

高能物理研究所第三次深度学习研讨会

# 高能同步辐射的物理大脑研究

赵丽娜

多学科研究中心

2023年6月15日

# 报告提纲

- 1. 高能同步辐射的数据驱动研究**
- 2. 高能同步辐射的物理大脑**
- 3. 相关的方法学与技术**
- 4. 研究工作基础概述**
- 5. 小结短期研究目标**

# 1. 高能同步辐射的数据驱动研究—背景概述

## □ 高能同步辐射大科学装置科学研究的数据驱动方法学

数据驱动物理研究	研究目标	数据来源	数据特征	研究方法
粒子物理	发现基本粒子、研究粒子间相互作用、探索新物理现象	粒子探测器 (如：大型强子对撞机)	高能粒子碰撞产生的 粒子轨迹数据、 能量沉积数据	粒子识别、模式识别、 事件重建、异常检测 (主要：ParticleNet)
同步辐射	支撑 <b>多学科</b> 基础研究、 分析 <b>材料结构</b> 、 <b>生物体系</b> 的物理化学性质， 探索物理新现象	同步辐射光源 (如：高能同步辐射光源)	<b>复杂多模态数据</b> 包括图像数据、 衍散射数据、 吸收谱学数据 具有噪声、需预处理	<b>多方法学</b> (近百种方法学) <b>多种机器学习方法并用</b> (如：CT图像LRCS-Net、 衍射FCU-Net、 SEDCNN、 谱学MPNN等)

# 2.高能同步辐射的物理大脑—提出

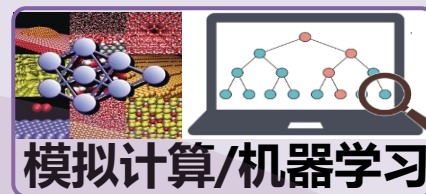
## 智能光源

- 光源操作系统
  - 统一框架、统一系统、统一接口
- 智能系统融合
  - 实体设施、软件系统、数据标准融合



## 物理大脑

- 数字孪生大脑
  - 实时交互、实时处理、自动调控
- 基于深度学习的数据挖掘
  - 特征提取、总结规律信息



## 科学大数据流控

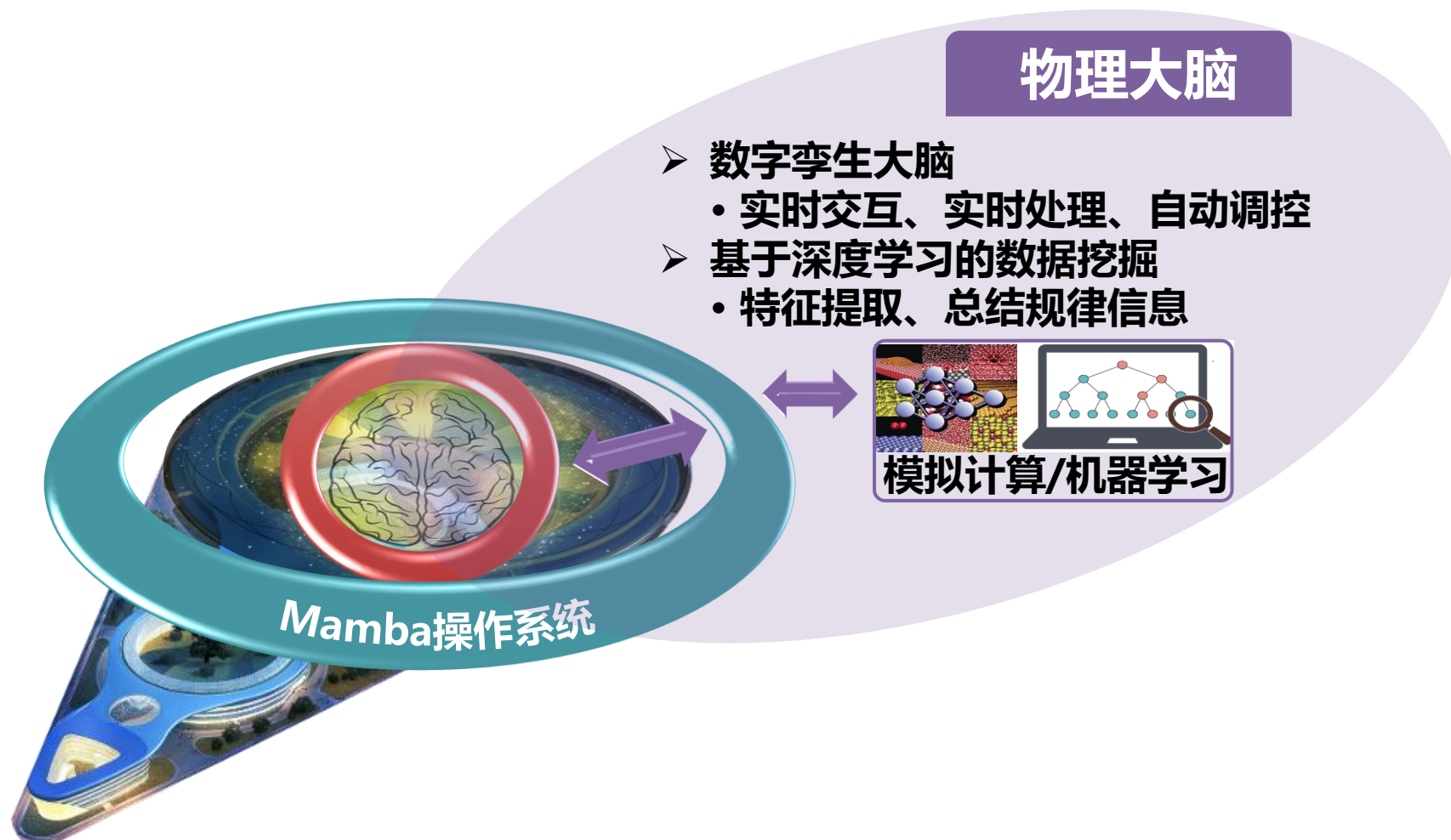
- 全生命周期的数据 workflow
  - 数据采集、数据调度、数据处理
- 基于科学数据流的科学大模型孵育

