Contribution ID: **51** Type: **not specified**

上海自由电子激光（SHINE）及其科学数据系统规划

Tuesday, 16 July 2019 11:25 (35 minutes)

硬 X 射线自由电子激光装置（SHINE）是国家重大科技基础设施建设“十三五”规划优先启动项目，其目标是建设一台电子能量为 8GeV 的超导直线加速器，3 条波荡器线、3 条光学束线以及首批 10 个实验站。装置提供能量范围为 0.4-25keV 的光子，具备超高峰值亮度和平均亮度、高重复频率（可达 1MHz）、飞秒级超快脉冲等优异特性，同时具备纳米级的超高空间分辨能力和飞秒级的超快时间分辨能力。首批实验站能够充分发挥高重复频率 XFEL 用户装置的多用户、超高空间分辨、超高能量分辨等突出优势，满足各类超快激光/超快电子/X 射线泵浦-X 射线探测高时间分辨实验的需求。

高重频 X 射线自由电子激光的技术挑战不仅来自于加速器、波荡器、精密 X 射线光学等关键技术，更来自于数据科学的挑战。高重频面探测器是 SHINE 高速数据流的主要来源，其工作频率可达 100kHz 至 1MHz，数据采集速率将比低重频探测器提高三个数量级。不同于同步辐射光源基于光子累积的成像原理，SHINE 装置将主要采用 SASE（自放大自发辐射）运行模式，需要记录每次与样品作用后的 X 射线衍射图样信息。SHINE 数据峰值流量的设计值为 100GB/s，并可能在远期达到 TB/s 的量级，亟需数据分析手段的重大革新以应对未来高通量大数据的挑战。现有基于“科学计算密集型”的超级计算机系统架构，不能满足“数据密集型”的数据分析的要求。

SHINE 的数据系统由数据采集与分析系统（DAQ）与高性能数据管理系统（DM）构成，目标是实现实验数据的快速缓存、高速传输、实时/离线处理以及高效管理和安全访问。数据采集与分析系统负责对探测器产生的数据进行采集、筛选、压缩和发送，同时对光束线与实验站产生的控制数据、诊断数据进行统一化管理与分析。高性能数据管理系统负责建设专用的数据中心，建立高速缓存系统、高速数据传输系统、高性能计算系统及大容量离线数据存储系统。它将 DAQ 系统采集的高通量数据通过在线高性能计算集群机进行处理和快速反馈，同时自动将数据通过高速网络传输到大容量离线存储系统中，在离线存储中由离线高性能计算集群机对数据进行分析处理，最终得到实验结果。

SHINE 数据系统中将部署数据处理软件和数据管理软件，对数据进行预处理、分析和管理，可以实现统一数据存储服务与集中式系统管理、存储服务多用户快速访问、数据实时处理、数据远程访问；并将开发部署智能分析软件，将人工智能技术用于与数据的实时分析与反馈，实现诊断数据、科学数据的筛选、压缩与数据挖掘。

目前 SHINE 已经完成了数据系统的总体规划及数据中心的建安设计，并开展了高速数据获取、高速数据传输、大数据存储和管理、云计算与高性能计算及先进算法等研发工作。本报告将汇报 SHINE 数据系统的设计规划及最新进展。

Primary author: Prof. 怀, 平 (ShanghaiTech University ; Shanghai Advanced Research Institute)

Co-authors: Prof. FAN, Haiwei (ShanghaiTech University); Dr SHI, Wujun (ShanghaiTech University); Dr ZHANG, Xiaofeng (ShanghaiTech University)

Presenter: Prof. 怀, 平 (ShanghaiTech University ; Shanghai Advanced Research Institute)

Session Classification: 大会报告 II

Track Classification: 科学数据存储与管理技术