# Exploring the Synergy of Kinematics and Dynamics for Collider Physics


## Myeonghun Park

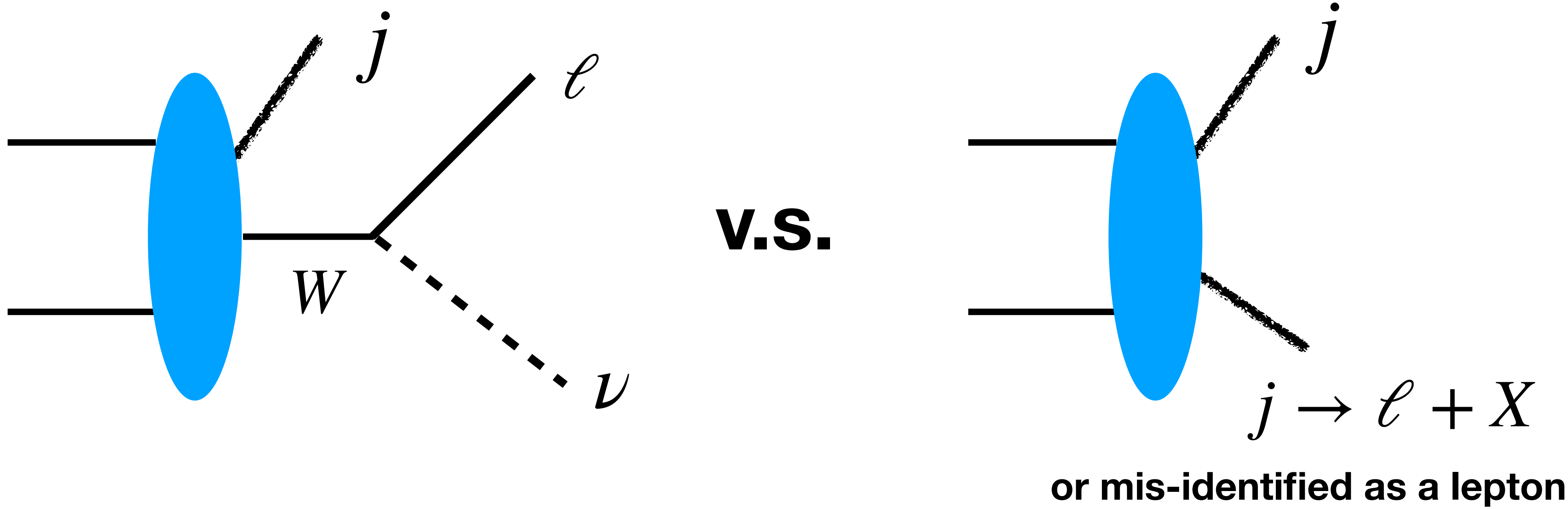
(Seoultech)


- Based on arXiv:2311.16674 with Kayoung Ban, Kyoungchul Kong and Seong Chan Park

---


**High1 Workshop on Particle, String and Cosmology**
**2024. 01. 21 - 01. 27**

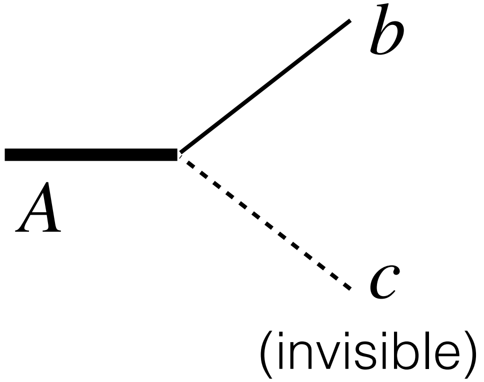
# Search for a "signal" at Colliders




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

# Extracting phase-space "features" of a signal

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



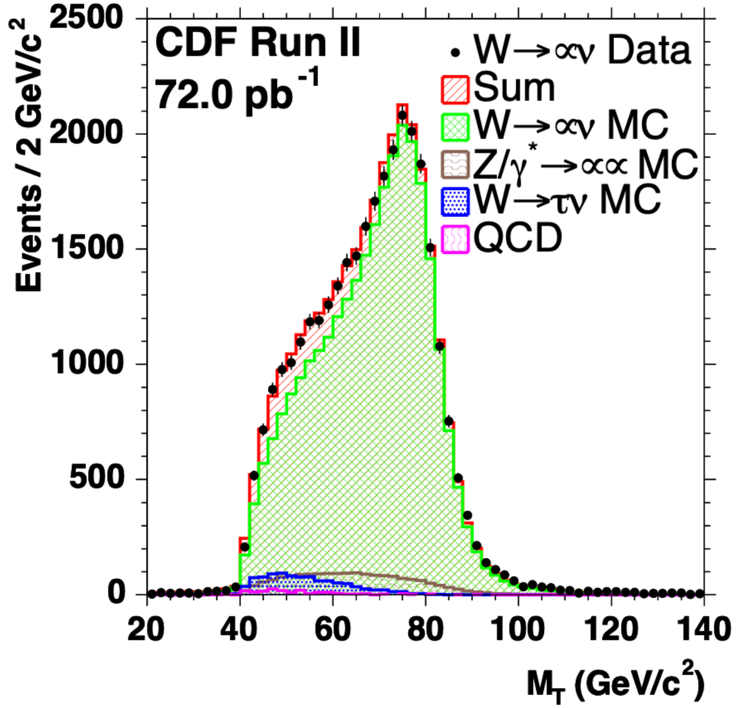

$$\theta = \{m_A\} \longrightarrow X = \{p_b^{\mu}, p_c^{T}\}$$

**3: 3-Momentum from visible particle**
**2: Transverse Momentum from imbalanced situation**

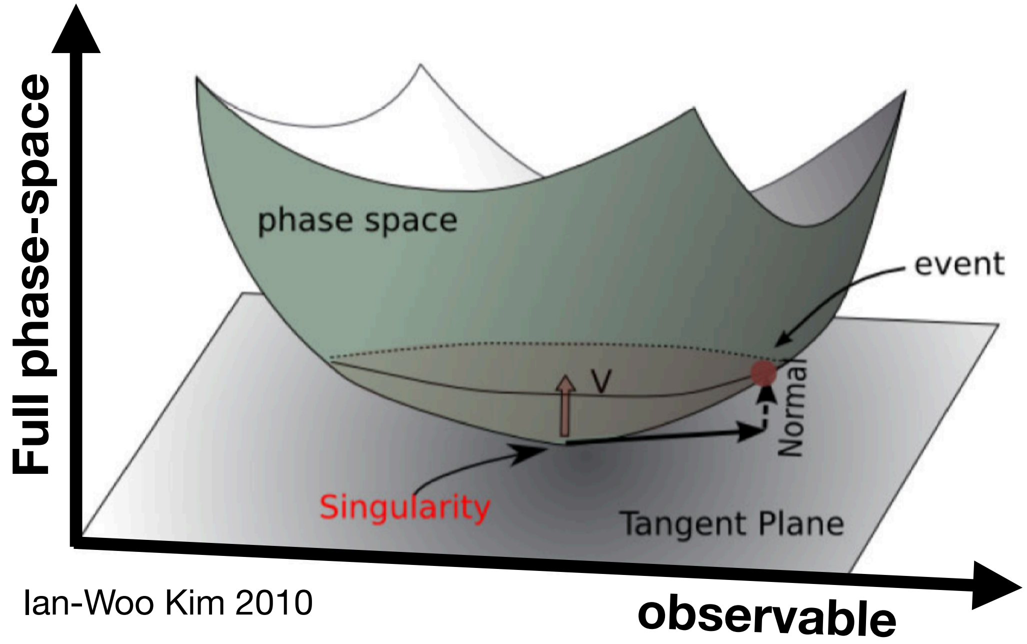
$$\dim(X) = 3 + 2 \rightarrow \dim(V) = 1$$

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

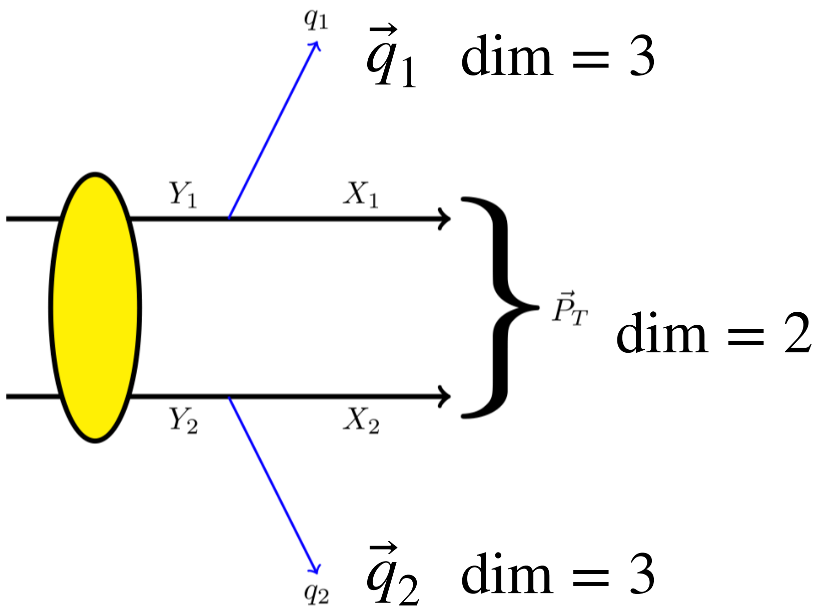
  (highly singular behavior due to its Jacobian peak)



- Constructing an "observable" from a multi-dimensional phase space is **non-trivial.**




Ian-Woo Kim 2010




$$\dim(X) = 8 = 2 \times 3 + 2$$
$$\rightarrow \dim(\text{observable}) = 1$$

- Kink-Structure in an observable: Won Sang Cho, Kiwoon Choi, Yeong Gyun Kim, Chan Beom Park (PRL 2008)
- Generic algorithm to find a Singularity observable: Ian-Woo Kim (PRL 2010)
- The LHC-robust observable: Konstantin Matchev, **MP** (PRL 2011)
- Detalied investigation on singularity observables: Chan Beom Park (JHEP 2021)
- For a recent review, "Kinematic variables and feature engineering for particle phenomenology" in Rev. Mod. Phys. 95 (2023) by Doojin Kim, **MP** et.al.