## AI数据银行

- 专业领域知识数据银行
- 智能装置群融合实验数据银行
- AI大模型

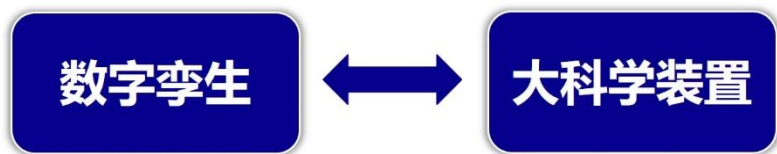


## 2.高能同步辐射的物理大脑—蓝图

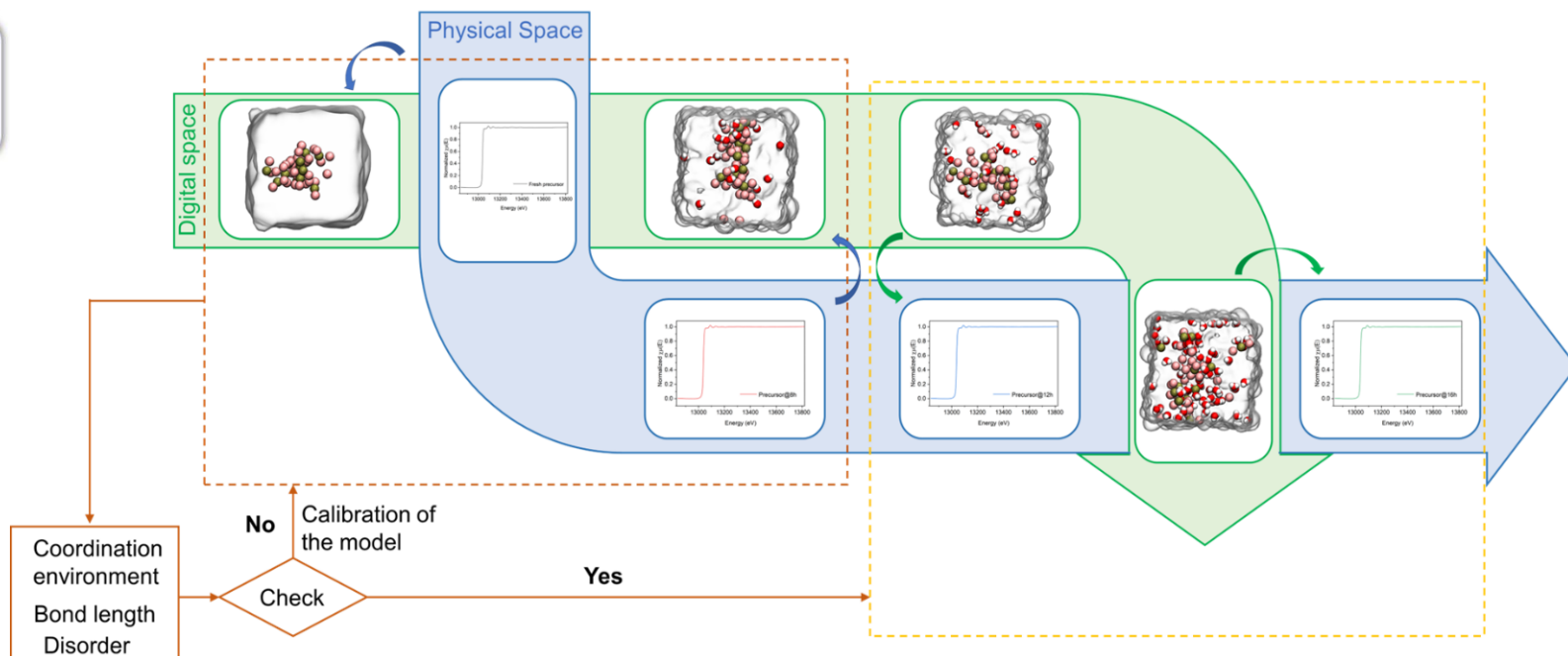


