



面向核数据智能体的模型驱动核数据研究

——基于张量模型的核数据评价与应用

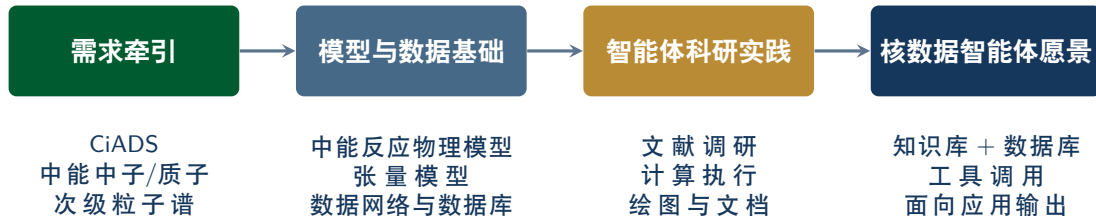
苏军

中山大学

AI 智能体赋能核物理研究研讨会 2026

从中能核反应物理模型、机器学习核数据方法到面向应用的数据智能体

报告思路：从张量模型到核数据智能体



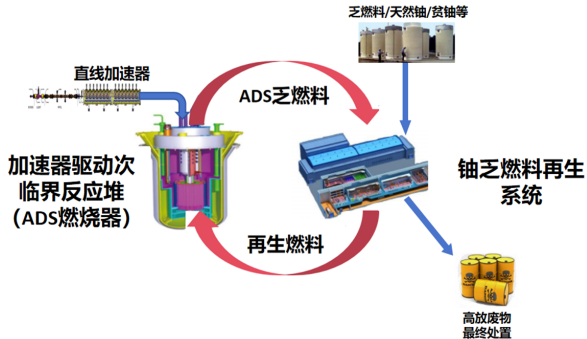
本报告以张量模型核数据评价为核心成果之一，进一步讨论如何把物理模型、数据库、机器学习方法和科研工作流整合为面向应用的核数据智能体。

01

需求牵引：CiADS 与 中能核反应数据

从 CiADS 应用需求出发，明确中能核反应数据的缺口、
难点与数据库建设目标。

CiADS 对核数据提出新的需求



宽能谱驱动

CiADS 中子能谱跨越热中子、快中子到中能区，核数据需求不再局限于传统反应堆能区。

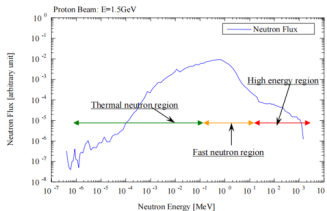


Fig. Typical spectrum in the ADS

需求牵引：中能中子核反应数据、质子诱发核反应数据，以及次级粒子能谱数据需要协同发展。

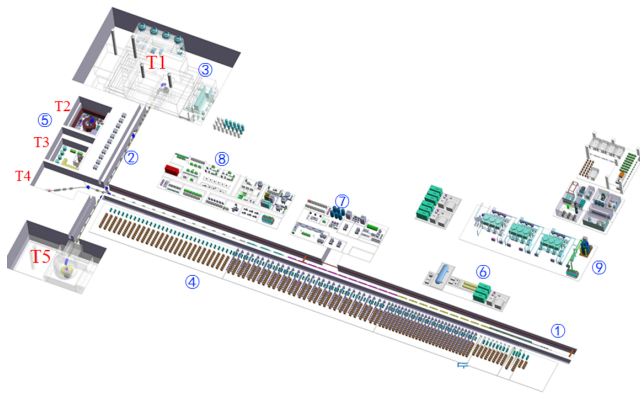
中高能核反应数据库发展现状

数据库	能区覆盖	特点与依据	地区
FENDL-3.2c	$E_n \leq 150 \text{ MeV}$; E_p 可至 GeV 区	面向聚变、加速器等应用; 中子数据扩展到 150 MeV, 质子数据子库常至 3 GeV	IAEA
TENDL	$E_n/E_p \leq 200 \text{ MeV}$	覆盖 n、p、d、t、 ^3He 、 α 等入射粒子, 形成较完整的计算评价库	Europe
JENDL/HE	$E_n/E_p \leq 3 \text{ GeV}$	高能文件包含中子、质子诱发反应评价数据, 覆盖百余核素量级	Japan
LA-150	$E_n/E_p \leq 150 \text{ MeV}$	面向加速器驱动系统应用, 发展 150 MeV 以下中子和质子数据库	USA

