Some Aspects of Deep Learning Frontier
in
Theoretical High Energy Physics

Partha Konar

Physical Research Laboratory

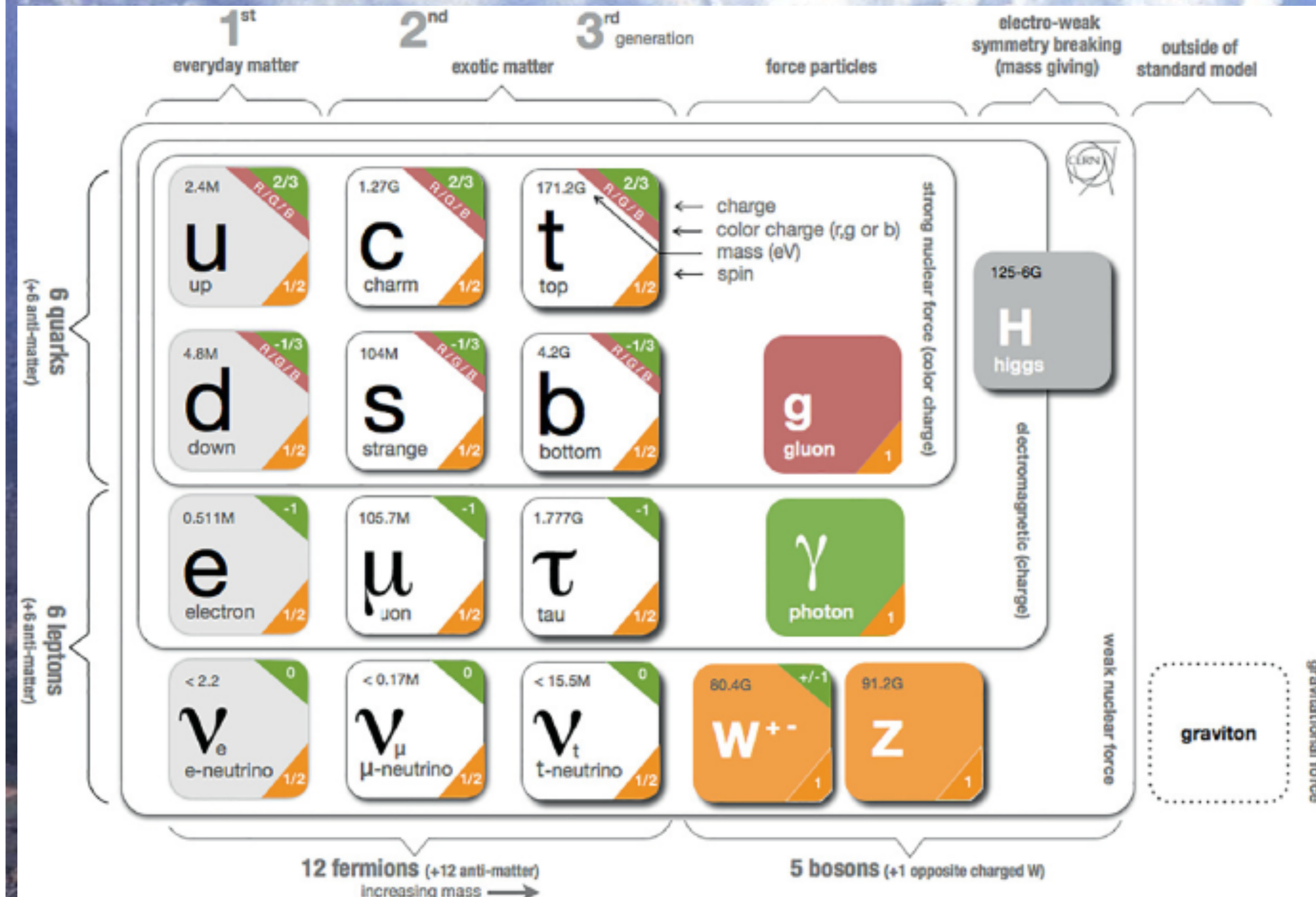
Ahmedabad, India

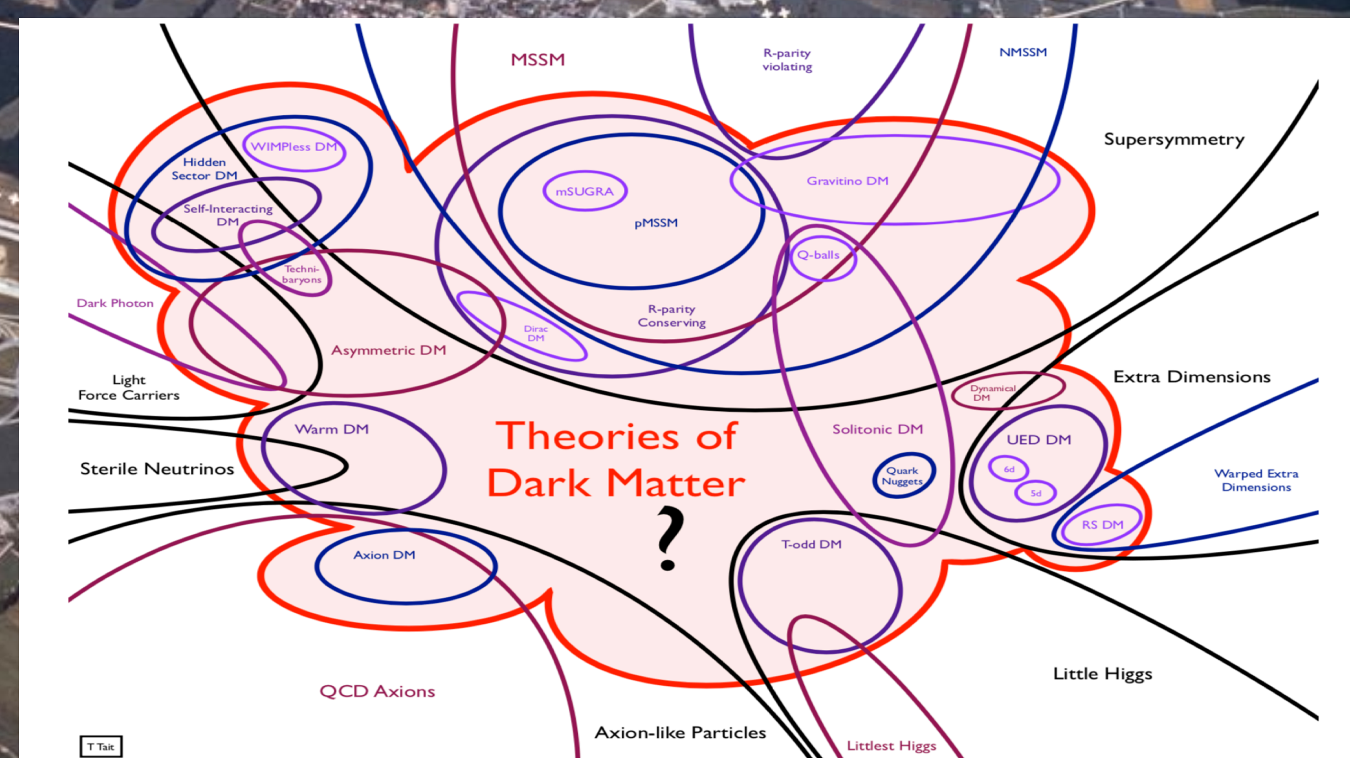
PHOENIX-2023 @ IIT Hyderabad

Dec 19, 2023

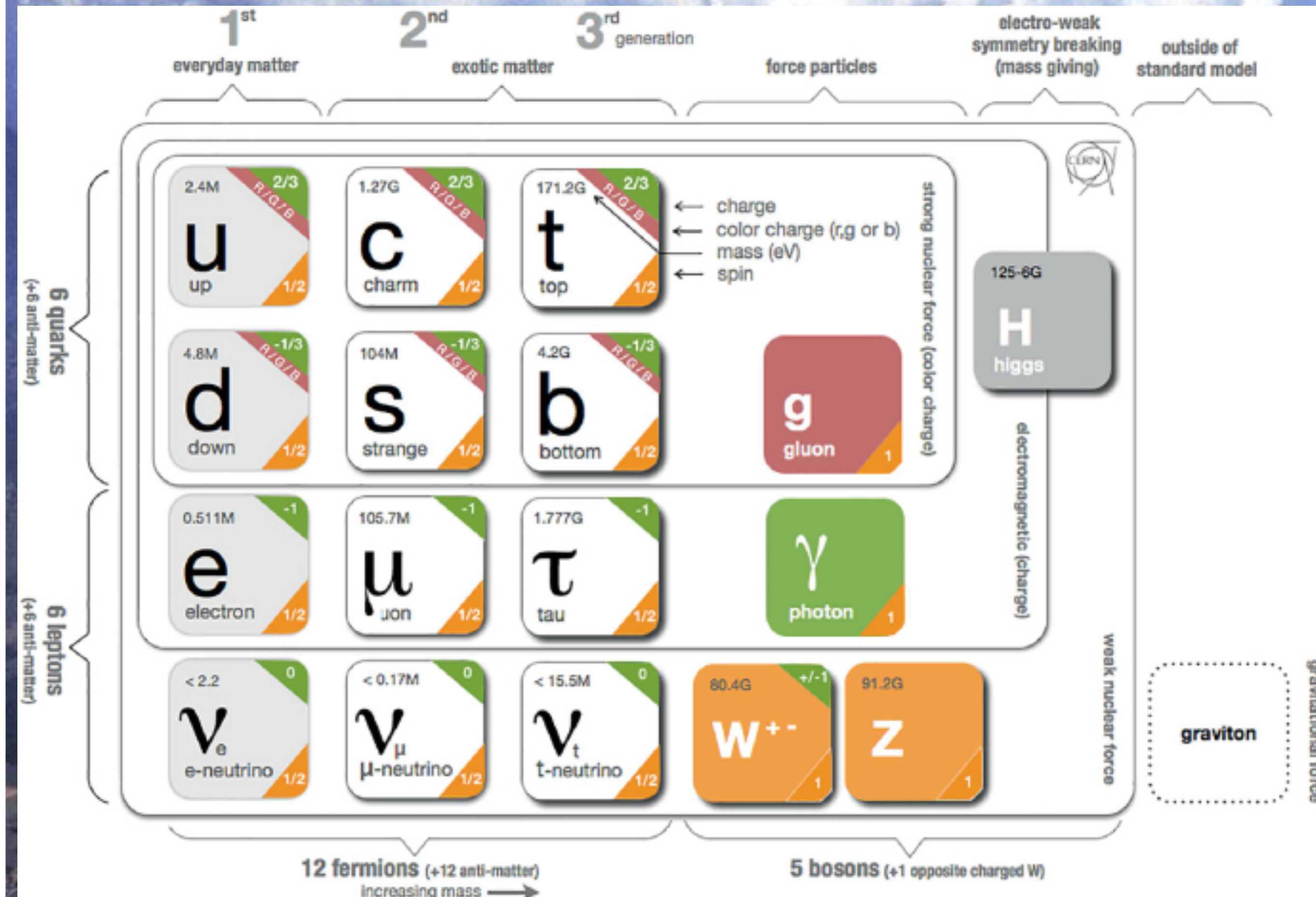
SEARCH AT LARGE HADRON COLLIDER (LHC)

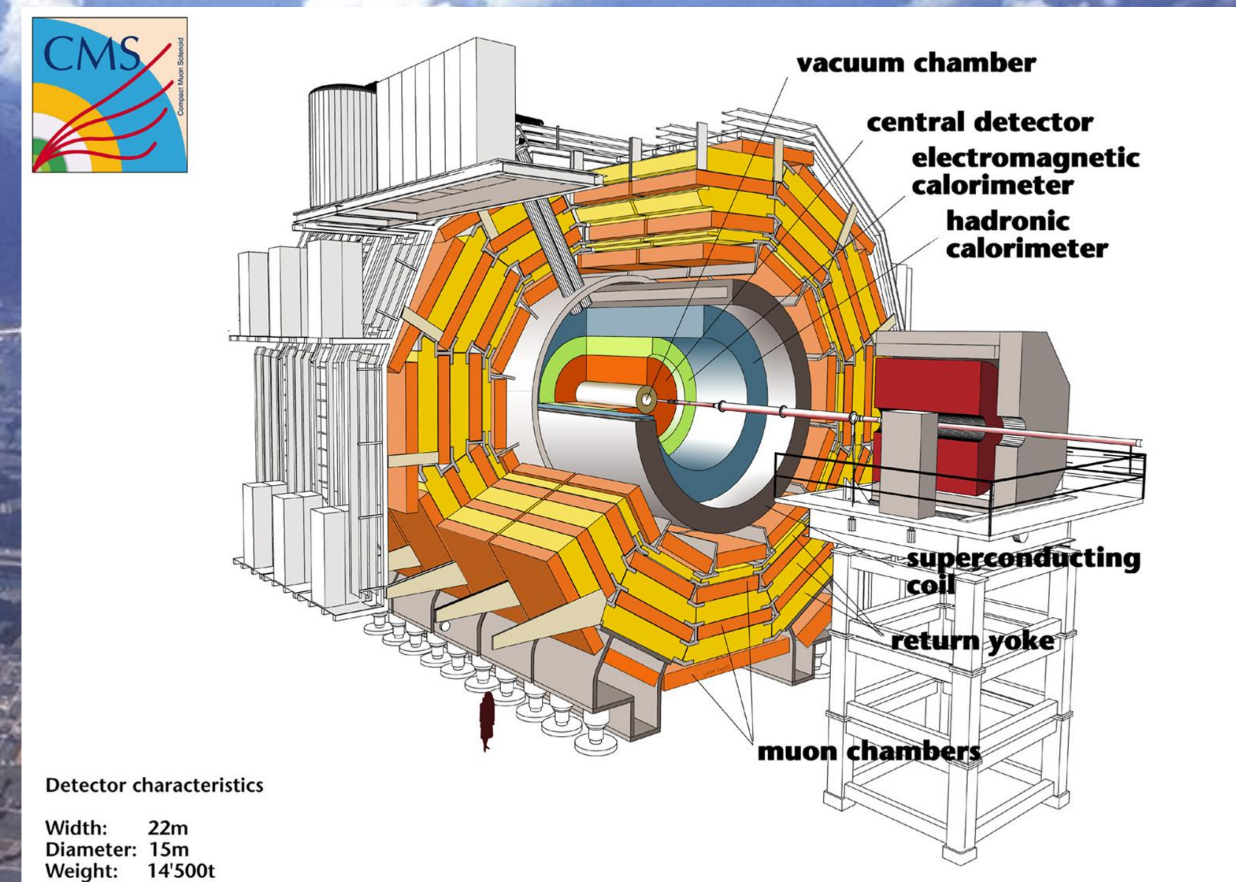
SEARCH AT LARGE HADRON COLLIDER (LHC)

