Studying Higgs Bosons in New Physics

—— Some recent studies by machine learning

Jin Min Yang

ITP, CAS

Nanjing 2021.8.30

Contents

Part 1 A mini review of Higgs in new physics

--- from technicolor to SUSY

Part 2 Machine learning for Higgs physics at LHC

- Introduction to machine learning
- Machine learning for CP property of top-Higgs coupling at LHC
- Machine learning for triple-Higgs coupling at LHC
- Machine learning for an ALP at LHC
- Conclusion

Part 1 A mini review of Higgs in new physics

--- from technicolor to SUSY

Possible New Physics

to solve fine-tuning problem

- Double the number of particles (Higgs as fundamental scalar): SUSY !
- Higgs is composite: Technicolor !

  For other composite Higgs,
  see the talk by Honghao Zhang and Fengfeng Cai
- Higgs as Pseudo-Goldstone boson: Little Higgs !

  Twin Higgs !

Part 1 A mini review of Higgs in new physics

--- from technicolor to SUSY

Possible New Physics

to solve fine-tuning problem

- Double the number of particles (Higgs as fundamental scalar): SUSY !
- Higgs is composite: Technicolor ! For other composite Higgs, see the talk by Honghao Zhang and Fengfeng Cai
- Higgs as Pseudo-Goldstone boson: Little Higgs ! Twin Higgs !
- Two-Higgs-doublet models (not solve fine-tuning)

  Three-Higgs-doublet models see the talk by Igor Ivanov

Two-Higgs-Doublet Models

Lei Wang, 2021 Pre-SUSY school

Ning Chen, Pre-school, this workshop

T. D. Lee, PRD8 (1973) 1226

H. E. Haber, G. L. Kane, T. Sterling, NPB161 (1979) 493

L. J. Hall, M. B. Wise, NPB187 (1981) 397

J. G. Donoghue, L. F. Li, PRD19 (1979) 945

V. D. Barger, J. L. Hewett, R. Phillips, PRD41 (1990) 3421

Twin-Higgs Models

Z. Chacko, H. S. Goh, R. Harnik, PRL 96, 231802 (2006)

JHEP 01, 108 (2006)

The twin Higgs mechanism is proposed as an interesting solution to the hierarchy problem. The SM Higgs emerges as a pseudo-Goldstone boson once a global symmetry is spontaneously broken, which is similar to what happens in the little Higgs models. An additional discrete symmetry is imposed, which ensures the absence of one-loop quadratic divergence of Higgs mass. The resulting Higgs boson mass is naturally around the electroweak scale when the cut-off scale of the theory is around 5-10 TeV.

Thanks to Guo-Li Liu

technicolor

S. Weinberg, PRD13(1976) 974; D19(1979) 1277

L. Susskind, PRD20 (1979) 2619

E. Farhi, L. Susskind, Phys. Rept. 74 (1981) 277

technicolor

Introduce TC-fermions which have QCD-like strong interaction. At some high scale (say TeV), TC-fermions condensate, replacing Higgs mechanism for EWSB.

基本思想

引入一种新的渐进自由的类似QCD的强相互作用和具有这种强相互作用的TC费米子，在一个高的能量标度（比如TeV）TC费米子发生凝聚，以此代替标准模型中的黑格斯粒子来实现电弱对称性到电磁对称性的动力学破缺

- 黑格斯玻色子是两个TC费米子的束缚态
- 最初由超导理论启发而发展起来

正是参照朗道-金兹伯格的超导自由能形式和强子物理中的手征性破缺的Gell-Man Levy西格玛模型，提出微观物理不需要基本标量场，类比于BCS理论中的库珀对凝聚，提出了夸克凝聚的TC理论

technicolor

- **Dynamical EWSB without fundamental Higgs**
- **New interaction appear at TeV, solving hierarchy problem**

解决了标准模型的什么问题

- 解释了SM的EWSB而无需基本标量场和$\mu^2 < 0$的条件
- 解决了SM的等级问题

technicolor

Challenges

1. How can new strong dynamics generate fermion masses?

   Pure technicolor does not; ETC is needed

2. How can the new dynamics generate fermion masses without large FCNC ?

   Changing behavior of TC dynamics (walking TC) can help

3. The dynamics splitting top and bottom masses would tend to affect ρ-parameter. How to keep $\delta\rho \sim 0$?

   Models with additional top strong interactions (topcolor)

technicolor

Until 2012, TC theory was allowed by experiments;

The only trouble is its inelegance, getting more and more complex.

直到2012年，TC理论基本上与实验符合

大多数诟病是针对它的inelegance：

- 为了解释电弱自发破缺，引入一种新的力TC，避开了基本标量场，可以给出合理的W和Z质量；
- 为了解释普通费米子质量，扩充了TC，成为ETC；
- 为了自洽-同时给出费米子质量而又避开FCNC，引入了walking机制
- 为了提供顶夸克质量，引入了topcolor相互作用

technicolor

In 2012 LHC discovery of 125 GeV Higgs give TC a fatal blow, because in TC2 theory the top-Higgs boson formed by the top quark condensation should be heavier than 300 GeV.

Improvement:

1 topcolor improved to top seesaw

Aad, et al., PLB (2012)
Chatrchyan, et al, PLB (2012)

2 conformal TC

Franzosi, Foadi, PRD 88, 015013(2013)

technicolor

Learn a lesson:

- **TC theory, starting from a very simple very beautiful idea, is getting more and more complex;**
- **TC/ETC is too ambitious, dynamical EWSB plus generating fermion masses, too good too big to be true.**

正确看待 TC理论

- TC理论发展过程中经历了一系列的修修补补拆东墙挖西墙捉襟见肘的样子
- 但是因为TC/ETC 理论野心很大（成功的理论将解释EWSB 并提供SM所有费米子的质量来源，而不像SM那样由任意参数给出），所以至今为止没有建立起完善的理论也并不令人吃惊

technicolor

Commemoration for TC：

- **Give up fundamental Higgs to solve naturalness problem**
- **A window to dynamical EWSB through strong interaction**
- **The discovery of 125 GeV Higgs gave a fatal blow to TC. Either give up or modify TC. The traditional TC predicts heavy composite Higgs (arXiv:1108.4000)**

Thanks to Lei Wang

Little Higgs Theory

Arkani-Hamed, Cohen, Nelson, hep-ph/0105239

Motivation:

No one-loop quadratic divergence of Higgs mass

What is little Higgs

- **Higgs boson is a pseudo Goldstone boson of an approximate global symmetry**
- **At least two sets of interactions are needed to break this global symmetry -- collective symmetry breaking**

小黑格斯的历史：

- 为了自然地得到一个较轻的希格斯粒子，人们早在二十世纪七十年代就提出了把希格斯粒子看作赝哥德斯通粒子，以保持它的小质量，这就是小希格斯理论最初的思想(Georgi)
- 八十年代Georgi和Kaplan以这个思想构造了一个这样的模型，但是从电弱标度到该模型的截断标度仍然需要精细调节
- 二十世纪初，受"Dimension (De)Construction"(Arkani-Hamed, Cohen, Georgi) 的启发，人们引进了协同对称破缺机制，将希格斯粒子构造成赝哥德斯通粒子的小黑格斯模型成功建立起来(Arkani-Hamed, Cohen, Georgi, Katz, Nelson)

Little Higgs Theory

- SUSY (symmetry between *opposite* spin & statistics)

  Natural cancellations: $\tilde{t}$ versus t

  $\tilde{W}$ versus W

  $\tilde{H}$ versus H

- The Little Higgs

  Quadratic divergences cancelled at one-loop level by new states:

  $$W, Z, B \leftrightarrow W_H, Z_H, B_H; \quad t \leftrightarrow T; \quad H \leftrightarrow \Phi$$

  (cancellation among same spin states!)

Little Higgs Theory

collective symmetry breaking

$$L = L_0 + \epsilon_1 L_1 + \epsilon_2 L_2$$

$$\delta m_h^2 \sim \frac{\varepsilon_1^2}{16\pi^2}\frac{\varepsilon_2^2}{16\pi^2}\Lambda^2$$

假设$\epsilon_1 L_1$仅仅破坏一部分整体对称性，而L仍剩下足够的整体对称性使Higgs粒子依然是严格的哥德斯通粒子（无质量）

再引进另外一个明显破缺项$\epsilon_2 L_2$

$$L = L_0 + \epsilon_1 L_1 + \epsilon_2 L_2$$

$\epsilon_1 L_1$和 $\epsilon_2 L_2$这两项一起才能破坏足够的整体对称性，使Higgs成为赝哥德斯通粒子，而黑格斯的质量项必须同时含有ϵ_1 和 ϵ_2

$$\delta m_h^2 \sim \frac{\varepsilon_1^2}{16\pi^2}\frac{\varepsilon_2^2}{16\pi^2}\Lambda^2$$

具有二次发散贡献的Higgs质量在两圈出现

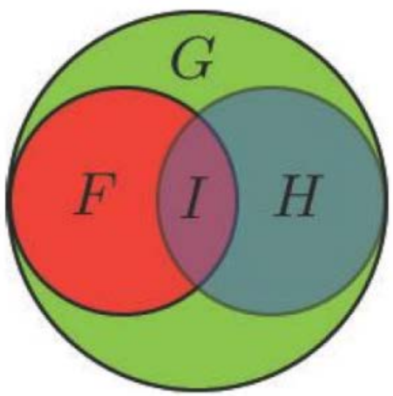
Little Higgs Theory

Littlest Higgs Model

Arkani-Hamed, Cohen, Katz, Nelson, JHEP07(2002)034

Global Symmetries	Gauge Symmetries
$SU(5)/SO(5)$	$[SU(2)\otimes U(1)]^2$

G H F $I = SM$



$$SU(5) \xrightarrow{f} SO(5)$$

14 Goldstones

Little Higgs Theory

Phenomenology

(1) **Higgs diphoton signal is suppressed**

Wang, JMY, PRD84, 075024(2011)

(2) **T-odd particles, top-partner**

Belyaev, Chen, Tobe, Yuan, PRD74, 115020 (2006)

(3) **Dark matter**

Wang, JMY, Zhu, PRD88 (2013) 075018

Supersymmetry

In fact, the concept of supersymmetry emerged historically at least in part because of its role in string theory.

Experimental discovery of supersymmetry would certainly give string theory a big boost, and learning how supersymmetry is broken might very well give string theorists crucial clues about how to proceed.

Edward Witten

2003.8.16

Supersymmetry

* SUSY can make a “small” Higgs mass natural
* SUSY is part of a larger vision of physics,
  not just a technical solution
* measured value of $\sin^2\theta$ favors SUSY GUT
* survives electroweak tests
* heavy top mass, as needed

Edward Witten

2003.8.16

Supersymmetry

Minimal SUSY (MSSM)

SM field	Super partner	$SU_L(2)$	$U_Y(1)$
$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	$\tilde{L} = \begin{pmatrix} \tilde{\nu}_L \\ \tilde{e}_L^- \end{pmatrix}$	$\mathbf{2}$	$-\frac{1}{2}$
e_L^c	$\tilde{E} = \tilde{e}_L^c \Rightarrow \tilde{e}_R^*$	$\mathbf{1}$	$+1$
$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\tilde{Q} = \begin{pmatrix} \tilde{u}_L \\ \tilde{d}_L \end{pmatrix}$	$\mathbf{2}$	$+\frac{1}{6}$
u_L^c	$\tilde{U} = \tilde{u}_L^c \Rightarrow \tilde{u}_R^*$	$\mathbf{1}$	$-\frac{2}{3}$
d_L^c	$\tilde{D} = \tilde{d}_L^c \Rightarrow \tilde{d}_R^*$	$\mathbf{1}$	$+\frac{1}{3}$
$H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix}$	$\tilde{H}_d = \begin{pmatrix} \tilde{H}_d^0 \\ \tilde{H}_d^- \end{pmatrix}$	$\mathbf{2}$	$-\frac{1}{2}$
$H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix}$	$\tilde{H}_u = \begin{pmatrix} \tilde{H}_u^+ \\ \tilde{H}_u^0 \end{pmatrix}$	$\mathbf{2}$	$+\frac{1}{2}$

Note: all fermion fields are two-component Weyl spinors !

Supersymmetry

SUSY needs 2 Higgs doublets

To give masses for both up and down type quarks, two Higgs doublets with opposite Y are needed since H_1^* term in W violates SUSY

$$y_{u,ij}\hat{Q}_i\hat{H}_u\hat{U}_j^c + y_{d,ij}\hat{Q}_i\hat{H}_d\hat{D}_j^c$$

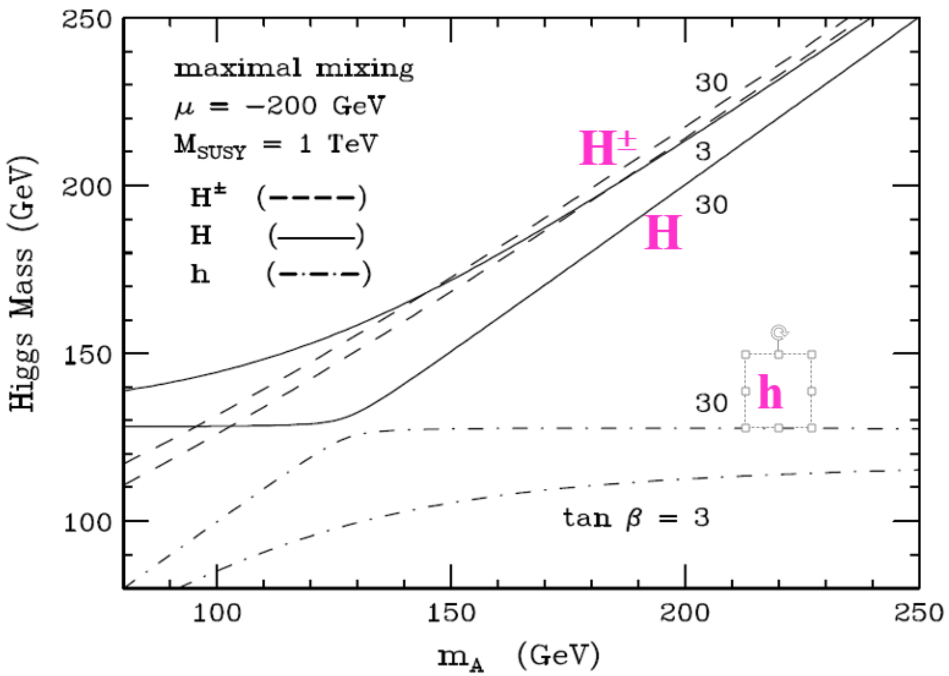
To cancel $SU(2)_L \times U(1)_Y$ anomaly, need one more higgsino and thus one more Higgs chiral superfield

$$H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix} \qquad \tilde{H}_d = \begin{pmatrix} \tilde{H}_d^0 \\ \tilde{H}_d^- \end{pmatrix} \qquad \mathbf{2} \qquad -\tfrac{1}{2}$$

$$H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix} \qquad \tilde{H}_u = \begin{pmatrix} \tilde{H}_u^+ \\ \tilde{H}_u^0 \end{pmatrix} \qquad \mathbf{2} \qquad +\tfrac{1}{2}$$