# 3. 高能同步辐射的物理大脑—分子模拟方法学

- 结合多尺度分子模拟，建立高能同步辐射的数字孪生系统



- 以能源材料构效关系调控科学问题为例

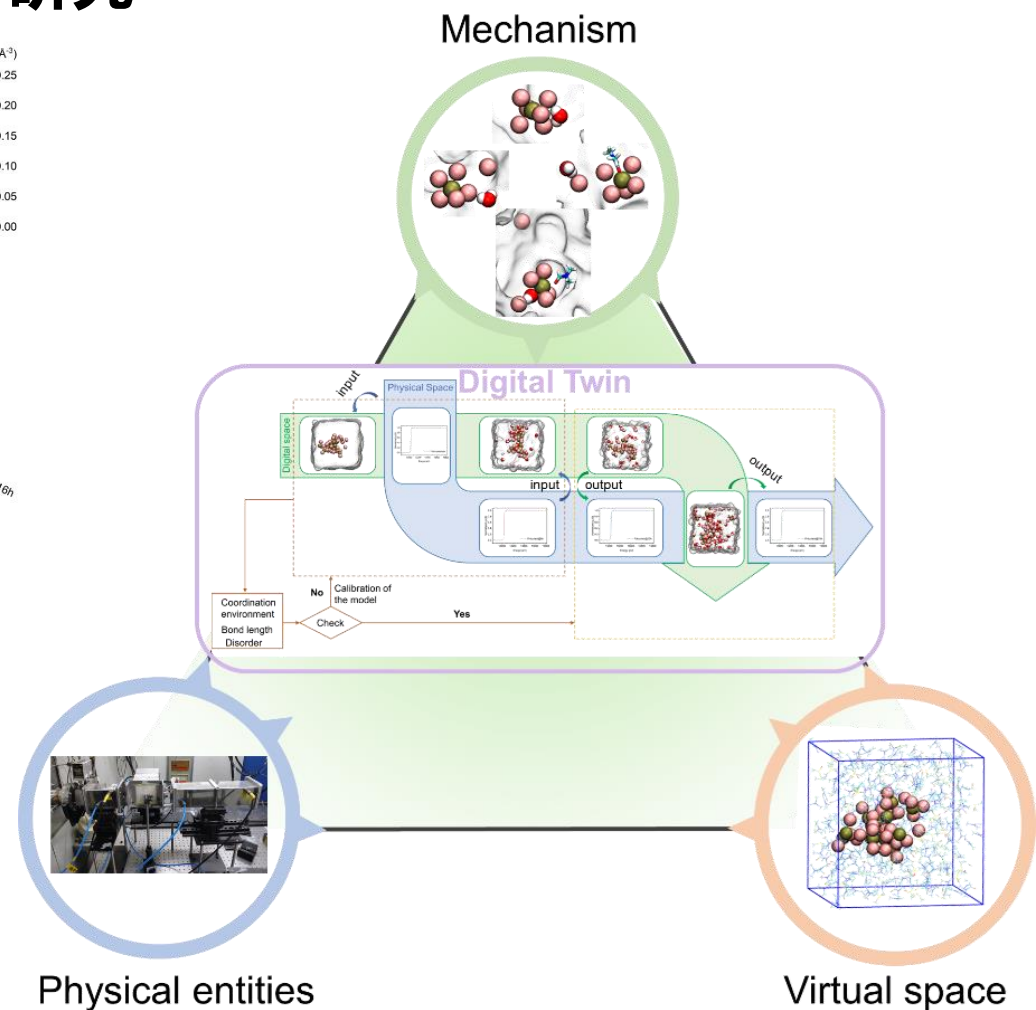
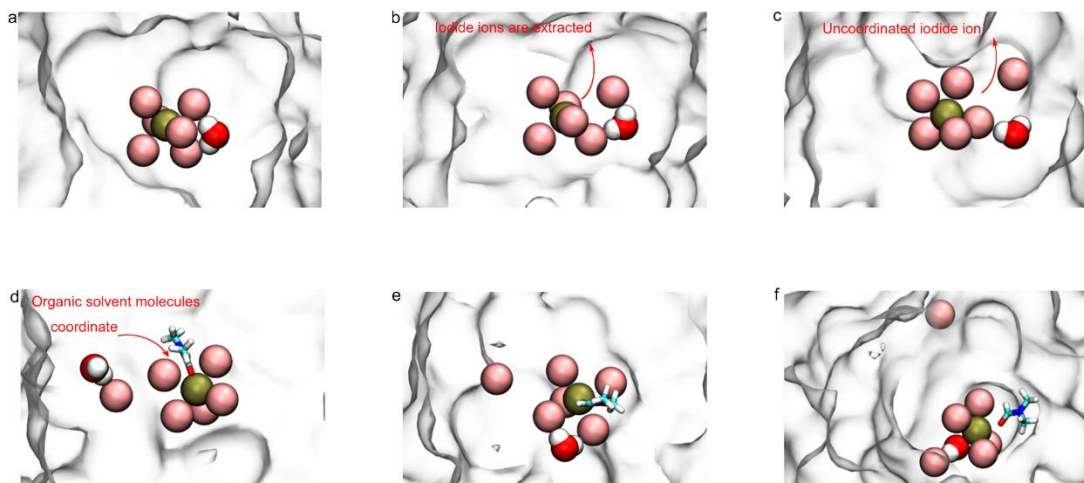
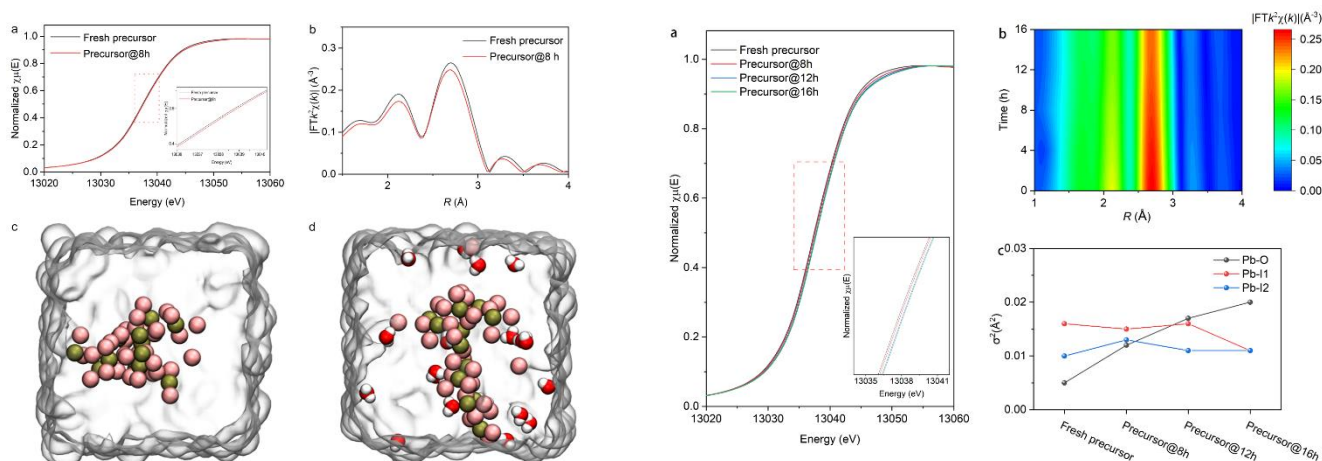


