p_T range"



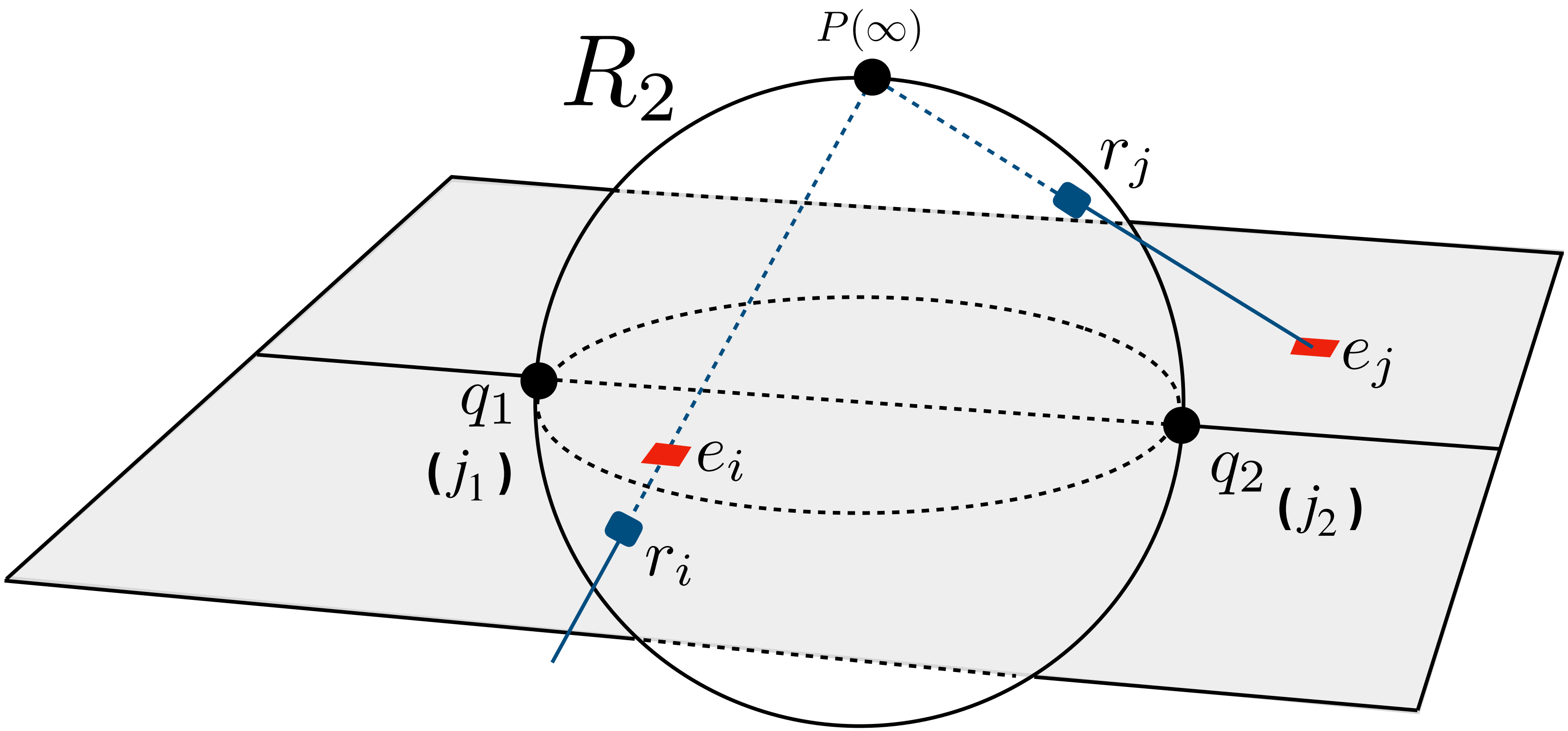
- "Easy" solution demands a **huge price**: the **statistics**. We want **to collect as many as events, statistics!!!**

- A **binary problem**, either "**inside**"or "**outside**" a circle.



Inverse stereographic projection

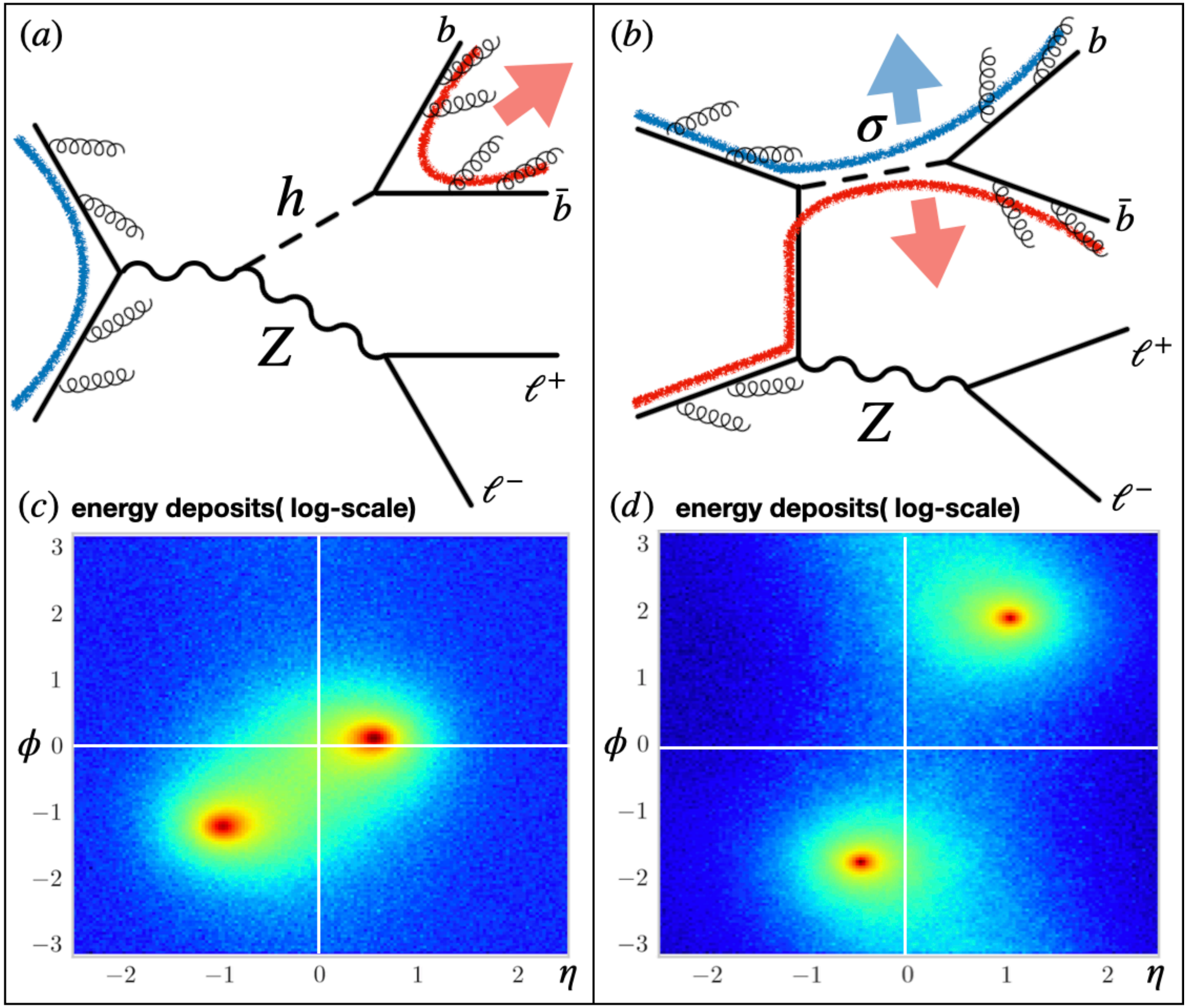
(a.k.a. "Riemannian" Kernel)