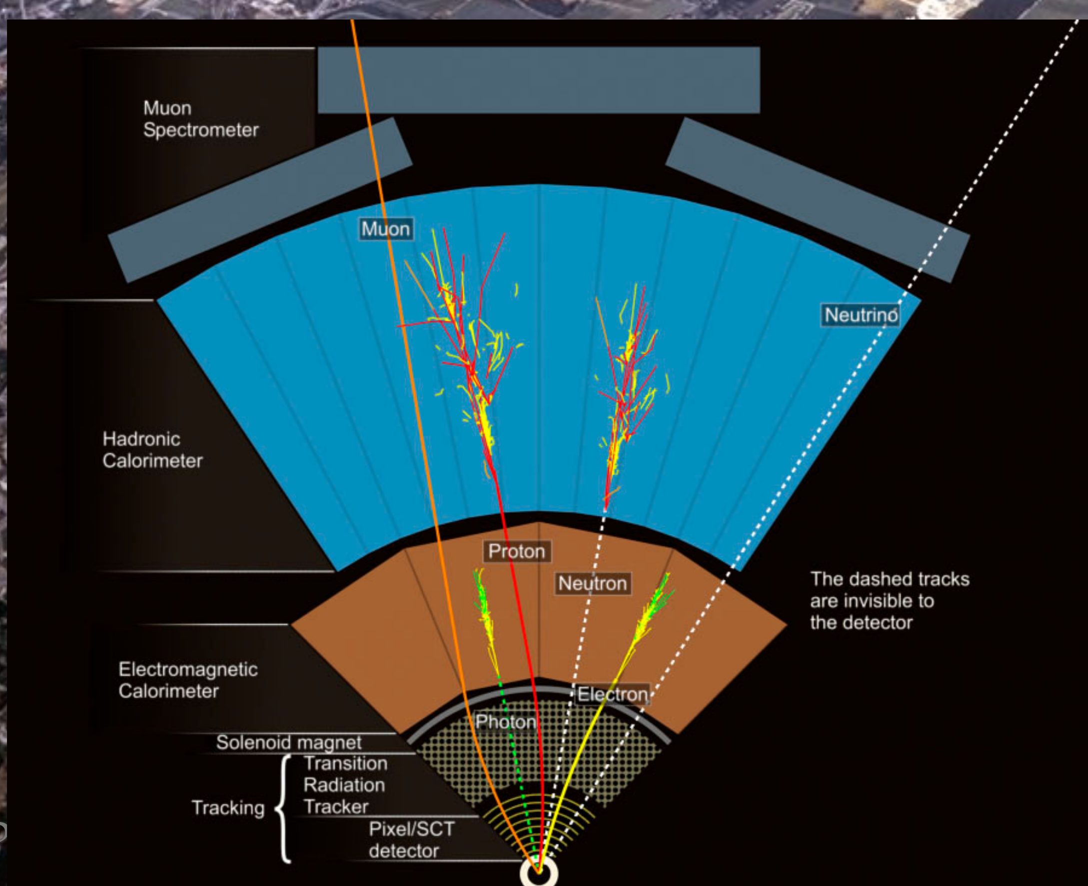


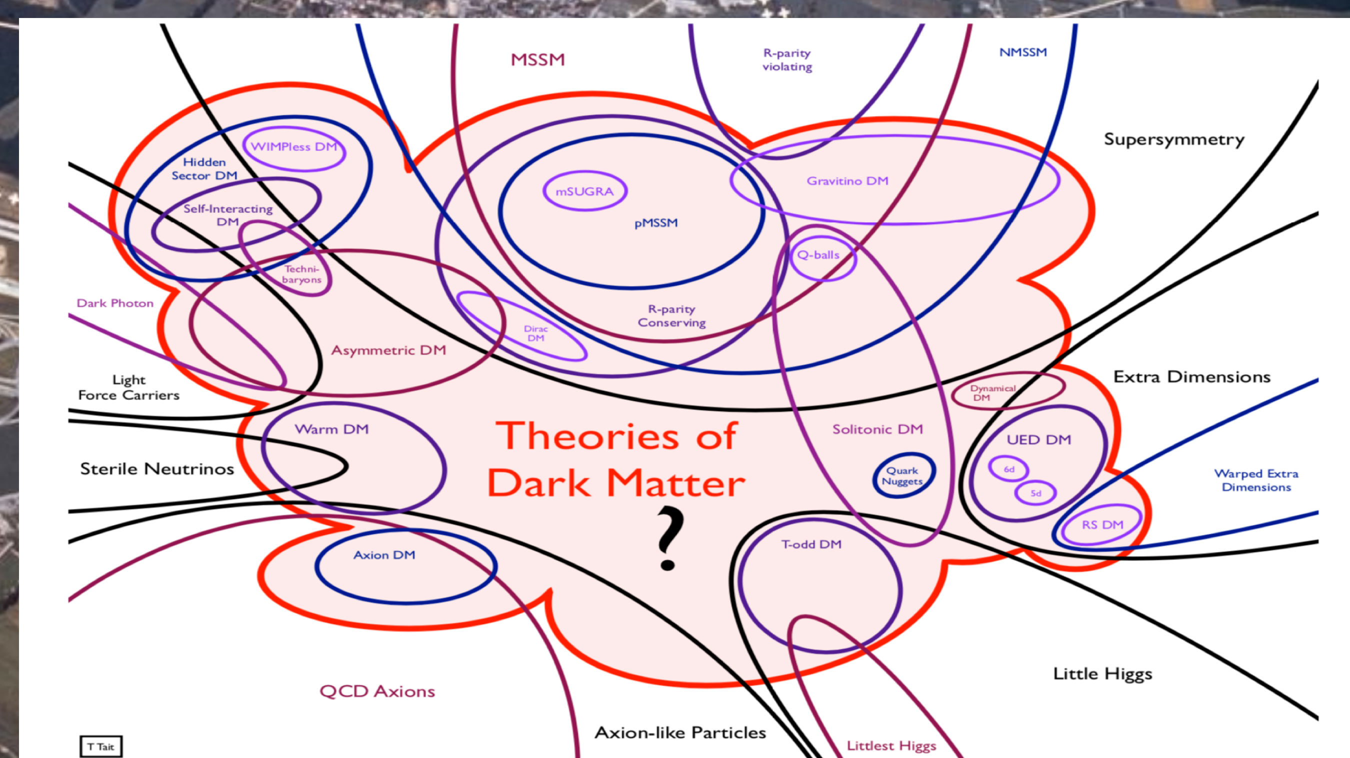


SEARCH AT LARGE HADRON COLLIDER (LHC)









MACHINE LEARNING

FOR HEP COMMUNITY

- Machine learning is not new for HEP community
- Used in low to high level experimental measurements with **track finding**, **calorimeter hit reconstruction, particle identification, energy/momenta reco**
- Multi Variate Analysis (**MVA**) & Boosted Decision Tree (**BDT**) used extensively on high level variables with primary focus as Classifier
  — Significant contribution in Higgs discovery
- **I focus from the viewpoint of the emergence of modern deep learning era** that greatly outperformed the previous state of arts in last one decade or so
- Driving forces -
  — Advent of graphics processor units (GPU) + Increased computing power
  — Large available data + Development of advanced ML architectures

MACHINE LEARNING

AND .. GOING DEEPER

- Universal function approximation: NN with a single hidden layer can approximate any continuous function to any desired precision!

- Deep learning models with **multiple hidden layers** solves the need for infinitely large no of nodes in shallow NN

- Learning **scalable** with data - larger data for better performance

- Deep learning models are now capable of **extracting feature directly from low level data**

  — End for physics intuitive high level variables from domain experts?

ARTIFICIAL NEURAL NETWORK

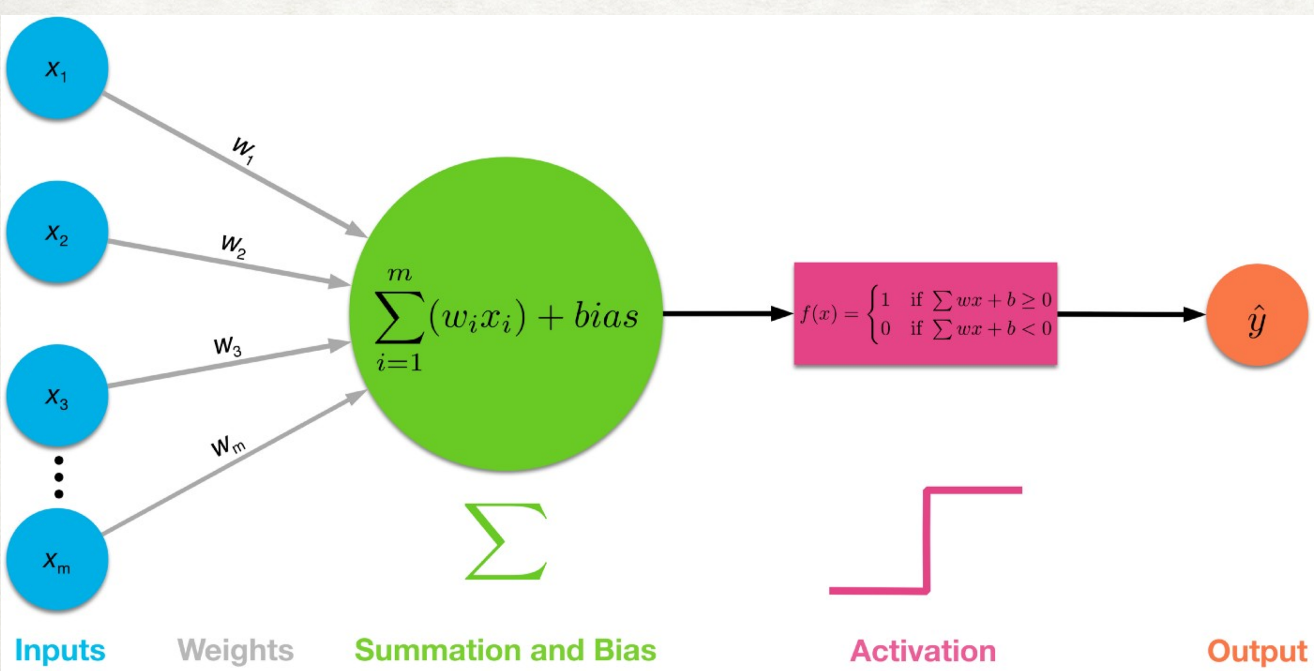
(ANN)

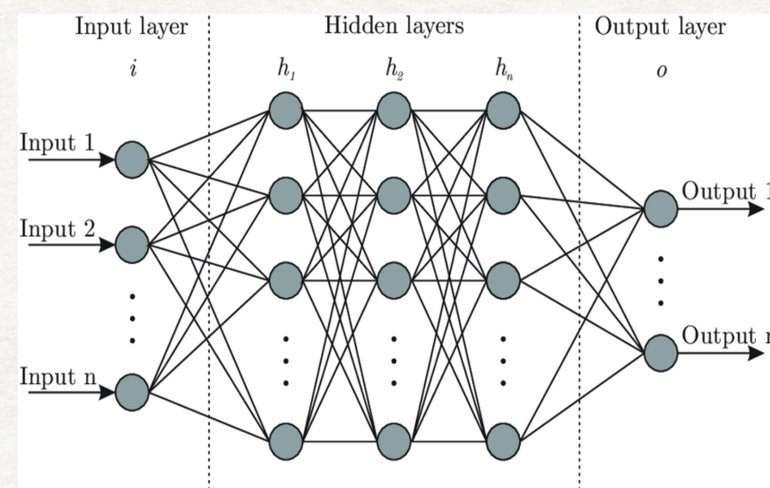
- **Search for a function** $f(\vec{x}, w) : X \to h_1 \cdots \to h_i \to h_{i+1} \cdots \to h_n \to Y$

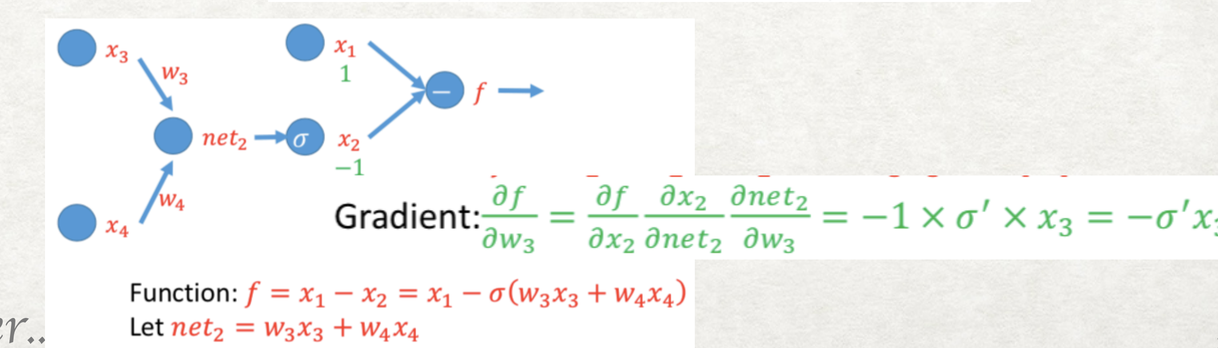
  X : **Input/obs. space;** Y: **Target space [low- dimensional space]**

  Optimize loss function $\mathscr{L}[y - f_w(x)]$; w - tunable parameters

- During training, trainable ***weight*** parameters (w) are ***learned*** by the back-propagation whose aim is to **minimize the loss function.**





