



《计算机系统结构》研究性学习

高性能计算机的最新发展趋势

目录

1 摘要	1
2 正文	1
3 个人观点	3
4 结论	3
参考文献	4

1 摘要

高性能计算（HPC）正从传统的并行计算向 AI、量子计算与绿色技术深度融合的新范式转型。全球范围内，云原生技术推动 HPC 服务普惠化，显著降低科研与产业算力门槛。AI 与 HPC 协同重塑科学研究的边界的同时，量子计算与 HPC 的融合路径逐渐清晰，异构架构分阶段推进，为各个领域提供颠覆性的算力支持。硬件革新聚焦高能效与绿色化。存算分离架构、光子芯片、液冷技术等突破显著降低能耗，推动数据中心能效比持续优化。边缘计算与云边协同架构的成熟，扩展了 HPC 应用边界。可持续性成为未来 HPC 发展核心命题。

HPC 正迈向智能化、低碳化、泛在化的新阶段，通过异构算力整合、算法-硬件协同优化及云边端三级网络，为科学突破与产业革新提供强大引擎，引领计算技术向更高能效、更广领域迈进。

2 正文

高性能计算（High Performance Computing, HPC）通过并行计算、分布式架构与超算系统实现复杂任务的高效执行，其算力规模已突破百亿亿次（Exascale）级别。HPC 作为国家战略高科技技术，是解决国家安全（如核武器模拟）、经济建设（如金融风险建模）、社会发展等一系列重大挑战性问题的重要手段，已经成为信息时代世界高科技竞争和大国技术竞争的焦点。

全球范围内，高性能计算集群的大规模新建和升级，加上云原生技术的推广，使得算力服务变得更加充足且易于获取，阿里云、AWS 等推出 HPC 云服务，支持用户按需调用数万核并行资源，微软 Azure HPC-AI 集群提供 ND H100 v5 虚拟机，其混合精度算力较前代提升 6 倍，可支持千亿参数大模型训练。

在智能计算时代，AI for Science 不仅推动了基础科学研究进步，也为高性能计算与人工智能技术发展带来了新的机遇与挑战。科学计算模拟的需求正经历爆发式增长：微观粒子模拟、血流和癌细胞模拟，核聚变、气象以及地理空间模拟等等。传统数值模拟算法经过 AI 优化后性能得到了显著提升，例如深度学习加速 CFD 求解，如牛津大学团队采用张量网格（TNS）压缩湍流模拟维度，在单 CPU 上实现高维流场解析，计算效率提升 10 倍。大大提高了科研团队的生产效率，并推动科学研究进入快速进步的新阶段。

GPU 技术的发展对 HPC 与 AI 的深度融合起到了至关重要的推动作用。最初用于游戏和比特币挖矿的 GPU，如今已成为 ML/AI 计算的核心硬件，其强大的并行处理能力显著加速了 HPC 系统对复杂模拟任务的处理速度。NVIDIA H100 GPU 采用台积电 4nm 工艺，FP8 算力达 3,979 TFLOPS，能效比较 A100 提升 3 倍，已部署于全球 90% 的 AI 超算中心；华为昇腾 910B 通过近存计算设计实现 1 TFLOPS/W 能效，支撑国家级科研项目。量子计算与 HPC 的融合路径逐渐清晰：IBM Quantum System Two 与经典超算互联，金融组合优化效率提升 40%；芬兰 VTT 技术研究中心的混合云平台实现量子纠错周期毫秒级，材料模拟加速比达 100 倍；本源量子“悟源”超导量子计算机与曙光超算协同，物流路径规划效率提升 30 倍。未来基于 CPU+GPU+QPU 的 HPC 3.0 模型将分三阶段推进——网络互联、算力协同调度、深度系统集成，最终实现量子比特与经典比特的指令级交互。

高性能计算机的发展不仅依赖于高带宽内存和低延迟通信技术的突破，更受益于 AI 驱动的硬件革新。存算分离架构正在成为主流，服务器聚焦计算，专用数据处理器与存储系统各司其职。如 Lightmatter Enviser 光子芯片通过光互连实现 1ns 延迟，能效为传统 GPU 的 10 倍，应用于 Meta 数据中心推荐系统；DPU 卸载存储操作，释放服务器算力，配合 NVMe/CXL 协议使数据中心灵活扩展。同时，新一代冷却技术的推广，也为大幅节省能耗和提高系统效率提供了可能，谷歌全浸没式数据中心 PUE 值降至 1.1，较风冷节能 40%；中国曙光数创液冷系统在 30 个国家级算力中心部署，单机柜功耗密度突破 100kW。GPU 与软件优化的结合，使得 HPC 系统在处理复杂模拟任务时远超传统计算平台，进一步推动了科学计算和人工智能技术的融合。