- 研究目标：水分子调制钙钛矿前驱体结构分子机制
- 通过反复信息交互，拓展从静态到动态的吸收谱学研究。
- 引入DeepMD、DeepKS提速数字空间模拟效率。



# 3. 高能同步辐射的物理大脑—线站物理方法学

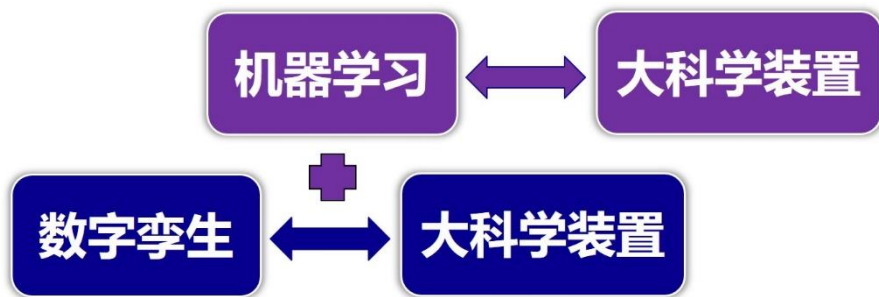
## □ 数字孪生赋能，拓展从静态到动态的吸收谱学研究



*J. Phys. Chem. Lett.* 14, 4876, 2023.

