Deep Learning Frontier
in
Theoretical High Energy Physics

Partha Konar
Physical Research Laboratory
Ahmedabad, India

ICHEPAP2023 @ SINP

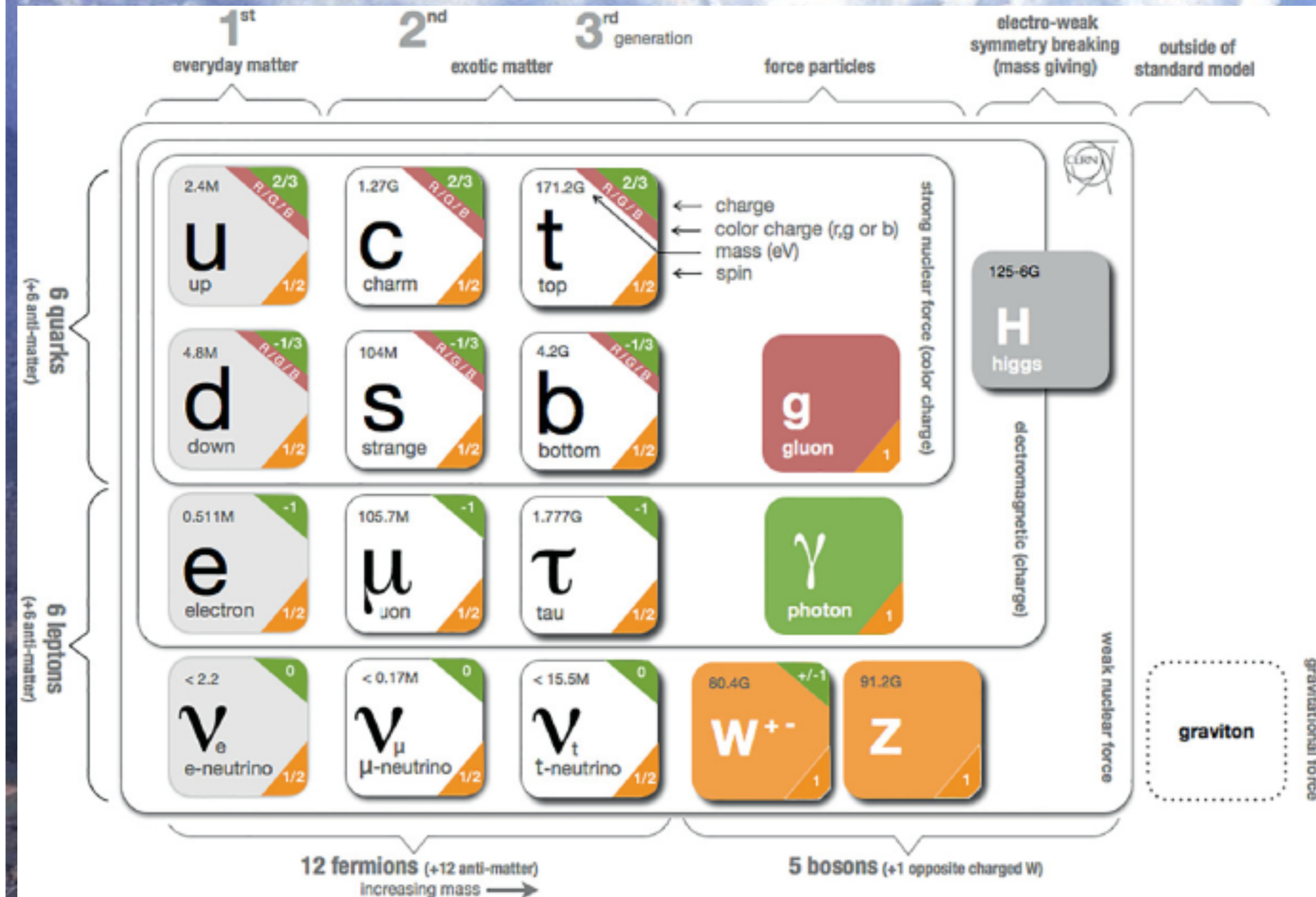
Dec 12, 2023

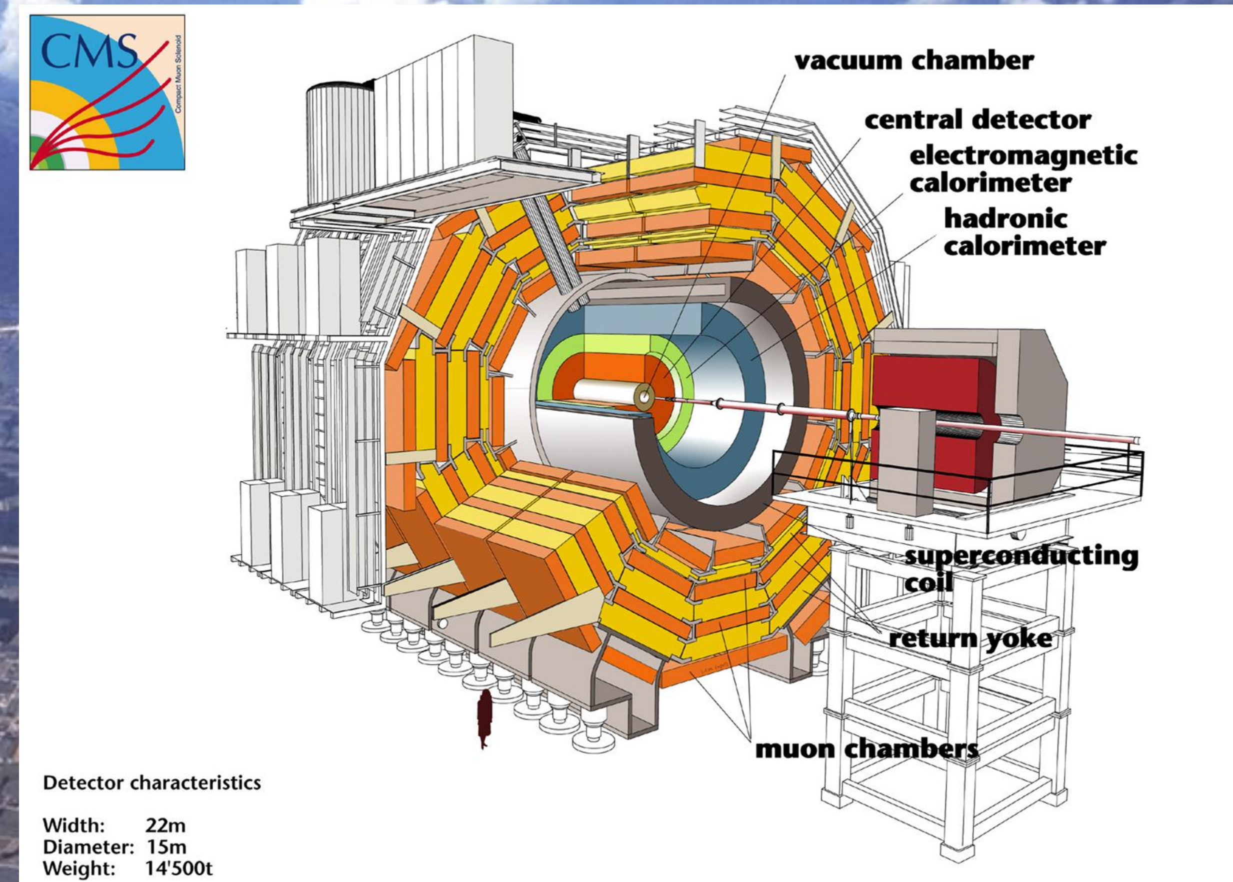
Deep Learning Frontier
in
Theoretical High Energy Physics

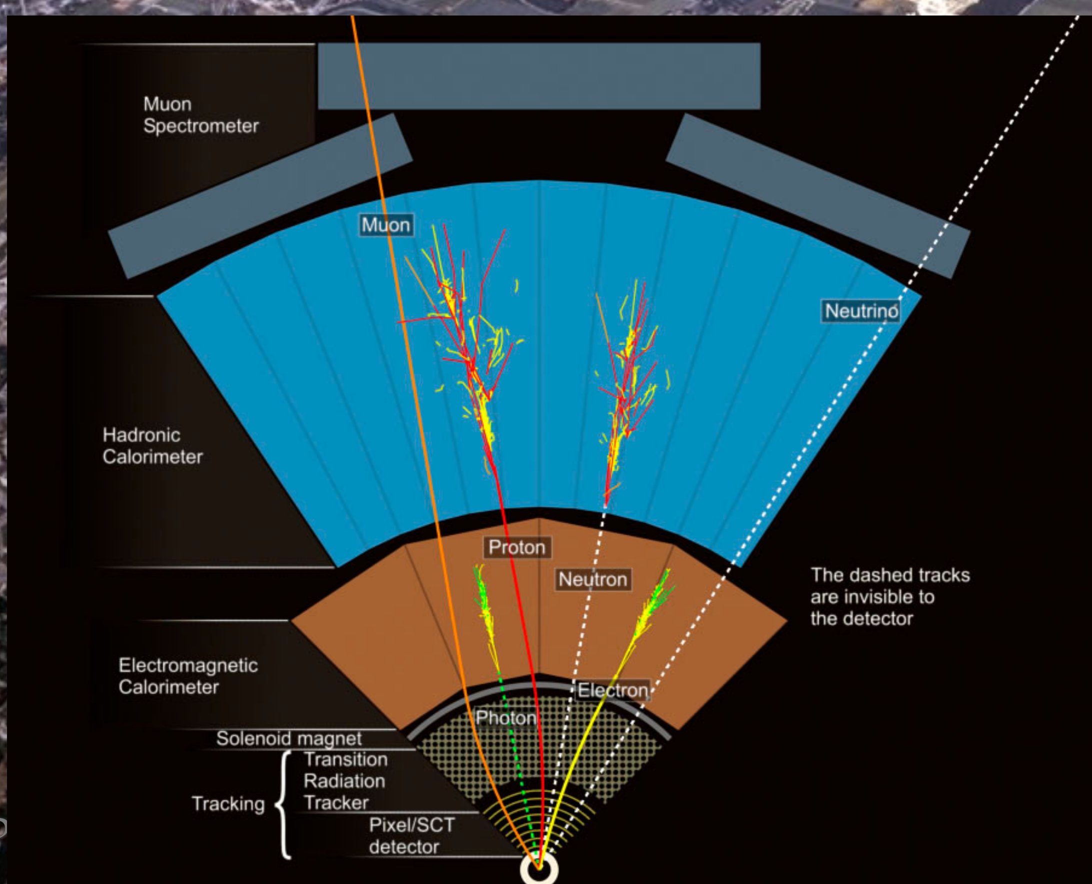
Partha Konar
Physical Research Laboratory
Ahmedabad, India