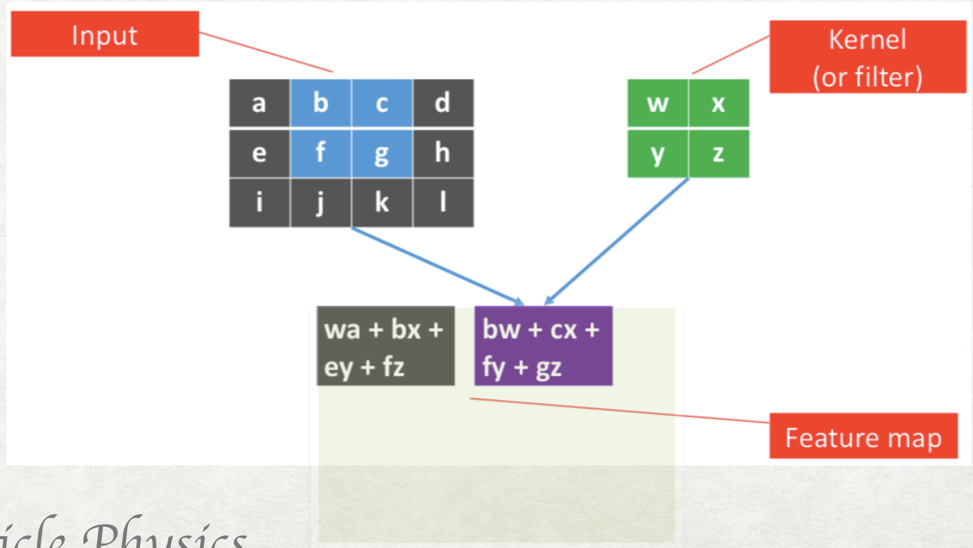


CONVOLUTIONAL NEURAL NETWORK (CNN)

- Most significant innovation in DNN - **Image processing**
- **Convolution architecture** rely on **local** and **global** features with **translation invariance**
- **Inductive biases based on locality and weight sharing**
- Image pixels are convoluted with no. of **kernel/filter** "k_j"

$$x_{i+1} = \sigma(wh + b) \quad \rightarrow \quad h_{i,j} = \sigma(k_j \,.\, h_i + b_j)$$

- Sharing same weights passing through full image
  => reduce tunable parameters drastically
  => translational symmetry on the network
- Algorithm first learn **edges** and **shapes**
  -> more **complex local features**
  -> leads to **global features**

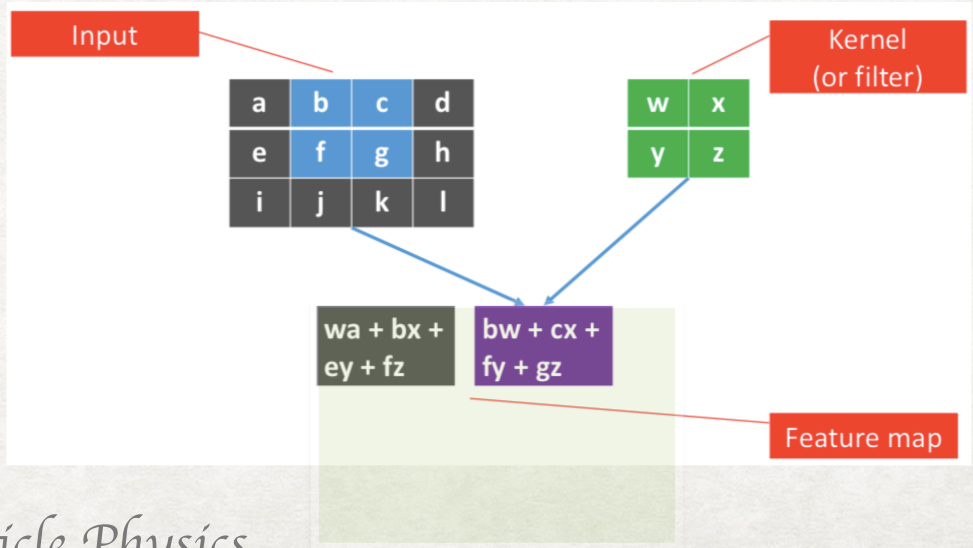


CONVOLUTIONAL NEURAL NETWORK (CNN)

- Most significant innovation in DNN - **Image processing**
- **Convolution architecture** rely on **local** and **global** features with **translation invariance**
- **Inductive biases based on locality and weight sharing**
- Image pixels are convoluted with no. of **kernel/filter** "k_j"

$$x_{i+1} = \sigma(wh + b) \quad \rightarrow \quad h_{i,j} = \sigma(k_j \,.\, h_i + b_j)$$

- Sharing same weights passing through full image
  => reduce tunable parameters drastically
  => translational symmetry on the network
- Algorithm first learn **edges** and **shapes**
  -> more **complex local features**
  -> leads to **global features**



DEEP MACHINE LEARNING

CATEGORY

Strategy —— *Representations* —— *Targets / tagging* —— *strategies*

Classification

- *Jet Image*
- *Event Image*
- *Sequence (Recurrent NN)*
- *Graph (Graph NN)*
- *Sets (Point cloud - Graph)*

- *Quarks vs gluons*
- *Boosted H / W/ Z/ Top tag*
- *New particles and models*
- *Particle tagging at detector*
- *Neutrino flavour*

- *Weak/ Semi/ Un-supervised*
- *Reinforcement Learning*
- *Quantum Machine Learn*
- *Feature Ranking*
- *Optimal Transport*

Regression

- *Parameter estimation*
- *Pileup mitigation*
- *Parton Distribution Func*
- *Symbolic Regression*
- *Function Approximation*

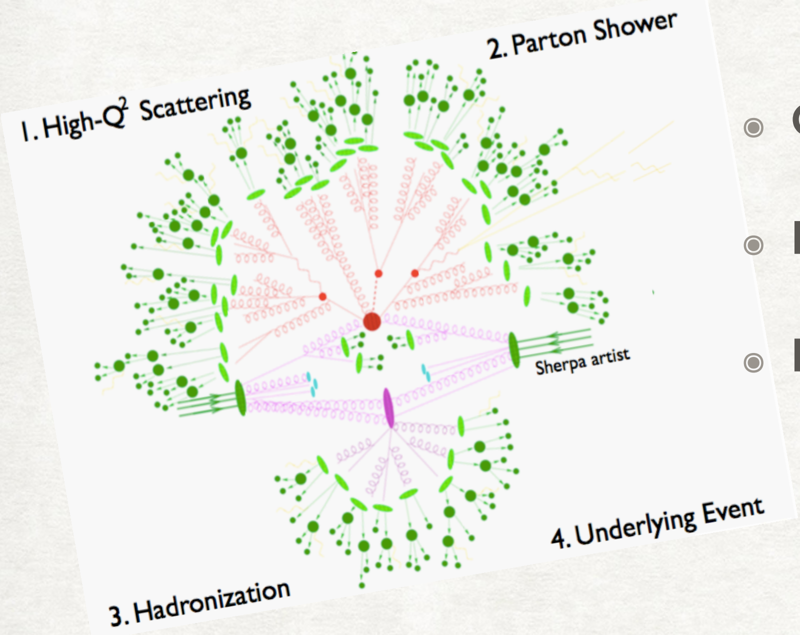
Generative models

- *GANs*
- *Autoencoders*
- *Phase space generation*
- *Normalizing flows*

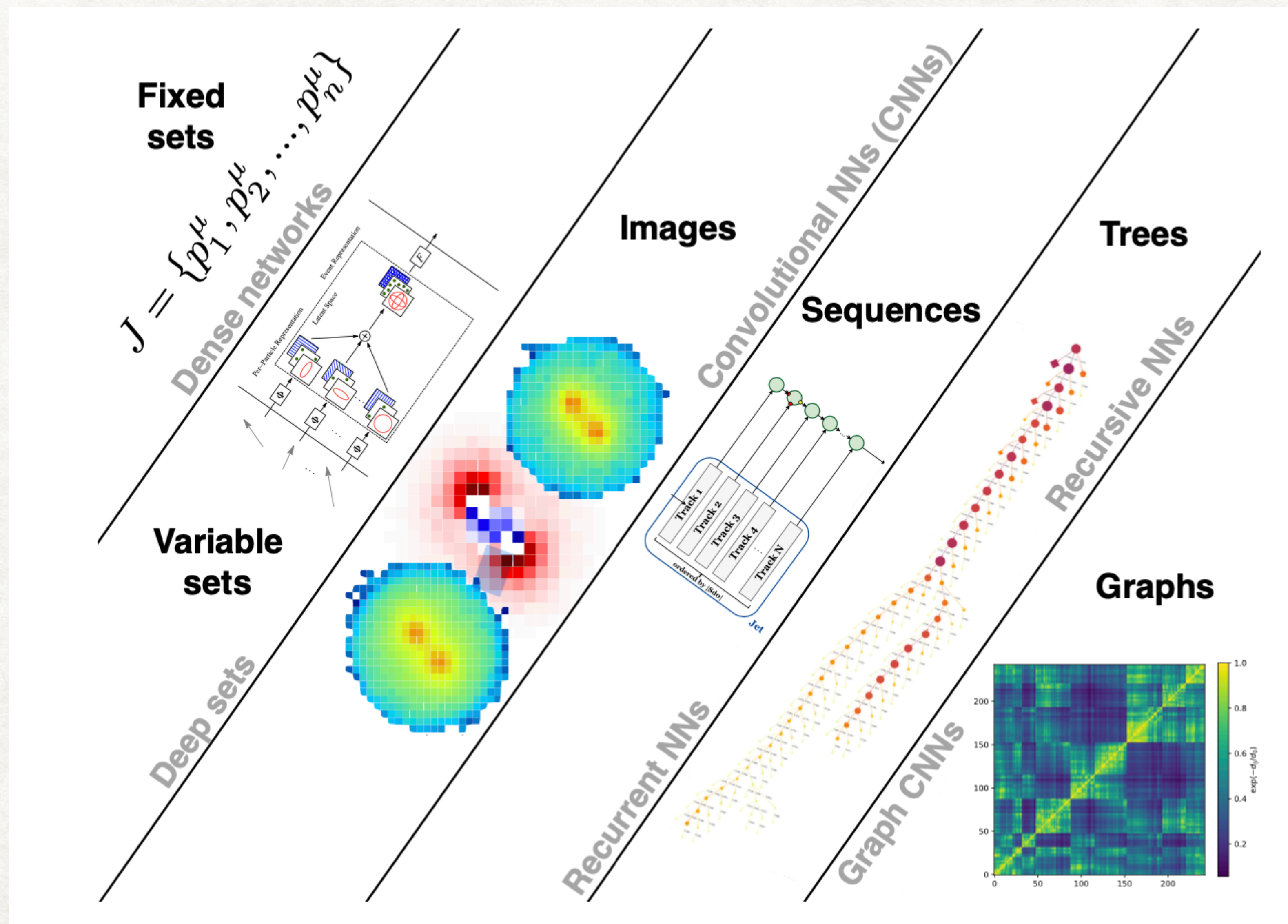
Anomaly detection

JET REPRESENTATION

JET DATA - IMAGES, SEQUENCES AND SETS



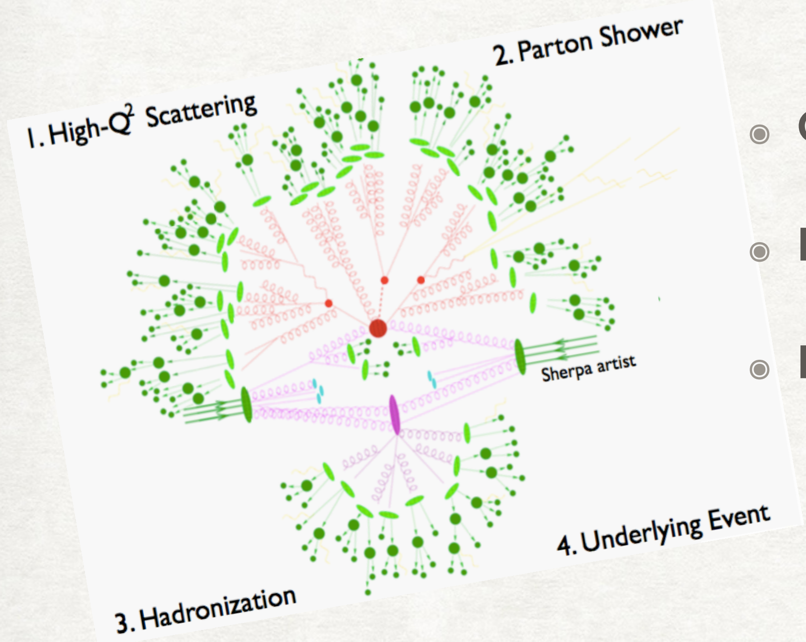
- QCD Jets have a rich & complex structure - perfect playing field
- How related to the first principles in Quantum Chromodynamics?
- No unique way for encoding radiation pattern into a particular data structure



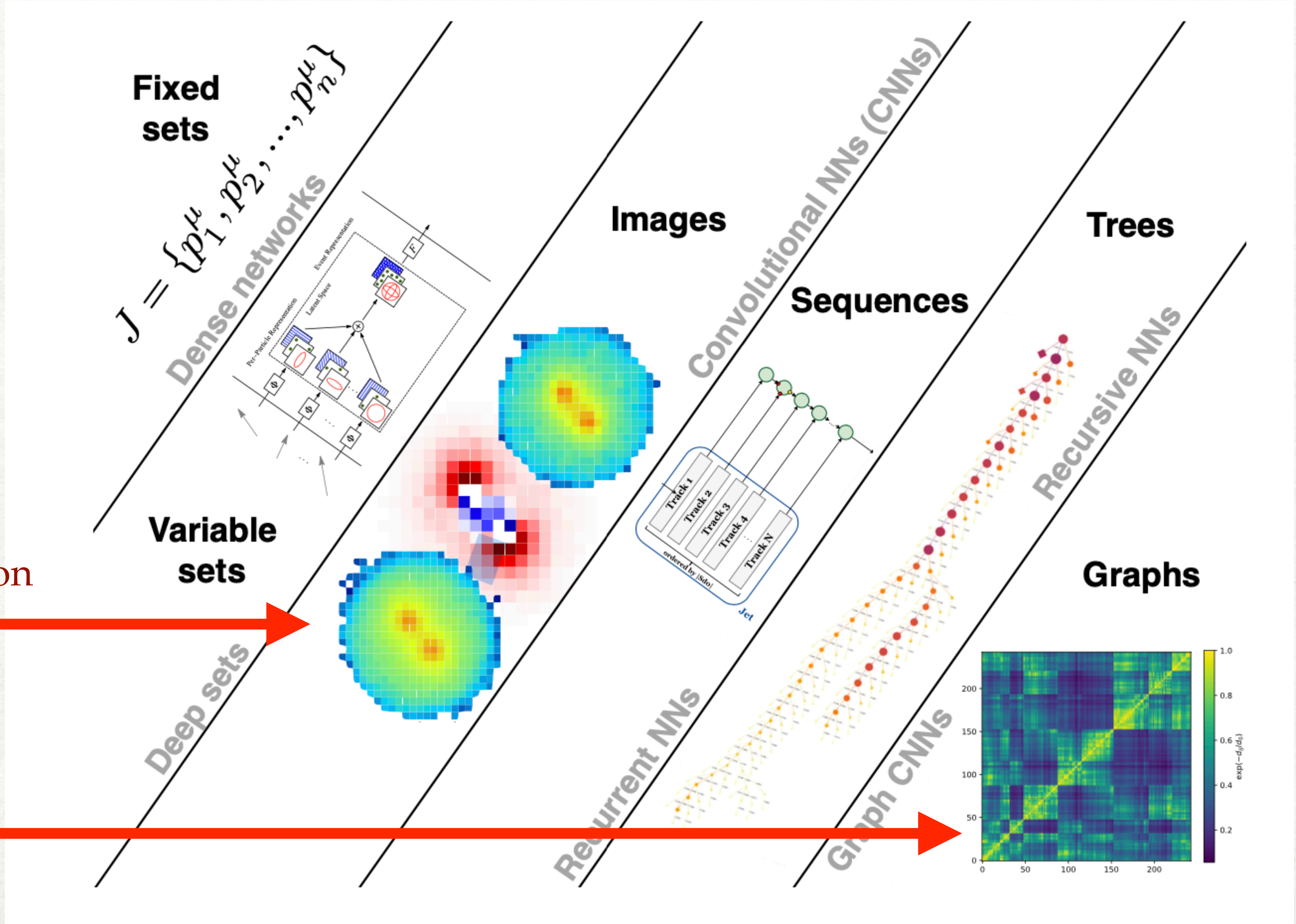
arXiv:1709.04464

JET REPRESENTATION

JET DATA - IMAGES, SEQUENCES AND SETS



- QCD Jets have a rich & complex structure - perfect playing field
- How related to the first principles in Quantum Chromodynamics?
- No unique way for encoding radiation pattern into a particular data structure