Supersymmetry

SUSY predicts 5 Higgs bosons



Supersymmetry

SUSY predicts 5 Higgs bosons

At tree level: $m_h < m_Z |\cos 2\beta| < m_Z$

At loop-level:

$$M_{\tilde{t}}^2 = \begin{pmatrix} M_{\tilde{Q}}^2 + m_Z^2 \cos 2\beta(\frac{1}{2} - \frac{2}{3}s_W^2) + m_t^2 & m_t(A_t - \mu \cot \beta) \\ m_t(A_t - \mu \cot \beta) & M_{\tilde{U}}^2 + \frac{2}{3}m_Z^2 \cos 2\beta s_W^2 + m_t^2 \end{pmatrix}$$

$$\equiv \begin{pmatrix} m_{\tilde{t}_L}^2 & m_t X_t \\ m_t X_t & m_{\tilde{t}_R}^2 \end{pmatrix}$$

$$m_{\tilde{t}_{1,2}}^2 = \frac{1}{2}\left(m_{\tilde{t}_L}^2 + m_{\tilde{t}_R}^2\right) \mp \frac{1}{2}\sqrt{\left(m_{\tilde{t}_L}^2 - m_{\tilde{t}_R}^2\right)^2 + 4m_t^2 X_t^2}$$

$$M_S^2 \equiv \left(m_{\tilde{t}_1}^2 + m_{\tilde{t}_2}^2\right)/2$$

$$m_h^2 \leq m_Z^2 + \epsilon = m_Z^2 + \frac{3m_t^4}{2\pi^2 v^2}\left[\log\frac{M_S^2}{m_t^2} + \frac{X_t^2}{M_S^2}\left(1 - \frac{X_t^2}{12M_S^2}\right)\right]$$

$$\leq 135\ GeV \quad (for\ M_S \leq 2TeV)$$

Supersymmetry

SUSY predicts 5 Higgs bosons

Neutral Higgs Couplings:

	h	H	A	type
$t\bar{t}$	$\cos\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\gamma_5\cot\beta$	$Hf\bar{f}$
$b\bar{b}\ \tau\bar{\tau}$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\cos\beta$	$\gamma_5\tan\beta$	
$WW,\ ZZ$	$\sin(\beta-\alpha)$	$\cos(\beta-\alpha)$	0	HVV
$Z\ A$	$\cos(\beta-\alpha)$	$\sin(\beta-\alpha)$	0	HHV

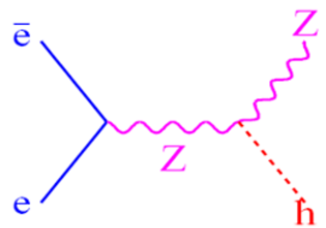
$\Downarrow\Downarrow\Downarrow$ $M_A \to \infty\ (\alpha \to \beta - \pi/2)$ $\Downarrow\Downarrow\Downarrow$

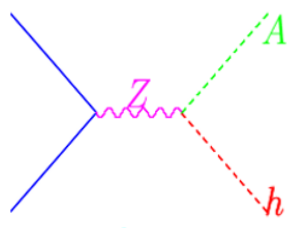
	h	H	A	type
$t\bar{t}$	1	$-\cot\beta$	$\gamma_5\cot\beta$	$Hf\bar{f}$
$b\bar{b}\ \tau\bar{\tau}$	1	$\tan\beta$	$\gamma_5\tan\beta$	
$WW,\ ZZ$	1	0	0	HVV
$Z\ A$	0	1	0	HHV

Supersymmetry

SUSY Higgs bosons have very rich pheno at colliders

At lepton colliders LEP， CEPC， ……





See, e.g.,

talks by Manqi Ruan et al

Li, Song, Su, Su, JMY, 2010.09782

Cao, Han, Ren, Wu, JMY, Zhang, 1410.1018

At hadron colliders LHC， SPPC， ……

See, e.g.,

Pre-school lecture by Yanwen Liu
Several talks in this workshop

Supersymmetry

Status of SUSY in light of 125 GeV Higgs and muon g-2:

cannot explain muon g-2

GMSB/AMSB:	can give 125 GeV Higgs, but with very heavy stop (fine-tuning)
CMSSM/mSUGRA:	can give 125 GeV Higgs; but cannot explain muon g-2
MSSM:	can fit all data well, but suffer from little fine-tuning
nMSSM:	nearly excluded (suppress diphoton rate too much)
NMSSM:	can fit all data well
Split-SUSY:	no problem (give up naturalness)
Stealth SUSY:	no problem (can always escape detections)
Compressed SUSY:	no problem (can escape detection at LHC)

Supersymmetry

Status of SUSY in light of 125 GeV Higgs and muon g-2:

cannot explain muon g-2

GMSB/AMSB: can give 125 GeV Higgs, but with very heavy stop (fine-tuning)

CMSSM/mSUGRA: can give 125 GeV Higgs; but cannot explain muon g-2

Improve them :

Li, Liu, Wang, JMY, Zhang, 2106.04466

Wang, Wu, Xiao, JMY, Zhang, 2104.03262

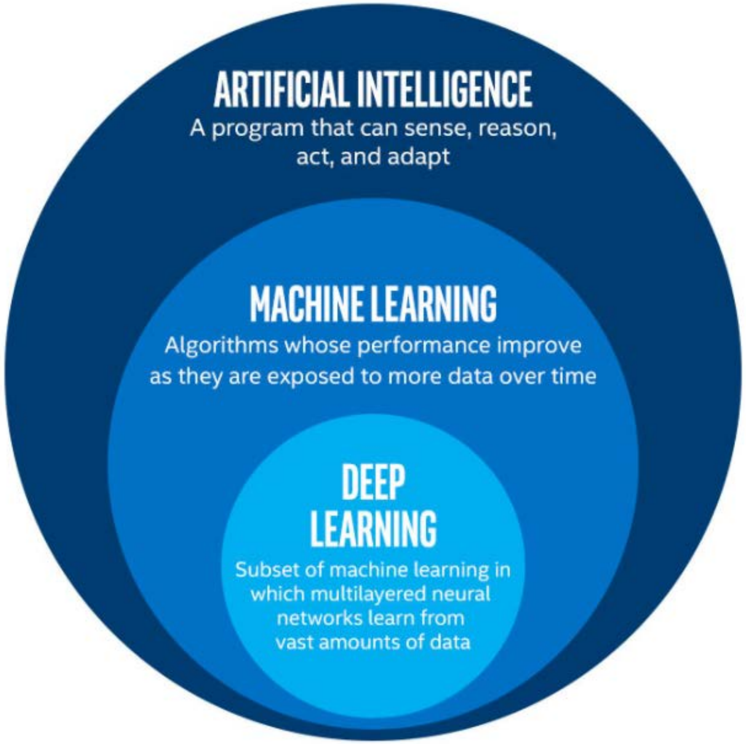
Wang, Wang, Xiao, JMY, Zhu, 1808.10851

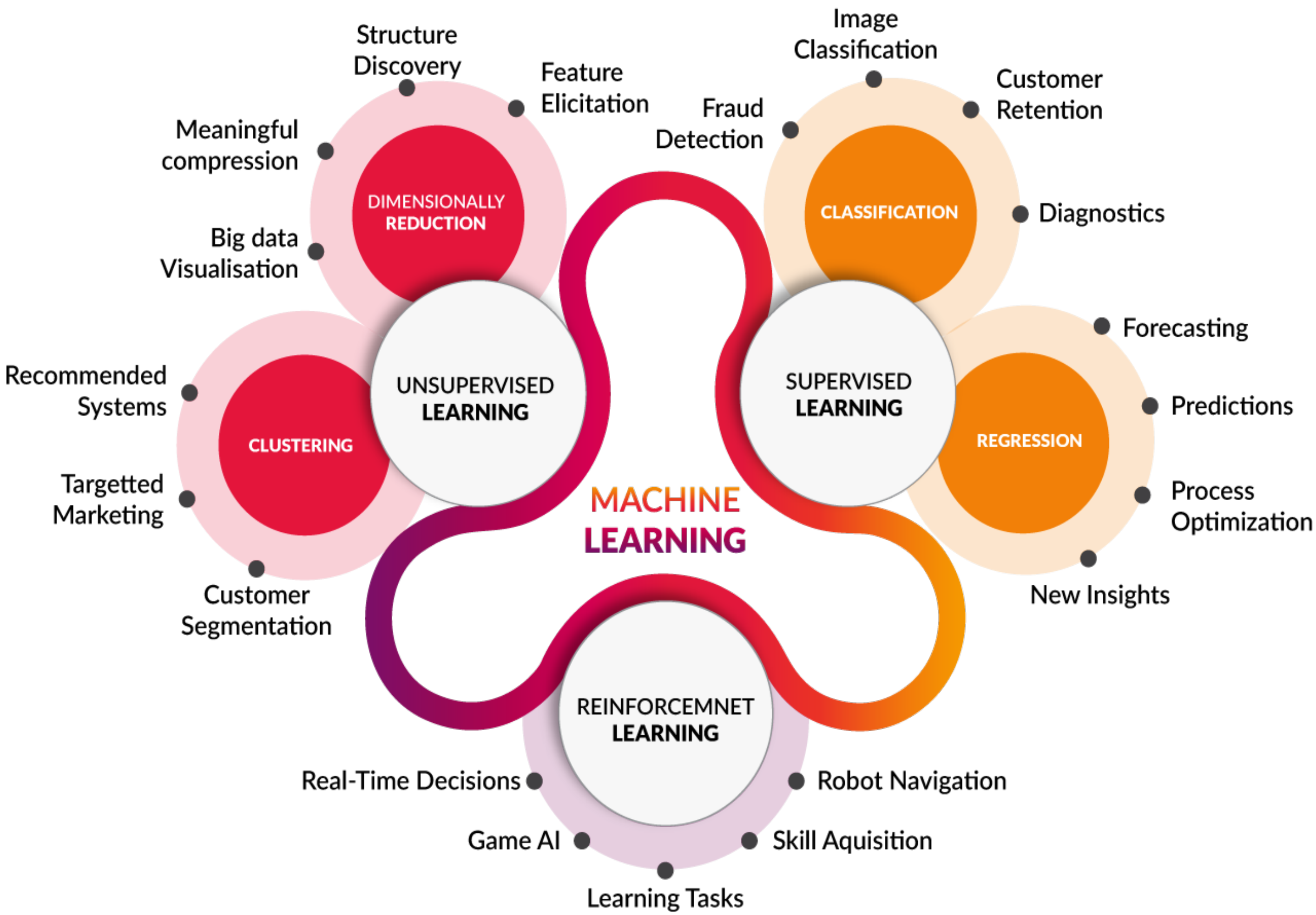
Wang, Wang, JMY, 1703.10894; 1504.00505

Wang, Wang, JMY, Zhang, 1505.02785

Part 2 Machine learning for Higgs physics at LHC

2.1 Introduction to machine learning

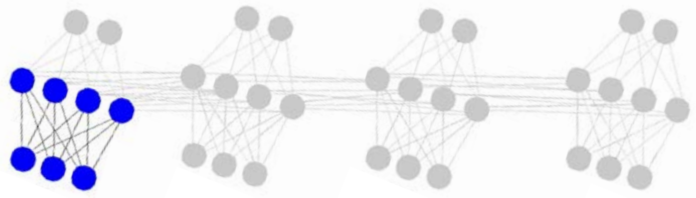




MACHINE LEARNING
UNSUPERVISED LEARNING
SUPERVISED LEARNING
REINFORCEMNET LEARNING
DIMENSIONALLY REDUCTION
CLUSTERING
CLASSIFICATION
REGRESSION
Structure Discovery
Feature Elicitation
Meaningful compression
Big data Visualisation
Recommended Systems
Targetted Marketing
Customer Segmentation
Image Classification
Customer Retention
Fraud Detection
Diagnostics
Forecasting
Predictions
Process Optimization
New Insights
Real-Time Decisions
Robot Navigation
Game AI
Skill Aquisition
Learning Tasks

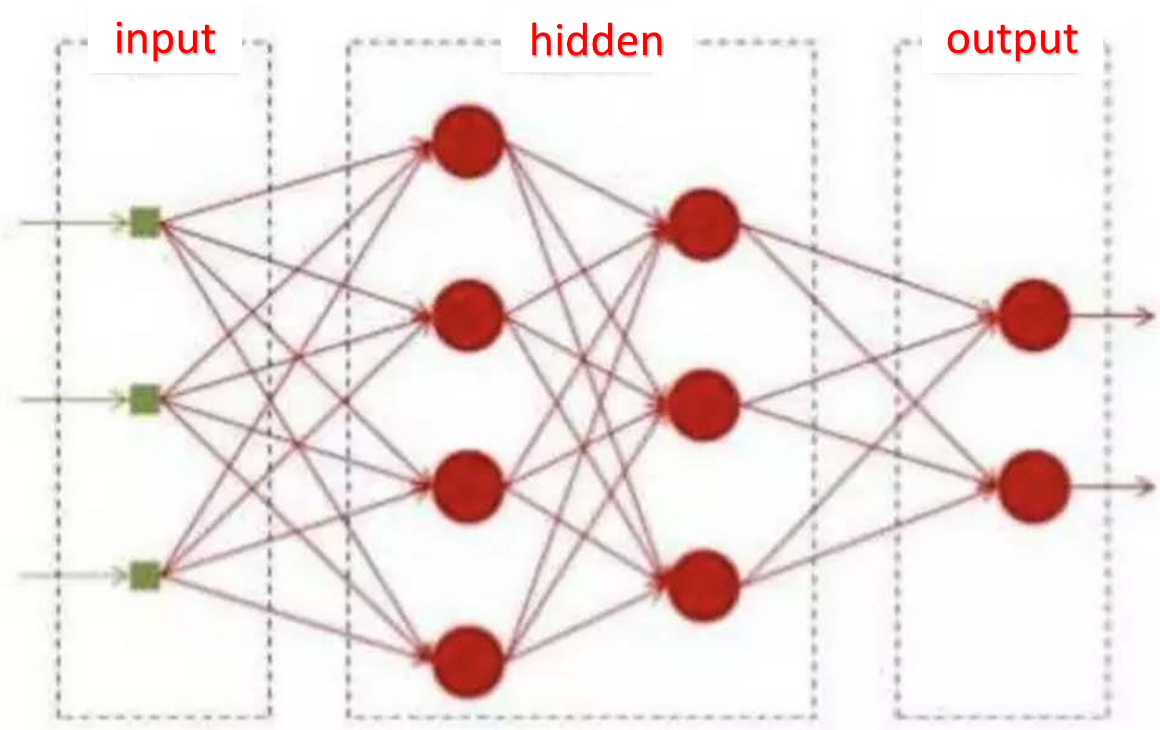
Neural Network

- Warren McCulloch and Walter Pitts (**1943**) created the **first neural network** based on mathematics and algorithms called threshold logic.
- The **perceptron algorithm** was invented in **1957** at the Cornell Aeronautical Laboratory by Frank Rosenblatt.
- For **multilayer perceptron** (feed-forward neural network), where at least one hidden layer exists, more sophisticated algorithms such as backpropagation (Rumelhart, Hinton and Williams, **1986**) must be used.



