

机器学习所创新课题进展和大模型思考

方亚泉 代表

所ML创新团队和所创新“机器学习在高能物理应用”的课题成员
(实验物理中心)

高能所第三次深度学习研讨会

2023/6/15

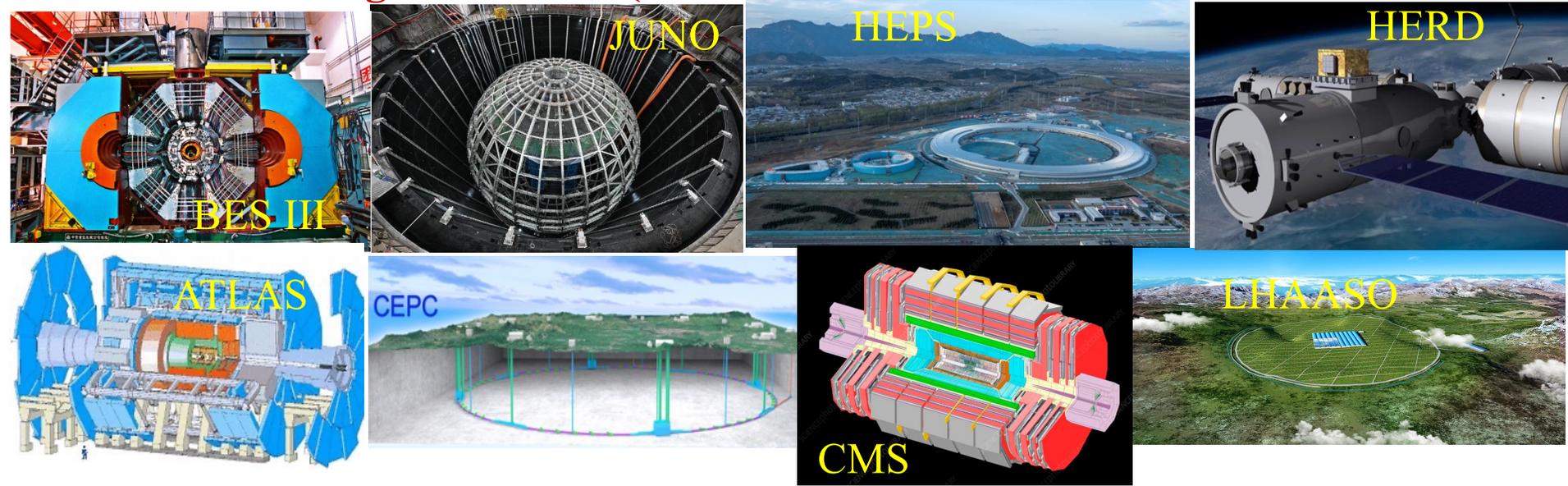
所机器学习创新团队（2022年10月10日成立）

动态开放，欢迎加入！！！！

- 目前团队成员有 > 50人：
- 职工占一半以上，有20名职工承诺花40%以上时间。
- 中心包括：实验物理、天体、多学科、加速器、计算中心、东莞.....
- 实验：ATLAS/CMS、BESIII、CEPC、HERD、LHAASO、光源.....
- 成立 Steering committee (~ 10人)



群聊: IHEP ML Innovation Group



ML创新团队活动

- 每周一次的例会，steering committee月会，一些邀请报告
- 一些tutorials: ParticleNet/PointNet 的tutorials
- 获得所创新课题的资助
 - 不断提炼共性的问题，理清创新的思路，明晰研究的方向，确定发展的目标！
- 邀请曲慧麟访问并做了深入的探讨和合作：
 - 计划与企业大厂加强合作
- 与美国SNOWMASS上机器学习的专家建立起一定的合作

kickoff meeting of IHEP ML Innovation group
Tuesday Oct 11, 2022, 2:00 PM → 4:00 PM Asia/Shanghai
202 (计算中心)
Description Zoom : 536 870 8448

2:00 PM → 2:20 PM Introduction (20m)
Speaker: Yaqun FANG Yaqun (高能所)
Introductio...

2:20 PM → 2:40 PM contributions (20m)
Speakers: Beijiang LIU (高能所), Cheng ZHANG (ihep), Dianshuai Zhang (高能所), Gang Li Gang (EPD.IHEP) (高能所), Jin Wang (IHEP), Lu WANG Lu (IHEP), UNKNOWN YAO Haodong, UNKNOWN ZHANG Kai (高能所), UNKNOWN ZHAO Guang (高能所), UNKNOWN 伍力源, UNKNOWN 孙明辉, UNKNOWN 张正德, UNKNOWN 张笑鹏 (高能所), UNKNOWN 罗武鸣 (高能所), UNKNOWN 赵丽娜 (高能所), Yaqun FANG Yaqun (高能所), Ye YUAN (高能所), 文兴方 (高能所), 浩凯 孙 (高能所), 琪 吴 (高能所), 睿 杜 (高能所), 誉 胡 (高能所), 镇轩 张 (高能所)
MLCMS_JL..., Wenxing F..., 赵丽娜-ML...

Weekly Meeting of IHEP ML Innovation group
Thursday Oct 20, 2022, 1:00 PM → 3:00 PM Asia/Shanghai
ParNet_Tuto_20221020_ShudongWANG.pdf
Description Zoom : 536 870 8448

1:00 PM → 1:10 PM Introduction (10m)
Speaker: Yaqun FANG Yaqun (高能所)

1:10 PM → 2:00 PM Tutorial on ParticleNet (50m)
Speaker: Shudong Wang (高能所)
ParNet_Tut...

2:00 PM → 2:20 PM contributions (20m)
Speakers: Beijiang LIU (高能所), Cheng ZHANG (ihep), Dianshuai Zhang (高能所), Gang Li Gang (EPD.IHEP) (高能所), Jin Wang (IHEP), Lu WANG Lu (IHEP), UNKNOWN YAO Haodong, UNKNOWN ZHANG Kai (高能所), UNKNOWN ZHAO Guang (高能所), UNKNOWN 伍力源, UNKNOWN 孙明辉, UNKNOWN 张正德, UNKNOWN 张笑鹏 (高能所), UNKNOWN 罗武鸣 (高能所), UNKNOWN 赵丽娜 (高能所), Yaqun FANG Yaqun (高能所), Ye YUAN (高能所), 文兴方 (高能所), 浩凯 孙 (高能所), 琪 吴 (高能所), 睿 杜 (高能所), 誉 胡 (高能所), 镇轩 张 (高能所)

Month	Date	Event
June 2023	Jun 01	Weekly Meeting of IHEP ML Innovation/production/generation group
May 2023	May 11	Weekly Meeting of IHEP ML Innovation group
April 2023	Apr 13	Weekly Meeting of IHEP ML Innovation/production/generation group
	Apr 06	Weekly Meeting of IHEP ML Innovation group
March 2023	Mar 23	Weekly Meeting of IHEP ML Innovation group
	Mar 16	Weekly Meeting of IHEP ML Innovation group
February 2023	Feb 28	Weekly Meeting of IHEP ML Innovation group
	Feb 23	Weekly Meeting of IHEP ML Innovation group
	Feb 09	Weekly Meeting of IHEP ML Innovation group
	Feb 02	Weekly Meeting of IHEP ML Innovation group
January 2023	Jan 12	Weekly Meeting of IHEP ML Innovation group
	Jan 05	Kick off meeting the ML application in HEP (IHEP Innovation project)
	Jan 05	Weekly Meeting of IHEP ML Innovation group
	Jan 03	particleTransformer tutorial
December 2022	Dec 22	Weekly Meeting of IHEP ML Innovation group
	Dec 01	Weekly Meeting of IHEP ML Innovation group
November 2022	Nov 24	Weekly Meeting of IHEP ML Innovation group
	Nov 17	Weekly Meeting of IHEP ML Innovation group
October 2022	Oct 27	Weekly Meeting of IHEP ML Innovation group
	Oct 20	Weekly Meeting of IHEP ML Innovation group
	Oct 11	kickoff meeting of IHEP ML Innovation group

所创新课题研究内容

探测器的模拟 速度 指数级提升

- ✓ CMS
- ✓ BESIII
- ✓

事例/粒子重建 精度/效率 显著提升

- ✓ JUNO 罗武鸣
- ✓ BESIII 袁野
- ✓ CEPC
- ✓ LHAASO
- ✓ HERD
- ✓

喷注/事例鉴别 效率vs排错 显著提升

- ✓ JUNO 罗武鸣
- ✓ BESIII
- ✓ CEPC
- ✓ LHAASO
- ✓ HERD
- ✓ ATLAS/CMS

徐达
李刚
王锦

分类和推断 决策的有效和优化

- ✓ BES分波分析 刘北江
- ✓ CEPC 李刚
- ✓ 加速器上优化
- ✓ 同步辐射应用 伍力源
- ✓

量子机器学习 算法速度提升

- ✓ 寻找量子场景
- ✓ 优化核算法

BESIII上探测器

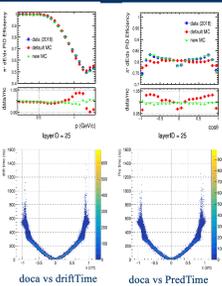
模拟

方文兴

方文兴 (高能所)

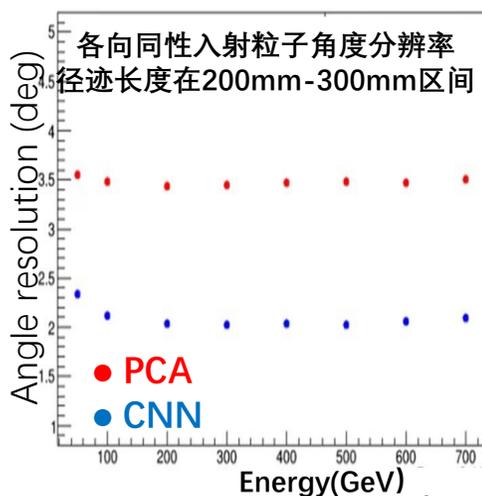
基于数据驱动的模拟

- 漂移室 dE/dx 模拟
- 漂移室时间模拟
- 量能器反中子模拟
- 可以明显降低系统误差

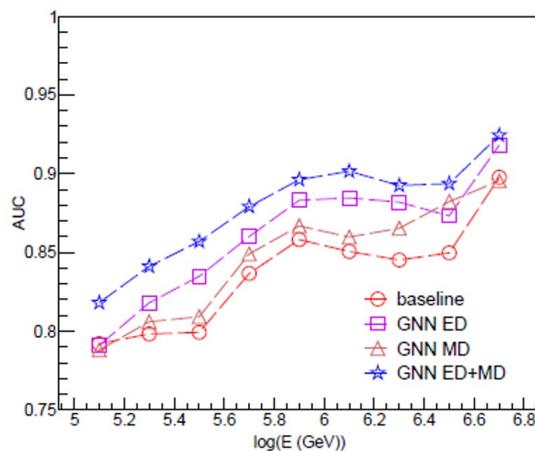


HERD入射角

重建 张诚

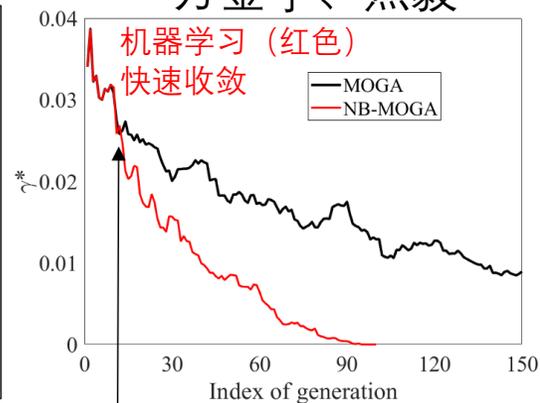


LHAASO上P鉴别张笑鹏



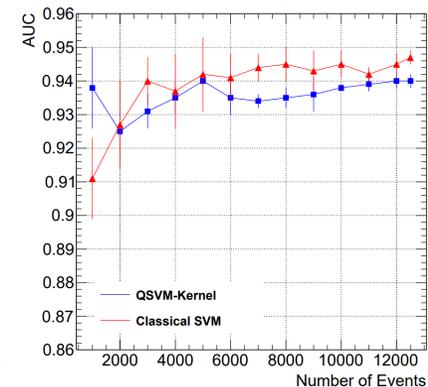
回速器：HEP存贮环 非线性动力学优化

万金宇、焦毅



量子机器学习vs 机器学习

方亚泉



进展1: W-tagger @ ATLAS

徐达

- 基于图形神经网络的机器学习方法 (ParticleNet, ParticleTransformer, PointNet etc) , 显著提升对撞机实验中 W tagging的鉴别效率, 助力HL-LHC上 di-Higgs 的首次观测。
- 人员: 王书栋、徐达、李刚、方亚泉
- 基于ATLAS实验, 完成基于constituent-based新型W/Z tagger, 使其成为成熟可供合作组使用的tagger。
- 基于ATLAS实验, 通过上述新型tagger方案研究HH—WW的全强子末态。
 - ✓ 目前正在准备一篇ATLAS-PUB-NOTE, 下周将进行ATLAS jet tagger组预审核,
 - ✓ 计划于8月BOOST2023国际会议发表, 同时本课题学生也将在BOOST2023展示该成果的Poster