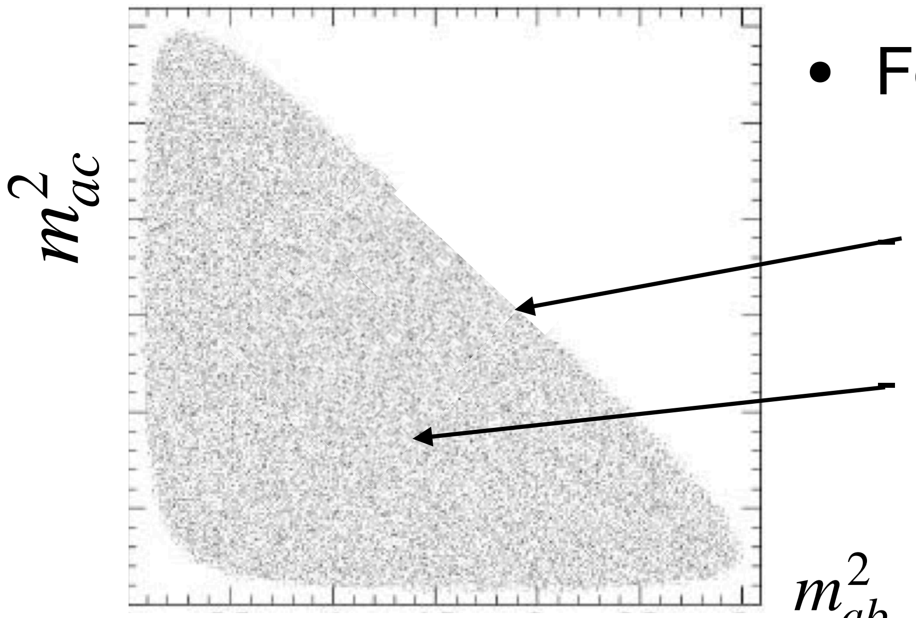
# Kinematics: Global information

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

**Classic example: Dalitz plot**



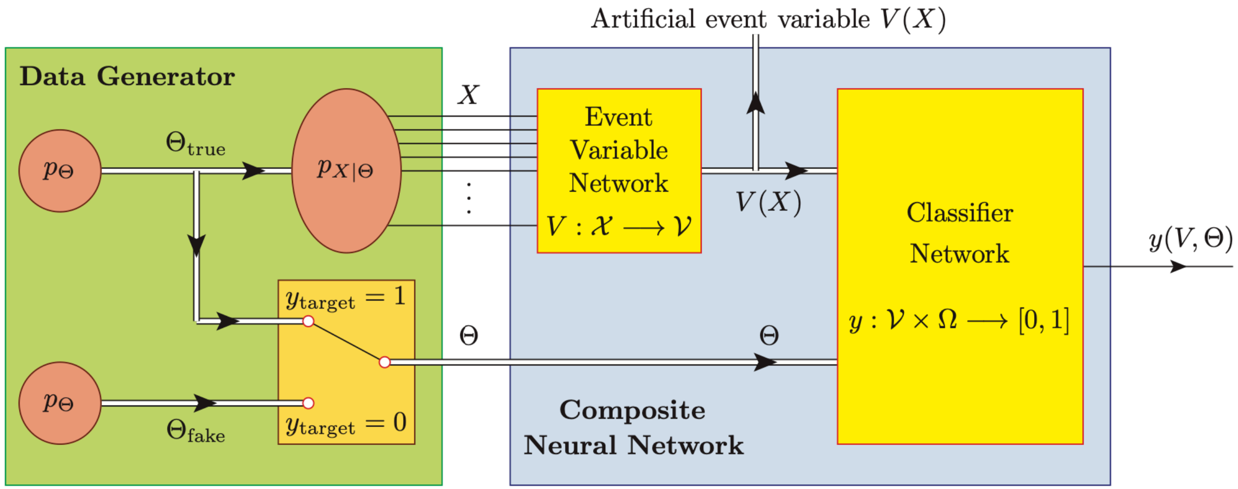

- For example, $X \to a, b, c$

  - Boundary: Structure of $1 \to 3$ phase space

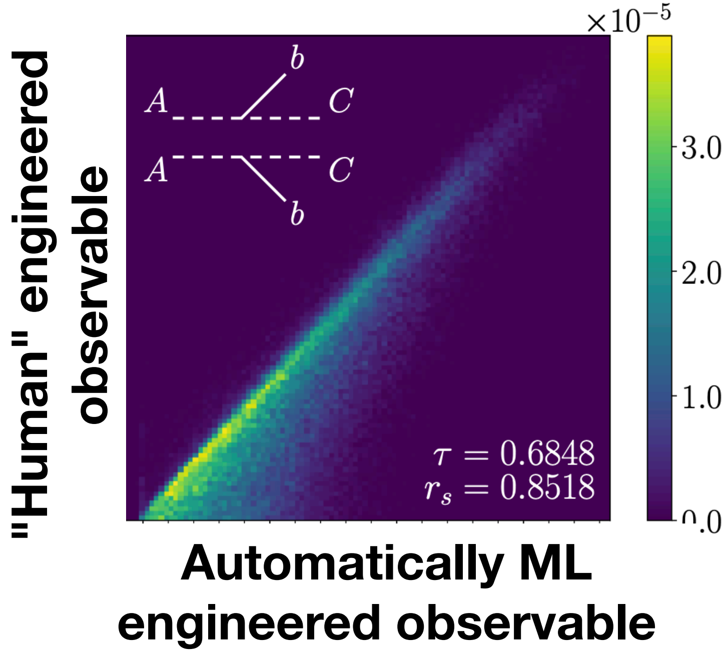
  - Density:
    information on intermediate particles, spin..

# Extracting "features" utilizing Machine Learning

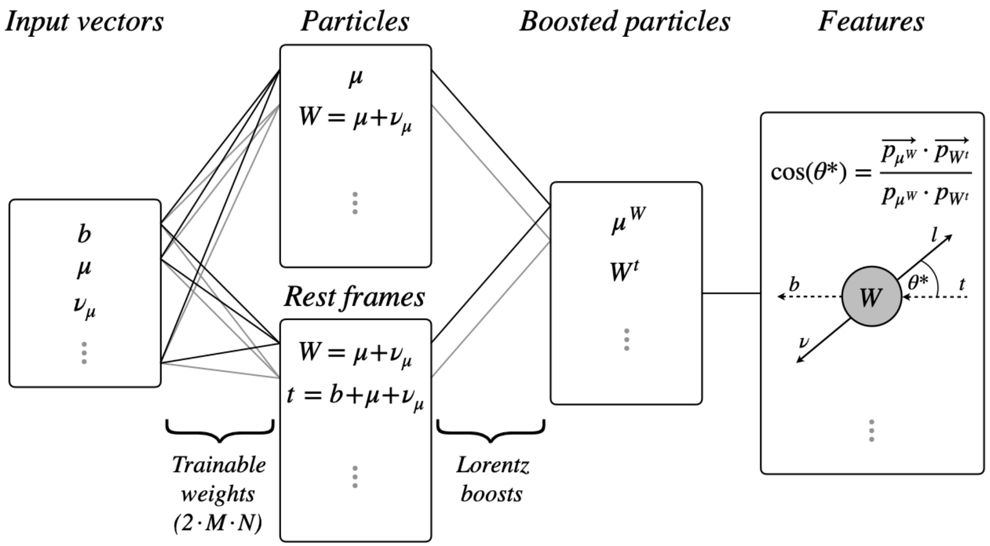
## A Neural-Network can design an event-variable
## (by enforcing information-bottleneck to NN)

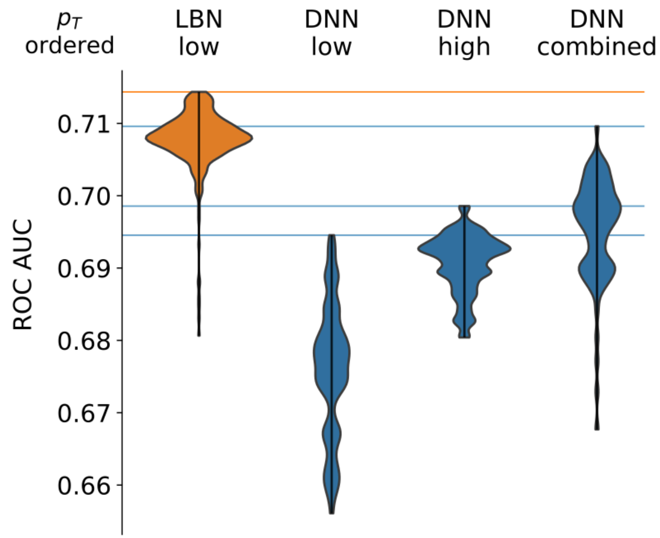



oojin Kim, KC Kong, Konstantin Matchev, Prasanth, **MP**. (PRD 2023)




## A NN with enforcing a relativistic kinematics



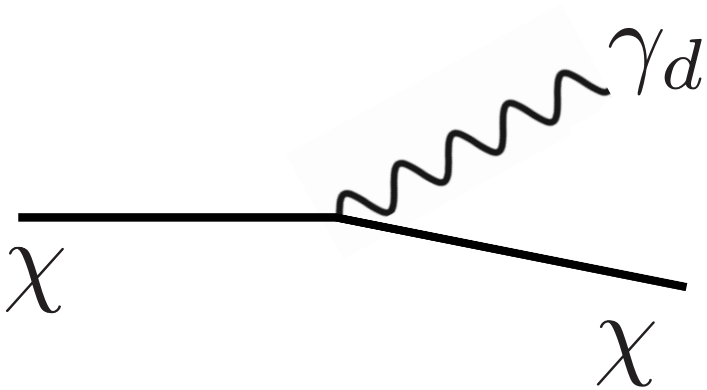

M. Erdmann, E. Geiser, Y. Rath and M. Rieger (2019)

# Orthogonal information to the Kinematics

- **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

: In a **chiral** case, the **longitudinal** component of a dark photon couples to a dark matter




**boosted (accelerated) DM**

**soft energy deposits**

- Minho Kim, Hye-Sung Lee, **MP**, Mengchao Zhang (2018)
- Junmou Chen, Pyungwon Ko, Hsiang-Nan Li, Jinmian Li, Hiroshi Yokoya (2019)

