

专题: AI 物质科学

## 多智能体在物理材料计算领域的应用\*

吴颖 卢仲毅 高泽峰†

(中国人民大学物理学院, 北京 100872)

(2025 年 12 月 8 日收到; 2026 年 1 月 13 日收到修改稿)

随着材料研发日益复杂, 传统“试错法”与零散计算模拟面临效率低下、资源消耗大的挑战. 人工智能, 特别是大语言模型的发展, 为材料计算领域带来新机遇, 其中多智能体系统通过模拟科研团队协作, 展现出处理复杂任务的潜力. 本文系统综述了多智能体在物理材料计算中的应用, 重点分析了 VASPilot, PhysAgent 等专业化系统如何通过角色分工与工具集成, 实现从结构建模、参数设置到结果分析的全流程自动化. 研究表明, 多智能体系统能够有效提升计算效率与可重复性, 并在物理定律自主发现、复杂材料模拟等方面取得初步成果. 然而, 通过对现有系统的客观分析和实际操作使用, 本文发现其在计算资源分配、物理一致性、科学创造性等方面仍存在局限. 未来通过深化物理机理融合、构建闭环科研生态, 多智能体有望从辅助工具演化为能够自主探索与发现的科研伙伴, 推动材料研究范式的根本变革.

**关键词:** 材料计算, 多智能体, 科学发现自动化**DOI:** 10.7498/aps.75.20251687**CSTR:** 32037.14.aps.75.20251687

## 1 引言

材料是现代社会的基石, 其研发效率直接关系到能源<sup>[1]</sup>、信息<sup>[2]</sup>、生命科学<sup>[3]</sup>、航空航天<sup>[4]</sup>等关键领域的进步. 新型材料可以应对新兴技术挑战、提高效率、降低成本以及降低各行业对环境的影响<sup>[5-7]</sup>.

然而, 传统的材料研发模式严重依赖于“试错法”、研究人员的直觉经验以及零散的计算模拟, 这是一个典型的“时间长、成本高、风险大”的过程. 随着计算机科学与量子力学理论的深度融合, 以密度泛函理论和分子动力学为代表的计算模拟方法, 已成为在原子尺度上揭示材料构效关系、预测其物理化学性质的强大工具<sup>[8-11]</sup>. 尽管如此, 这些计算方法本身仍面临着严峻挑战: 其工作流程高度复杂<sup>[12]</sup>, 涉及结构建模、参数设置、计算执行、错误诊断、数据提取与分析等多个专业环节; 计算资源需求巨

大<sup>[13-15]</sup>; 且领域知识壁垒高筑<sup>[16,17]</sup>, 也在很大程度上阻碍了高通量计算与大规模材料筛选的实现.

近年来, 人工智能, 特别是基于 Transformer 架构的大语言模型 (LLMs) 的革命性进展, 为科学计算领域注入了新的活力<sup>[18,19]</sup>. 我们所耳熟能详的 GPT-4<sup>[20]</sup> 和 DeepSeek<sup>[21]</sup> 都是 LLMs 的杰出代表. 这些模型通过在大量通用文本语料库上进行预训练, 并在对话数据上进一步微调, 展现出卓越的代码生成、自然语言理解与逻辑推理能力, 使其不再仅仅是对话工具, 而是进化为能够规划并执行复杂任务的“智能代理”<sup>[22,23]</sup>.

基于其强大的对话能力, LLM 开发者已将结构化输入输出格式融入微调过程中, 使这些模型具备了调用外部程序的能力, 即所谓的函数调用或工具使用<sup>[24]</sup>. 这一突破性进展使得单个 AI 代理已经能够通过强化学习 (RL) 从基于结果的奖励中进行训练, 以使用外部工具<sup>[25-28]</sup>. 这种单一、整体策略的方法在专门的单工具场景中 (如代码执行解决数

\* 北京市自然科学基金 (批准号: Z250005) 和国家自然科学基金 (批准号: 62476278) 资助的课题.

† 通信作者. E-mail: zfgao@ruc.edu.cn

学问题, 或网络搜索处理知识密集型问题) 已被证明是有效的<sup>[29-32]</sup>. 然而, 面对材料计算中常见的多步骤、多目标、跨尺度的复杂研究任务<sup>[33-35]</sup>, 单代理系统往往因上下文长度限制、任务过载和专业分工不明而显得力不从心. 特别是在执行复杂的多步骤任务时, 来自多个工具调用的长上下文累积使得单智能体系统难以保持最佳性能. 这种方法的根本挑战在于长视野信用分配困难——即使有细粒度的回合级奖励, 将最终结果归因于特定的中间工具调用仍然极具挑战<sup>[36,37]</sup>, 导致训练不稳定和脆弱的推理时泛化, 表现为工具过度使用或“认知卸载”<sup>[38,39]</sup>等策略缺陷.

在此背景下, 多智能体系统应运而生, 其通过模拟人类科研团队的协作模式, 将复杂的宏观任务分解并分配给多个具备特定角色、专业知识和工具集的专用代理, 通过协同合作与信息交互, 最终实现整体目标. 代表性框架包括 MetaGPT<sup>[40]</sup>, 其通过角色代理协作建模软件工程过程; AutoGen<sup>[41]</sup>, 支持可定制的多代理对话以增强推理能力; 以及 CrewAI<sup>[42]</sup>, 一个广泛采用的构建基于 LLM 的代理工具包. AgentFlow<sup>[43]</sup>, 作为可训练的流式智能体框架, 通过规划器、执行器、验证器、生成器 4 个模块协调, 在多轮交互中优化规划与工具使用. 这些框架将复杂的任务分解为一系列更简单的子任务, 然后分配给配备不同工具、领域知识和独立上下文状态的专用智能体, 通过实现协同执行来提高鲁棒

性和可扩展性.

多智能体系统的特性使其适合应对材料设计的复杂性<sup>[44,45]</sup>. 这些系统能够整合来自学术文献、在线数据库和尖端物理模拟等多种外部来源的多样化数据模态<sup>[46]</sup>, 并提取新知识. 如图 1 所示的多模型多智能体系统架构<sup>[47]</sup>, 其核心包含 3 个主要组件: 大脑、感知和行动<sup>[48]</sup>. 大脑作为 AI 代理模型的核心, 主要由前沿 LLM 构成, 负责决策、推理和规划等关键任务; 感知模块收集和处理多模型数据; 而行动模块则基于大脑的指导执行决策. 通过这种分工协作, 多智能体系统能够克服传统以人为中心的方法可能难以解决的复杂问题.

在材料科学这一具体领域, 多智能体系统的应用已展现出巨大的潜力和初步的成功. LAMMPS<sup>[49]</sup> 是一款开源的分子动力学模拟器, 支持从原子到连续尺度的材料建模, 具备高度灵活性和并行计算能力, 为材料模拟提供了基础支持. TopoMAS<sup>[50]</sup> 基于大语言模型, 通过人机协作实现从查询到验证的拓扑材料发现流程. MatPilot<sup>[51]</sup> 作为 LLM 驱动的 AI 材料科学家, 借助多智能体协作增强团队能力, 实现科学假设生成与自动化实验. ChemAgents<sup>[52]</sup> 基于 Llama-3.1-70B, 可自主执行复杂化学实验, 满足个性化研究需求. El Agente<sup>[53]</sup> 是多智能体量子化学系统, 通过自然语言界面动态生成 workflow, 任务成功率超 87%. DREAMS<sup>[54]</sup> 基于密度泛函理论, 通过分层智能体协作实现 DFT 模拟和材料

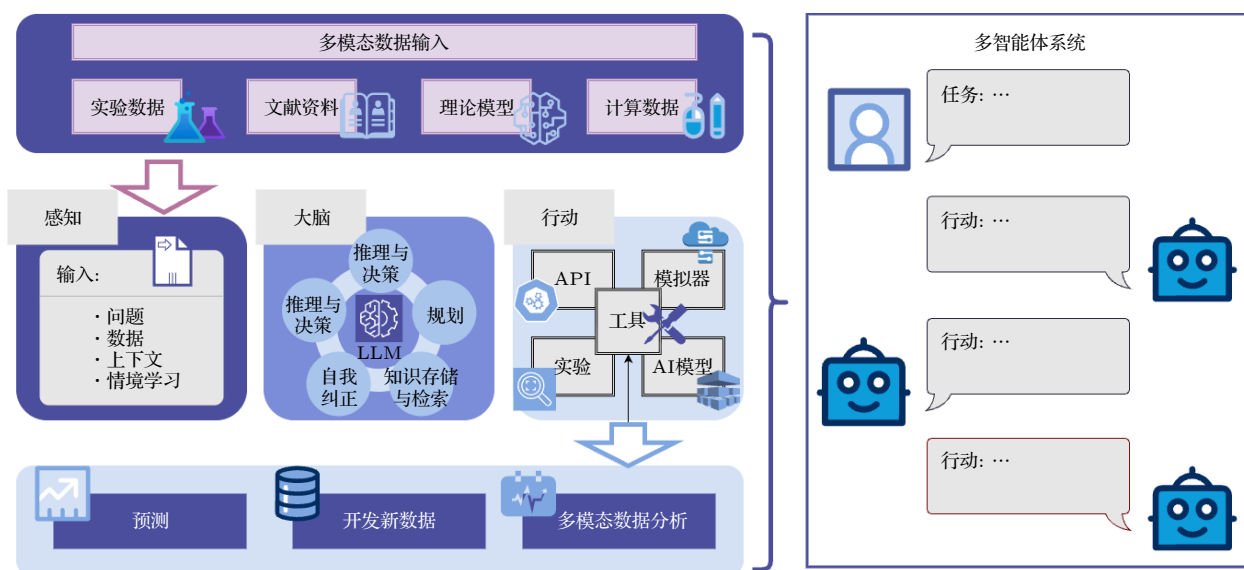


图 1 多模型多智能体方法作为材料发现、建模和预测的灵活建模策略

Fig. 1. Multi-modal multi-agent approach as a flexible modeling strategy for materials discovery, modeling, and prediction.

筛选. AtomAgents<sup>[55]</sup> 作为物理感知的生成式 AI 平台, 整合知识检索、多模态数据集成和物理模拟, 用于合金设计. SciLink<sup>[13]</sup> 通过将实验观察、新颖性评估和理论模拟直接、自动化地联系起来, 旨在提高材料表征的效率并促进意外科学发现. 这些系统推动了材料研究的自动化与智能化, 为未来复杂研究场景提供了新思路 and 工具, 有望进一步提升研究效率与质量<sup>[56]</sup>.

面向第一性原理计算领域的专用多智能体平台也取得了一定成果. 中国科学院物理研究所吴泉生和翁红明团队 (2025)<sup>[12]</sup> 开发的 VASPilot 系统, 聚焦于自动化广受欢迎的 VASP 软件工作流. 该平台基于 CrewAI 框架和标准化的模型上下文协议<sup>[57]</sup>, 部署了包括晶体结构代理、VASP 代理和结果验证代理在内的多个专业化角色. VASPilot 实现了自动获取结构、生成输入文件、到最终进行数据可视化的一站式自动化处理. 该框架同时实践了平面波截断能收敛性测试、不同范德瓦耳斯修正下的晶格常数优化等复杂基准测试, 极大地降低了研究者的技术负担, 提升了计算的可靠性与可重复性.

更进一步, 中国人民大学的高泽峰团队 (2025)<sup>[58]</sup> 提出了 PhysAgent 系统, 其愿景不再局限于执行预设的计算任务, 而是致力于自动化整个科学发现过程. PhysAgent 引入了“导师-学生-领导者”这一富有启发性的代理协作架构. 令人瞩目的是, 该系统被证明能够仅从模拟的观测数据中, 在无任何先验知识的前提下, 自主地重新发现开普勒行星运动定律和牛顿第二定律. 此外, 其还能完成 GaAs 半导体材料的全自动电子结构计算, 甚至模拟高速列车车窗上雨滴流动这类复杂的物理现象. 这预示了多智能体系统已经从辅助计算的工具, 演进为能够参与基础科学发现的自主科研伙伴.

综上所述, 以 VASPilot 和 PhysAgent 为代表的多智能体系统, 正以前所未有的方式重塑着物理与材料计算的研究范式. 本综述将系统性地梳理多智能体系统在该领域的最新进展, 深入剖析其核心架构, 详细探讨其广泛应用; 并且对比多智能体在材料计算领域的框架——VASPilot, PhysAgent 和其他代表性多智能体——MetaGPT 三者之间的区别; 最后, 对该技术当前面临的挑战进行批判性审视, 并对其未来发展方向进行展望.

## 2 前沿工作

近年来, 多智能体系统在物理与材料计算领域实现了从概念验证到实际应用的跨越. 本节将深入剖析前沿研究的两个具有代表性的框架: VASPilot 和 PhysAgent. 它们分别面向第一性原理计算和通用物理科学发现, 展现了多智能体架构的多样性与在材料计算领域应用的强大潜力. 然后, 介绍代表较为成熟的多智能体代表——MetaGPT, 并且对这 3 个框架进行对比.

### 2.1 VASPilot: 基于多智能体系统的全自动 VASP 模拟平台

密度泛函理论 (DFT) 作为计算材料科学的基石, 已在材料设计、物性预测和新材料发现中发挥着不可替代的作用<sup>[59-61]</sup>. 然而, 传统 DFT 计算流程复杂且高度依赖人工干预——从晶体结构准备、参数设置、任务提交到结果分析, 研究人员需要投入大量时间处理技术细节而非科学问题本身. 尽管 VASPKIT<sup>[62]</sup>, ASE<sup>[63]</sup> 和 Pymatgen<sup>[64]</sup> 等工具在一定程度上提升了计算效率, 但在面对复杂多步骤任务时, 仍缺乏真正的端到端自动化解决方案.

在此背景下, 中国科学院物理研究所吴泉生和翁红明团队开发的 VASPilot 平台<sup>[12]</sup>, 通过融合 CrewAI 多智能体框架与模型上下文协议 (MCP), 首次实现了 VASP 计算的全流程自动化, 为材料计算领域带来了范式变革的可能.

#### 2.1.1 VASPilot 架构设计

如图 2 所示, VASPilot 采用 3 层模块化架构, 巧妙地将用户交互、智能决策与工具执行分离, 确保了系统的灵活性与可扩展性.

前端交互层基于 Flask<sup>[65]</sup> 框架构建, 提供直观的 Web 界面. 用户可通过自然语言提交计算任务, 系统实时展示任务执行流程图、代理决策日志和工具调用详情, 实现了计算过程的监控.

