ML at IHEP ATLAS: Status and plans

Yaquan Fang (IHEP) on behalf of IHEP ATLAS team

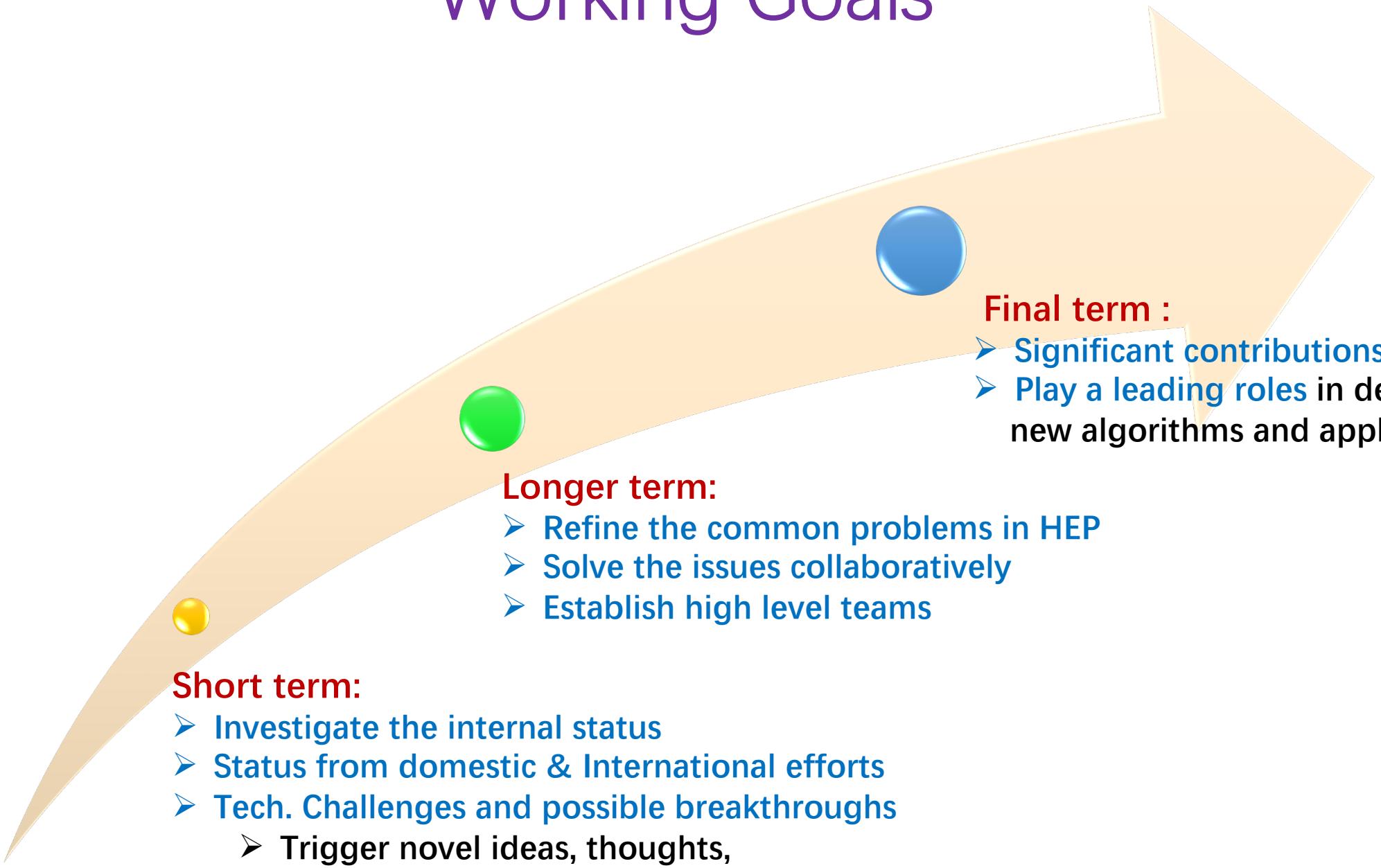
机器学习技术在高能物理中的应用研讨会

September 18th,2022

Motivation

- Various developments and progresses are distributed in different projects (e.g. LHC, BES, Juno …).
  - However, the efforts are scattering.
- Is it possible to work together to solve the common needs systematically for each project?
  - Identify the frontiers of ML in particle physics, especially in the facilities of the big sciences?
  - Are there any common issues & technologies?
- Can we play a leading role/roles in ML/AI applications in high energy physics?
  - So far, we are mostly following
  - What can we do to make significant scientific contributions to ML/AI at the worldwide level?
  - How to quantify our leaderships?

Working Goals



Final term :

- Significant contributions
- Play a leading roles in developing new algorithms and applications.

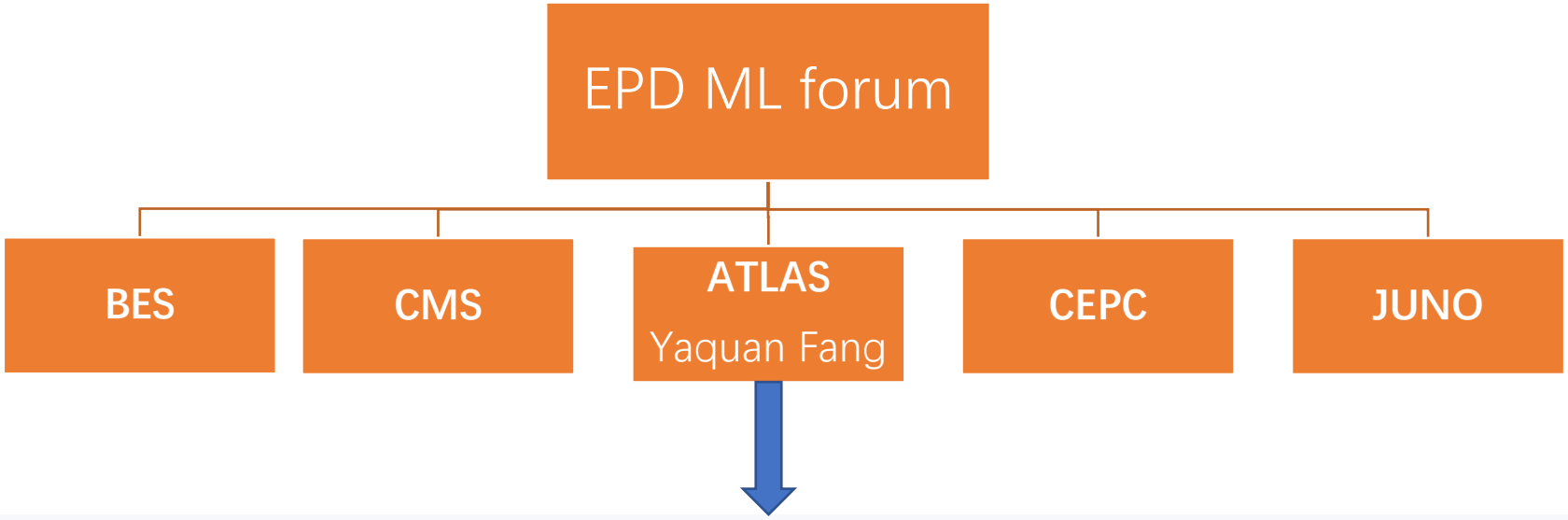
Longer term:

- Refine the common problems in HEP
- Solve the issues collaboratively
- Establish high level teams

Short term:

- Investigate the internal status
- Status from domestic & International efforts
- Tech. Challenges and possible breakthroughs
  - Trigger novel ideas, thoughts,

Plan to establish the efforts



Participate the discussions:

- ✓ Identify physics issues.
- ✓ Human resources : welcome junior faculties, postdocs and students.
- ✓ Computing resources needed.
- ✓ Cooperation with ATLAS ML forum.
- ✓ Cooperation with other ML societies.
- ✓ Others

Mattermost: https://mattermost.ihep.ac.cn/atlas/channels/machine-learning

ATLAS Machine Learning Forum (AML)

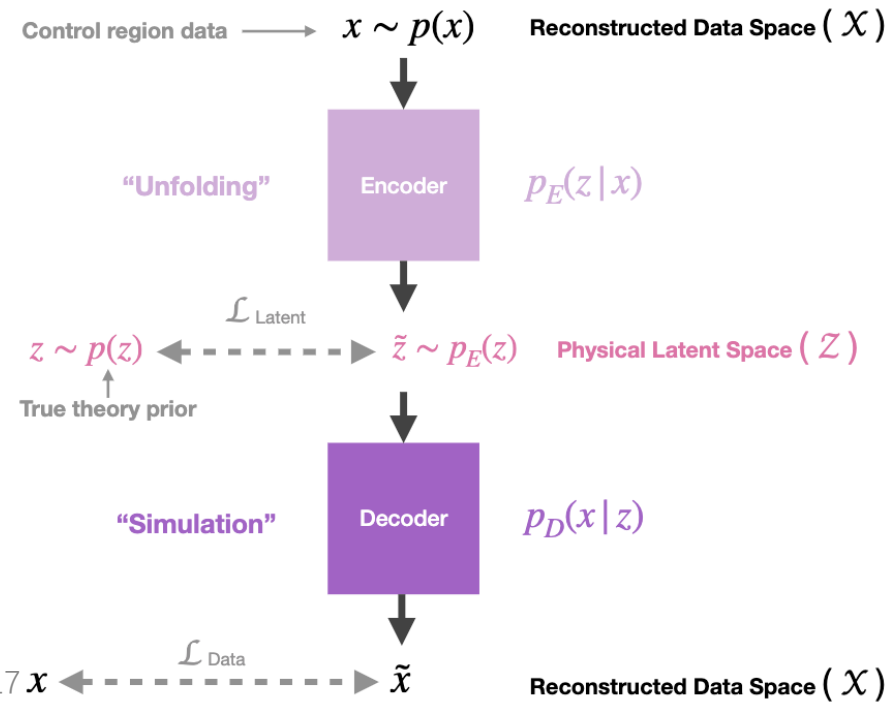
- **This forum is jointly organized by Software and Computing (S&C) and Physics and Computing Coordinators for the goals of:**
  - Communications and liaisons inside/outside ATLAS:
    - **Organizing ML workshops every year and the bi-week meetings.**
      - **Topics in the bi-weekly meetings: Reports from physics group liaisons; specific applications.**
    - **Organizing tutorials for ATLAS users.**
    - **Liaising with other groups inside and outside ATLAS, facilitating collaboration with Machine Learners.**
    - **Forum for all ATLAS people: atlas-talk, twiki.**
  - Supporting ML tools and techniques:
    - **Coordinating development of common ML tools.**
    - **Establishing standards where useful.**
    - **Providing advice on new ML tools and best tools for specific applications.**

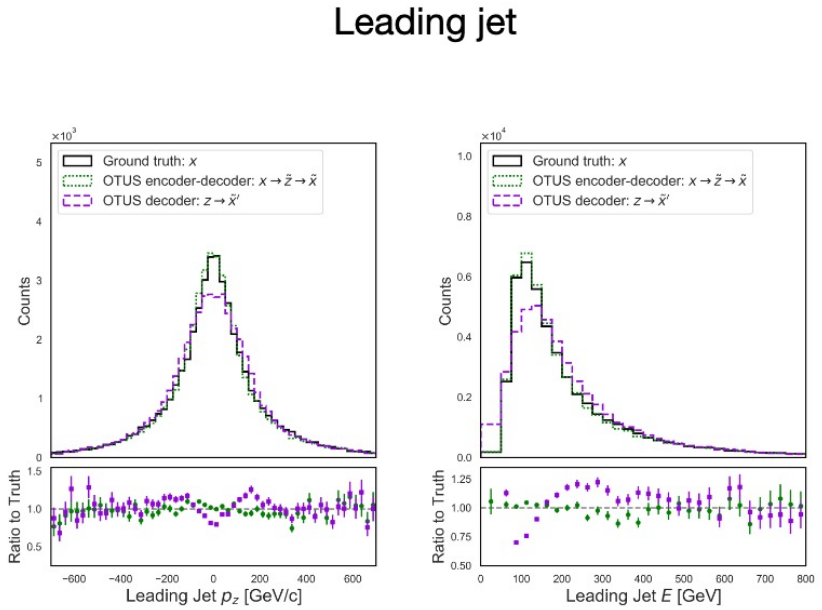
Machine Learning application (ATLAS)

- **Method and packages: encourage people to use non-HEP tools [twiki].**
  - Interexperimental Machine Learning Pages and resource lists.
  - **Do NOT recommend running trainings within analysis environment, but:**
    1. Produce a minimal "flat" dataset using the standard ATLAS workflow
    2. Run training in a stand-alone environment
    3. (Optionally) port trained algorithms back to Analysis/Reconstruction releases via ONNXruntime. [tutorial]
  - **Computing resources at CERN:**
    - Tools for flat dataset production: HDF5Utils for DAOD/AOD and uproot for ntuples.
    - For distributed training: intelligent Data Delivery Service.
    - Docker images: ATLAS Machine Learning docker, minimalist images.
    - GPU and notebook: JupyterHub at CERN and GPU queues.