国际数据库已覆盖到 150 MeV、200 MeV 乃至 GeV 区, 但应用对象、反应道和不确定度仍存在差异。

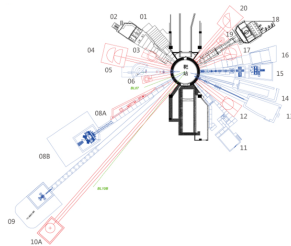
对 CiADS 而言, 需要面向具体应用场景筛选、补全、验证和追踪数据来源。

国内发展现状：起步较晚，发展加快



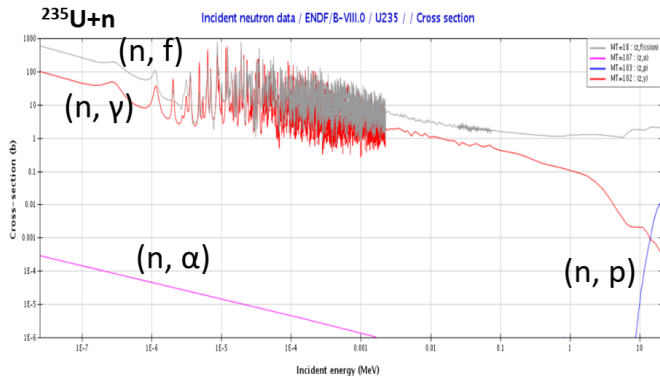
装置牵引

强流质子加速器、靶站和多用途实验终端，
为中高端核数据测量提供平台。



从装置建设到数据库发展，国内中高端核数据研究正在形成“实验测量 + 模型计算 + 数据评价”的连续能力。

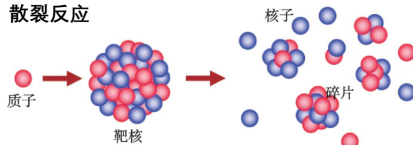
从热中子到高能中子反应道



反应道更复杂

随入射能升高，反应道从低能俘获、裂变逐步扩展到多粒子发射和散裂过程。

散裂反应



中高能核数据评价不仅要补全截面，还要解释反应机制、适用能区和模型外推边界。

为什么中能核数据问题适合智能体？

数据分散、多源、持续增长

实验、EXFOR、文献、评价库、模型输出与应用反馈需要长期汇聚和追踪。

评价周期长，需要缩短迭代

从实验测量到工程应用常以年为单位；智能体可加速检索、整理、计算、比较和报告。

需求场景差异大

工程应用需要成套标准数据；核物理探索常关注单条关键数据或反应机制。

工具链复杂，适合受控调用

数据库查询、物理模型、张量模型、数据网络、输运计算和绘图可由智能体组织。

必须可追踪、可解释

核数据推荐不仅要给数值，还要说明来源、适用能区、模型依据和不确定度。

适合人机协同，而非完全自动化

智能体负责连接流程和形成候选方案，专家负责物理判断、取舍和最终评价。

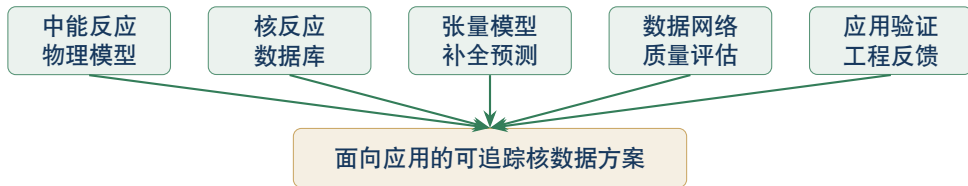
因此，中能核数据研究适合发展“知识库 + 数据库 + 模型工具链 + 专家判断”的核数据智能体。