智能决策层是系统的核心, 基于 CrewAI 框架构建了分工明确的智能体团队, 其中包括了 4 个智能代理: 管理代理 (manager agent) 作为“大脑”, 负责解析用户指令、规划任务流程并协调各专业代理工作; 晶体结构代理 (crystal structure agent) 专司材料结构检索与处理, 可自动从 Materials Project

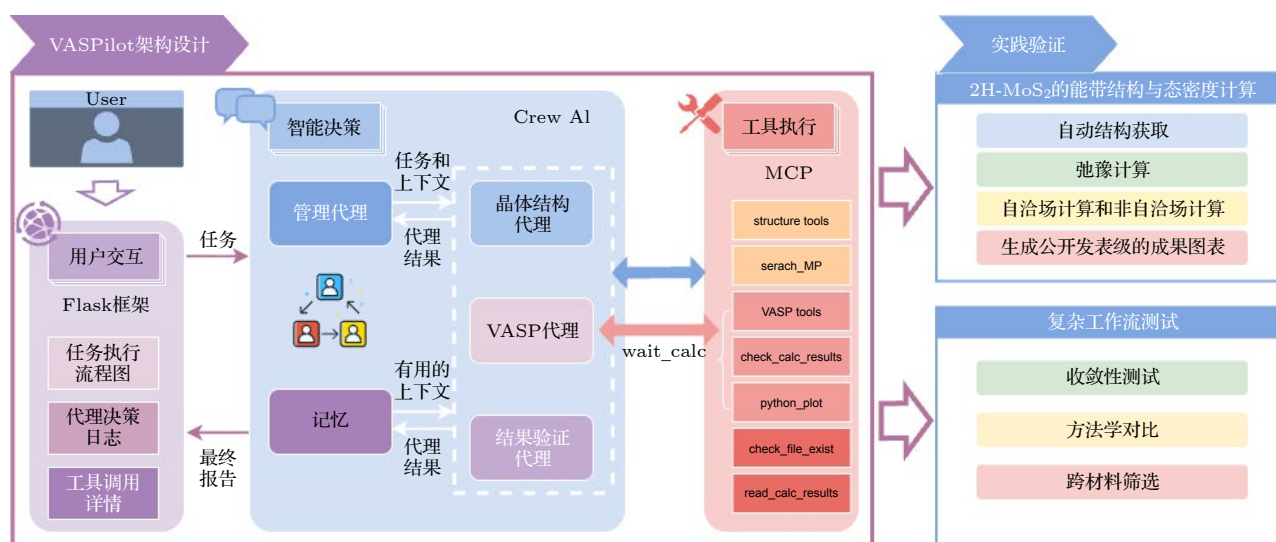


图 2 VASPilot 架构设计和实践验证示意图

Fig. 2. Schematic diagram of VASPilot architecture design and experimental verification.

等数据库获取结构文件<sup>[66]</sup>; VASP代理 (VASP agent) 作为“执行者”, 掌控从结构优化、自洽场计算到能带结构分析的全流程; 结果验证代理 (result validate agent) 扮演“质检员”角色, 对计算结果进行自动校验与可信度评估。

工具服务层基于模型上下文协议 (MCP), 将 Pymatgen, VASP 命令行工具等专业软件封装为标准化的 API 服务, 为智能体提供可靠的工具调用环境. VASPilot 的核心创新在于其基于 MCP 的工具服务器设计. 该设计不仅实现了计算工具的标准封装, 更通过一系列精巧的工具函数, 赋予了系统强大的自适应能力。

此外, 系统通过“wait\_calc\_tool”实现计算状态的定期轮询, 有效避免了长时任务中的网络超时问题; 而基于检索增强生成 (RAG) 的记忆池则让智能体能够借鉴历史经验, 做出更加明智的决策<sup>[12]</sup>。

### 2.1.2 实践验证

如图 2 所示, 研究团队通过多维度测试验证了 VASPilot 的实用性与鲁棒性。

在实例实践方面, 系统成功完成了 2H-MoS<sub>2</sub> 的能带结构与态密度计算全流程, 包括自动结构获取、弛豫计算、自洽场计算和非自洽场计算, 并生成了公开发表级的成果图表. 其中, 在 2H-MoS<sub>2</sub> 能带计算案例中, 当 VASP 返回“Bravais 晶格不一致”错误时, 系统能够自动解析错误信息, 调整对称性容差参数, 并重新提交任务. 这种“理解-调整-

重启”的闭环处理模式, 显著降低了人工干预的需求。

在复杂工作流测试中, VASPilot 同样表现出色. 首先在收敛性测试中, 其自动完成平面波截断能 (300—500 eV) 扫描与能量收敛分析; 在方法学对比中, 系统比较了多种范德瓦耳斯修正方法 (DFT-D2<sup>[67]</sup>, DFT-D3<sup>[68,69]</sup>, rVV10<sup>[70]</sup> 等) 对 MoS<sub>2</sub> 晶格常数优化的影响; 最后在跨材料筛选时, 同步计算 4 种过渡金属硫族化合物 (MoS<sub>2</sub>, MoSe<sub>2</sub>, WS<sub>2</sub>, WSe<sub>2</sub>) 的带隙并生成对比图表。

值得注意的是, 所有这些任务均在零样本设定下完成, 无需针对特定任务提供示例, 展现了系统强大的泛化能力。

## 2.2 PhysAgent: 面向通用物理科学发现的自主科研智能体

科学发现, 尤其是物理学领域的规律探索, 长期以来依赖于人类的直觉、严密的数学推理以及反复的实验验证. 然而, 随着研究体系日趋复杂、数据规模爆炸式增长以及计算工作流高度碎片化, 传统研究模式正面临前所未有的效率瓶颈<sup>[71,72]</sup>. 尽管大型语言模型及其驱动的智能体技术发展迅速, 但将其直接应用于需要深度因果推理和领域知识集成的物理学研究, 仍存在显著差距。

现有主流多智能体框架, 如 MetaGPT 和 Auto Gen, 虽在任务分解与协作上取得了进展, 但其核心设计并非针对科学探究. 它们普遍缺乏对物理第

一性原理的深入理解,难以无缝集成专业的科学计算工具(如第一性原理模拟软件),并且在处理跨尺度的动态任务调度方面存在不足<sup>[73]</sup>.为此,中国人民大学的高泽峰团队<sup>[58]</sup>提出了 PhysAgent——一个专为物理学研究设计的 LLM 驱动的多智能体系统,旨在实现从假设生成、计算建模到数据分析与规律发现的端到端自主科学 workflow.

### 2.2.1 PhysAgent 架构设计

如图 3 所示, PhysAgent 的核心创新在于其仿照人类科研体系构建的角色分工型多智能体架构.该系统由 3 个功能明确、各司其职的智能体构成,通过协同合作完成复杂研究任务.

导师代理 (mentor agent) 扮演经验丰富的物理学教授角色.其核心职能并非直接提供答案,而是通过苏格拉底式提问启发思考,引导研究走向.例如,在推导牛顿第二定律时,导师代理会提示学生代理“避免在假设中引入加速度概念”,从而确保定律是从数据中自然涌现,而非先验预设.学生代理 (student agent) 作为研究任务的直接执行者.学生代理具备强大的技术实现能力,能够熟练运用 Python 及科学计算库(如 NumPy, Matplotlib),并调用 Quantum ESPRESSO, VASP 等专业软件进行第一性原理计算,完成从代码编写、数据可视化到学术报告撰写的全流程工作.领导代理 (leader agent) 充当项目管理者.负责动态任务调度与资源分配,确保导师与学生代理之间高效、有序的迭代

循环,避免同一代理被连续分配任务,从而优化整体工作效率.这种“引导-执行-管理”的三元结构,将人类的科研协作模式成功复刻到 AI 系统中,实现了科学思维与技术操作的有效分离与融合.

PhysAgent 的强大功能,深植于其精心设计的提示词工程与深度领域工具集成.3 个核心代理的提示词 (prompt) 并非通用模板,而是深度内化了各自角色的专业行为和学术标准.例如,导师代理的提示词要求其“以 *Phys. Rev. Lett.* 期刊的标准审视图表”,而学生代理的提示词则明确规定其代码需符合 PEP8 规范,并能处理计算中的错误调试.

在工具层面, PhysAgent 超越了通用框架仅集成基本 API 和网络搜索的能力, PhysAgent 通过 ASE 库将 Quantum ESPRESSO, VASP 等第一性原理计算工具无缝嵌入其 workflow.这种领域专用的工具集成,是 PhysAgent 能够执行高质量科学计算任务的根本保障.与面向软件工程的多智能体相比, PhysAgent 在代理角色、 workflow 机制、处理任务和最终输出上都进行了彻底的物理学重塑.

### 2.2.2 实践验证

研究团队通过一系列跨越经典与量子物理的案例,全面验证了 PhysAgent 的自主发现与执行能力.

在经典物理发现方面, PhysAgent 展现了其“从零开始”的归纳推理能力.系统在完全不知情的

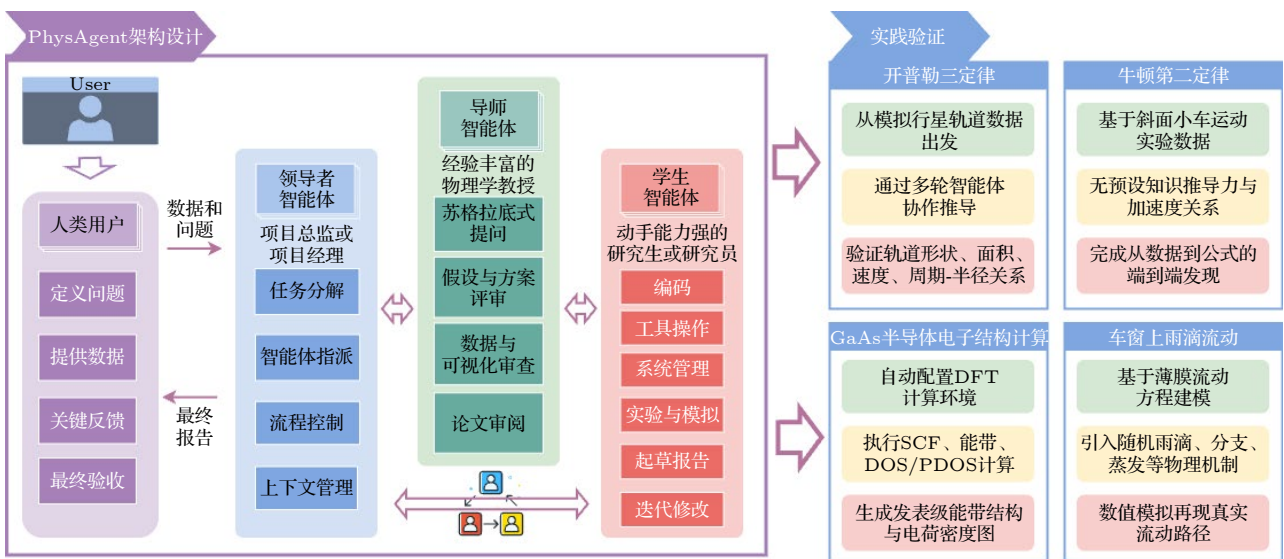


图 3 PhysAgent 架构设计和实践验证示意图

Fig. 3. Schematic diagram of PhysAgent architecture design and experimental verification.

前提下, 仅依据人类用户提供的行星轨道时序数据, 通过多轮智能体协作, 成功推导出开普勒三定律. 同样, 该系统也从斜面小车的运动数据中, 自主发现了力、质量与加速度之间的定量关系, 即牛顿第二定律 ( $F = ma$ ), 复现了人类科学史上的关键飞跃.

在量子材料计算领域, PhysAgent 成功自动化了 GaAs 半导体电子结构的第一性原理计算全流程. 从接收 CIF 文件开始, 学生代理自主完成环境配置、结构建模、自洽场计算、能带与态密度计算, 直至生成电荷密度三维等值面和发表级的综合图表. 这标志着 PhysAgent 已能够处理需要专业计算工具和复杂工作流的现代科研任务.

此外, PhysAgent 还将其能力拓展至复杂现实物理现象的模拟. 针对高速列车车窗上的雨滴流动这一涉及重力、表面张力、空气剪切应力的薄膜流动问题, 系统自主构建了控制方程 (基于 Navier-Stokes 方程简化), 并进行了数值求解与可视化, 其结果与真实观测高度吻合, 展现了其解决非传统、跨学科物理问题的强大适应力.

### 2.3 MetaGPT: 面向软件工程的元编程协作框架

为了更好地了解和对比多智能体在材料计算领域的发展, 本文介绍一个有代表性的多智能体框架——MetaGPT 框架. 其核心思想是将人类在软件工程等专业领域积淀的标准化操作程序 (SOPs) 编码到多智能体协作流程中<sup>[39,42]</sup>, 从而实现复杂任务的高效、连贯且准确的分解与执行.

MetaGPT 的核心理念是“元编程”, 即“编程的编程”<sup>[39]</sup>. 在此框架下, 智能体社会被组织成一个模拟的软件公司, 每个智能体被赋予特定的角色, 如产品经理、架构师、项目经理、工程师和测试工程师等. 每个角色都有明确的职责、目标、约束和专属技能. 这种角色专业化使得复杂的软件工程任务能够被系统地分解为一系列更小、更具体的子任务. 其工作流程严格遵循软件开发的 SOP: 产品经理首先分析用户需求并撰写结构化的产品需求文档 (PRD); 架构师根据 PRD 进行系统设计, 输出包括文件列表、数据结构和接口定义等; 项目经理将设计拆解为具体的任务列表; 工程师负责编写代码; 最后, 测试工程师生成测试用例以保障代码质量. 这种流水线式的协作模式确保了开发过程的

有序性和产出物的标准化.

为了保障协作效率, MetaGPT 引入了结构化的通信协议. 与以往依赖自由形式自然语言对话的框架不同<sup>[75-79]</sup>, MetaGPT 要求智能体以结构化的文档、图表等规范形式进行输出和沟通. 例如, 架构师需要提供系统接口设计和序列流程图. 这种方式极大地减少了自然语言模糊性带来的信息失真, 类似于避免了“传话游戏”中的信息扭曲. 同时, 框架采用了一个全局共享的消息池和基于发布-订阅机制的信息共享模式. 智能体将其结构化产出发布到消息池中, 其他智能体可以根据自身角色和任务需求订阅相关信息, 从而高效地获取所需内容, 避免了点对点通信的拓扑复杂性和信息过载问题<sup>[80,81]</sup>.