ML application in ATLAS （1）

- Optimal Transport based Unfolding and Simulation (OTUS) arxiv: 2101.08944
  - A data-driven ML simulator with altering Variational Auto-encoders (VAEs): predict the **reco-level data (X)** from **parton interactions (Z)**. GANs only mimic the X but not learn the transformation $Z \rightarrow X$
  - Design the loss: Latent loss + Data loss + any additional physically motivated constraints.
  - Extra bonus: unfolding mapping from data to truth.

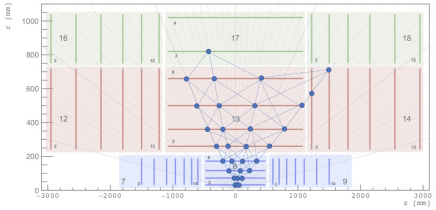


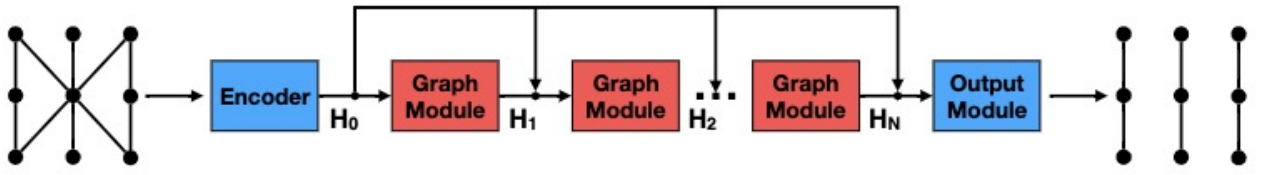


$pp \rightarrow t\bar{t}$

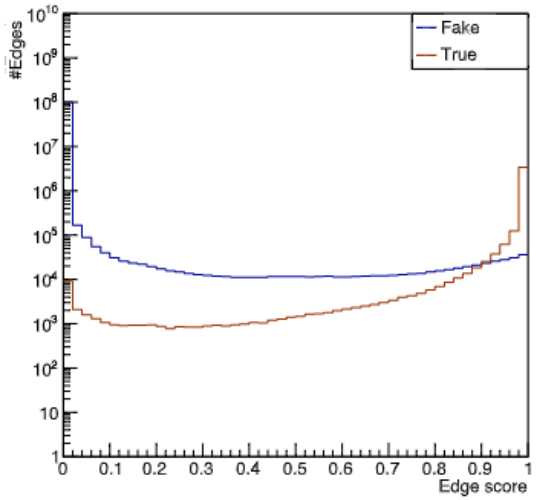
ML application in ATLAS （2）

- Track reconstruction algorithm with GNN in HL-LHC:
  - Connect the hits and select the truth track from all connections (edge in graph). arXiv:2103.0091
  - Reduce the connections with detector module maps and geometric cuts for less memory.

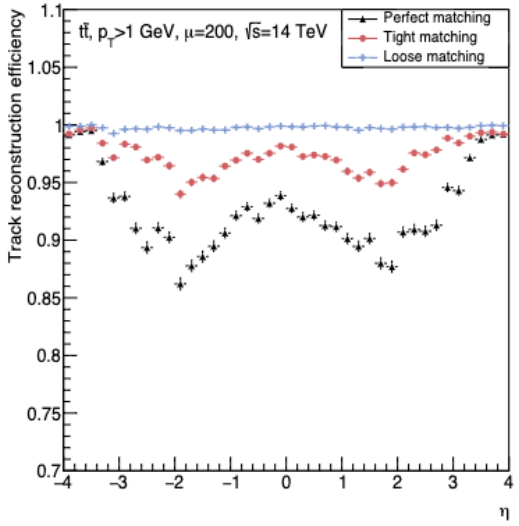




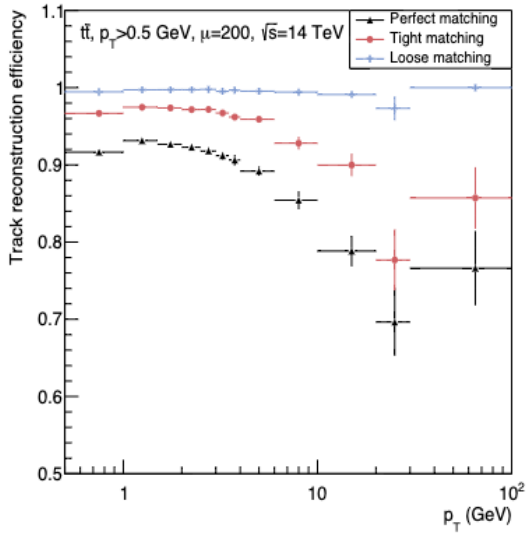
(arXiv:2003.11603)



Edges classification score (on a test dataset)



