

·科技建议·

加快材料基因组工程信息化 基础设施的建设

2016年2月,中国科学技术部发布了国家重点研发计划“材料基因工程关键技术与支撑平台”重点专项^[1],其主要目标是:融合高通量计算(理论)/高通量实验(制备和表征)/专用数据库3大技术,变革材料研发理念和模式,实现新材料研发由“经验指导实验”的传统模式向“理论预测、实验验证”的新模式转变,实现新材料“研发周期缩短一半、研发成本降低一半”的目标,为实现“中国制造2025”的目标做出贡献。

2016年初,美国国家自然科学基金委员会(NSF)下面的一系列研发计划(如CIF21,SI²,DMREF)强调应重视支撑材料科学的科学软件基础设施/科学软件基础生态系统的研发。

在经历了始于2008年的全球金融危机之后,美国政府意识到实体经济的重要性,提出重振制造业。2011年6月24日,美国总统奥巴马宣布了一项超过5亿美元的“先进制造业伙伴关系”计划,“材料基因组计划”(Material Genome Initiative, MGI)是该计划中的重要部分。MGI的核心理念,就是强调通过计算,数据,实验“三位一体”的结合,把发现、开发、生产和应用新材料的速度提高2倍。

2016年初,NSF“21世纪科学和工程信息化基础设施框架”(Cyberinfrastructure Framework for 21st Century Science and Engineering, CIF21)^[2]研究计划下的“可持续创新的软件基础设施”^[3](Software Infrastructure for Sustained Innovation, SI²)研发计划“数学和物理科学”方向的材料专题中,再次强调了SI²与“设计材料以革新和设计我们的未来”^[4](Designing Materials to Revolutionize and Engineer our Future, DMREF)的结合,在接下来的3~4年中,预计将有总计750000~1600000美元资助DMREF。

美国NSF下面的CIF21、SI²、DMREF 3个不同的研发计划,本质上都有其相同点,都从一个方面强调了对材料基因组计划的资助,即注重信息化基础设施相关研发。由此引发对中国材料基因工程专项实施的思考。

1 21世纪科学和工程信息化基础设施框架(CIF21)

认识到信息化基础设施支撑当代科学研究的重要性,美国NSF自20世纪80年代就开始重视支撑科学研究的信息化基础设施(Cyberinfrastructure)的建设,如NCSF资助的超级计算中心计划,PAACI计划,TeraGrid计划等。在这些计划下,美国开发了许多支撑科学研究的软件基础设施,如Open Science Grid, the National Virtual Observatory, Data Activities,以及相关的研究合作Network for Earthquake Engineering Simulation (NEES), The National Ecological Observatory Network (NE-

ON), Ocean Observatories Initiative (OOI), Large Hadron Collider (LHC)等。在此阶段,英国设立了e-Science计划,欧盟设立了ESRFI, EGI等。

2010年,NSF又提出了更为长远的战略规划,即CIF21。CIF21旨在开发并部署全面,综合,可持续,安全的信息化基础设施(CI)以加快计算和数据密集型科学和工程的研究等,从而转化为有效应对和解决面对科学和社会许多复杂问题的能力。

2 可持续创新的软件基础设施研究计划(SI²)

SI²是CIF21下的一个研究计划,主要强调研发支撑各学科领域研究的可持续发展的软件系统,或软件基础设施,也就是人们所说的领域信息学(X-Informatics)或科研信息化(e-Science)。SI²主要支持3类软件的开发:1)科学软件元素,2)科学软件集成,3)科学软件创新研究院。SI²强调与学科的交叉。在其“数学和物理科学”方向中,又进一步分为材料、物理、化学、天文等。材料方向重点强调了与DMREF的结合,支持材料基因组计划下的软件使能和支撑技术研究。

3 设计材料以革新和设计我们的未来(DMREF)

DMREF最初是NSF为响应材料基因组计划而设立的一个专项,着重强调支持那些能加快材料发现和研发的各种活动,比如通过构建需要的基础知识库来设计和制造由第一原理预测出的材料功能或性能。实现这一目标涉及建模,分析和计算机模拟,通过样品制备,表征和设备进行验证;涉及新的数据分析工具和统计算法,充分利用机器学习,数据挖掘和稀疏逼近等发展预测模型;涉及与新设备功能相结合的材料性能模拟等。

这些都需要研发方便的、可扩展的、可伸缩的和可持续的数据基础设施;开发,部署,和维护用于下一代材料设计的可靠,可互操作的,可重复使用的软件基础设施;以及用于管理大规模,复杂,异构的分布式材料数据从而帮助材料的设计、合成,及纵向研究协同创新能力的发展。

在当今数据密集型科学研究的背景下,计算机和信息科学技术对当代科学研究正起着不可或缺的作用。上述美国NSF的CIF21, SI²及DMREF研发计划,甚至包括材料基因组计划本身,其共同点都是强调通过信息化基础设施,加快包括新材料在内的科学发现。科技部发布的“材料基因工程关键技术与支撑平台”,也体现了信息化基础设施建设重要性。实际上,一些用于帮助新材料发现的理论方法和手段,如结构筛选,元素替代,性能与成分优化等,均涉及到大规模,高并发的材料计算任务协同,以及材料计算数据的自动归档,典藏(data curation)和计算

数据分析等,尤其需要信息化工具的支持。Ceder小组的工作之所以能引起业界关注,就在于他们通过高通量第一原理计算系列信息化工具的帮助,从3万余种化合物中理论上筛选出了高效安全的锂电池阴极材料^[6]。

然而,中国目前还缺乏有自主知识产权的各类材料软件和材料数据库。核心软件均依赖进口,一些关键材料数据库对中国封锁。以高通量计算为例,Ceder小组的工具,以及AFLOW等^[6],在中国使用不是特别方便(为此,中国科学院计算机网络信息中心也初步研发了支撑高通量材料集成计算基础平台和软件框架MatCloud)。

目前,中国的材料研究多以跟踪为主,工程化不够,系统性不强,特别是融合计算、测试和表征、开发、优化的全链条材料开发模式尚未形成。因此该专项的实施,应重视和加快包括材料软件,数据库,模型,工具,和平台等在内的信息化基础设施的研发及建设,提供从材料理论设计、制备与表征、组织与工艺优化、到性能评价等全链条的材料快速、低成本研发的信息化支撑。

参考文献(References)

[1] 科技部关于发布国家重点研发计划高性能计算等重点专项2016年

度项目申报指南的通知[EB/OL]. 2016-02-19. http://www.most.gov.cn/fggw/zfwj/zfwj2016/201602/t20160218_124156.htm.

[2] Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)[EB/OL].[2016-01-25]. http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504730.

[3] Software Infrastructure for Sustained Innovation (SI²)[EB/OL]. [2016-01-25]. http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf11539.

[4] Designing Materials to Revolutionize and Engineer our Future (DMREF)[EB/OL]. [2016-01-25]. https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505073.

[5] The Material project[EB/OL]. [2016-01-25]. <https://www.materialsproject.org>.

[6] Aflow[EB/OL]. [2016-01-25].<http://www.aflowlib.org>.

文/杨小渝

作者单位 中国科学院计算机网络信息中心。

(编辑 田恬)