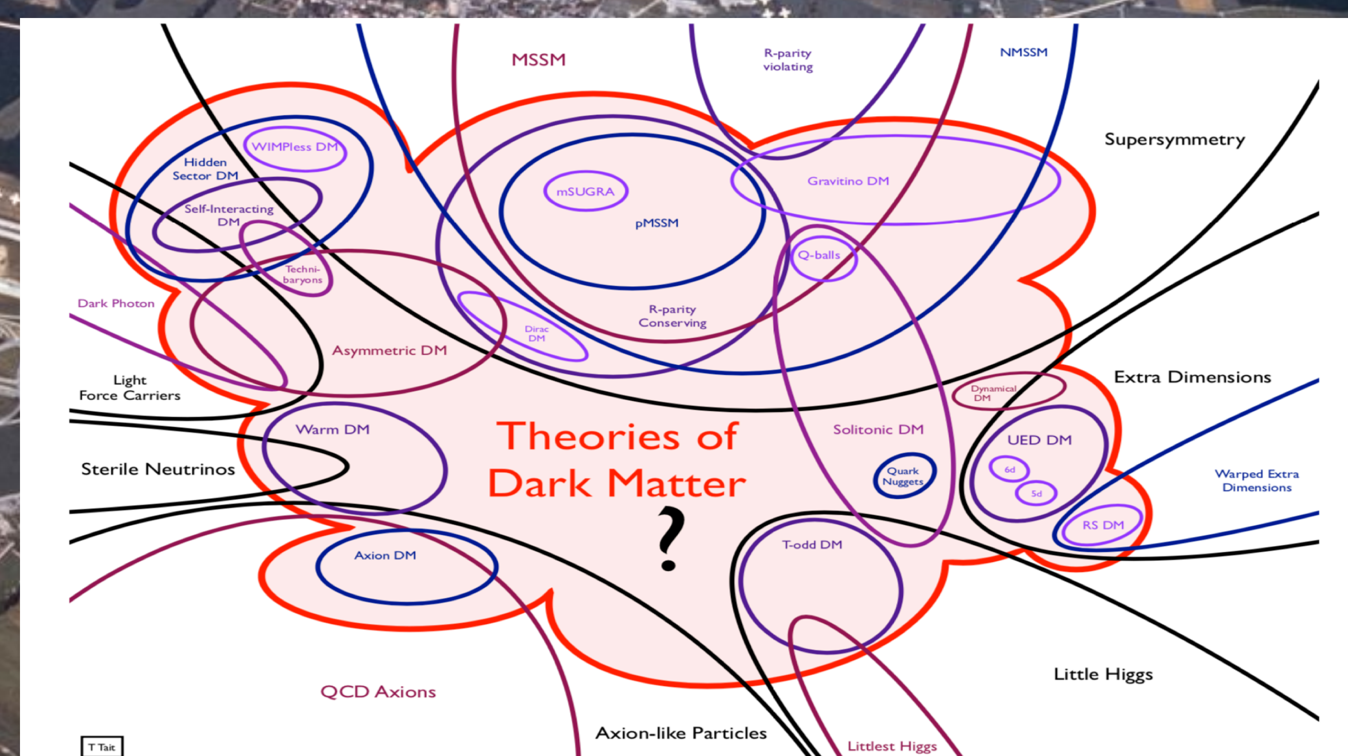
- **Sanmay Ganguly -:**
  Application of Machine Learning Techniques in HEP experiments
- **Daniel Nieto Castano -:**
  Applications of Machine Learning in Astroparticle Physics

SEARCH AT LARGE HADRON COLLIDER (LHC)









MACHINE LEARNING

RELEVANCE

- With around 40 mHz branch crossing LHC taking ~ 40 million snaps/s
- Each snapshot encounter large no of particles compounding ~10^8 sensors at different parts of detector
- ML takes role from low level reconstruction, identification, underlying event mitigation to high level identification, extraction, classification and anomaly detection
- Crucial roles in
  (i) Data reduction in real time - triggers
  (ii) Anomaly detection
  (iii) Fast accurate reconstruction, identification with multi-sensor data
  (iv) Improvements in classification, regression, statistical analysis

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

- Classification: Find faint signal against a large background
- Move into higher dimensional space —
  - ☑ Multivariate analysis with High Level Variables
  - ☑ Low Level Variables from detectors (number of dimensions very large)
- Find the Division Boundary in this higher dimensional space
  — Best possible [under-fitting?] but Trustworthy [over-fitting?] way
- Neural Networks based on interconnected nodes in layered structure
  — In analogy with brain neurones
  — Connects different input/ derived data
  — Involve free parameters (weight and bias) [inductive bias?]
  — Optimise “free parameters” using labeled data [Model]

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**

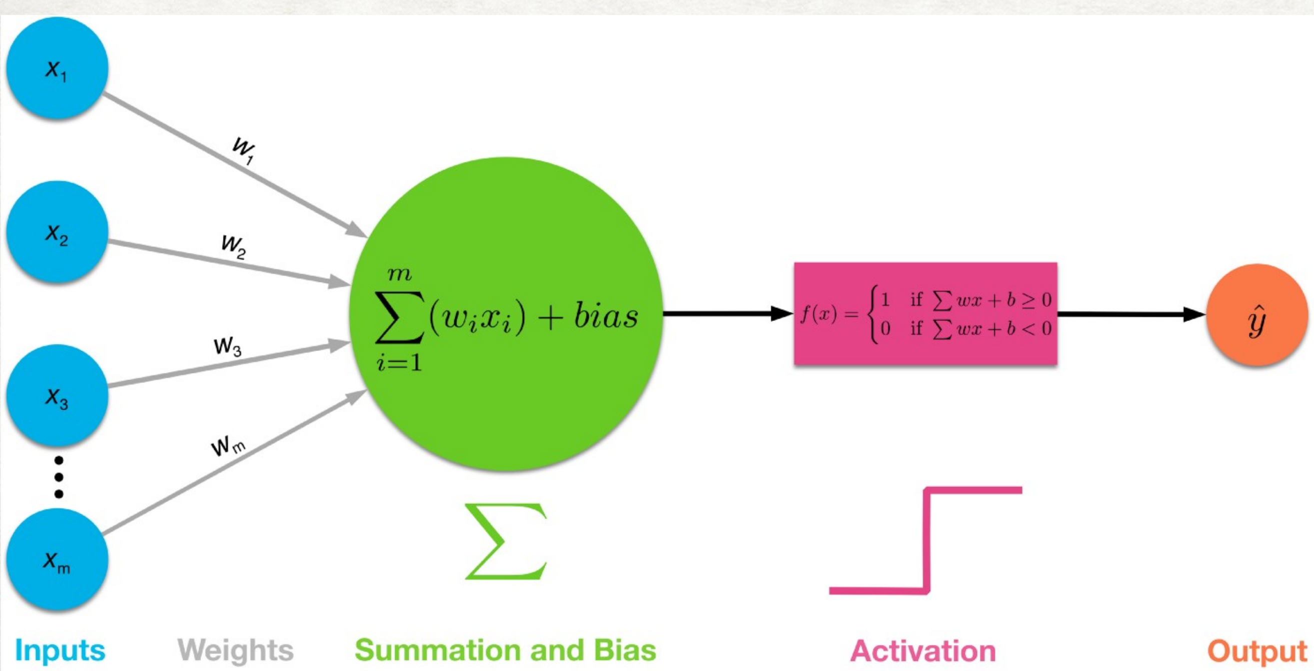
  — End for physics intuitive high level variables from domain experts?

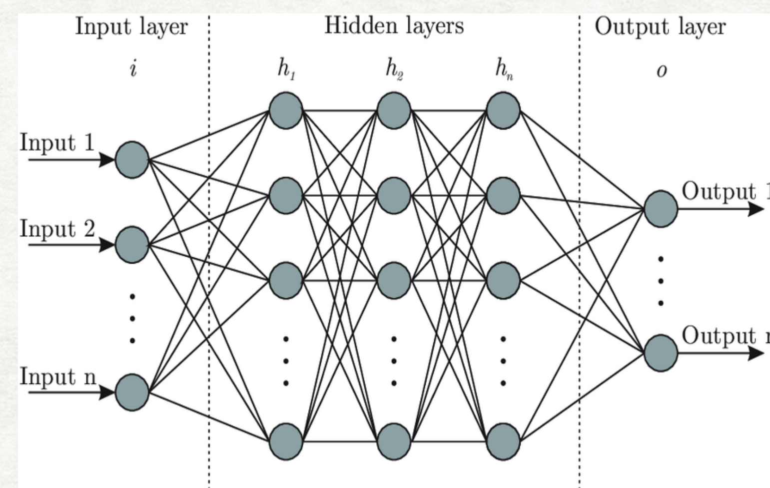
ARTIFICIAL NEURAL NETWORK

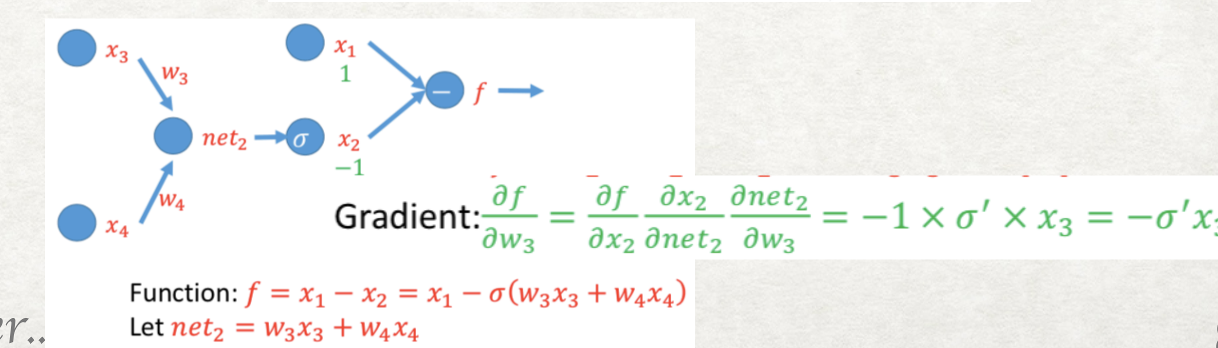
(ANN)