此外, MetaGPT 集成了一个创新的可执行反馈机制来进一步提升代码生成质量. 在工程师生成初始代码后, 框架会要求其实际运行代码并进行测试. 如果运行中出现错误, 工程师将根据错误信息、历史执行记忆以及之前生成的需求和设计文档进行调试, 并迭代地修改代码, 直至通过测试或达到重试上限. 这一机制将传统的、非执行性的代码审查或自我反思<sup>[82,83]</sup>, 转变为基于实际运行结果自我修正, 显著提高了最终代码的可执行性和正确性.

在实验验证方面, MetaGPT 在多个代码生成基准测试上展现了卓越的性能. 在 HumanEval<sup>[84]</sup> 和 MBPP<sup>[74]</sup> 基准上, MetaGPT 与 GPT-4 协作分别达到了 85.9% 和 87.7% 的 Pass@1 准确率, 创造了新的最先进水平. 在一个更具挑战性的自建软件工程基准 SoftwareDev 上, MetaGPT 在代码可执行性、代码量以及人工修订成本等指标上均显著优于 AutoGPT<sup>[85]</sup>, LangChain<sup>[86]</sup> 等基线框架. 消融研究进一步证实了角色分工和可执行反馈机制的有效性, 表明它们对于生成高质量、可工作的软件解决方案至关重要.

### 2.4 三个系统比较分析

表 1 给出了三个系统的对比. VASPilot 是一个垂直领域专业化的系统, 其设计理念聚焦于解决计算材料科学中的具体工程问题. 该系统本质上是一个“智能执行引擎”, 旨在将研究人员从繁琐的 VASP 软件操作中解放出来. 其核心价值在于流程优化和效率提升, 通过自动化重复性工作来加速科研进程. 而 PhysAgent 则是一个通用科学探索型

系统, 其设计理念更加宏大——要创建一个能够自主进行科学发现的 AI 伙伴. PhysAgent 不仅关注任务的执行, 更强调科学思维的模拟和知识的创造性发现. 系统的目标是复现人类科学家的认知过程, 从数据中归纳出新的物理规律. 而 MetaGPT 则是一个面向软件工程领域的多智能体协作框架. 其核心思想是将人类在软件工程等专业领域积累的标准化操作程序 (SOPs) 编码到多智能体协作流程中, 从而实现对复杂任务的高效、连贯且准确的分解与执行. MetaGPT 的核心目标是实现软件项目开发的全流程自动化, 从需求分析到代码生成与测试.

在代理角色设计上, 3 个系统体现了不同的思路. VASPilot 采用功能分解架构, 其代理角色基于 workflows 的技术环节划分: 晶体结构代理负责数据获取; VASP 代理负责计算执行; 结果验证代理负责质量把控. 这种设计反映了工程思维中的“分而治之”原则. PhysAgent 采用认知分工架构, 其代理角色基于科学发现过程中的思维模式: 导师代理体现专家学者的指导思维; 学生代理体现研究人员的执行思维; 领导代理体现项目管理的协调思维. 这种设计更接近真实科研团队的协作模式. MetaGPT 则严格遵循软件行业的角色模型, 其代理被赋予产品经理、架构师、工程师、测试工程师等现实世界中的职业角色. 每个角色拥有明确的专

业职责, 通过生成结构化的中间产物 (如 PRD、设计文档、代码文件) 进行协作, 确保了产出物的标准化和流程的有序性.

工具集成层面, VASPilot 通过 MCP 协议实现了深度但狭窄的工具集成, 专门针对 VASP 生态进行优化, 确保在特定领域达到工业级的可靠性. PhysAgent 的集成策略是广度优先, 覆盖多个第一性原理软件和科学计算库, 以保持学科应用的灵活性. MetaGPT 则集成了完整的软件开发工具链, 其工具集围绕代码生成、执行、测试和文档管理展开, 支撑从设计到交付的整个软件生命周期.

人机交互方面, VASPilot 提供了工程化的 Web 监控界面, 适合专业人员对计算过程进行精细控制. PhysAgent 则采用自然语言对话接口, 降低了使用门槛, 使非专业人士也能进行复杂的物理研究. MetaGPT 主要采用文档驱动的交互方式, 用户以需求规格说明作为输入, 系统则输出完整的软件代码和相关技术文档.

从实际应用来看, VASPilot 在结构化问题的处理上表现出色, 能够可靠地完成材料计算的标准 workflow, 在效率提升方面成效显著. PhysAgent 的强项在于处理非结构化问题, 如从观测数据中归纳物理定律、对复杂现象进行建模解释等, 展现了科学发现的潜力. MetaGPT 的优势则体现在处理复

表 1 VASPilot, PhysAgent 和 MetaGPT 三者对比表格  
Table 1. Comparison among VASPilot, PhysAgent, and MetaGPT frameworks.

特性维度	VASPilot	PhysAgent	MetaGPT
应用领域	计算材料科学	通用物理科学发现	软件工程
核心目标	VASP 工作流全自动化	端到端自主科学发现	软件项目全流程自动化
架构基础	CrewAI+MCP 协议	LLM+自定义协作机制	角色代理协作
代理角色设计	功能专业化 管理代理 晶体结构代理 VASP 代理 结果验证代理	认知分工 导师代理(引导) 学生代理(执行) 领导代理(调度)	软件工程角色 产品经理 架构师 工程师 测试员
工具集成	专业计算工具 Pymatgen VASP Slurm 材料数据库	科学计算生态 Quantum ESPRESSO VASP ASE Python 科学栈	软件开发工具 代码执行 文档生成 API 管理
人机交互方式	Web 界面监控 实时流程可视化	自然语言交互 多轮反馈优化	文档驱动 需求规格输入
典型应用场景	1) 能带结构计算 2) 晶格优化 3) 材料筛选 4) 收敛性测试	1) 物理定律发现 2) 复杂现象模拟 3) 电子结构计算 4) 科学假设验证	1) 软件原型开发 2) API 设计生成 3) 技术文档编写

杂软件项目的能力上,其通过角色分工和 SOPs,能够将模糊的自然语言需求转化为具体、可执行且质量较高的软件代码,在 HumanEval 和 MBPP 等代码生成基准上达到了最先进的水平。

VASPIlot 代表的是“专家系统深化”路线,通过在特定领域深度优化来提供可靠的解决方案; PhysAgent 代表的是“通用智能拓展”路线,通过模拟人类认知来开拓 AI 在科学研究中的新边界;而 MetaGPT 代表的则是“流程标准化”路线,通过将成熟的人类工作流程和规范引入多智能体协作,来解决复杂工程任务的自动化问题. 前两者是 AI for Science 领域的工作成果,而后者主要聚焦于 AI for Engineering 领域的技术路径。

### 3 讨论和展望

尽管 VASPIlot, PhysAgent 等框架已经清晰地展示了多智能体系统在物理材料计算中的巨大潜力,标志着该领域从概念验证迈向了初步应用,但将其广泛应用于复杂的科研实践仍面临诸多挑战。

#### 3.1 当前面临的挑战与局限性

当前,面向材料计算的多智能体系统在迈向实际应用的过程中,正面临一系列交织关联的核心瓶颈. 这些挑战不仅源于计算资源的硬约束,更深刻地植根于其学习范式、数据依赖以及系统架构的内在局限性。

首先,计算效率与资源分配是一个根本性瓶颈. 这些系统严重依赖 VASP<sup>[8]</sup>, LAMMPS<sup>[49,87]</sup> 等第一性原理或分子动力学模拟引擎进行核心计算,这些模拟本身已是计算密集型任务. 当系统进行高通量筛选或迭代优化时,会产生海量的计算作业,对计算资源提出极高要求<sup>[73]</sup>. 此时,系统的整体效率高度受制于底层物理模拟的效率. 尽管部分框架(如 PhysAgent)引入了“领导者代理”来优化资源分配,但其调度策略仍处于初级阶段,远未达到人类专家那般智能权衡计算成本与科学收益的水平<sup>[58]</sup>,因此极易成为系统性能瓶颈或导致资源浪费。

其次,系统存在显著的模拟与现实差距及物理一致性问题. 目前部分系统对物理规律的“发现”与验证,都是在完美、洁净的模拟或理想化生成数据

上完成的<sup>[58]</sup>. 这导致其处理真实实验中不可避免的噪声、误差和不完整数据时的鲁棒性存疑. 更重要的是,现有的多智能体框架本质上是数据驱动的,缺乏对能量守恒、对称性等物理基本定律的硬约束<sup>[89]</sup>. 在复杂的决策链中,智能体可能生成数学上正确但物理上荒谬的方案,而大模型固有的幻觉与推理不确定性在脱离流程护栏后会进一步放大此风险<sup>[90]</sup>. 尽管执行过程可追溯,但智能体内部的“思维过程”仍是一个黑箱,当其做出颠覆性发现或严重错误时,研究人员难以理解其深层逻辑,这严重阻碍了科学上的信任与采纳<sup>[91]</sup>.

特别值得关注的是,多智能体在工具学习领域的研究仍处于起步阶段,系统在处理诸如密度泛函理论(DFT)等具有相对明确标准流程的问题时,尚缺乏稳定性与唯一性. 一旦面对需要多步推理与精确计算的任务,智能体往往难以生成唯一正确且可复现的操作路径,从而导致整个多智能体框架在执行层面出现不稳定甚至失效的情况. 针对这一问题,一种常见的思路是通过人为设定规则或进行适度干预,以引导智能体沿着预定义的有效路径完成任务<sup>[88,93]</sup>. 也就是预设一个固定的工作流. 此类方法在短期内可以显著提高结果的正确率与收敛速度,但同时也带来了副作用——多智能体在强约束条件下逐渐丧失自主探索、独立决策和创造性解决问题的能力,最终削弱了系统的智能演化潜力. 为了在稳定性与自主性之间取得平衡,近年来有研究者提出引入强化学习机制<sup>[92]</sup>,通过环境反馈信号与奖励设计的方式,使多智能体能够在反复交互与试错过程中自主学习最优策略<sup>[93-95]</sup>. 这类方法在理论上有望提升模型在复杂任务环境下的自适应能力与鲁棒性,但目前仍主要处于探索与验证阶段,尚需在算法的收敛性、样本效率以及跨任务迁移能力等方面进行深入研究。

此外,多智能体系统还面临着面向多样化计算方法的架构挑战. 当前介绍的多智能体系统(如 VASPIlot)主要面向第一性原理计算(DFT). 然而,计算物理领域广泛使用的其他数值模拟方法(如有限元、相场模拟、分子动力学等)在引入智能体时面临独特挑战。

首先是先验参数问题. 与 DFT 计算不同,有限元、相场等宏观或介观模拟高度依赖一系列先验参数(如弹性常数、梯度能量系数等),这些参数难以自动获取. 为解决这一“参数鸿沟”,需要设计分

层的“参数化智能体”架构。该智能体作为连接不同尺度模拟的桥梁,能够根据宏观模拟的需求,自动规划并触发微观尺度(如DFT)的计算 workflow,以“按需”方式在线生成精确、自洽的材料参数,实现跨尺度模拟的自动化耦合<sup>[96,97]</sup>。

其次是长时序任务管理难题。分子动力学和相场模拟等计算耗时极长。若采用持续的同步轮询机制来监控任务状态,将导致大量无效的Token消耗和查询。为此,需设计基于“回调”的异步事件驱动架构。在这一架构中,一个任务调度与监控代理将扮演HPC资源与计算代理之间的“中间件”角色,其工作机制类似于AiiDA, Fireworks等成熟的科学 workflow 管理平台<sup>[63,98,99]</sup>。仅在作业完成事件触发时,才唤醒后续的数据分析代理,从而在根本上杜绝冗余消耗,优化长时序、高延迟计算任务的管理效率。

对于初学者,VASP等软件参数设置与收敛性判断是关键难题。一个高级的计算代理能够通过集成标准化的收敛性测试协议(如利用Pymatgen, ASE库<sup>[62,63]</sup>),自动、高效地完成如ENCUT和KPOINTS的系统性扫描、数据解析与收敛性判断,其效率和标准化程度远超人工操作,并能自动生成包含收敛曲线的分析报告。然而,在面对因物理原因导致的计算崩溃等“非常规”问题时,当前智能体缺乏人类专家的物理直觉来诊断并调整高级参数(如ALGO, AMIX)。这揭示了当前智能体在“物理推理”能力上的差距。

最后, workflow 的前瞻性验证也是一个挑战,因为真正的测试需要大量真实、未发表且具有代表性的实验数据,而这在现实中往往难以获取<sup>[54]</sup>。因此,对多智能体系统进行全面、可靠的基准测试将需要更广泛的社区协作与数据共享机制。目前在材料计算领域中,多智能体的应用成果大多集中于复现或验证人类科学家已有的发现,例如DREAMS在CO/Pt(111)吸附位点预测中成功复现了文献中FCC位点更稳定的结论。然而,在面对人类尚未解决的科学难题时,如某些复杂催化反应路径的确定或强关联电子体系中的基态预测,多智能体系统仍显得力不从心,尚未展现出超越人类专家的创新能力<sup>[53]</sup>。这反映出当前智能体在自主探索与机理推断方面的局限,也凸显了在保持系统稳定性的同时提升其自主决策与创造性解决问题能力的迫切需求。

为了更好地说明目前多智能体在材料计算领域的应用中存在的共性问题,本文尝试调用了前文中提到的Scilink框架<sup>[13]</sup>。详细调用过程和代码见附录A。我们试运行了两个具有代表性的 workflow,它们分别测试了智能体在几何结构精确性和缺陷电子结构物理修正方面的极限——高几何约束任务:在Au(210)这种高指数表面的1×1小晶胞上吸附大分子C6H6;高物理要求任务构建GaAs(001)极性表面并在顶层引入Ga空位。

其中,附录B展示了任务一的分析结果。研究发现,智能体在处理极易造成原子间距过近的Au(210)组合时,难以自主生成化学上稳定的起点结构。尽管 workflow 最终报告成功,但对POSCAR文件的后续分析(见附录B)显示,吸附的C6H6分子发生了严重的几何畸变(例如,苯环内的C—C键长出现多个值,平均键长偏离预期值1.39 Å,甚至出现额外的C—C键),这证明智能体在强大的几何约束下,牺牲了分子的化学稳定性来满足结构要求,生成了物理上不合理的初始构型。

而附录C则展示了任务二的失败案例。要求智能体构建GaAs(001)极性表面并引入Ga空位,这不仅需要精确的结构控制,还会带来电荷不平衡和缺陷态,对VASP输入文件的NELECT和ISPIN参数提出极高要求。智能体在此任务中表现出双重缺陷:首先,生成的POSCAR文件存在灾难性的几何错误,例如原子数(215个)和晶胞Z轴尺寸(61 Å)严重偏离任务要求的6层2×2 Slab (~47个原子, ~20 Å晶胞);其次,其完全忽略了处理Ga空位缺陷所必需的NELECT(确保电荷中性)和ISPIN = 2(开启自旋极化)参数,这使得后续的DFT计算将缺乏基本的物理可信度。

这两个案例验证了,在面对几何约束复杂或电子结构需要特殊修正的零样本任务时,即使是专为材料计算设计的智能体,也极有可能因缺乏深层次的物理认知或逻辑推理失败而生成错误或无法收敛的初始配置,验证了其在处理复杂任务时缺乏稳定性和科学正确性的局限。

### 3.2 未来发展方向与展望

在未来,多智能体科研系统将朝着深度融合物理机理与人工智能模型的方向演进,超越当前对仿真工具的简单调用,转而将物理定律更深层次地嵌入智能体的推理逻辑中。具体而言,系统将通过物

理信息增强的学习方法,在规划与决策过程中引入物理约束作为奖励或惩罚机制,从而规避违反基本物理规律的行为.使其在灵活决策的同时始终符合物理常识.