02

模型与数据基础：物理模型、张量模型与数据网络

以物理机制、实验数据和机器学习模型为基础，形成可用于评价、预测与不确定度分析的核数据能力。

模型与数据基础：核数据智能体的可信基础



核数据智能体依靠什么基础推荐、解释和评估数据？

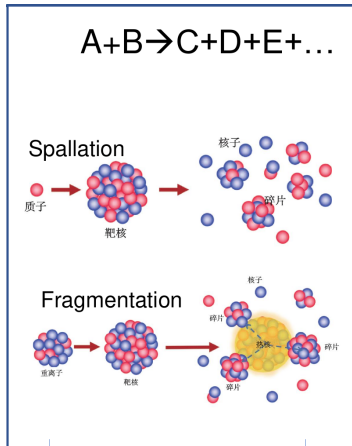
核心主线：

物理约束 + 数据基础 + 机器学习 + 应用验证。

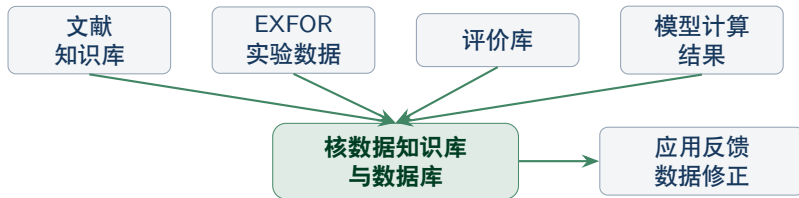
中能核反应机制：数据评价的物理约束

- 中能区反应包含散裂、碎裂、预平衡发射与复合核退激等过程
- CiADS 需要的不只是截面数值，还需要能区、反应道和次级粒子谱的物理解释
- 物理模型为机器学习结果提供合理性约束和外推边界

在智能体中，物理模型提供模型依据、适用范围和机理约束。



核反应数据库：从数据收集到智能体基础



智能体不是替代数据库，而是在数据库之上组织查询、证据追踪、模型调用和面向场景的数据方案。

数据问题的本质：稀疏、多维、不一致

- 核数据依赖核素、能量、角度、出射粒子、反应道等多个变量
- 关键应用对象常常缺少直接实验数据
- 不同实验和评价库之间可能存在系统差异
- 需要同时完成“补全”和“可信度判断”

这正是张量模型和数据网络进入核数据评价的原因。

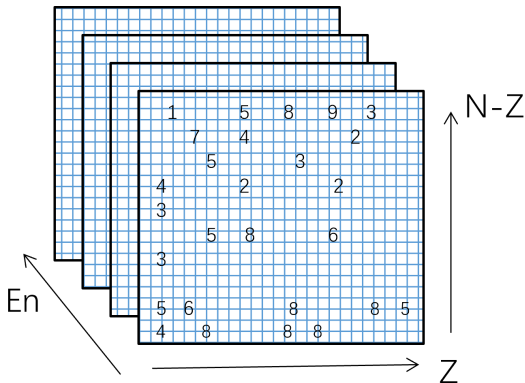
	Thermal	Fast	HE
Th-227	ENDF		
Th-229	ENDF		
Th-232		ENDF	ENDF
Pa-231		ENDF	
U-232	ENDF		
U-233	ENDF	ENDF	ENDF
U-234		ENDF	ENDF
U-235	ENDF	ENDF	ENDF
U-236		ENDF	ENDF
U-237		ENDF	
U-238		ENDF	ENDF
Np-237	ENDF	ENDF	ENDF
Np-238		ENDF	
Pu-238		ENDF	
Pu-239	ENDF	ENDF	ENDF
Pu-240	ENDF	ENDF	ENDF
Pu-241	ENDF	ENDF	
Pu-242	ENDF	ENDF	ENDF
Am-241	ENDF	ENDF	ENDF
Am-242m	ENDF		
Am-243		ENDF	
Cm-242		ENDF	
Cm-243	ENDF	ENDF	
Cm-244		ENDF	
Cm-245	ENDF		
Cm-246		ENDF	

Isotopes	$T_{1/2}$
^{93}Zr	$1.61 \times 10^6 \text{ y}$
^{79}Se	$2.95 \times 10^5 \text{ y}$
^{107}Pd	$6.5 \times 10^6 \text{ y}$
^{126}Sn	$2.3 \times 10^5 \text{ y}$
^{90}Sr	28.8 y
^{135}Cs	$2.3 \times 10^6 \text{ y}$
^{137}Cs	30 y
^{129}I	$1.57 \times 10^7 \text{ y}$
^{99}Tc	$2.11 \times 10^5 \text{ y}$