- **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.**







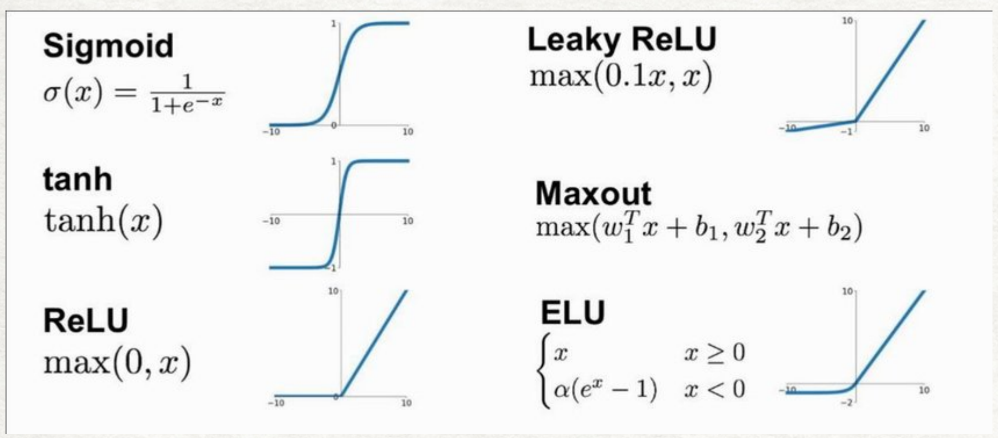
ARTIFICIAL NEURAL NETWORK

(ANN)

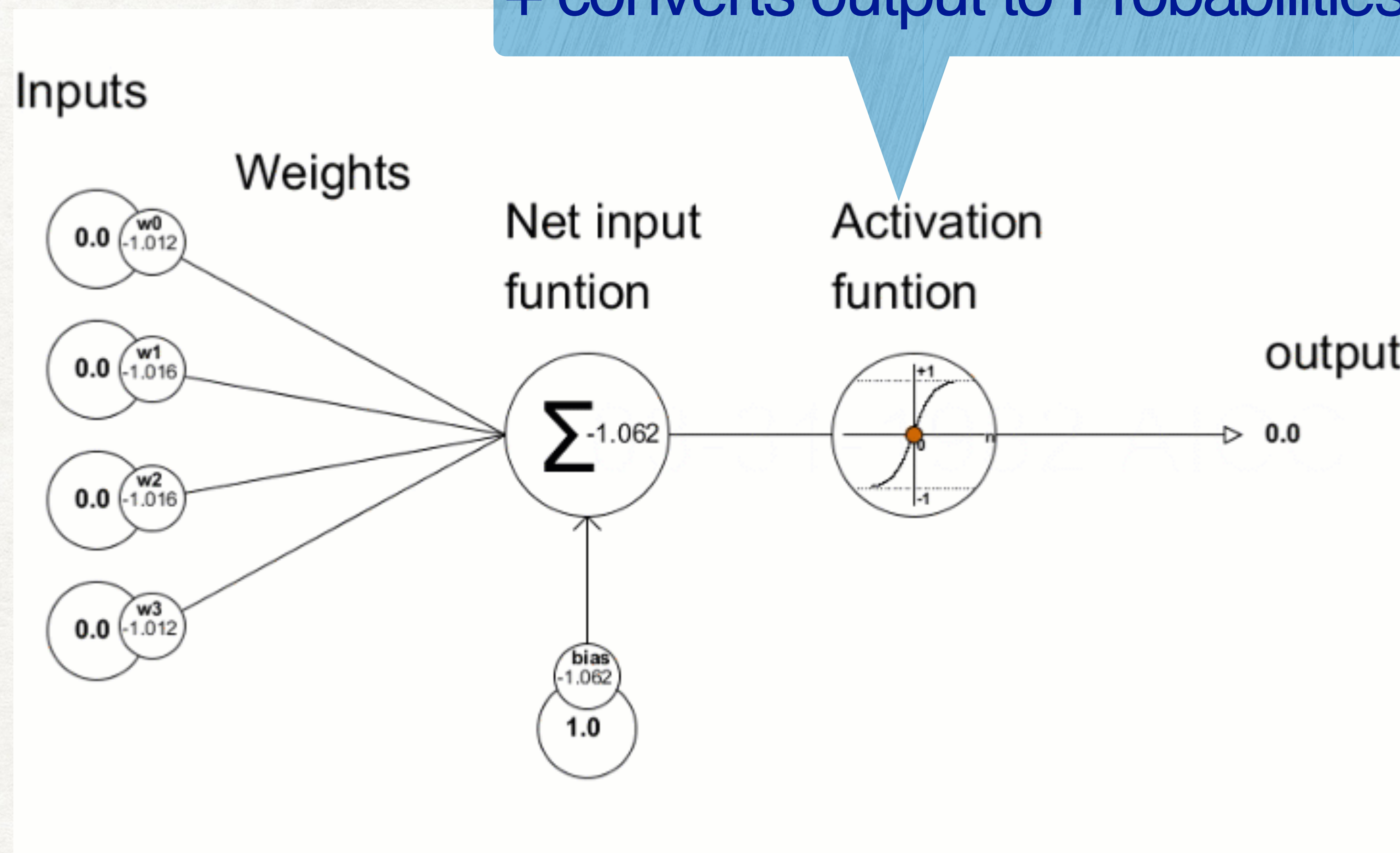
$$h_{i+1} = \sum_i w_i h_i + b_i = w^T h$$

Activation/response: $\sigma(w^T h)$

<ReLu, sigmoid, tanh..>



Introduce non-linearities
+ converts output to Probabilities



Loss/Cost fn : $\mathscr{L}[y - f_w(x)]$ [Mean Squared Error, Cross-Entropy/logarithmic loss>]

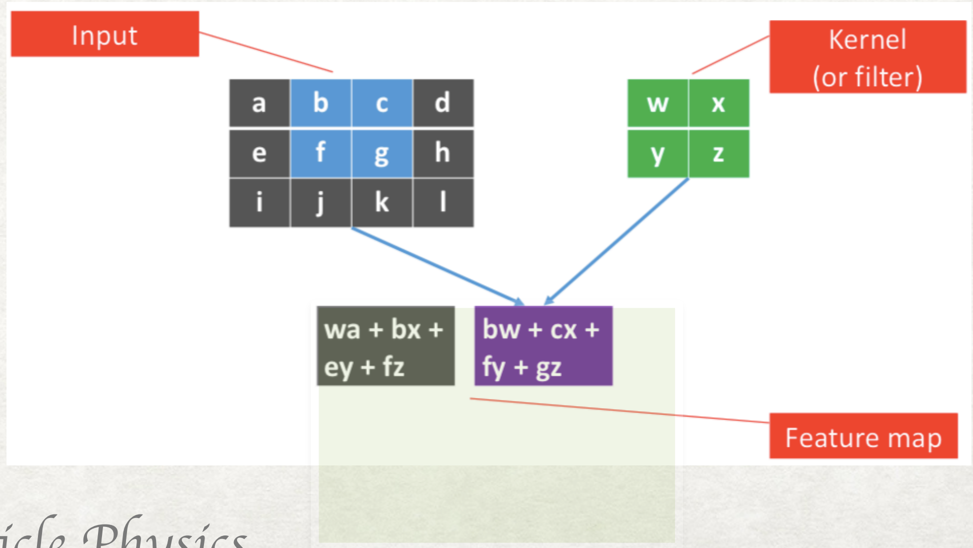
so that [gradient descent] $\nabla_w \mathscr{L}[y - f_w(x)] \to 0$