# 3. 高能同步辐射的物理大脑——机器学习方法学

应对实时预测需求，建立对应大科学装置的机器学习系统。



同步辐射海量衍射数据在线解析的AI for Science研究取得进展

原创 科技兴国的 FUTURE远景 2023-04-15 06:03 发表于上海

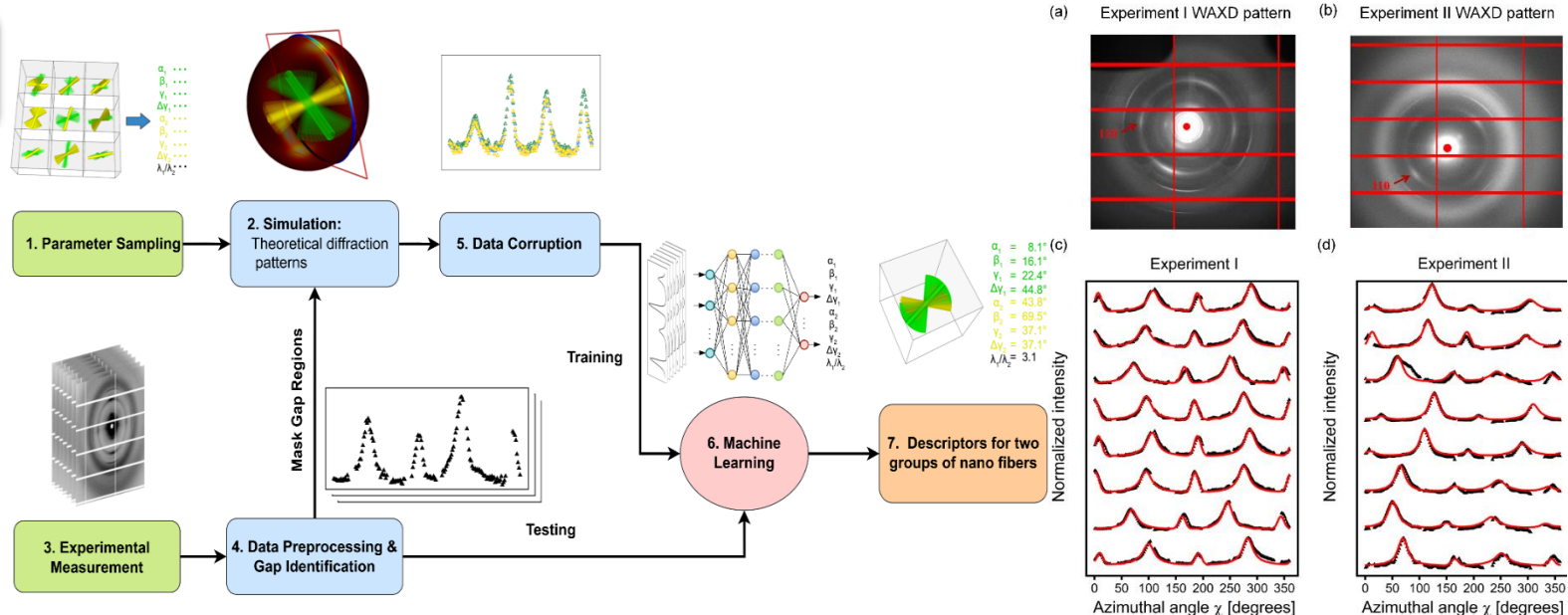


FUTURE | 远景

FUTURE | 远景 闵青云 编辑

撰稿 | 张一、赵丽娜

近日，面向高能同步辐射光源（HEPS）海量衍射图像降噪、实验数据在线解析的迫切研究需求，HEPS光束线软件系统张一团队与中科院高能物理研究所赵丽娜AI研究团队密切合作，取得对应的AI for Science研究阶段性进展，研究成果在领域重要学术期刊 IUCrJ 与npj Computational Materials 上在线发表。



贡献：实现实验数据高精度 ( $R^2 > 0.91$ )、快速（比现有拟合方法快 $\sim 10^4$ 倍）解析，并能够跨光源、线站应用，为知识驱动机器学习解析奠定基础。

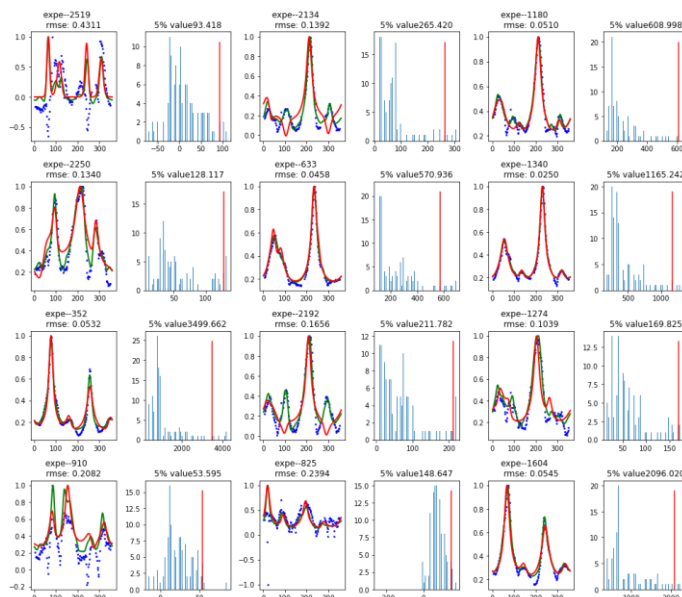
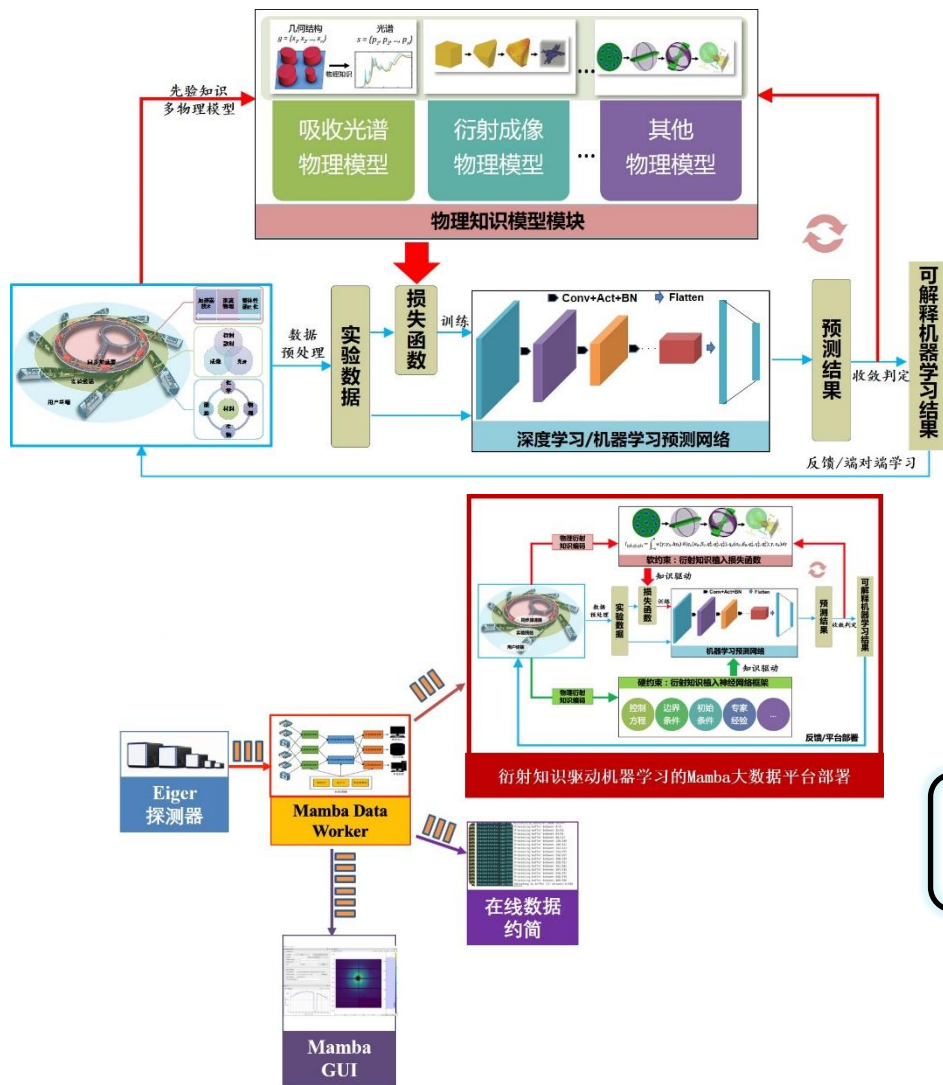
大科学装置 → 大科学装置 → 科学大数据解析