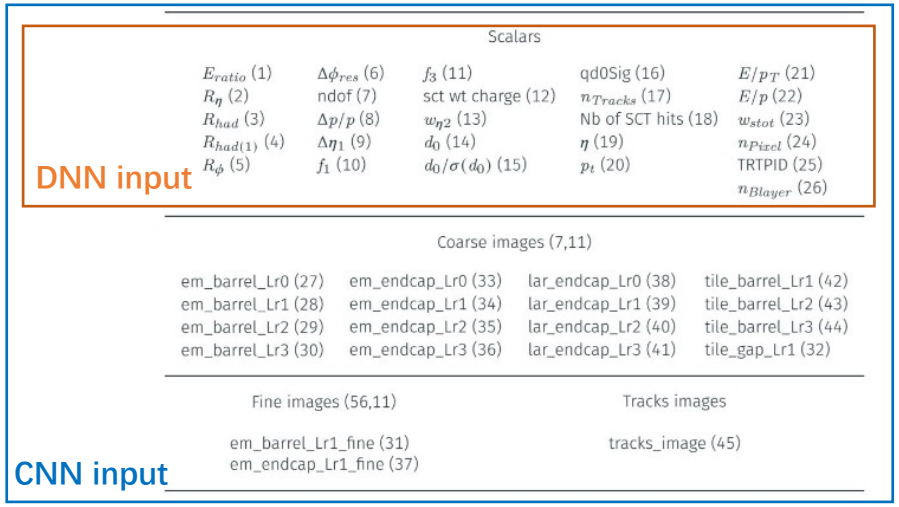
Track reconstruction efficiency vs. η

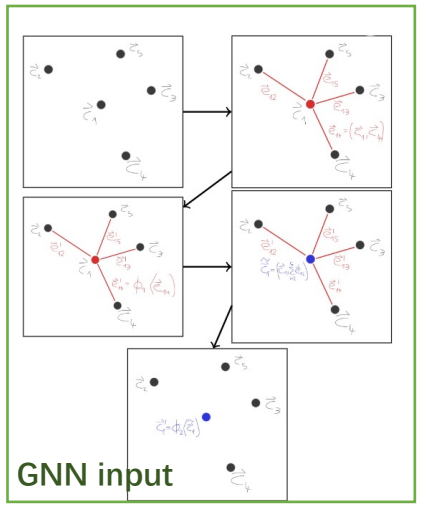


Track reconstruction efficiency vs. p_T

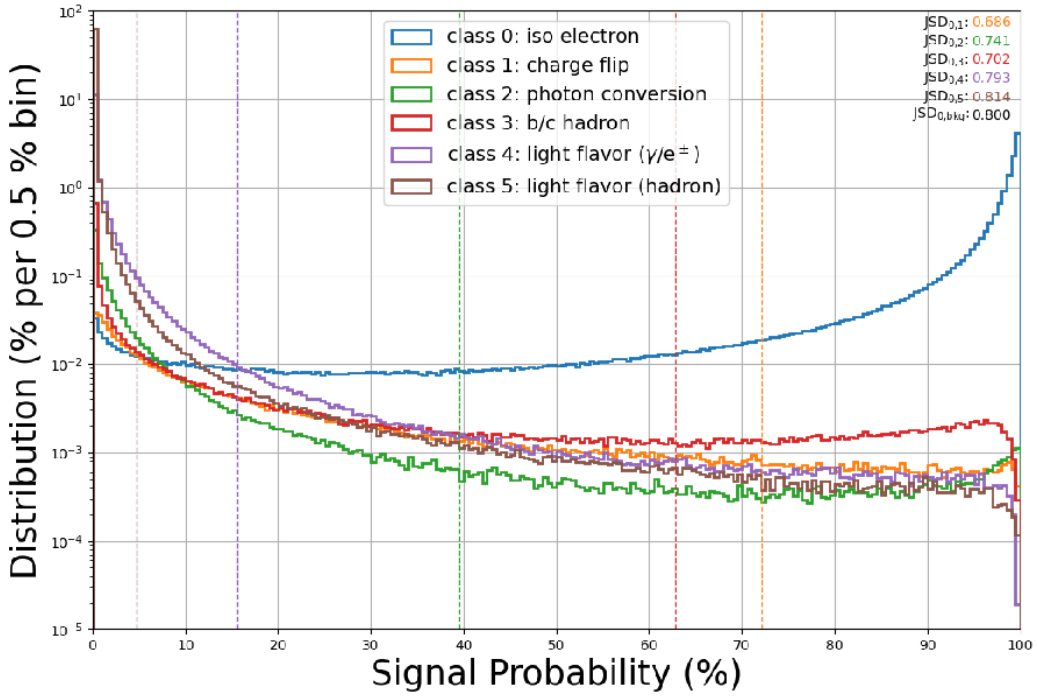
ML application in ATLAS (3)

- Object identification with ML:
  - Electron ID with high level features + DNN, low level features + CNN/GNN.
  - Jet tagging for b-jet, c-jet, di-tau, gluon and bosons.
  - Similar procedure: select the input information and proper method, tune it and get the result.
  - Take e-ID as example:
    - High level info + multi-class: DNN.
    - Low level cell info (hit maps): CNN.
    - Low level cell info (hit structure): GNN.





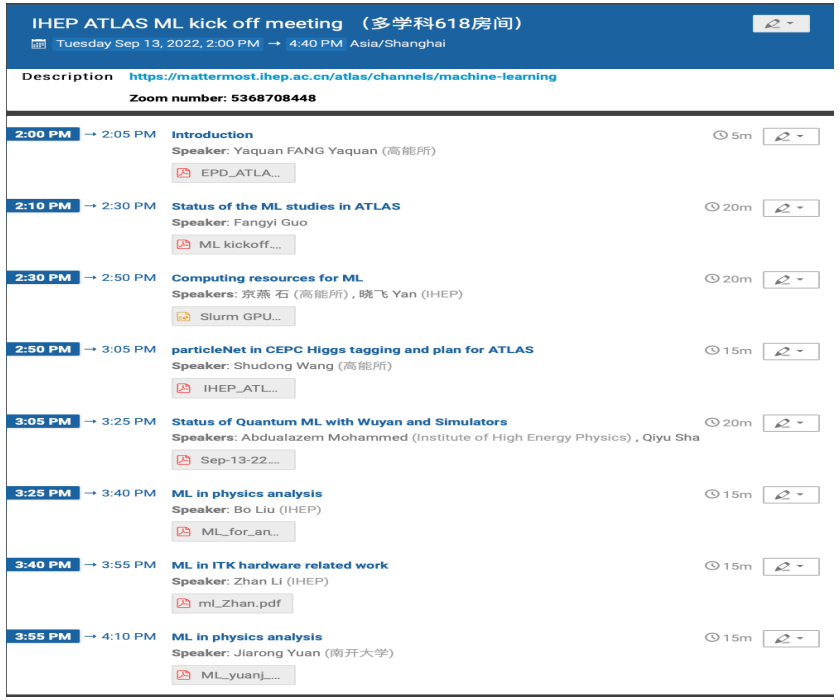
CNN classification performance



Activities

IHEP ATLAS ML kickoff meeting

https://indico.ihep.ac.cn/event/17736/ **Sep 13th**



Tutorial of particleNet **Sep 12th**

Wechat group:

Mattermost channel:

https://mattermost.ihep.ac.cn/atlas/channels/machine-learning

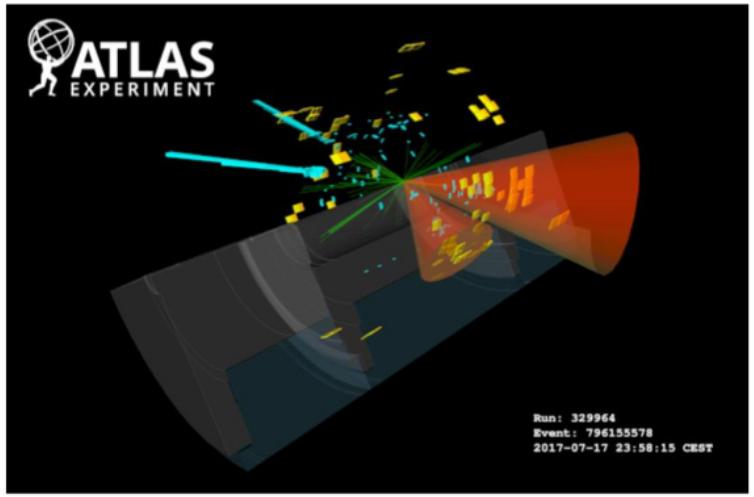
Some thoughts and plans (Need to be expanded by you)

- **ML with low statistics.**
  - VBF Higgs→ $\gamma\gamma$ analysis
- **Quantum ML** (talk from Abdualazem/Qiyu)
  - Hardware vs Simulator
  - Develop effective algorithms
- **Implementation of latest ML methods to particle tagging** (talk from Sudong)
- **Add more ideas at the Mattermost.**
  - ……
  - ……
- **Hardware related ML** (talk from Zhan)
- **Computing resources** （talk from Jingyan）

Plan for boson tagger

Shudong Wang

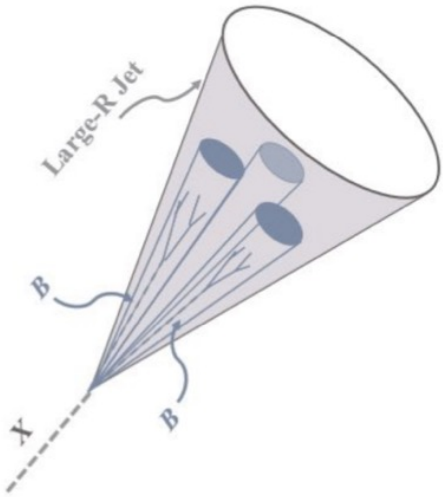
- Get in touch with jet tagging conveners (done)
  - Request for useful information (data samples, standard procedure, current situation, etc.)
- Get familiar with related works in the past (on-going)
  - boson taggers in the past used in both ATLAS & CMS
  - newly released ATLAS note about constituent-based top tagger
  - …
- Get started to work on it
  - Understand physics behind the samples and data structure of them
  - Use ParticleNet or even more powerful & state-of-the-art techs (e.g. Particle Transformer) to explore the performance of these tools on boson tagging tasks
  - …
- Get involved in ATLAS ML community
  - Participate the meetings, forums
  - …
- Get it used in γγWW(hadronic decay) / γγML analyses

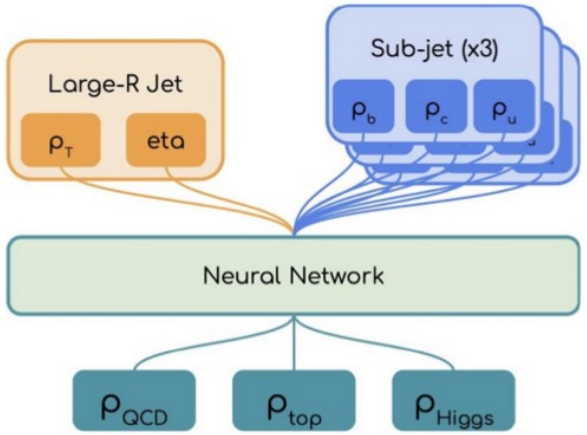


ATLAS event display: candidate pair of Higgs bosons decay in ATLAS

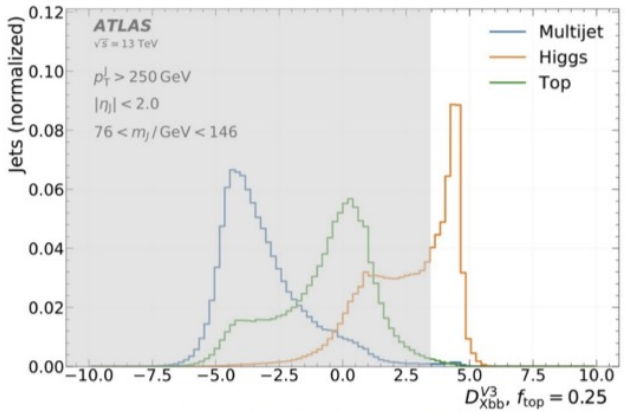
Boosted X→bb

Talk from Bo Liu





- Double bjet tagging in boosted regime
- Reject top and QCD events



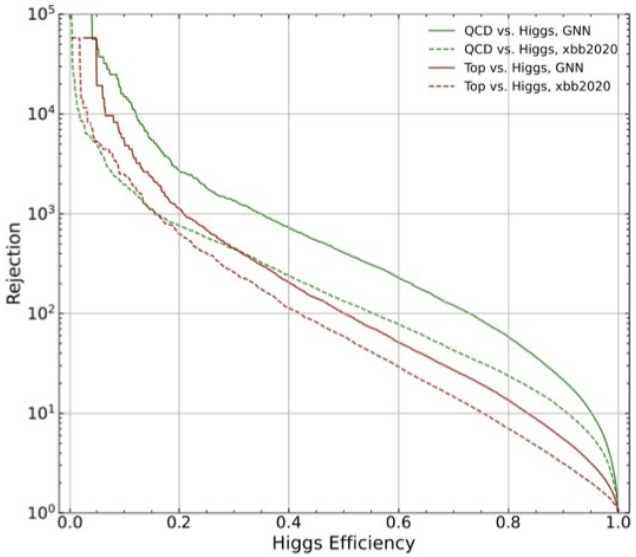
Grey shaded area 60% WP

$$D_{\text{Xbb}} = \ln \frac{p_{\text{Higgs}}}{f_{\text{top}} \cdot p_{\text{top}} + (1 - f_{\text{top}}) \cdot p_{\text{multijet}}}$$