# Case of the Standard Model Gauge group

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

**Jason Gallicchio, Matthew D. Schwartz 2010**

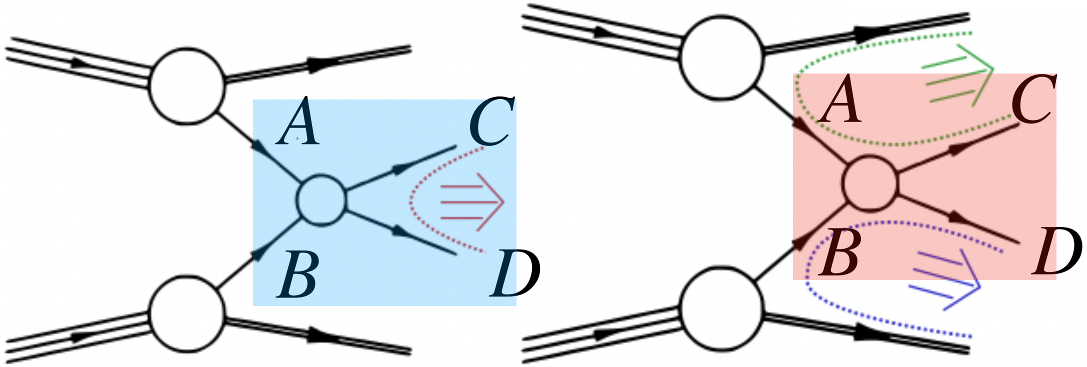
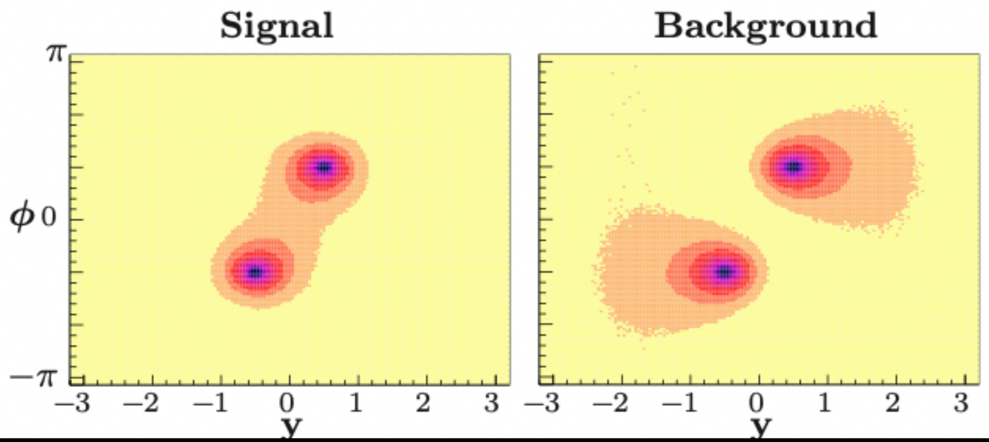



FIG. 1: Possible color connections for signal ($pp \to H \to b\bar{b}$) and for background ($pp \to g \to b\bar{b}$).




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

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

V . S .

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

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

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

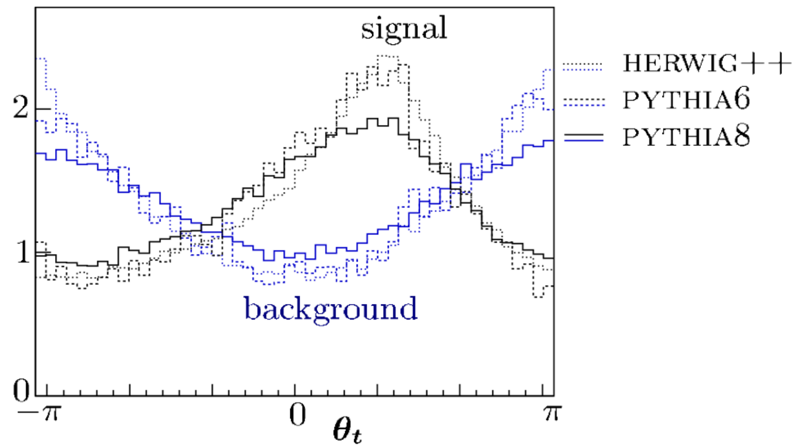
# Utilizing localized information

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

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

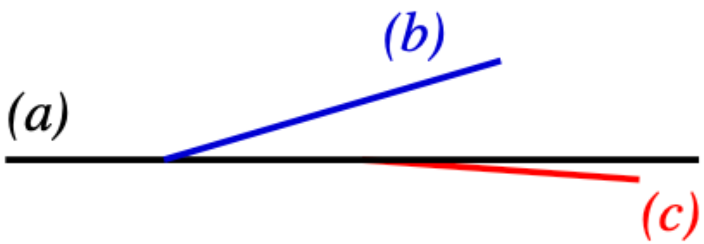
(Jason Gallicchio, Matthew D. Schwartz 2010)

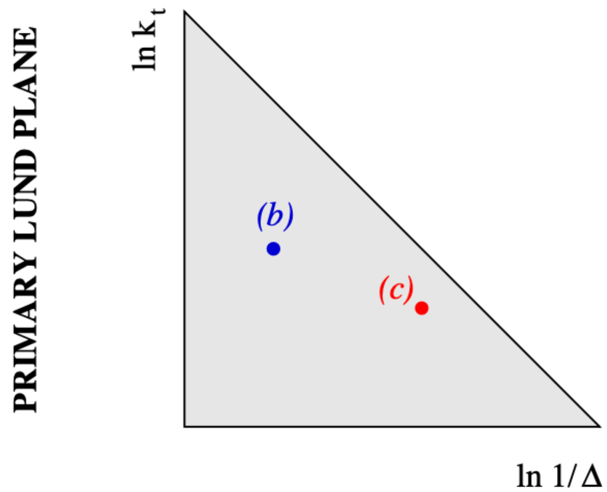
provides an one-dimensional feature




- Or one can get two-dimensional features,

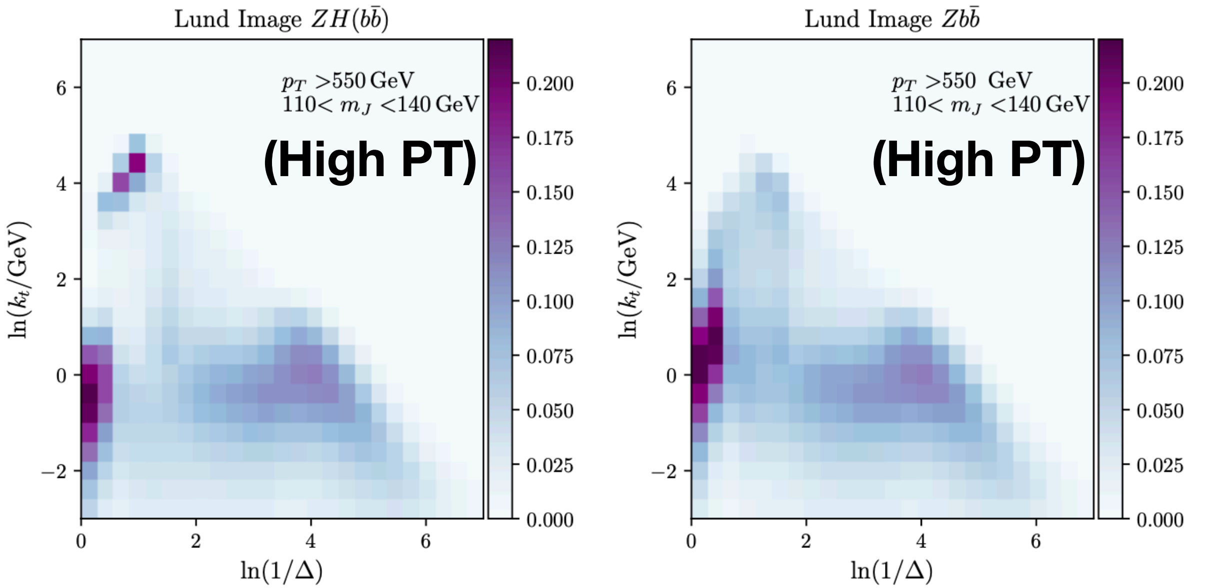
(Frederic A. Dreyer, Gavin P. Salam, Gregory Soyez 2018)