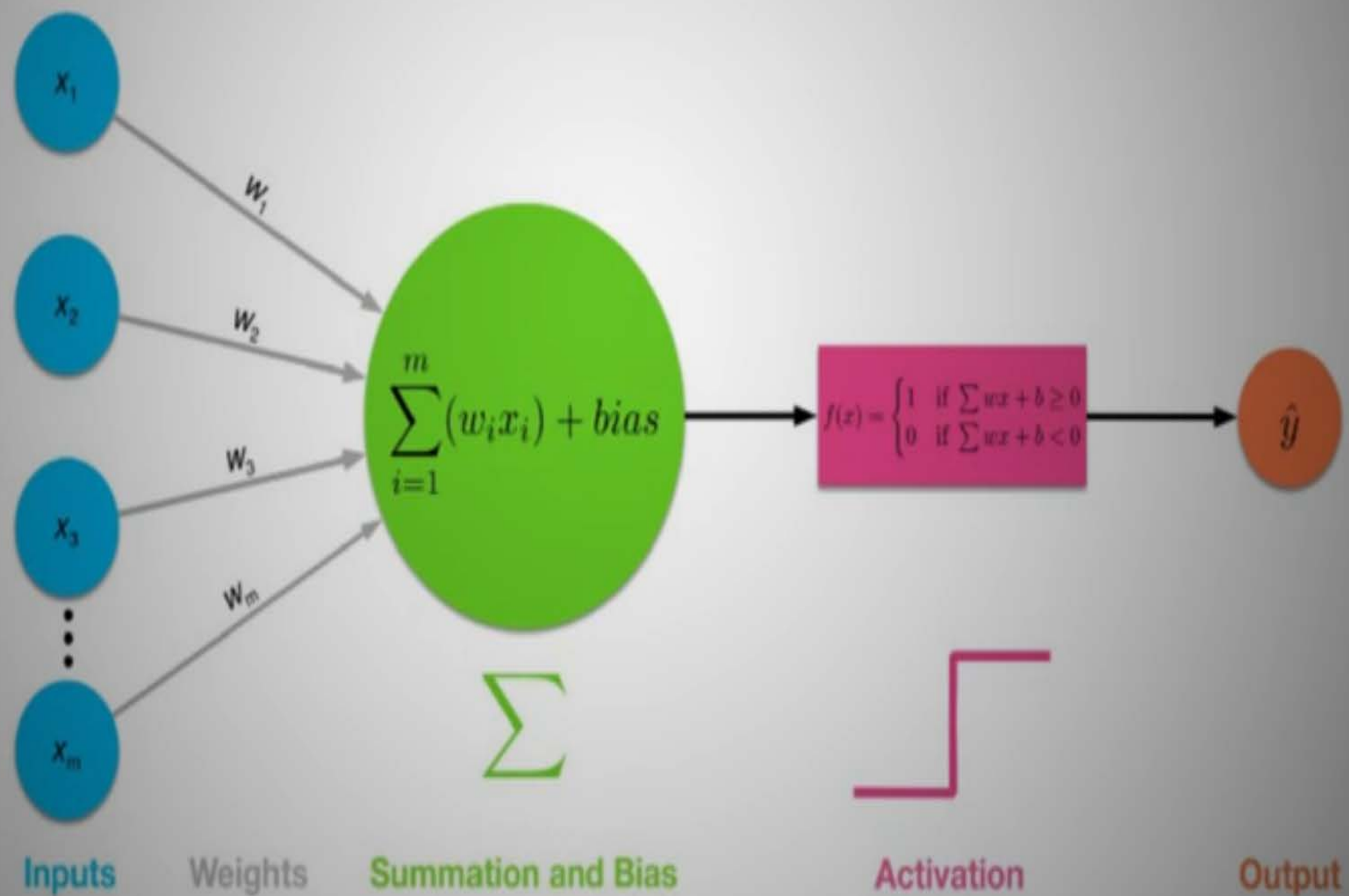
Neural Network

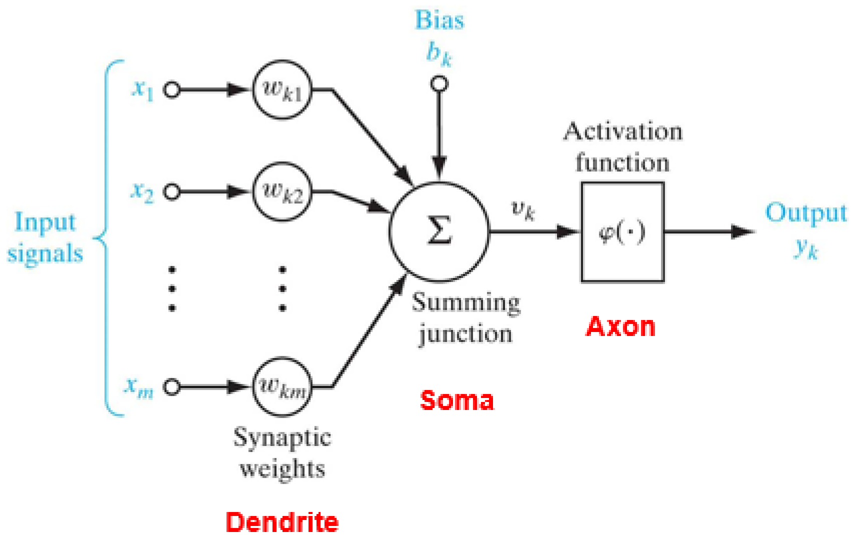
input x



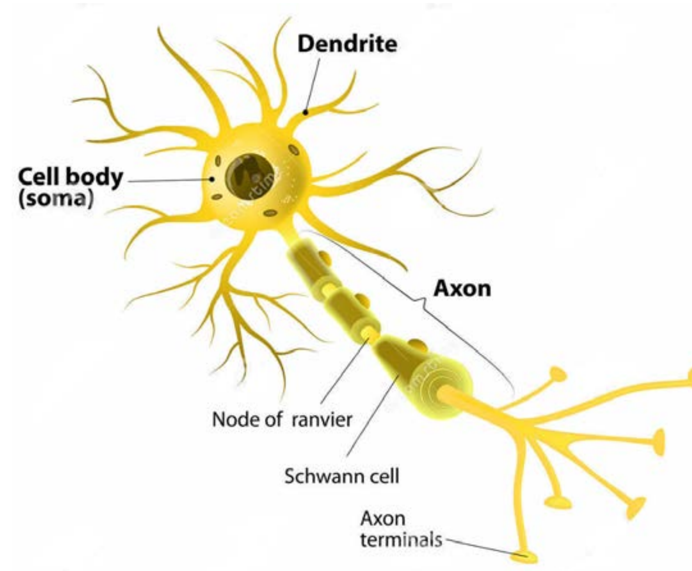
output y

ARTIFICIAL NEURON – THE HEART OF A NEURAL NETWORK





Bias
b_k
x_1
x_2
x_m
w_{k1}
w_{k2}
w_{km}
Input signals
Σ
v_k
Activation function
$\varphi(\cdot)$
Output y_k
Summing junction
Synaptic weights
Axon
Soma
Dendrite



Dendrite
Cell body (soma)
Axon
Node of ranvier
Schwann cell
Axon terminals

Machine Learning in HEP

Abdughani, Ren, Wu, JMY, Zhao, 1905.06047
Supervised deep learning in high energy phenomenology: a mini review

GOAL

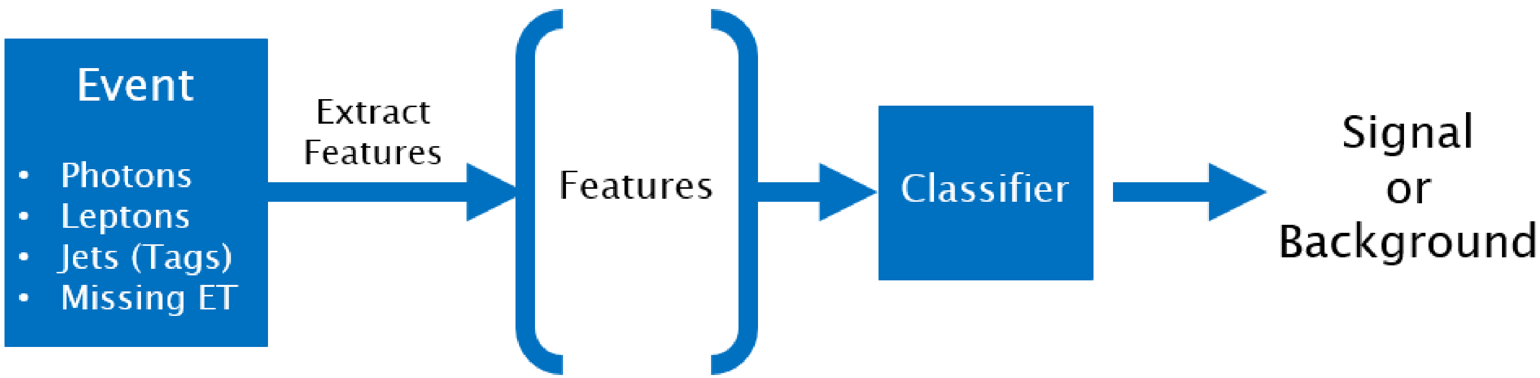
- **“Solve” HEP problems using DATA**

EXAMPLE

- Physics model selection
  - Scan (e.g. 1011.4306, 1106.4613, 1703.01309, 1708.06615)
- Collider
  - Parton distribution function (e.g. 1605.04345)
  - Object reconstruction (e.g. NIPS-DLPS)
  - Pileup mitigation (e.g. 1512.04672, 1707.08600)
  - Jet tagging (e.g. 1407.5675, 1501.05968, 1612.01551, 1702.00748)
  - Event selection (e.g. 1402.4735, 1708.07034, 1807.09088)
  - Decayed object reconstruction
  - Anomaly event detection (e.g. 1807.10261)

Machine Learning in HEP

An **event** is a **signal** or **background** ?



HIGH–LEVEL FEATURES

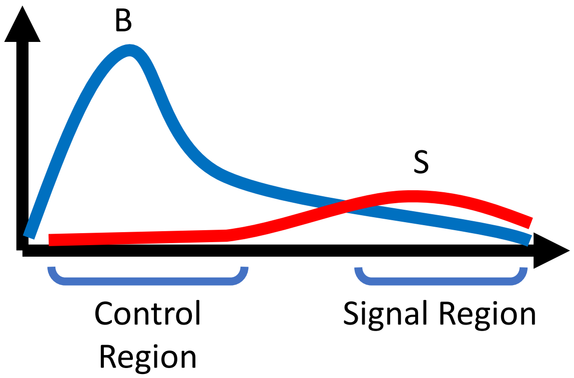
- Number of jets
- p_T of the leading lepton
- $\Delta\phi$ between the leading jet and missing ET
- Reconstructed top mass
- ...

LOW–LEVEL FEATURES

- four–momenta of reconstructed objects
- ...