https://indico.cern.ch/event/1132691/timetable/#b-457079-tagging-techniques

For future: GNN

✓ An overall better performance



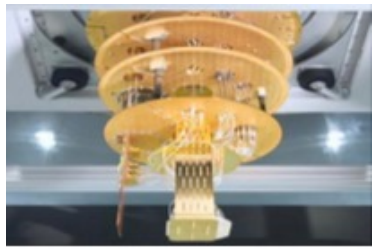
Quantum Machine Learning

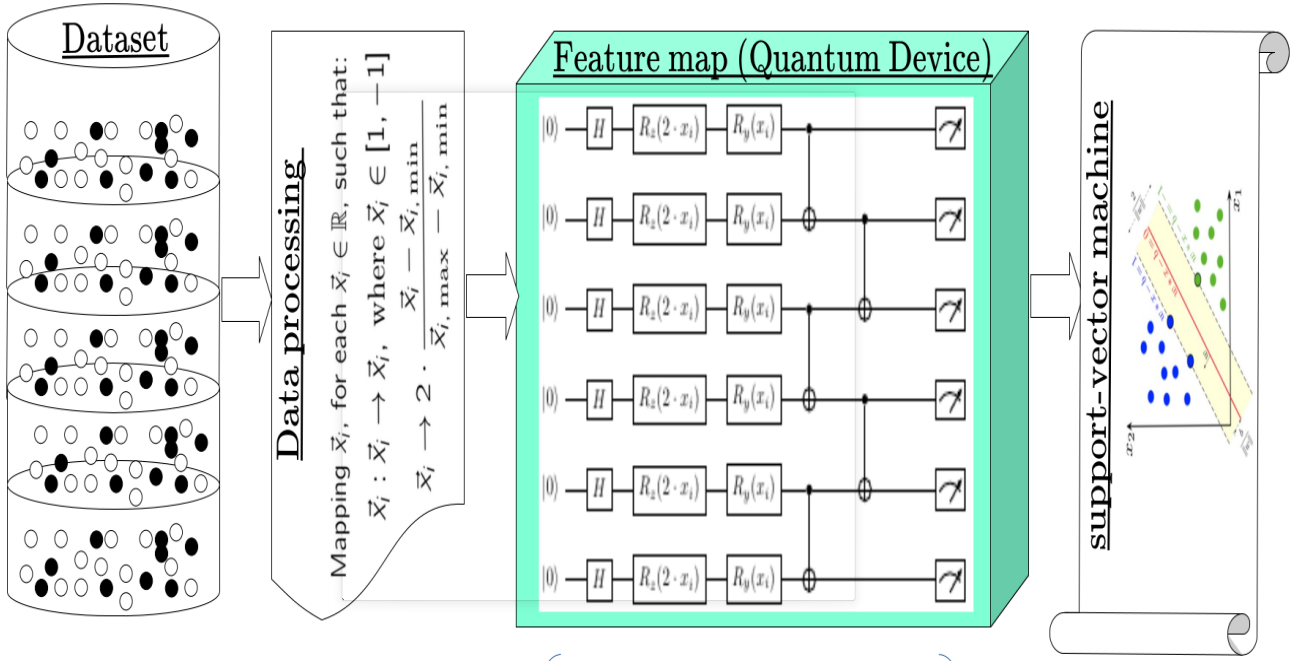
Abdualazm Fadol
Qiyu Sha

QM hardware

IBM

Wuyuan



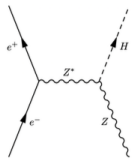


- ✓ Use CEPC H→ $\gamma\gamma$ as an example.
- ✓ QSVM is used as ML algorithm.
- ✓ Try both Simulator and Hardware (IBM/Wuyuan).

CEPC signatures

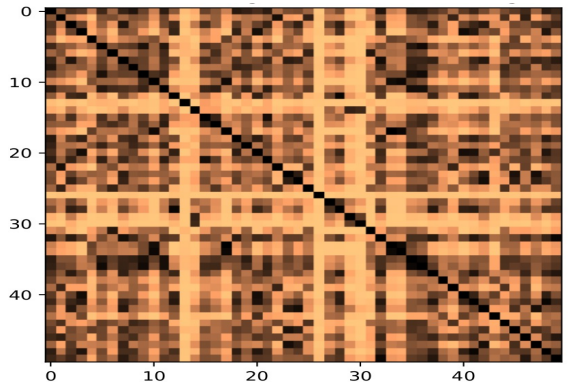
- ☐ Use Monte Carlo simulation for the signal and backgrounds:
  - Signal process $e^+e^- \to ZH \to \gamma\gamma q\bar{q}$
  - Backgrounds $e^+e^- \to (Z/\gamma^*)\gamma\gamma$
- ☐ The samples are generated with the CEPC configurations;
- ☐ at a centre-of-mass energy of 240 GeV with an integrated luminosity of 5.6 ab^{-1}.
- ☐ Up to 25k signal and background events are used.
- ☐ Six variables are used in both SVM and QSVM (6 qubits).

$$\vec{x}_i \to 2 \cdot \frac{\vec{x}_i - \vec{x}_{i,\min}}{\vec{x}_{i,\max} - \vec{x}_{i,\min}}$$