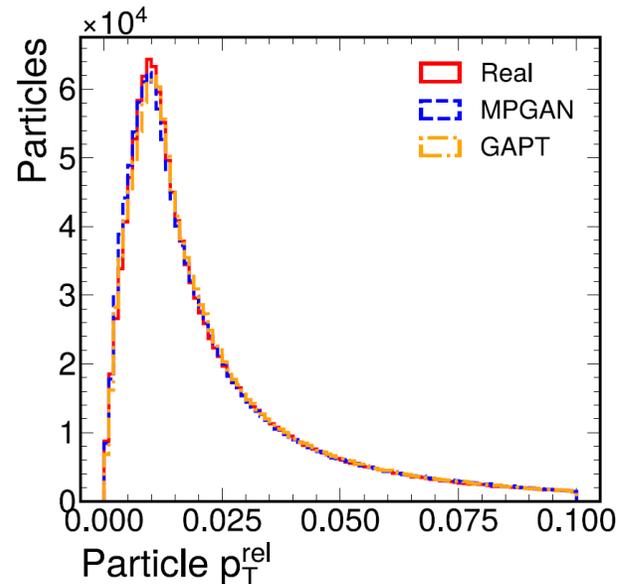
ParticleTransformer/Net
比现存的tagger性能好>3倍

进展2: Generative Model for FastSim 和 Pileup 排除 @ CMS

➤ CMS电磁量能器性能组 (Convener: 王锦, Thomas Reis) 与CMS机器学习组研究改进前沿生成模型以进行探测器快速准确模拟

- 基于对抗生成网络模型的改进 (Generative Adversarial Networks, GANs) : MPGAN, EPiC-GAN
- 基于变分自编码器模型的改进 (Graph Variational Autoencoders)
- 基于扩散模型的改进 (Diffusion models)

王锦



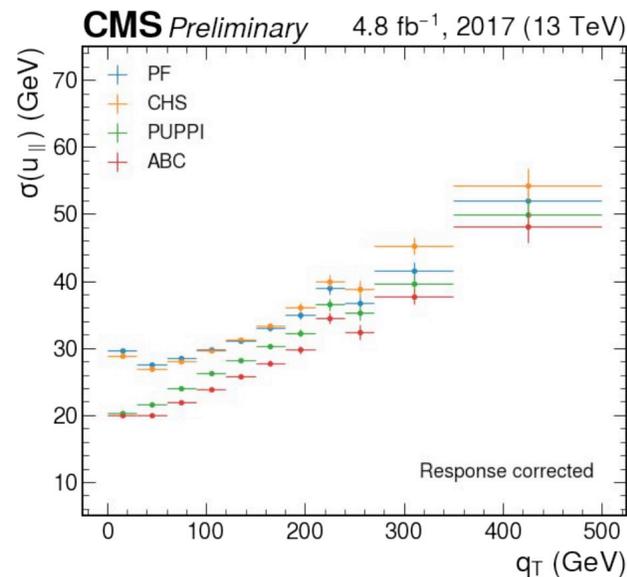
高能对撞实验中的堆积事例将显著降低物理对象的能量分辨率

① 使用最优传输成本函数及基于注意力机制图神经网络 (ABCNet) 来进行堆积事例排除 (TOTAL) - Fabio Lemmi (IHEP)

- ① 目前唯一可以直接在CMS实验中实际应用的堆积事例排除机器学习方法
 - ① 其它ML方法依赖于粒子的真实标记, 在实际模拟事例中无法取得
- ① 性能优于现在基于选择的堆积事例排除方法 (PUPPI)
 - ① 更好地重建事例中强子射流和缺失的横向能量, 提高这些物理对象的分辨率
- ① Proof-of-concept已经发表[arxiv](#)

① 最新进展

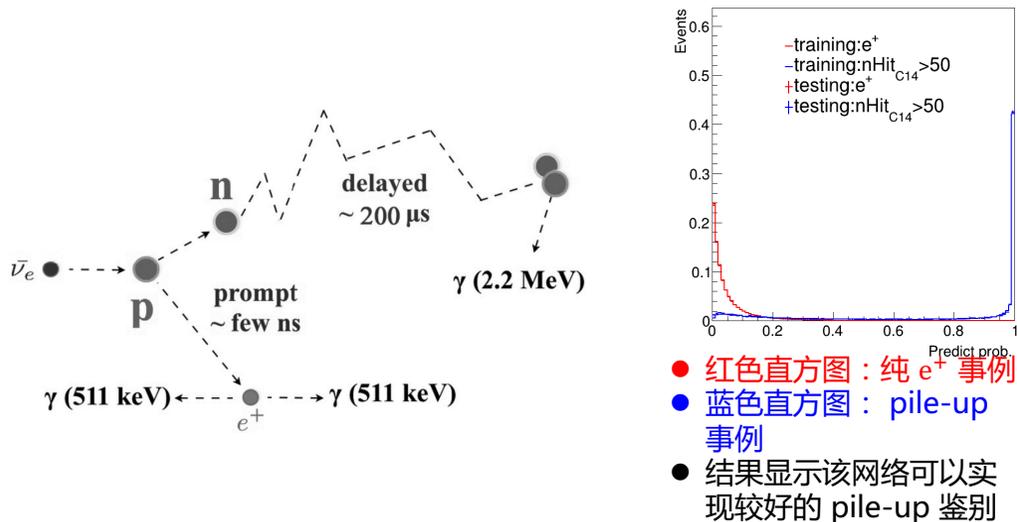
- ① 完成了CMS软件框架内相应的模块的编写
- ① 进行了更细致的模型优化, 改进了运行时间与内存使用
- ① 完成了更多的基于数据的性能研究



进展3: JUNO上机器学习的应用

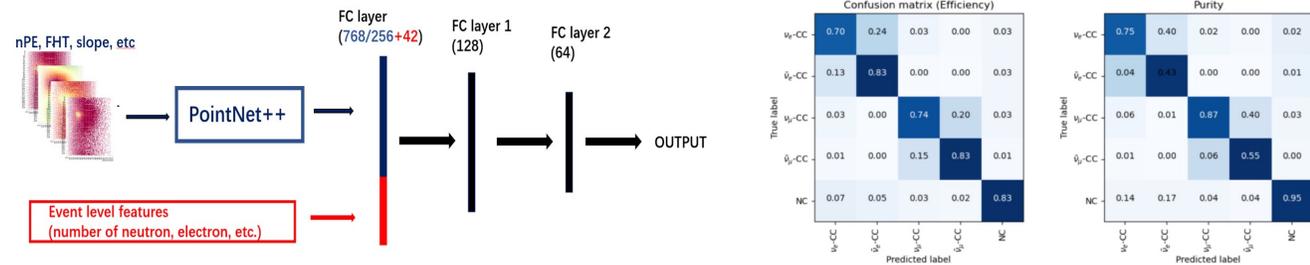
JUNO ^{14}C pile-up 鉴别 方文兴

- 背景：反 beta 衰变 (IBD) 是 JUNO 中微子质量顺序测量的信号事例 ($\bar{\nu}_e + p \rightarrow n + e^+$)。其中 e^+ 重建是实验的关键，包括能量、位置重建。JUNO 20千吨 LAB 液闪的 ^{14}C 放射性活度为4万 Bq, 导致 e^+ 能量分辨率相对降低 1.4%@1MeV, 不能忽略
- 难点： ^{14}C 的信号幅度比 e^+ 小一个量级以上，当 ^{14}C 和 e^+ 在时间和空间上接近时， ^{14}C pile-up 鉴别具有困难
- 基于机器学习方法的 ^{14}C pile-up 鉴别：
 - 利用一维卷积神经网络，识别该事例是否为 pile-up 事例
 - 网络输入：重建 pulse 在时间上的分布
 - 网络输出：事例类型 (是否为 pile-up 事例)

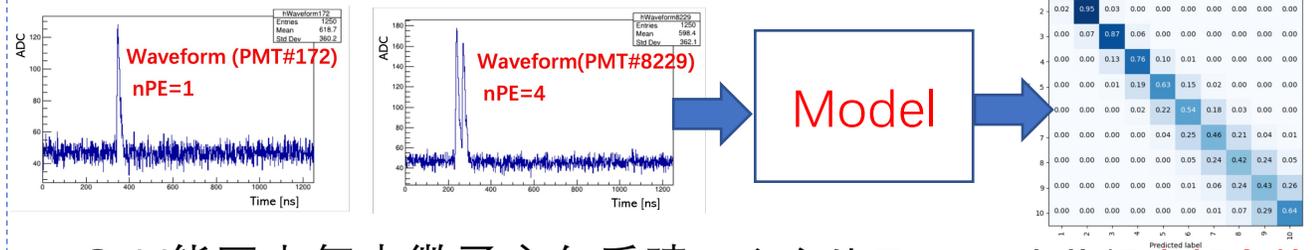


JUNO: GeV能区大气中微子研究

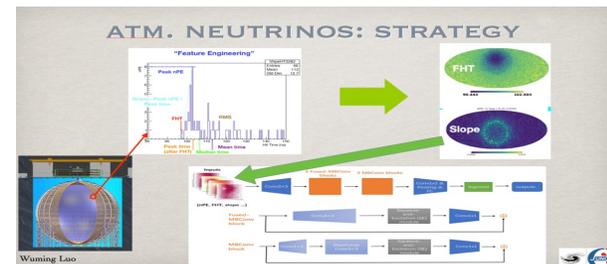
- GeV能区大气中微子PID：计划7/15号之前完成分析



- MeV能区反应堆中微子：PMT波形光子数识别



- GeV能区大气中微子方向重建：文章处于JUNO合作组内部审核



First attempt of directionality reconstruction for atmospheric neutrinos in a large unsegmented liquid scintillator detector

Z. Yang,¹ J. Liu,² H. Duysang,¹ W. Guo,² X. He,² Y. Hu,² T. Li,¹ Z. Liu,² W. Luo,² X. Luo,² X. Tan,¹ F. Zeng,² and Y. Zhang²

¹Shandong University, Jinan, China, and ²Key Laboratory of Particle Physics and Particle Irradiation of Ministry of Education, Shandong University, Qingdao, China

³Institute of High Energy Physics, Beijing, China

The directionality information of incoming neutrinos is essential to atmospheric neutrino oscillation analysis since it is directly related to the oscillation baseline length. Large unsegmented liquid scintillator detectors, while offering excellent energy resolution, are traditionally very limited in their capabilities of measuring event directionality. In this paper, we present a novel directionality reconstruction method for atmospheric neutrino events in large unsegmented liquid scintillator detectors based upon waveform analysis and machine learning techniques. We demonstrate for the first time that such detectors can achieve good direction resolution and potentially play an important role in future atmospheric neutrino oscillation measurements.

I. INTRODUCTION

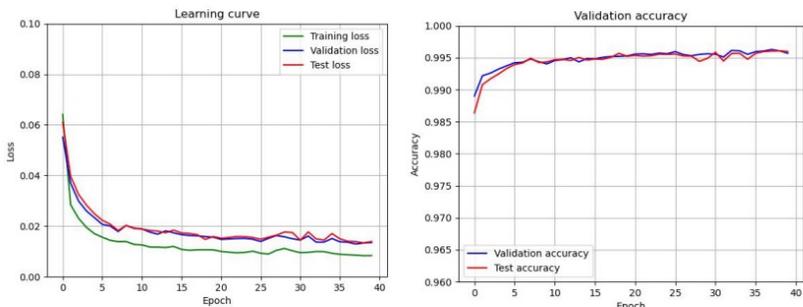
Liquid scintillator (LS) detectors play an important role in neutrino physics. They typically offer low solar neutrinos, they certainly can be adapted for atmospheric neutrino oscillation measurements if successful. However, major hardware upgrades are required for those techniques to be applicable to existing detectors.