- Theoretically motivated Qn

> Low level image of jet and QCD radiation
Using CNN network

> Hadronic jet and QCD radiation
Using GNN network

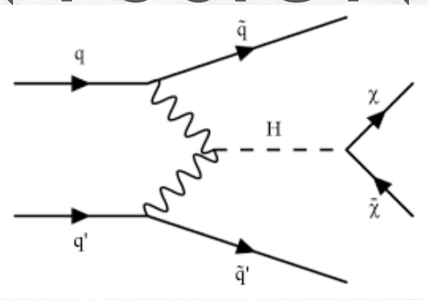
Deep-learning techniques for VBF Higgs searches – a case study in the invisible decay channel

Based on: 2008.05434 [Eur.Phys.J.C 80 (2020) 11, 1055]
2201.01040 [Phys.Rev.D 105 (2022) 11, 113003]

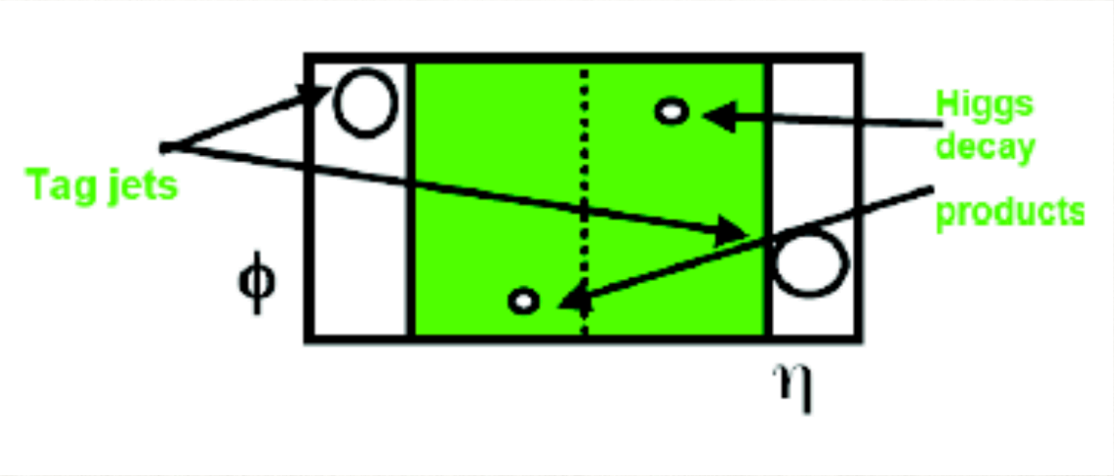
(Akanksha Bhardwaj, Partha Konar, Aruna K Nayak, Vishal Ng)

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

CONVOLUTIONAL NEURAL NETWORK



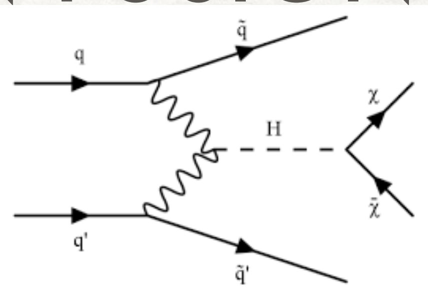
- ✓ Vector Boson Fusion (VBF) was a novel proposal for Higgs search
- ✓ Interesting topology for a VBF
  Two forward jets + large inv. Mass
  No central jet activity between them
  Decay products at the central region



- Qn. Can CNN learn feature for such event selection?
- Problem is even more difficult if Higgs is decaying invisibly!

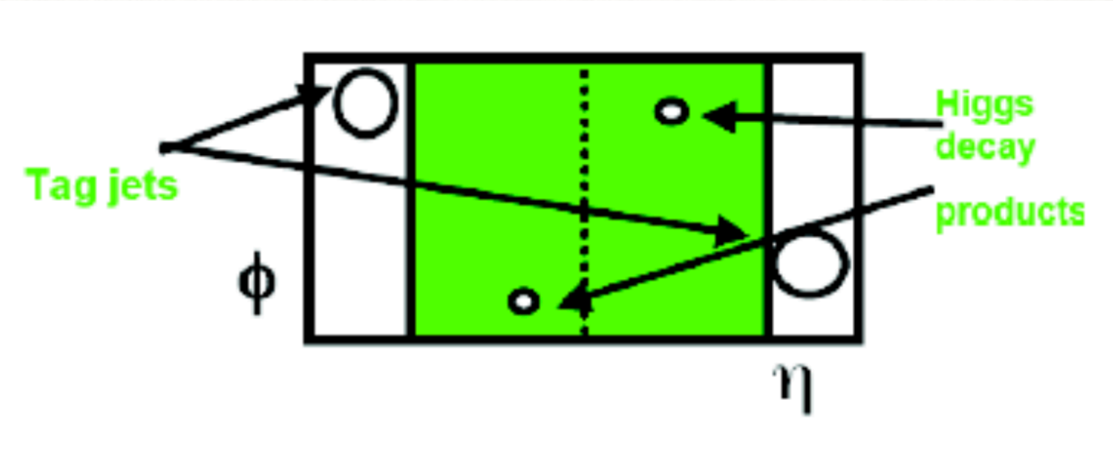
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

CONVOLUTIONAL NEURAL NETWORK



- ✓ Vector Boson Fusion (VBF) was a novel proposal for Higgs search

- ✓ Interesting topology for a VBF
  Two forward jets + large inv. Mass
  No central jet activity between them
  Decay products at the central region



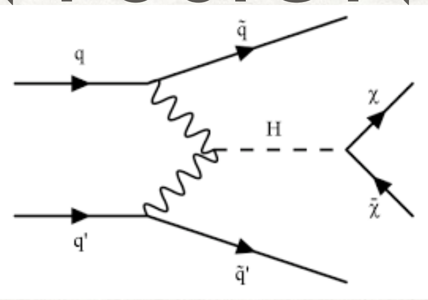
- Qn. Can CNN learn feature for such event selection?

- Problem is even more difficult if Higgs is decaying invisibly!

Br(h->inv) : VBF is most sensitive channel to give max bound on invisible BR

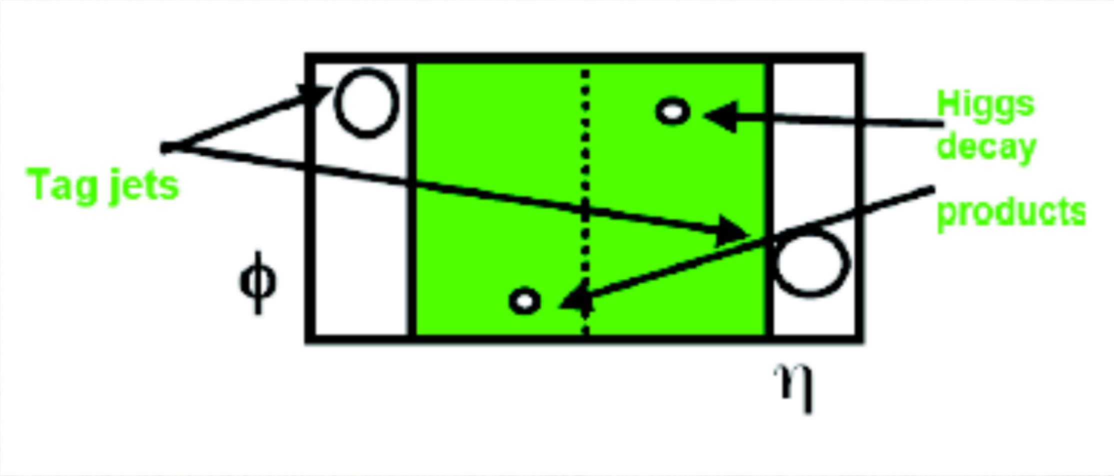
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

CONVOLUTIONAL NEURAL NETWORK



- ✓ Vector Boson Fusion (VBF) was a novel proposal for Higgs search

- ✓ Interesting topology for a VBF
  Two forward jets + large inv. Mass
  No central jet activity between them
  Decay products at the central region



- Qn. Can CNN learn feature for such event selection?

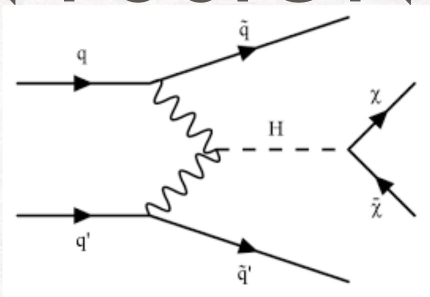
- Problem is even more difficult if Higgs is decaying invisibly!

Br(h->inv) : VBF is most sensitive channel to give max bound on invisible BR

Collider bounds on Br(h->inv) [23% (36fb), 10% (140fb)]>> SM prediction (<.01%)!!

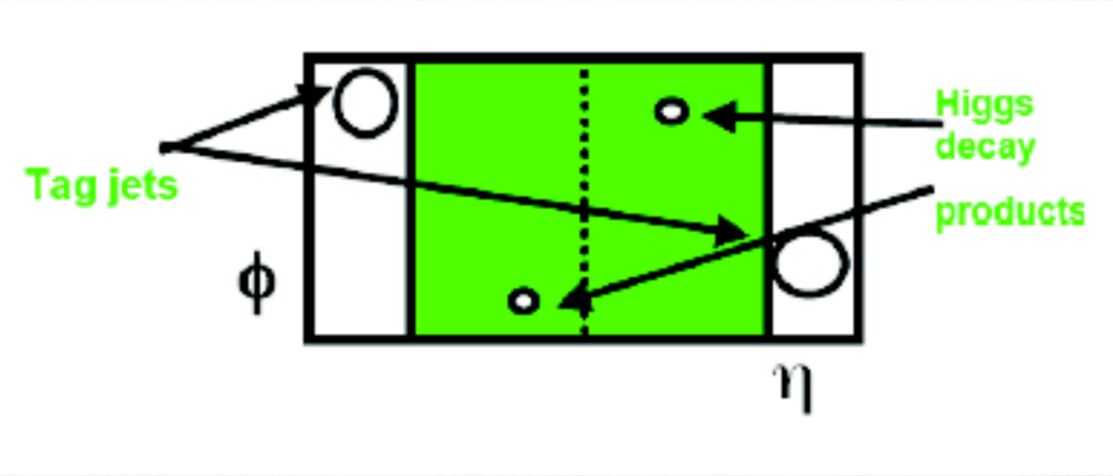
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

CONVOLUTIONAL NEURAL NETWORK



✓ Vector Boson Fusion (VBF) was a novel proposal for Higgs search

✓ Interesting topology for a VBF
Two forward jets + large inv. Mass
No central jet activity between them
Decay products at the central region



- Qn. Can CNN learn feature for such event selection?

- Problem is even more difficult if Higgs is decaying invisibly!

Br(h->inv) : VBF is most sensitive channel to give max bound on invisible BR

Collider bounds on Br(h->inv) [23%(36fb), 10% (140fb)]>> SM prediction (<.01%)!!