云计算和大数据技术的不断进步，正在推动 HPC 融入云端和大数据平台，为用户提供更加高效、灵活和可靠的计算服务。高性能边缘计算（HPEC）也在逐步成长，其单点算力已达到千万亿次每秒的级别，典型应用场景包括自动驾驶和机器人。新兴的神经形态芯片和光电计算芯片等技术的成熟，不仅不断提升边缘算力和能效，还推动了数据中心与边缘计算的协同工作，为实时决策和信息安全提供了有力保障。

不可忽视的一点是，能源效率与可持续性问题的在超级计算机的发展中日益突出。随着超级计算机逐步向量子计算过渡，如何以更高效、更节能的方式处理复杂运算已成为关注焦点。当前全球数据中心年耗电量已达 650-1050 太瓦时，占全球电力需求的 3.5% 以上，而量子计算正成为破解能耗困局的关键技术：以 QuEra 的 256 量子比特中性原子量子计算机 Aquila 为例，其功耗仅 10kW，仅为传统超算的千分之一；中国玻色量子的光量子计算机整机功耗更低于 2 千瓦。未来，高性能计算机的设计、建设与运行将更加注重绿色环保理念，通过采用更高能效的软硬件、先进的冷却技术以及更合理的算力供需匹配，实现计算能效的整体提升，并将“量子效率”作为重要的评估指标。

3 个人观点

在当前大模型 AI 迅速发展的背景下，参数规模的飞速膨胀必将引发质的飞跃。大模型 AI 对高性能计算系统提出了极高要求，特别是在高算力芯片、高带宽互连芯片和存储芯片方面，这不仅要求硬件性能的不断突破，也催生了对系统整体协同效能的更高期待。与此同时，科学计算的深入应用和深度学习驱动的人工智能正在不断重新定义 HPC 的创新和应用边界。通过将 HPC 与 AI 深度融合，我们可以采用更加快速、便捷且成本更低的方式，最大化地利用现有 HPC 基础设施的优势，从而推动更多领域的应用创新，实现技术与产业的双重飞跃。

我认为，AI 与 HPC 的融合主要体现在算法创新、架构创新、平台管理软件以及数据中心创新这四个维度。然而，目前国内超算发展的现状显示，大多数超算项目仅停留在平台融合管理层面，甚至存在 AI、HPC 与大数据完全独立运行的情况，这样的分割模式难以释放两者深度融合所应有的潜力。只有打破现有的界限，挖掘 AI 与 HPC 更深层次的价值，才能真正实现强强联手，为各领域带来更广阔的创新空间和实际应用价值。

在算法创新方面，我认为这不仅仅是技术的提升，更是应用创新的核心。当前，随着 CPU 和 GPU 在发展过程中逐渐遇到能量墙问题，虽然 GPU 在并行计算方面表现出色，但其高功耗和复杂的使用难度也在一定程度上制约了整体性能的发挥。因此，AI 与 HPC 超算系统并非简单地依赖 CPU 和 GPU 的叠加，而需要在软件代码和算法优化上持续下功夫。只有在硬件和软件之间实现更紧密的协同，突破各自的瓶颈，才能充分发挥出两者结合所带来的综合优势，推动超算技术迈向更高效、更节能、更智能的新时代。

4 结论

综上所述，当前高性能计算机正处于一个从传统模拟向融合人工智能、量子计算及绿色环保技术转型的重要阶段。不断升级的硬件技术、日益成熟的软件优化以及云边协同的新型架构，将为未来科学研究和实际应用提供强大的算力支持，推动计算技术向更高能效、更低能耗和更广应用领域迈进。

参考文献

陈左宁. 我国高性能计算进展及国际高性能计算发展趋势

张云泉, 李根, 孙凝晖. 高性能计算技术演进与量子混合架构研究[J]. 计算机学报, 2023, 46(5): 1023-1035.

数字中国发展报告 2024 年

Yuval Boger, QuEra Computing. How Quantum Computing Could Reshape Energy Use in High-Performance Computing

NVIDIA. (2023). *NVIDIA H100 Tensor Core GPU Architecture White Paper* [Technical Report].

The Evolution, Convergence and Cooling of AI & HPC Gear. TOP 500