### 3.2.1 拓展前沿应用场景的广度与深度

多智能体的核心优势在于通过任务分解与角色协同解决复杂问题,这为材料计算中的多个前沿领域带来范式变革.

在复杂催化反应机理研究方面,面对寻找复杂反应过渡态与路径的挑战,可构建由“规划者”、“知识库”、“执行者”、“分析者”等角色专业化的多智能体团队.有代理起到“规划者”的作用.作为团队的大脑,它负责制定全局搜索策略.它可以调用经典的过渡态搜索算法(如 Dimer 方法)<sup>[100]</sup>,或利用深度强化学习来训练智能体自主学习寻找过渡态的最优策略<sup>[101]</sup>,从而动态规划下一步行动.还有起到“知识库”作用的代理是加速反应路径发现的关键.它不再是一个被动的数据库,而是集成了先进的生成式 AI 模型,例如仅需 2D 分子图即可直接预测 3D 过渡态几何构型的 TSDiff 模型<sup>[102]</sup>.当“规划者”需要一个初始猜测时,它能够迅速提供化学上合理且结构多样化的候选过渡态,将搜索从“盲目”变为“有指导”.有代理作为团队的执行单元,进行计算的功能.它们封装了 VASP, Gaussian 等多种量子化学计算程序,负责接收和执行来自“规划者”的具体计算指令.最后,由分析者代理负责深度解析计算代理返回的原始数据,例如判断计算是否正常收敛、过渡态是否具有唯一的虚频,并将结构化的高层信息反馈给“规划者”,供其进行下一轮决策.通过这种方式,整个反应路径的探索过程将转变为一个自适应的闭环 workflow.这种分工明确、信息闭环的协作模式,有望将复杂的机理探索任务从“手工作坊”式的研究,转变为一个稳健、高效的自动化科学发现流程.

功能材料的“逆向设计”——根据期望的性能反向推导出材料的组分与结构——是实现材料“按需设计”的终极目标.我们设想,一个基于多智能体的“生成-评估”闭环框架,将构成一个高效的自主材料发现平台<sup>[103]</sup>.首先,第 1 个代理的核心是先进的材料生成模型,如晶体扩散变分自编码器(CDVAE)<sup>[104]</sup>.它接收一个或多个目标性能作为输

入,通过在巨大的化学空间中学习“结构-性能”关系,直接生成全新的、具有高可能性满足目标性能的候选材料结构.一旦“生成者代理”提出候选材料,评估者代理将立即启动一个高通量的自动化计算流程,分派多个并行的工作智能体代理高效地调用 DFT 等工具完成关键性能的计算.最后,再由智能体代理起到决策者的作用.这是实现闭环智能的核心.它收集“候选结构-计算性能”数据对,并运用主动学习策略来智能地决定下一轮探索的方向.这种方法已在全自动化的实验平台中成功加速了新材料的发现<sup>[105,106]</sup>.这一流程形成了一个数据驱动、持续迭代的自主学习闭环.每一轮循环,系统都会获得新的知识,并利用这些知识更智能地指导下一轮的材料生成与评估,最终实现从“高通量筛选”到“智能化设计”的飞跃.

将原子尺度的微观行为与工程级别的宏观性能无缝连接,是计算材料科学长期追求的目标.一个以桥梁代理为核心的多智能体系统,能够完美解决这一耦合难题.微观代理负责在原子尺度进行高精度计算.为了提高效率,该智能体可以利用近年来发展迅速的机器学习原子间势 (MLIPs) 来替代昂贵的 DFT 计算,从而在保持精度的同时极大扩展模拟的尺度和时间<sup>[107]</sup>.宏观代理负责在连续介质尺度上运行有限元 (FEA) 等模拟,依赖于由微观模拟提供的精确材料本构关系作为输入.桥梁代理作为连接两个尺度的关键枢纽,其核心职责是实现信息的双向、自动化“握手”.它可以实现正向耦合,能够自动解析微观代理的输出,并利用这些离散的微观数据点,训练一个轻量级的机器学习代理模型来代表材料的本构关系<sup>[108]</sup>,为宏观模拟提供连续、光滑的输入.同时进行反向自适应耦合——当宏观代理的模拟显示器件的某个局部区域出现应力集中时,桥梁代理能够识别该关键区域,并自动触发一个新的、更高精度的微观代理计算任务.计算出的新参数再通过桥梁代理实时更新回宏观模型中.这种通过智能体实现自动化、分布式工程分析的设想,不仅在早期已有探索,并且在当下正与多学科设计思想深度融合,形成了更先进的智能设计框架<sup>[109]</sup>.通过这种自适应的、双向耦合的闭环,多尺度模拟将从一个静态、线性的流程,演变为一个能够根据宏观需求动态调用微观计算的“活”系统,从而真正解决跨尺度建模中的一致性与保真度难题.

### 3.2.2 深化“科学发现”能力：从执行者到假设提出者

多智能体系统的终极愿景是实现从“计算执行者”到“科学假设提出者”的认知跃迁<sup>[110]</sup>。这需要构建一个由多个认知角色组成的智能体生态系统。

首先是知识的内化与表征。“知识挖掘与整合代理”利用科学 NLP 模型 (如 SciBERT<sup>[111]</sup>) 阅读海量文献, 整合材料数据库, 并构建动态更新的多模态知识图谱, 形成系统的“集体智慧”<sup>[112]</sup>。

其次是假设的孕育与提出。“假设生成代理”在知识图谱上进行复杂推理, 主动提出假设: 通过分析图谱拓扑发现“知识空白区”<sup>[113]</sup>; 在数据中寻找隐藏关联——如识别强关联材料的结构“指纹”; 进行“类比推理”, 例如, 基于铜酸盐超导体的特征, 类比提出镍氧化物可能成为新超导材料。

最后是进行验证与迭代闭环。提出的假设将由规划代理设计计算 workflow 进行验证。无论结果如何, 新产生的知识都将反馈更新知识图谱, 形成一个永不停止、自我完善的科学发现螺旋, 使系统演化为能够自主探索前沿的“智能科研伙伴”。

### 3.2.3 融合实验闭环: 构建自驱动实验室

进一步地, 系统将致力于构建覆盖“设计-计算-实验”全流程的闭环自主科研生态<sup>[13]</sup>。未来的多智能体不仅局限于理论预测, 还将跨越至真实世界的实验操作, 通过集成自动化合成与表征平台——如 SciN $\Sigma$ T 所设想的“分布式实验室集合体”<sup>[14]</sup>。当多智能体系统进化为一个完全自主的“自驱动实验室”<sup>[114,115]</sup>的中央“大脑”时, 其真正的变革性力量才会得以完全释放。这要求系统不仅能提出科学假设, 还必须能主动地规划实验、实时地解读数据, 并动态地调整其科研策略。我们设想, 将由一组专门的智能体协同工作, 以管理整个实验 workflow。

一旦假设生成代理提出了一个目标材料, 实验规划代理将接管后续工作。其核心职责是将一个抽象的科学目标——例如, “合成掺杂 5% Sr 的 BaTiO<sub>3</sub>, 以实现 250 K 的顺电-铁电相变温度”, 转化为一个具体的、机器可执行的协议。该智能体访问一个包含各种合成程序, 如溶胶-凝胶法、固相反应法、水热法等的综合知识库。它会选择最合适的合成方法, 并生成一份详尽的实验“配方”, 精确指定前驱体化学品、摩尔比、溶剂体积、升温/降温速率和保温时间等。

至关重要的是, 这份“配方”并非人类可读的文本, 而是一个结构化的命令文件 (如 JSON 或 XML 格式)。该文件随后被分派给实验室的自动化硬件。多智能体系统可以直接指挥液体处理工作站来分配前驱体溶液、控制自动管式炉, 并管理样品转移的物流, 从而在无需人工干预的情况下, 高效地执行整个高通量合成任务<sup>[105]</sup>。

当样品合成完毕, 它们将被自动传送至表征工作站。此时, 一个数据分析代理将开始主导, 启动测量程序并执行即时的、在线的数据分析。无需等待人类专家进行可能耗时数小时甚至数天的手动分析, 数据分析代理会调用专门的机器学习模型。例如, 对于原子力显微镜 (AFM) 图像, 一个集成了神经网络的自动化方法就能够自主地发现其中的分子结构<sup>[116]</sup>; 而对于 XRD 图谱, 同样有 AI 模型可以进行快速的物相鉴定<sup>[117]</sup>。同样, 对于透射电子显微镜 (TEM) 图像, 另一个深度学习模型可以自主地识别和分类晶体缺陷、测量晶粒尺寸, 或分割出不同形貌的纳米颗粒<sup>[118]</sup>。

从原始数据到物理描述符: 该智能体的输出不仅仅是一个分类标签, 而是一组结构化的、量化的物理描述符, 例如, {"phase": "perovskite", "purity": "98%", "lattice\_constant\_a": "3.99 Å"}。这些被处理过的信息将立刻被用于下一阶段的认知决策。

最后是自适应反馈与动态的方案调整。这是最关键的阶段, 它闭合了整个发现的回路, 实现了真正的自主科学发现。一个更高层级的策略代理会持续地将实时的实验结果与最初的计算预测进行比较。

策略代理评估预测性质与数据分析代理所提供结果之间的差异。如果实验成功地合成了预期的物相并展现出期望的性质, 该智能体可以自主决定进行更深入的二次表征, 例如测量电学或磁学性质。

基于贝叶斯优化的动态调整: 如果实验失败或产生了意外的结果, 策略师并不会简单地丢弃这次结果, 而是将其视为一个有价值的新数据点。它会运用贝叶斯优化或其他主动学习的策略<sup>[113,119]</sup>, 来更新其内部关于高维合成参数空间的代理模型。基于这个更新后的模型, 它会做出一个智能化的决策——即为了最大化信息增益, 下一步应该执行哪个“最优”的实验。它可能会判断, 稍微提高煅烧温度是下一个最有希望的尝试; 或者, 它也可能判定

当前的合成路线是一个死胡同, 并果断地转向一个全新的前驱体组合.

这个“规划-执行-分析-调整”的持续迭代和自我修正的循环, 彻底改变了传统的科研过程. 它使多智能体系统能够在复杂的、高维的实验空间中自主导航, 从每一次成功和失败中学习, 从而以前所未有的速度和效率, 快速地收敛到具有所需功能的新型材料.

### 3.2.4 精细化 workflow 和科研范式变化

为了具体阐明智能体如何优化参数设置与收敛性测试, 我们以 VASP 计算中最经典的 ENCUT (平面波截断能) 和 KPOINTS ( $K$  点网格) 收敛性测试为例, 详细拆解一个高级智能体的 workflow. 其表现不仅旨在媲美, 更在效率、标准化和可复现性方面超越了人类.

一个设计精良的 VASP 计算代理内部集成了一套基于成熟计算材料学库的、标准化的“收敛性测试协议”. 当接收到“对硅 (Si) 晶体进行收敛性测试”的高级指令后, 协议启动与自动编排, 智能体首先启动内置的收敛性测试协议. 该协议的知识库中固化了最佳实践: “先进行  $K$  点收敛, 再用收敛后的  $K$  点进行 ENCUT 收敛”. 这种严格的顺序编排, 避免了初学者常犯的同时更改多个参数的错误.

之后, 智能体自动创建一系列 KPOINTS 文件 (例如, 从  $2 \times 2 \times 2$  到  $10 \times 10 \times 10$ ), 同时保持其他所有 INCAR 参数 (包括一个适中的、未收敛的 ENCUT) 固定不变. 并且它将这一系列任务通过 3.1 节讨论的“异步事件驱动”机制提交到计算集群, 并高效等待所有任务完成. 所有计算结束后, 智能体自动调用解析器 (如 Pymatgen 的 Vasprun) 从 vasprun.xml 中提取每个  $K$  点密度对应的总能量. 随后, 它将能量数据进行归一化 (例如, 转换为 meV/atom), 并计算相邻数据点之间的能量差. 当能量差连续 (例如, 连续两个点) 小于一个预设的、物理意义明确的阈值 (如 1 meV/atom) 时, 智能体判定  $K$  点已收敛<sup>[98]</sup>.

一旦确定了收敛的  $K$  点密度 (例如,  $8 \times 8 \times 8$ ), 智能体立即将其设置为后续计算的固定参数. 接着, 它自动生成一系列新的 INCAR 文件, 其中仅 ENCUT 参数从一个较低值 (如 POTCAR 中 ENMAX 的 0.8 倍) 系统性地增大到一个较高值 (如 ENMAX 的 2.0 倍), 并再次异步提交计算. 它

重复上述的数据解析和收敛性判断过程, 最终确定一个收敛的 ENCUT 值. 完成所有测试后, 智能体不仅得出了结论 (“ $K$  点收敛于  $8 \times 8 \times 8$ , ENCUT 收敛于 520 eV”), 更重要的是, 它会自动生成一份包含收敛曲线图和核心输入/输出参数的标准化报告, 如一个 Jupyter Notebook 或 PDF 文档.

与人类专家的比较, 在“常规”任务上, 智能体全面占优. 对于标准的收敛性测试, 智能体的操作比任何人类都更快、更不知疲倦、更不容易出错, 例如, 不会有手动修改 INCAR 时的“笔误”. 它能在数小时内完成一个研究生可能需要断续几天才能完成的工作, 且保证了流程的绝对标准化.