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**

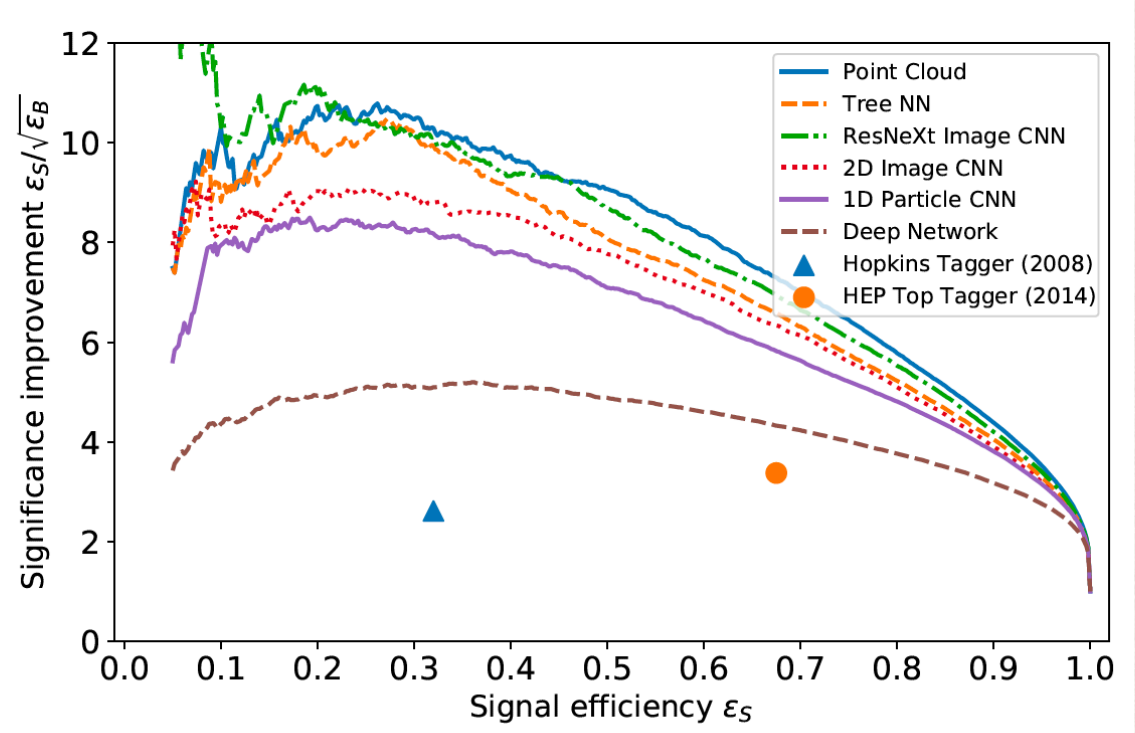


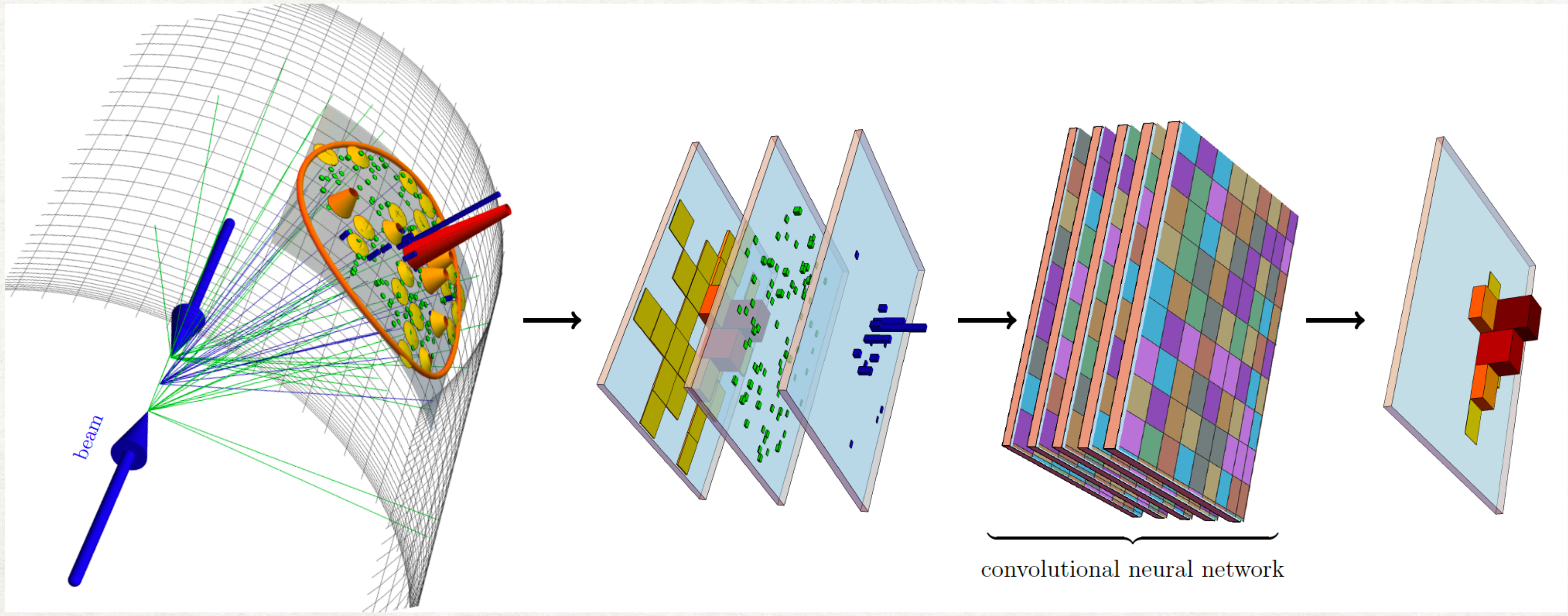
CONVOLUTIONAL NEURAL NETWORK

WORKING PRINCIPLE AT LHC DATA

- Detectors calorimeter tower => pixels of an image

- Powerful image classification network proved to be extremely successful in jet-substructure studies





Review :1806.11484, 2103.12226

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*

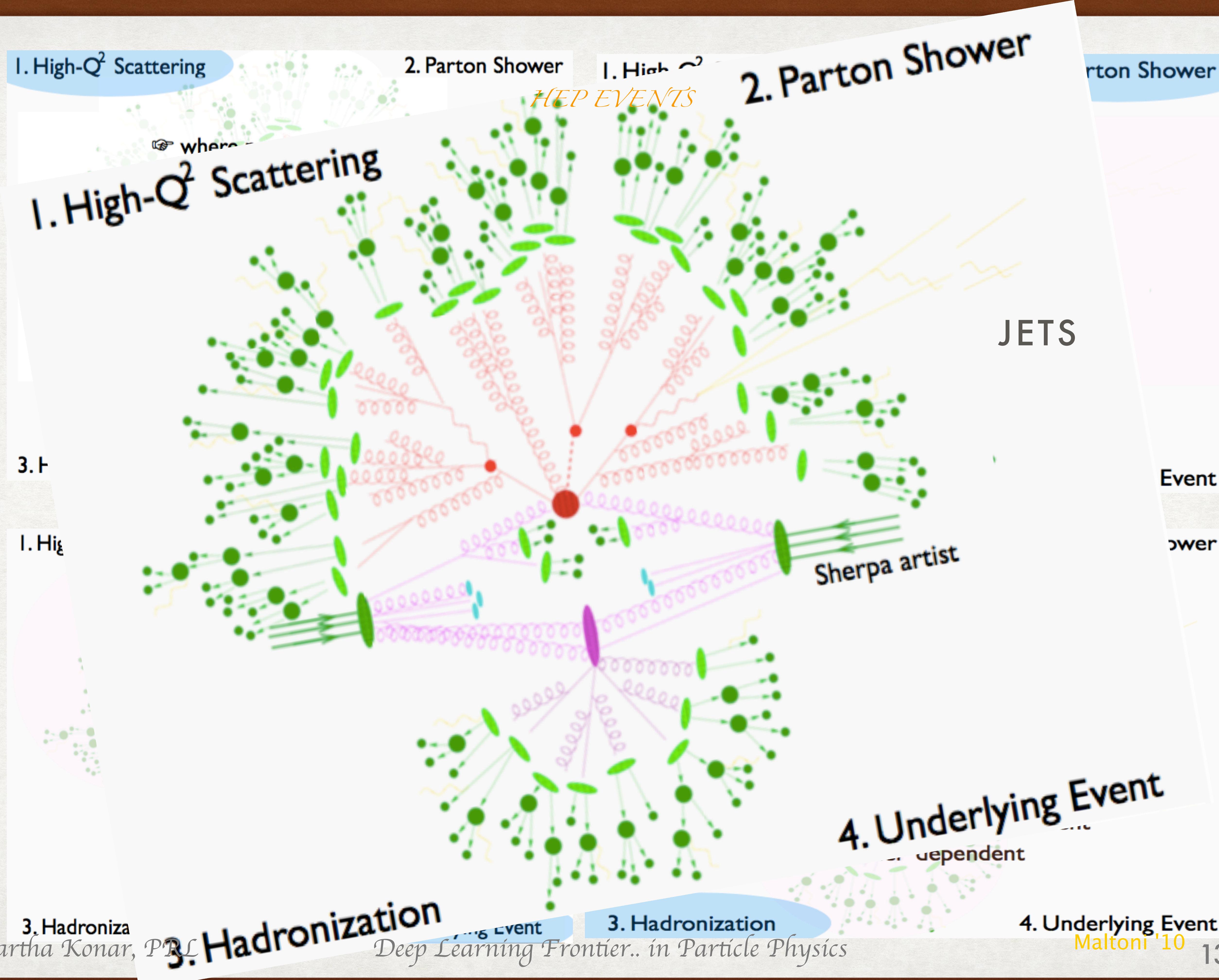
Generative models

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

Anomaly detection

HEP ML Living Reviews

HEP EVENTS



JETS

REPRESENTATION OF DATA

- 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

★ Set of one-dimensional physically-motivated observables [e.g. Gallicchio, Schwartz 2011]

✦ Jets as images - pixelated grayscale image. intensity ~ energy (or PT) of all particles that deposited energy in a particular location [e.g 1603.09349, 1407.5675]

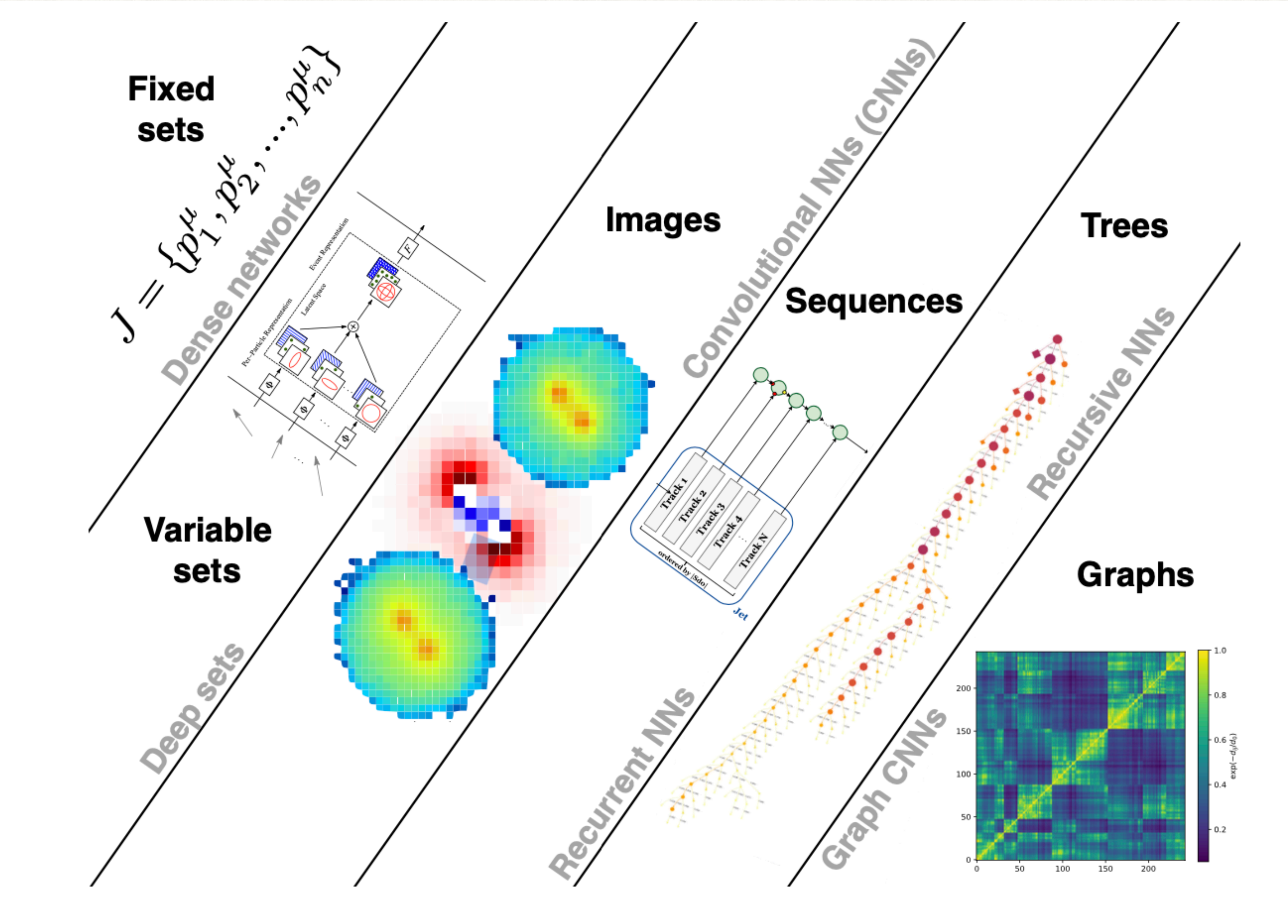
✦ Include additional layers ('RGB') to encode more information such as charge-energy versus neutral-energy

✳ jet clustering history as an image that mimics the QCD splitting function [Lund Jet Plane - 2018]

- Constituents as a sequence - clustering history as input [e.g. 1702.00748, 1711.02633]
- jet as a graph - nodes and edges in point cloud : node property & connection strength between the various nodes of the graph. [Deep sets:1810.05165]

JET REPRESENTATION

JET DATA - IMAGES, SEQUENCES AND SETS



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

- **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**

INFRA-RED AND COLLINEAR (IRC)

SAFE OBSERVABLES

Set of hard jets in a event should remain unchanged

— under a collinear splitting or addition of soft emission

For an observable $\mathcal{O}_n$ defined on n particles.

$$\mathcal{O}_{n+1}(p_a, \cdot, p_b, p_r, p_s, p_c, \cdot) \to \mathcal{O}_n(p_a, \cdot, p_b, p_q, p_c, \cdot)$$

In the infra-red $(z_r \to 0 \text{ or } z_s \to 0)$ or collinear limits $(\Delta_{rs} \to 0)$

Calculable in pQCD!!