IUCrJ 10, 3, 2023.

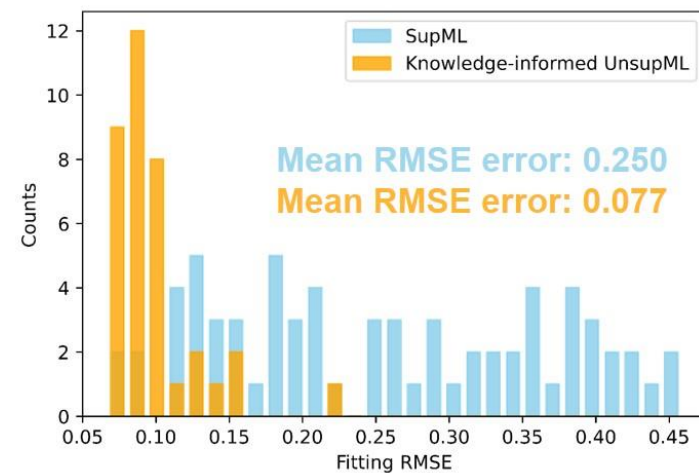


# 3. 高能同步辐射的物理大脑—线站物理方法学

## 解决多物理知识与机器学习一致性嵌入的问题



物理驱动 vs 纯数据驱动

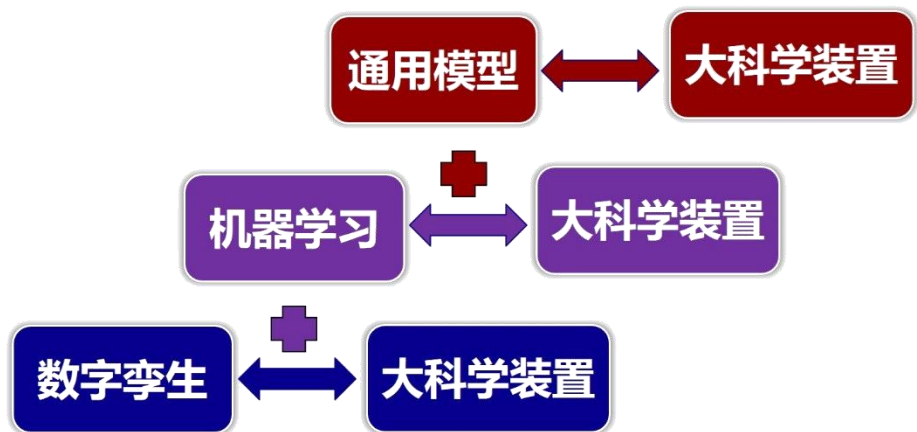


贡献：X-射线衍射知识软驱动的机器学习方法，实现高精度、自监督、可解释的解析研究，为系统的知识驱动科学大数据解析奠定基础。

# 3. 高能同步辐射的物理大脑——线站物理方法学

## □ 大语言模型能够从语料学习中嵌入物理知识，解决相应领域问题

..... 物理大脑



- 数据集构建
  - 文献筛选
    - 时间区间：2019-2023
    - 文献类型：期刊（综述）论文、会议论文
    - 语言：英文
    - 期刊要求：期刊影响因子>10
    - 领域范围：实验物理中心（41），加速器中心（44篇），粒子天体物理中心（44），计算中心（40），多学科中心（45），理论物理室（36）
  - 问答对提取
    - 对筛选出的文献使用HEPAI根据一定的Prompt构建问答对。
    - 对问答数据在文献中相应位置进行定位，进行正确性判断。
    - 问答对格式：标注DOI，标注人，关键词，领域信息。
    - 问答对数量：4918对