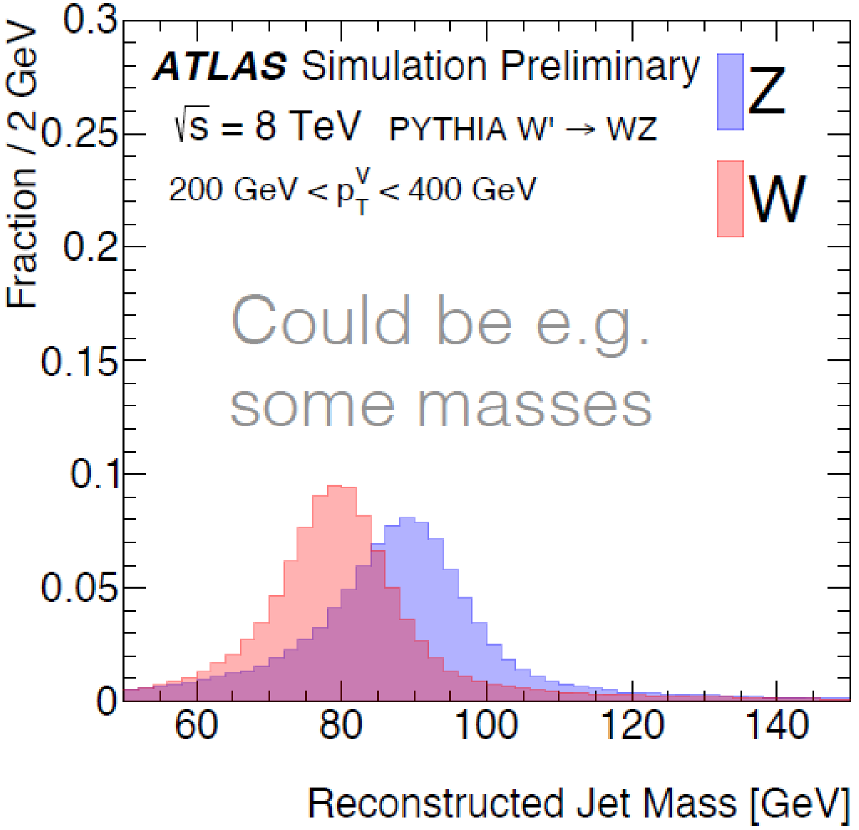
Methods for Event Selection

- **Cut-flow**



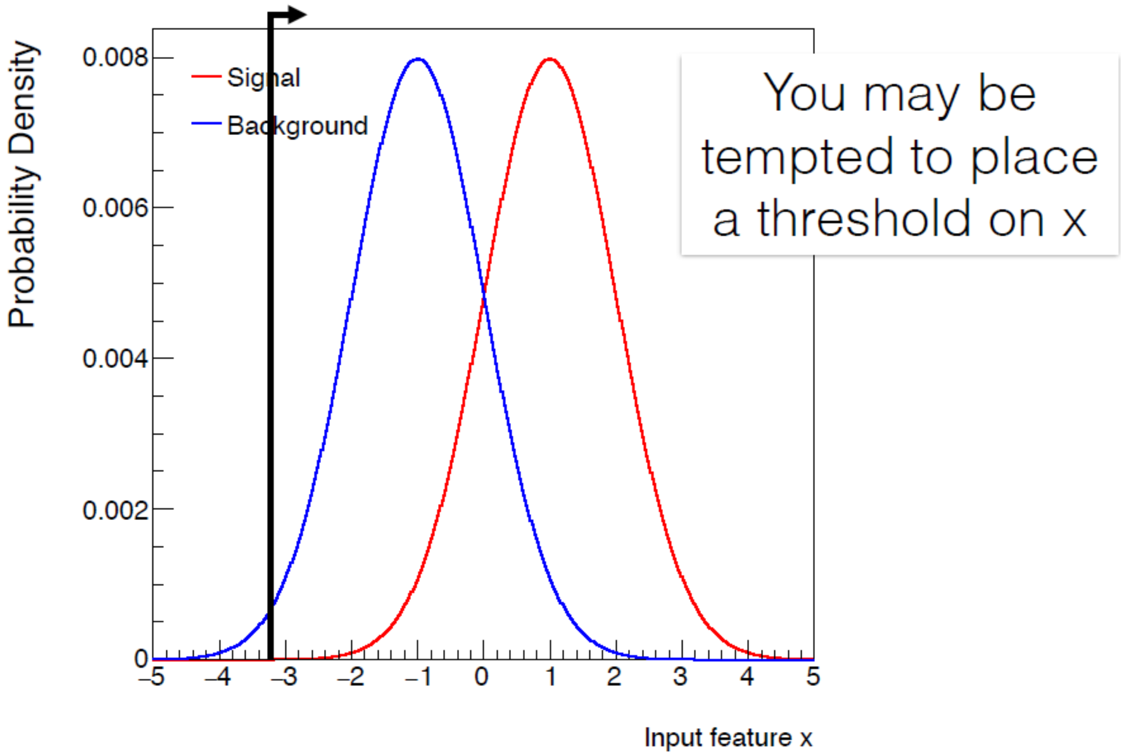
Methods for Event Selection

- **Cut-flow**



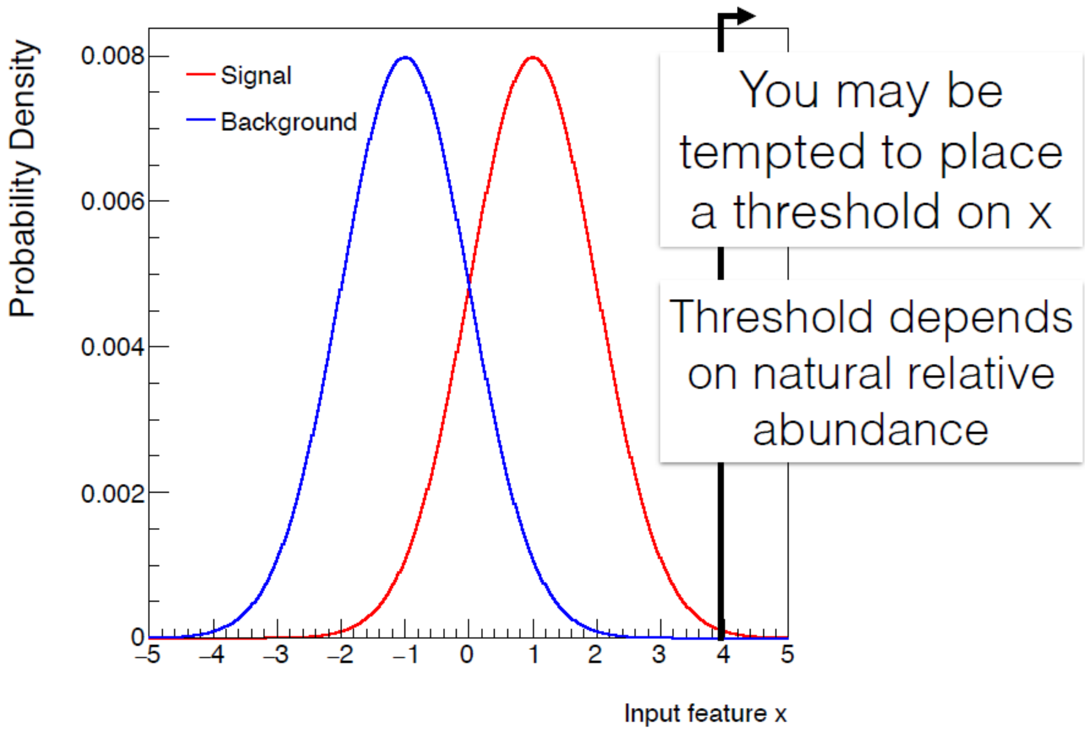
Methods for Event Selection

- **Cut-flow**



Methods for Event Selection

- **Cut-flow**

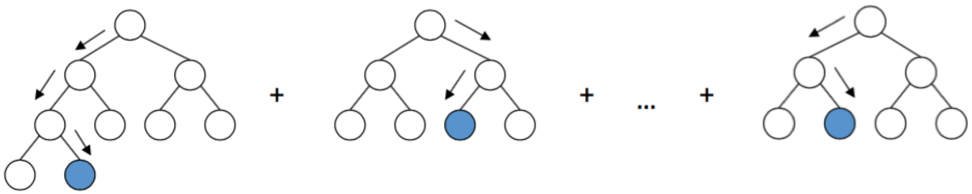


Is the simple threshold cut **optimal**?

Benjamin Nachman

Methods for Event Selection

- **Cut-flow**
- **Machine Learning**
  - **Boosted Decision Tree (BDT)**

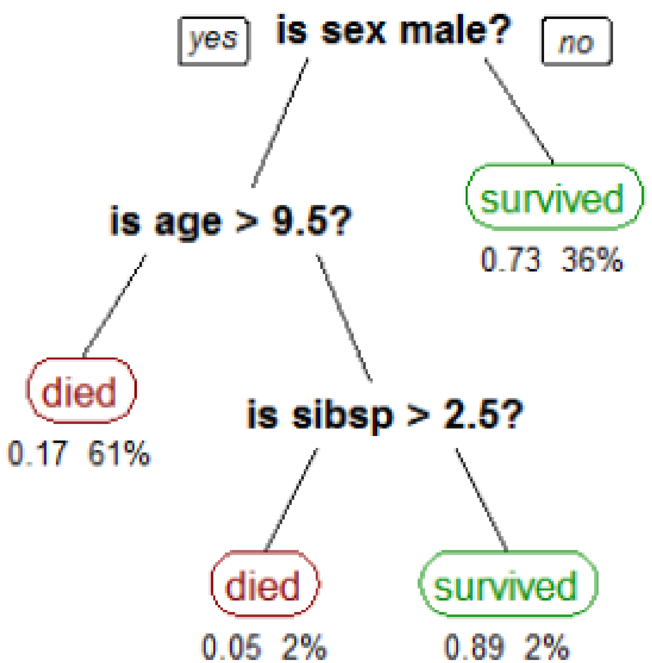


a lot of trees → a forest

When an event comes, it passes each tree and is valued 1(signal) or 0(background). Finally, these values are averaged.

Methods for Event Selection

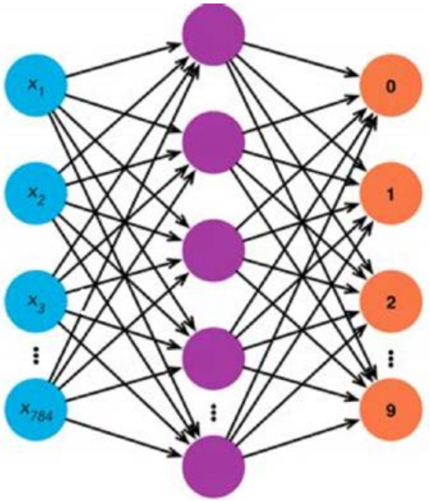
- **Cut-flow**
- **Machine Learning**
  - **Boosted Decision Tree (BDT)**



泰坦尼克号乘客能否幸存的决策树

Methods for Event Selection

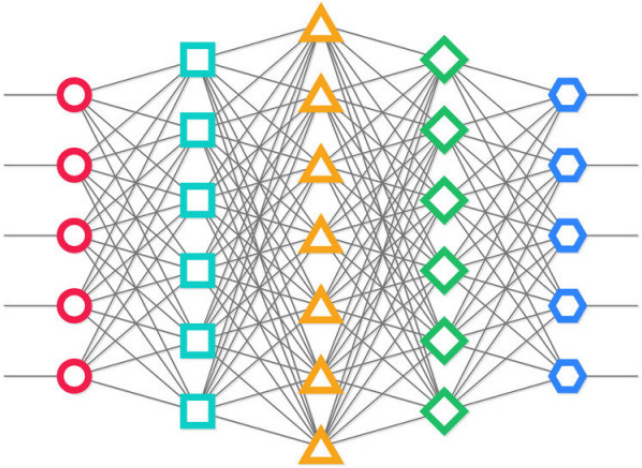
- **Cut-flow**
- **Machine Learning**
  - **Boosted Decision Tree (BDT)**
  - **Neural Networks**
    - Shallow Neural Network (NN)



Methods for Event Selection

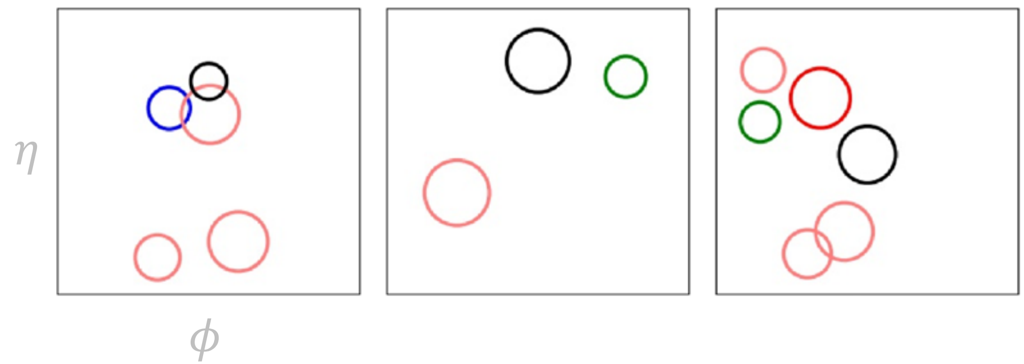
- **Cut-flow**
- **Machine Learning**
  - **Boosted Decision Tree (BDT)**
  - **Neural Networks**
    - Shallow Neural Network (NN)
    - Deep Learning
      - Deep Neural Network (DNN)

1410.3469, 1402.4735, 1803.01550



Methods for Event Selection

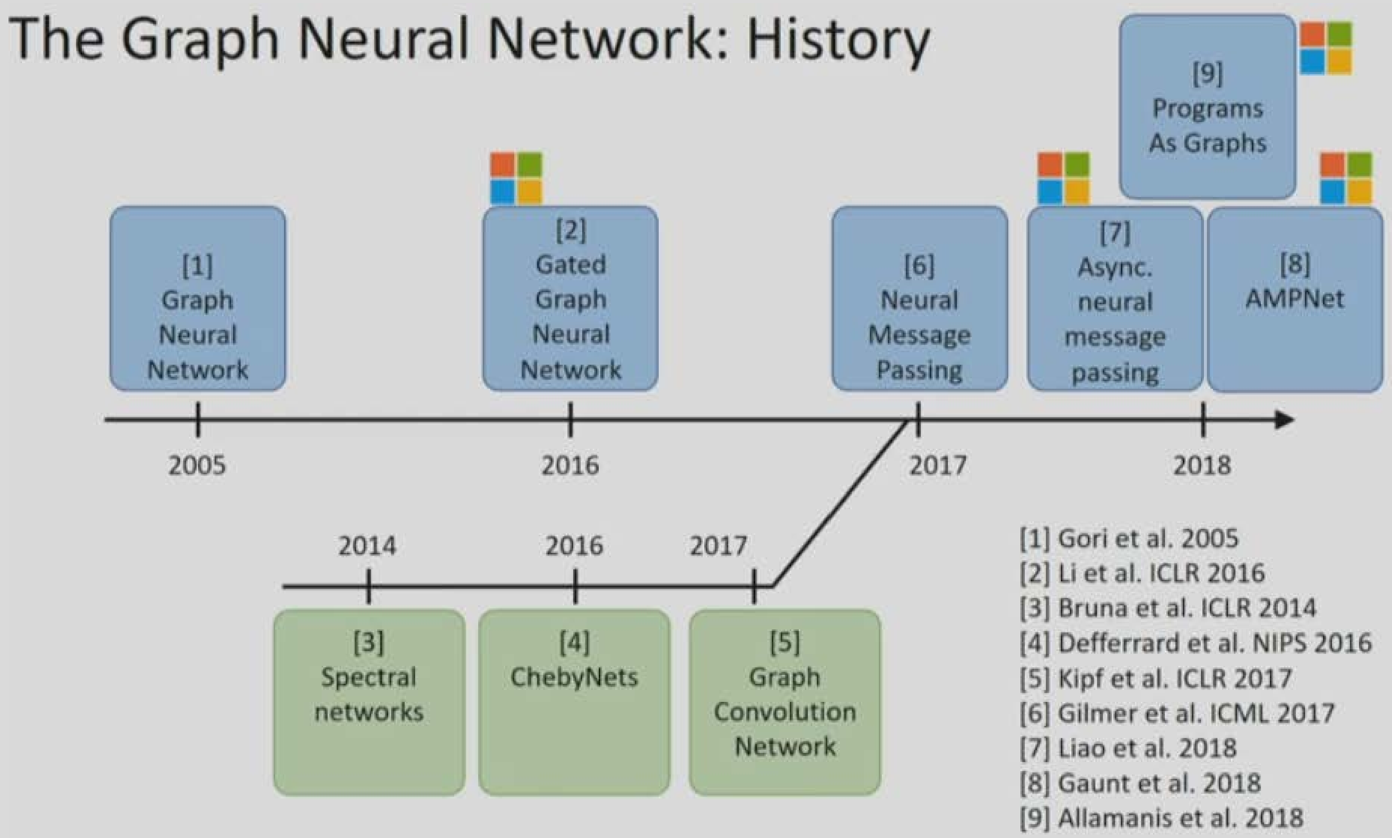
- **Cut-flow**
- **Machine Learning**
  - **Boosted Decision Tree (BDT)**
  - **Neural Networks**
    - Shallow Neural Network (NN)
    - Deep Learning
      - Deep Neural Network (DNN) 1410.3469, 1402.4735, 1803.01550
      - Convolutional Neural Network (CNN) 1708.07034



