EPD- ML forum

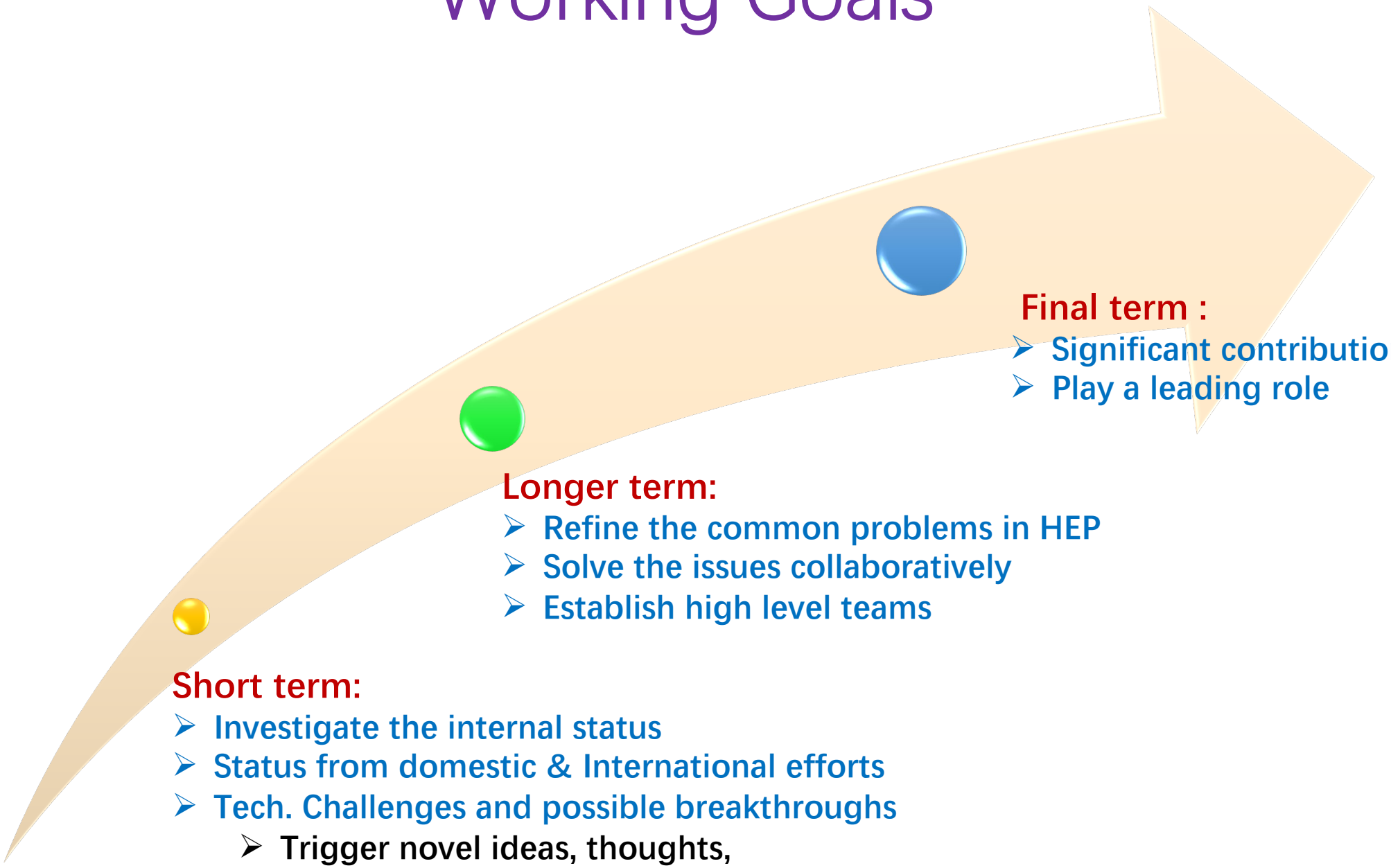
Yaquan Fang (IHEP)

IHEP ATLAS group meeting
August 30th,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 role