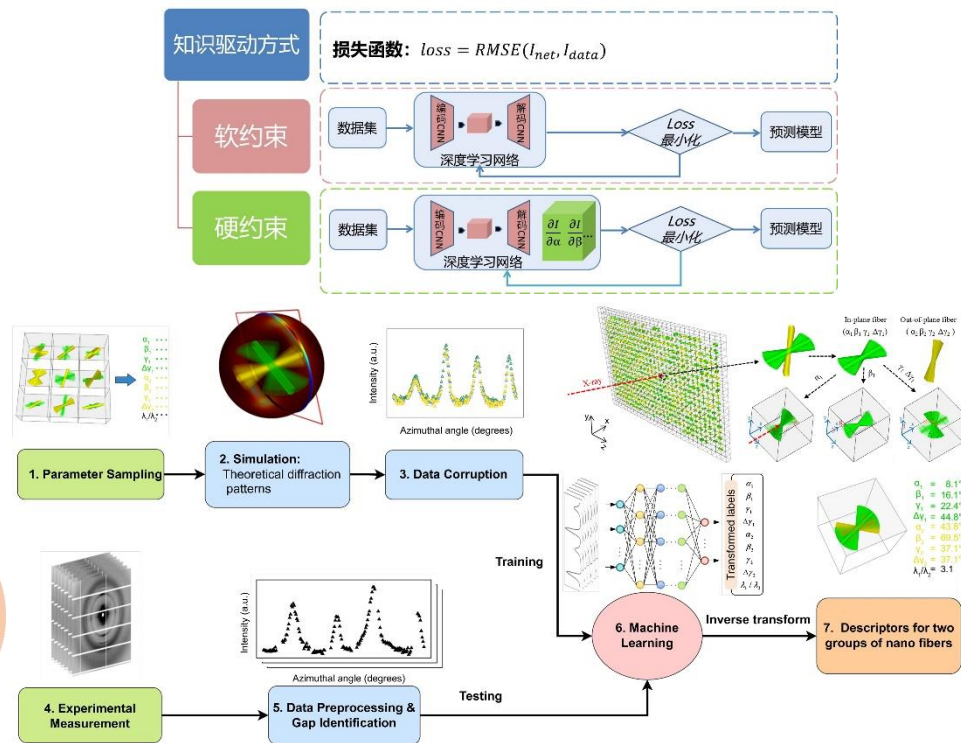
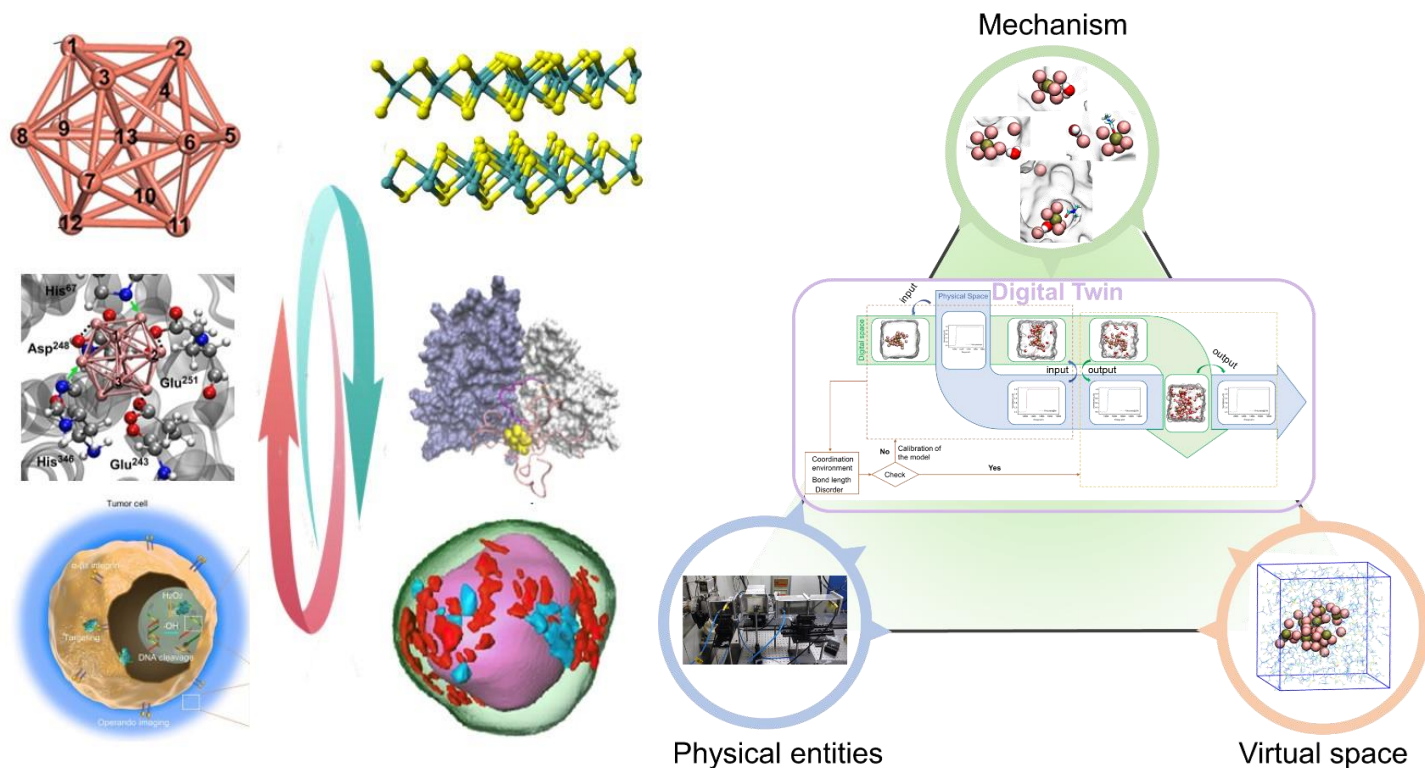
```
"id": "00000798",  
"question": "What can the properties of high-energy LFQPO provide insights into?",  
"answer": "The properties of high-energy LFQPO can provide insights into the nature of the accretion flow in the central regions. In particular, the LFQPO phase lag can provide insights into the geometry and the radiation processes of the accretion flow, and has important implications for the LFQPO origin.",  
"category": null,  
"source": "DOI: 10.1038/s41550-020-1192-2",  
"labeler": "赵锐",  
"label_time": "2023-05-05 22:25:17",  
"artificial_answer": null,  
"answer_quality": null,  
"checked": false,  
"locked": false,  
"misc": {  
  "questioner": null,  
  "question_time": null,  
  "origin_num": "10",  
  "title": "Discovery of oscillations above 200keV in a black hole X-ray binary with Insight-HXMT",  
  "domain": "多学科研究中心",  
  "keywords": [  
    "Insight-HXMT",  
    "X-ray binary"  
  ]  
}
```