进展4: 机器学习用于径迹的重建@ BESIII

袁野

• 信号与噪声的分类

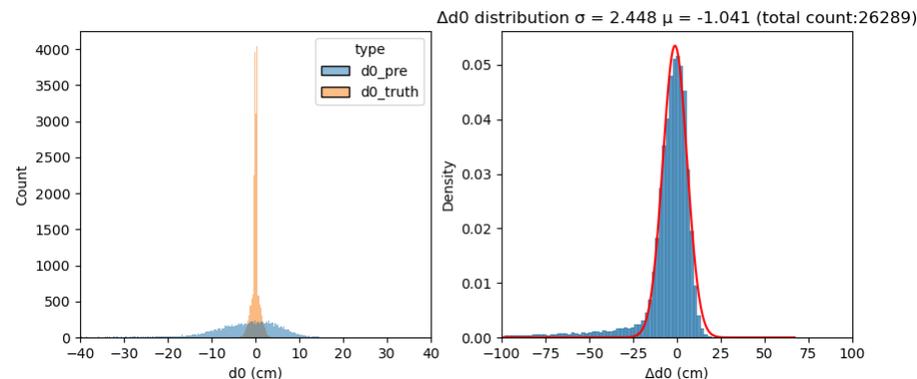
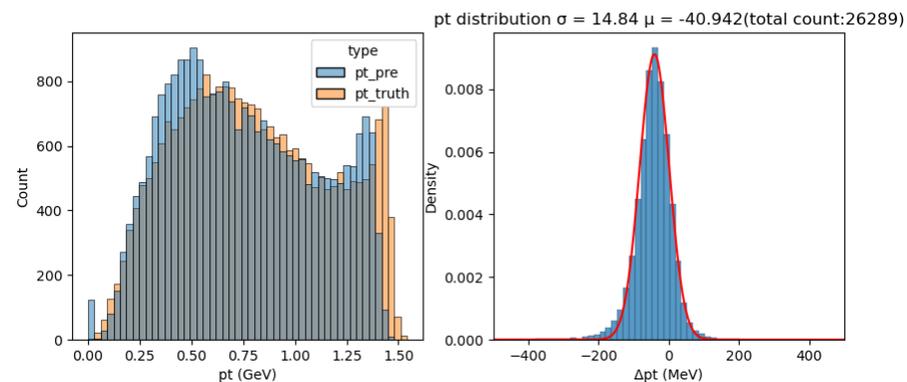
- 利用图神经网络, 区分噪声与信号的准确率已经达到99%以上 [\[link\]](#)



~99.5% classification accuracy

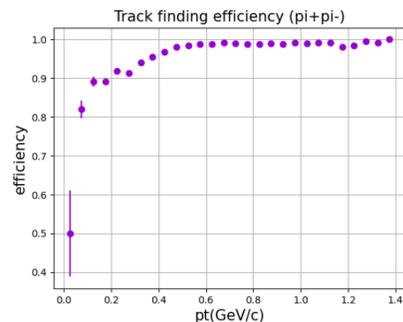
• 径迹参数的拟合

- 目前还没有针对该任务的直接研究工作
- 但是寻迹工作中, 有提出利用径迹拟合辅助寻迹的想法, 该工作中的拟合效果如下图所示 [\[Link\]](#)



• 信号击中的聚类

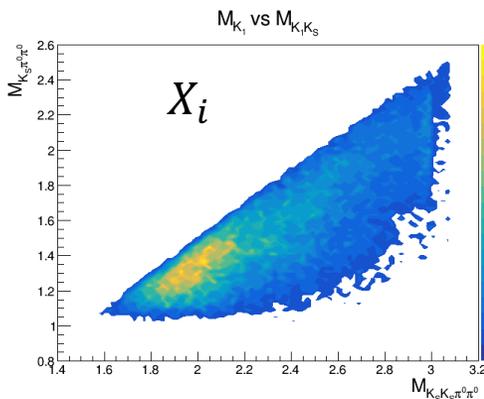
- 该任务在机器学习领域并没有直接可用的方法, 目前的难点
- 基于GNN的聚类寻迹效率如右图所示 [\[link\]](#)
- 基于PointNet的聚类寻迹效率平均为97.02%



进展5：机器学习用于多维拟合优度检验工具@ BESIII

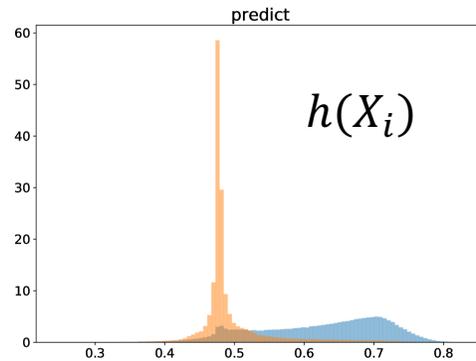
刘北江

- 开发了基于分类器的拟合优度检验算法
 - 拟合优度检验 → 两组多维数据的比较
 - 机器学习中的异常检测
 - 使用分类器在多维空间提取“实验数据”和“拟合结果”的特征并加以区分
 - 分类器得到1维输出 (LRT, AUC, ...) 用来判断拟合优度



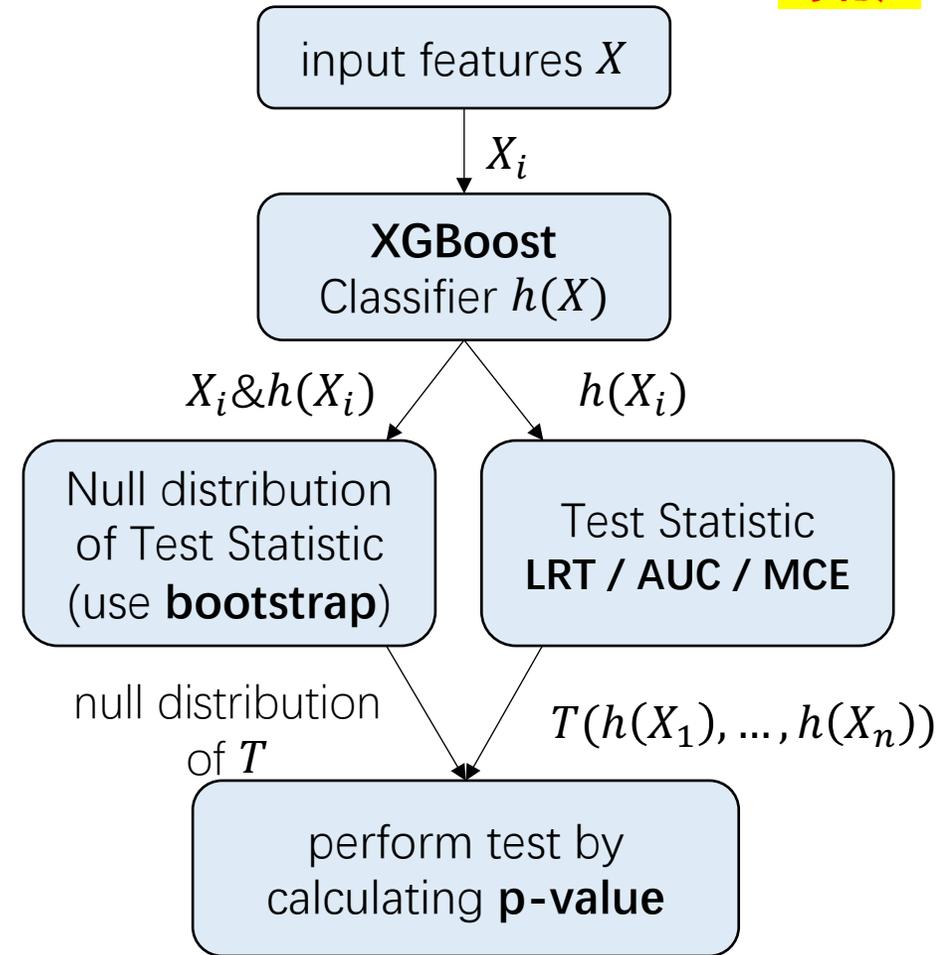
五体末态相空间 (10维数据)

分类器



似然比 (1维输出)

一维检验 p值



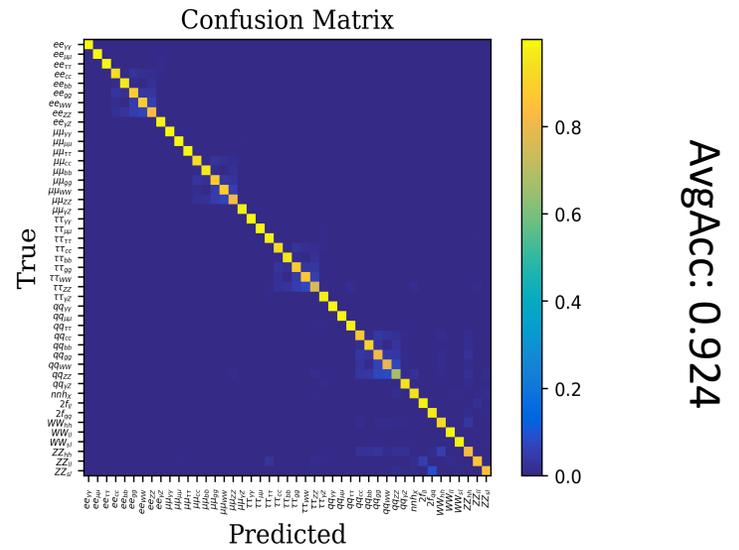
完成情况：已经实现。使用基于贝叶斯优化的超参数自动调参工具，进一步提高分类器性能
使用相同数据集情况下，经过贝叶斯优化调参后的分类器性能及拟合优度检验效果得到明显提升
计划：准备撰写文章，研究应用transformer的异常检测，探索在新信号（新共振态）寻找的应用

进展6-8: CEPC, HERD和LHAASO上机器学习的应用

CEPC Global Higgs Analyses

王书栋、李刚

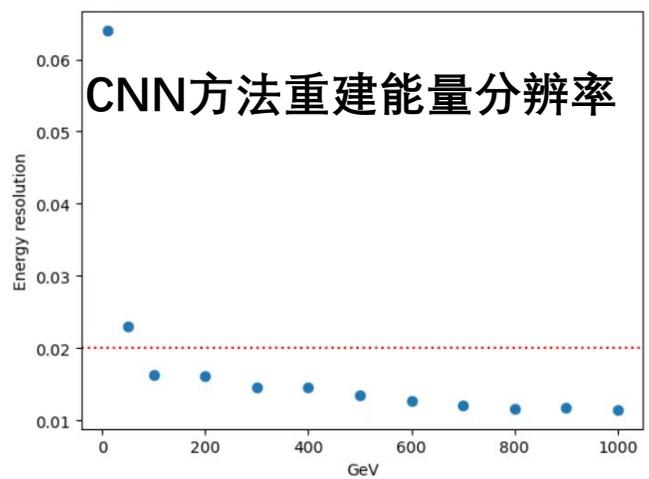
用ParticleTransformer进行45分类:



应用ParticleTransformer对CEPC上4个Higgs产生模式($e^+e^-H, \mu^+\mu^-H, \tau^+\tau^-H, q\bar{q}H$)和相关本底($l^-l^+, q\bar{q}, nnH, WW_l, WW_h, WW_{sl}, ZZ_l, ZZ_h, ZZ_{sl}$)进行多分类, 作为对Higgs衰变末态($H \rightarrow b\bar{b}, c\bar{c}, s\bar{s}, \mu^+\mu^-, \tau^+\tau^-, gg, \gamma\gamma, WW^*, ZZ^*, Z\gamma$)的分支比进行高效联合拟合测量的基础 (正在进行)。