Lots of Dark Matter models (higgs portal) still exist because of this large limit

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

3 SET OF ANALYSIS

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

3 SET OF ANALYSIS

A. **CMS analysis with 36 fb^-1 data [Based on expert level VBF feature]**
Simulated Signal and BG => Reproducing CMS "BR upper limit" result

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

3 SET OF ANALYSIS

A. **CMS analysis with 36 fb^-1 data [Based on expert level VBF feature]**
   Simulated Signal and BG => Reproducing CMS "BR upper limit" result

B. **DML with sets of three different high level data [ANN]**
   1. Kinematic data : Event-kinematics from reconstructed objects

$$\mathcal{K} \equiv (\ |\Delta\eta_{jj}|,\ |\Delta\phi_{jj}|\ ,\ m_{jj}\ ,\ MET\ ,\ \phi_{MET}\ ,\ \Delta\phi^{j_1}_{MET}\ ,\ \Delta\phi^{j_2}_{MET}\ ,\ \Delta\phi^{j_1+j_2}_{MET}\ )$$

   2. Radiative: Contains information about the QCD radiation pattern

$$\mathcal{R} \equiv (H_T^{\eta_C} | \eta_C \in \mathcal{E}) \quad , \quad H_T^{\eta_C} = \sum_{\eta < |\eta_C|} E_T$$

   3. Combination of above two

$$\mathcal{H}$$

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

3 SET OF ANALYSIS

A. **CMS analysis with 36 fb^-1 data [Based on expert level VBF feature]**
Simulated Signal and BG => Reproducing CMS "BR upper limit" result

B. **DML with sets of three different high level data [ANN]**

1. Kinematic data : Event-kinematics from reconstructed objects

$$\mathcal{K} \equiv (\ |\Delta\eta_{jj}|,\ |\Delta\phi_{jj}|\ ,\ m_{jj}\ ,\ MET\ ,\ \phi_{MET}\ ,\ \Delta\phi^{j_1}_{MET}\ ,\ \Delta\phi^{j_2}_{MET}\ ,\ \Delta\phi^{j_1+j_2}_{MET}\ )$$

2. Radiative: Contains information about the QCD radiation pattern

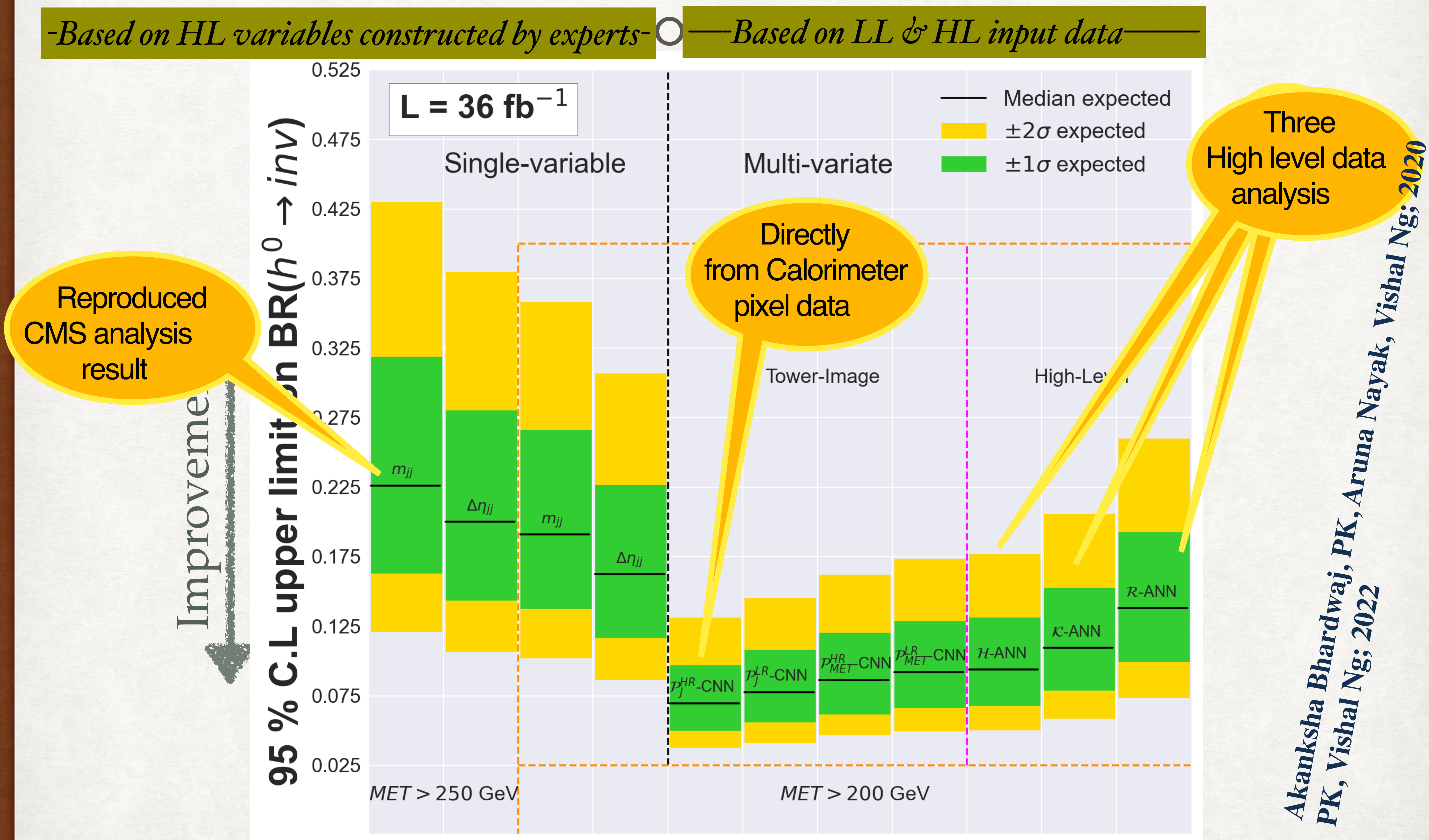
$$\mathcal{R} \equiv (H_T^{\eta_C} | \eta_C \in \mathcal{E}) \quad , \quad H_T^{\eta_C} = \sum_{\eta < |\eta_C|} E_T$$

3. Combination of above two

$$\mathcal{H}$$

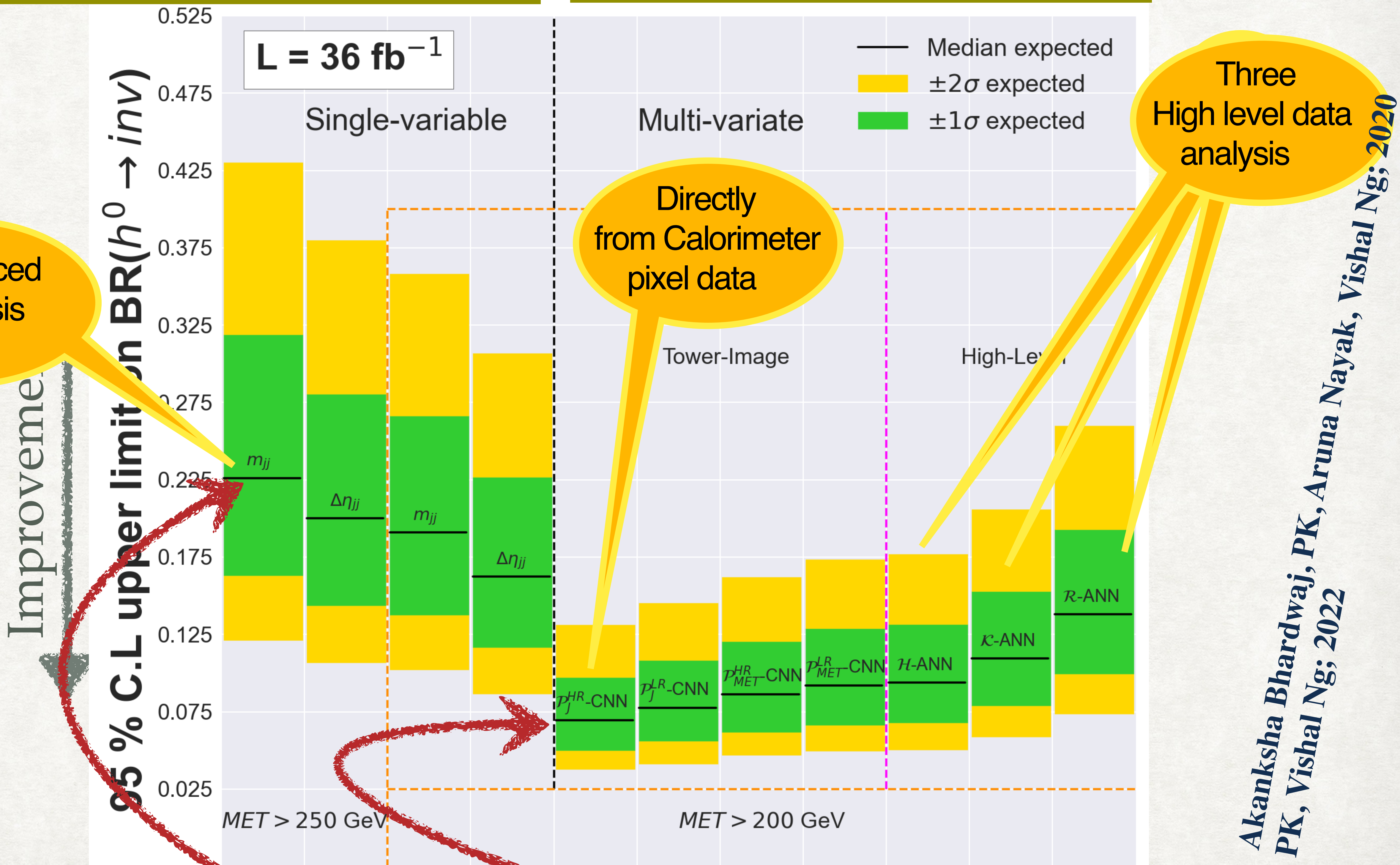
C. **DML with low level calorimeter input data [CNN]**
- Hi & Low resolution Calorimetry

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION



INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

-Based on HL variables constructed by experts- —Based on LL & HL input data—



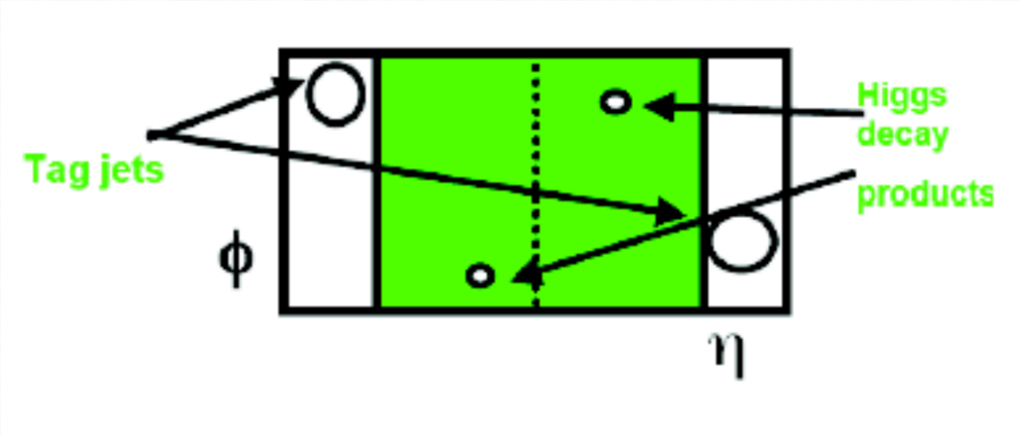
Akanksha Bhardwaj, PK, Aruna Nayak, Vishal Ng; 2020
PK, Vishal Ng; 2022

Factor of three improvement using the same data!

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

ROLE OF PARTON SHOWER

★In this simple setup with just two jets : NN minutely learned the **kinematic relation** & **radiation pattern** from the data



★Extra QCD radiation between two tag jets extremely significant!!

★Central-jet Veto:
Efficiently rejects large QCD backgrounds by vetoing events with additional central jet

★Qn. How *faithful the distribution function which NN learn?*

- *Perturbative Accuracy of Matrix Element Simulation :*
  LO vs NLO => Important for any process

- *Parton Shower recoil Scheme [*Dipole parton shower*]*
  => Wrong global scheme (for spacelike shower) used in most analysis

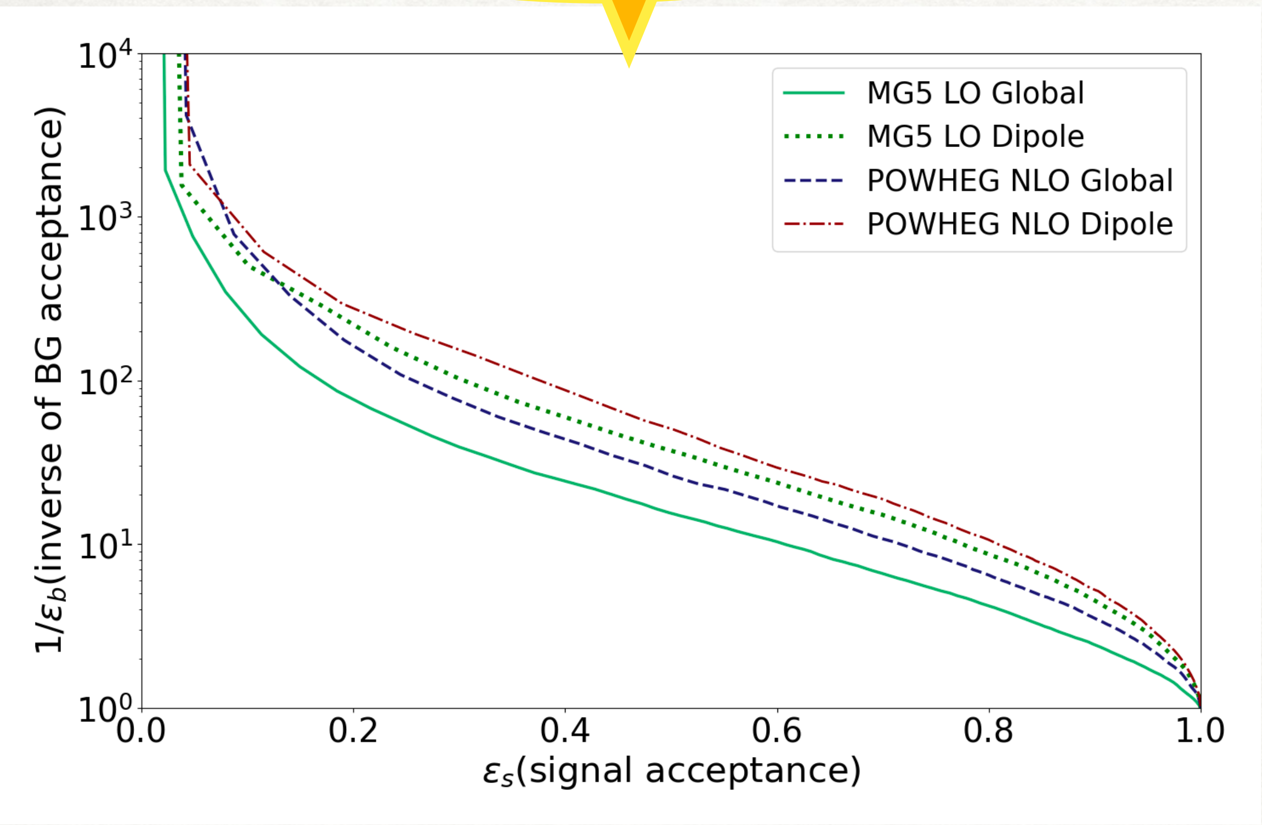
★True potential unfolds if theoretical predictions are accurate enough.