核数据天然可以组织为张量

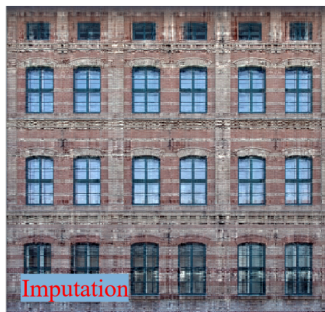
- 张量是多维数组
- 核数据天然具有多个物理维度
- 例：核素、入射能、散射角、出射能、产物核
- 缺失数据就是张量中的空缺元素

关键转变：从“单条曲线拟合”走向“多维关系学习”。



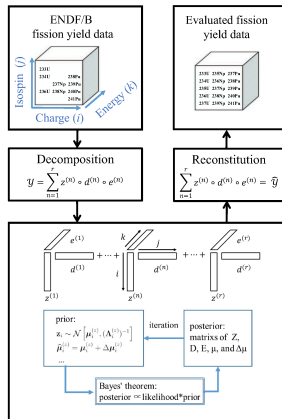
稀疏核数据是张量补全问题

- 已测数据提供张量中的观测元素；未测能区、未测核素和未测角度构成缺失元素
- 张量补全利用全局低秩结构重构缺失值，适合中能核数据“覆盖不全但相关性强”的特点



CP 分解：从多维数据中提取低秩结构

- 复杂张量可表示为若干简单分量之和
- 每个因子矩阵描述一个物理维度
- 少量分量捕捉全局趋势
- 低秩结构让稀疏数据之间能够“互相借力”

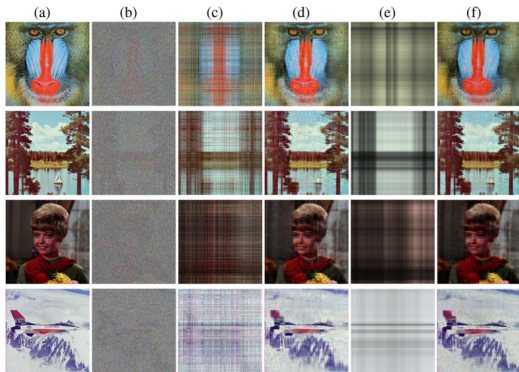


贝叶斯张量补全：稳定预测与不确定度

- 真实核数据同时存在缺失和噪声
- 贝叶斯推断把先验知识与观测数据结合
- 模型输出不只是点估计，也可服务不确定度分析
- 这为智能体的“可信推荐”提供方法基础

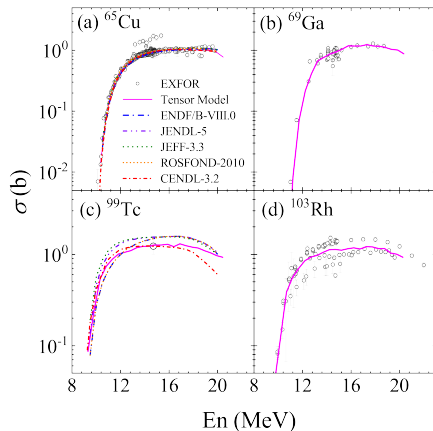
方法基础：Bayesian robust tensor completion via CP decomposition。

(a) Original image; (b) Noisy image; (c) BGCP; (d) MoG GWLRTF-CP; (e) BRTF; (f) MoG BTC-CP.