For a splitting: $q \to r + s$

$$p_q = p_r + p_s$$

$$p_q = (z_q, \hat{p_q})$$

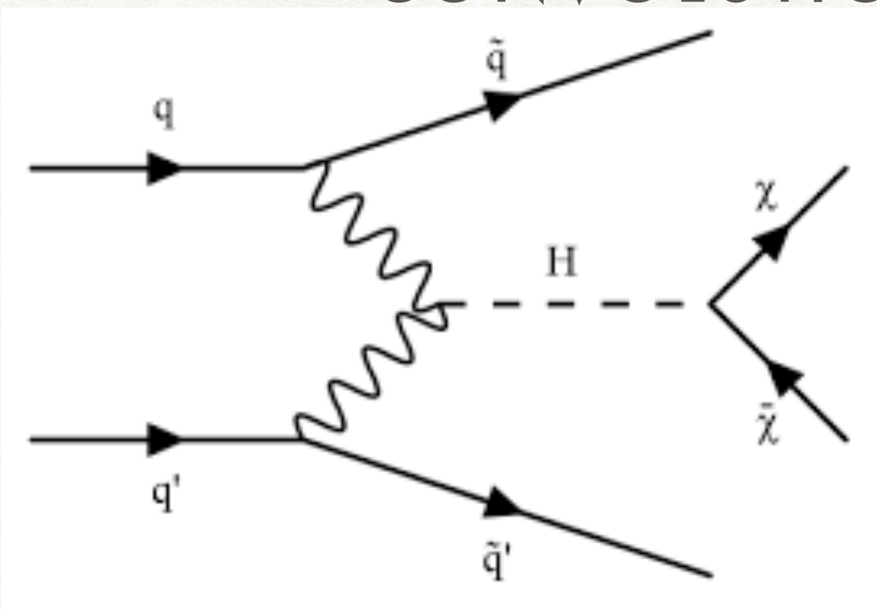
$$p_r = (z_r, \hat{p_r})$$

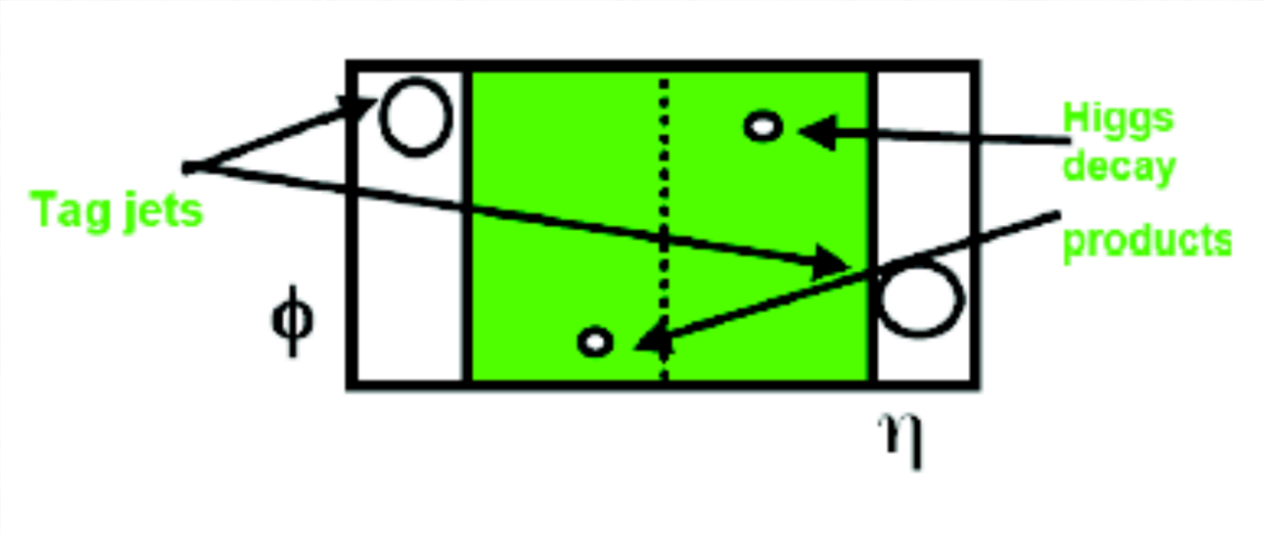
$$p_s = (z_s, \hat{p_s})$$

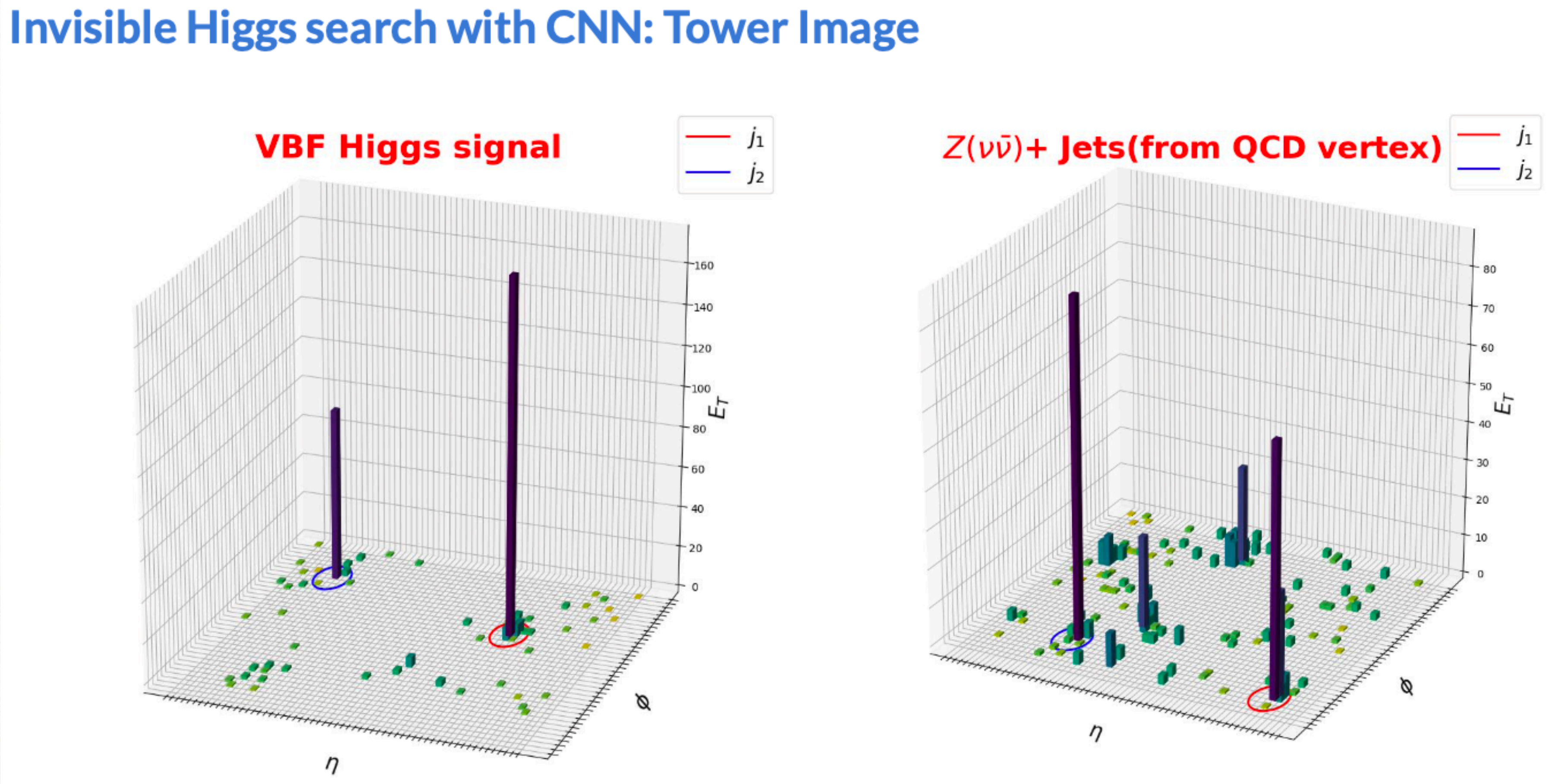
How can we make neural networks aware of this physics input?
So that, it treats all hadronic/jet analysis in a IRC safe way.

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

CONVOLUTIONAL NEURAL NETWORK





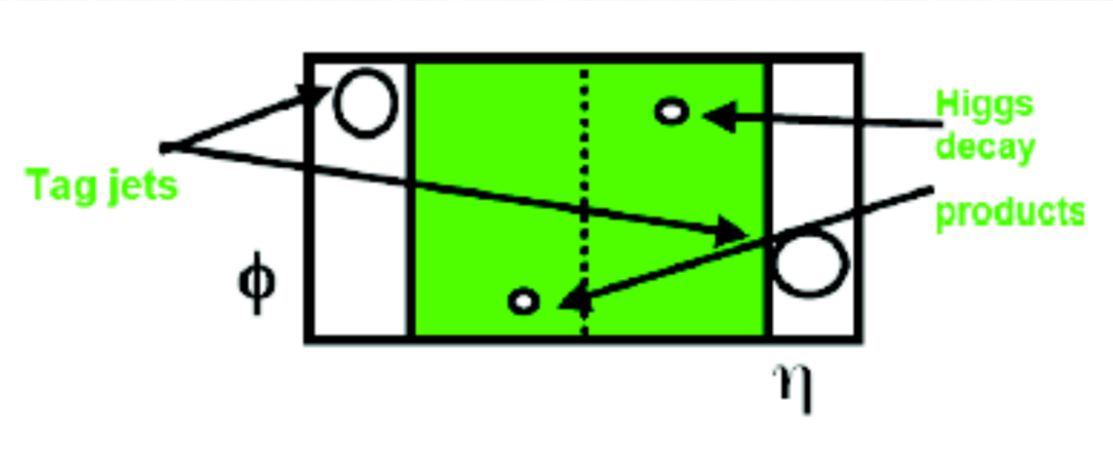


INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

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 — No additional features from decay product!