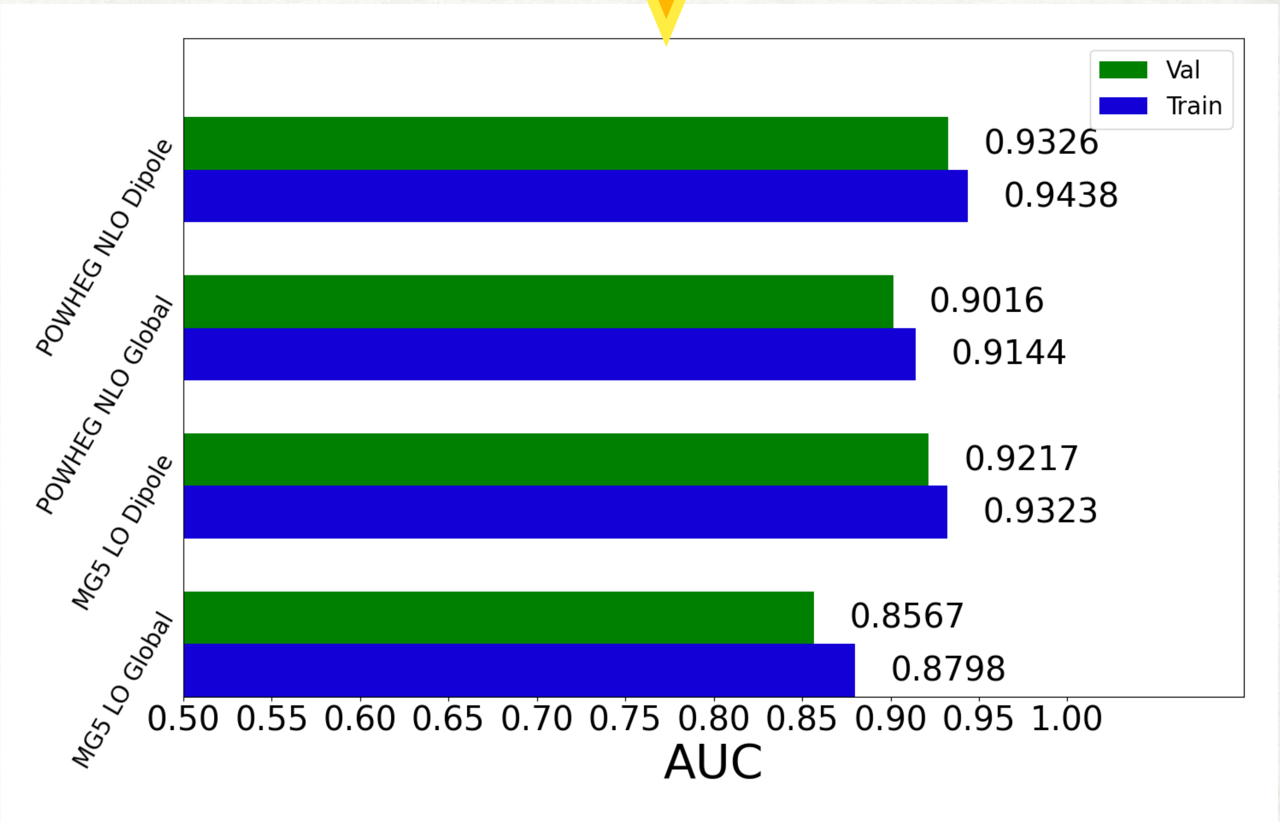
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

ROLE OF PARTON SHOWER

Receiver Operator Characteristics (ROC) Curve

AUC





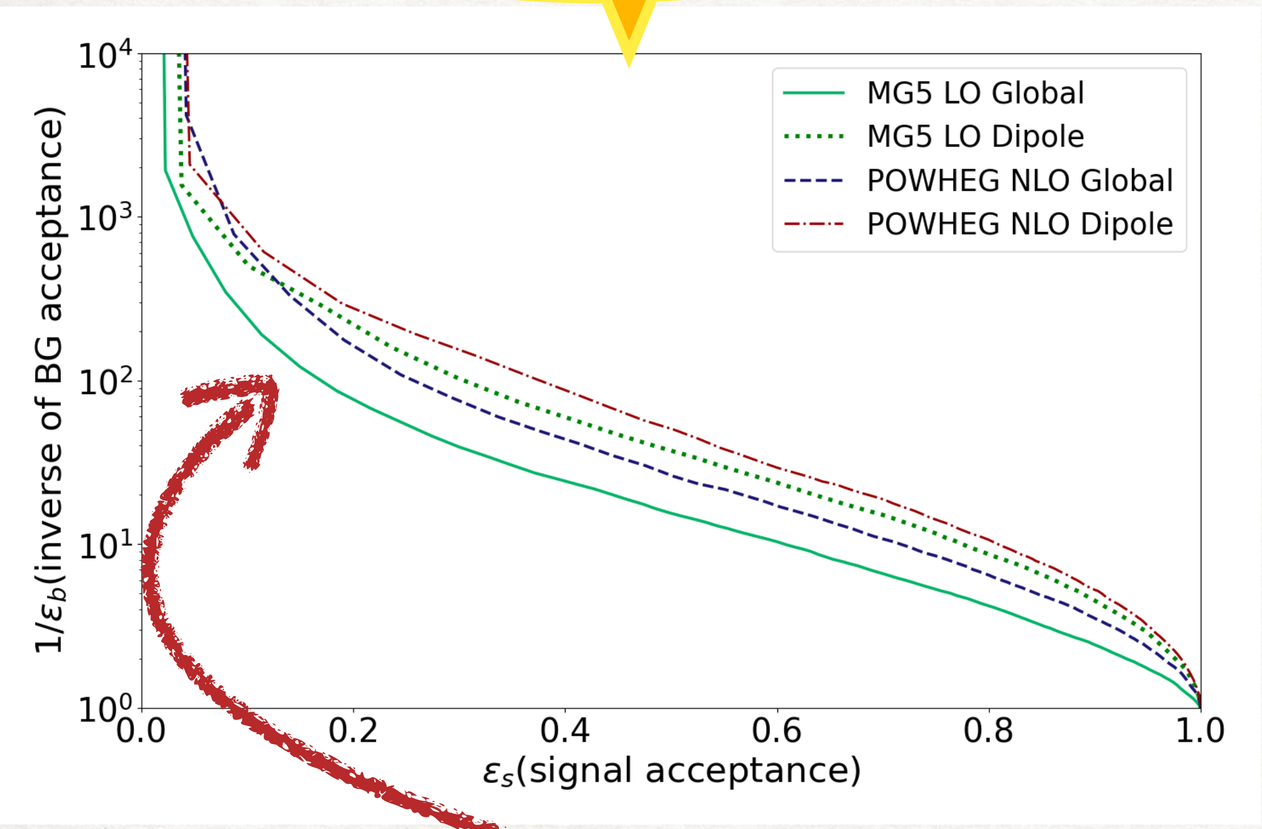
- ✓ LO + Global parton shower scheme shows lowest performance
- ✓ NLO + Dipole parton shower scheme shows best performance
- ✓ Rest two (LO+ Dipole & NLO+ Global) shows intermediate performance

Vishal Ng, PK (2021)

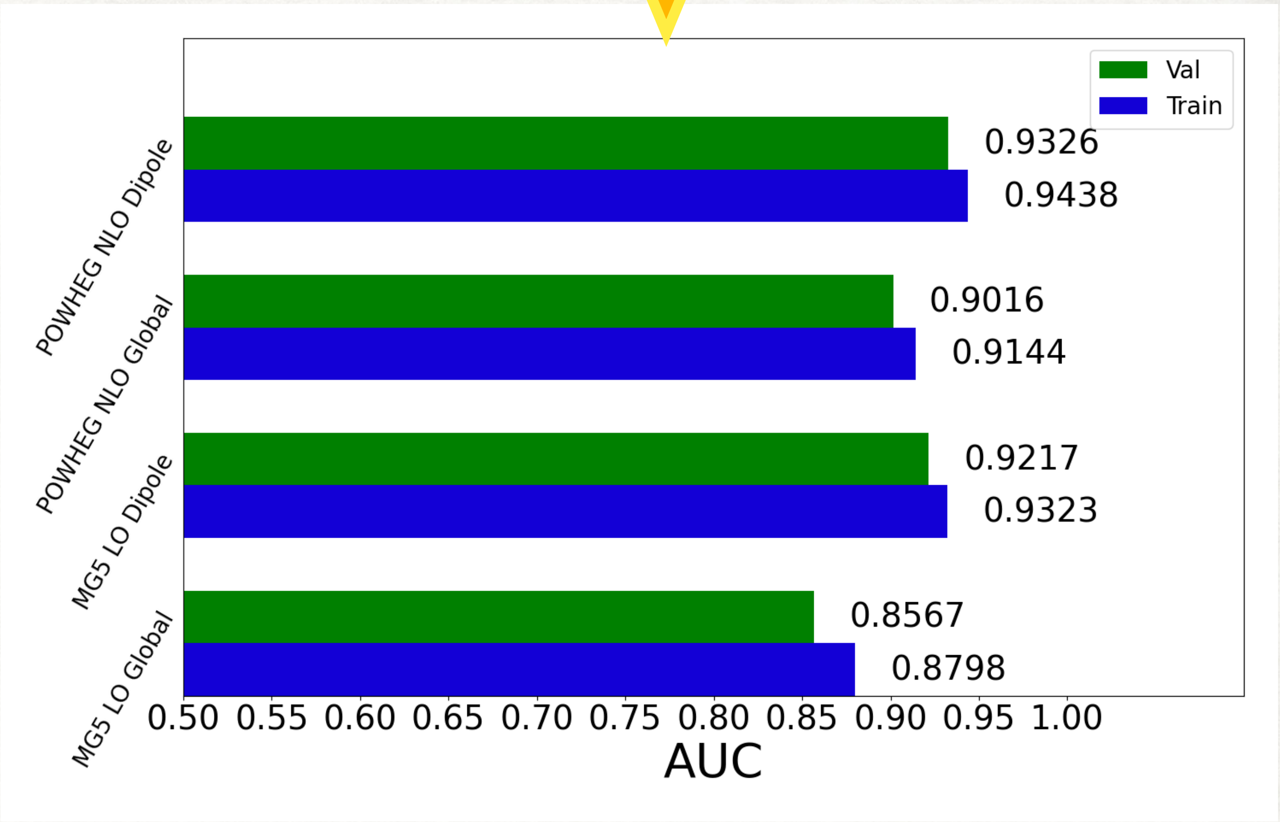
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

ROLE OF PARTON SHOWER

Receiver Operator Characteristics (ROC) Curve



AUC



- ✓ LO + Global parton shower scheme shows lowest performance
- ✓ NLO + Dipole parton shower scheme shows best performance
- ✓ Rest two (LO+ Dipole & NLO+ Global) shows intermediate performance

Common in analysis

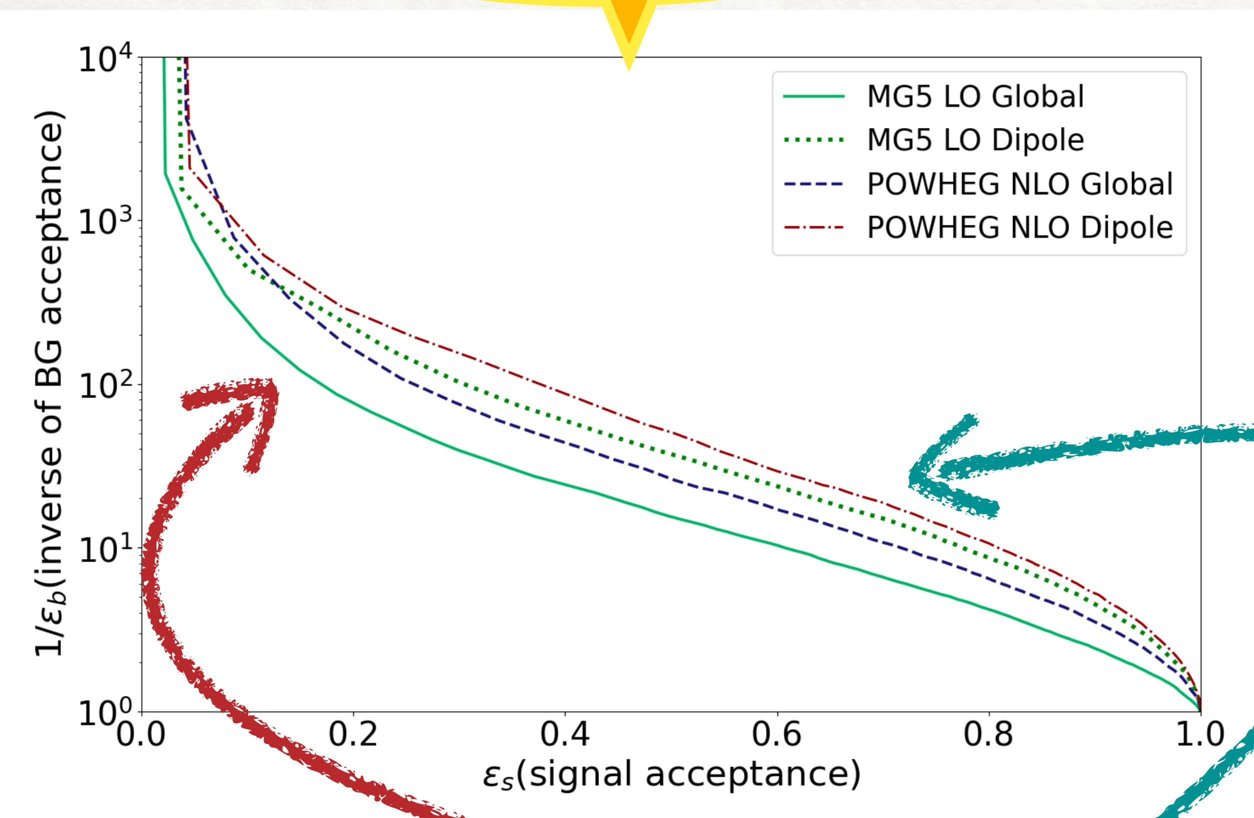
Vishal Ng, PK (2021)