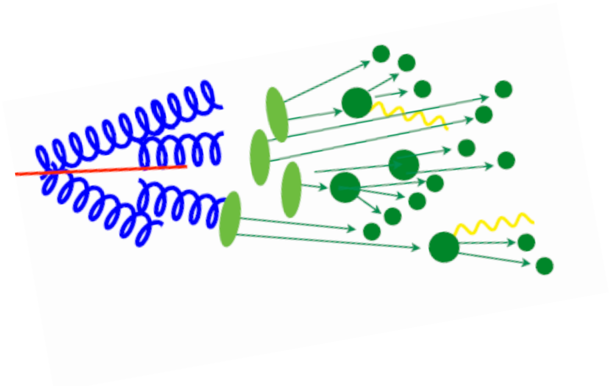
color—particle type

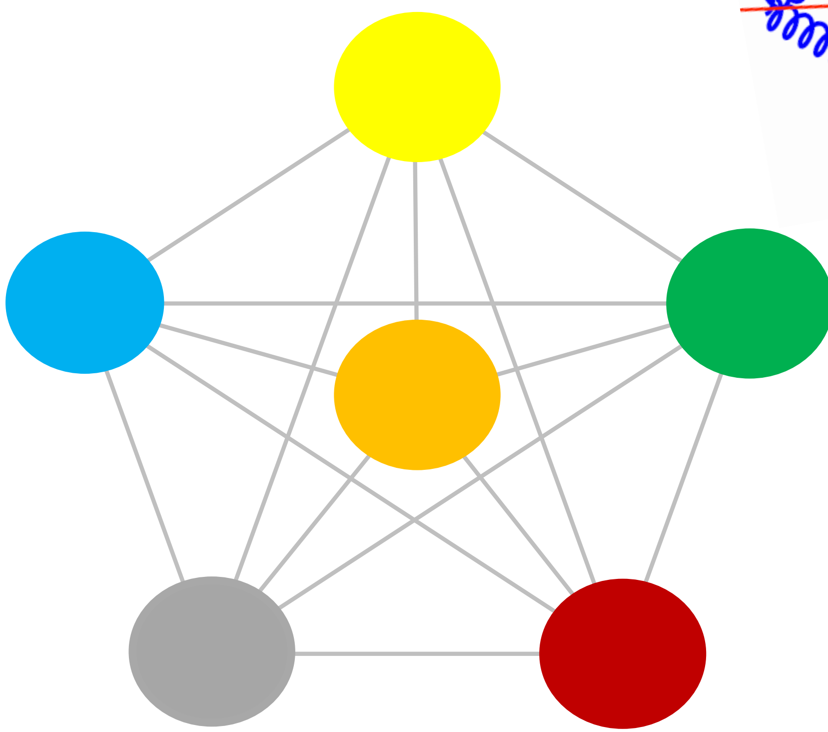
size —energy or tansverse momentum

Graph Neural Network

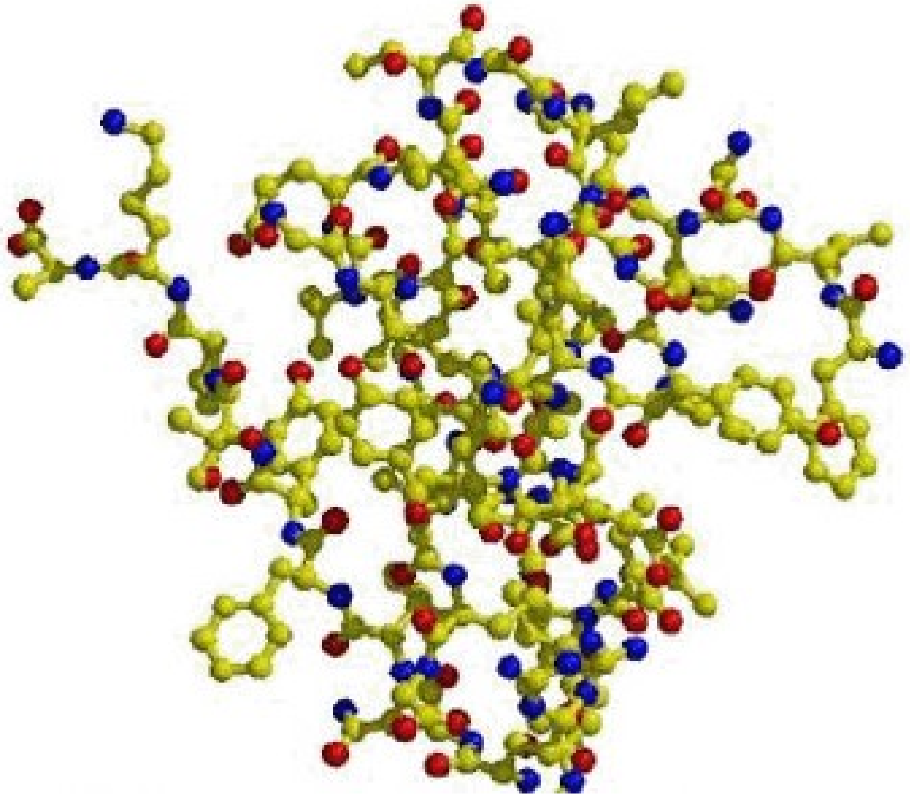


Graph Neural Network





Graph Neural Network



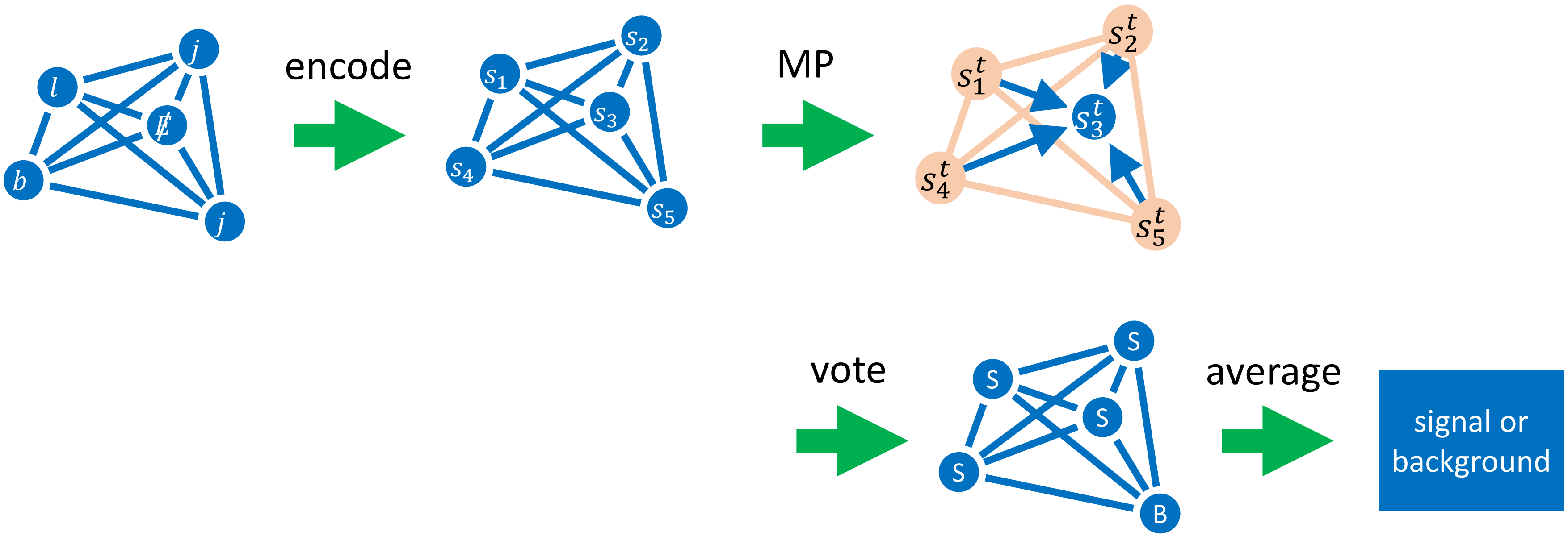
Event as Graph

Our Idea

- Represent an event as a ***graph*** $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$
- Encode each vertex into a ***state vector***
- ***Message passing*** between vertices
- Each vertex ***votes*** the signal/background
- ***Average the votes*** as the final result

$\boldsymbol{E}$: $d_{ij} \equiv \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}$

$\boldsymbol{V}$: $(0,0,1,0, m, E, P_T)$



Performance Index

Expected discovery significance is

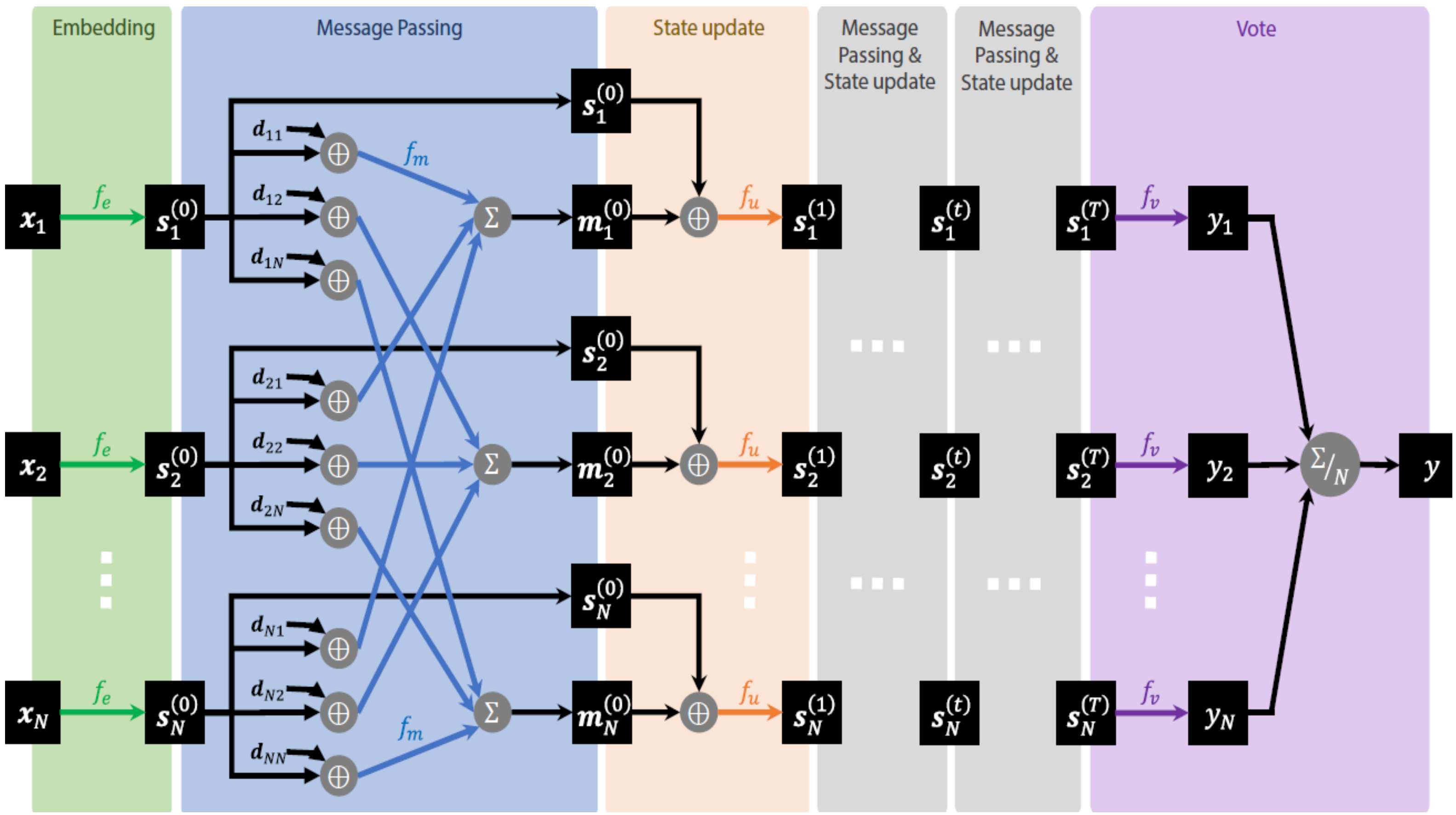
$$\frac{S}{\sqrt{B}} = \frac{\sigma_S L \epsilon_S^0}{\sqrt{\sigma_B L \epsilon_B^0}} \cdot \frac{\epsilon_S}{\sqrt{\epsilon_B}}$$

- S, B: the number of selected signal and background events
- σ: cross section
- L: integrated luminosity
- ϵ^0, ϵ: efficiencies of preselection cuts and classifier

We define the expected relative discovery significance as $\epsilon_S / \sqrt{\epsilon_B}$

Detailed operation （1）

Message Passing Neural Network



Detailed operation （2）

Neural Network Model

- Use ***one-hot-like*** encoding for object identity.
- ***30-dim*** feature vectors
- ***Distance*** measure using $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$
- Pair distances are expanded in a Gaussian basis (linearly distributed in [0, 5]) as vectors of length 21.
- Use separate message and update functions for each iteration.