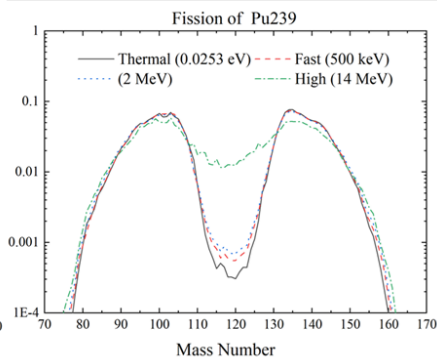
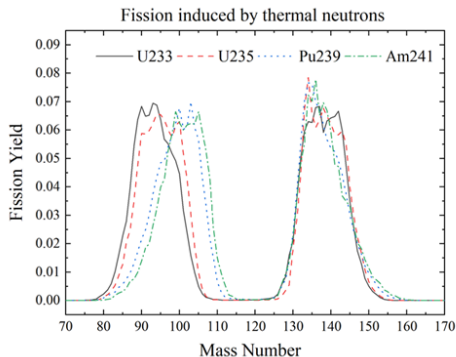
协同学习：让相关核数据互相借力

- 缺失点可从同位素链、同位旋、相邻能区和相似反应趋势中学习
- 数据少的核素不再只依赖自身的少量实验点
- 这种思想与推荐系统类似，但物理维度明确
- 对长寿命裂变产物和放射性靶核尤其重要

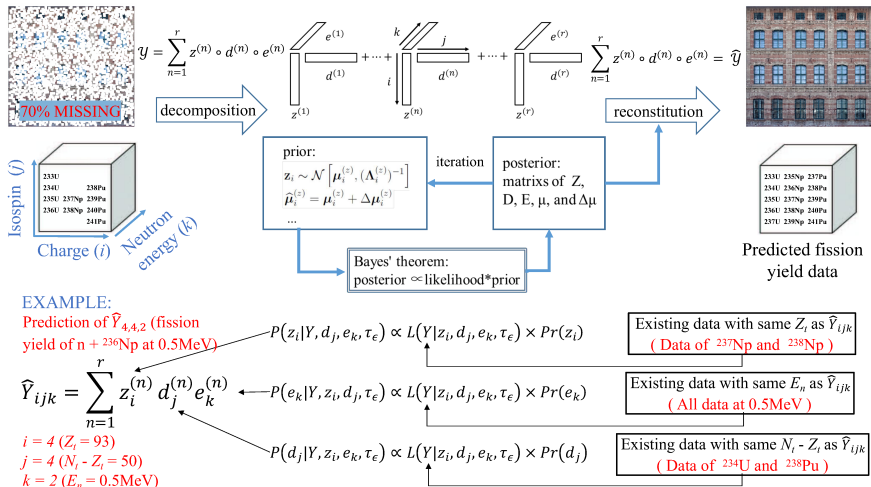


裂变产额数据：重要但不完整

- 裂变产额影响燃耗、衰变热、毒物积累和乏燃料组成
- 独立产额依赖靶核、入射能、产物核和同核异能态，是张量补全的典型核数据问题

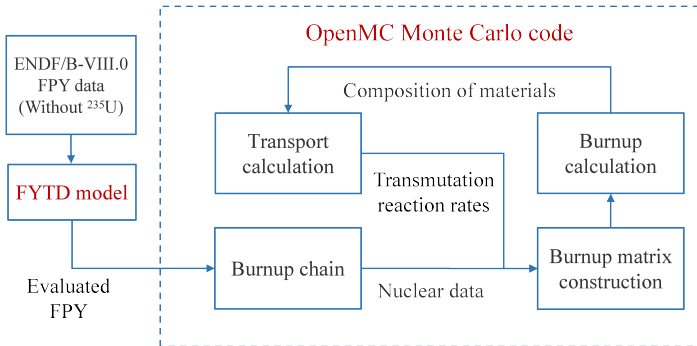


张量模型预测裂变产额



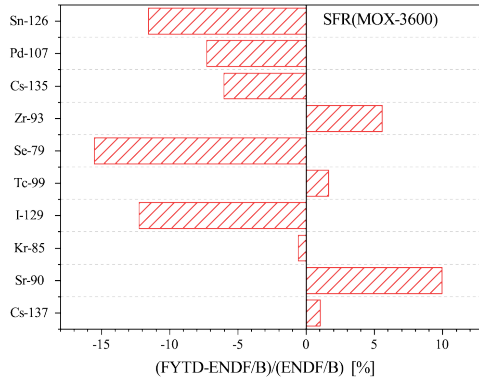
从数据比较到燃耗验证

- 核数据评价不能只看单点误差
- 产额误差会通过燃耗链影响反应性、衰变热和核素库存
- 使用 OpenMC 输运-燃耗计算验证评价数据应用效果
- 只改变裂变产额数据，隔离数据影响