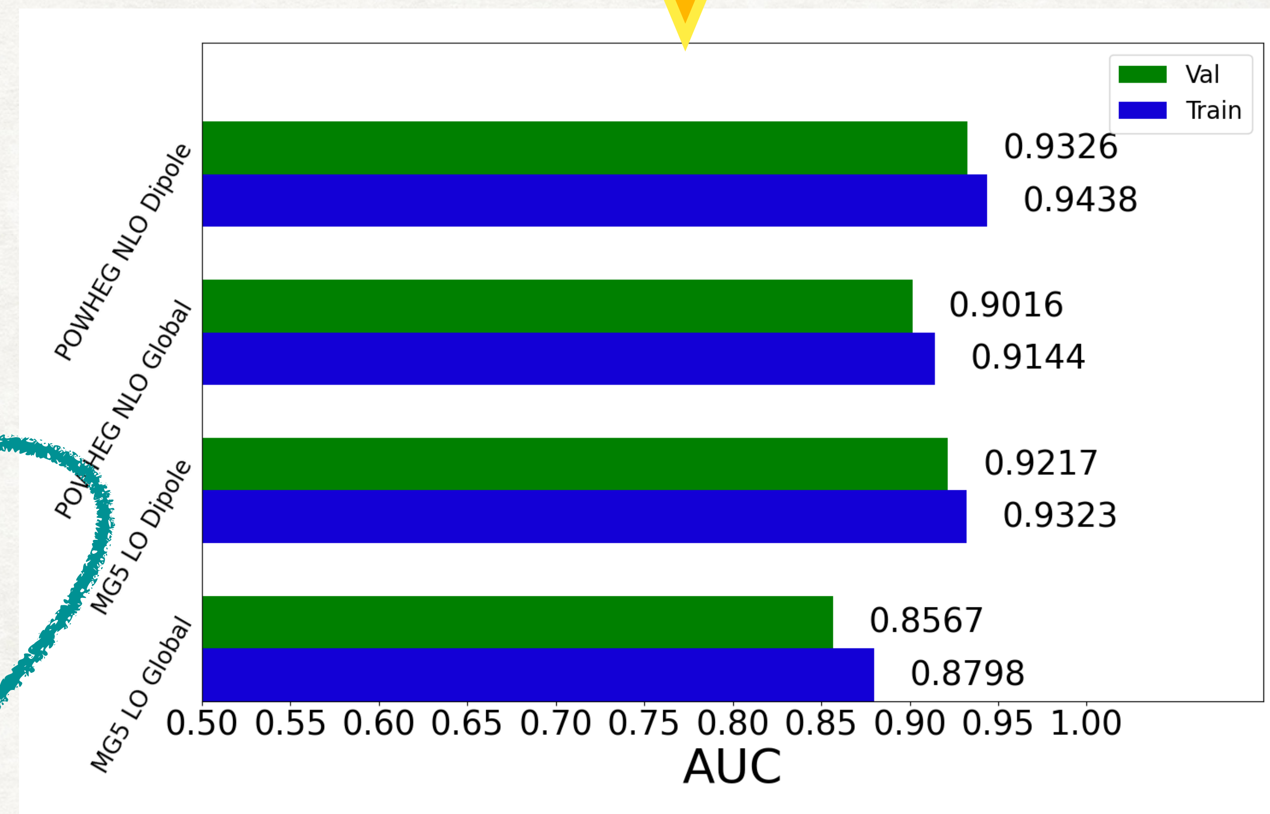
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

ROLE OF PARTON SHOWER

Receiver Operator Characteristics (ROC) Curve

AUC



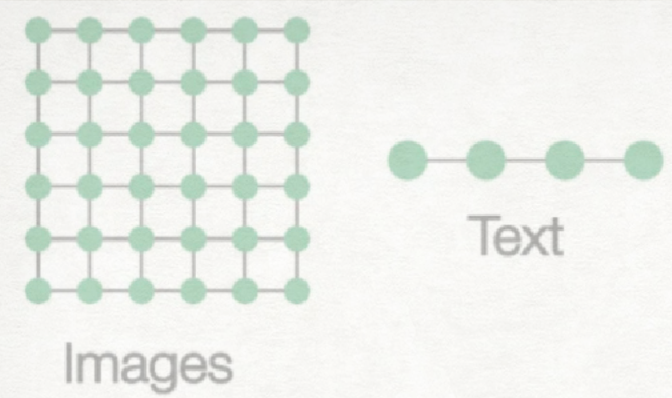


- ✓ LO + Global parton shower scheme shows lowest performance
- ✓ NLO + Dipole parton shower scheme shows best performance
- ✓ Rest two (LO+ Dipole & NLO+ Global) shows intermediate performance

Accurate description

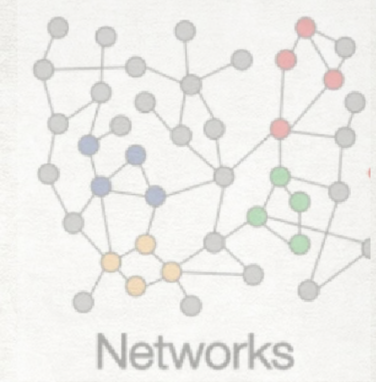
Vishal Ng, PK (2021)

Common in analysis



BEYOND CNN

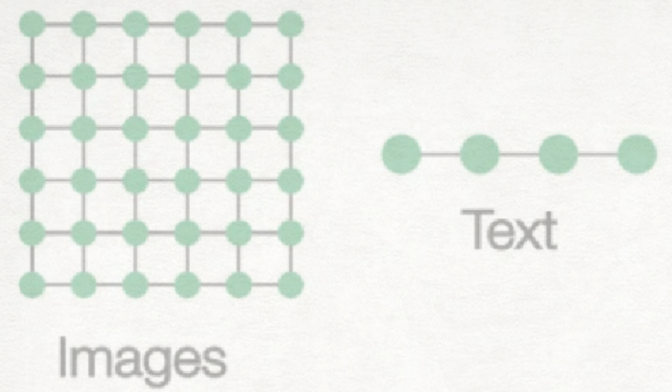
GRAPH NEURAL NETWORK



- ➡ Detectors calorimeter hits are typically very sparse and unstructured
- ➡ Varying number of reconstructed constituents
- ➡ Large number of tunable parameters

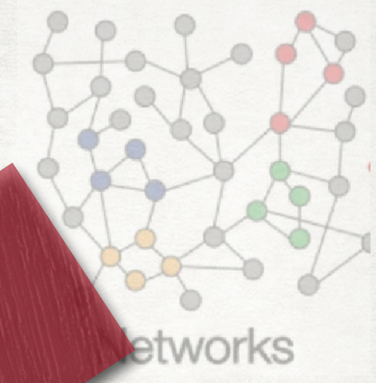
✓ Euclidean image (CNN) => general non-Euclidean domain (GNN) : Geometric deep learning

- Graph: Event as point cloud with each entry containing a vector composed of observables
- Graph == Nodes (data point) + Edges (connections are as important as the data itself)
- Message passing operation: nodes features and edge features are exchanged and provide a sophisticated feature extraction
- GNN is very powerful recent concept - mostly unexplored!!



BEYOND CNN

GRAPH NEURAL NETWORK



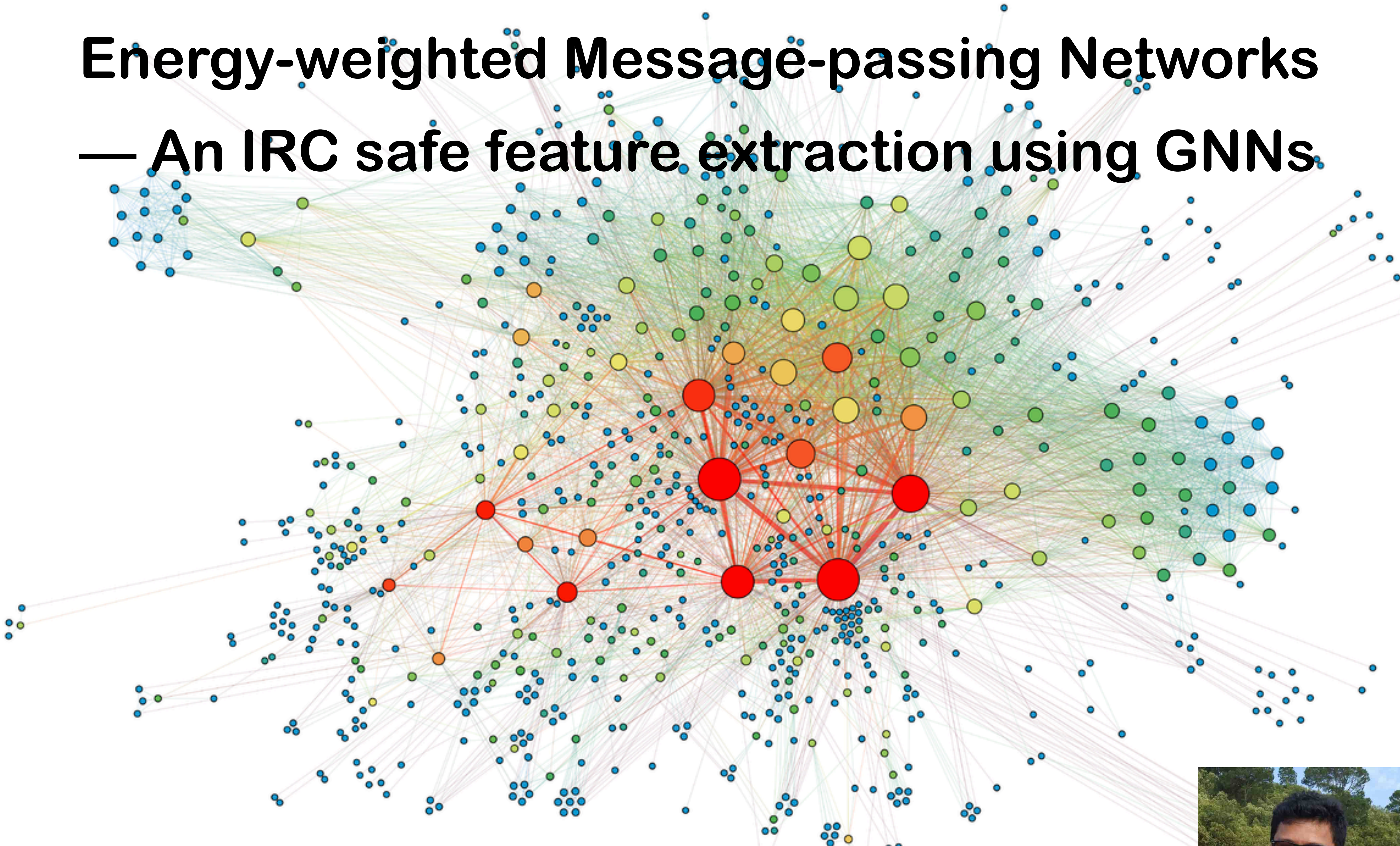
- ➡ Detectors calorimeter hits are typically very sparse and unstructured
- ➡ Varying number of reconstructed constituents
- ➡ Large number of tunable parameters

✓ Euclidean image (CNN) => general non-Euclidean domain (GNN) : Geometric deep learning

- Graph: Event as point cloud with each entry containing a vector composed of observables
- Graph == Nodes (data point) + Edges (connections are as important as the data itself)
- Message passing operation: nodes features and edge features are exchanged and provide a sophisticated feature extraction
- GNN is very powerful recent concept - mostly unexplored!!

Heterogeneous data with variable length & Inherently non-Euclidean
=> Point clouds provide a flexible geometric representation

Energy-weighted Message-passing Networks
— An IRC safe feature extraction using GNNs



Based on: arxiv: 2109.14636 [JHEP 02 (2022) 060]
Cited in newly introduced AI chapter - PDG’22
arxiv: 2309.17351 [JHEP xx (2023) xxx]
(Partha Konar, Vishal Ng, Michael Spannowsky)

INFRA-RED AND COLLINEAR (IRC)

Any QCD jet observable should be
— sensitive to the physics you want to probe
— calculable from first principles in Quantum Chromodynamics (QCD)

- Kinoshita-Lee-Nauenberg (KLM) theorem: Divergences exactly cancel between the real and virtual contributions to the observable at each perturbative order when the soft and collinear regions of phase space are inclusively summed over.
- IRC safety ensures that the phase space restrictions that the measured value of an observable imposes do not disrupt this cancellation [Sterman and Weinberg]
- IRC safe Jet mass & thrust observable [early beginning of jet sub-structure]
- Catani etal (CTTW) large log resummed jet substructure observable

INFRA-RED AND COLLINEAR (IRC)

- High-energy partons lead to collimated bunches of hadrons
- jet definition: project from large no of hadrons => few parton-like objects
- Provide link between experimental observables and the theoretical construction
- Def of jet must be invariant with respect to certain modifications of the event

  -> collinear splitting -> infrared emission

- Effort went into constructing IRC safe jet : Sequential recombination in KT, Anti-KT

How can we make neural networks aware of this physics input?
So that, it treats all hadronic/jet analysis in a IRC safe way.

POINT CLOUD

.

Set of points sampled from an underlying space (**not necessarily Euclidean**)

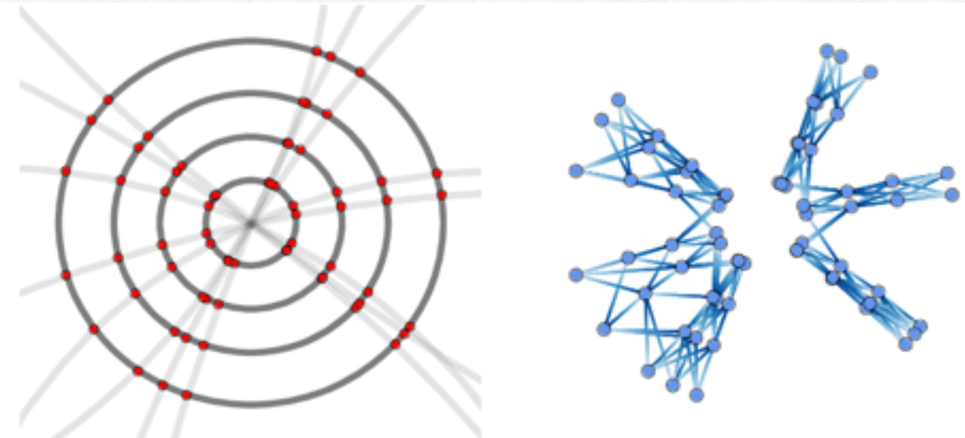
Each data sample is a set with variable cardinality:

$$\mathcal{S}_\alpha = \{p_1, p_2, \ldots\ldots, p_{n_\alpha}\}$$

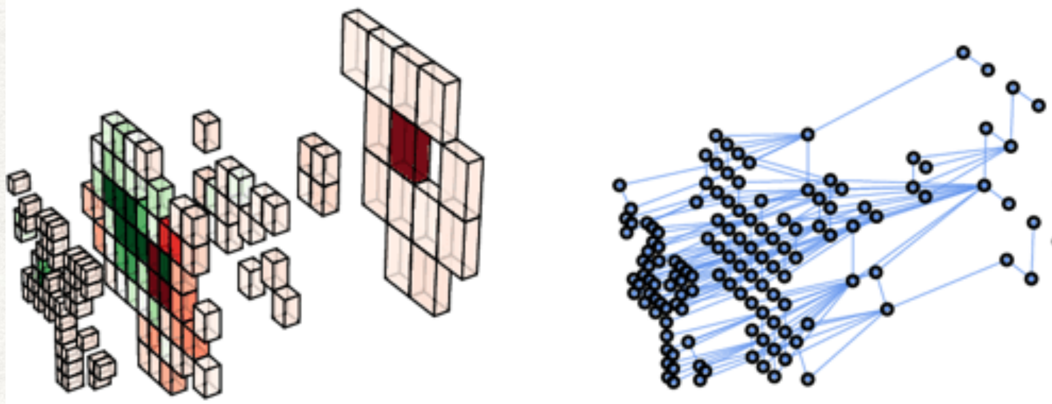
Can also be a collection of sets:

$$\mathcal{S}_\alpha^{all} = \{\mathcal{S}_\alpha^{jets}, \mathcal{S}_\alpha^{leptons}, \mathcal{S}_\alpha^{photon}, \ldots\}$$

$$\alpha = \text{ Event index}$$



(a)