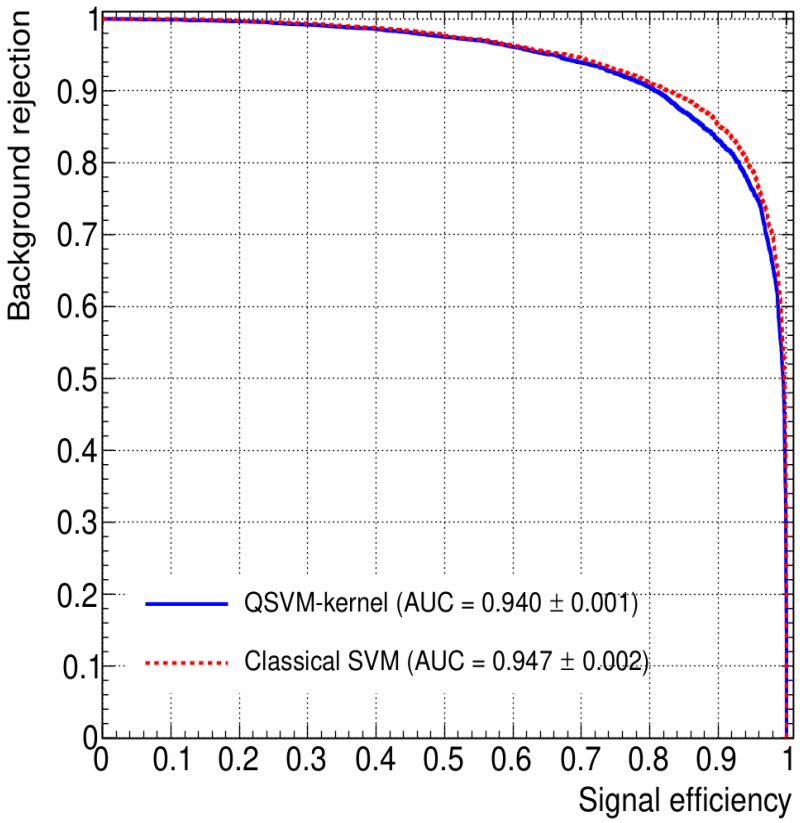


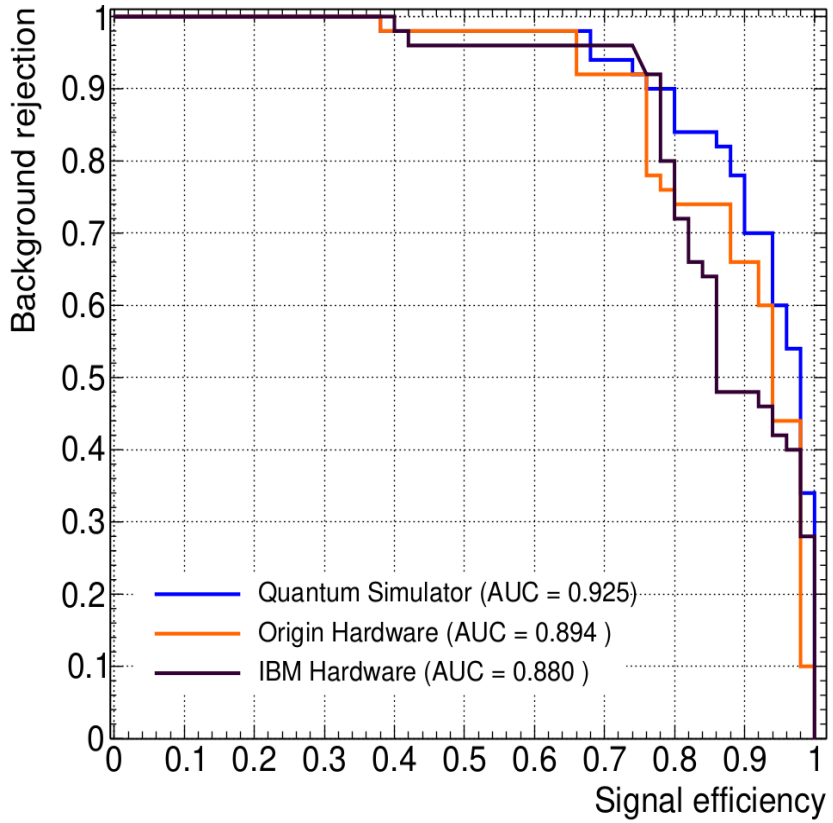
Quantum Kernel estimation

$$k(\vec{x}_i, \vec{x}_j) = \left| \langle \Phi(\vec{x}_i) | \, \Phi(\vec{x}_j) \rangle \right|^2$$



The results for simulator and quantum hardware

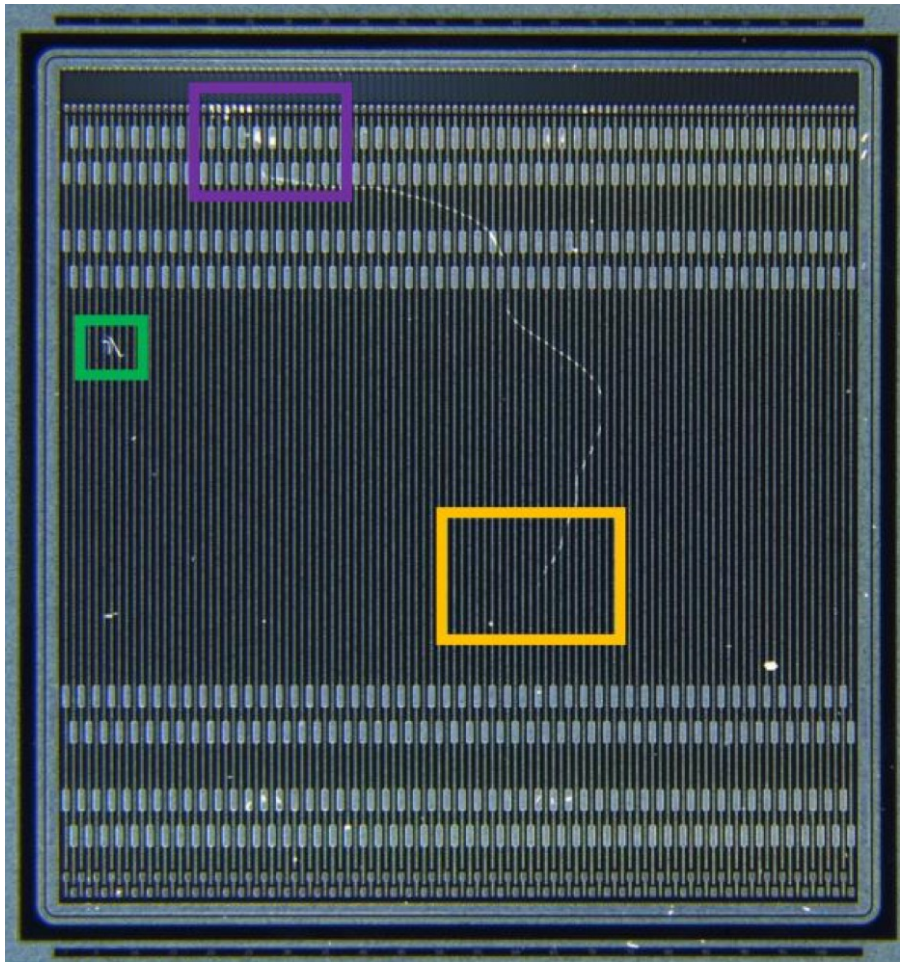




- Using 12000 events for training and testing datasets for the simulator (left).
- This is compared to the classical support-vector machines.
- And only 100 events are tested on Wuyuan and IBM hardware (right).

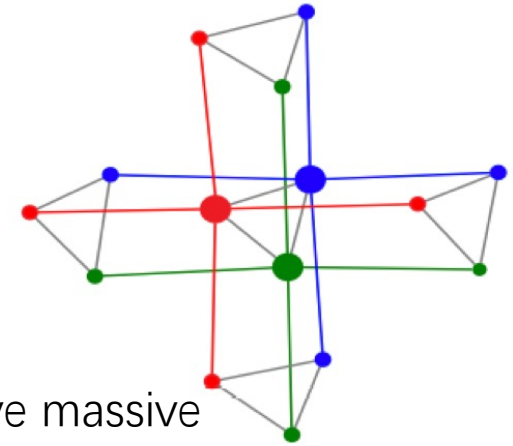
Identify Defects for ATLAS ITK Sensors with ML

Zhan Li



- So for, use microscope and scan over the sensor with eyes. Future, take a picture and inspect with ML algorithms.