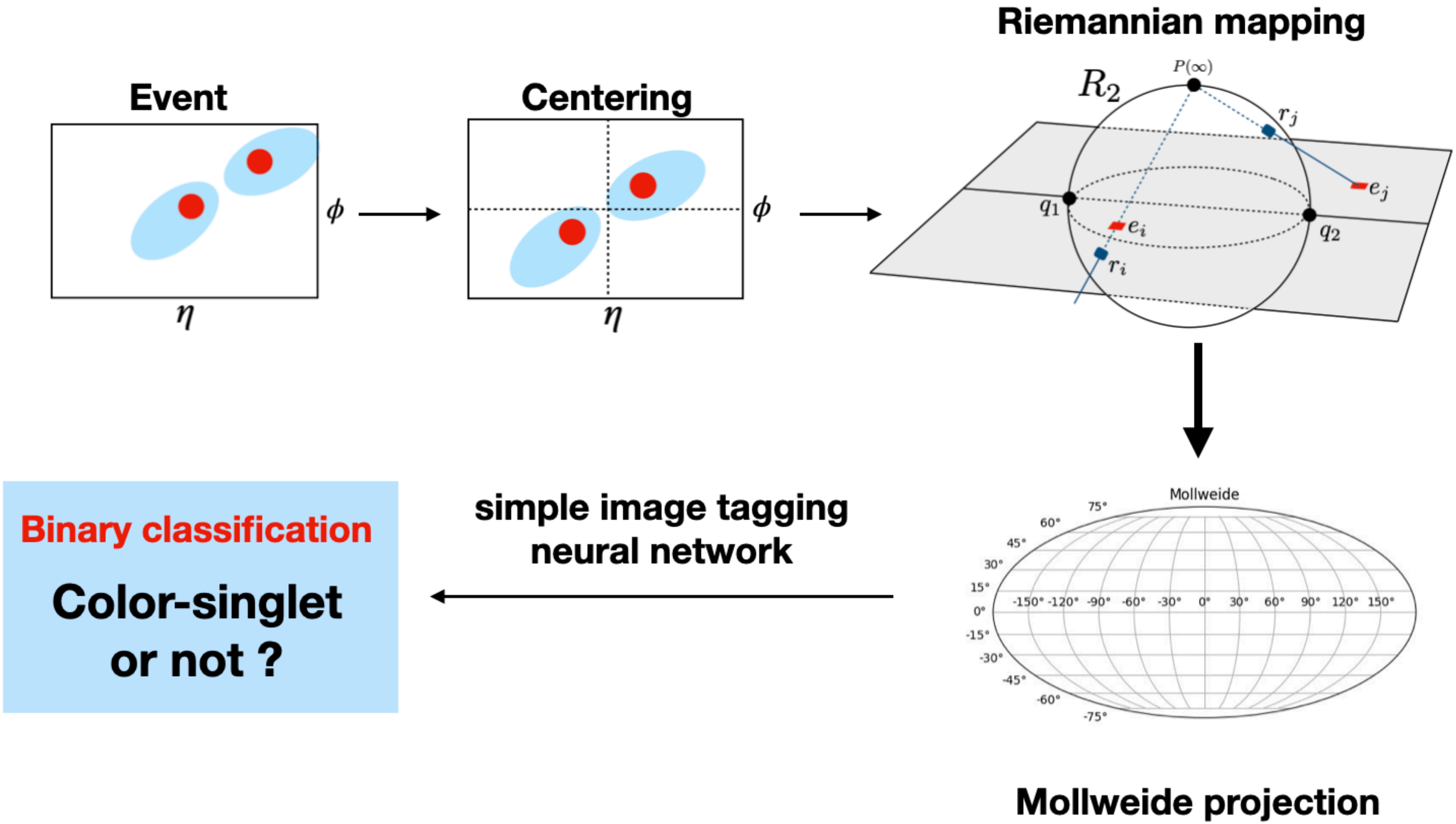
- Soft radiations which are
  inside of a circle $\rightarrow$ Southern hemisphere (H)
  outside of a circle $\rightarrow$ North hemisphere (Color octet status)

- Consider **only angular positions**, **totally** **independent from a radius which is proportional to** $P_T(jj)$.

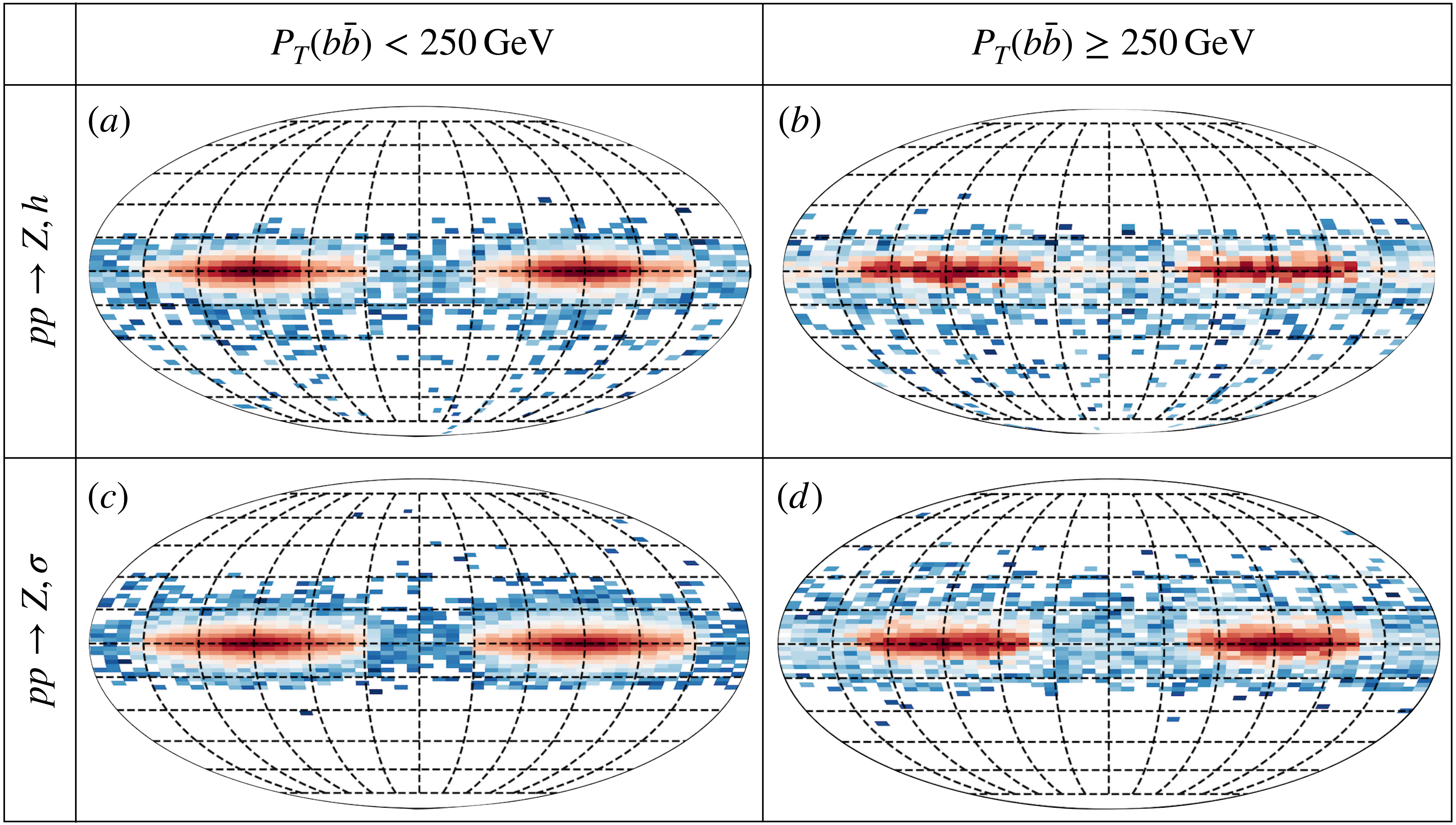
- A toy model of color octet "scalar" particle with $m_\sigma = m_h$ to focus on checking the performance on "QCD". (Also QCD backgrounds, $pp \to (g \to b\bar{b}), Z$ is in this case)