- $f_e(\mathrm{id}, E, p_T) = \mathrm{relu}\left(W_e \begin{bmatrix} \mathrm{onehot}(\mathrm{id}) \\ p_T \\ E \end{bmatrix} + b_e\right)$
- $f_m^{(t)}(s, d) = \mathrm{relu}\left(W_m^{(t)} \begin{bmatrix} s \\ \mathrm{expand}(d) \end{bmatrix} + \boldsymbol{b}_m^{(t)}\right)$
- $f_u^{(t)}(s, m) = \mathrm{relu}\left(W_u^{(t)} \begin{bmatrix} s \\ m \end{bmatrix} + \boldsymbol{b}_u^{(t)}\right)$
- $f_v(\boldsymbol{s}) = \sigma(W_v \boldsymbol{s} + \boldsymbol{b}_s)$

relu: rectified linear unit (non linear trans)

expand: expand to Gausian basis to form a vector

$\boldsymbol{b_e}, \boldsymbol{b_m}, \boldsymbol{b_u}, \boldsymbol{b_s}$: parameters

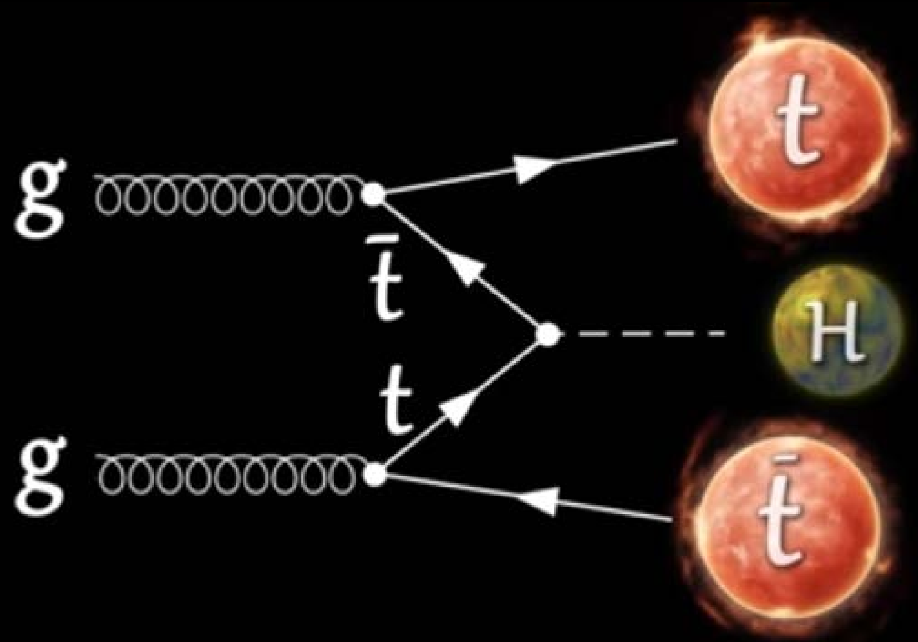
$\boldsymbol{W_e}, \boldsymbol{W_m}, \boldsymbol{W_u}, \boldsymbol{W_v}$: linear transformations

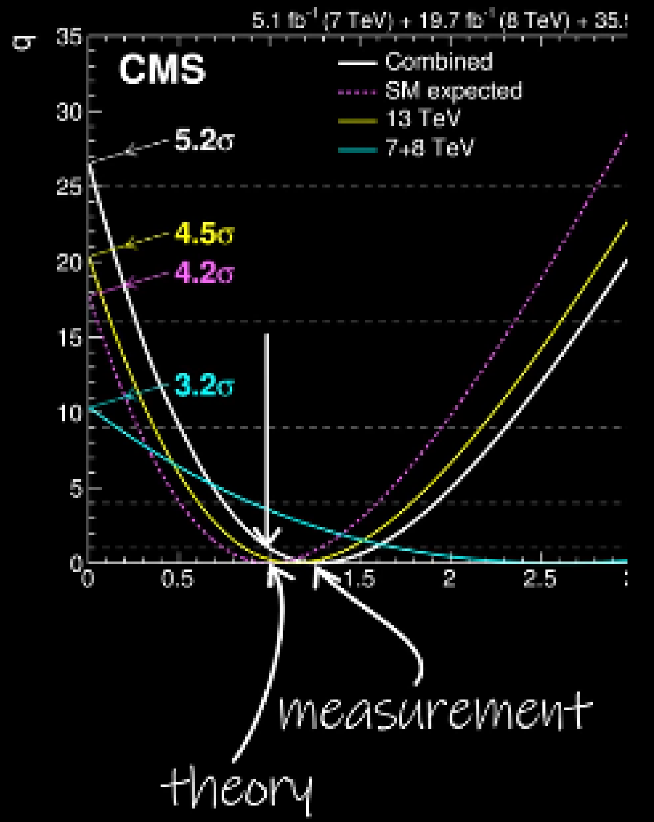
Training

- Binary Cross-Entropy (BCE) as loss function.
- Calculate gradients using error back-propagation.
- Optimize network parameters using Adam algorithm.
- Training with mini-batch of examples.
- Adopt early stopping to prevent overfitting.

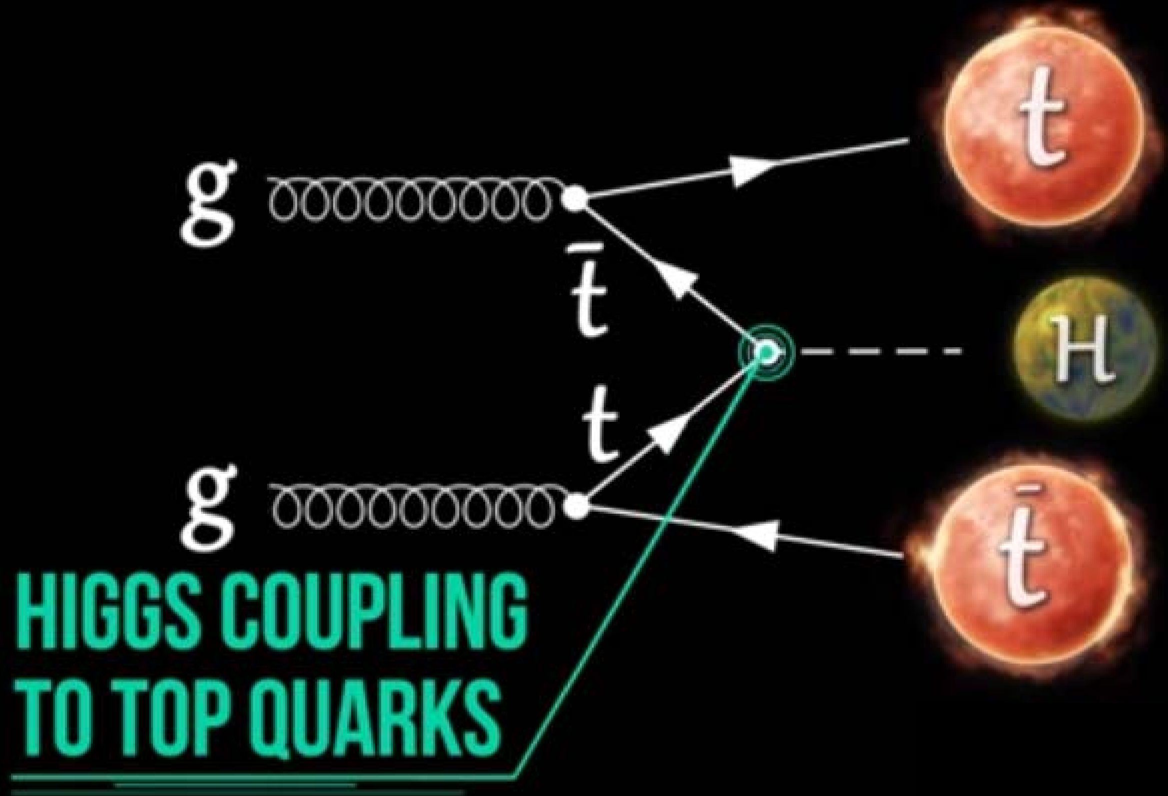
2.2 Machine learning for CP property of top-Higgs coupling at LHC

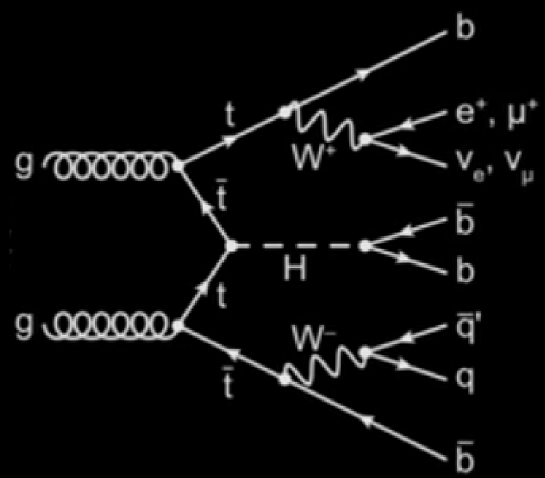
1901.05627 J Ren, L Wu, JMY

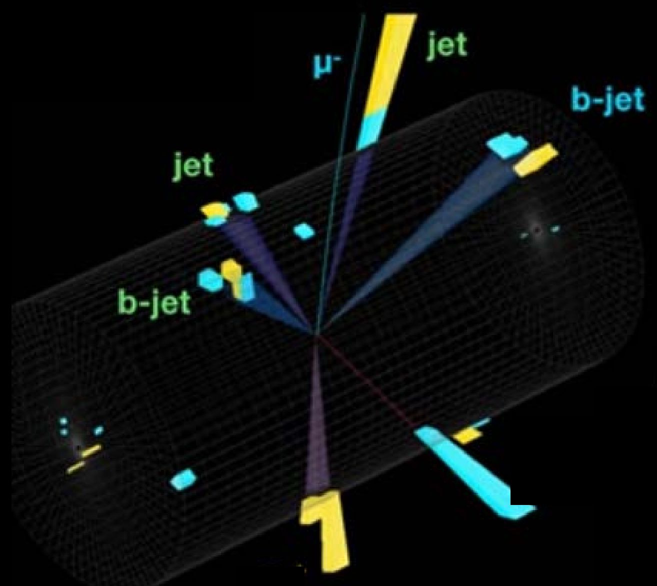




See the talks by Hualin Mei, Hongtao Yang, et al


g
g
t̄
t
t
H
t̄
HIGGS COUPLING
TO TOP QUARKS

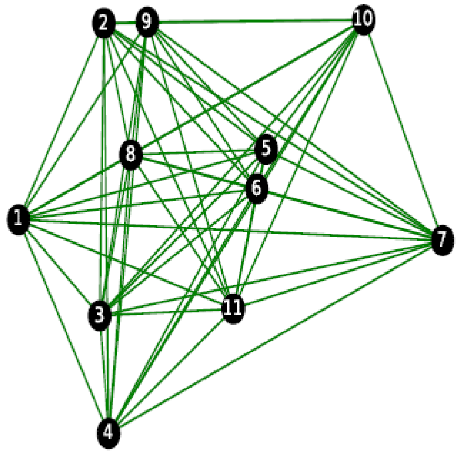

b
t
e+, μ+
W+
ν_e, ν_μ
g
t̄
b̄
H
b
t
g
q̄'
W−
q
t̄
b̄


μ-
jet
b-jet
jet
b-jet

$$pp \rightarrow t\bar{t}H \;\; (H = h) \rightarrow \bar{t}t\, b\bar{b}$$

$$pp \rightarrow t\bar{t}H \;\; (H = A) \rightarrow \bar{t}t\, b\bar{b}$$

$pp \rightarrow t\bar{t}\;\; b\bar{b}$ (background)



x

1	0	-1	0	0	0.0160	0.0679	-0.0001
2	0	0	1	0	0.3265	0.6360	0.0587
3	0	0	1	0	0.1956	0.3934	0.0187
4	0	0	-1	0	0.1703	0.3179	0.0215
5	0	0	-1	0	0.1520	0.1605	0.0113
6	0	0	-1	0	0.1442	0.1491	0.0174
7	0	0	-1	0	0.0827	0.3333	0.0110
8	0	0	1	0	0.0452	0.0709	0.0068
9	0	0	1	0	0.0373	0.0519	0.0059
10	0	0	-1	0	0.0282	0.0552	0.0043
11	0	0	0	1	0.2154	0.2154	0.0000

d

	1	2	3	4	5	6	7	8	9	10	11
1	0.0000	2.8578	1.5566	3.1012	2.6385	2.3965	4.2066	1.4305	3.0275	4.4020	2.4592
2	2.8578	0.0000	2.2198	0.5844	2.3785	2.7535	4.5119	1.8490	0.4306	2.5726	2.6467
3	1.5566	2.2198	0.0000	1.6393	2.8377	2.3509	3.5506	2.2553	2.2532	3.3985	1.3257
4	3.1012	0.5844	1.6393	0.0000	2.8085	3.2339	4.2561	2.4220	0.6836	2.5808	2.1275
5	2.6385	2.3785	2.8377	2.8085	0.0000	0.5545	2.1564	1.3419	2.1250	2.0419	2.2379
6	2.3965	2.7535	2.3509	3.2339	0.5545	0.0000	1.9747	1.3322	2.5569	2.5738	1.6836
7	4.2066	4.5119	3.5506	4.2561	2.1564	1.9747	0.0000	3.3068	4.2135	3.1631	2.2780
8	1.4305	1.8490	2.2553	2.4220	1.3419	1.3322	3.3068	0.0000	1.8527	2.9715	2.3663
9	3.0275	0.4306	2.2532	0.6836	2.1250	2.5569	4.2135	1.8527	0.0000	2.1421	2.4522
10	4.4020	2.5726	3.3985	2.5808	2.0419	2.5738	3.1631	2.9715	2.1421	0.0000	2.6109
11	2.4592	2.6467	1.3257	2.1275	2.2379	1.6836	2.2780	2.3663	2.4522	2.6109	0.0000

FIG. 1. Event graph with detailed node features and edge weights for a specific simulated $t\bar{t}h$ event.