(b)

CONSTRUCTION OF GRAPH

LEARNING HOW DIFFERENT POINTS RELATE

A graph $G(\mathcal{S}, \mathcal{E})$ defined on a set $\mathcal{S}$, with edge-set $\mathcal{E}$

$G(\mathcal{S}, \mathcal{E})$ $\quad$ $\mathcal{S} = \{a, b, c\}$

Node features: $\{\mathbf{h}_a, \mathbf{h}_b, \mathbf{h}_c\}$

$$\mathcal{E} = \{(b,a),(c,a),(a,b),(c,b),(b,c),(a,c)\}$$

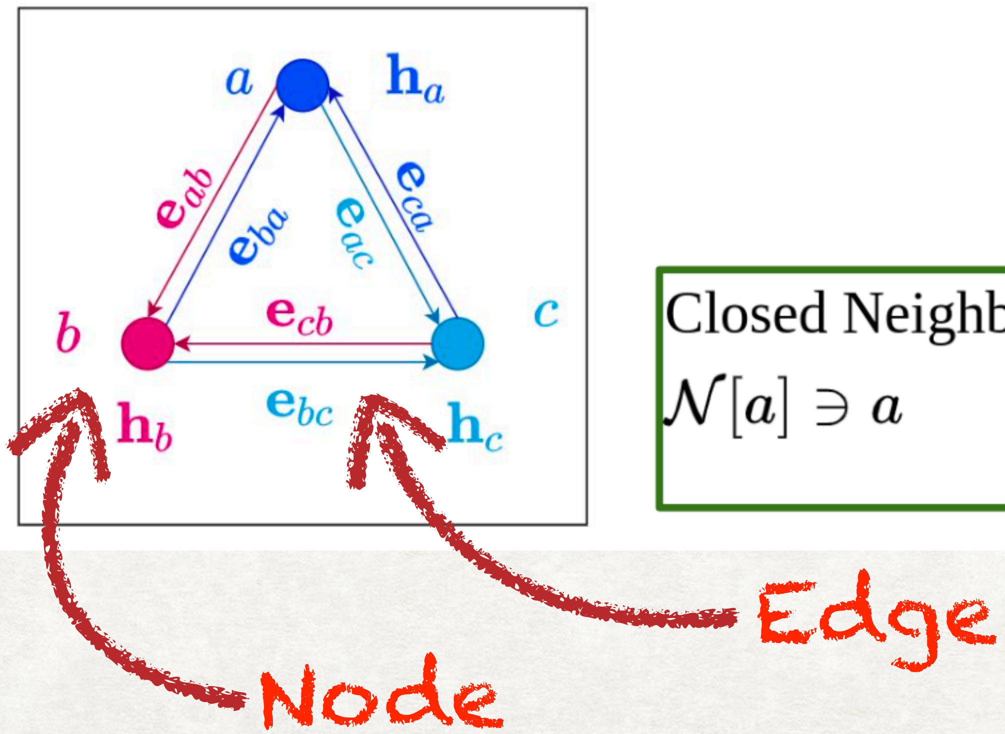
Edge features: $\{\mathbf{e}_{ba}, \mathbf{e}_{ca}, \mathbf{e}_{ab}, \dots\}$

Neighbourhood sets:

$\mathcal{N}(a) = \{b, c\}$

$\mathcal{N}(b) = \{a, c\}$

$\mathcal{N}(c) = \{c, a\}$



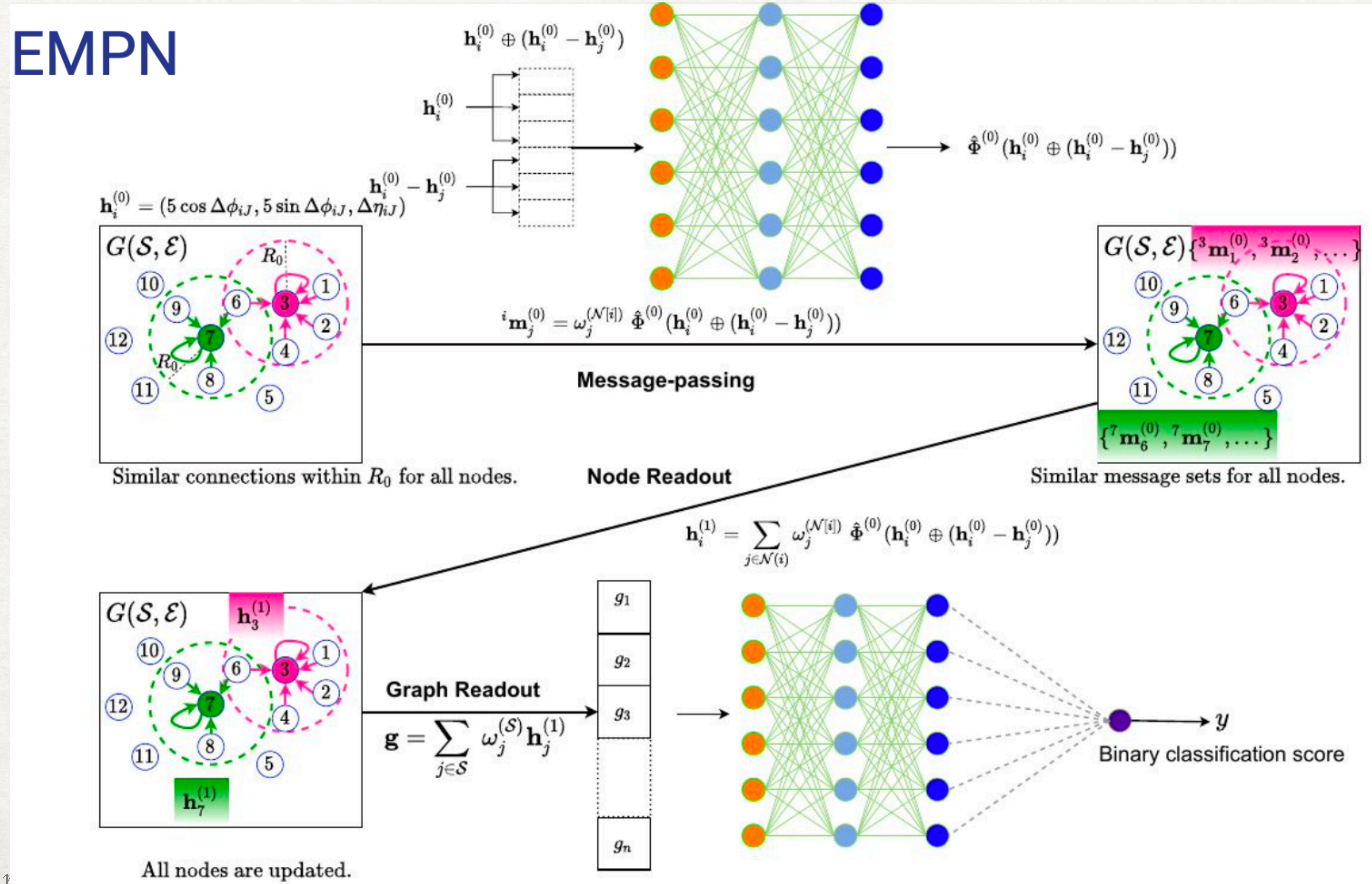
Closed Neighbourhood sets for IRC safety:

$\mathcal{N}[a] \ni a$

7

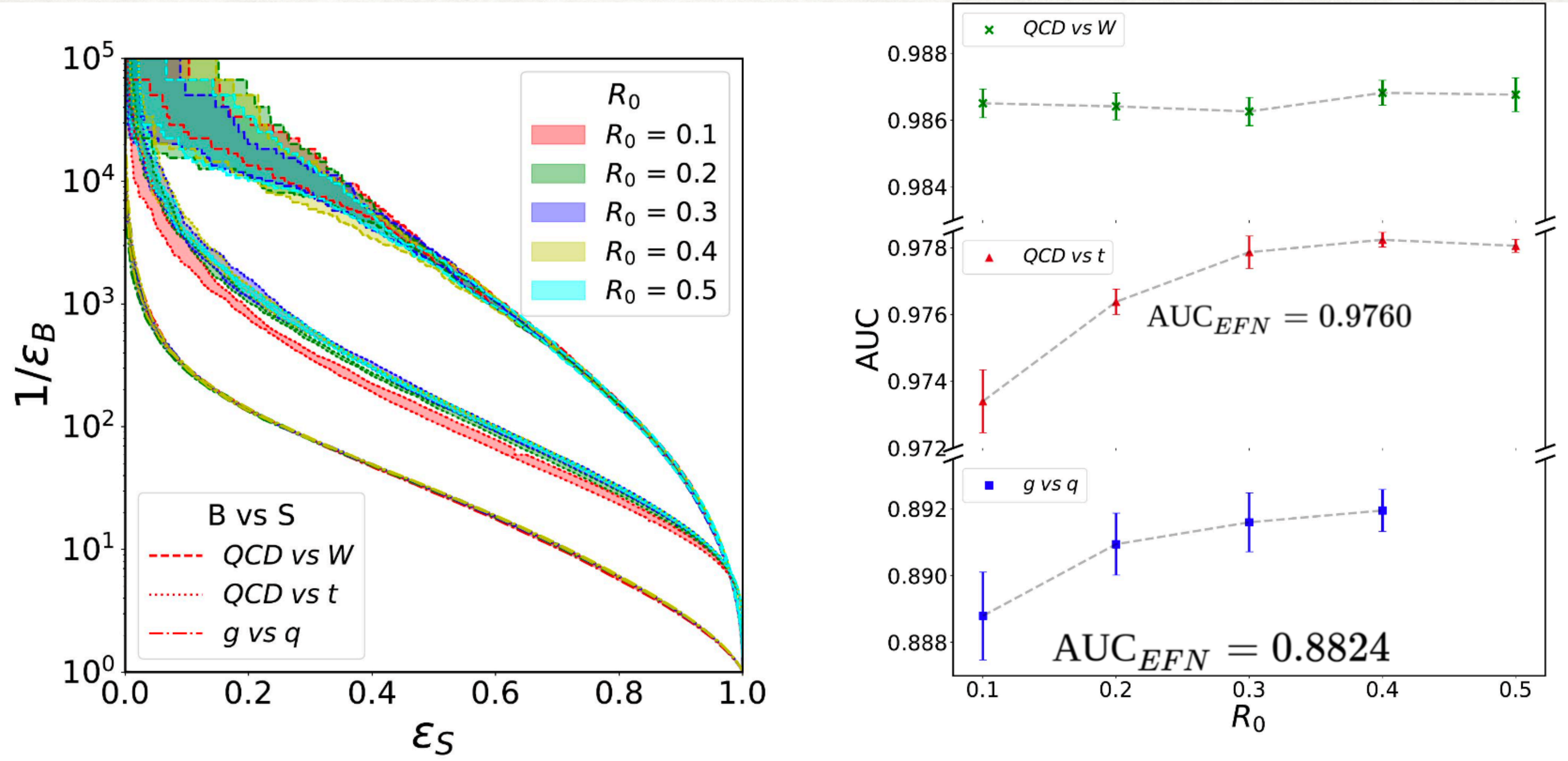
ENERGY-WEIGHTED MESSAGE-PASSING

IRC SAFE WAY



NETWORK PERFORMANCE

EMPN - IRC SAFE WAY

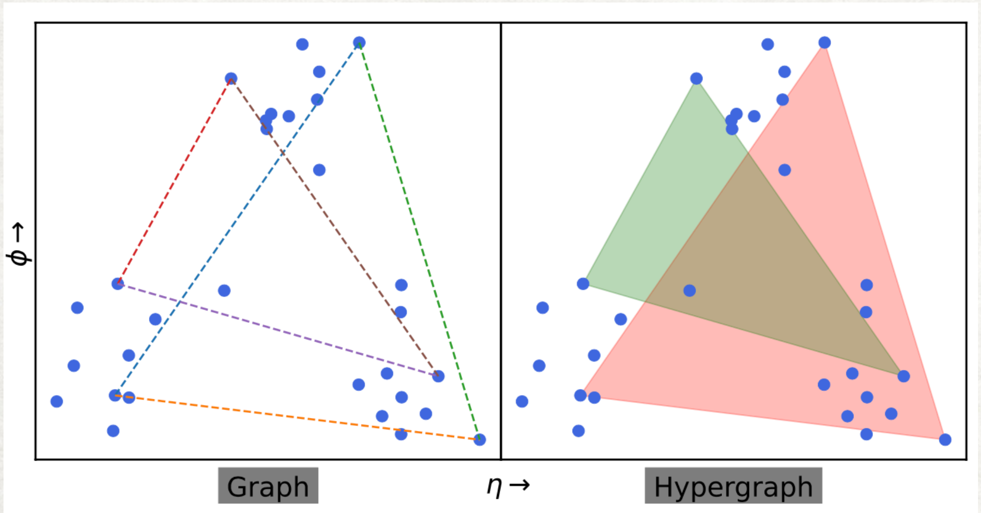


PK, Vishal Ng, Michael Spannowsky; 2022

TOWARDS IRC SAFE H-EMPN

HYPER - GRAPH NEURAL NETWORK

- Extracting features from any N-point correlation
- Construct IRC safe higher-point correlations
- Hypergraph Energy-weighted Message Passing Networks (H-EMPNs) - designed to capture any N -point correlation among particles
- Order-three hyperedges simultaneously link properties of three jet constituents at a time
- Access higher-order correlations amongst jet constituents



PK, Vishal Ng, Michael Spannowsky; 2023

MACHINE LEARNING

CHALLENGES

- Interpretability: Relevant physics knowledge learned by the model
  : Physics intuitive high-level features capture real insights,
  but clearly sacrifice some useful information
- Prejudice: Decades-old research by human mind must be supreme
  (After all, NN tried to mimic the neurones??)
- Status quo: are "we" and "journals" evolving slowly to catch up!

- In research: Dealing with different kinds of abstract data
- Overreach: Is not effective in all kinds of problems!

- Involved cost:
  — Data science skill development + domain knowledge expertise
  — Order of magnitude higher computation power requirement
  — Opaque transition between knowledge & learning


Thank
you