Riemannian preprocessing



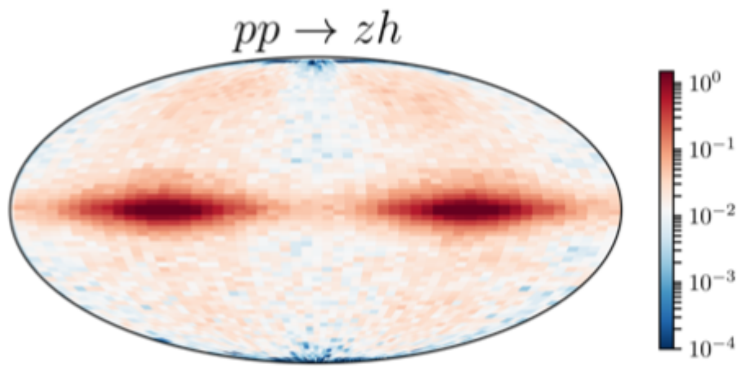
Mollweide projection of Riemannian preprocessing

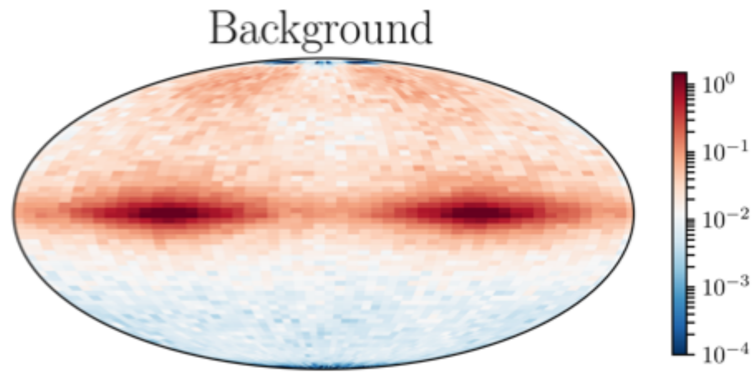


- The distribution of soft patterns does not show a dependency on $P_T(b\bar{b})$

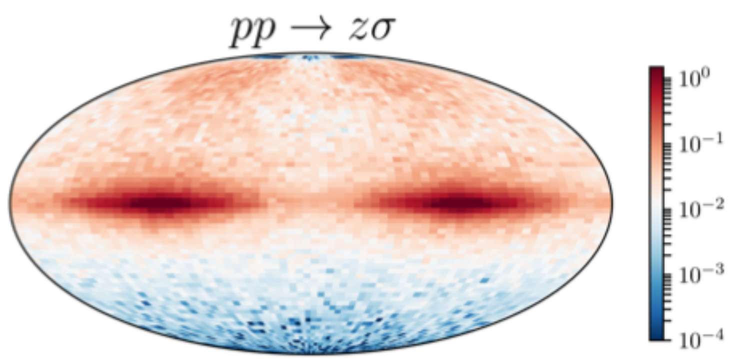
Landscape of Color activity

- Accumulated 5000 events shot



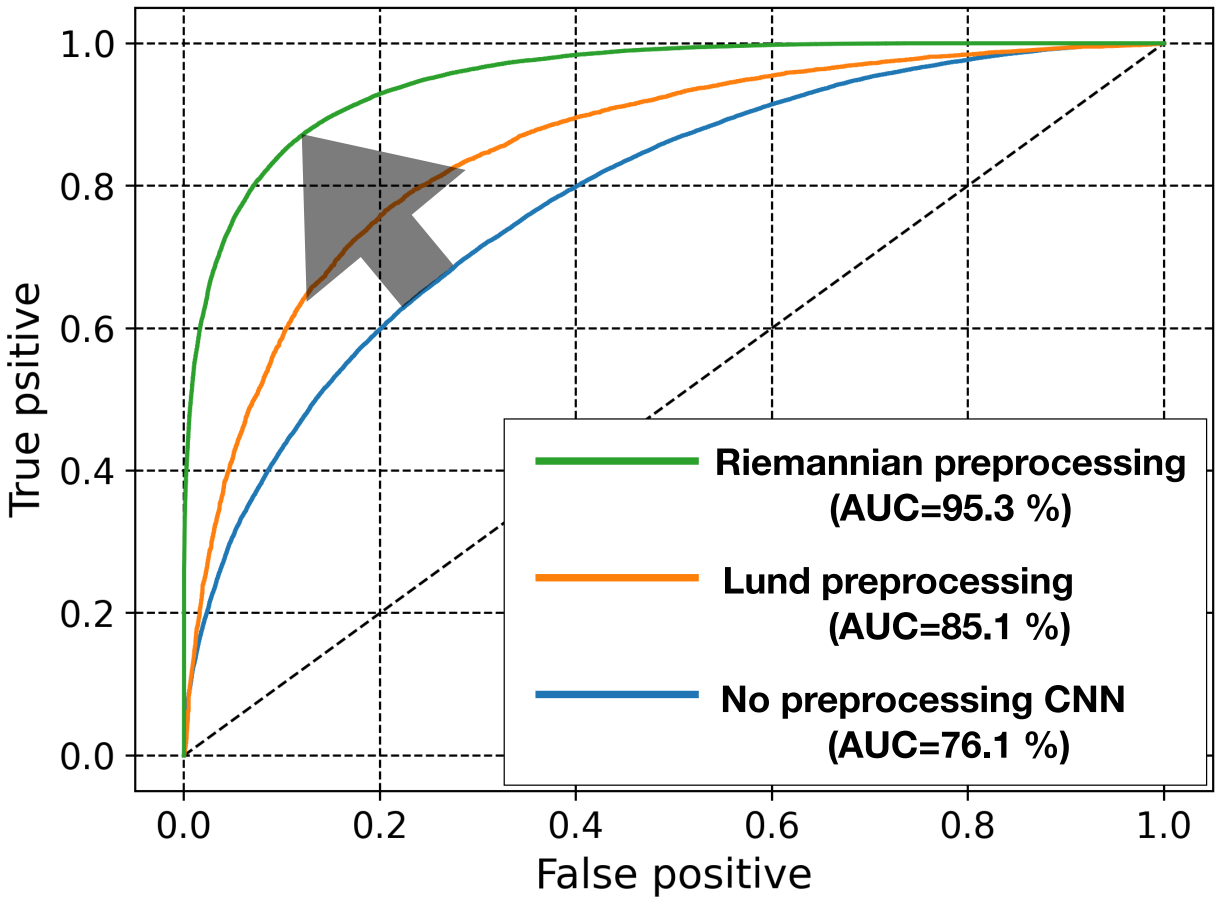


$(V + js,\ t\bar{t},\ t(\bar{t}),\ VV,\ \cdots)$



- Corruptions in North hemisphere are from ISR / MPI QCD activities.