HERD机器学习应用

- ✓ 聚类算法用于量能器簇射团寻找
- ✓ 量能器粒子入射角度的重建
- ✓ 量能器能量重建工作

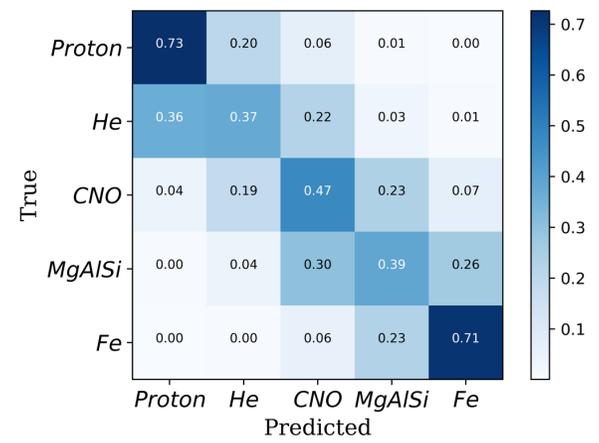


- 量能器簇射团寻找工作已基本完成, 持续完善优化中。
- 量能器粒子入射角度的重建已完成CNN模型的重建, 效果达到预期, 正在尝试使用CoAtNet、PointNet++等先进算法进一步提升重建效果。
- 量能器能量重建工作已完成CNN模型的重建, 目前着重解决模型泛用性问题, 在Fine tuning、DANN和数据增强三个方向上进行尝试。
- 量能器粒子鉴别工作正在使用CNN、ResNet50和ParticleNet模型进行训练调试中。

LHAASO机器学习应用

✓ 利用LHAASO-KM2A全阵列模拟+重建数据对5种原始成分进行分类

- P任务 (信号: Proton; 背景: He+CNO+MgAlSi+Fe)
- L任务 (信号: Proton+He; 背景: CNO+MgAlSi+Fe)
- 在10-100TeV能段上, ParticleNet对P任务和L任务的判别能力均超过传统方法 (基线), CNN和GNN
- 首次尝试直接对5种成份进行区分

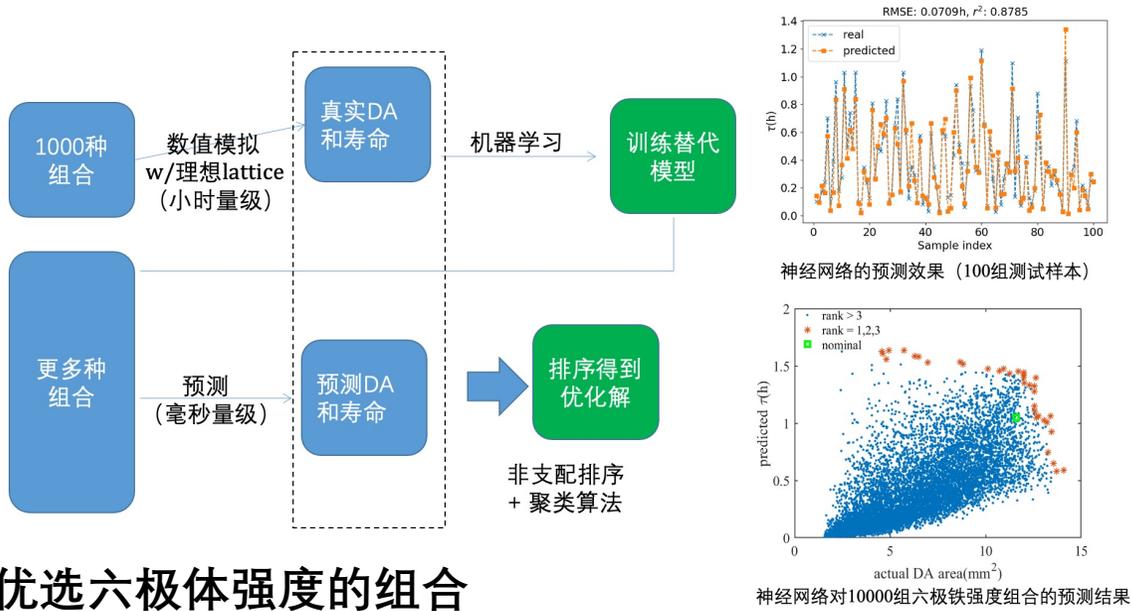


进展9-10: HEPS上相关的应用

焦毅
伍力源、赵丽娜

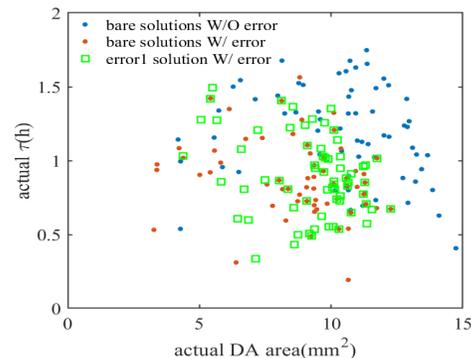
调束中后期色品调节/六极铁强度组合确定

- 优化HEPS设计色品, 提高预测的准确度

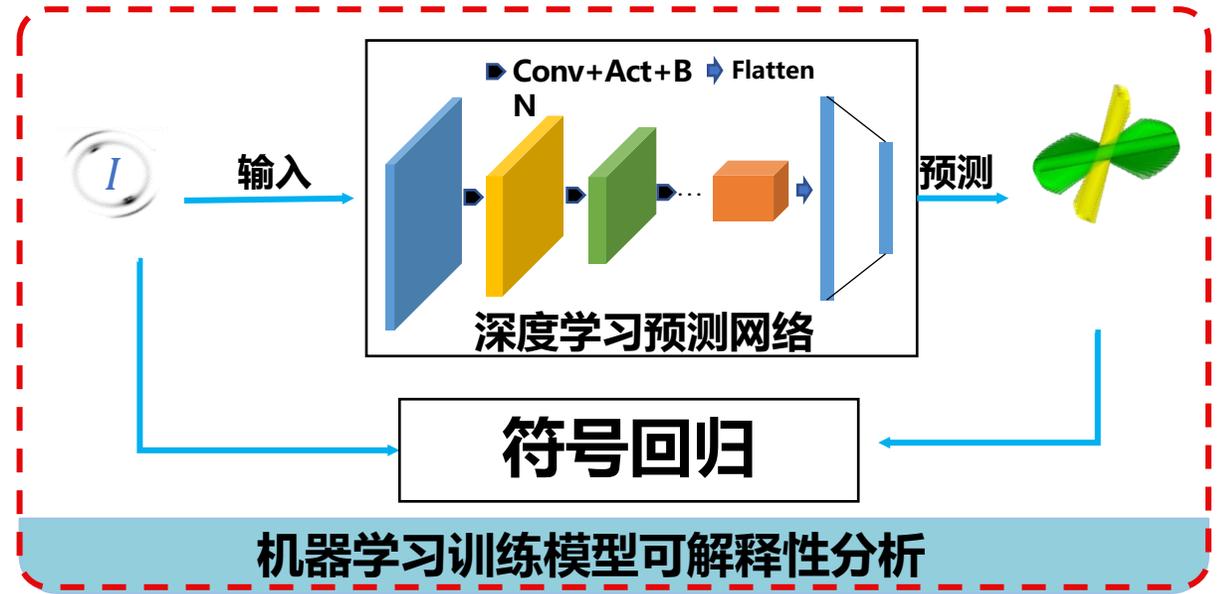


优选六极体强度的组合

模型只需训练一次
在数分钟内快速地给出机器需要的几十种不同 (DA、寿命) 优选的六极铁强度组合, 供实际机器测试, 并选定最终组合



纳米纤维取向度与衍射强度的关系探索



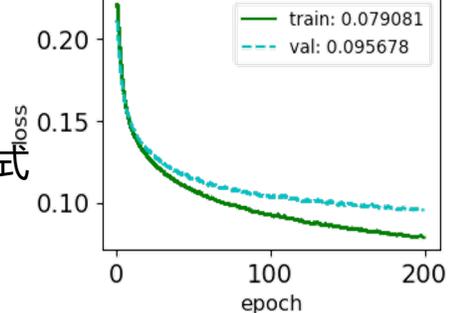
初步实现通过符号回归将训练模型表达为解析式:

- ✓ 机器学习训练数据得到“黑箱”预测模型
- ✓ 将符号回归嵌入深度学习, 实现输出解析关系式

下一步:

- 增加数据量, 探索符号回归的极限
- 调整符号回归的参数, 探索是否能得到更简易的数据表达式

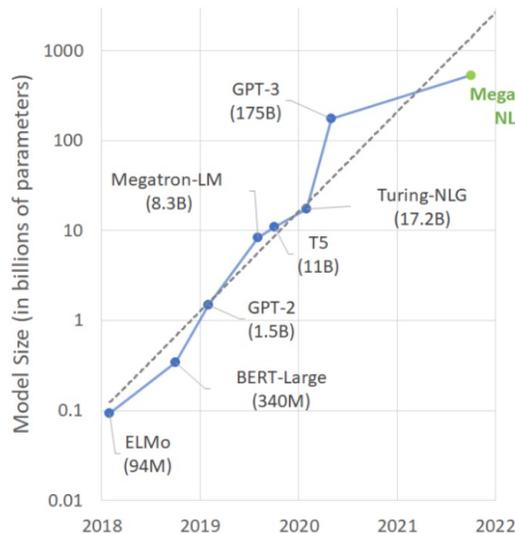
process:: 200/200 epoch, 11th batch



Foundation models

A **foundation model** is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks ; current examples include BERT [Devlin et al. 2019], GPT-3 [Brown et al. 2020], and CLIP [Radford et al. 2021].

- 2021年就由Stanford研究团队提出，作为趋势，已经有好几年了 (NLP)
- 生成型模型对没有标记的数据进行训练，再进行多任务作业
- 大模型对应大的计算资源和大数据集
- 为此，CERN专门争对FM与IBM进行了研讨会
 - 探讨了在模拟、物理分析、加速器等方面的FM研究。
 - 但目前只在聊天ChatGPT-like和探测器快速模拟方面有些想法和进展



Contents	2
1 Introduction	3
1.1 Emergence and homogenization	3
1.2 Social impact and the foundation models ecosystem	7
1.3 The future of foundation models	9
1.4 Overview of this report	12
2 Capabilities	21
2.1 Language	22
2.2 Vision	28
2.3 Robotics	34
2.4 Reasoning and search	40
2.5 Interaction	44
2.6 Philosophy of understanding	48
3 Applications	53
3.1 Healthcare and biomedicine	54
3.2 Law	59
3.3 Education	67
3.4 Technology	73
4.1 Modeling	74
4.2 Training	81
4.3 Adaptation	85
4.4 Evaluation	91
4.5 Systems	97
4.6 Data	101
4.7 Security and privacy	105
4.8 Robustness to distribution shifts	109
4.9 AI safety and alignment	114
4.10 Theory	118
4.11 Interpretability	123
5 Society	129
5.1 Inequity and fairness	130
5.2 Misuse	136
5.3 Environment	140
5.4 Legality	146
5.5 Economics	149
5.6 Ethics of scale	152
6 Conclusion	161
Acknowledgments	161
References	161

arXiv:2108.07258v3 [cs.LG] 12 Jul 2022

On the Opportunities and Risks of Foundation Models

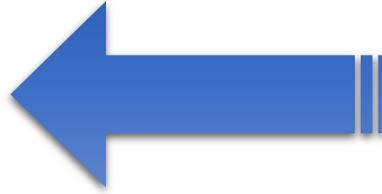
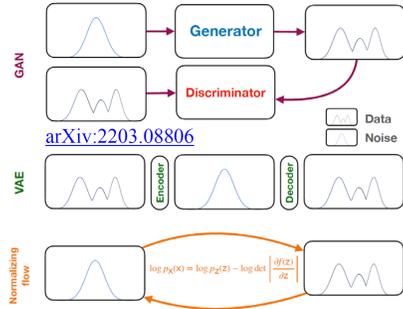
Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kudithipudi Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avnika Narayan Deepak Narayanan Ben Newman Allen Nie Juan Carlos Nieves Hamed Nilforoshan Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou Percy Liang*¹

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.