在“非常规”问题上, 人类经验仍是关键. 如果某个计算因物理原因而崩溃, 当前的智能体可能仅能根据预设规则进行处理. 而一个有经验的研究生则能凭借物理直觉, 通过分析 OUTCAR 中的详细信息, 判断出可能需要调整 ALGO 或 AMIX 等高级参数. 这正是未来智能体需要通过引入更强的“物理推理”能力来弥补的差距.

智能体工具的普及, 绝不意味着对学生科研能力要求的降低, 而是标志着一次深刻的能力模型转型. 课题组和教育者应主动拥抱这一变革, 对研究生培养方式进行前瞻性设计.

研究生培养要从“操作者”转换到“战略家”. 教学的重点应从“如何手动修改 INCAR 中的每一个参数”, 转变为“如何提出一个有价值的科学问题, 并将其分解为智能体可以理解的一系列高层指令”. 学生的精力应从繁琐的、重复性的计算执行中解放出来, 投入到更具创造性的前端科学问题设计和后端结果深度分析中. 大语言模型正加速从知识挖掘到智能设计的全链条材料研发过程<sup>[123]</sup>, 而目标性质导向的材料生成方法正在推动按需构筑的逆向设计<sup>[124]</sup>.

同时要培养研究生的“智能体驾驭能力”. 未来的研究生必须具备“AI-Literacy”的人工智能素养. 他们需要理解智能体的能力边界: 知道它擅长什么——如高通量筛选、标准化流程, 材料科学专用 LLM 平台如 MatChat 已在无机材料合成路径预测中展现强大能力<sup>[120]</sup>; 不擅长什么——如非标准问题的物理诊断等; 学会如何验证和批判智能体的输出, 而不是盲目信任; 并掌握与智能体高效交互的技能, 例如, 通过编写简单的 Python 脚本来调用 Agent API, 实现更复杂的自定义 workflow.

我们也要强化研究生“科学批判性思维”与“数据科学”技能训练. 当获取数据的成本急剧下降时, 诠释数据、从海量结果中发现科学规律的能力就变得前所未有的重要<sup>[123]</sup>. 大语言模型已在电池科研全流程中证明其价值<sup>[121]</sup>, 而实验与理论数据融合平台如 MaterialsGalaxy 通过 AI 工具整合多源数据, 为材料发现提供了新范式<sup>[56]</sup>. 研究生的培养应强化科学直觉. 面对智能体产出的“最优结构”, 学生需要能判断其在物理上或化学上是否合理, 其结果与现有理论或实验是否存在矛盾. 其次是培养数据科学技能. 学生的“实验记录本”正越来越多地从纸质变为 Jupyter Notebook. 他们需要掌握利用 Python 或其他软件对智能体产生的大规模数据集进行可视化、统计分析和二次挖掘的能力. 实验与理论数据融合平台如 MaterialsGalaxy, 通过 AI 工具整合多源数据, 为材料发现提供了新范式<sup>[122]</sup>.

综上所述, 智能体并非要取代研究生, 而是要成为一个强大的“增强智能”工具, 将他们从“计算工匠”的角色中解放出来, 使他们能更专注于作为“科学家”的核心使命——提出深刻问题, 并做出创造性发现.

### 3.3 结论

多智能体系统正深刻地变革着物理与材料计算的研究范式, 将研究人员从繁琐的操作中解放出来, 并赋予了探索前所未有的复杂问题的能力. 尽管在可靠性、通用性和智能化程度方面仍面临挑战, 但通过将物理机理深度融入 AI、构建人机协同的闭环科研生态系统, 多智能体系统有望从一位得力的“助理”成长为能够提出创见、引领发现的“科研伙伴”, 最终加速人类对物质世界的理解和创新进程.

感谢中国人民大学高性能计算物理实验室提供计算资源.

#### 附录A Scilink 包调用代码示例

```
import scilink
#任务一
from scilink.workflows import DFTWorkflow
complex_task_request = "Generate a 1x1 Au(210) surface with a C6H6 (Benzene) molecule adsorbed flat on a bridge site, using a 4-layer slab and 15 Å of vacuum.
```

```
Ensure the C-C bond length of Benzene is initially set to 1.39 Å. "
```

```
# 创建 DFT 工作流
dft_workflow_failure1 = DFTWorkflow(output_dir="dft_1")
```

```
# 运行完整工作流
```

```
Result1=dft_workflow_failure1.run_complete_workflow(complex_task_request)
```

```
#任务二
```

```
from scilink.workflows import DFTWorkflow
```

```
gaas_failure_request = "Generate an unreconstructed 2x2 GaAs(001) slab, 6 layers thick, with an Arsenic (As) termination on the top surface. Introduce a single Gallium (Ga) vacancy in the top-most layer of the slab and prepare VASP inputs for a structural relaxation. Use 12 Å of vacuum."
```

```
# 创建 DFT 工作流
```

```
dft_workflow_failure2 = DFTWorkflow(output_dir="dft_2")
```

```
# 运行完整工作流
```

```
Result2=dft_workflow_failure2.run_complete_workflow(gaas_failure_request)
```

#### 附录B 任务一——Au(210) 的 1×1 小晶胞上吸附 C6H6

任务描述: Generate a 1×1 Au(210) surface with a C6H6 (Benzene) molecule adsorbed flat on a bridge site, using a 4-layer slab and 15 Å of vacuum. Ensure the C—C bond length of Benzene is initially set to 1.39 Å.

在生成的日志 workflow\_log.txt 中, 我们可以看到初始代码生成并执行了 4 次. (10:56:32, 396 (Attempt 1) → 10:56:57, 285 (Attempt 2) → 10:57:22, 697 (Attempt 3) → 10:57:47, 074 (Attempt 4)). 说明智能体缺乏生成唯一正确、可执行代码的能力, 必须依赖多轮代码生成、执行和重试才能产出第一个结构文件 (Au210\_C6H6\_bridge.poscar).

外部 Validator Agent 一次性发现了 4 个逻辑错误. 包括苯分子所有原子被错误地固定 (F F F); 错误地将 2×1 超胞当做 1×1 表面; 约束标签 (tag 0) 错误地包含了苯分子. 16 个金原子中的 4 个被固定以代表 '底部两层', 但固定底部 8 个原子是模拟体相更常见且稳健的方法.

虽然智能体在接收到 Validator Agent 的 4 个错误修正提示后, 生成了新脚本并执行成功. 但是从生成的 Au210\_C6H6\_bridge.poscar 文件来看, 结果仍然存在一些问题.

```
# 获取原子索引
```

```

2025-11-26 10:56:07,873 - INFO: root: Sending request to LLM...
2025-11-26 10:56:32,396 - INFO: root: Saved script for attempt 1 to: dft_unstable_adsorption_failure/
script_Generate_a_1x1_Au_210_surface_attempt1_20251126_105632.py
2025-11-26 10:56:32,399 - INFO: root: Attempting to execute generated script...
2025-11-26 10:56:32,400 - WARNING: root: ⚠ EXECUTING AI-GENERATED CODE - Security sandbox verified
2025-11-26 10:56:33,853 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE tool
2025-11-26 10:56:33,855 - INFO: root: Sending request to LLM...
2025-11-26 10:56:57,285 - INFO: root: Saved script for attempt 1 to: dft_unstable_adsorption_failure/
script_Generate_a_1x1_Au_210_surface_attempt1_20251126_105657.py
2025-11-26 10:56:57,326 - INFO: root: Attempting to execute generated script...
2025-11-26 10:56:57,327 - WARNING: root: ⚠ EXECUTING AI-GENERATED CODE - Security sandbox verified
2025-11-26 10:57:00,955 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE tool
2025-11-26 10:57:00,961 - INFO: root: Sending request to LLM...
2025-11-26 10:57:22,697 - INFO: root: Saved script for attempt 1 to: dft_unstable_adsorption_failure/
script_Generate_a_1x1_Au_210_surface_attempt1_20251126_105722.py
2025-11-26 10:57:22,717 - INFO: root: Attempting to execute generated script...
2025-11-26 10:57:22,718 - WARNING: root: ⚠ EXECUTING AI-GENERATED CODE - Security sandbox verified
2025-11-26 10:57:26,968 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE tool
2025-11-26 10:57:26,971 - INFO: root: Sending request to LLM...
2025-11-26 10:57:47,074 - INFO: root: Saved script for attempt 1 to: dft_unstable_adsorption_failure/
script_Generate_a_1x1_Au_210_surface_attempt1_20251126_105747.py
2025-11-26 10:57:47,080 - INFO: root: Attempting to execute generated script...
2025-11-26 10:57:47,080 - WARNING: root: ⚠ EXECUTING AI-GENERATED CODE - Security sandbox verified
2025-11-26 10:57:48,233 - INFO: scilink.agents.sim_agents.val_agent: Starting LLM-based validation with

```

图 B1 任务一 workflow\_log.txt 文件

Fig. B1. Task 1 workflow\_log.txt files.

```

2025-11-26 10:57:48,577 - INFO: scilink.agents.sim_agents.val_agent: Sending request to Validator LLM for
full validation and script hints...
2025-11-26 10:58:40,177 - INFO: scilink.agents.sim_agents.val_agent: LLM full validation feedback and
script hints received successfully.
2025-11-26 10:58:40,183 - WARNING: scilink.agents.sim_agents.val_agent: LLM validation for '/Users/wing/
Desktop/dft_unstable_adsorption_failure/Au210_C6H6_bridge.poscar' found issues requiring script correction:
['All adsorbate atoms (Carbon and Hydrogen) are incorrectly constrained (fixed) in the output POSCAR file,
as indicated by the 'F F F' flags on their coordinate lines. These atoms must be free to move during
relaxation.', 'The script error stems from applying constraints after combining the slab and adsorbate.
When the untagged benzene atoms are added to the tagged slab, they are assigned a default tag of '0'. The
subsequent constraint selection `if a.tag in (0, 1)` incorrectly includes all benzene atoms along with the
bottom slab layer (tag 0).', 'The user requested a '1x1' surface, but the generated structure is a 2x1
supercell of the primitive Au(210) surface. This is evidenced by the 16 Au atoms in the 4-layer slab,
whereas a primitive cell would contain 8 atoms (2 atoms/layer).', 'Only 4 out of 16 Au atoms are fixed to
represent the 'bottom two layers'. For a 16-atom slab, fixing the bottom 8 atoms is a more common and
robust approach to simulate the bulk, although fixing 4 may be a reasonable choice for the highly
corrugated (210) surface.']
2025-11-26 10:58:40,186 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE tool
2025-11-26 10:58:40,187 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE tool
2025-11-26 10:58:40,188 - INFO: root: Sending request to LLM...

```

图 B2 任务一 workflow\_log.txt 文件

Fig. B2. Task 1 workflow\_log.txt files.

```

c_indices = [i for i, symbol in enumerate(atoms.get_
chemical_symbols()) if symbol == 'C']
au_indices = [i for i, symbol in enumerate(atoms.get_
chemical_symbols()) if symbol == 'Au']
# 遍历所有 C 原子对
for i in range(num_c):
    for j in range(i + 1, num_c):
        # 计算原子 i 和 j 之间的距离 (使用最小像约定
mic=True)
        dist = atoms.get_distance(c_indices[i], c_indices[j],
mic=True)
        # 假设 C—C 键应该在 1.0 Å 到 1.8 Å 之间
        if 1.0 < dist < 1.8:
            all_cc_lengths.append(dist)
# 找到所有 C—Au 原子对的最小距离
min_au_c_distance = atoms.get_distance(c_indices[0],
au_indices[0], mic=True) # 初始化
found_distance = False

```

for c\_idx in c\_indices:  
 for au\_idx in au\_indices:  
 dist = atoms.get\_distance(c\_idx, au\_idx, mic=True)  
 min\_au\_c\_distance = min(min\_au\_c\_distance, dist)  
 found\_distance = True

通过验证 C—C 键长 (苯环内部) 和 C—Au 最小距离, 发现苯环 C—C 键总数为 8 (预期为 6)。我们预期 C—C 键长为 1.39 Å, 但是实际结果是 4 个 1.3934 Å 的 C—C 键和 2 个 1.6725 Å 的 C—C 键, 平均 C—C 键长为 1.461 Å。这说明发生了严重的几何畸变。智能体为了满足吸附在 1×1 小晶胞上的约束, 将苯环压扁或拉伸, 导致整个分子畸形, 甚至使本不相连的原子靠得过近, 形成 C—C 键。说明 C—C 键长出现多个值或严重偏离目标, 证明智能体未能精确执行几何约束。此外, C—Au 最小距离为 2.583 Å, 接近化学合理最小距离 2.21 Å, C—Au 最小距离在合理范围内。

虽然智能体最终“成功”生成了文件, 但其结果是物理上不合理的。

```

Au210_C6H6_bridge_corrected.poscar
1 Au C H
2 1.0000000000000000
3 9.1231573481991415 0.0000000000000000 0.0000000000000000
4 0.0000000000000000 4.0800000000000001 0.0000000000000000
5 0.0000000000000000 0.0000000000000000 23.5862101437393967
6 Au C H
7 16 6 6
8 Selective dynamics
9 Direct
10 0.0500000000000000 0.2500000000000000 0.3179824123627069 F F F
11 0.1500000000000000 0.2500000000000000 0.5887427466775659 T T T
12 0.3499999999999999 0.7500000000000000 0.5500626989183003 T T T
13 0.8499999999999999 0.2500000000000000 0.5500626989183003 T T T
14 0.5499999999999999 0.2500000000000000 0.5113826511590347 T T T
15 0.7500000000000000 0.7500000000000000 0.4727026033997691 T T T
16 0.0500000000000000 0.7500000000000000 0.5113826511590346 T T T
17 0.2499999999999999 0.2500000000000000 0.4727026033997691 T T T
18 0.6499999999999999 0.7500000000000000 0.5887427466775659 T T T
19 0.1500000000000000 0.7500000000000000 0.3953425078812380 T T T
20 0.4500000000000000 0.7500000000000000 0.4340225556405036 T T T
21 0.6499999999999999 0.2500000000000000 0.3953425078812380 T T T
22 0.3500000000000001 0.2500000000000000 0.3566624601219724 F F F
23 0.5500000000000002 0.7500000000000000 0.3179824123627069 F F F
24 0.8500000000000001 0.7500000000000000 0.3566624601219724 F F F
25 0.9500000000000000 0.2500000000000000 0.4340225556405036 T T T
26 0.4761797668805074 0.2049570315538310 0.6820175876372931 T T T
27 0.3238202331194922 0.2049570315538310 0.6820175876372931 T T T
28 0.2476404662389848 0.5000000000000000 0.6820175876372931 T T T
29 0.3238202331194924 0.7950429684461691 0.6820175876372931 T T T
30 0.5523595337610149 0.5000000000000000 0.6820175876372931 T T T
31 0.4761797668805074 0.7950429684461691 0.6820175876372931 T T T
32 0.2640821425441301 -0.0264075983787762 0.6820175876372931 T T T
33 0.6718357149117391 0.5000000000000000 0.6820175876372931 T T T
34 0.5359178574558695 1.0264075983787764 0.6820175876372931 T T T
35 0.2640821425441304 1.0264075983787766 0.6820175876372931 T T T
36 0.1281642850882607 0.5000000000000001 0.6820175876372931 T T T
37 0.5359178574558695 -0.0264075983787763 0.6820175876372931 T T T
    
```

图 B3 Au210\_C6H6\_bridge.poscar 文件

Fig. B3. Au210\_C6H6\_bridge.poscar file.