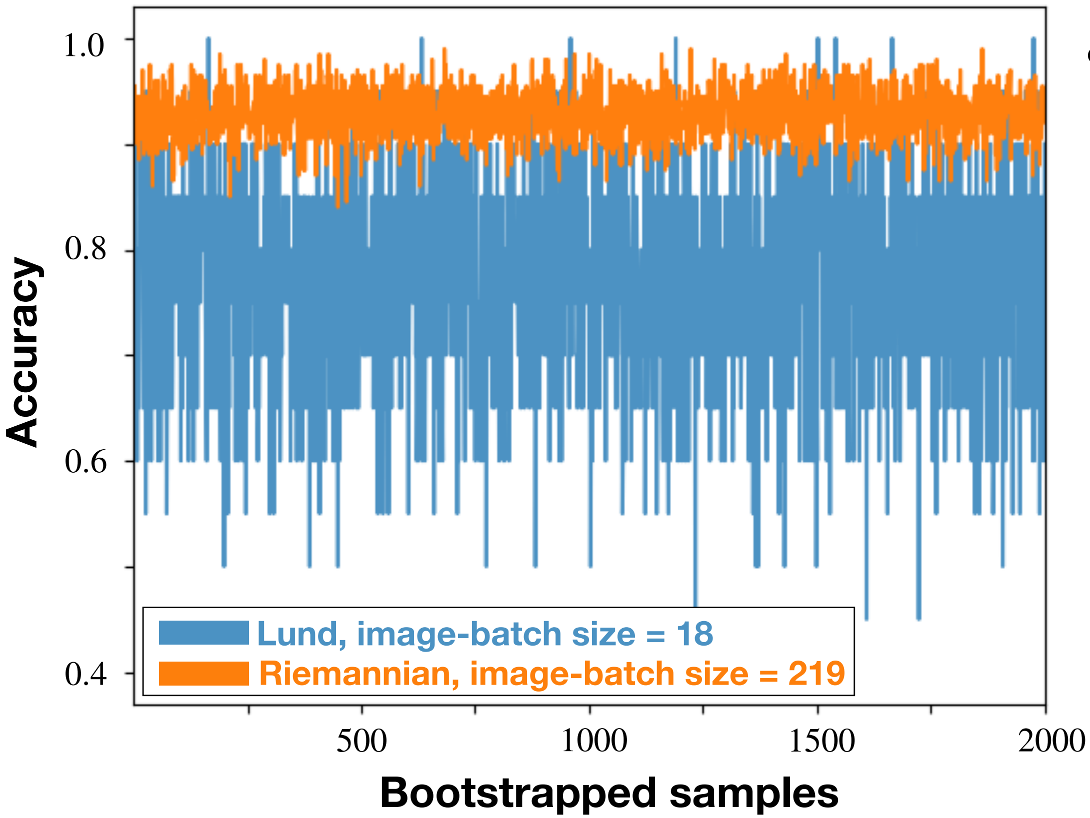
Performance test



- With 100,000 MC data sample each for (1) whole p_T range and for (2) boosted p_T "Riemann" preprocessing has a outperformance.

- Lund preprocessing ("double-logarithmic plane") is from [arXiv:2105.03989] for a boosted Higgs (Data preprocessing with selected QCD features)

Applying to the LHC test



- Based on the ATLAS work (Measurement of WH/ZH in $H \to b\bar{b}$, 13TeV with $139\mathrm{fb}^{-1}$: arXiv:2007.02873)

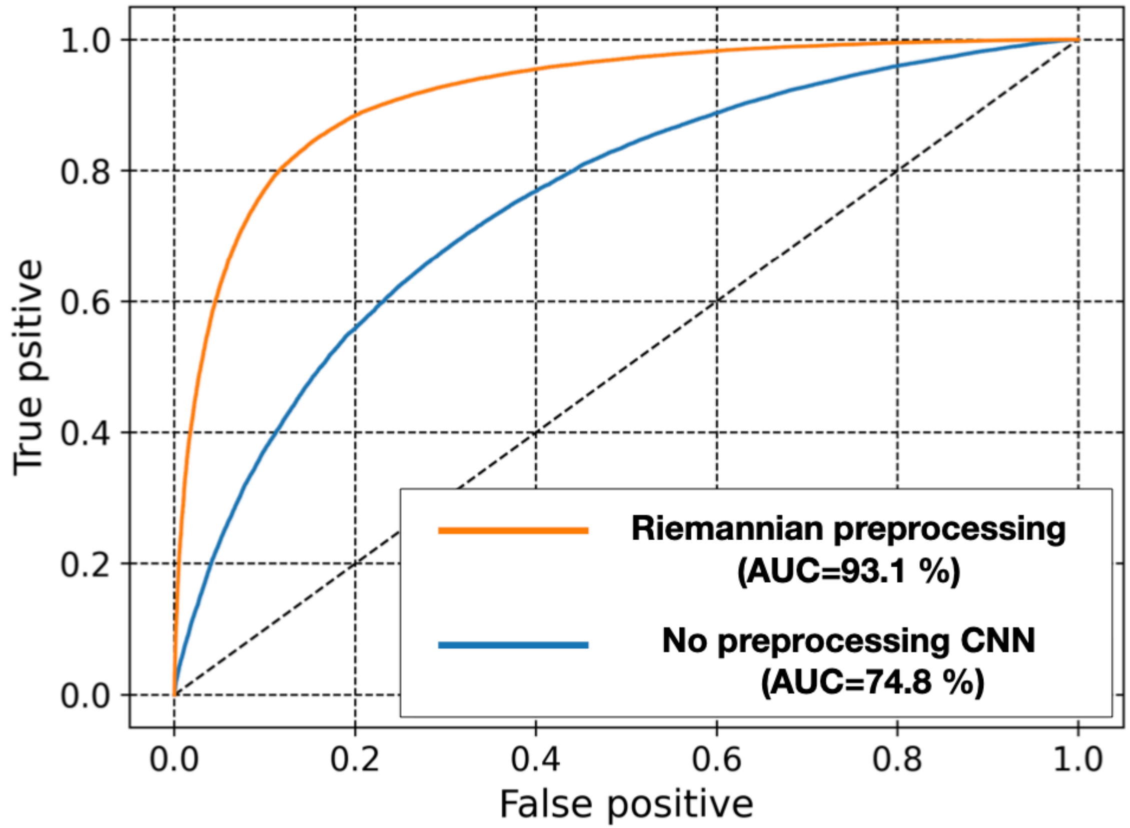
  - **Number of Higgs samples after selection cuts : 219**

  - **Number of Higgs samples in the boosted region ($p_T > 250\mathrm{GeV}$) : 18**

- With well-trained Neural Network, **analysis only with High P_T region** will **suffer from "statistical fluctuation"** in the real battle of the LHC.