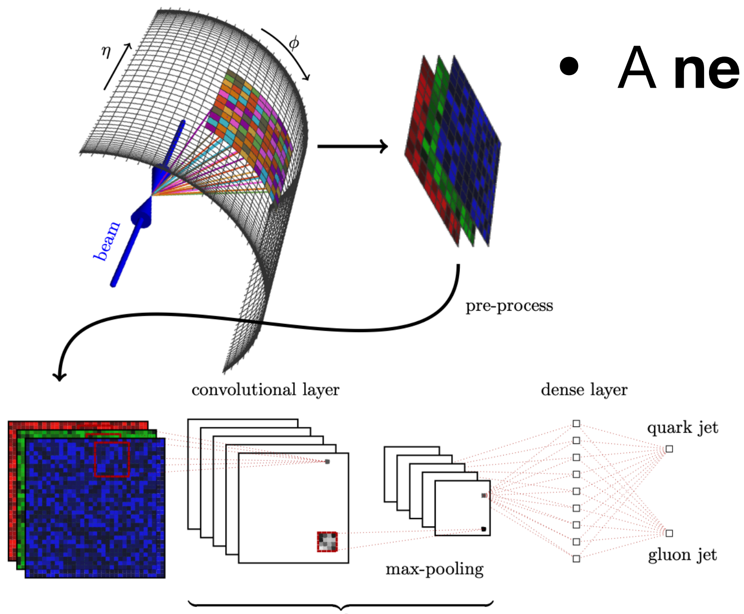



# Fully utilizing localized information

- One needs to understand differences in **"the full information"**



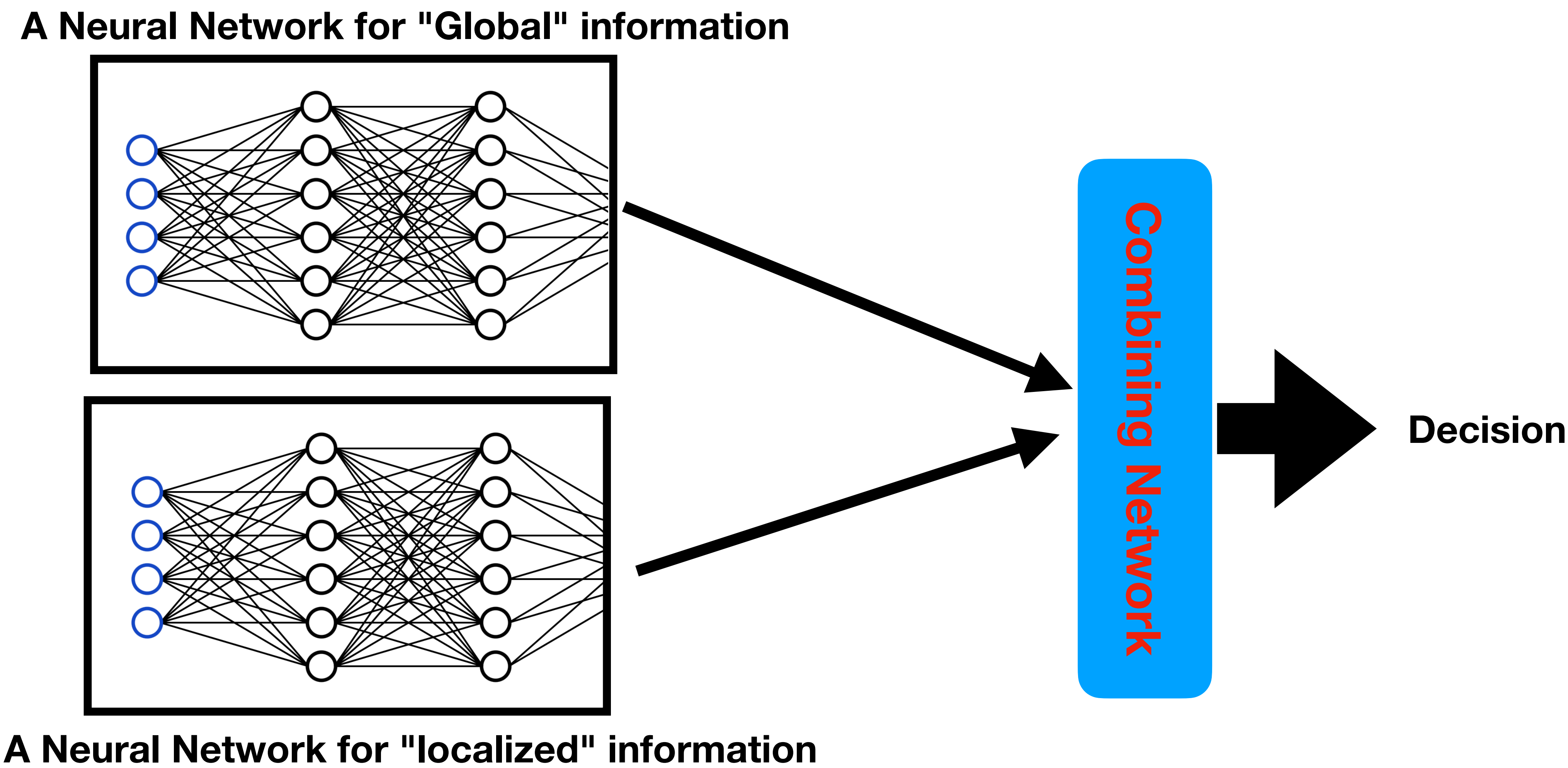

(Charanjit K. Khosa, Simone Marzani, 2021)




M. Schwartz et.al. (JHEP 2017)

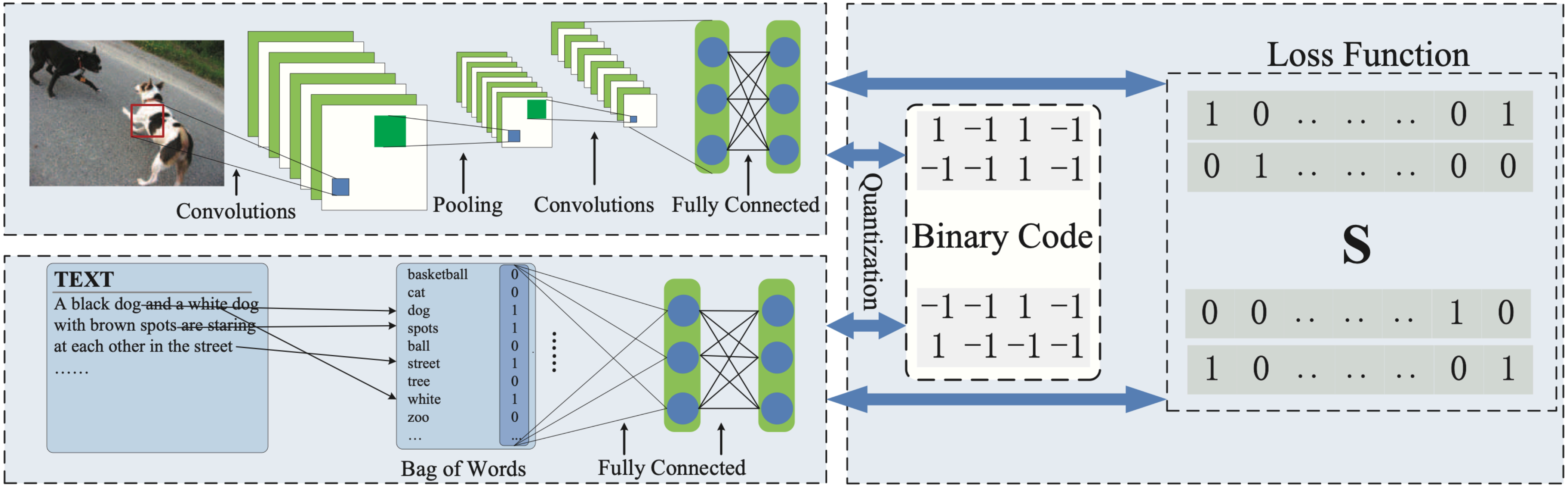
- A **neural network** can tell differences in soft-patterns
  - Leandro G. Almeida, Mihailo Backović, Mathieu Cliche, Seung J. Lee, Maxim Perelstein (JHEP 2015)
  - Won Sang Cho, Hyung Do Kim, Dongsub Lee (PRD 2020)
  - Sung Hak Lim, Mihoko M. Nojiri (PRD 2020), …

# Global and Local information




- Jeong Han Kim, Minho Kim, Kyoungchul Kong, Konstantin T. Matchev, **MP** (JHEP 2019)
- Thomas Flacke, Jeong Han Kim, Manuel Kunkel, Pyungwon Ko, Jun Seung Pi (JHEP 2023)
- Daohan Wang, Jin-Hwan Cho, Jinheung Kim, Soojin Lee,  Prasenjit Sanyal, Jeonghyeon Song (PRD 2024)

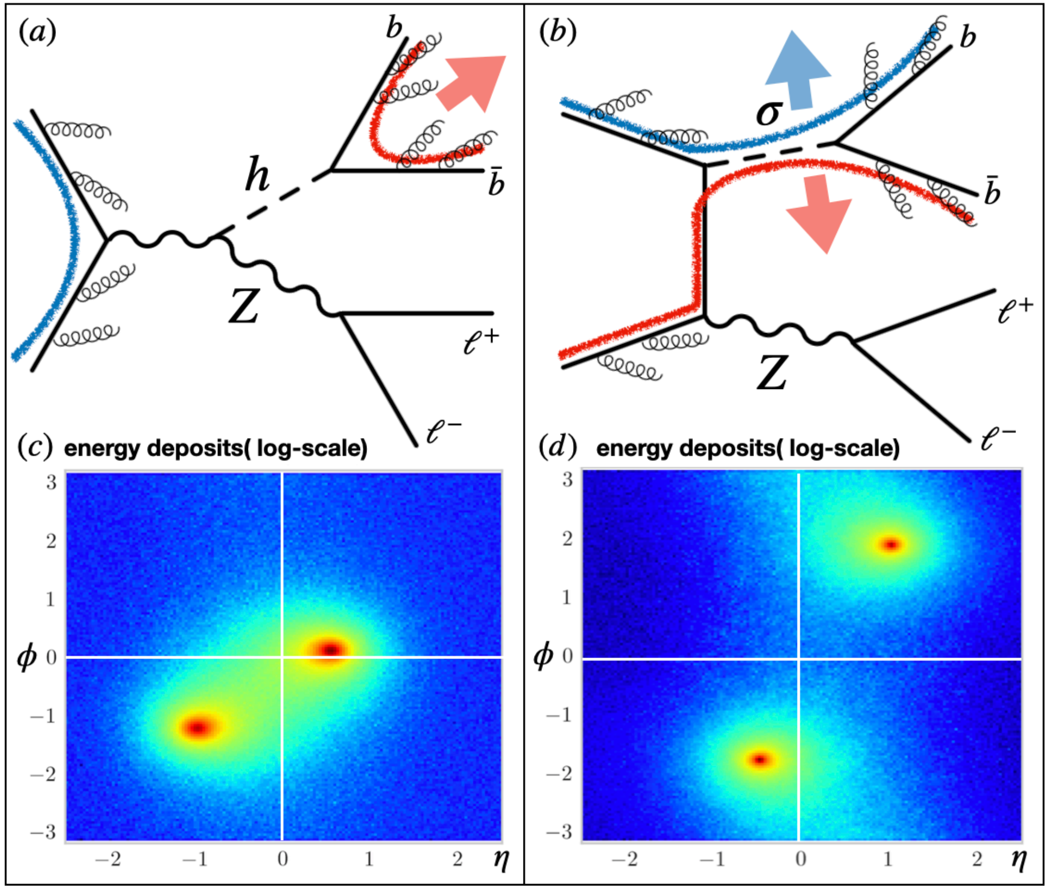
# Multi-modal Network




- In the commercial applications of Machine Learning, various sources of information (for example, different images, text) are utilized to interpret a situation in a consistent way.

# Multi-modal Network in collider physics

- In the collider physics, "kinematics" and "localized pattern (for example QCD) are **orthogonal information**, so that we need to take a **complementary** approach.




- For example,

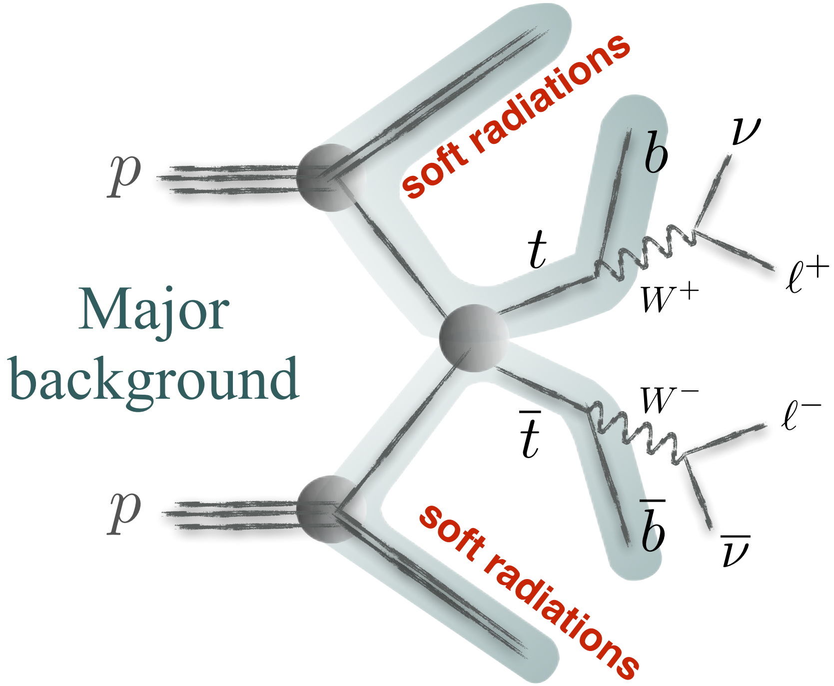
  - similar kinematics (distance between two jets if $m_h \simeq m_\sigma$)