- Different algorithms can been tried :
  - ✓ LAD
  - ✓ U-net
  - ✓ Colour-Based Segmentation
- Check the difference between one point and its neighthoods.



- Our IHEP ATLAS ITK team is going to have massive Production of the ITK chips and can use these mthodes

Future Activities and plans

- **A kickoff meeting at the Lab level in one week（done）.**
  - Tutorial on particleNet (done)
  - Tutorial on Quantum ML
- **Workshop : Half a year**
- **Monthly meeting**
  - have one before EPD ML Forum?
- **Invitation of experts to present ML talks.**
  - Both from industry and high energy physics community.
- **Active communication with ATLAS ML forum**
  - learn how to run the efforts efficiently.
  - Latest progresses and hot topics.
- **Short terms: Choose some topics, e.g. particleNet in ATLAS, QML etc..**
- **How to play leading roles: novel algorithms, new implementations…**
- **Manpower : so far 5-7 students and a couple of postdocs show interests, it has to be improved.**
- **Some Needs to be discussed.**
  - GPU resources and platform to do studies (See the talk from Jingyan)?
  - Fundings for machine usage?

backups

ATLAS Machine Learning Forum (AML)

Thanks to Fangyi Guo

- **This forum is jointly organized by Software and Computing (S&C) and Physics and Computing Coordinators for the goals of:**

  Communications and liaisons inside/outside ATLAS:
  - **Organizing ML workshops every year and the bi-week meetings.**
    - **Topics in the bi-weekly meetings: Reports from physics group liaisons; specific applications.**
  - **Organizing tutorials for ATLAS users.**
  - **Liaising with other groups inside and outside ATLAS, facilitating collaboration with Machine Learners.**
  - **Forum for all ATLAS people: atlas-talk, twiki.**

- **Supporting ML tools and techniques:**
  - Coordinating development of common ML tools.
  - Establishing standards where useful.
  - Providing advice on new ML tools and best tools for specific applications.

Machine Learning application (ATLAS)

- **Method and packages: encourage people to use non-HEP tools [twiki].**
  - Interexperimental Machine Learning Pages and resource lists.

  - **Do NOT recommend running trainings within analysis environment, but:**
    1. Produce a minimal "flat" dataset using the standard ATLAS workflow
    2. Run training in a stand-alone environment
    3. (Optionally) port trained algorithms back to Analysis/Reconstruction releases via ONNXruntime. [tutorial]

  - **Computing resources at CERN:**
    - Tools for flat dataset production: HDF5Utils for DAOD/AOD and uproot for ntuples.
    - For distributed training: intelligent Data Delivery Service.
    - Docker images: ATLAS Machine Learning docker, minimalist images.
    - GPU and notebook: JupyterHub at CERN and GPU queues.

VH(cc)

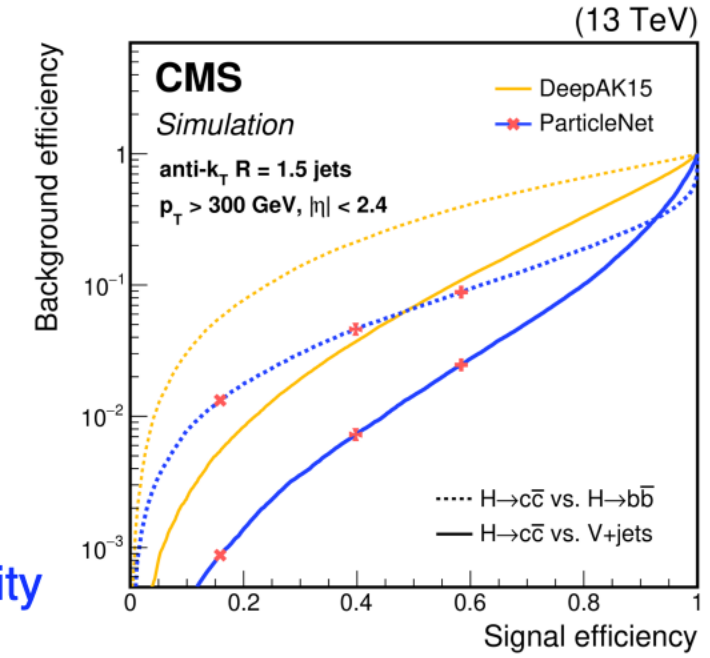
ATLAS: EPJC 82, 717 (2022)
CMS: arXiv:2205.05550

- Three lepton channels: $ZH \to vvcc$(0-lep),$WH \to lvcc$(1-lep), $ZH \to llcc$(2-lep)

VH (full Run2)	ATLAS	CMS	
$H \to cc$ reconstruction	resolved	boosted	resolved
c-tagging	DL1, a deep neural network	ParticleNet, a graph neural network DGCNN	DeepJet
Categorization	2 p_T^V bins x (1-c-tag and 2-c-tag)	BDT x 3 cc-tagging regions	
Fitting discriminants	m_{Hcc}	m_{Hcc}	BDT

ATLAS uses a c-tagger that includes a b-tag veto on MV2c10@70% to establish orthogonality with VH(bb) and allows the combination

CMS ParticleNet tagger:
~5x better H→bb rejection
~5x better V+jet rejection
>2x improvement in the final sensitivity



Some thoughts
(Need to be expanded by you)

- **ML with low statistics.**
  - Meet in VBF Higgs→ $\gamma\gamma$ analysis
- **Quantum ML**
  - Hardware vs Simulator
  - Develop effective algorithms
- **Please feel free to add more now or later at the Mattermost.**
  - ……
  - ……

ML Applications in physics (ATLAS)

- **Interesting topics about ML we can do:**

  Many general topics are still remaining:

  - Event generation and simulation: GAN …
  - Jet identification/quark-gluon tagging: CNN, GNN …
  - Modeling: Gaussian Processes, DNN…
  - Event classification: all kinds of ML methods.

- **Fresh topics:**
  - Unsupervised learning for new physics.
  - Tools for ML: feature extraction, hyper-parameter optimization, robustness quantification, etc.
  - Hardware-based ML.
  - Quantum computing and quantum ML.