- Thus, **the method with wide range of $p_T(h)$ would be better**

Reducing QCD backgrounds

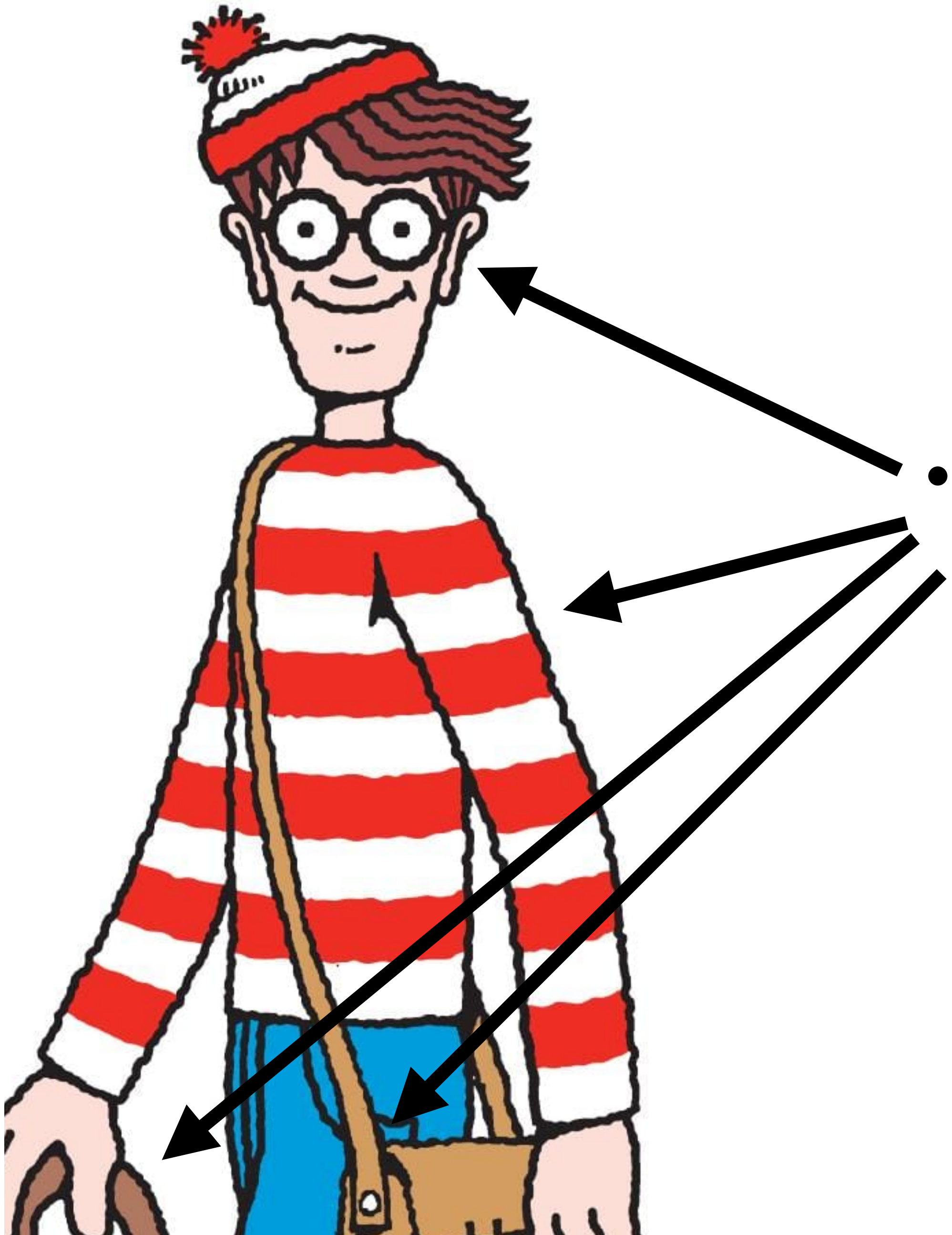


- Applying "color singlet ML tagger", we can achieve **"factor 2"** (CNN: 25%) enhancement compared to conventional cut-and-counting based only on kinematic features.