¹Corresponding author: pliang@cs.stanford.edu

Generative models in HEP



Diffusion

Autoregression



基于ML的

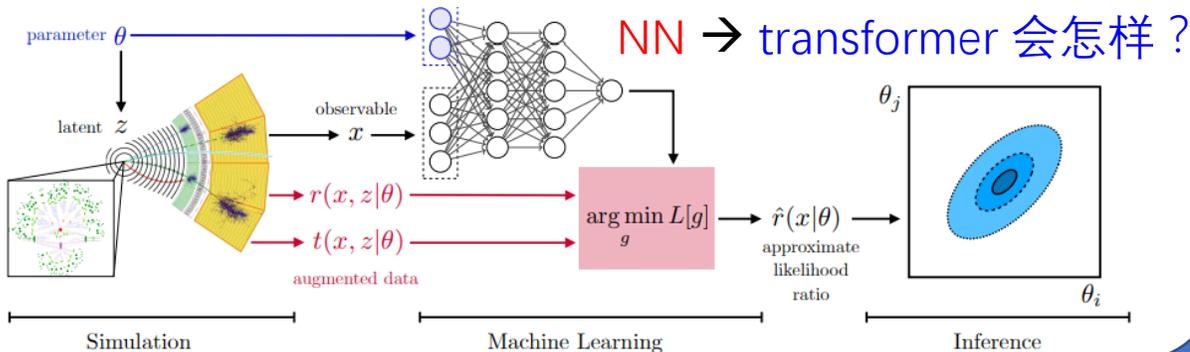
- 快速模拟
- 事例重建
- 参数估计
- 反卷积
- 异常检测

把HEP问题映射为工业界问题，用先进工具加以解决

Foundation models :

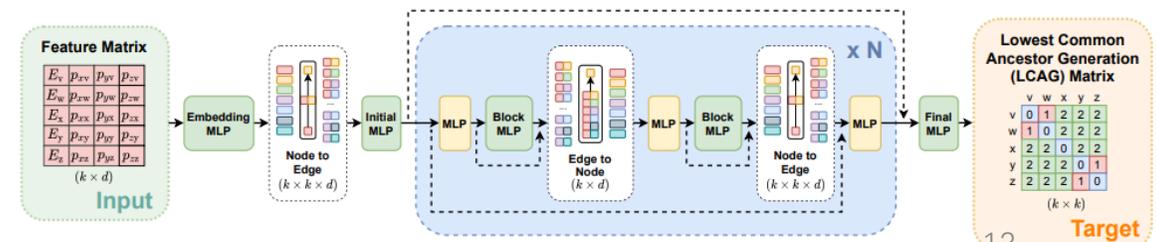
- 算法
 - Diffusion, Autoregression, ...
- 架构
 - Transformer, ...
- 通用数据表示

例一：Sim.based Inference [PRL 121 111801 (2018)]



例二：Full Event Interpretation at Belle II

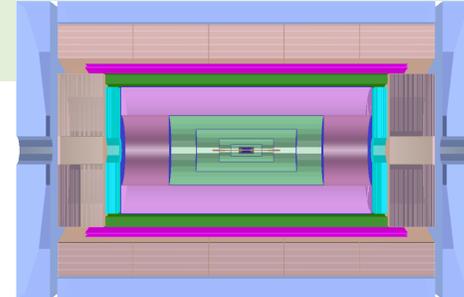
- cut \rightarrow BDT \rightarrow transformer



chatGPT/GPT-4 inspired 探测器设计、研发和新物理寻找

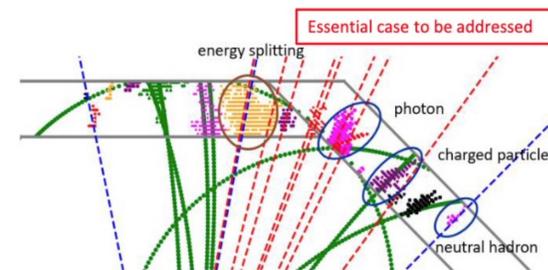
(ChatGPT/GPT-4 like detector design, R&D and search for new physics)

探测器的设计



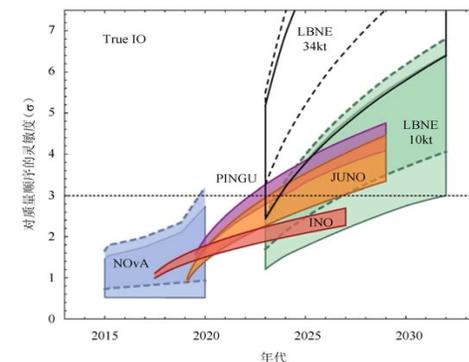
+

探测器的模拟

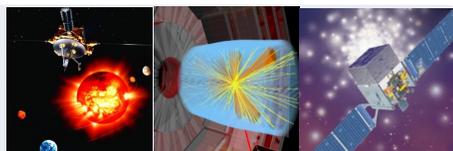


+ +

新物理的发现



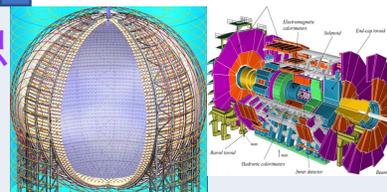
告诉AI
设计目标



开发 ChatGPT/GPT-4 inspired
高能物理探测器/新物理寻找
AI设计系统



各种物理知识
设计经验、
探测器例子、
新物理模型



积累了大量的粒子与物质
相互作用的经验

Geant 4

Born from the requirements of large scale HEP experiments

Widely used also in

- Space science and astrophysics
- Medical physics, nuclear medicine
- Radiation protection
- Accelerator physics
- Pest control, food irradiation
- Humanitarian projects, security
- etc.
- Technology transfer to industry, hospitals...

Most cited "engineering" publication in the past 2 years!

强大算力和存贮系统
(超算中心和数据中心)

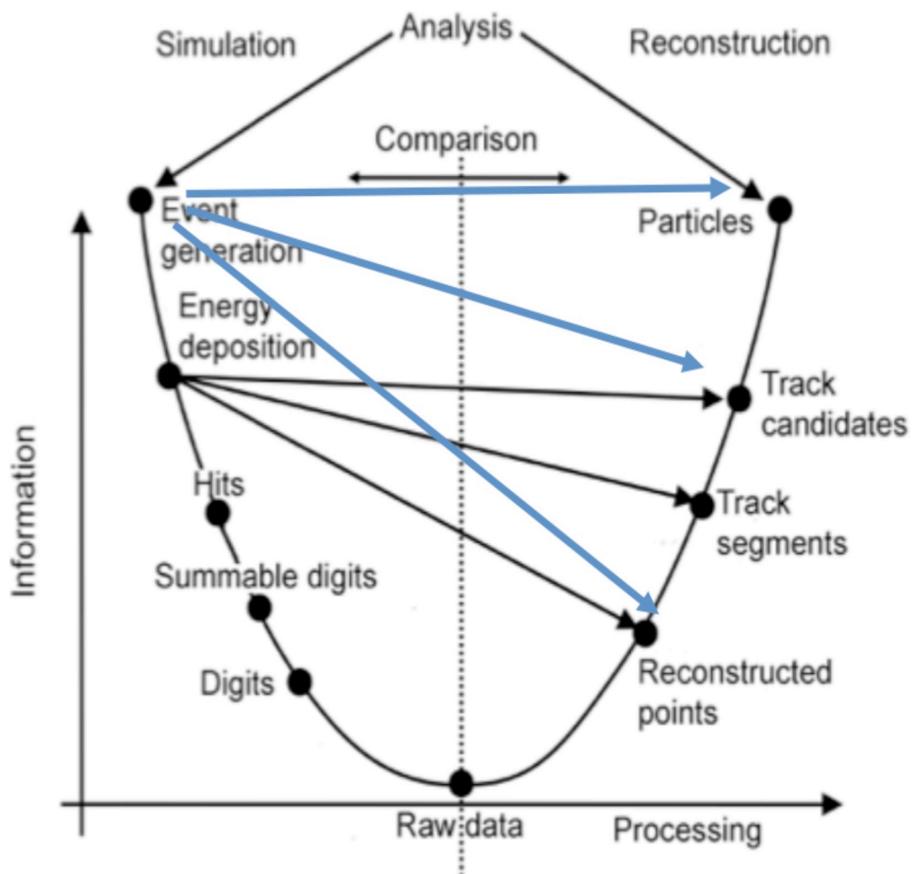


人工智能驱动的科学
研究 (AI for Science)

- 利用AI进行探测器的设计、研发、新物理寻找等一系列过程；
- 建立实验高能物理研究的新范式 and 标准。

传统流程

Generative ML as shortcuts



加速现有研究：

- 传统基于MC的模拟计算量巨大（占离线处理的一半以上）
- 基于ML的模拟可以兼顾速度和精度
 - ML全部或部分替代Geant4 → 加速模拟
 - ML模型可微 → 探测器设计优化

新研究方法：

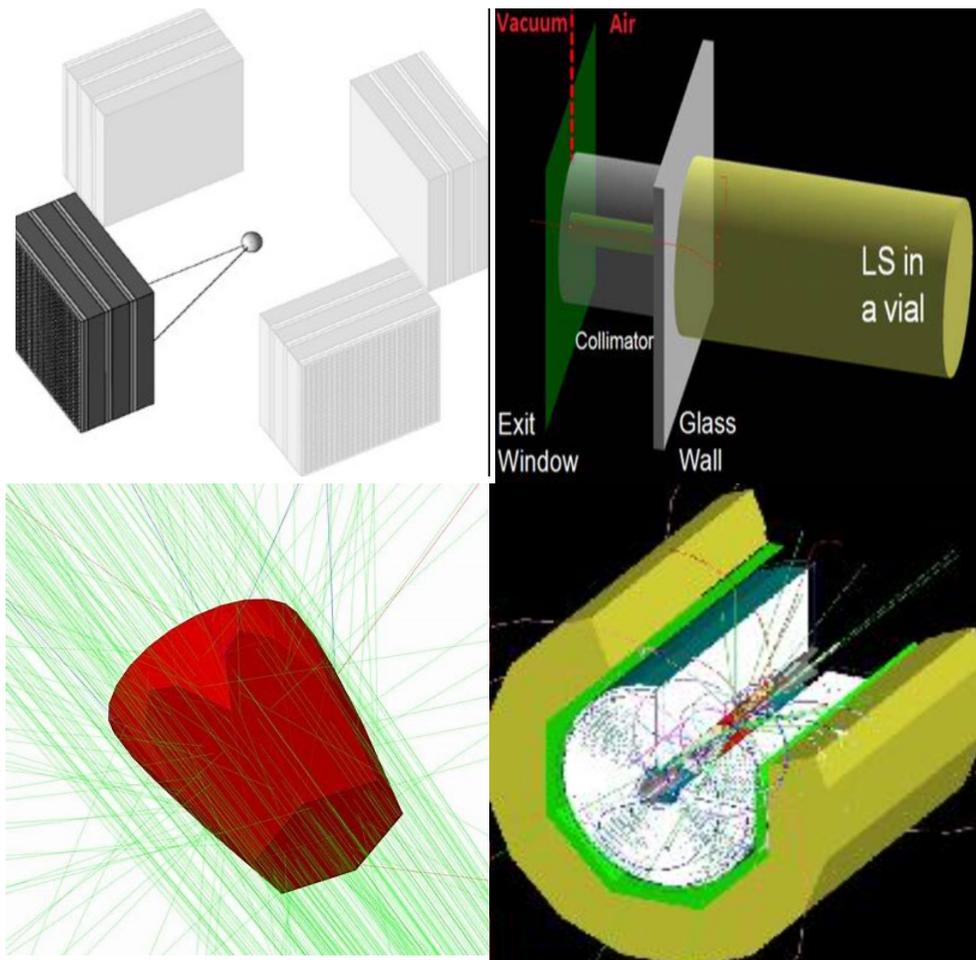


基于ML模拟的

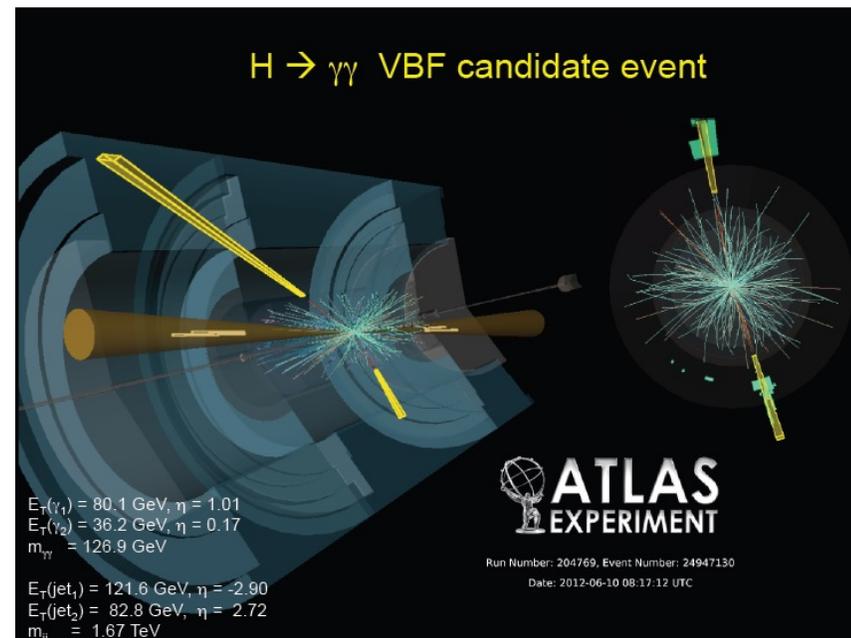
- 参数估计：分波分析
- 反卷积：分波分析
- 异常检测：拟合优度，稀有信号寻找