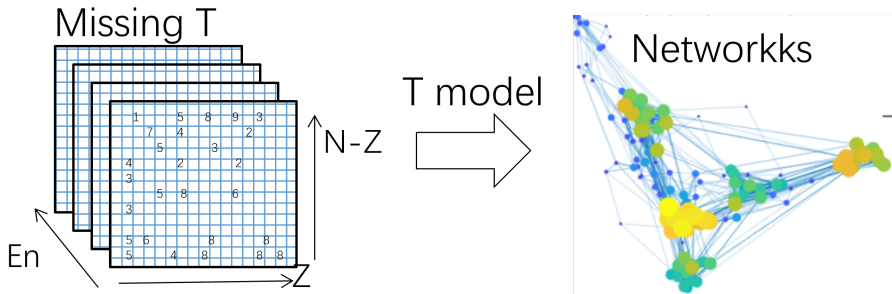
应用验证：反应性与停堆量保持接近

- 在热谱和快谱算例中测试裂变产额评价数据
- 运行阶段反应性偏差保持在较小范围
- 停堆后衰变热、Xe-135、Sm-149 等关键量接近参考结果
- 应用验证证明模型数据可进入工程计算链条



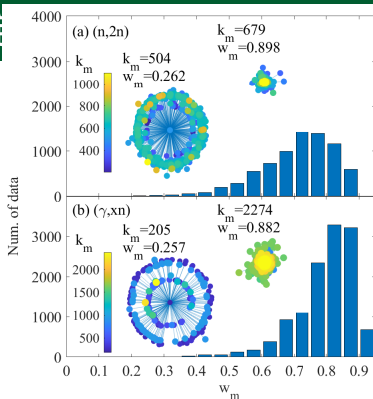
数据网络：从数据点到数据关系

- 每个实验数据点可看作网络节点；数据点之间的相似性或偏差定义连接权重
- 网络结构反映数据库覆盖、增长和一致性，为智能体提供“数据质量”视角



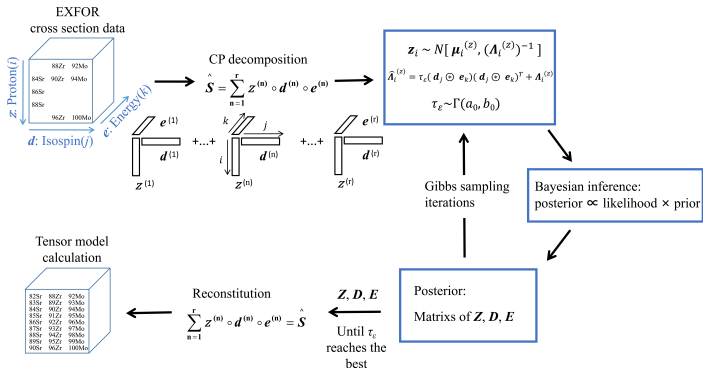
数据网络用于质量与不确定

- 网络性质可描述探索效率、整体不确定度和单点一致性
- 弱连接或异常权重提示需要进一步检查的数据
- 数据网络不是替代专家判断，而是提供全局诊断信号
- 与张量模型结合，可同时服务补全和质量控制



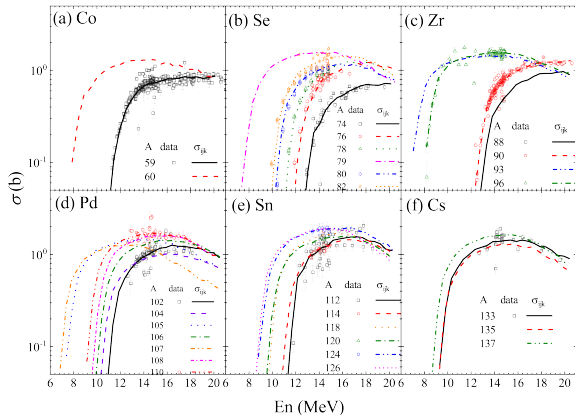
$(n, 2n)$ 反应：嬗变应用中的关键截面

- 长寿命裂变产物嬗变需要可靠的中能中子反应截面
- 放射性靶核实验困难，数据稀缺
- 张量模型使用质子数、同位旋和约化入射能组织数据
- 约化能量有助于处理反应阈值效应