For each event:

- each node i gives 3 probabilities $(p_i)_k$ for $t\bar{t}h$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$
- average over all the nodes as the final output

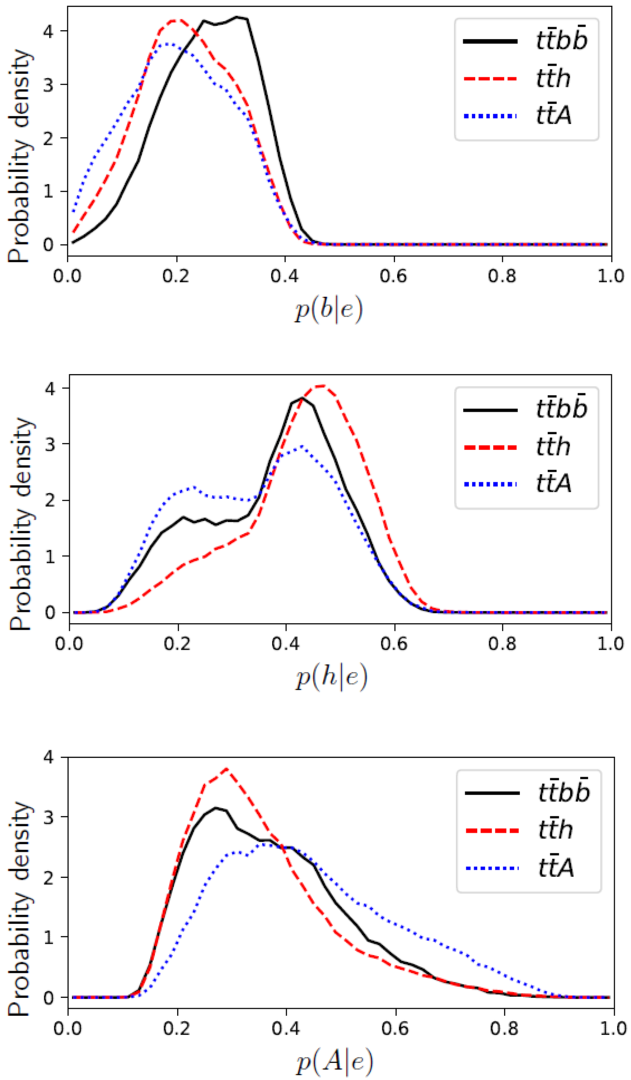
$$\frac{1}{N}\sum_i (p_i)_k \quad \left\{ \begin{array}{l} p(h|e) \\ \\ p(A|e) \\ \\ p(b|e) \end{array} \right.$$

For each event sample D:

$$L_h(D) = \prod_{e \in D}{}' p(h|e)$$

$$L_A(D) = \prod_{e \in D}{}' p(A|e)$$

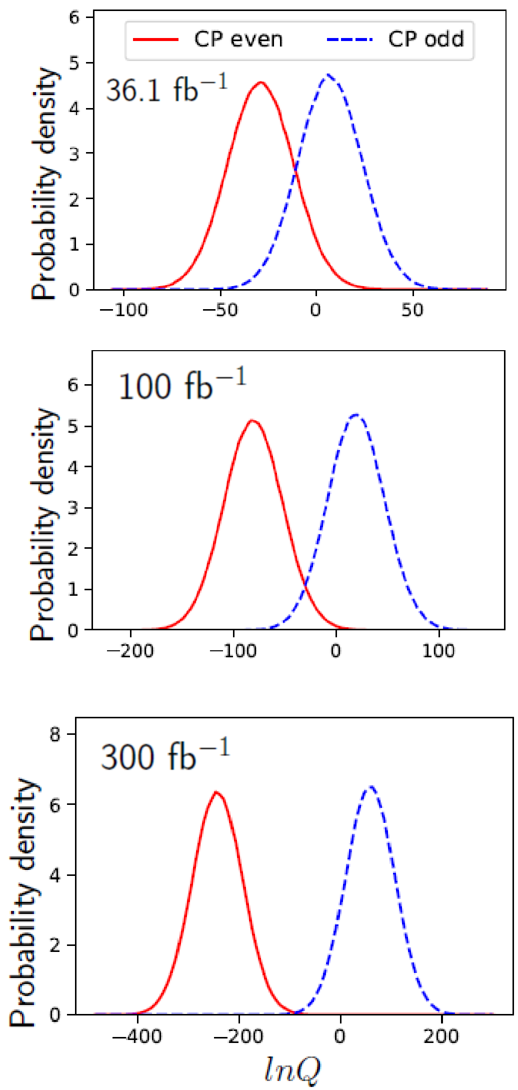
$$Q(D) = \frac{L_A(D)}{L_h(D)}$$



- each node i gives 3 probabilities $(p_i)_k$ for $t\bar{t}h, t\bar{t}A$ and $t\bar{t}b\bar{b}$
- average over all the nodes as the final output

$$\frac{1}{N}\sum_i (p_i)_k \left\{ \begin{array}{l} p(h|e) \\ p(A|e) \\ p(b|e) \end{array} \right.$$

The MPNN has indeed learned some discriminative features for different processes:
The background $t\bar{t}b\bar{b}$ events tend to have higher p(b|e);
The $t\bar{t}h$ events tend to have higher p(h|e);
The $t\bar{t}A$ events tend to have higher p(A|e)



For each event:

- each node i gives 3 probabilities $(p_i)_k$ for $t\bar{t}h$, $t\bar{t}A$ and $t\bar{t}b\bar{b}$
- average over all the nodes as the final output

$$\frac{1}{N}\sum_i (p_i)_k \quad \begin{cases} p(h|e) \\ p(A|e) \\ p(b|e) \end{cases}$$

For each event sample D:

$$L_h(D) = \prod_{e\in D}{}' p(h|e)$$

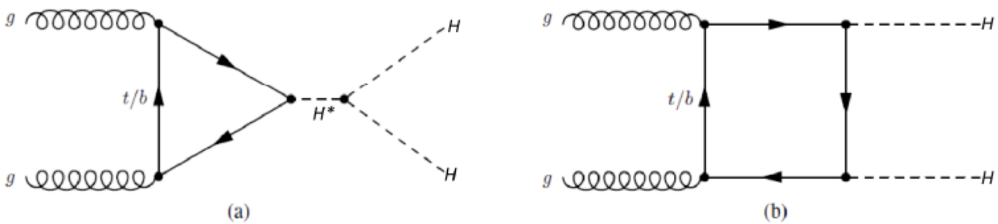
$$L_A(D) = \prod_{e\in D}{}' p(A|e)$$

$$Q(D) = \frac{L_A(D)}{L_h(D)}$$

The overlap between the two distributions reduces with increasing luminosity. When the luminosity is 300 fb^{-1} , the two distributions have nearly no overlap, which means that the CP nature of top-Higgs coupling can be determined.

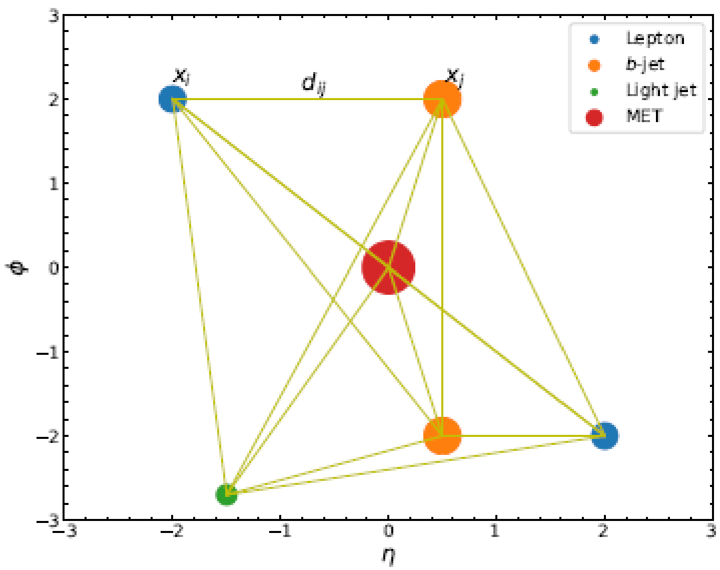
2.3 Machine learning for triple-Higgs coupling at LHC

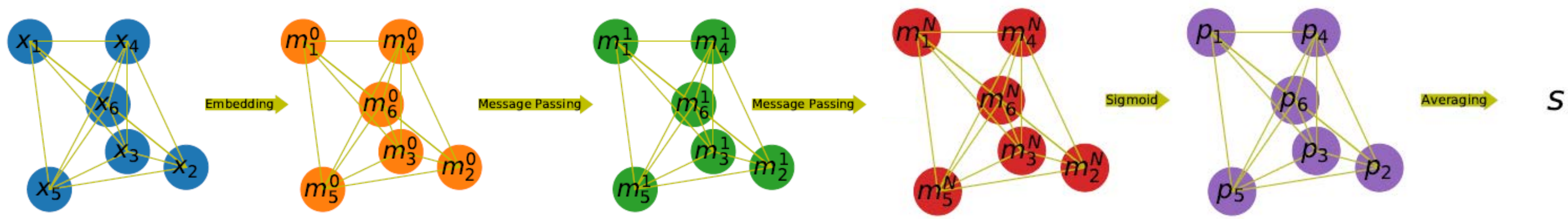
Abdughani, Wang, Wu, JMY, Zhao, 2005.11086

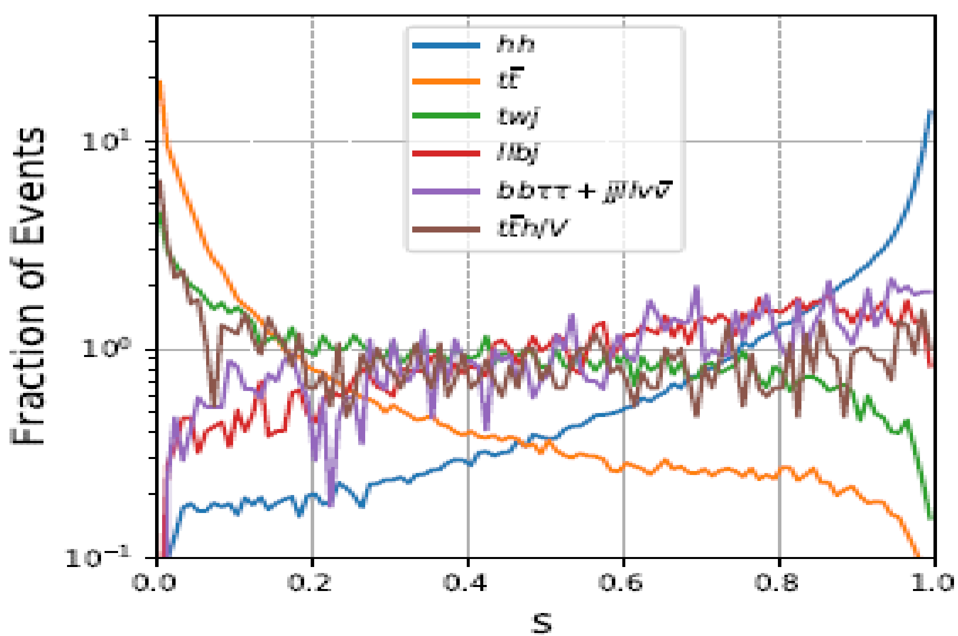


See the talks by Kunlin Ran, Zihang Jia, Junmou Chen, Mellado et al

$$pp \to hh \to b\bar{b}WW* \quad \to \quad 2b + 2\ell + E_{\mathrm{T}}^{\mathrm{miss}}$$





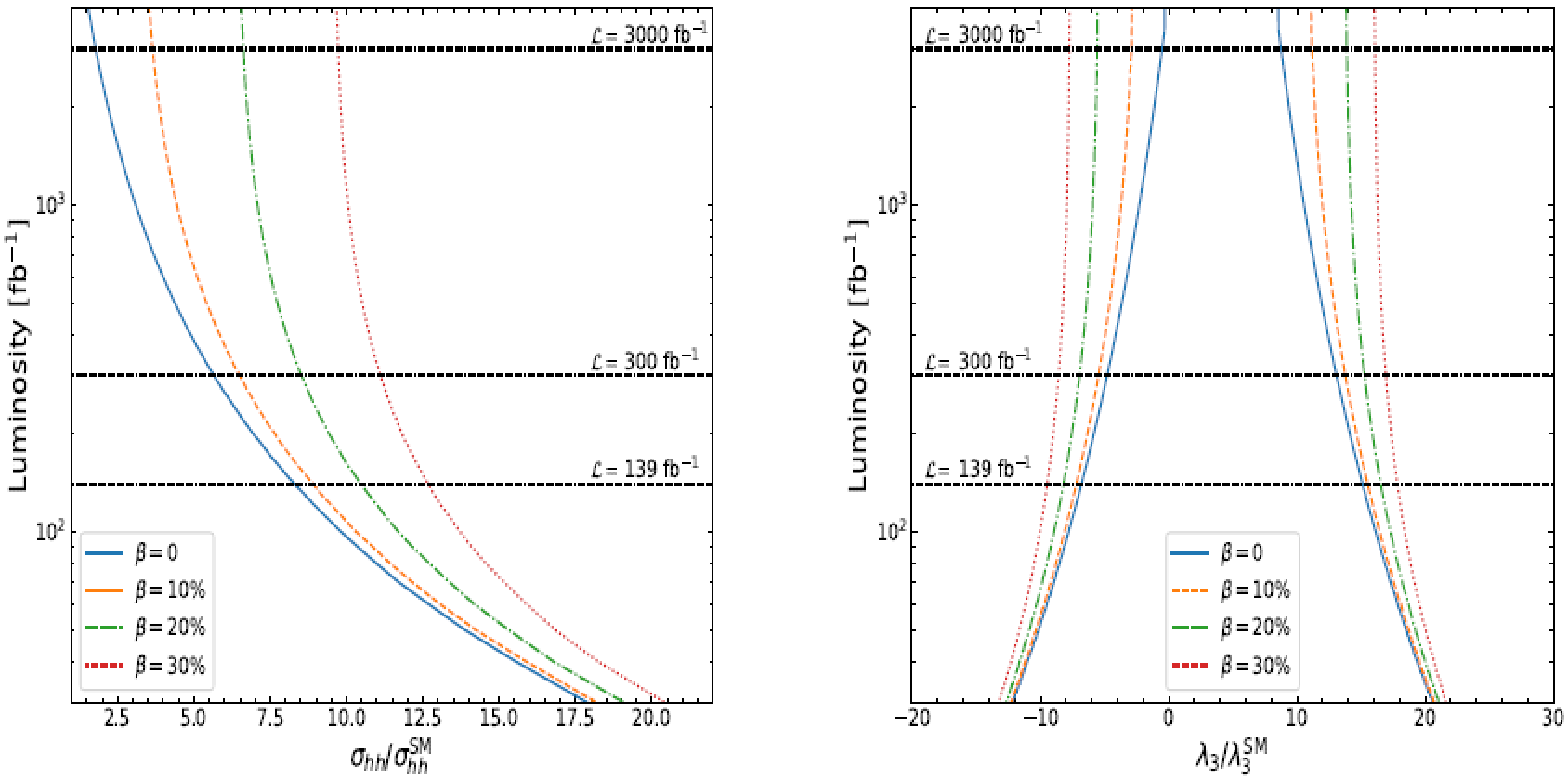


The MPNN training results for the signal (hh) and backgrounds.
The event fractions of signal and background versus the final score s