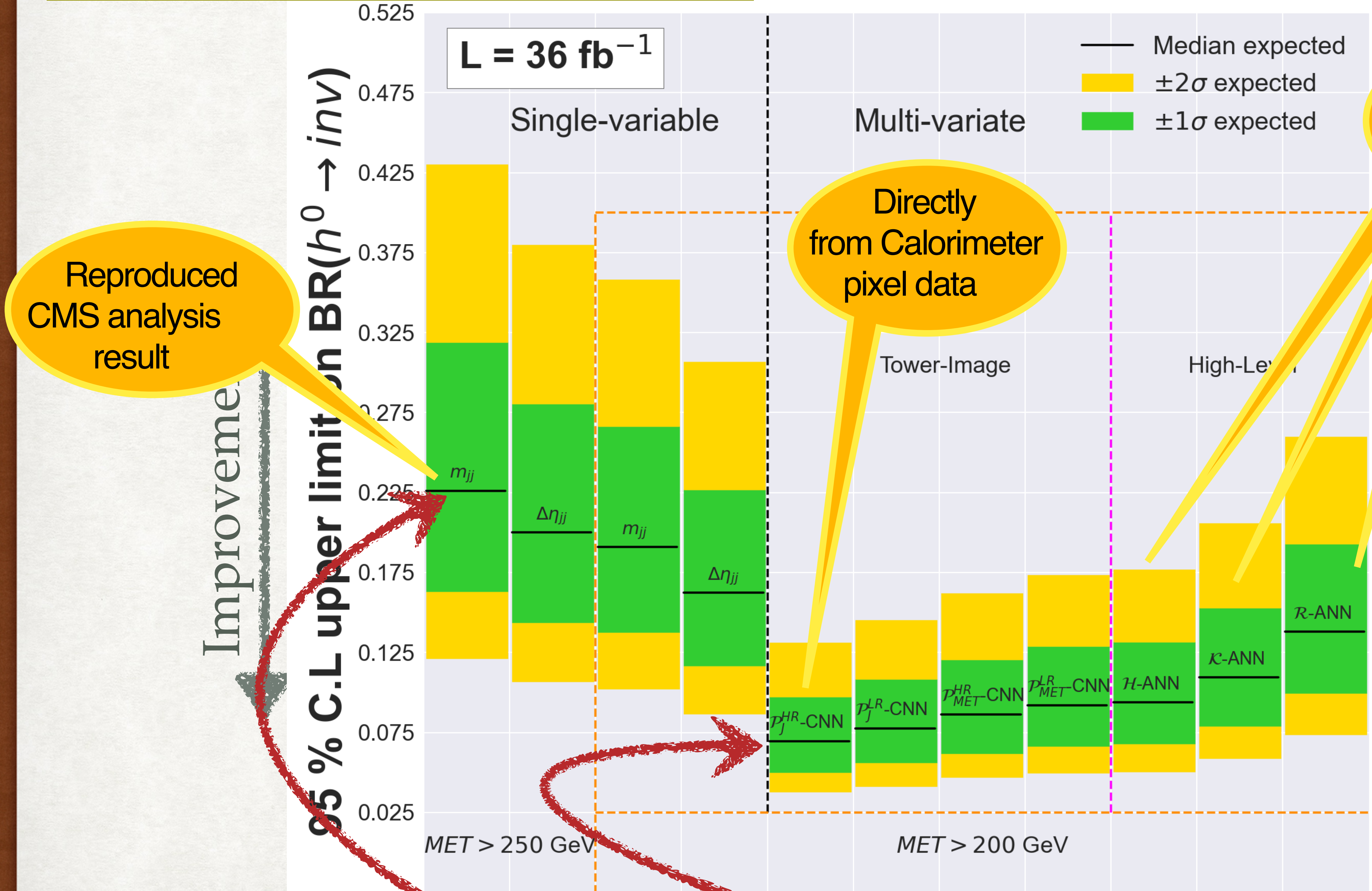
- Let us try that!

VBF is most sensitive channel for invisible Higgs search

Collider bounds on invisible branching ratio of Higgs much higher than SM prediction!!

INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

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



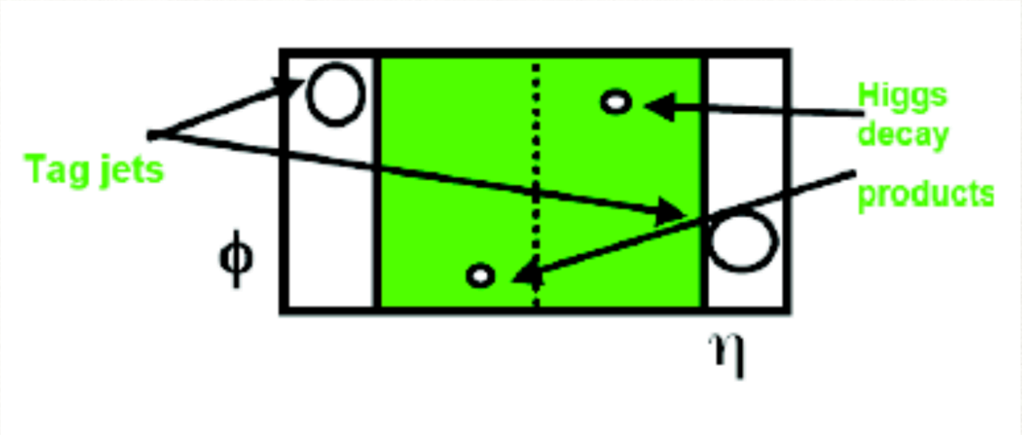
Akanksha Bhardwaj, PK, Aruna Nayak, Vishal Ng; 2020
PK, Vishal Ng; 2022

Factor of three improvement using the same data!
Hours of CNN training just extracted the relevant underlying feature better than our decades of research!

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

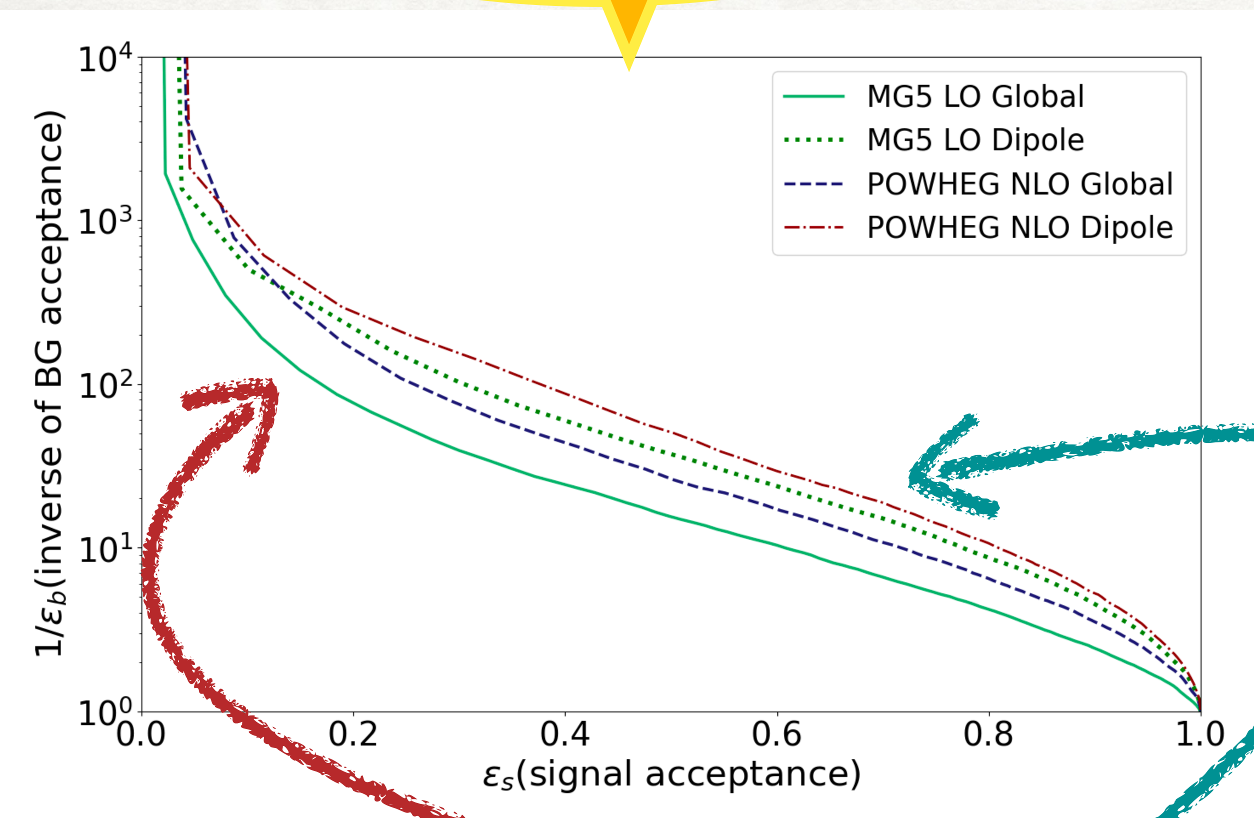
- *Parton Shower recoil Scheme [*Dipole parton shower*]*
  => Wrong global scheme (for spacelike shower) used in most analysis

★True potential unfolds if theoretical predictions are accurate enough.