张量模型预测缺失 $(n, 2n)$ 截面

- 模型可复现实验激发函数的主要趋势
- 对数据稀缺核素，可借助相关核素与相邻能区信息
- 预测长寿命裂变产物缺失截面
- 可为选择性嬗变能区和数据库补全提供候选数据
- 该算法被核数据中心收录使用



质子诱发反应：从角分布到源项数据

- 堆靶耦合需要质子诱发反应数据和次级中子谱
- 弹性散射角分布是典型多维核数据
- proton-natPb DDX 直接连接散裂靶源项计算
- 张量模型可补全稀疏的能量-角度双微分数据



IAEA

Atoms for Peace and Development

الوكالة الدولية للطاقة الذرية
国际原子能机构
International Atomic Energy Agency
Agence internationale de l'énergie atomique
Международное агентство по атомной энергии
Organismo Internacional de Energía Atómica

Vienna International Centre, PO Box 100, 1400 Vienna, Austria
Phone: (+43 1) 2600 • Fax: (+43 1) 26007
Email: Official.Mail@iaea.org • Internet: <https://www.iaea.org>

In reply please refer to: EVT2404177
Dial directly to extension: (+43 1) 2600-21708/21711

Mr Su Jun

Sun Yat-sen University
Sino-French Institute of Nuclear Engineering
and Technology (IFCN)
ZHUHAI 519082
CHINA

2026-05-25

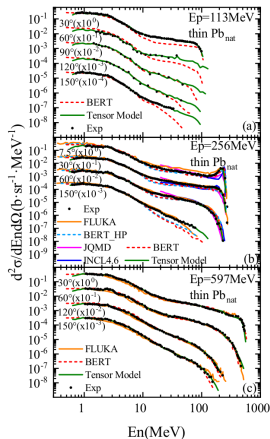
Dear Mr Su Jun,

The Secretariat of the International Atomic Energy Agency (IAEA) is pleased to inform you that your designation by the IAEA as a participant in the **Technical Meeting on Uncertainty Quantification in Nuclear Data** (hereinafter referred to as "event") at its Headquarters in Vienna, Austria, from

与物理模型比较，检查物理合理性

- 张量模型预测需要接受独立物理模型检验
- DDX 预测与 BERT、FLUKA、PHITS 等模型比较
- 比较可发现物理模型和实验数据的差异区域
- 物理模型 + 张量模型形成互补评价证据

面向应用的数据方案应同时给出数值、来源、模型依据和不确定度说明。



小结：从模型与数据基础到智能体能力

已有基础	已展示成果	支撑的智能体能力
中能物理模型	散裂、碎裂、输运-退激机制	解释数据、约束外推、判断适用范围
核数据与数据库	文献、EXFOR、评价库、模型结果 汇聚	查询、证据追踪、版本管理
张量模型	裂变产额、 $(n, 2n)$ 、DDX 补全预测	缺失数据推荐、候选评价数据生成
数据网络	数据关系、质量诊断、不确定度线 索	异常发现、数据可信度评估
应用验证	OpenMC 能耗与源项相关应用	面向场景判断数据是否可用

03

科研实践：智能体辅助 核数据研究流程

智能体工具推动核数据研究从人工串联流程走向人机协同闭环。工具环境：VScode+Codex。



智能体做什么？

检索、拆解、执行、检查、记录、形成候选方案

专家做什么？

判断物理机制、模型边界和最终结论

学生训练目标：不是只会跑脚本，而是看懂证据链

不是替代核物理判断，而是把耗时、易错、难交接的流程变成可追踪资产。

小时级 → 分钟级

文献表格结构化

PDF 表格到 CSV 初稿，保留单位、误差和来源

上万文件内定位

找代码 / 找数据

快速找到脚本、输入、输出和图件路径

成套更新

脚本-CSV-图件闭环

一次修改同时追踪数据、图像和论文用途

可交接

课题决策日志

把为什么继续做、边界在哪里写成团队资产

把重复性科研组织工作压缩一个量级；科学结论仍由专家把关。

过去：靠人记、靠手工、靠反复问

- ✗ PDF、数据、脚本、图件分散
- ✗ 换学生后细节容易断档
- ✗ 图能看，但难追溯到脚本
- ✗ 文档可能混入旧模型版本



现在：可追踪、可复现、可接续

- ✓ 文件、脚本、CSV、图件形成链条
- ✓ brief / decision log 支撑交接
- ✓ 一张图对应脚本、数据和论文用途
- ✓ 逐项检查文档与当前代码是否一致