类chatGPT/GPT-4 探测器设计、研发和新物理寻找

(ChatGPT/GPT-4 like detector design, R&D and search for new physics)



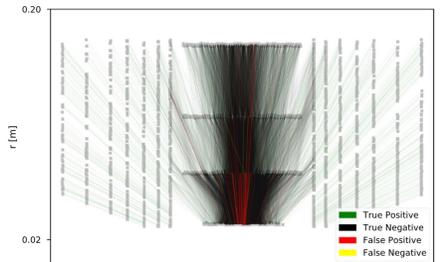
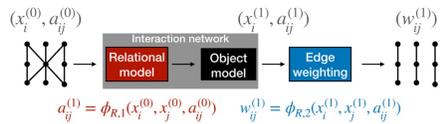
- 利用Geant4建立各种几何形状探测器原型和粒子相互作用的信息，并喂给AI。
- 将设计的要求输入给AI
- 训练并优化。



- 以新粒子（比如希格斯粒子作为新物理信号）和相关事例在探测器行为特征（比如Hit）、物理知识输入给AI进行训练。
 - 取代物理分析这一步。
- 进行可行性验证
- 算力是关键

粒子物理实验：end-to-end探测器重建&分析

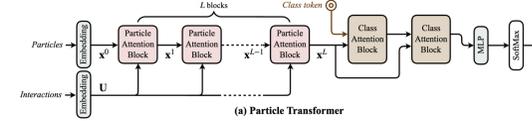
郭方毅



Comput.Softw.Big.Sci. 5, 26 (2021)

All of these can be combined into Large Model in the future!

arxiv 2202.03772



	All classes		$H \rightarrow b\bar{b}$	$H \rightarrow c\bar{c}$	$H \rightarrow gg$
	Accuracy	AUC	Rej _{50%}	Rej _{50%}	Rej _{50%}
PFN	0.772	0.9714	2924	841	75
P-CNN	0.809	0.9789	4890	1276	88
ParticleNet	0.844	0.9849	7634	2475	104
ParT	0.861	0.9877	10638	4149	123
ParT (plain)	0.849	0.9859	9569	2911	112

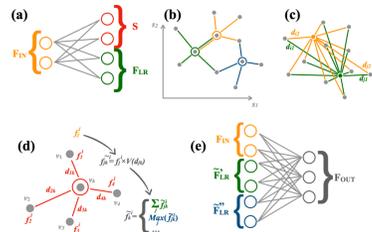
探测器原始信号

径迹室：寻迹+拟合
量能器：聚类

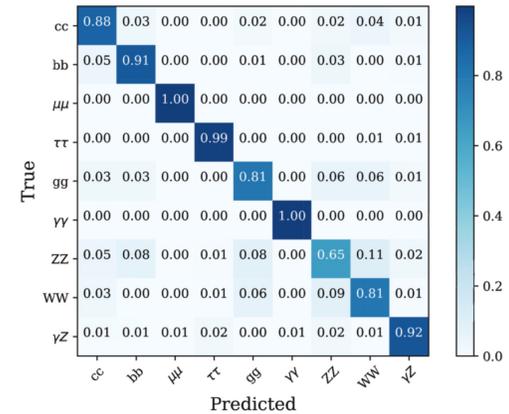
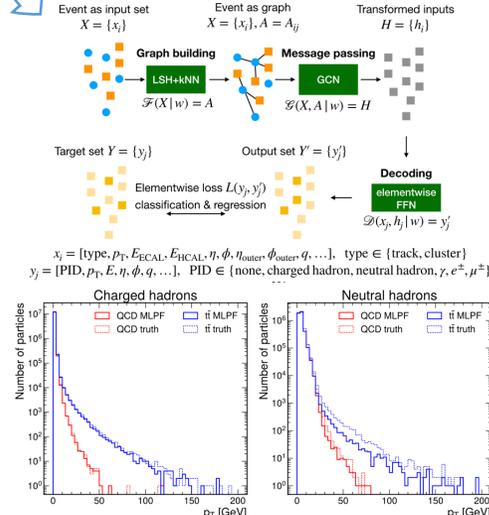
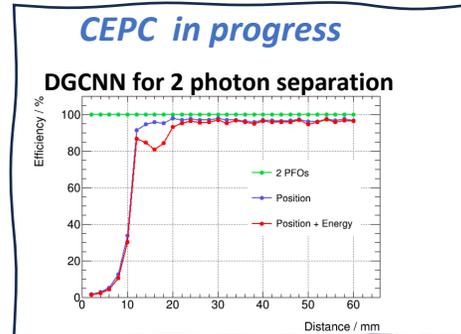
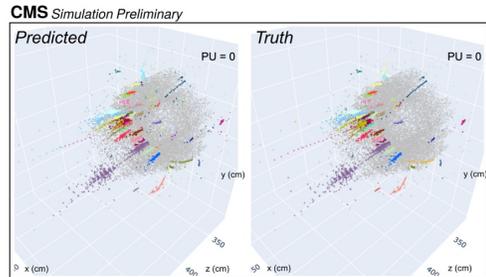
PFA重建

粒子鉴别事例分类

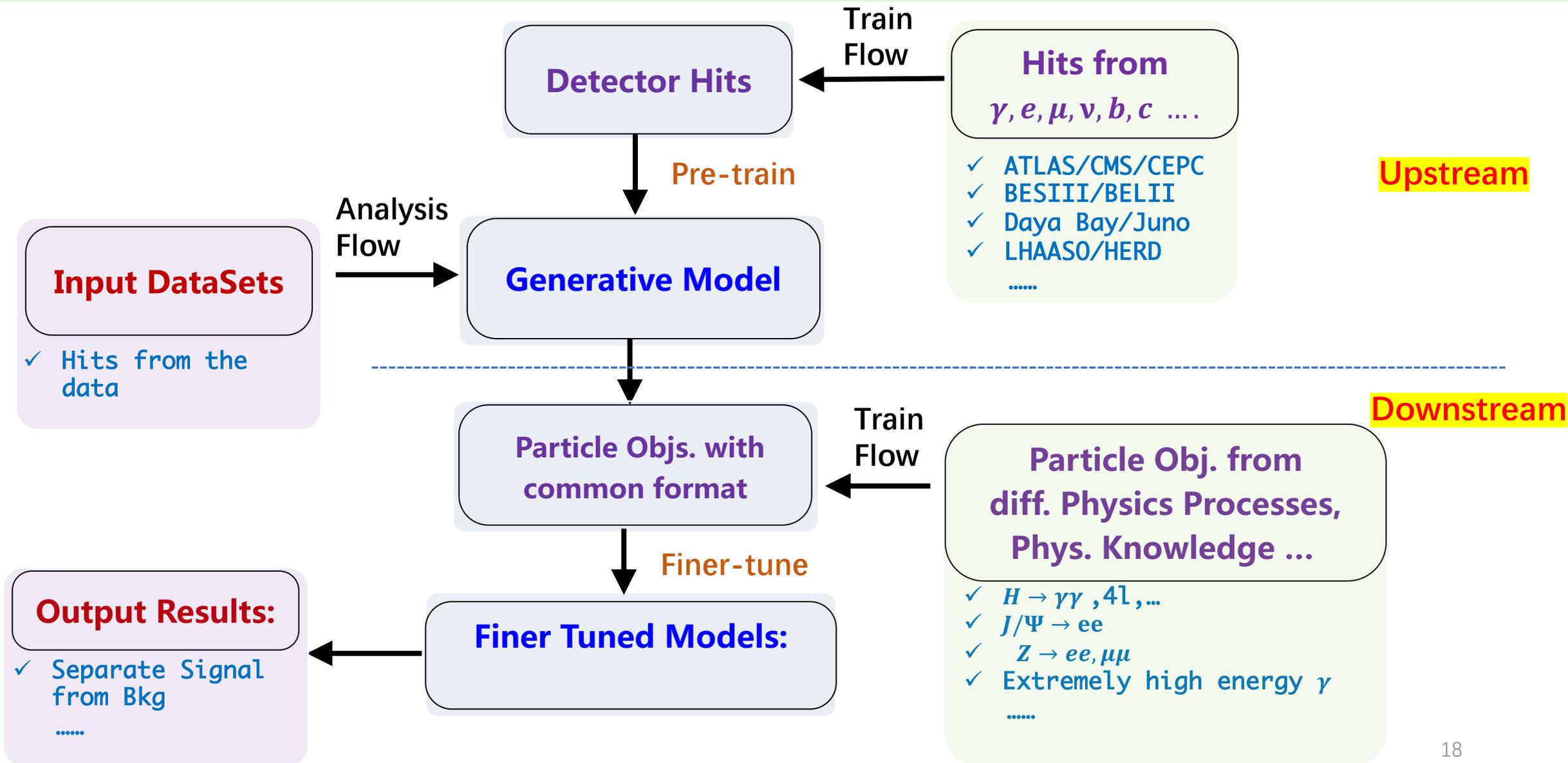
物理分析



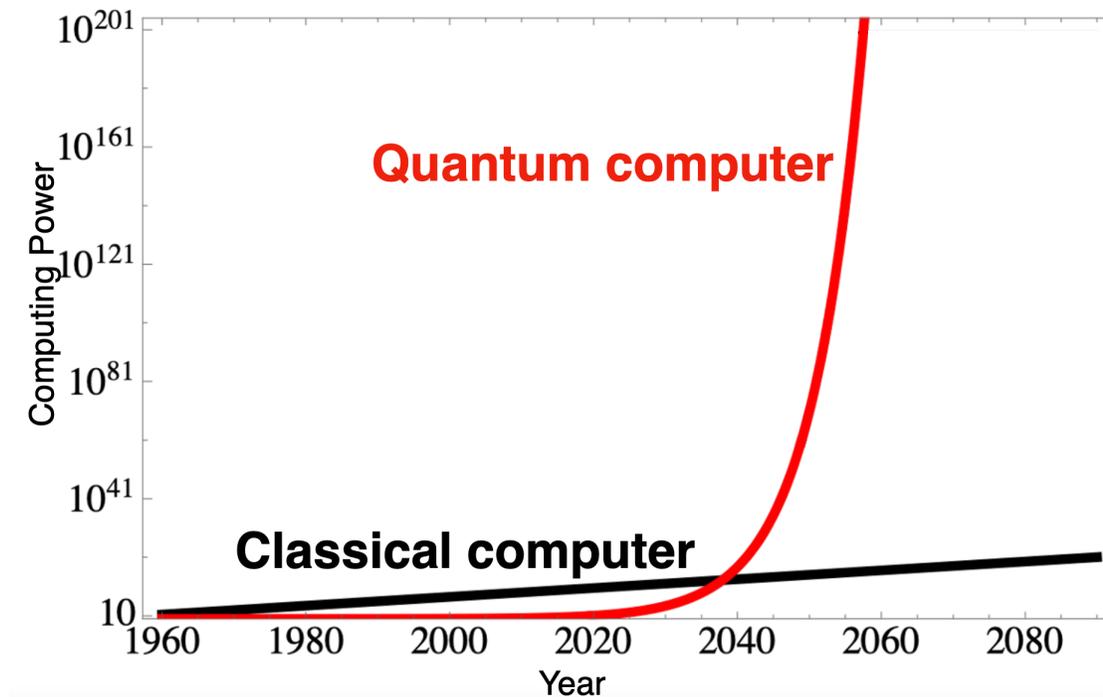
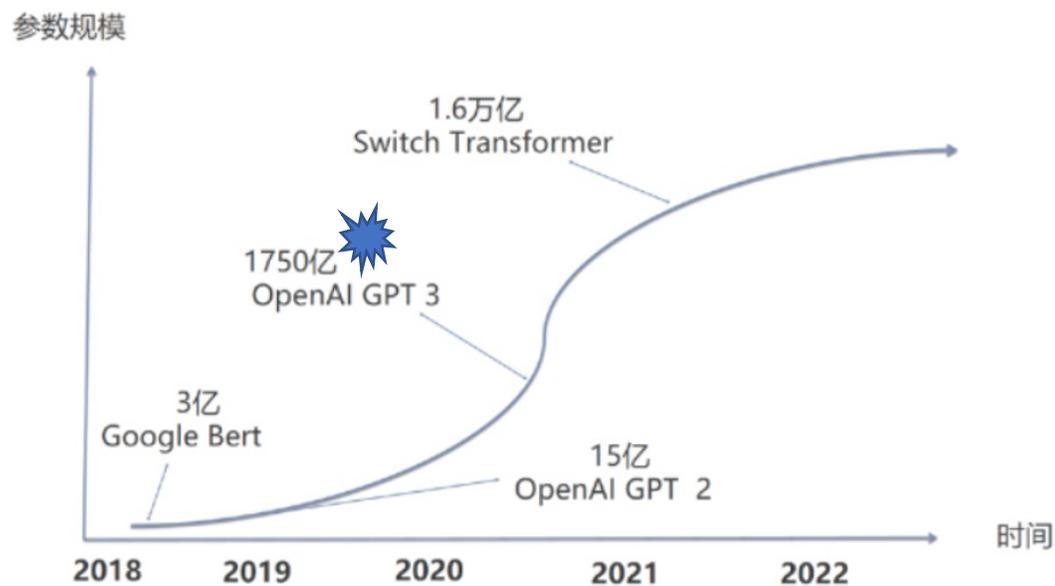
从Hit出发，进入了一次尝试：CEPC 4th detector, 2-photon分离