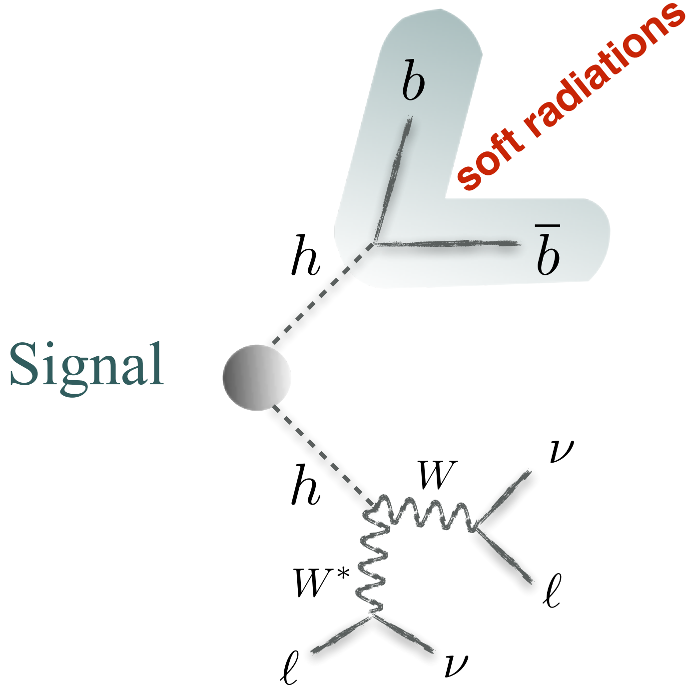
  - totally different soft QCD patterns with color singlet ($h$) v.s. color octet ($\sigma$)

# Combining two different information

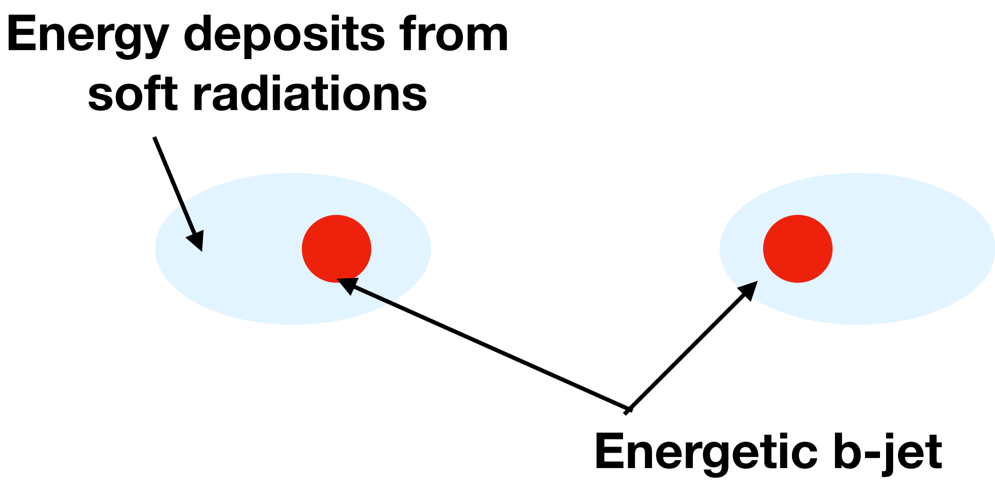
?

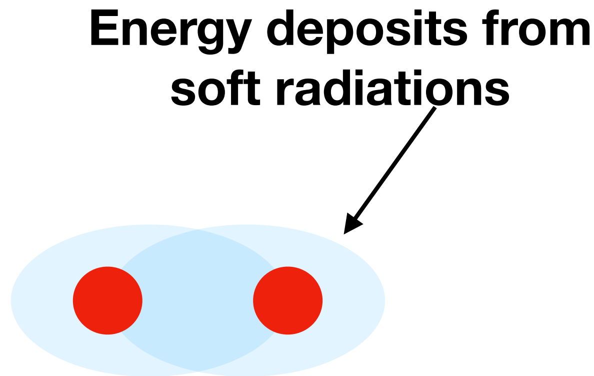
- Are we sure that our combination is the **optimized** one ?



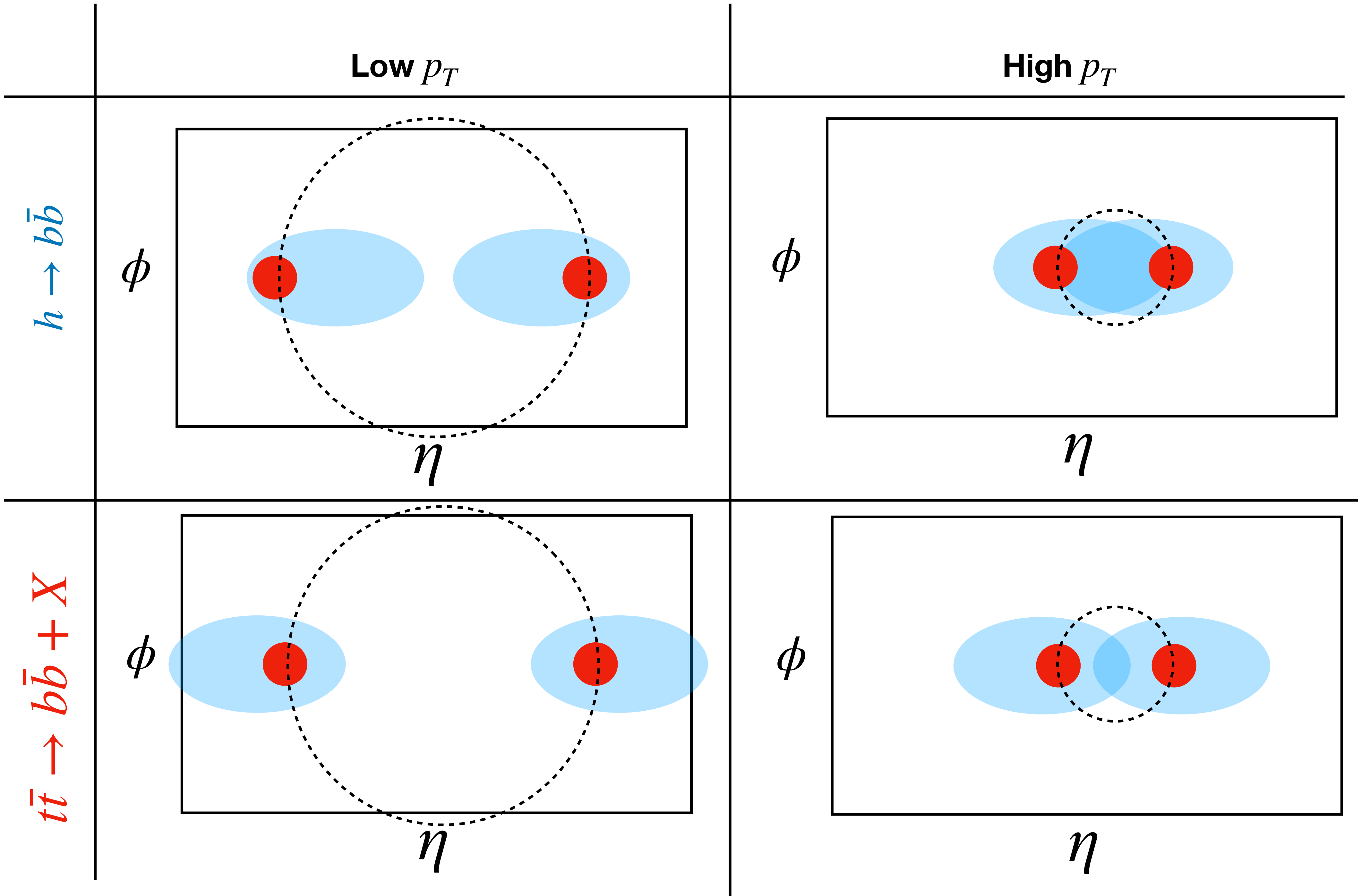

- **Attentions are on hot cores**



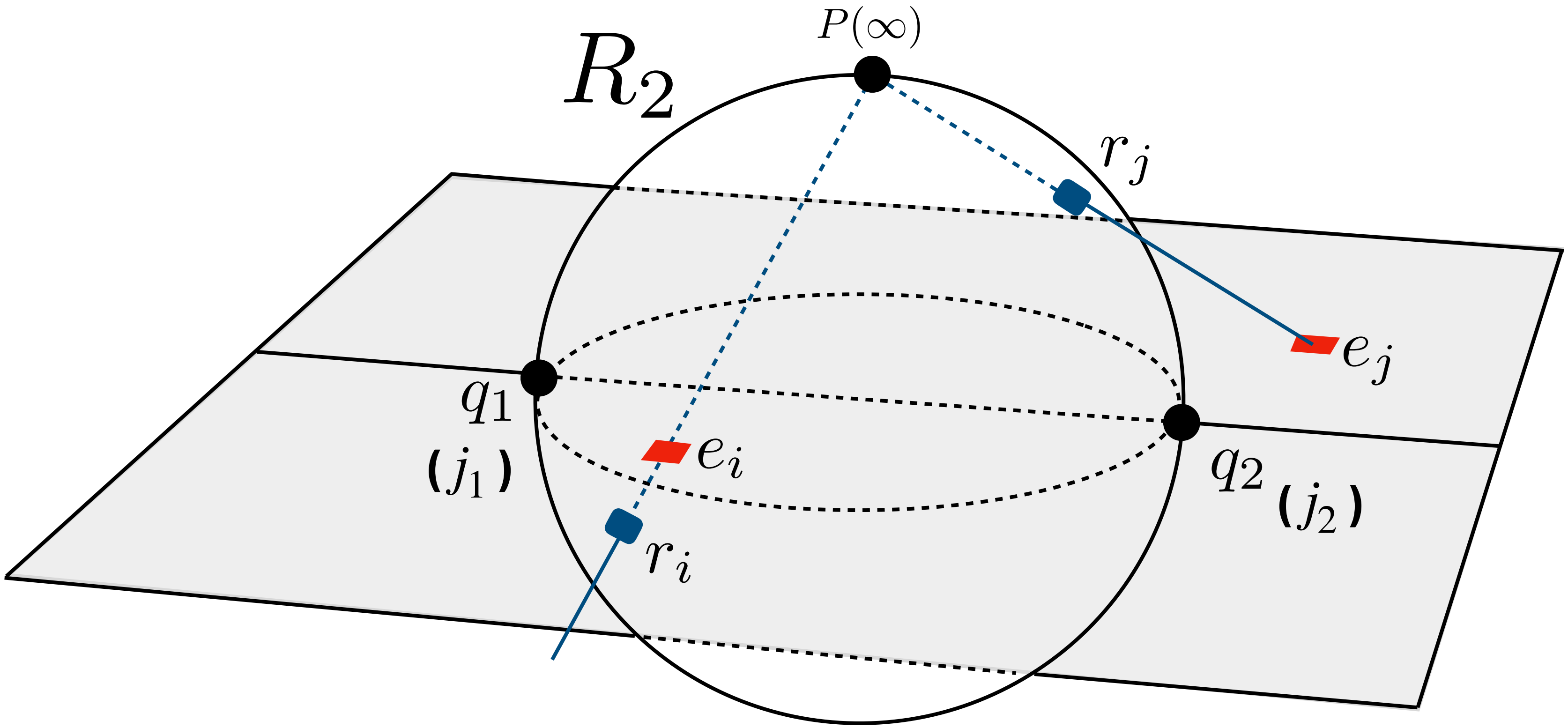




- Possibility to **shadow information from soft radiations** by distinct **differences in kinematics from jets**.