把“个人经验”转化为“团队资产”

文献→数据

PDF表格提取工作流程.md

PRX table1.csv

课题决策

02_decision_log.md

brief / research questions

图件闭环

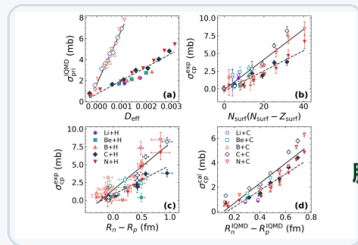
fig4p.py / fig4p.csv

fig4p.png / EPS

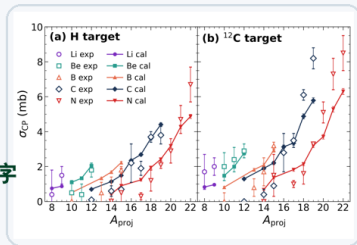
模型审查

pdf_to_script_revision_notes

脚本-文档一致性



脚本 / 数据 / 图件 / 文字



真实图件输出

真实论文图闭环

一个真实任务：从数据表到论文图，成套更新

脚本 / CSV / PNG / EPS

输入

PRX table1.csv

脚本

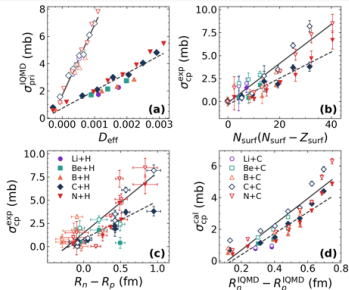
fig4p.py

数据

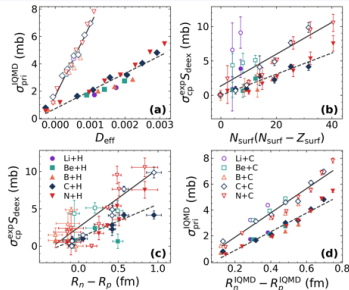
fig4p.csv

图件

fig4p.png / EPS



版本 A：模型比较



版本 B：投影/变量更新

不是只生成一张图，而是保留可复现的文件链

智能体擅长

- ✓ 检索资料
- ✓ 定位脚本
- ✓ 执行命令
- ✓ 检查路径/空图/缺失值
- ✓ 记录参数和下一步



协同

专家必须把关

- ! 物理机制是否合理
- ! 模型外推边界
- ! 不确定度来源
- ! 能否进入评价/应用
- ! 论文主张是否过强

学生培养目标

从“会跑脚本”走向“会提出问题、会检查证据链、会说明边界”

04

愿景：面向应用的 核数据智能体

整合知识库、数据库、模型和工具调用，面向 CiADS 等应用场景提供可追踪的数据方案。

愿景图：面向应用的核数据智能体



大模型用于知识库建设、资料更新和候选方案生成；轻量模型承担查询、路由、工具调用和模板化服务。

总结与致谢

本报告总结

- CiADS 等应用牵引中能核数据需求
- 物理模型、数据库、张量模型和数据网络构成可信基础
- VScode+Codex 等智能体工具可辅助文献整理、代码定位、图件复现和证据追踪
- 核数据智能体应服务候选方案生成，最终判断仍由专家负责

可推广场合

数据来源分散、工具链较长、结果需要复现和交接的科研任务：核数据评价、数据库更新、模型比较、论文图件复现、学生课题管理。

致谢

- 博士生：黄英格、冯玉洁、肖尔熙、顾富昌
- 硕士生：黄嘉励、王慧、刘惠子、陈洽锋
- 中国核数据中心：舒能川、续瑞瑞
- 北航：孙保华

谢谢各位老师！

敬请批评指正