### 附录C 任务二——构建 GaAs(001) 极性表面且在顶层引入 Ga 空位

任务描述: Generate an unreconstructed 2x2 GaAs (001) slab, 6 layers thick, with an Arsenic (As) termination on the top surface. Introduce a single Gallium (Ga) vacancy in the top-most layer of the slab and prepare VASP inputs for a structural relaxation. Use 12 Å of vacuum.

在生成的日志 workflow\_log.txt 中, 多智能体进行了 4 次尝试, Validator Agent 反复指出结构文件 (POSCAR) 与智能体脚本中处理的材料 (SiC, MoS2/WS2, Fe2O3) 不一致。

智能体试图修正前期的结构错误, 但却生成了一个原子数高达 215 个 (预期约 23 个) 的结构。晶胞 Z 轴 (~61 Å) 和初始化学计量比均完全错误。

检查生成的 POSCAR 结构文件, 可发现实际原子数为 215 个 (112+103), 但是我们的预期原子数是 47 个

```

2025-11-26 11:39:11,736 - INFO: scilink.agents.sim_agents.val_agent: LLM full validation
feedback and script hints received successfully.
2025-11-26 11:39:11,795 - WARNING: scilink.agents.sim_agents.val_agent: LLM validation for '/
Users/wing/Desktop/dft_gaas_defect_failure/POSCAR' found issues requiring script correction:
[["The most critical issue is that the output structure (GaAs) does not match the material
specified in the request and script (SiC). The elements, stoichiometry, and atom counts in the
POSCAR are entirely inconsistent with the script's logic.", "The structure in the POSCAR file is
severely non-stoichiometric, with 28 Arsenic atoms and 23 Gallium atoms. This extreme vacancy
concentration was not part of the original request and makes the structure's physical relevance
questionable without further justification.", "There is a discrepancy in the applied
constraints. The script logic aims to fix the four central layers, which is appropriate for a
grain boundary. The POSCAR file, however, has the bottom two layers fixed, which is a convention
used for surface slab calculations.", "The in-plane lattice vectors of the POSCAR correspond to
a 2x2 supercell of a (111) surface, but the atomic positions along the z-axis are highly
disordered and do not resemble the expected clean twin boundary structure of a Σ3(111) grain
boundary.", "The user request for a Carbon-terminated interface cannot be fulfilled, as the
structure is composed of Gallium and Arsenic."]]
2025-11-26 11:39:11,827 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool
2025-11-26 11:39:11,831 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool
    
```

```

2025-11-26 11:40:54,676 - INFO: scilink.agents.sim_agents.val_agent: LLM full validation
feedback and script hints received successfully.
2025-11-26 11:40:54,693 - WARNING: scilink.agents.sim_agents.val_agent: LLM validation for '/
Users/wing/Desktop/dft_gaas_defect_failure/POSCAR' found issues requiring script correction:
["The provided POSCAR file describes a Gallium Arsenide (GaAs) structure, which is fundamentally
inconsistent with the user's request for a MoS2/WS2 heterostructure and the Python script's
logic. This appears to be an error where the wrong file was provided.", "The script's logic for
creating 'R-type' stacking is flawed. It applies the stacking shift to the MoS2 and WS2 blocks
separately before joining them. This will create a stacking fault (e.g., ...C-A|-A-B...) at the
interface, rather than a continuous R-type sequence (...C-A-B...) across the entire
heterostructure.", "The method used to identify layers within the 'make_r_type' function is not
robust. The calculation 'layer_index = int(np.floor(z_coord / LAYER_SPACING))' can incorrectly
assign atoms from the same S-M-S trilayer to different layer indices, which would result in a
scrambled and physically incorrect structure.", "The implementation of atomic constraints is
incorrect. The user requested to fix the 'bottom 3 layers' (three S-M-S trilayers), but the
script only fixes the atoms belonging to the single bottom-most trilayer (the first three atomic
planes)."]
2025-11-26 11:40:54,700 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool
2025-11-26 11:40:54,703 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool

2025-11-26 11:42:04,210 - INFO: scilink.agents.sim_agents.val_agent: LLM full validation
feedback and script hints received successfully.
2025-11-26 11:42:04,245 - WARNING: scilink.agents.sim_agents.val_agent: LLM validation for '/
Users/wing/Desktop/dft_gaas_defect_failure/POSCAR' found issues requiring script correction:
["Elemental Composition Mismatch: The most critical issue is that the provided POSCAR file
contains Gallium (Ga) and Arsenic (As), while the user request and Python script explicitly
specify Iron (Fe) and Oxygen (O) for a Hematite (Fe2O3) slab. This discrepancy means the
structure file is completely unrelated to the script's purpose.", "Incorrect Slab Thickness
Specification: The user requested a slab '15 layers thick', but the script implements this using
'min_slab_size=15'. In pymatgen's 'SlabGenerator', this parameter defines the minimum thickness
in Angstroms, not layers, failing to meet a key requirement of the request.", "Unverified
Surface Termination: The script filters for symmetric and stoichiometric slabs and simply
selects the first result. It does not contain any logic to confirm that this slab is actually
Fe-terminated. For the Fe2O3(0001) surface, the most common symmetric and stoichiometric slab is
O-terminated, meaning the script would likely fail to generate the requested structure.",
"Contradictory Request for Stoichiometry: The user requested a 'stoichiometric composition' but
also an 'Fe vacancy'. While the script correctly interprets this as creating a vacancy in a
stoichiometric slab (resulting in a non-stoichiometric final structure), this inherent
contradiction in the request should be noted.", "Incorrect POSCAR Stoichiometry: As a direct
result of the elemental mismatch, the structure in the POSCAR file has a composition of 28 As
and 23 Ga atoms, which is inconsistent with the requested 2:3 ratio of Fe:O."]
2025-11-26 11:42:04,260 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool
2025-11-26 11:42:04,263 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool

2025-11-26 11:43:36,391 - INFO: scilink.agents.sim_agents.val_agent: LLM full validation
feedback and script hints received successfully.
2025-11-26 11:43:36,446 - WARNING: scilink.agents.sim_agents.val_agent: LLM validation for '/
Users/wing/Desktop/dft_gaas_defect_failure/POSCAR' found issues requiring script correction:
["Incorrect Surface Orientation: The script was requested to generate a (001) surface, which
should have a square or rectangular in-plane lattice. The resulting POSCAR file's lattice
vectors define a hexagonal-like cell in the xy-plane, which is characteristic of a (111)
surface, not a (001) surface. The top-down image (Z-axis view) confirms this hexagonal atomic
arrangement.", "Incorrect Stoichiometry: A perfect 2x2 6-layer GaAs(001) slab should contain 24
Ga and 24 As atoms. After introducing a single Ga vacancy, the expected composition is 23 Ga and
24 As. The generated POSCAR file incorrectly contains 23 Ga and 28 As atoms, making the slab
unphysically arsenic-rich.", "Incorrect Number of Layers: The user requested a 6-layer slab.
However, a review of the unique z-coordinates in the generated POSCAR file reveals more than 10
distinct atomic planes, indicating the slab is much thicker than requested.", "Script-Level Root
Cause: The likely source of all issues is the use of the default primitive cell from
'ase.build.bulk()' as input to the 'ase.build.surface()' function. To correctly generate a (001)
surface with the expected square symmetry and layer structure, the conventional (cubic) unit
cell must be used instead."]
2025-11-26 11:43:36,457 - INFO: scilink.agents.sim_agents.structure_agent: Selected default ASE
tool

```

图 C1 任务二 workflow\_log.txt 文件

Fig. C1. Task 2 workflow\_log.txt files.

```

2025-11-26 11:45:25,217 - INFO: scilink.agents.sim_agents.val_agent: LLM full validation
feedback and script hints received successfully.
2025-11-26 11:45:25,276 - WARNING: scilink.agents.sim_agents.val_agent: LLM validation for '/
Users/wing/Desktop/dft_gaas_defect_failure/POSCAR' found issues requiring script correction:
["Incorrect Slab Thickness and Atom Count: The primary issue is that the generated structure is
not a 6-layer slab. The POSCAR file lists 215 atoms (112 As, 103 Ga), whereas a 2x2 6-layer slab
with one vacancy should contain approximately 23 atoms. The visualization clearly shows a very
thick slab.", "Incorrect Initial Stoichiometry: The script's logic relies on creating a
symmetric 7-layer slab and then removing the bottom layer. However, the atom counts (112 As vs.
104 Ga before the vacancy is created) show that the initial 'symmetric' slab was non-
stoichiometric, which is incorrect for a symmetric GaAs(001) slab and indicates a flaw in the
generation process.", "Excessive Z-Lattice Vector: The cell height is ~61 Å. For a 6-layer GaAs
slab (~7-8 Å thick) and 12 Å of vacuum, the total z-vector should be around 20 Å. The generated
cell is unnecessarily large, which would waste computational resources, and it is a direct
consequence of the incorrectly thick slab being generated."]

```

图 C2 任务二 workflow\_log.txt 文件

Fig. C2. Task 2 workflow\_log.txt files.

```

≡ POSCAR
1  As Ga
2  1.0000000000000000
3  11.305999999999992  0.000000000000000  0.000000000000000
4  0.000000000000000  11.305999999999992  0.000000000000000
5  0.000000000000000  0.000000000000000  60.744499999999951
6  As Ga
7  112 103
8  Selective dynamics

```

图 C3 POSCAR 结构文件

Fig. C3. POSCAR structural file.

```

≡ INCAR
1  SYSTEM = GaAs(001) 2x2 slab with Ga vacancy - Relaxation
2
3  # --- Start Parameters ---
4  ISTART = 0          # Start from scratch
5  ICHARG = 2         # Charge from atomic superposition
6  PREC = Accurate    # Precision level
7
8  # --- Electronic Relaxation ---
9  ENCUT = 450        # Energy cutoff in eV
10 EDIFF = 1E-06      # SCF energy convergence criterion
11 LREAL = Auto       # Automatic optimization for real-space projection
12 ALGO = Normal      # Blocked Davidson algorithm
13 NELM = 100         # Max number of SCF steps
14
15 # --- Ionic Relaxation ---
16 IBRION = 2          # Conjugate Gradient algorithm for ion relaxation
17 NSW = 200           # Max number of ionic steps
18 ISIF = 2            # Relax atomic positions, keep cell shape/volume fixed
19 EDIFFG = -0.02     # Force convergence criterion (eV/Angstrom)
20
21 # --- Smearing/DOS ---
22 ISMEAR = 0          # Gaussian smearing
23 SIGMA = 0.05        # Smearing width in eV
24
25 # --- Special Settings for Slab ---
26 IVDW = 11           # DFT-D3 van der Waals correction (Grimme's method with zero damping)
27 LDIPOL = .TRUE.     # Include dipole corrections for potential
28 IDIPOL = 3          # Dipole correction along the z-axis
29
30 # --- Output Control ---
31 LWAVE = .FALSE.     # Do not write WAVECAR
32 LCHARG = .TRUE.     # Write CHGCAR
33

```

图 C4 INCAR 文件

Fig. C4. INCAR file.

( $2 \times 2 \times 6$  Slab, 1 Ga vacancy). 原子数相差巨大 (4.6 倍), 结构彻底错误。

计算晶胞  $Z$  轴尺寸为  $60.74 \text{ \AA}$ , 而根据任务要求的预期值应该是约  $20 \text{ \AA}$  (6 层厚度 +  $12 \text{ \AA}$  真空). 真空层冗余, 这将导致  $Z$  轴方向的平面波计算急剧增加, 浪费大量的计算资源。

在 INCAR 文件中, 缺失电子数修正 NELECT、自旋极化 ISPIN. 会导致无法确保计算体系的电荷中性, 基态能量和缺陷能级计算结果将缺乏物理可信度; 无法识别缺陷导致的自旋极化现象, 计算可能无法收敛到真正的基态。

尽管多智能体 workflow 在形式上通过了代码执行和语法验证 (workflow\\_status: "success"), 但其结构构建逻辑链彻底崩溃, 导致生成的 POSCAR 文件原子数和晶胞体积存

在灾难性偏差. 更深层次的失败在于, 它在 INCAR 文件中完全忽略了缺陷计算所需的关键电子结构参数 NELECT 和 ISPIN). 这证明了当前智能体在没有人工干预的情况下处理需要精确几何控制和高阶物理知识转化的复杂计算任务时, 缺乏科学合理性与鲁棒性。

## 参考文献

- [1] Xiao X, Xu N S, Tian X Y, Zhang T K, Wang B Z, Wang X M, Xian Y M, Lu C Y, Ou X Y, Yan Y F, Sun L C, You F Q, Gao F 2025 *Nature* **638** 670
- [2] de Leon N P, Itoh K M, Kim D, Mehta K K, Northup T E, Paik H, Palmer B S, Samarth N, Sangtawesin S, Steuermand W 2021 *Science* **372** eabb2823
- [3] Lin S, Liu J, Liu X, Zhao X 2019 *Proc. Natl. Acad. Sci. U. S. A.* **116** 10244