- Decorrelating localized information from kinematics



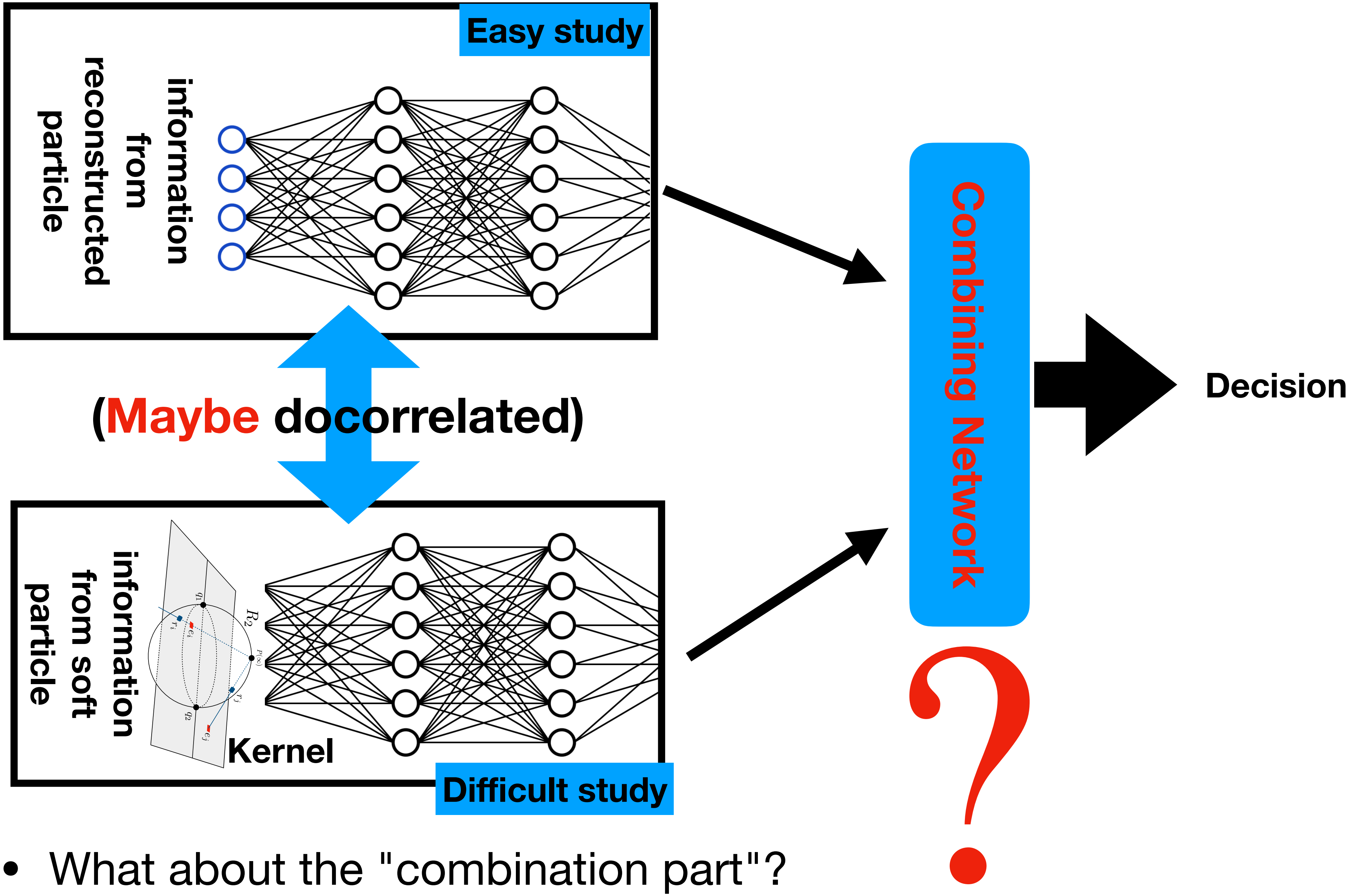
# Decorreation using a Kernel trick




MP (2023)

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

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




- What about the "combination part"?

  - Does a Neural Network learn both from kinematics and QCD in the equal basis (without mostly relying on "Easy" part) ?

- What about the "combination part"?

  - Does a Neural Network learn both from kinematics and QCD in the equal basis (without mostly relying on "Easy" part) ?

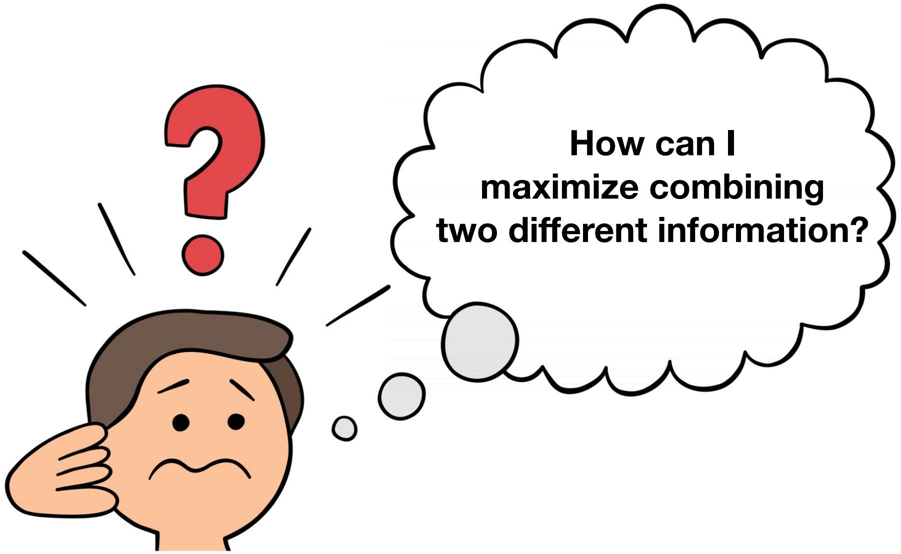
- **I have suffered from this problem more than a year...**

  **-** Above question is related to the **core question of ML**
  : How do you know what a Machine does learn ?

  **The explainable A.I.**

- Sunghoon Jung, Dongsub Lee, Ke-Pan Xie (Eur. Pys, 2020)
- Jin Choi, Un-Ki Yang (progress)

# Fortunately, I have a chance to discuss with Kayoung Ban




MP


- In Machine learning, a simple solution is always welcomed
as it provides the better convergence !
(with a few training data)

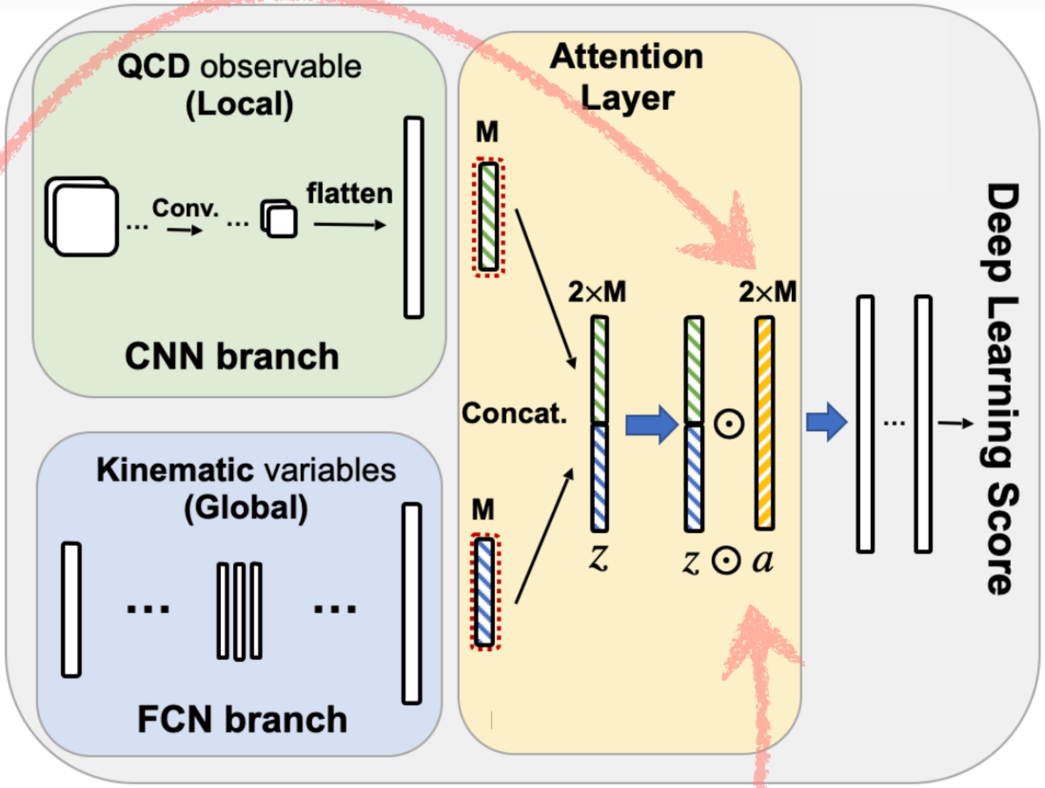
Kayoung Ban


We can look into ML with attention mechanism

# Method - Attention Layer

* Minh-Thang Luong, Hieu Pham, and Christopher D. Manning,*"**Effective approaches to attention-based neural machine translation**," (2015), arXiv:1508.04025 [cs.CL].*
* Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, *"**Neural machine translation by jointly learning to align and translate**," arXiv preprint arXiv:1409.0473*