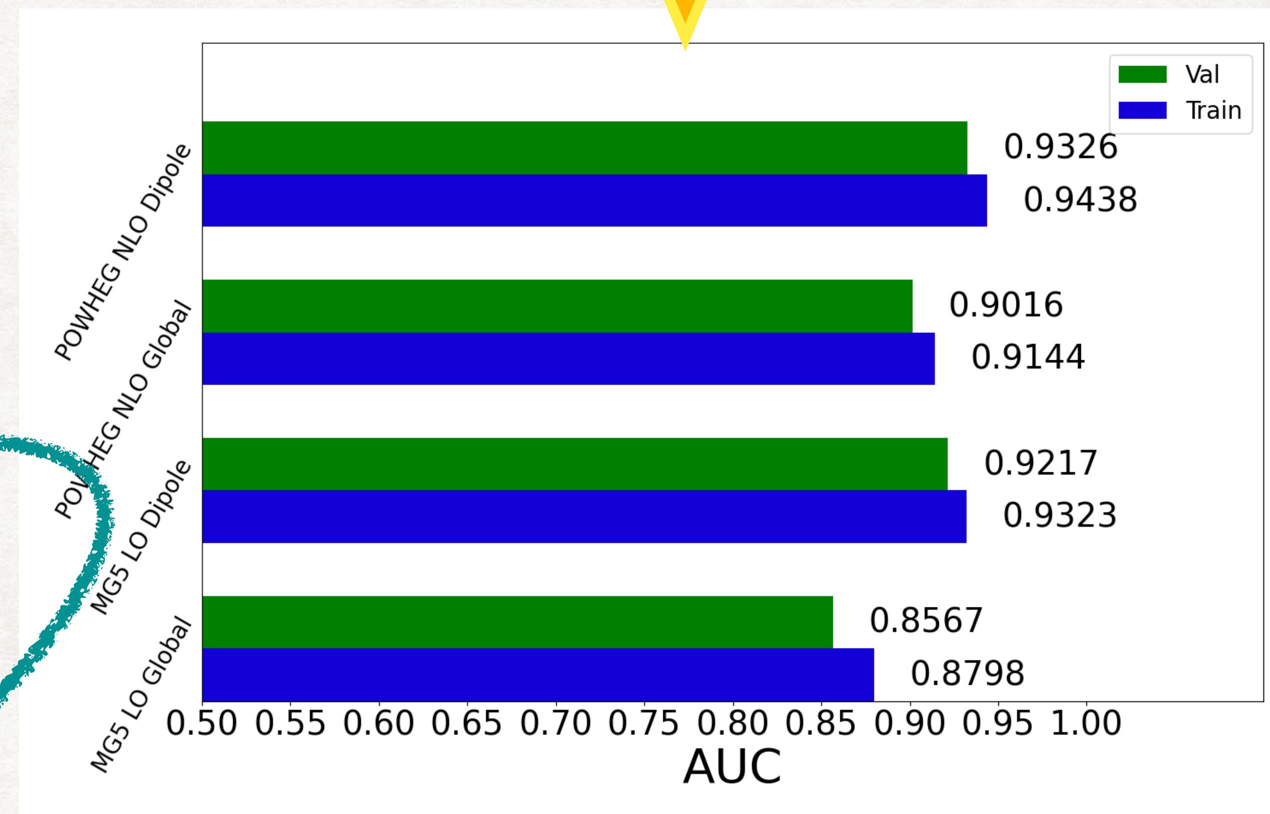
INVISIBLE HIGGS DECAY @ VECTOR-BOSON FUSION

ROLE OF PARTON SHOWER

Receiver Operator Characteristics (ROC) Curve



AUC

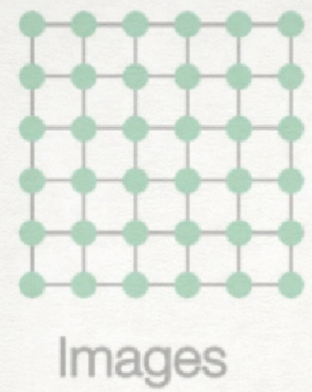


- ✓ LO + Global parton shower scheme shows lowest performance
- ✓ NLO + Dipole parton shower scheme shows best performance
- ✓ Rest two (LO+ Dipole & NLO+ Global) shows intermediate performance

Accurate description

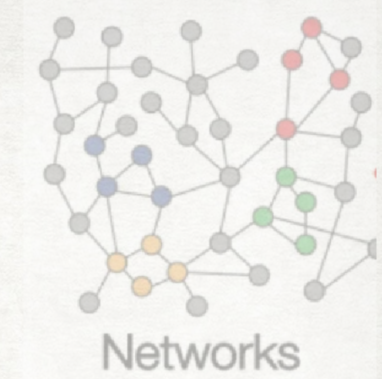
Common in analysis

Vishal Ng, PK (2021)



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!!

POINT CLOUD

.

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

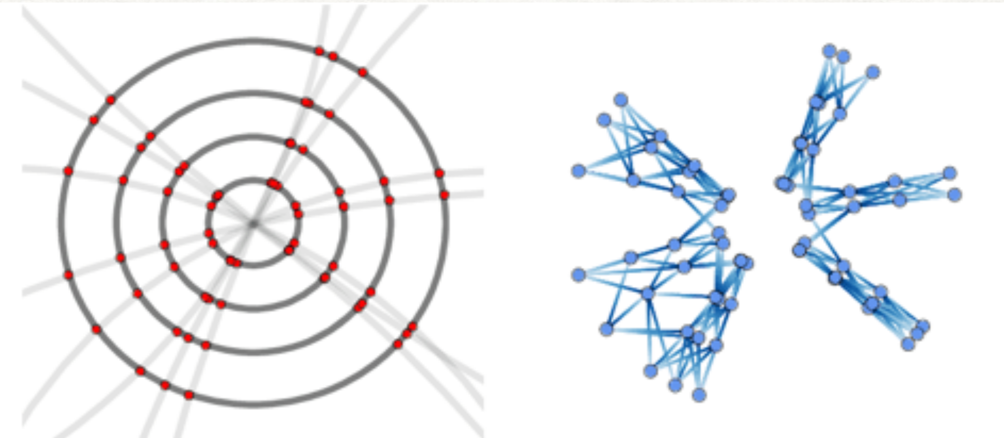
Each data sample is a set with variable cardinality:

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

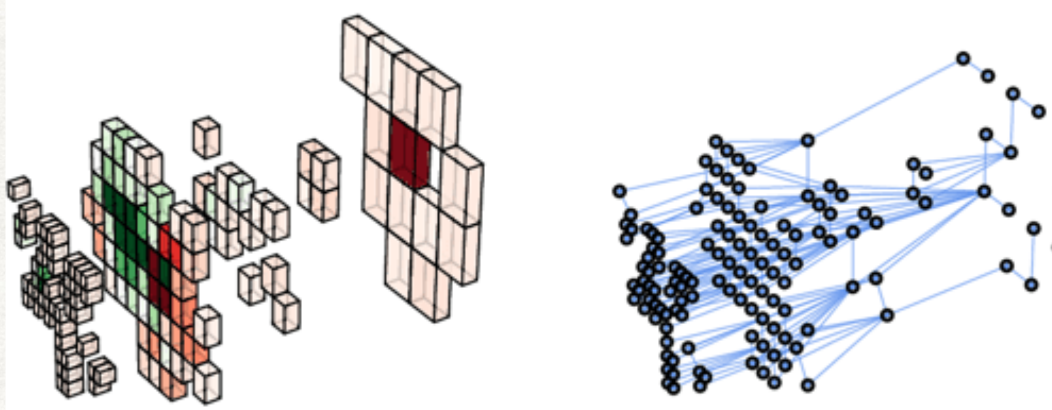
Can also be a collection of sets:

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

$\alpha =$ 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: $\{\mathbf{h}_a, \mathbf{h}_b, \mathbf{h}_c\}$

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

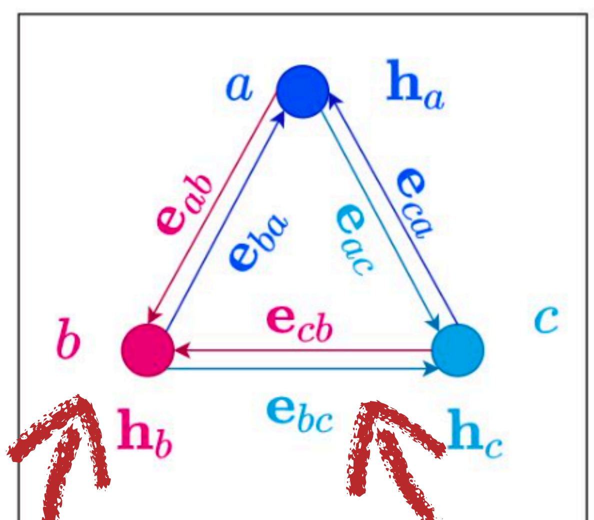
Edge features: $\{\mathbf{e}_{ba}, \mathbf{e}_{ca}, \mathbf{e}_{ab}, \ldots\}$

Neighbourhood sets:

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

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

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



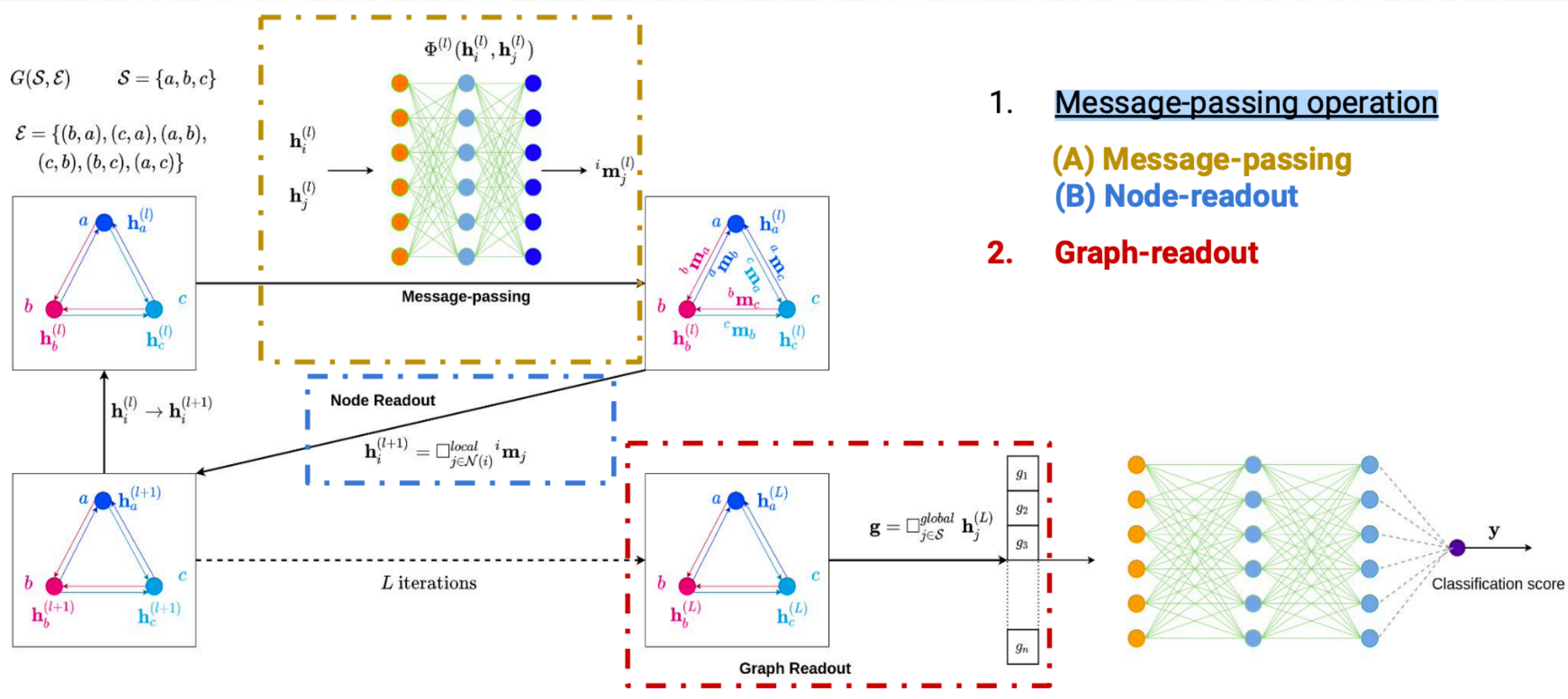
Closed Neighbourhood sets for IRC safety:
$\mathcal{N}[a] \ni a$

7

Edge

Node

MESSAGE-PASSING OPERATIONS



1: Each node in the graph has associated (hidden states) feature vector.