	hh	$t\bar{t}$	$tW+j$	$\ell^+\ell^-bj$	$t\bar{t}h$	$\tau^+\tau^-b\bar{b}$	$t\bar{t}V$	$jj\ell^+\ell^-\nu\bar{\nu}$	$\alpha(\sigma)$	S/B
No cut	40.7 [127]	953600 [128]	123200	117100 [140]	661.3 [141]	29070 [140]	1710 [142]	48200 [a]	$\simeq 0$	$\simeq 0$
Baseline cuts	0.0105	1.8568	0.2189	0.0675	0.0247	0.0246	0.0153	0.0101	0.3876	0.0047
MPNN	0.0067	0.0581	0.0180	0.0152	0.0080	0.0025	0.0018	0.0017	1.13	0.06

$$\alpha = S/\sqrt{B + (\beta B)^2}$$

Signal and background cross sections in fb at 14 TeV HL-LHC with luminosity 3000 fb^{-1}, before hadron-level cuts but after baseline cuts and after MPNN validation process requiring the signal events number $Nsig = 20$ to have reasonable statistics. The significance is calculated by using $\beta = 0$.



The 2σ upper bounds on production cross section of the Higgs pair and triple Higgs coupling at 14 TeV LHC

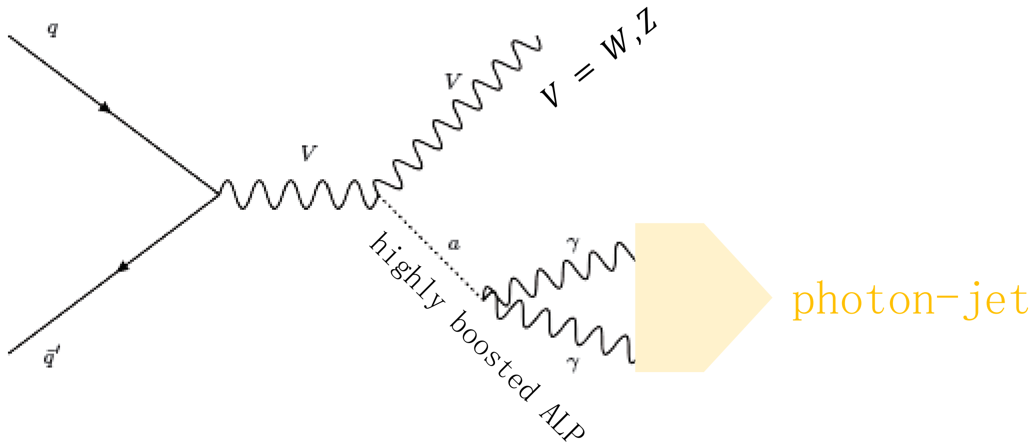
Summary

We apply the Message Passing Neural Network (MPNN) to the study of non-resonant Higgs pair production process $pp \to hh$ in the final state with $2b + 2\ell + E_T^{\rm miss}$ at the LHC. Although the MPNN can improve the signal significance, it is still challenging to observe such a process at the LHC. We find that a 2σ upper bound (including a 10% systematic uncertainty) on the production cross section of the Higgs pair is 3.7 times the predicted SM cross section at the LHC with the luminosity of 3000 $\rm fb^{-1}$, which will limit the triple Higgs coupling to the range of $[-3, 11.5]$.

图形网络用于LHC上的双希格斯产生，发现用模拟的信号和背景数据训练出来的图形网络可以提高信号的统计显著性，但是发现在高亮度的HL-LHC上信号的统计显著性还是达不到发现的标准而只能给出产生截面的上限（标准模型所预言截面的3.7倍）和自耦合的限制区域（自耦合大于-3小于11）

2.4 Machine learning for an ALP at LHC

Ren, Wang, Wu, JMY, Zhang, 2106.07018



$$pp \to W^{\pm}(\to \ell^{\pm}\nu)a(\to \gamma\gamma) \qquad pp \to Z(\to \ell^{+}\ell^{-})a(\to \gamma\gamma)$$

SM backgrounds: $V+\gamma$, $V+j$, QCD di-jets

How to distinguish a photon-jet from a single photon or QCD jet ?

We use convolutional neural network (CNN) to identify photon-jet

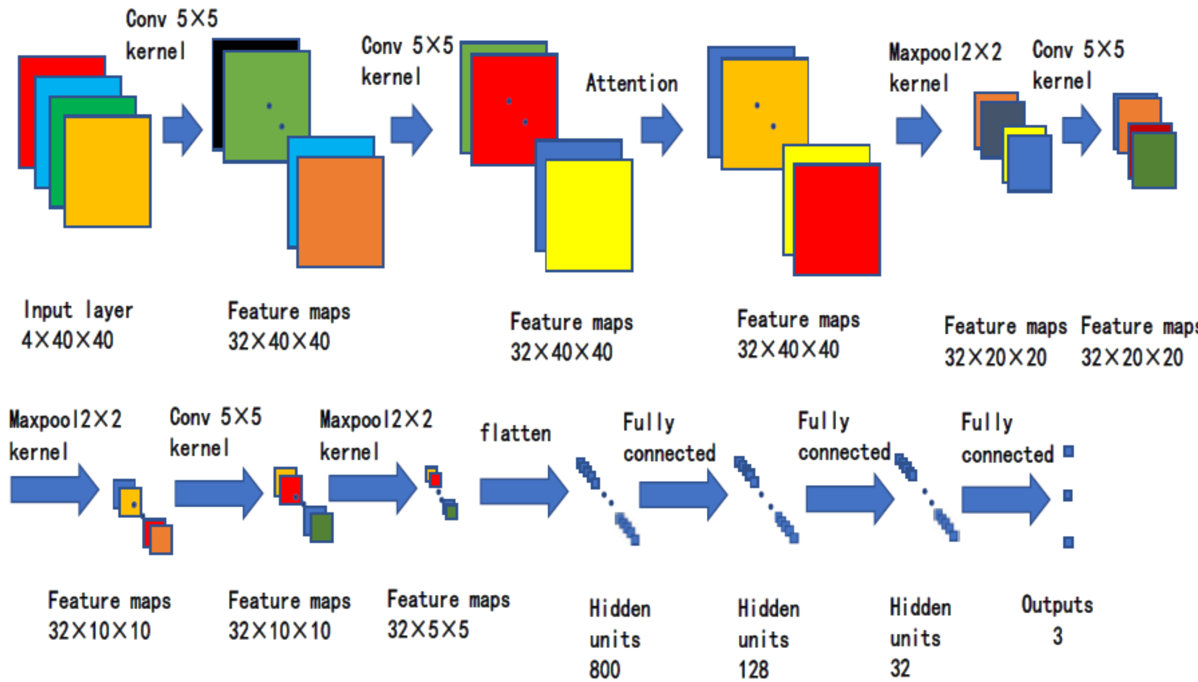
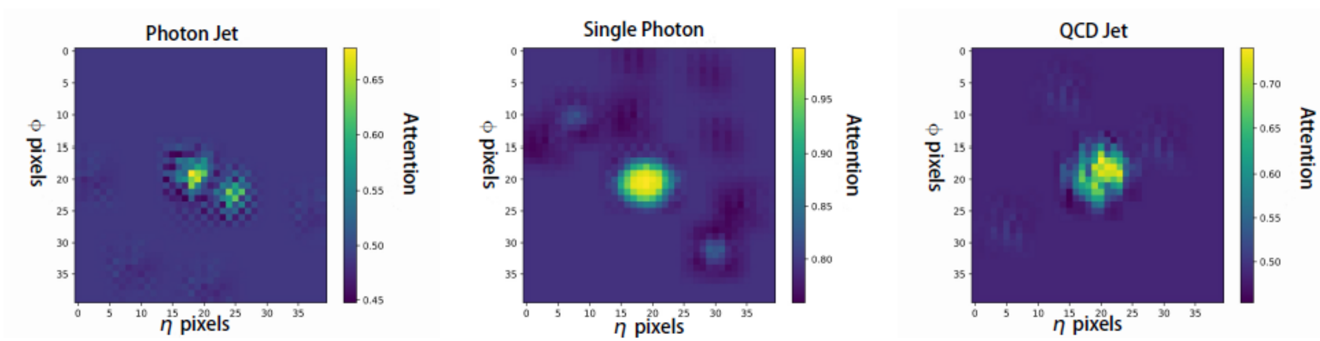
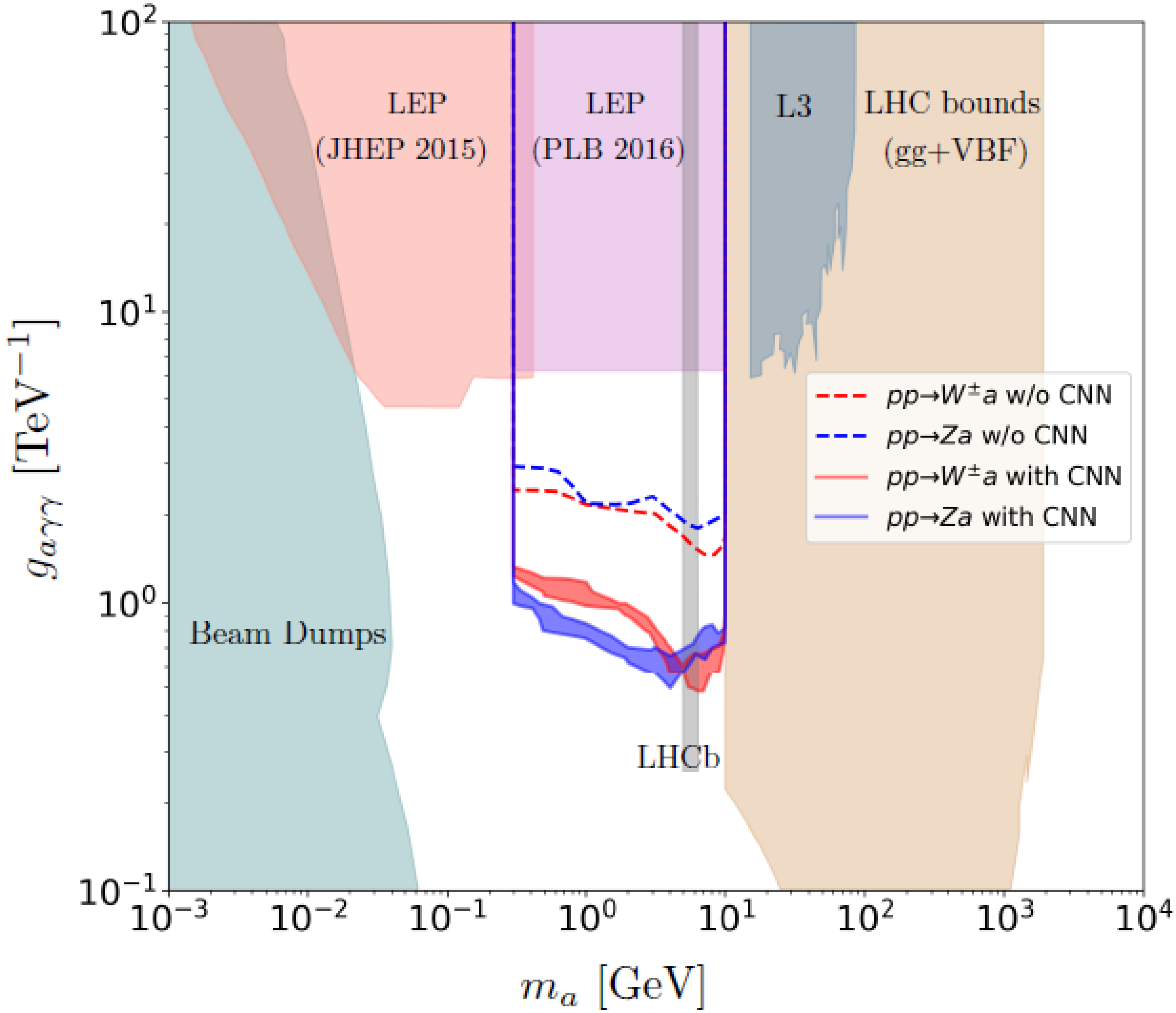


Illustration of the jet-tagging CNN



Attention obtained by the training samples for ALP of 3 GeV



LEP
(JHEP 2015)
LEP
(PLB 2016)
L3
LHC bounds
(gg+VBF)
Beam Dumps
LHCb
$pp \to W^{\pm}a$ w/o CNN
$pp \to Za$ w/o CNN
$pp \to W^{\pm}a$ with CNN
$pp \to Za$ with CNN
$g_{a\gamma\gamma}$ [TeV^{-1}]
m_a [GeV]
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}

Conclusion

Machine learning is useful for Higgs physics

-- Graph neural network (GNN) for $Ht\bar{t}$ production at LHC
can help distinguish CP-even h from CP-odd A

-- Graph neural network (GNN) for HH production at LHC
can enhance signal significance

-- Convolutional neural network (CNN) for ALP production at LHC
can enhance signal significance

Thank you for your attention !