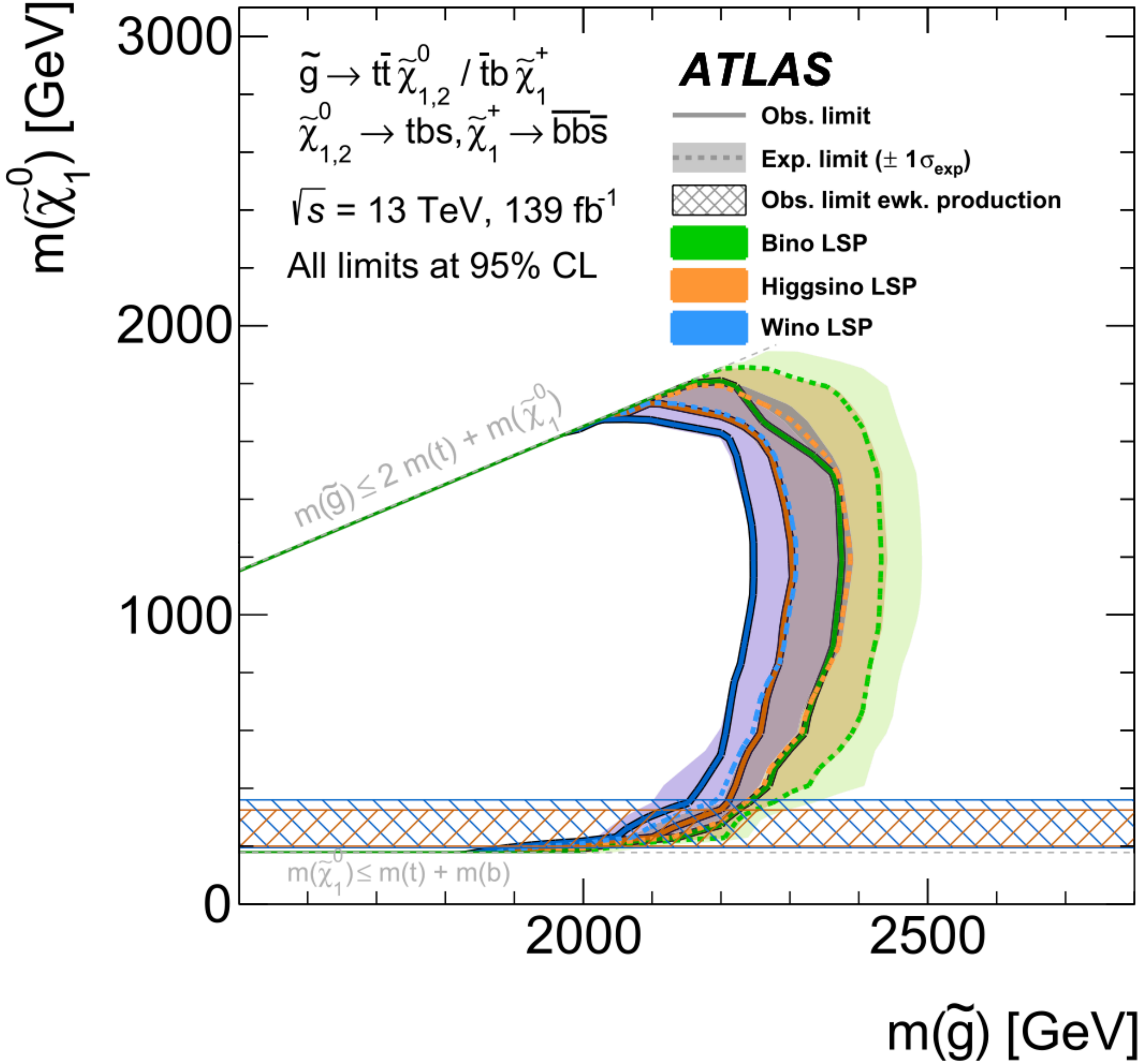
Untold story

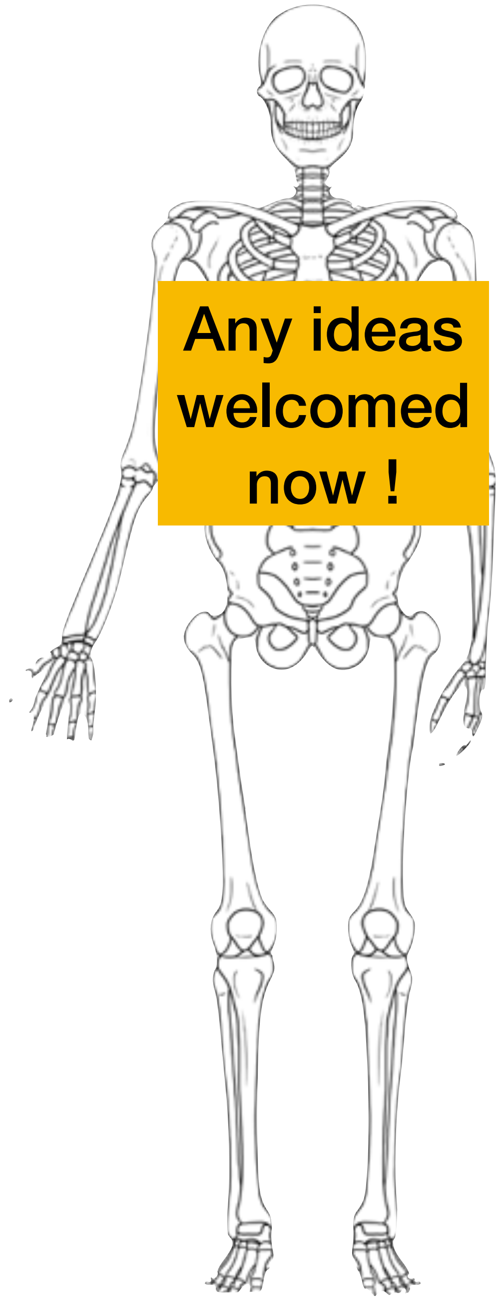
Q) Do we really expect "Wally" looks like this?

who said???



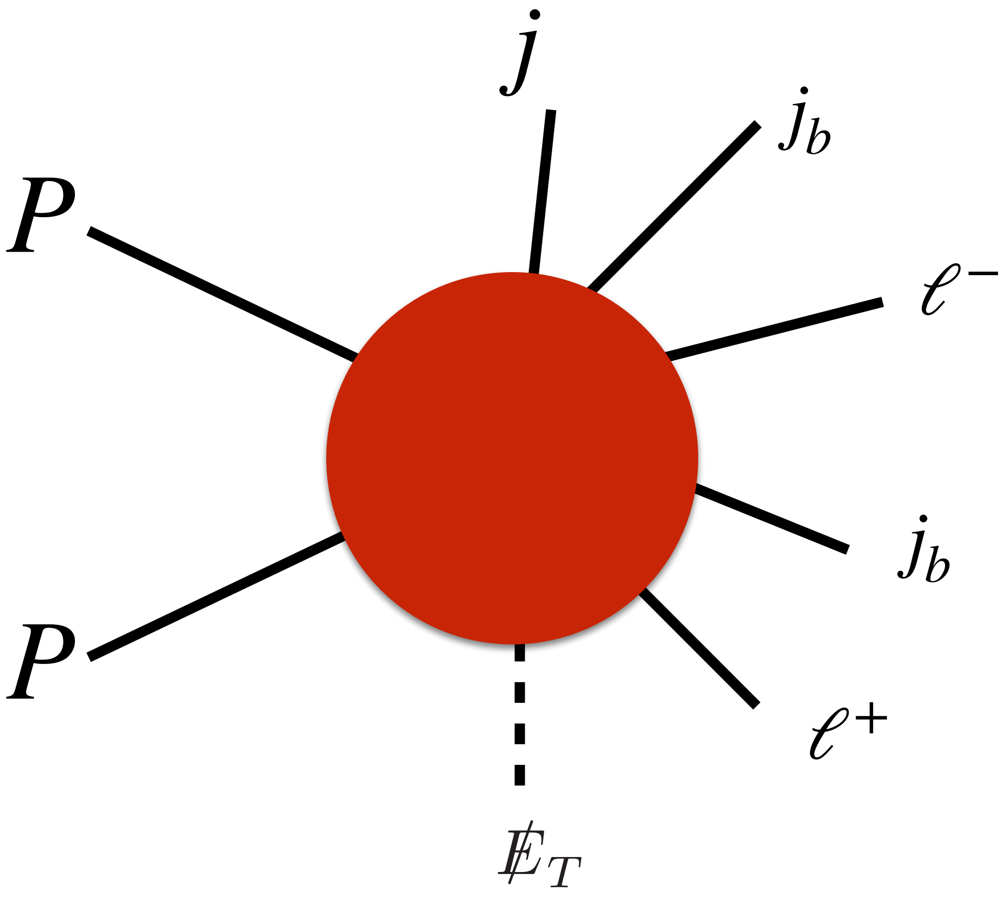
- Focusing on features of "**signal**"??????????????

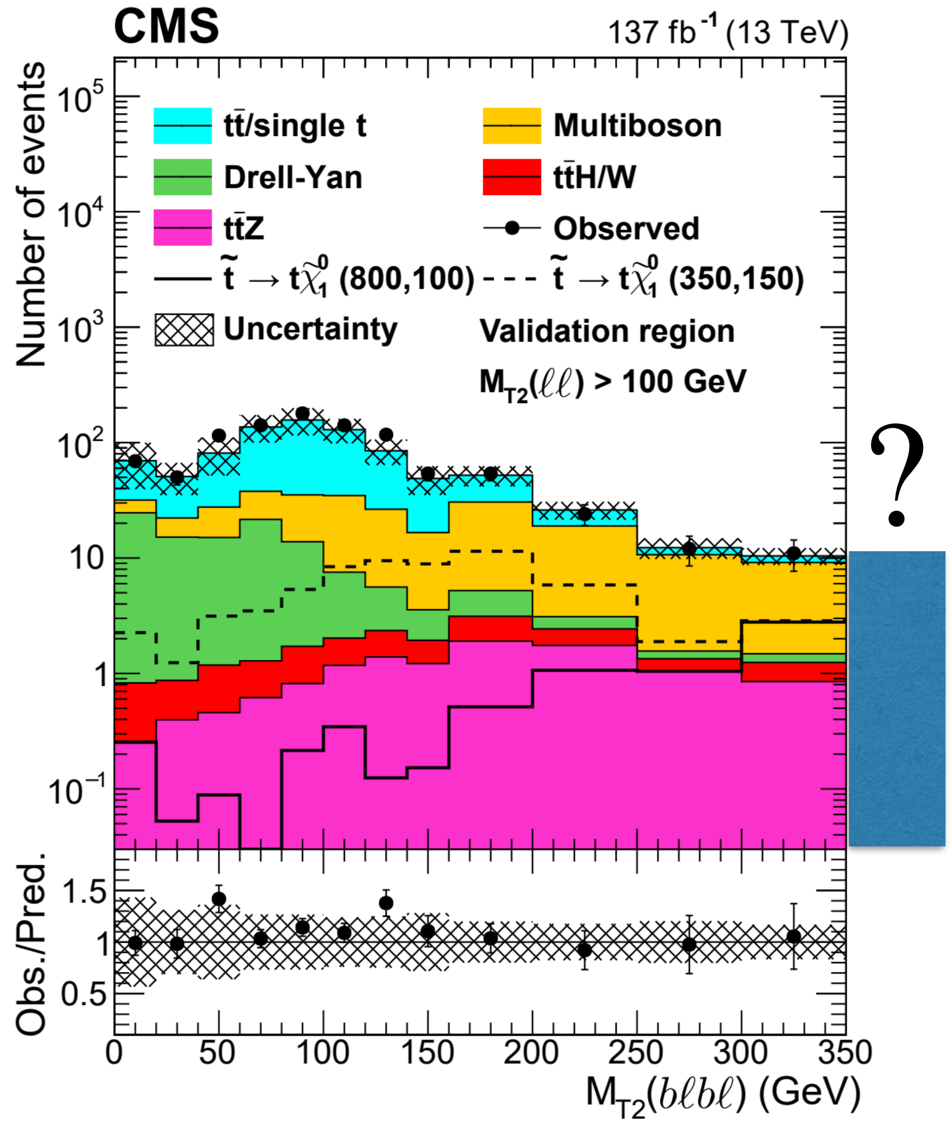




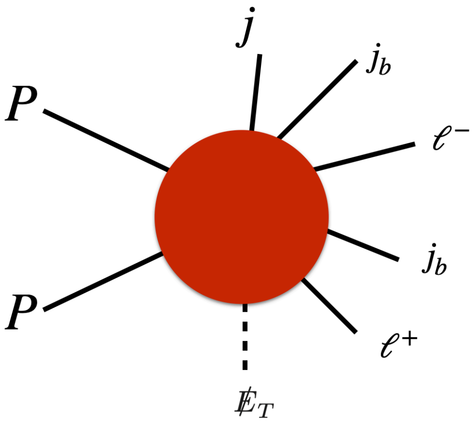
Any ideas welcomed now !