# 4. 高能同步辐射的物理大脑—工作基础概述

□ 扎实积累多尺度物理模型基础 → 支撑In house研究 → 运用算法不断突破硬件效能上限

基于多尺度物理模型的装置数字孪生分析大脑

基于深度学习的装置实时在线解析终端



*Sci. Adv.* 6, 1421, 2020.  
*Nat. Biomed. Eng.* 4, 686, 2020.  
*Nat. Nanotechnol.* 16, 708, 2021.

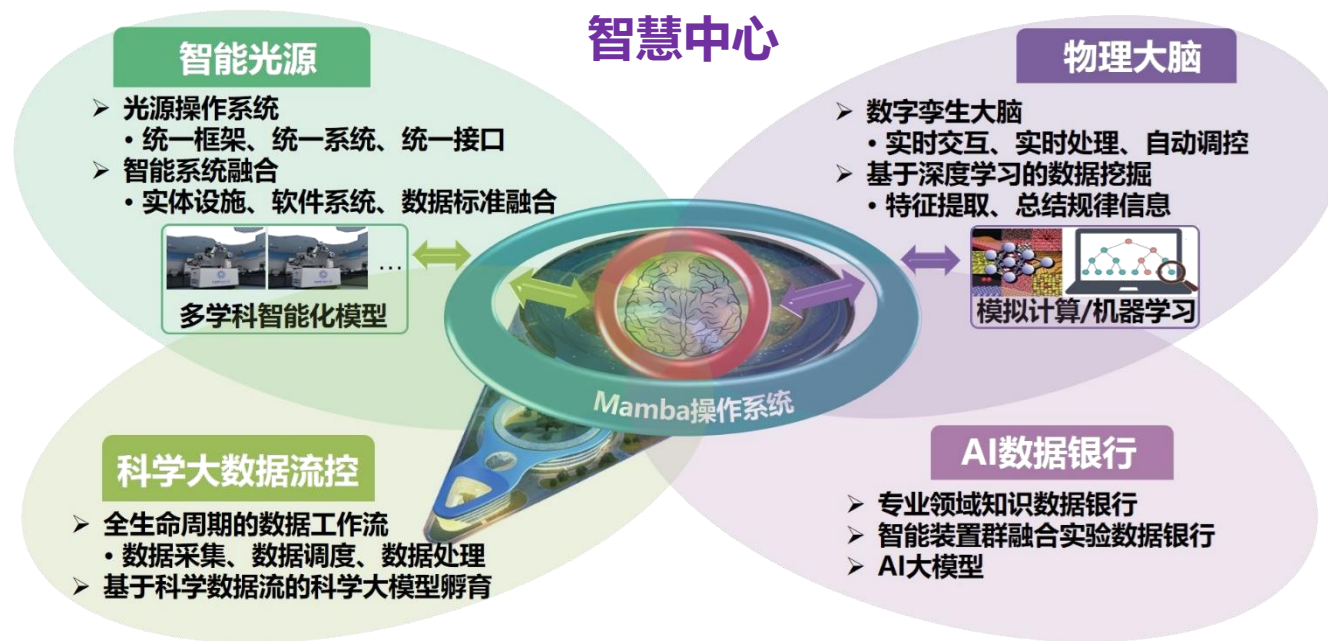
*J. Phys. Chem. Lett.* 14, 4876, 2023.  
*J. Am. Chem. Soc.* 145, 4279, 2023.  
*Nat. Commun.* 13, 418, 2022.

*IUCrJ* 10, 297, 2023.  
*npj Comput. Mater.* 9, 58, 2023.



# 5. 高能同步辐射的物理大脑——短期研究目标

## □ 智慧中心建设1~2年内实现的短期战略目标（蝴蝶效应）



- 纳米束线智能调光系统
- Nano-Arpes智能化实验系统
- 动态衍射实验智能操控系统
- 衍散射知识驱动AI在线解析系统
- AI辅助增材制造原位动态系统

- 提升高能同步辐射光源**软实力**，保障一期线站的高效运行。
- 一体化智能操控体系，助力重大基础科学产出，实现**开门红**。
- 在全球领域率先建立光源智慧中心，引领大科学装置科学研究的新范式。

# 致谢

感谢各位专家老师的批评指正~