ChatGPT inspired 物理分析

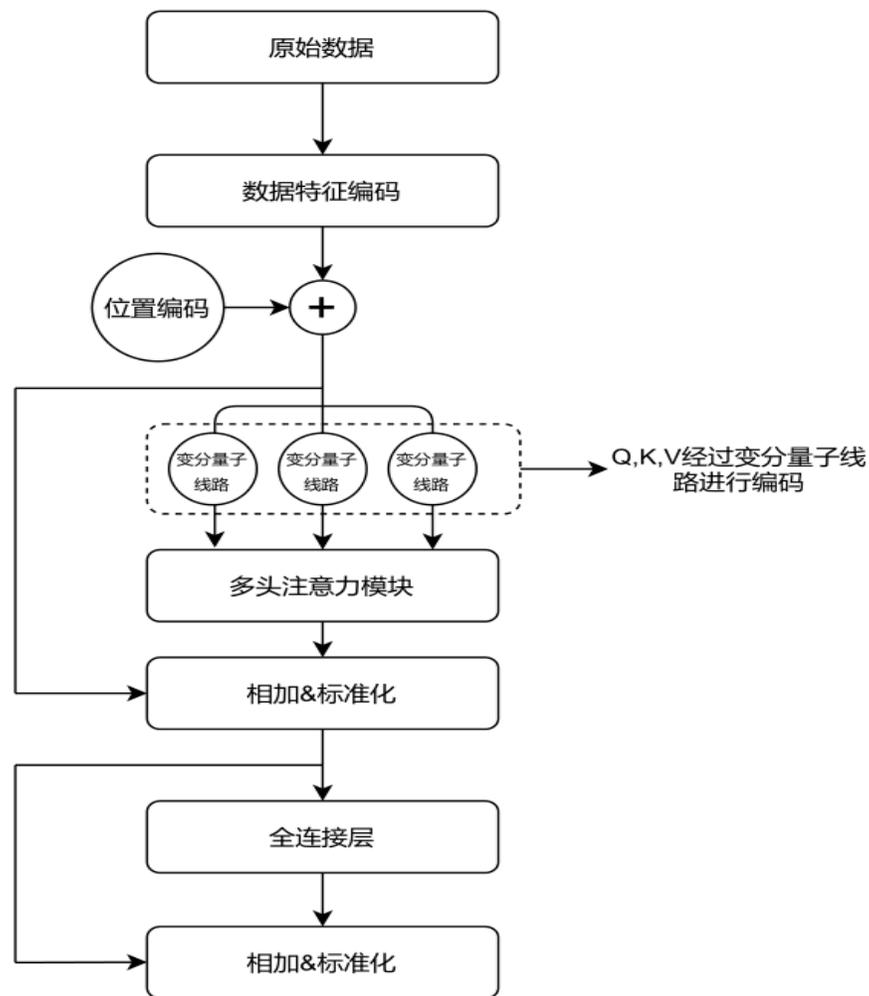


量子机器学习：有望满足AI对计算能力的需求



- 传统计算机的算力增长速度可能难以满足AI参数指数增长的速度。
- 量子计算机算力指数增长有望解决这一挑战。

Quantum Particle Transformer的开发

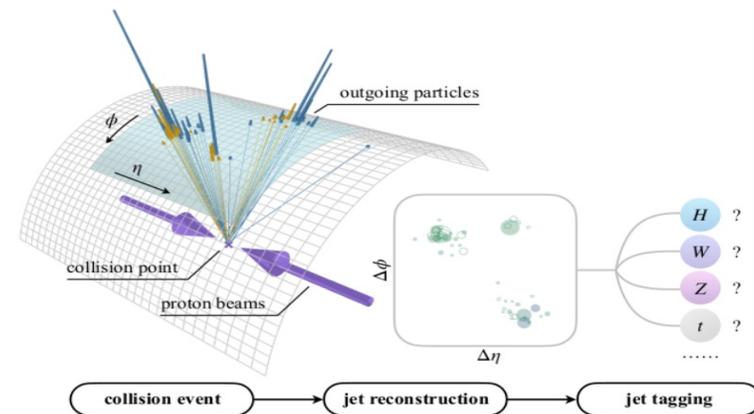


➤ Transformer是目前最先进的自注意机制的机器学习算法

Q : Query
K : Key
V : Value

$Q \cdot K$ 越高给的注意力权重越高

➤ ParticleTransformer是将Transformer+物理知识应用于实验高能物理领域，并在CMS c-tagging上获得显著成功。



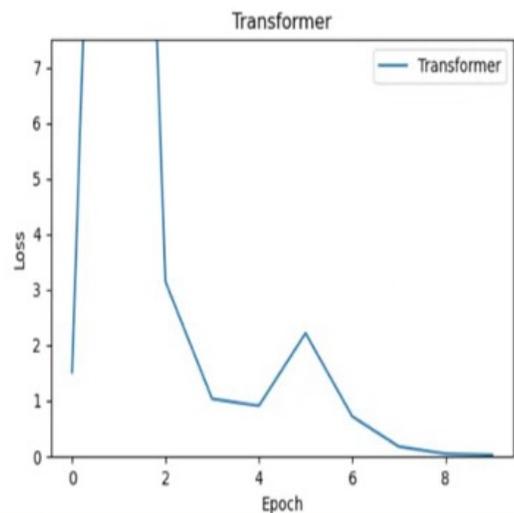
Quantum ParticleTransformer的研发：

Q, K, V 这块用量子计算的方法来计算：有望算力显著提升

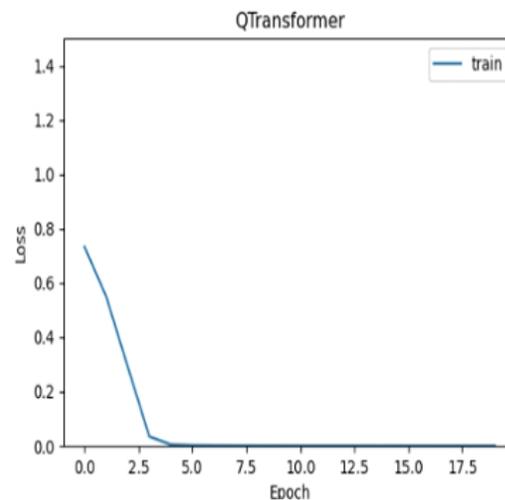
量子人工智能 (Q AI)

本源量子李蕾

经典Transformer



Quantum Transformer



算法模型	Accuracy	precision	recall	F1-score
QTransformer算法	0.875	0.86	1.0	0.92
经典Transformer算法	0.625	1.0	0.5	0.67

- 在本源量子的关于情感测试的研究中，Quantum Transformer比经典的Transformer具有明显性能提升。
- 量子人工智能是否能: 更好地模拟人脑的运作模式？模拟人类思维中的灵感？模拟人类思维的非逻辑部分而更加人性化？
- 人工智能中的“涌现”在量子人工智能中会出现？甚至有不同的特点？

未来人与量子AI (Q AI) 的关系

AI² Authentic Intelligence
Artificial Intelligence

量子计算机/群：
训练模型：
知识化、个性化
超强算力
实时迭代微调



人的言行、举止、甚至思考

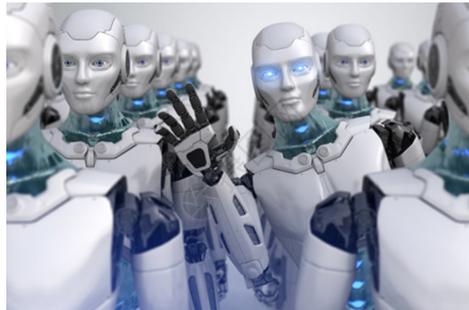
向量子机群实时传递形成
“个性化模型”

人类：从纷杂的事务中解脱出来，
思考从0到1的问题，探索未知，
增强社会主义生活幸福感……



无线实时传输
“个性化模型”和知识等

服务、分身等功能



指令

分身机器人：

Q AI 情感优势可能产生
与主人性格类似的机器人
(模型性格化)：

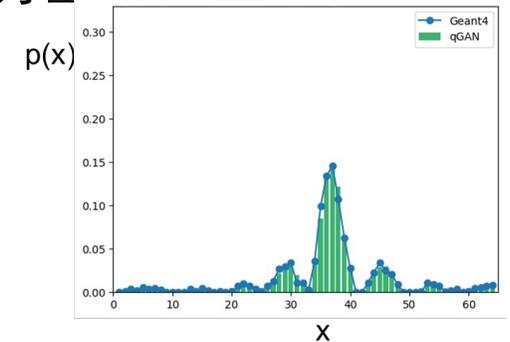
可以给主人提供各种服务，
并代表主人办理个性化事务。

国内外合作

- 所计算中心量子平台的有力保障
- 与国内外量子计算机硬件团队的合作
 - **本源量子**：平台对机器学习成熟的。只6量子比特硬件免费。Quantum SVM, Quantum Transformer, Quantum GAN（方亚泉，大川英希等）
 - **夸父（北京量子研究院）**：多达136量子比特硬件免费。在10量子比特硬件用QSVM成功了。136量子比特硬件对径迹重建很重要。现在和DESY的同事们准备VQE径迹重建。Quantum GAN应用也准备中。（方亚泉，大川英希等）
 - **科大上海研究院**：主要问题是如何把应用程序量子线路转换成平台支持的高效率等效量子线路（李卫东）
 - 与所外大学在量子计算理论/实验方面合作：李英英（科大）、黄性涛团队（山大）、周辰（北大）
- **将加强与企业大厂的合作以及基金部门的联系**

量子机器学习研究进展

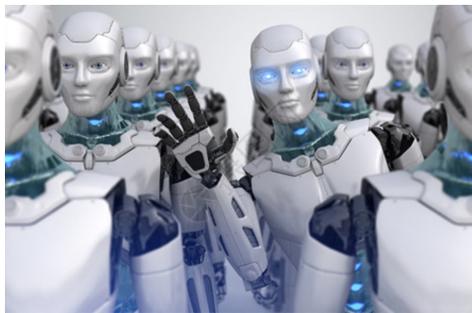
- 信号判别问题（量子支持矢量机QSVM、量子变分电路QVC等）
 - CEPC实验物理分析（沙其雨, Abdualazem Fadol, 方亚泉等）
 - $e^+e^- \rightarrow ZH \rightarrow q\bar{q}\gamma\gamma$: 文章已完成, 待发表 ([arxiv:2209.12788](https://arxiv.org/abs/2209.12788)) (课题进展11)
 - BESIII实验物理分析（李卫东、邹佳恒等）
 - Muon/Pion粒子鉴别：文章已投出, 待发表
 - BESIII上 $Z_c(3900)$ 的寻找：5月份CHEP2023会议Poster, 开始准备文章
- 其它：量子ParticleTransformer算法（方亚泉, 大川英希等）
- 量子机器学习算法的优化（量子控制比特法、傅立叶变换的方法）（方亚泉, 童宇）
- 国际合作和德国DESY
 - 在未来对撞器径迹重建 [VQE, QGNN等]（大川英希）
 - 数据生成问题（量子生成对抗网络QGAN）基于CLIC公开的3D量能器数据
 - 1D、2D量能器平均能量分布的快速模拟（黄晓忠）
 - 类似研究用本源和夸父的平台（大川英希）
- 今年3月召开量子计算的研讨会；