**An innovative network for Natural language**

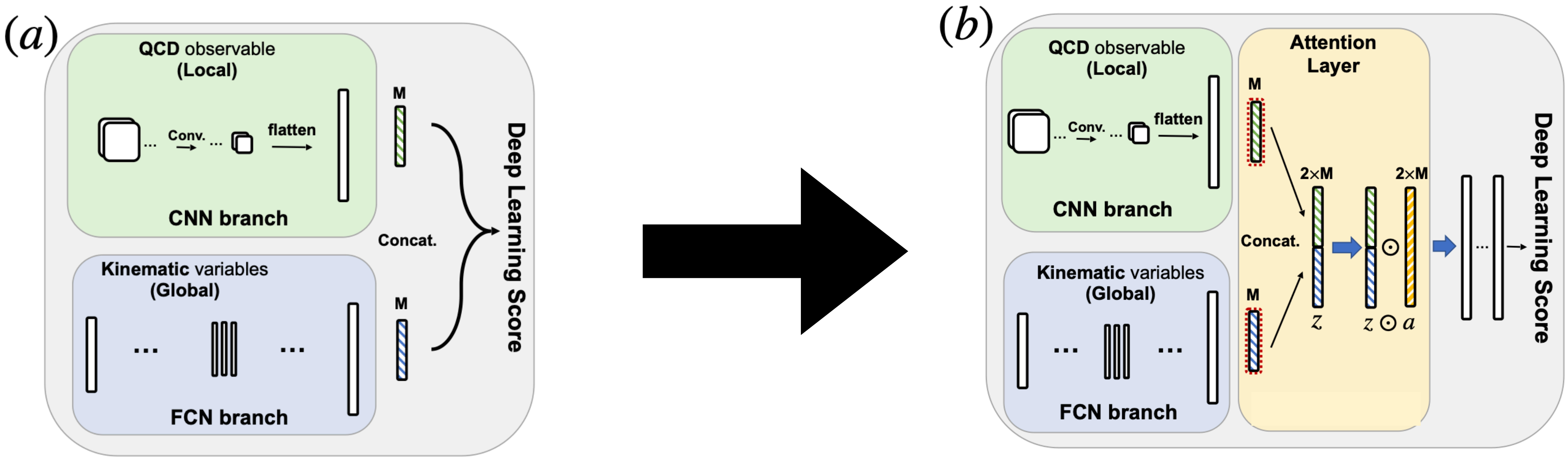



Attention score $\rightarrow a(z) = \dfrac{e^{f_{attn}(z)}}{\sum_{j=1}^{2M} e^{f_{attn}(z)_j}}$

where the layer $z = [z_{1:M} : z_{M+1:2M}]$ which is the concatenated layer from CNN and FCN respectively, and $f_{attn}(x) = W_{attn}x + b_{attn}$ is a trainable linear transformation.

Attention value $\rightarrow z \odot a$

- ✓ The attention values are computed as Hadamard product between the attention score ($a$) and the concatenated layer ($z$).
- ✓ We can interpret how much the model concentrates the two base models to classify signal and background.
- ✓ Attention values are then connected with a fully connected layer for classification.



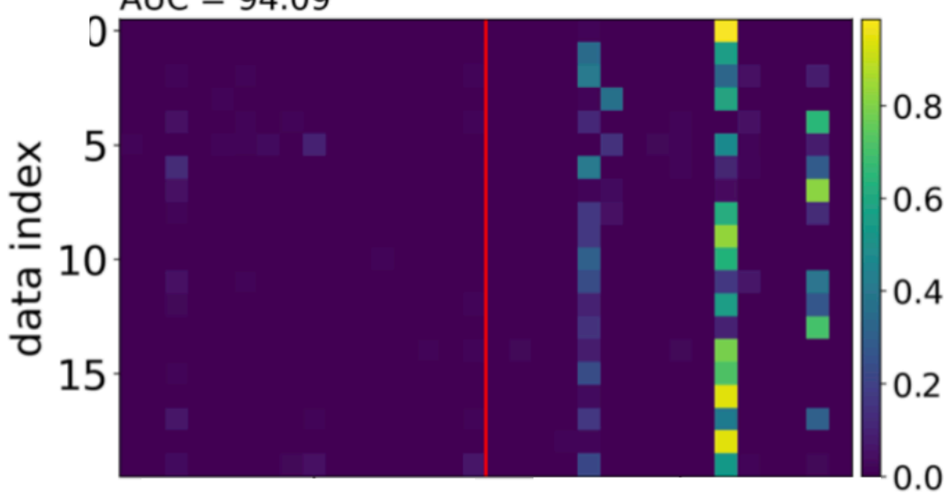

**Conventional**
**Black-box NN**

**Transparent NN**

- With **an attention layer**, <u>we can observe</u> that

  Network with the conventional combination **focus mostly on the easier material to study. (kinematics)**