Example: anomaly



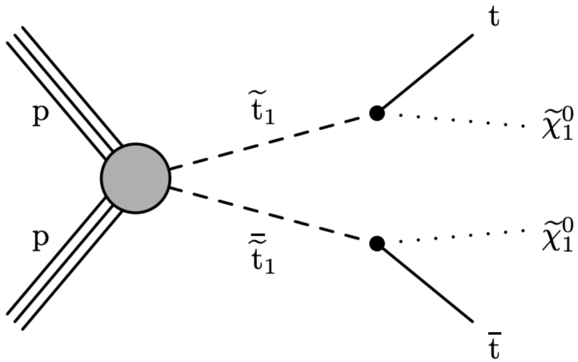


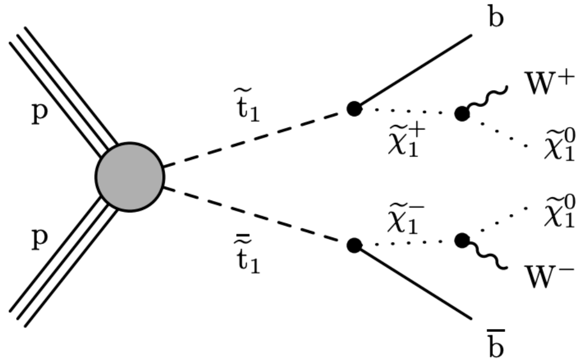
a simple kinematic variable

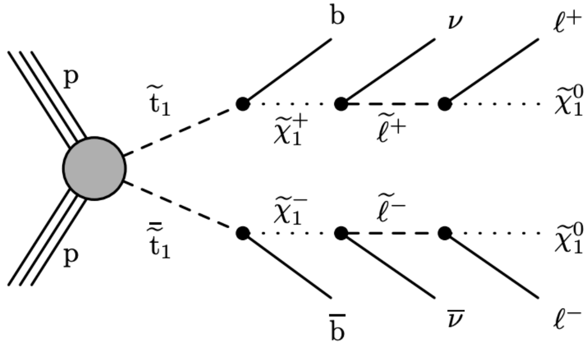


$\in$

$\{$

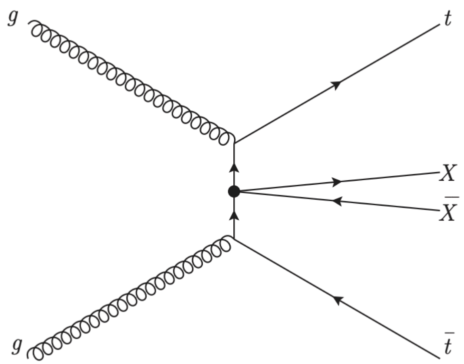






Sorry for my poor imagination, still SUSY....

• • •



$\}$

diagram from Lian-Tao Wang et.al. arxiv:1303.6638

Identifying Event-topology

- Once we **"get"** signals over **"expected" backgrounds,** (with, for example anomaly detection methods)

- If we can "**identify**" an event topology behind signal events

  - We can further **increases signal efficiencies** with various **"supervised" Machine Learning** methods

  - We can check the candidates for a model for this signatures, Very important, at least to give a novel prize :)

- Importantly, there are not so many studies on this ! (as far as I know)

If yo are interested, please check our other works (Minho Kim, Jae-hyeon Park, Pyungwon Ko, MP, arXiv:2111.07806)

Conclusion

- We are interested in maximizing the discovery chance with Machine Learning

  - We want to squeeze outputs from the (future) colliders

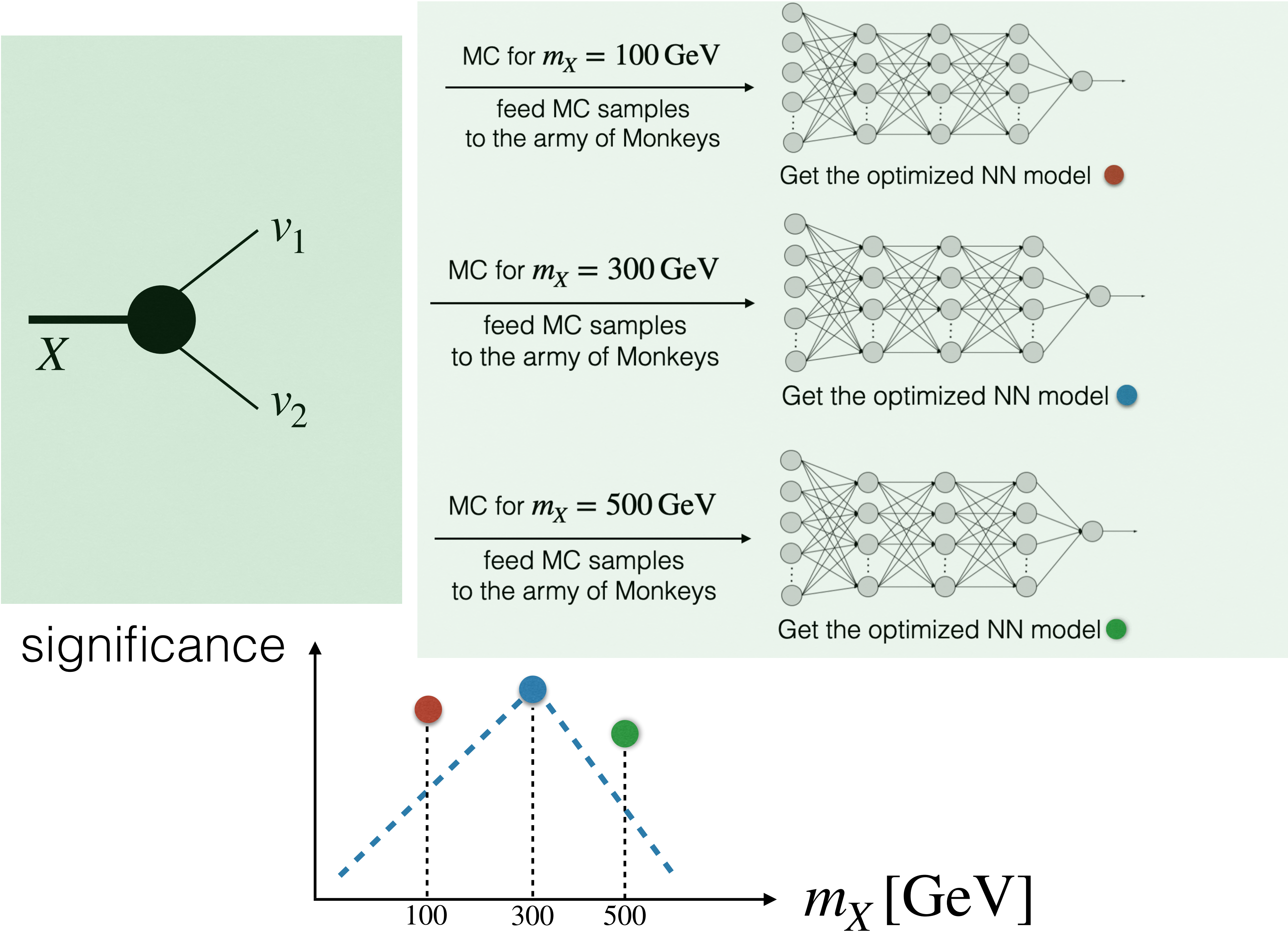
- Utilizing various information, including kinematics, **QCD information would be very helpful. ML can do this job nicely.**