量子AI 研究规划

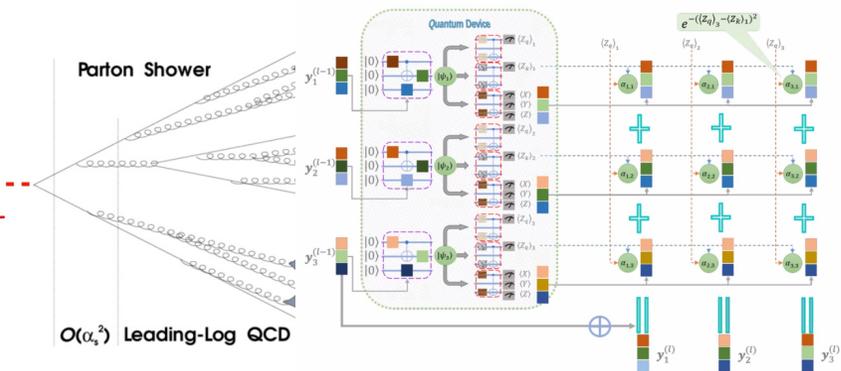
量子AI + 大模型 + 机器人：AI²

- 真正智能和人工智能的完美结合
- 实现分身机器人
- 个性化模型让人类实现智能/情感永恒
- **建立研究中心（比如在海南）**



量子模拟的应用和量子机器学习算法的研发

- 主要量子计算在HEP应用
- 量子ML算法的研发，面向未来布局
- **建立创新团队**



(量子) 机器学习和大模型的结合

- 实现算力的暴涨
- 人类情感有可能在量子AI实现
- 量子计算机的保真性能是关键
- **专门编制的人员**

开发 ChatGPT/GPT-4 like
高能物理探测器/新物理寻找
AI设计系统



总结

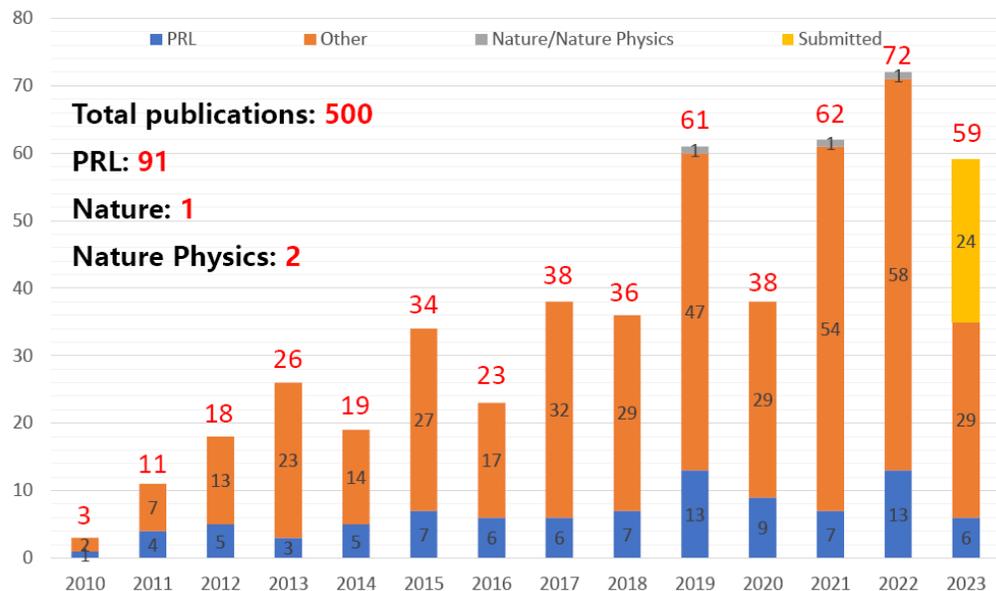
- 所机器学习的创新课题进展良好：
 - 出了一些在各合作组有显著性的成果：
 - 收获Transformer等先进机器学习算法的红利
 - 尽快发表相关的文章（时间窗口）
 - 生成型模型相关的研究应纳入咱们所创新课题的内容
- ChatGPT inspired的探测器设计、探测器模拟、物理分析是在实验高能物理可能用到的一些方向，应该有更加广泛的应用
 - 设计了相应的物理分析的思路
 - 特别想测试一下在物理分析中究竟有没有“涌现”
- 量子机器学习有潜力解决：
 - 算力问题
 - 有些运作模式（尤其是人脑）可能是量子过程的有效模拟问题
 - 分身机器人是未来发展的一个思路：敢想敢做，“再不努力我们可能成为纪录片”了
- 8月11-14日将在青岛召开量子计算和机器学习的[研讨会](#);
- 12月4-6日，在海南三亚的人工智能大会（高能物理section）

backup slides

课题进展

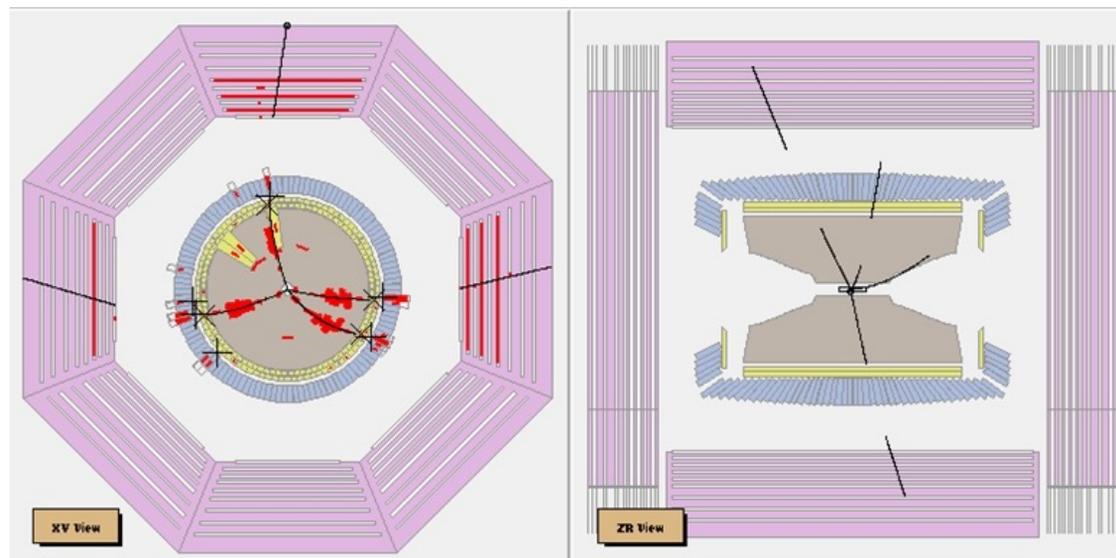
- 北京谱仪III (BESIII) 自2009年运行至今, 已采集了国际上最大的 τ -粲能区数据样本, 取得了一系列国际领先的物理成果。
- BESIII实验中, 末态带电粒子的径迹探测是极其关键的环节, 目的是获取带电粒子的**空间位置**和**动量**, 并重建**事例顶点**。

BESIII publications (May 9, 2023)



[BESIII Physics Journal Publications](#)

- 探测器记录下入射粒子留下的原始信息, 需要在离线数据处理中通过模式识别和径迹拟合, 精确地重现出入射粒子的飞行轨迹, 这就是**径迹重建**, 其性能对BESIII物理结果有着非常重要的影响。

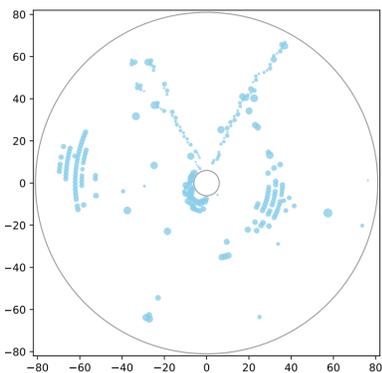


The first charmed-meson pair event seen in BESIII

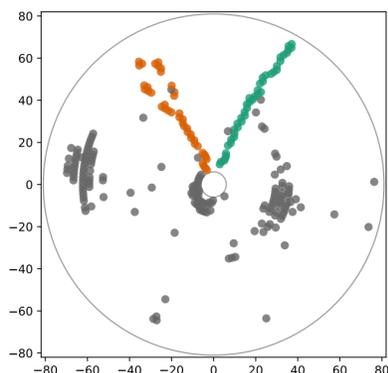
课题进展

传统径迹重建方法

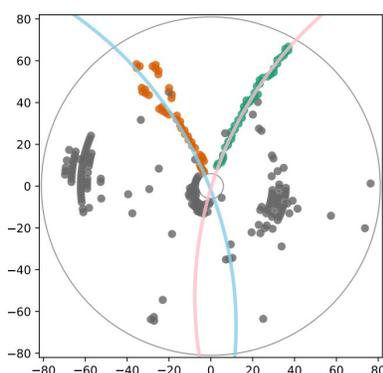
- **径迹寻找** - 寻找可能的入射粒子径迹以及属于该径迹的击中
 - 模版匹配
 - 数学变换- 共形变换(Conformal mapping), 勒让德变换(Legendre transform), 霍夫变换(Hough transform)
- **径迹拟合** - 通过拟合确定径迹参数及误差矩阵
 - 全局最小二乘拟合 - 将径迹视为标准螺旋线, 快速地得到径迹参数
 - 分段拟合 - 便于考虑多次散射、磁场不均匀、能量损失等效应, 如Kalman filter方法



探测数据



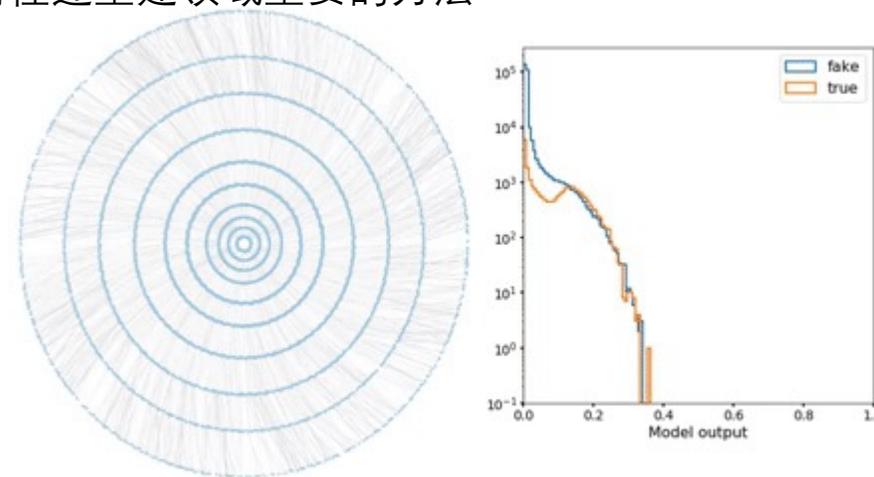
径迹寻找



径迹拟合

• 径迹重建引入机器学习的原因

- 提升特殊事例的寻迹效率
- 应对BEPCLII升级后更高的噪声比例
- 训练好的机器学习模型可以大幅减少重建所需的时间, 减少重建时候所需的计算资源
- 径迹重建是一个模式识别的问题, 是机器学习所擅长的任务
- 在径迹重建中应用机器学习在国外已经有相当多的研究, 是目前径迹重建领域重要的方法



Performance of a track finding Graph Neural Network from [Exa.TrkX](#)

量子线路的并行运行算法研究

- 现状与动机：目前计算核函数量子线路是单个串行模式，耗时低效。并行运算的算法有望显著降低整体运行时间，提高效率。

- 控制量子比特法：

- 增加控制量子比特
- 每个量子比特控制不同的U门
- 在模拟器上已经实现。
 - 但真机上受限于保真度

- 傅立叶变换的方法：

- 基于量子傅立叶变换
- 将特征空间分为网格并计算网格上点的内核函数
- 将事例的内核函数近似于网格上点的内核函数
 - 目前完成电路设计，但难点在于随着维度的增加，覆盖全网格的难度。
 - 计划利用经典机器学习算法来作为分区，实现降维。