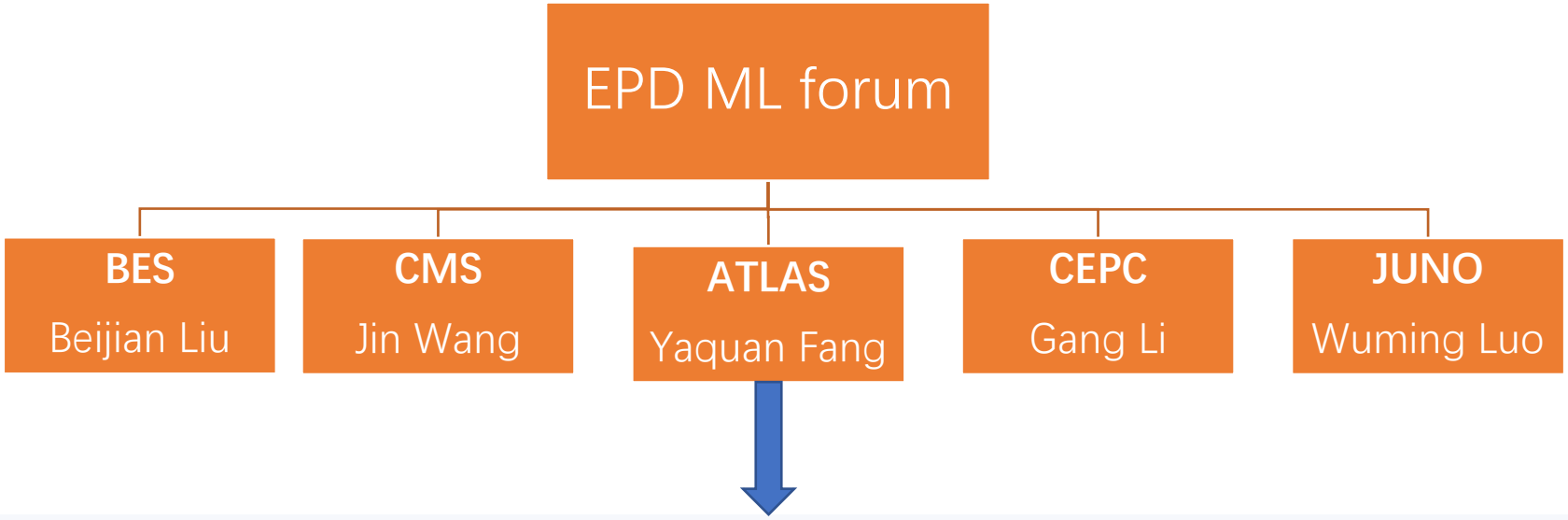
Longer term:

- 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**

Thanks to Kaili

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

Future activities

- **A kickoff meeting at the Lab level in one week.**
  - Will post the announcement at the Mattermost.
- **Workshop : Half a year**
- **Monthly meeting in the EPD ML Forum**
  - Should we have one before EPD ML Forum?
- **Invitation of experts to present ML talks.**
  - Both from industry and high energy physics community.
- **Some Needs to be discussed.**
  - GPU? Fee for machine usage?

Some thoughts
(Need to be expanded by you)

- **ML with low statistics.**
  - VBF Higgs→ $\gamma\gamma$ analysis
- **Quantum ML** (*Abdualazem/Qiyu*)
  - Hardware vs Simulator
  - Develop effective algorithms
- **Implementation of latest ML methods to particle tagging** (*Sudong*)
- **Please feel free to add more now or later at the Mattermost.**
  - ……
  - ……
- **Hardware relaed ML** (*Zhan*)
- **Computing resources** （*Jingyan*）

Wechat group:

IHEP ATLAS ML discussion

该二维码7天内(9月20日前)有效，重新进入将更新

Mattermost:

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

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)

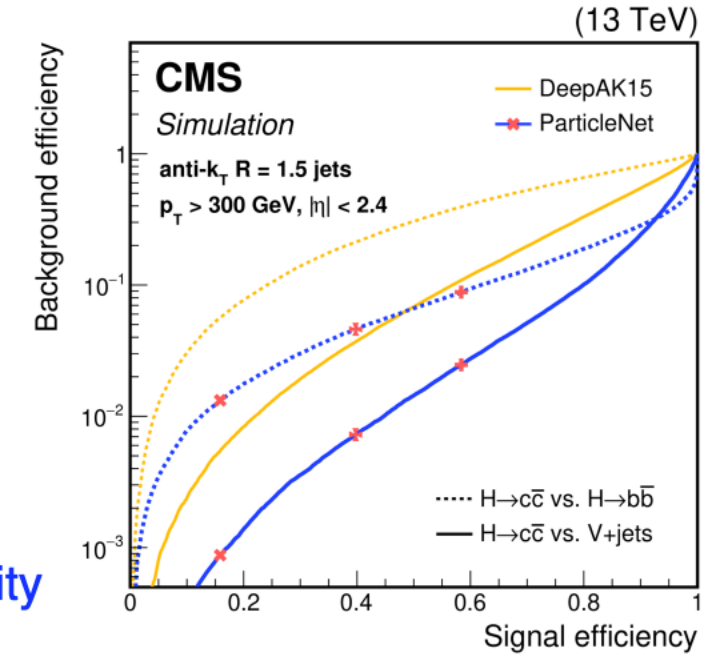
ATLAS: EPJC 82, 717 (2022)
CMS: arXiv:2205.05550

- 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.