- **Still we can design nicer networks or preprocessing method (Riemannian) based on our domain knowledge (physics)**

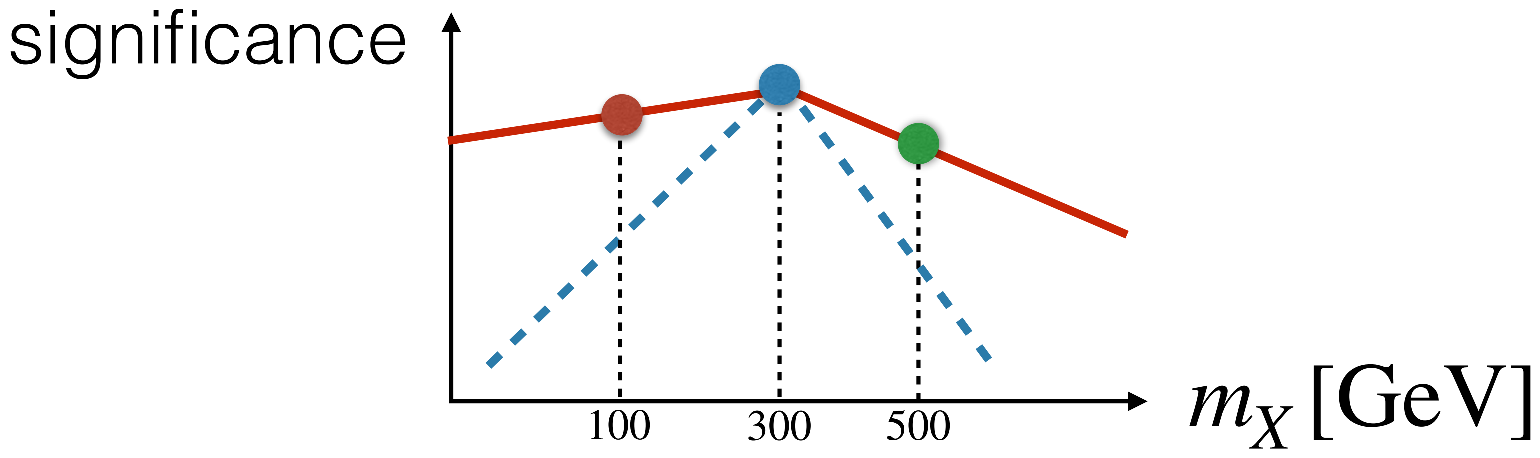
  - More efficient to minimize the training data size, quick learning convergence, etc...

Details on Kinematic ML

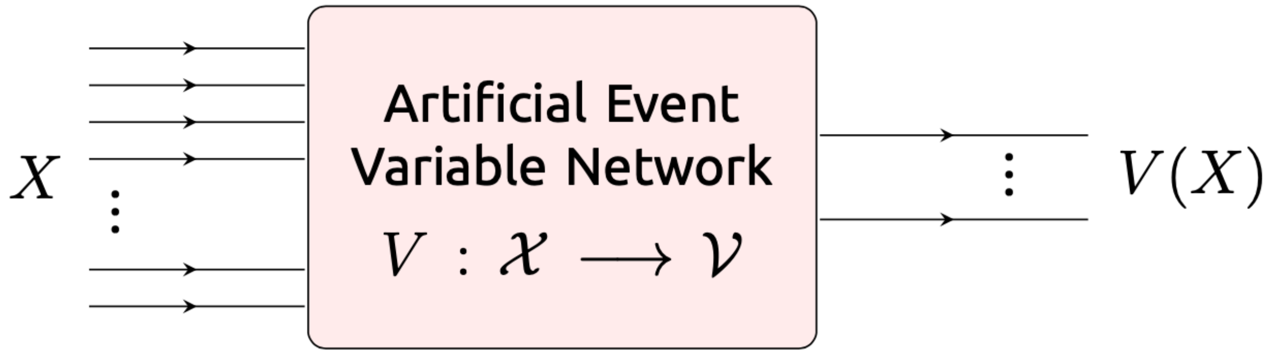
- Conventional supervised ML,



- **Our NN method provide**

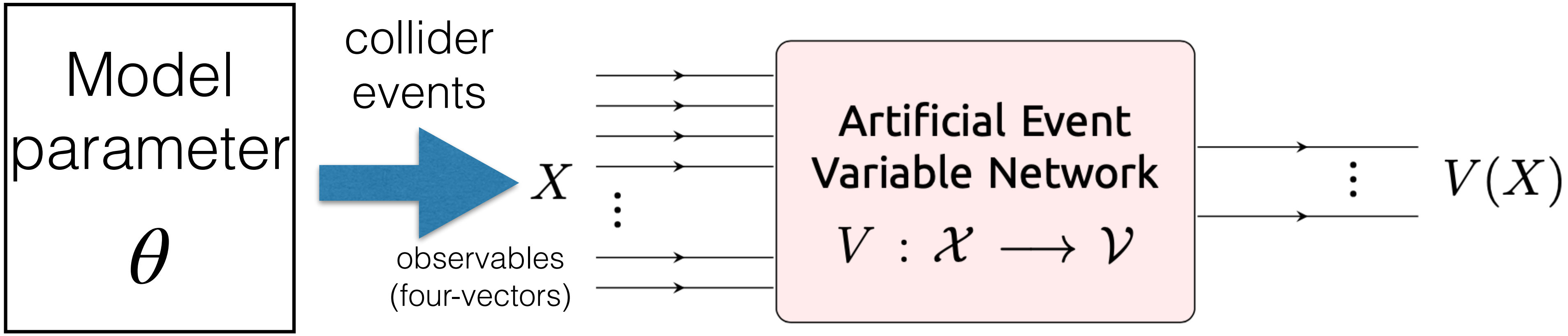


by synthesizing event variable with ML



- We want to make V **to focus on the kinematics.**

- We train the network so that V carries (focus) information on the phase space, independently on the underlying unknown parameter θ



dim(X) —— Dimensional reduction ——→ dim(V)

- Some information would be lost due to $\dim(X) > \dim(V)$
  - Try to minimize the information loss
  - Efficiently retain the underlying parameters θ

$$I(V;\Theta) = \int dv \int d\theta \; p_{V,\Theta}(v,\theta) \; \ln\left[\frac{p_{V,\Theta}(v,\theta)}{p_V(v)\, p_\Theta(\theta)}\right]$$

- This is KL divergence between $p_{(V,\theta)}$ and $p_V \otimes p_\theta$

Train V so that $p_{(V,\theta)}$ and $p_V \otimes p_\theta$ are highly distinguishable

How to train V ?

- Use two classes
  1. $p_{X|\theta}$: events X **from** θ
  2. $p_X \otimes p_\theta$: events X **not from** θ (totally uncorrelated)