- [4] Zhang X S, Chen Y J, Hu J L 2018 *Prog. Aerosp. Sci.* **97** 22
- [5] Wang C Y, Xia K L, Wang H M, Liang X P, Yin Z, Zhang Y Y 2019 *Adv. Mater.* **31** 1801072
- [6] Shen S C, Khare E, Lee N A, Saad M K, Kaplan D L, Buehler M J 2023 *Chem. Rev.* **123** 2242
- [7] Miracle D B, Miller J D, Senkov O N, Woodward C, Uchic M D, Tiley J 2014 *Entropy* **16** 494
- [8] Kresse G, Furthmüller J 1996 *Phys. Rev. B* **54** 11169
- [9] Jain A, Ong S P, Hautier G, Chen W, Richards W D, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G, Persson K A 2013 *APL Mater.* **1** 011002
- [10] Buehler M J, Hartmaier A, Gao H, Duchaineau M, Abraham F F 2004 *Comput. Methods Appl. Mech. Eng.* **193** 5257
- [11] Taheri Mousavi S M, Zhou H, Zou G, Gao H 2019 *npj Comput. Mater.* **5** 2
- [12] Liu J X, Zhu T N, Ye C Y, Fang Z, Weng H M, Wu Q S 2025 *Chin. Phys. B* **34** 117106
- [13] Schmidt J, Cerqueira T F T, Romero A H, Loew A, Jäger F, Wang H C, Botti S, Marques M A L 2024 *Mater. Today Phys.* **48** 101560
- [14] Merchant A, Batzner S, Schoenholz S S, Aykol M, Cheon G, Cubuk E D 2023 *Nature* **624** 80
- [15] Barroso-Luque L, Shuaibi M, Fu X, Wood B M, Dzamba M, Gao M, Rizvi A, Zitnick C L, Ulissi Z W 2024 *arXiv*: 2410.12771
- [16] Zhu C K, Mo Y, Chen Y A, Wang X 2024 *Phys. Rev. Lett.* **133** 030801
- [17] Chang Y J, Zhang J H, Lu Y H, Yang Y Y, Mei F, Ma J, Jia S T, Jin X M 2025 *Phys. Rev. Lett.* **134** 053601
- [18] Yang A, Li A F, Yang B S, Zhang B C, Hui B Y, Zheng B, Yu B W, Gao C, Huang C G, Lv C X, Zheng C J, Liu D H, Zhou F, Huang F, Hu F, Ge H, Wei H R, Lin H, Tang J L, Yang J, Tu J H, Zhang J W, Yang J X, Yang J X, Zhou J, Zhou J R, Lin J Y, Dang K, Bao K Q, Yang K X, Yu L, Deng L H, Li M, Xue M F, Li M Z, Zhang P, Wang P, Zhu Q, Men R, Gao R Z, Liu S X, Luo S, Li T H, Tang T Y, Yin W B, Ren X Z, Wang X Y, Zhang X Y, Ren X C, Fan Y, Su Y, Zhang Y C, Zhang Y E, Wan Y, Liu Y Q, Wang Z K, Cui Z Y, Zhang Z R, Zhou Z P, Qiu Z H 2025 *arXiv*: 2505.09388 [cs. CL]
- [19] Gemma Team 2025 *arXiv*: 2503.19786 [cs. CL]
- [20] OpenAI 2023 *arXiv*: 2303.08774 [cs. CL]
- [21] DeepSeek-AI 2024 *arXiv*: 2412.19437 [cs. CL]
- [22] Pandya K, Holia M 2023 *arXiv*: 2310.05421 [cs. CL]
- [23] Liu B, Li X F, Zhang J Y, Wang J L, He T J, Hong S R, Liu H Z, Zhang S K, Song K T, Zhu K L, Cheng Y H, Wang S Y, Wang X Q, Luo Y Y, Jin H B, Zhang P Y, Liu O, Chen J Q, Zhang H, Yu Z Y, Shi H C, Li B Y, Wu D K, Teng F W, Jia X J, Xu J W, Xiang J Y, Lin Y Z, Liu T M, Liu T L, Su Y, Sun H, Berseth G, Nie J Y, Foster I, Ward L, Wu Q Y, Gu Y, Zhuge M C, Liang X B, Tang X R, Wang H H, You J X, Wang C, Pei J, Yang Q, Qi X L, Wu C L 2025 *arXiv*: 2504.01990 [cs. AI]
- [24] Wang X D, Chen Z R, Guo P J, Gao Z F, Mu C, Lu Z Y 2025 *arXiv*: 2507.16307
- [25] Mai X J, Xu H T, Li Z Z, Wang X, Wang W N, Hu J, Zhang Y Y, Zhang W Q 2025 *arXiv*: 2505.07773 [cs. AI]
- [26] Xue Z H, Zheng L T, Liu Q, Li Y R, Zheng X S, Ma Z J, An B 2025 *arXiv*: 2509.02479 [cs. LG]
- [27] Feng J Z, Huang S J, Qu X W, Zhang G, Qin Y J, Zhong B Q, Jiang C Q, Chi J X, Zhong W J 2025 *arXiv*: 2504.11536 [cs. CL]
- [28] Li X F, Zou H Y, Liu P F 2025 *arXiv*: 2503.23383 [cs. CL]
- [29] Chen M Y, Sun L Z, Li T P, Sun H Z, Zhou Y J, Zhu C Z, Wang H F, Pan J Z, Zhang W, Chen H J, Yang F, Zhou Z N, Chen W P 2025 *arXiv*: 2503.19470 [cs. AI]
- [30] Jin B W, Zeng H S, Yue Z R, Yoon J, Arik S, Wang D, Zamani H, Han J W 2025 *arXiv*: 2503.09516 [cs. CL] [31] Song H T, Jiang J H, Min Y Q, Chen J, Chen Z P, Zhao W X, Fang L, Wen J R 2025 *arXiv*: 2503.05592 [cs. AI]
- [31] Song H, Jiang J, Min Y, Chen J, Chen Z, Zhao W X, Fang L, Wen J R 2025 *arXiv*: 2503.05592 [cs. AI]
- [32] Sun H, Qiao Z L, Guo J Y, Fan X B, Hou Y Y, Jiang Y, Xie P J, Zhang Y, Huang F, Zhou J R 2025 *arXiv*: 2505.04588 [cs. CL]
- [33] Roters F, Eisenlohr P, Hantcherli L, Tjahjanto D D, Bieler T R, Raabe D 2010 *Acta Mater.* **58** 1152
- [34] Fish J, Wagner G J, Keten S 2021 *Nat. Mater.* **20** 774
- [35] de Pablo J J, Curtin W A 2007 *MRS Bull.* **32** 905
- [36] Wei Q, Zeng S L, Li C L, Brown W, Frunza O, Deng W, Schneider A, Nevmiyaka Y, Zhao Y K, Garcia A, Hong M Y 2025 *arXiv*: 2505.11821 [cs. LG]
- [37] Wang Z L, Zheng X H, An K, Ouyang C J, Cai J L, Wang Y H, Wu Y C 2025 *arXiv*: 2505.15107 [cs. CL]
- [38] Wang H R, Qian C, Zhong W J, Chen X S, Qiu J H, Huang S J, Jin B W, Wang M D, Wong K F, Ji H 2025 *arXiv*: 2504.14870 [cs. AI]
- [39] Qian C, Acikgoz E C, Wang H R, Chen X S, Sil A, Hakkani-Tür D, Tur G, Ji H 2025 *arXiv*: 2502.11435 [cs. AI]
- [40] Hong S R, Zhuge M C, Chen J Q, Zheng X W, Cheng Y H, Zhang C Y, Wang J L, Wang Z L, Yau S K S, Lin Z J, Zhou L Y, Ran C Y, Xiao L F, Wu C L, Schmidhuber J 2023 *arXiv*: 2308.00352 [cs. AI]
- [41] Wu Q Y, Bansal G, Zhang J Y, Wu Y R, Li B B, Zhu E K, Jiang L, Zhang X Y, Zhang S K, Liu J L, Awadallah A H, White R W, Burger D, Wang C 2023 *arXiv*: 2308.08155 [cs. AI]
- [42] Moura J <https://www.crewai.com> [2025-10-12]
- [43] Li Z F, Zhang H X, Han S J, Liu S, Xie J W, Zhang Y, Choi Y J, Zou J, Lu P 2025 *arXiv*: 2510.05592 [cs. AI]
- [44] Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, Yogatama D, Bosma M, Zhou D, Metzler D, Chi E H, Hashimoto T, Vinyals O, Liang P, Dean J, Fedus W 2022 *arXiv*: 2206.07682 [cs. CL]
- [45] Chang Y P, Wang X, Wang J D, Wu Y, Yang L Y, Zhu K J, Chen H, Yi X Y, Wang C X, Wang Y D, Ye W, Zhang Y, Chang Y, Yu P S, Yang Q, Xie X 2024 *ACM Trans. Intell. Syst. Technol.* **15** 1
- [46] de Jong M, Chen W, Angsten T, Jain A, Notestine R, Gamst A, Sluiter M, Ande C K, van der Zwaag S, Plata J J, Toher C, Curtarolo S, Ceder G, Persson K A, Asta M 2015 *Sci. Data* **2** 150053
- [47] Zheng T S, Deng Z Y, Tsang H T, Wang W Q, Bai J X, Wang Z H, Song Y Q 2025 *arXiv*: 2505.13259 [cs. CL]
- [48] Ghafarollahi A, Buehler M J 2024 *arXiv*: 2407.10022 [cs. AI]
- [49] Thompson A P, Aktulga H M, Berger R, Bolintineanu D S, Brown W M, Crozier P S, in 't Veld P J, Kohlmeyer A, Moore S G, Nguyen T D, Shan R, Plimpton S J 2022 *Comput. Phys. Commun.* **271** 108171
- [50] Zhang B H, Li X, Xu H C, Jin Z, Wu Q S, Li C 2025 *arXiv*: 2507.04053 [cond-mat. mtrl-sci]
- [51] Ni Z Q, Li Y H, Hu K J, Han K Y, Xu M, Chen X Y, Liu F Q, Ye Y C, Bai S X 2024 *arXiv*: 2411.08063 [physics. soc-ph]
- [52] Song T, Luo M, Zhang X L, Chen L J, Huang Y, Cao J Q, Zhu Q, Liu D B, Zhang B C, Zou G, Zhang G Q, Zhang F, Shang W W, Fu Y, Jiang J, Luo Y 2025 *J. Am. Chem. Soc.*

- 147 12534
- [53] Zou Y H, Cheng A H, Aldossary A, Bai J R, Leong S X, Campos-Gonzalez-Angulo J A, Choi C H, Ser C T, Tom G, Wang A, Zhang Z J, Yakavets I, Hao H, Crebolder C, Bernales V, Aspuru-Guzik A 2025 *Matter* **8** 102263
- [54] Wang Z Q, Huang H S, Zhao H C, Xu C W, Zhu S, Janssen J, Viswanathan V 2025 arXiv: 2507.14267 [cs. AI]
- [55] Yao L, Samantray S, Ghosh A, Roccapriore K, Kovarik L, Allec S, Ziatdinov M 2025 arXiv: 2508.06569 [cs. AI]
- [56] Han X Q, Wang X D, Xu M Y, Feng Z, Yao B W, Guo P J, Gao Z F, Lu Z Y 2025 *Chin. Phys. Lett.* **42** 027403
- [57] Model Context Protocol <https://modelcontextprotocol.io> [2025-11-05]
- [58] Han X Q, Gao Z F, Guo P J, Lu Z Y 2025 ResearchGate 10.13140/RG.2.2.36573.12009 [physics.comp-ph]
- [59] Giannozzi P, Baroni S, Bonini N, et al. 2009 *J. Phys.: Condens. Matter* **21** 395502
- [60] Giannozzi P, Andreussi O, Brumme T, et al. 2017 *J. Phys.: Condens. Matter* **29** 465901
- [61] Clark S J, Segall M D, Pickard C J, Hasnip P J, Probert M I J, Refson K, Payne M C 2005 *Z. Kristallogr.* **220** 567
- [62] Wang V, Xu N, Liu J C, Tang G, Geng W T 2021 *Comput. Phys. Commun.* **267** 108033
- [63] Larsen A H, Mortensen J J, Blomqvist J, et al. 2017 *J. Phys.: Condens. Matter* **29** 273002
- [64] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A, Ceder G 2013 *Comput. Mater. Sci.* **68** 314
- [65] Pallets <https://flask.palletsprojects.com> [2025-11-05]
- [66] Ceder G, Persson K 2010 *The Materials Project: A Materials Genome Approach* (Washington, DC: U. S. Department of Energy)
- [67] Grimme S 2006 *J. Comput. Chem.* **27** 1787
- [68] Grimme S, Antony J, Ehrlich S, Krieg H 2010 *J. Chem. Phys.* **132** 154104
- [69] Grimme S, Ehrlich S, Goerigk L 2011 *J. Comput. Chem.* **32** 1456
- [70] Sabatini R, Gorni T, de Gironcoli S 2013 *Phys. Rev. B* **87** 041108
- [71] Rao C P, Ren P, Wang Q, Buyukozturk O, Sun H, Liu Y 2023 *Nat. Mach. Intell.* **5** 765
- [72] Li Z Y, Han W, Zhang Y, Fu Q F, Li J X, Qin L Z, Dong R Y, Sun H, Deng Y, Yang L J 2024 *Nat. Mach. Intell.* **6** 1566
- [73] Zhang D, Liu X Z, Zhang X Y, Zhang C Q, Cai C, Bi H R, Du Y M, Qin X J, Peng A Y, Huang J M, Li B W, Shan Y F, Zeng J Z, Zhang Y Z, Liu S Y, Li Y F, Chang J H, Wang X Y, Zhou S, Liu J C, Luo X S, Wang Z Y, Jiang W R, Wu J, Yang Y D, Yang J Y, Yang M Y, Gong F Q, Zhang L S, Shi M C, Dai F Z, York D M, Liu S, Zhu T, Zhong Z C, Lv J, Cheng J, Jia W L, Chen M, Ke G L, E W N, Zhang L F, Wang H 2024 *npj Comput. Mater.* **10** 293
- [74] Austin J, Odena A, Nye M, Bosma M, Michalewski H, Dohan D, Jiang E, Cai C, Terry M, Le Q, Sutton C 2021 arXiv: 2108.07732 [cs. PL]
- [75] Elazar Y, Kassner N, Ravfogel S, Ravichander A, Hovy E, Schütze H, Goldberg Y 2021 *Trans. Assoc. Comput. Linguist.* **9** 1012
- [76] Fernando C, Banarse D, Michalewski H, Osindero S, Rocktäschel T 2023 arXiv: 2309.16797 [cs. CL]
- [77] Schmidhuber J 2015 arXiv: 1511.09249 [cs. AI]
- [78] Li G, Hammoud H, Itani H, Khizbullin D, Ghanem B 2023 *Adv. Neural Inf. Process. Syst.* **36** 51991
- [79] Park J S, O'Brien J C, Cai C J, Morris M R, Liang P, Bernstein M S 2023 arXiv: 2304.03442 [cs. HC]
- [80] Qian C, Cong X, Yang C, Chen W, Su Y, Xu J, Liu Z, Sun M 2023 arXiv: 2307.07924 [cs. SE]
- [81] Zhao X F, Li M D, Weber C, Hafez M B, Wermter S 2023 arXiv: 2303.08268 [cs. RO]
- [82] Yao S Y, Zhao J F, Yu D, Du N, Shafran I, Narasimhan K R, Cao Y 2022 arXiv: 2210.03629 [cs. CL]
- [83] Dong Y, Jiang X, Jin Z, Li G 2024 *ACM Trans. Softw. Eng. Methodol.* **33** 1
- [84] Chen M 2021 arXiv: 2107.03374 [cs. LG]
- [85] Significant Gravitass <https://github.com/Significant-Gravitass/AutoGPT> [2025-11-05]
- [86] Chase H <https://github.com/langchain-ai/langchain> [2025-11-05]
- [87] Plimpton S 1995 *J. Comput. Phys.* **117** 1
- [88] Shahriari B, Swersky K, Wang Z, Adams R P, de Freitas N 2015 *Proc. IEEE* **104** 148
- [89] Raissi M, Perdikaris P, Karniadakis G E 2019 *J. Comput. Phys.* **378** 686
- [90] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang Y J, Madotto A, Fung P 2023 *ACM Comput. Surv.* **55** 1
- [91] Jiménez-Luna J, Grisoni F, Schneider G 2020 *Nat. Mach. Intell.* **2** 573
- [92] Mnih V, Kavukcuoglu K, Silver D, et al. 2015 *Nature* **518** 529
- [93] Boiko D A, MacKnight R, Kline B, Gomes G 2023 *Nature* **624** 570
- [94] Jaderberg M, Czarnecki W M, Dunning I, Marris L, Lever G, Garcia Castañeda A, Beattie C, Rabinowitz N C, Morcos A S, Ruderman A, Sonnerat N, Green T, Deason L, Leibo J Z, Silver D, Hassabis D, Kavukcuoglu K, Graepel T 2019 *Science* **364** 859
- [95] Degraeve J, Felici F, Buchli J, et al. 2022 *Nature* **602** 414
- [96] Tadmor E B, Miller R E 2011 *Modeling Materials: Continuum, Atomistic and Multiscale Techniques* (Cambridge: Cambridge University Press) p500
- [97] Tadmor E B, Ortiz M, Phillips R 1996 *Philos. Mag. A* **73** 1529
- [98] Pizzi G, Cepellotti A, Sabatini R, Marzari N, Kozinsky B 2016 *Comput. Mater. Sci.* **111** 218
- [99] Jain A, Ong S P, Chen W, Medasani B, Qu X, Kocher M, Brafman M, Petretto G, Rignanese G M, Hautier G, Gunter D, Persson K A 2015 *Concurr. Comput. Pract. Exp.* **27** 5037
- [100] Henkelman G, Jónsson H 1999 *J. Chem. Phys.* **111** 7010
- [101] Zhang J, Lei Y K, Zhang Z, Han X, Li M D, Yang L J, Yang Y I, Gao Y Q 2021 *Phys. Chem. Chem. Phys.* **23** 6888
- [102] Kim S, Woo J, Kim W Y 2024 *Nat. Commun.* **15** 341
- [103] Sanchez-Lengeling B, Aspuru-Guzik A 2018 *Science* **361** 360
- [104] Xie T, Fu X, Ganea O E, Barzilay R, Jaakkola T 2021 arXiv: 2110.06197 [physics.chem-ph]
- [105] Burger B, Maffettone P M, Gusev V V, Aitchison C M, Bai Y, Wang X Y, Li X B, Alston B M, Li B Y, Clowes R, Rankin N, Harris B, Sprick R S, Cooper A I 2020 *Nature* **583** 237
- [106] Szymanski N J, Rendy B, Fei Y, Kumar R E, He T, Milsted D, McDermott M J, Gallant M, Cubuk E D, Merchant A, Kim H, Jain A, Bartel C J, Persson K, Zeng Y, Ceder G 2023 *Nature* **624** 86
- [107] Behler J, Parrinello M 2007 *Phys. Rev. Lett.* **98** 146401
- [108] Kirchdoerfer T, Ortiz M 2016 *Comput. Methods Appl. Mech. Eng.* **304** 81
- [109] Ebrahimi B, Bataleblu A A, Roshanian J 2024 *Expert Syst. Appl.* **248** 123363