A: Evaluate message (for each edge) by exchanging information between connected nodes

B: [Sum of] all messages obtained from the neighbour update at each node : Update feature

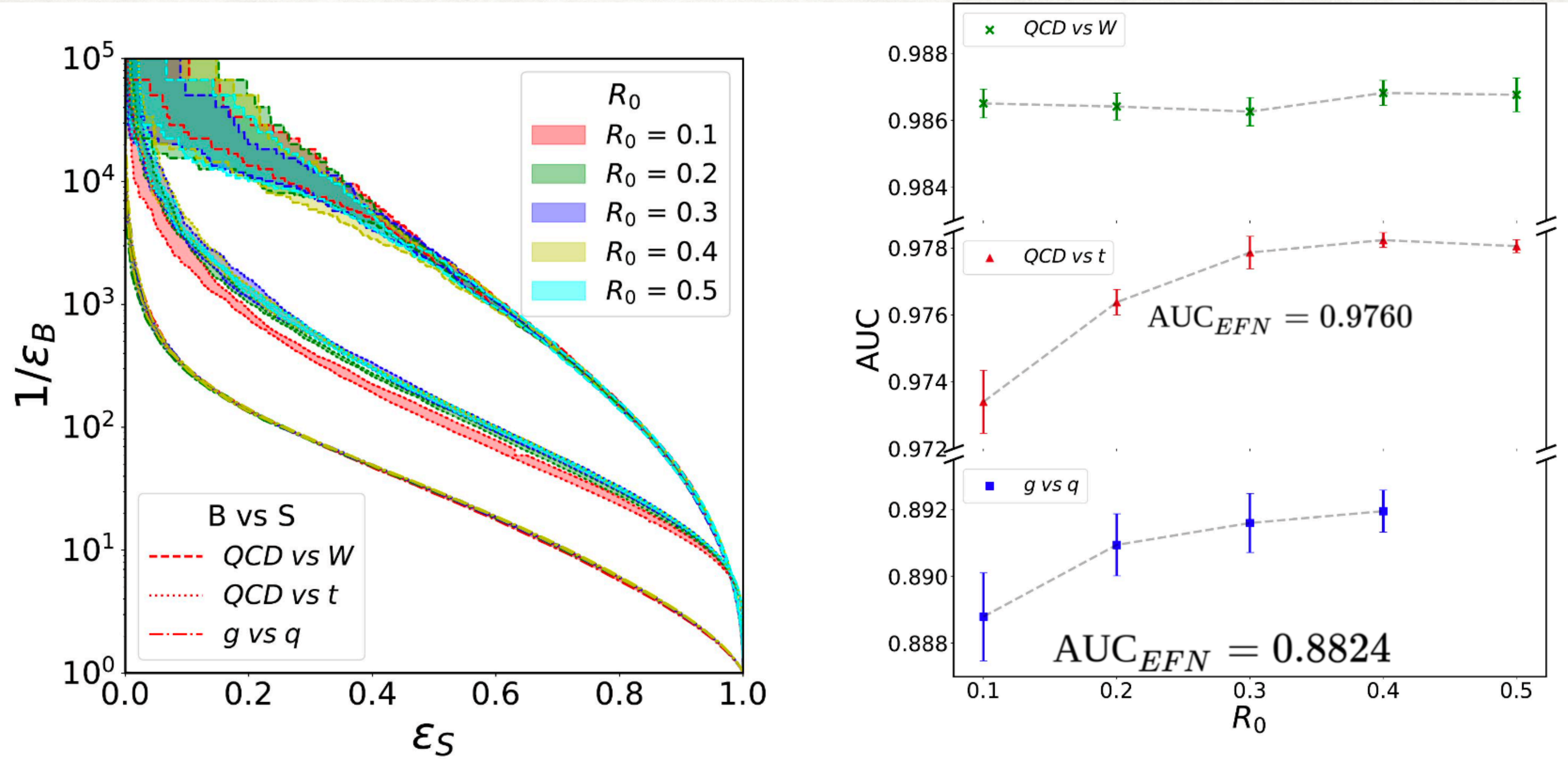
=> Has information about the neighbourhoods

Can go to step A for successive run (l = 1,2,..L) local to global information

2: [Extract] all updated hidden states and create a final feature vector describing the whole graph. This feature vector can be then used as input to a standard machine learning model.

NETWORK PERFORMANCE

EMPN - IRC SAFE WAY



PK, Vishal Ng, Michael Spannowsky; 2022

UNIVERSAL APPROXIMATION OF IRC SAFE OBSERVABLES

★ Any IRC safe observable $\mathcal{O}$ can be expanded in a basis of C-correlators

$$\mathcal{O} \approx \sum_{N=0}^{N_{max}} C_N^{f_N} \quad , \quad C_N^{f_N} = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} E_{i_1} E_{i_2} \dots E_{i_N} f_N(\hat{p}_{i_1}, \hat{p}_{i_2}, \dots \hat{p}_{i_N})$$

✓ Energy Flow Networks : Deep sets model which learns a per-particle map of each particle's directional coordinates $C_1 = \sum_i z_i \, g_1(\hat{\mathbf{p}}_i)$

✓ Energy-weighted Message Passing Networks $\mathbf{h}_i^{(\alpha+1)} = \sum_j z_j \, \mathbf{g}^{(\alpha+1)}(\mathbf{h}_i^{(\alpha)}, \mathbf{h}_j^{(\alpha)})$

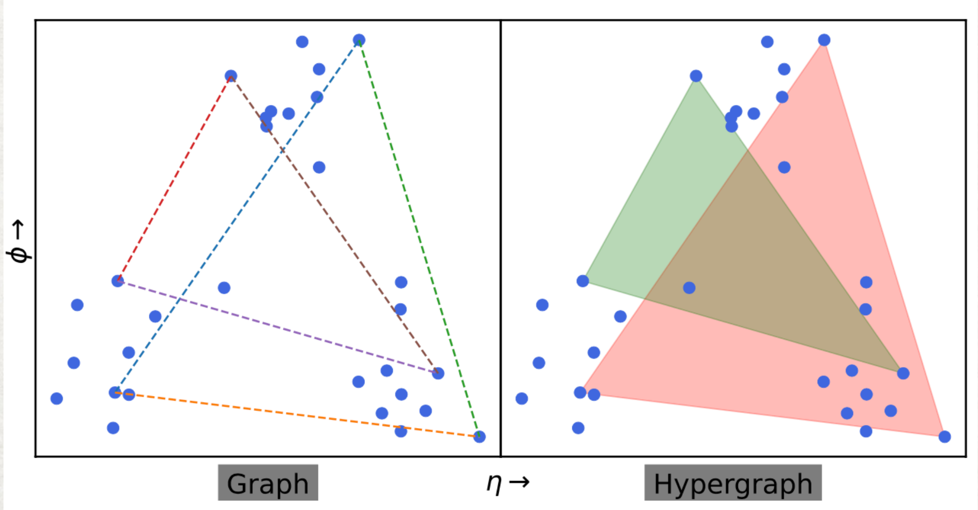
➡ The IRC safe graph representation is obtained as $\mathbf{G}^{(L)} = \sum_{i=1}^{n_{part}} z_i \, \mathbf{h}_i^{(L)}$

➡ For a particular L, we can efficiently extract 2^L angular arguments

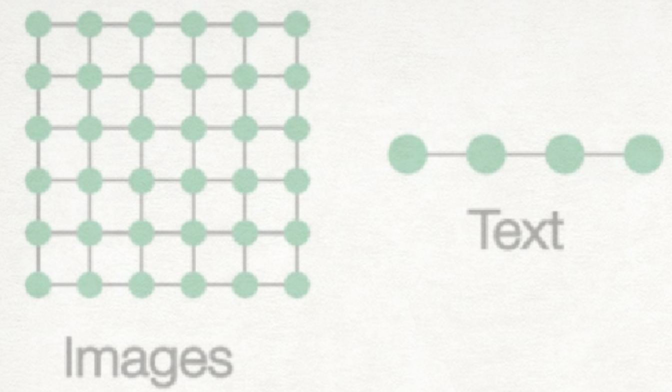
TOWARDS IRC SAFE H-EMPN

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

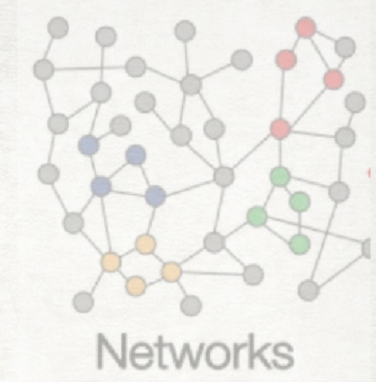


PK, Vishal Ng, Michael Spannowsky; 2023



IRC SAFE EMPN

GRAPH NEURAL NETWORK



- ➡ Generalised Energy Flow Networks (EFNs) to extract local correlations via message-passing operations
- ➡ Single Energy-weighted message passing improves upon EFNs
- ➡ Iterative application does not spoil IRC safety, performs better with reduced sensitivity to soft and collinear emissions
- ✓ **Devised generic graph construction algorithms which give invariant graph structure in the *deletion* of a soft or collinear vertex**
- ◉ **Graph: embed general structure for intuitive physics input**
- ◉ General enough to study inclusive event shapes
- ◉ **Infra-red and collinear safe GNN mechanism is constructed for QCD jet study**

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