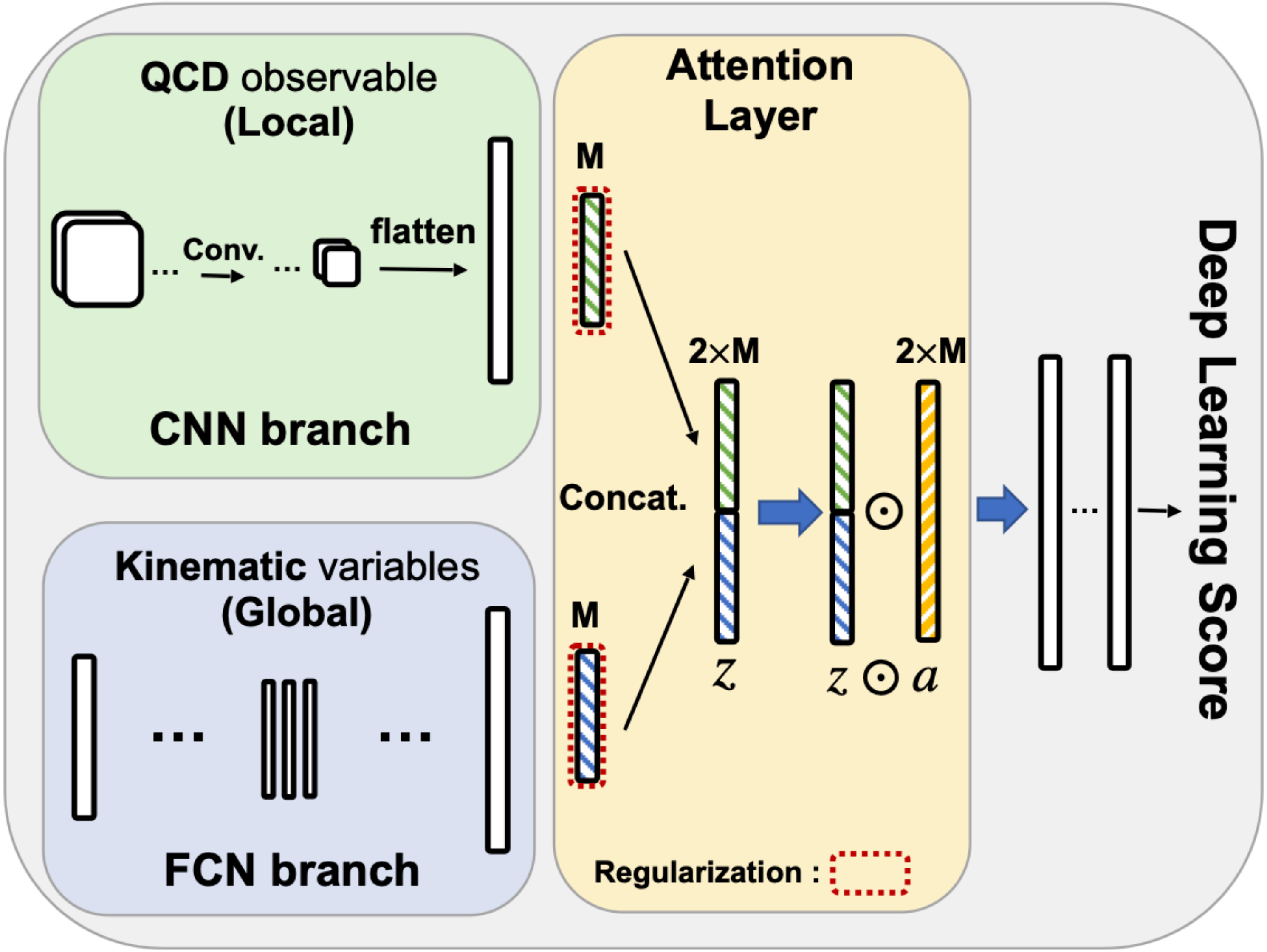
**Attention score**




**Network with QCD soft patterns**

**Network with Kinematics**


Now, let's do the magic part

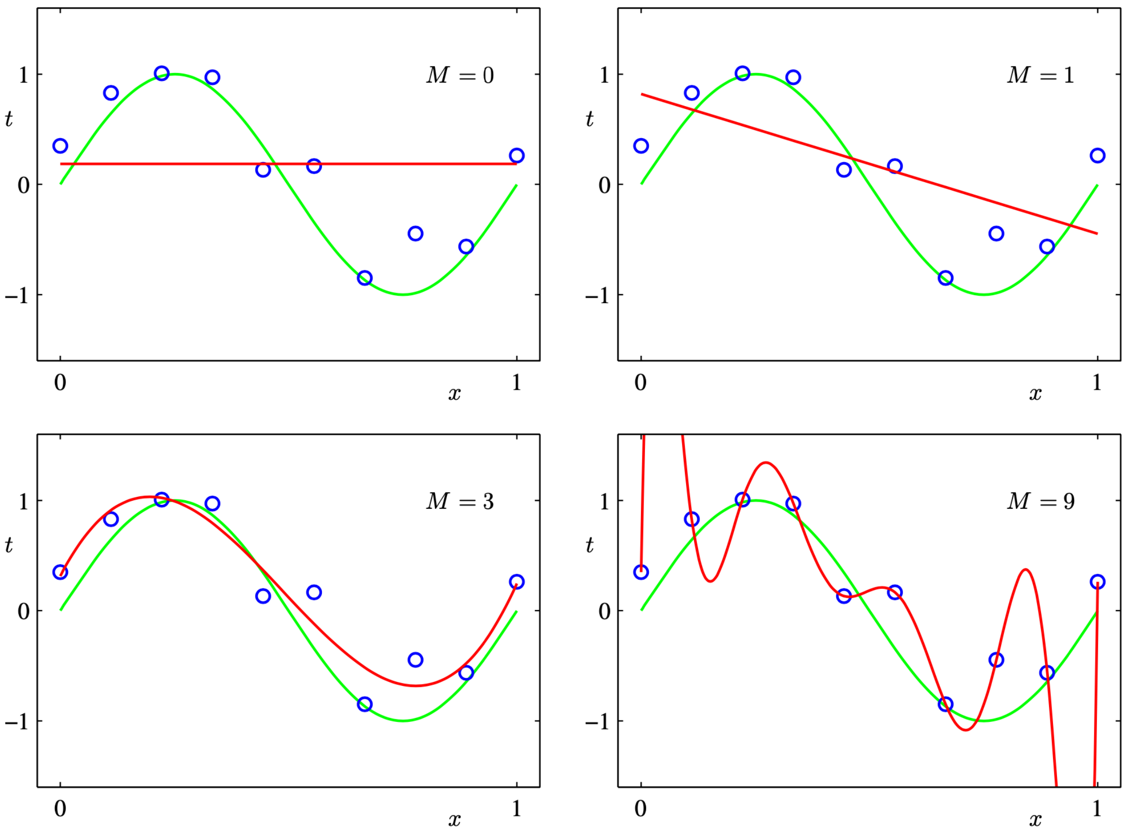


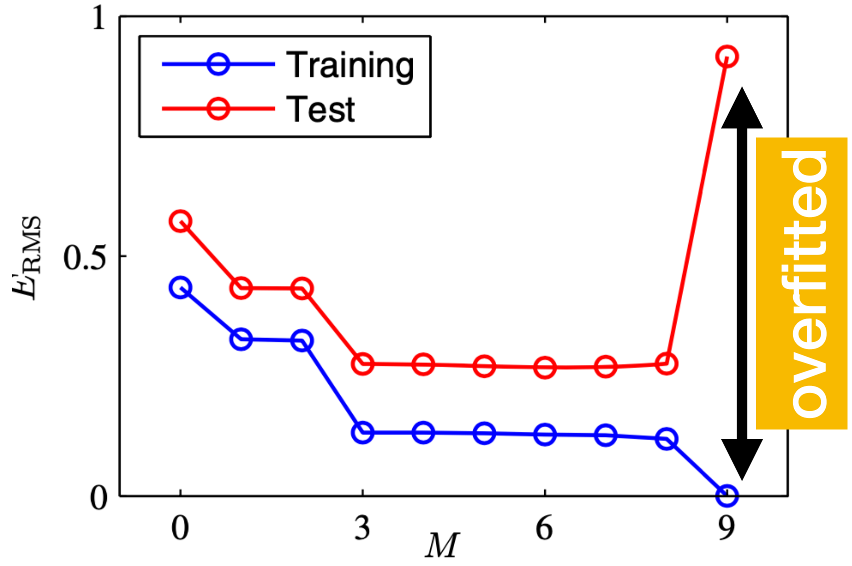

$$\text{Regularization term} = l_2 \times \sum_{k=1}^{M} W_k^2$$

# Actually, this is related to the core of ML - protecting overfitting

- With complicated networks, the overfitting can occur !

$$y(x, \omega) = \omega_0 + \omega_1 x + \omega_2 x^2 + \cdots + \omega_M x^M$$



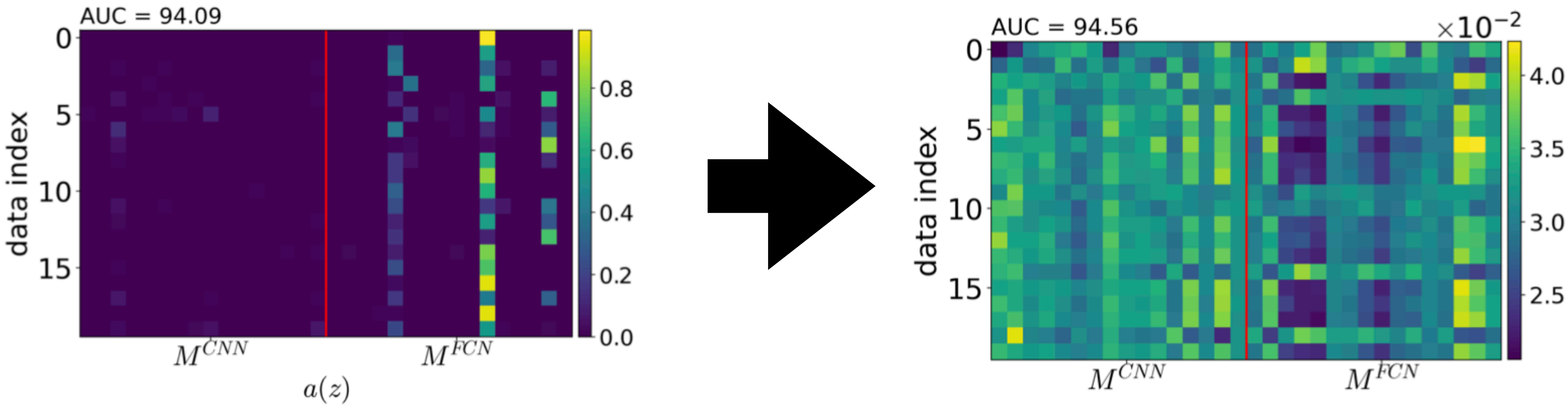




| | $M=0$ | $M=1$ | $M=6$ | $M=9$ |
|---|---|---|---|---|
| $w_0^\star$ | 0.19 | 0.82 | 0.31 | 0.35 |
| $w_1^\star$ | | -1.27 | 7.99 | 232.37 |
| $w_2^\star$ | | | -25.43 | -5321.83 |
| $w_3^\star$ | | | 17.37 | 48568.31 |
| $w_4^\star$ | | | | -231639.30 |
| $w_5^\star$ | | | | 640042.26 |
| $w_6^\star$ | | | | -1061800.52 |
| $w_7^\star$ | | | | 1042400.18 |
| $w_8^\star$ | | | | -557682.99 |
| $w_9^\star$ | | | | 125201.43 |

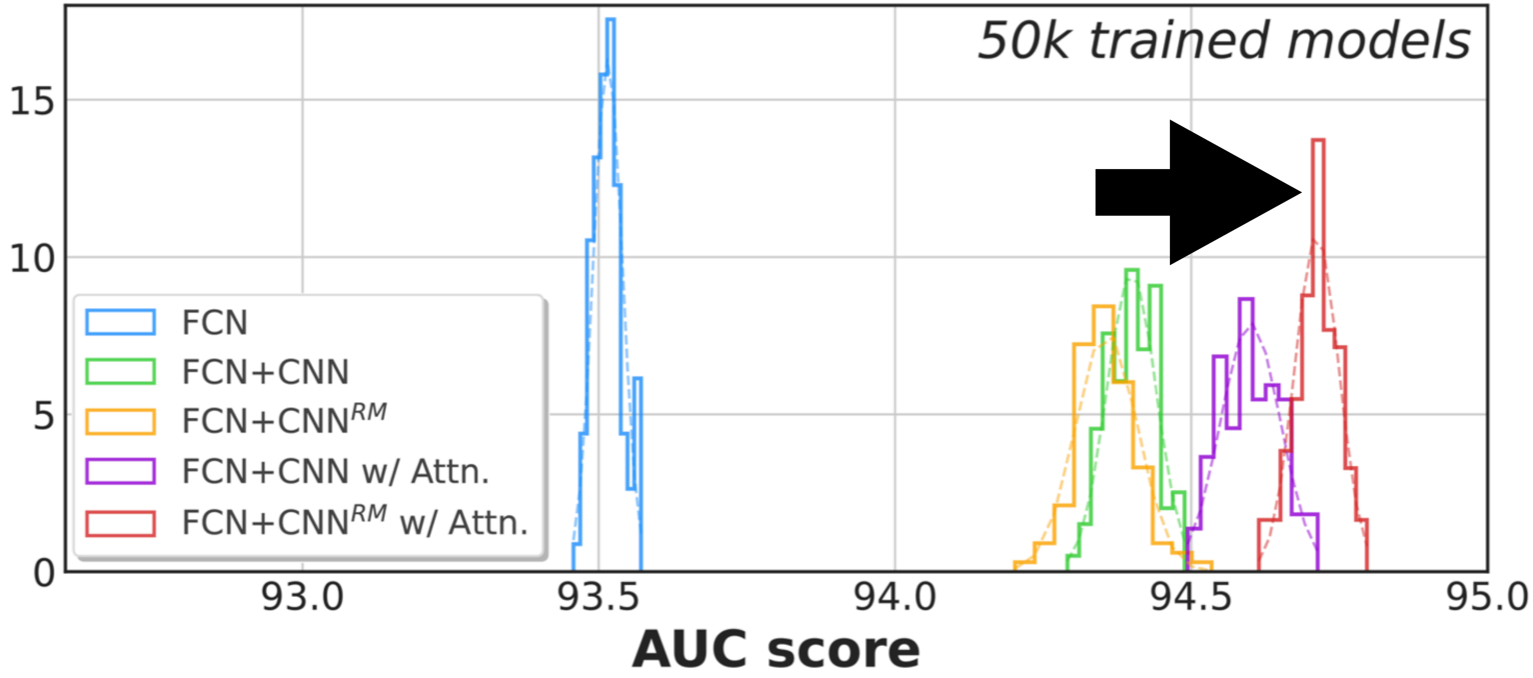
- Utilize a **regularizer**, a"**Lagrangian Multiplier**" in ML

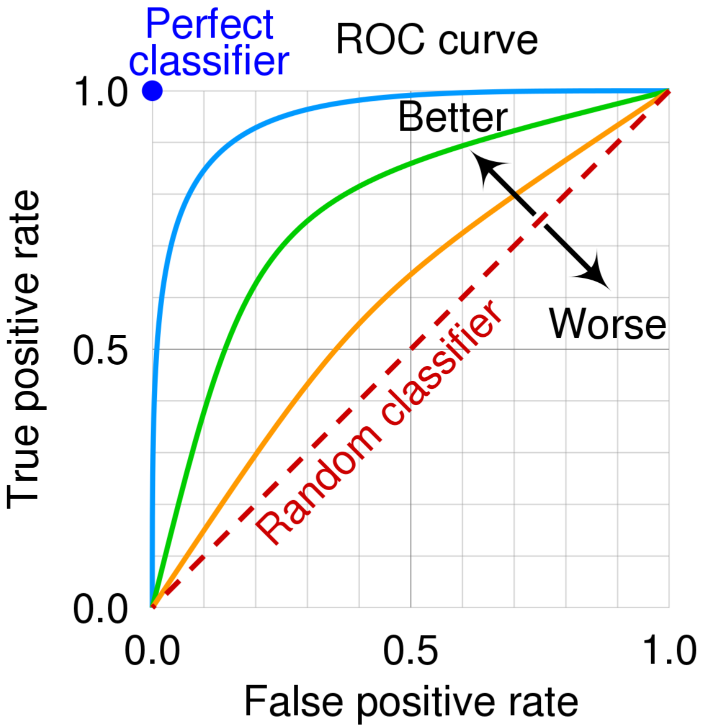
$$\boxed{\text{Regularization term} = l_2 \times \sum_{k=1}^{M} W_k^2}$$

- this **simple technique** works to **balance** information between local and global information!




- Eureka! I have missed this simple solution. This is an universal method.




- AUC score is the area under the ROC curve. Perfect discrimination = AUC score=1
- We get the improvement in the performance as a result.

# Conclusion

- There have been "huge" wave in applying machine learning methods in a particle phenomenology.
- So far, there was no serious study how to maximize the combination of two different types of information.
- To maximize the performance of collider analysis, one needs to utilize both "global" ( = full phase space) information and "localized" activities (= soft radiation patterns).
- We present a **simple but universal / effective** method to achieve the balanced training to maximize the performance.
  - This issue has not been taken seriously in the commercial advanced machine learning ( as they take a consistent approach)