- [110] Agrawal A, Choudhary A 2016 *Apl Mater.* **4** 053208
- [111] Beltagy I, Lo K, Cohan A 2019 arXiv: 1903.10676 [cs. CL]
- [112] Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z Q, Kononova O, Persson K A, Ceder G, Jain A 2019 *Nature* **571** 95
- [113] Raccuglia P, Elbert K C, Adler P D F, Falk C, Wenny M B, Mollo A, Zeller M, Friedler S A, Schrier J, Norquist A J 2016 *Nature* **533** 73
- [114] Li D, Lee K, Wang B Y, Osada M, Crossley S, Lee H R, Cui Y, Hikita Y, Hwang H Y 2020 *Nature* **581** 392
- [115] Qiu J, Zhu L W, Feng Z, Luo Z Z, Wang L L 2025 *Chem. Commun.* **61** 10026
- [116] Alldritt B, Hapala P, Oinonen N, Urtev F, Krejci O, Canova F F, Kannala J, Schulz F, Liljeroth P, Foster A S 2020 *Sci. Adv.* **6** eaay6913
- [117] Kusne A G, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, Sarker S, Oses C, Toher C, Curtarolo S, Davydov A V, Agarwal R, Bendersky L A, Li M, Mehta A, Takeuchi I 2020 *Nat. Commun.* **11** 5966
- [118] Nord M, Vullum P E, MacLaren I, Tybell T, Holmestad R 2017 *Adv. Struct. Chem. Imag.* **3** 9
- [119] Lookman T, Balachandran P V, Xue D, Yuan R H 2019 *npj Comput. Mater.* **5** 21
- [120] Chen Z Y, Xie F K, Wan M, Yuan Y, Liu M, Wang Z G, Meng S, Wang Y G 2023 *Chin. Phys. B* **32** 118104
- [121] Wu S Y, Li H 2025 *Acta Phys. Sin.* **74** 160701 (in Chinese) [吴思远, 李泓 2025 物理学报 **74** 160701]
- [122] Zhu T N, Fang Z, Wu Q S, Weng H M 2025 *Chin. Phys. B* **34** 120702
- [123] Huang Y D, Xia W J, Du J M, Jiang Y, Wang X, Chen Y Z, Wang H Y, Zhao J J, Guo C S 2025 *Acta Phys. Sin.* **74** 188101 (in Chinese) [黄钰丹, 夏琬钧, 杜俊梅, 蒋渝, 汪鑫, 陈元正, 王红艳, 赵纪军, 郭春生 2025 物理学报 **74** 188101]
- [124] Liu Z H, Chen X Y, Zhou Q H, Wang J L 2025 *Acta Phys. Sin.* **74** 240701 (in Chinese) [刘章赫, 陈新宇, 周登桦, 王金兰 2025 物理学报 **74** 240701]

SPECIAL TOPIC — AI + Physical Science

## Applications of multi-agent systems in computational materials science\*

WU Ying LU Zhongyi GAO Zefeng<sup>†</sup>

(Department of Physics, Renmin University of China, Beijing 100872, China)

(Received 8 December 2025; revised manuscript received 13 January 2026)

### Abstract

The complexity of modern materials research poses significant challenges to traditional “trial-and-error” methods and fragmented computational simulations, often resulting in low efficiency and high resource consumption. The emergence of artificial intelligence, especially large language models (LLMs), has brought transformative opportunities to computational materials science. Among these, multi-agent systems (MASs), which simulate the collaborative dynamics of human research teams, demonstrate exceptional promise for managing complex, multi-step scientific tasks.

This paper systematically reviews the applications of MAS in physical materials computation. We analyze how specialized MAS frameworks can achieve end-to-end automation of complex workflows, from initial structure modeling and parameter configuration to final simulation execution and result analysis. A detailed focus is placed on two pioneering systems: VASPIlot and PhysAgent. The VASPIlot is built on the CrewAI framework, and specifically designed to automate the entire workflow of the widely-used VASP (Vienna *ab initio* Simulation Package) software. It deploys a team of specialized agents (such as structural, calculational, validation agents) to streamline tasks such as band structure calculations, lattice optimization, and convergence testing, greatly reducing manual intervention and improving reproducibility. PhysAgent adopts a more ambitious “Mentor-Student-Leader” cognitive architecture, aiming not only at task execution but also at

\* Project supported by the Natural Science Foundation of Beijing, China (Grant No. Z250005) and the National Natural Science Foundation of China (Grant No. 62476278).

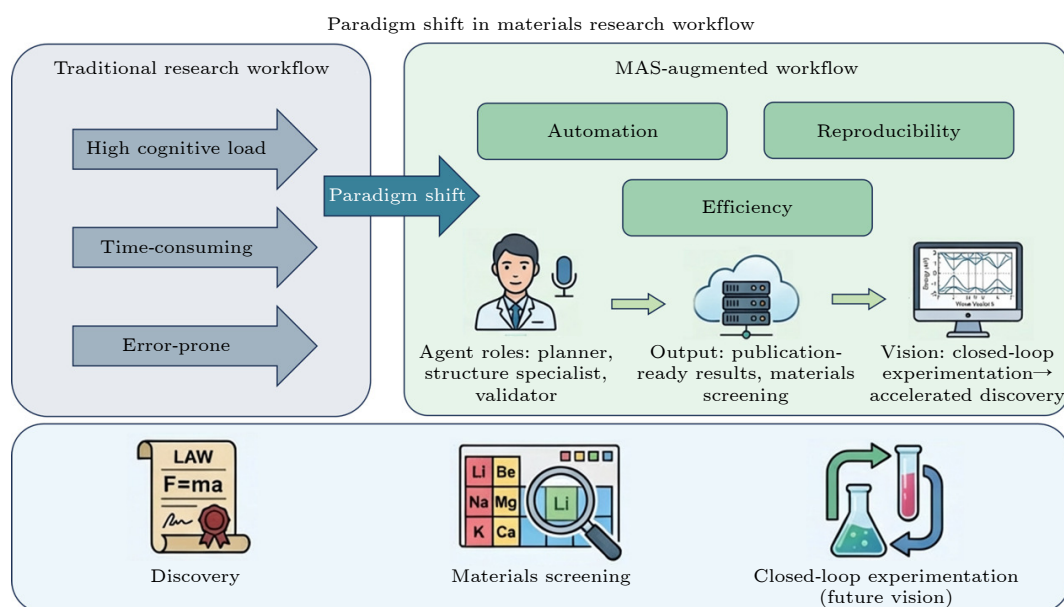
<sup>†</sup> Corresponding author. E-mail: zfgao@ruc.edu.cn

autonomous scientific discovery. It has demonstrated the ability to autonomously rediscover fundamental physical laws from simulated observational data, such as Kepler's laws of planetary motion and Newton's second law. Furthermore, PhysAgent successfully automates complex workflows such as the first-principles electronic structure calculations of GaAs and the simulations of complex phenomena such as raindrop flow on high-speed train windows.

Our analysis demonstrates that the MAS can substantially enhance computational efficiency, standardize procedures, and handle complex, non-routine problems beyond the scope of single-agent systems. The Key physical results and capabilities highlighted include the automated convergence testing of plane-wave cutoff energy (ENCUT) and k-point grids, comparative studies of van der Waals correction methods on material properties, high-throughput screening of transition metal dichalcogenides, and the autonomous derivation of physical laws from data.

However, through critical evaluation and practical case studies (such as using the SciLink framework for complex surface adsorption and defect calculations), we identify persisting limitations. These limitations include challenges in computational resource allocation, ensuring physical consistency in agent decisions, handling "parameter gaps" in multi-scale simulations, managing long-time-scale tasks, and a general lack of true scientific creativity for unsolved problems. The "black-box" nature of agent reasoning also hinders deep scientific trust.

Looking forward, we envision MAS evolving from efficient assistants to genuine scientific partners. This evolution will require 1) deeper integration of physical principles and constraints into agent reasoning, 2) the development of closed-loop research ecosystems including "hypothesis-design-calculation-experiment" cycles, potentially leading to self-driving laboratories, 3) expansion into frontier areas like catalytic reaction exploration, inverse materials design, and automated cross-scale modeling, and 4) fostering new researcher skills, with focusing on strategic problem formulation and critical analysis of AI-generated outputs. By addressing current challenges and leveraging these future directions, MAS hold the potential to fundamentally reshape the research paradigm in computational materials science and physics, accelerating the discovery and understanding of novel materials.



**Keywords:** computational materials science, multi-agent systems, automated scientific discovery

**DOI:** [10.7498/aps.75.20251687](https://doi.org/10.7498/aps.75.20251687)

**CSTR:** [32037.14.aps.75.20251687](https://cstr.net.cn/detail/32037.14.aps.75.20251687)



## 多智能体在物理材料计算领域的应用

吴颖 卢仲毅 高泽峰

### Applications of multi-agent systems in computational materials science

WU Ying LU Zhongyi GAO Zefeng

引用信息 Citation: *Acta Physica Sinica*, 75, 050801 (2026) DOI: 10.7498/aps.75.20251687

CSTR: 32037.14.aps.75.20251687

在线阅读 View online: <https://doi.org/10.7498/aps.75.20251687>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

大语言模型加速材料设计——从知识挖掘到智能设计的全链条赋能

Material design accelerated by large language models: end-to-end empowerment from knowledge mining to intelligent design

物理学报. 2025, 74(18): 188101 <https://doi.org/10.7498/aps.74.20250497>

高质量的材料科学文本挖掘数据集构建方法

A high-quality dataset construction method for text mining in materials science

物理学报. 2023, 72(7): 070701 <https://doi.org/10.7498/aps.72.20222316>

热智能材料及其在空间热控中的应用

Thermal smart materials and their applications in space thermal control system

物理学报. 2022, 71(1): 014401 <https://doi.org/10.7498/aps.71.20211889>

基于二维材料光电器件的传感器内计算与应用进展

Progress in in-sensor computing and applications based on photodetectors of two-dimensional materials

物理学报. 2025, 74(22): 228501 <https://doi.org/10.7498/aps.74.20251093>

高通量计算二维材料界面摩擦

High-throughput calculation of interfacial friction of two-dimensional material

物理学报. 2023, 72(2): 026801 <https://doi.org/10.7498/aps.72.20221676>

“机器微纳光学科学家”：人工智能在微纳光学设计的应用与发展

“Machine micro/nano optics scientist” : Application and development of artificial intelligence in micro/nano optical design

物理学报. 2023, 72(11): 114204 <https://doi.org/10.7498/aps.72.20230208>