

计算机学院科研团队情况介绍表

团队名称	智能存算系统实验室		团队负责人	孙贤和
联系人	何水兵	Email	heshuibing@zju.edu.cn	电话 1898621 1726
团队主要成员				
姓名	职称		研究方向	联系方式
孙贤和	求是讲座教授, IEEE Fellow		智能计算、并行和分布式 处理、内存和 I/O 系统、 大数据应用程序的软件 系统及性能评估和优化 等	sun@iit.edu
何水兵	百人计划研究员， 博士生导师		智能计算、并行和分布 式计算、信息存储、非 易失性内存、存算一 体、操作系统等	heshuibing@zju.edu.cn
<p>团队负责人孙贤和教授是浙江大学“求是讲座”教授，美国伊利诺伊理工大学计算机科学系杰出教授，国际电气电子工程师协会会士（IEEE Fellow）。研究领域包括智能计算，数据密集型高性能计算，内存和 I/O 系统，大数据软件系统，以及计算机系统性能评估和优化。曾提出“Sun-Ni 定律”（可扩展计算内存限制加速模型），统一并扩展了著名的 Amdahl 定律与 Gustafson 定律。在相关国际会议和期刊上发表学术论文 250 余篇，担任 IEEE 并行和分布式计算系统汇刊（TPDS, CCF A）副主编，IEEE 计算机学会金核心成员（Golden Core Member），IEEE 可扩展计算技术委员会的前副主席，同时是并行处理领域多个顶级期刊编委会成员。孙贤和教授是中国科学院海外评审专家，中国科协海智专家。曾于 2018 年获得</p>				

中国计算机学会（CCF）颁发的海外杰出贡献奖。更多信息参见网站
www.cs.iit.edu/~sun/。

团队成员何水兵是浙江大学计算机学院“百人计划”研究员，博士生导师。2009 年获华中科技大学计算机科学与技术专业博士学位，2011 年至 2015 年在美国伊利诺伊理工大学从事博士后研究。长期从事计算机体系结构、智能计算、并行分布式计算、信息存储、存算一体等方面研究。在相关国际期刊和会议上发表论文 70 余篇，包括多篇 ASPLOS、TC、TPDS 等 CCF A 类论文，获 SPEED’ 08 最佳论文奖。现为中国计算机学会(CCF)会员、CCF 信息存储技术专业委员会委员、IEEE 会员、ACM 会员。担任 CCF A 类国际期刊 TPDS 编委(Associate Editor)，获 2020 年 TPDS 优秀编辑奖。作为负责人主持多项国家自然科学基金面上项目、科技创新 2030 重大项目子课题、教育部重点研究项目子课题和之江实验室重点项目，参与多项 973、863、中国自然科学基金和美国自然科学基金项目，与阿里巴巴等企业具有深度项目合作。更多信息参见网站
<https://person.zju.edu.cn/heshuibing>。

研究方向

智能存算系统实验室围绕智能计算、并行和分布式计算、信息存储、非易失性内存、存算一体、资源智能调度等方向展开研究生培养和创新研究。

目前承担的主要项目

一、新型异构智能计算系统

为满足大规模人工智能应用的算力需求，研发一套系统集成与计算技术研究平台。针对不同应用的需求，研究面向智能计算应用的应用感知技术，使得系统能进行更合理的适配；设计异构计算系统，满足 GPU、CPU 与多种 AI 芯片之间的协同计算；针对“存储墙”问题，研发多层次异构存储系统，设计存储 I/O 系统的性能优化方案；针对智能系统所面临的“通信墙”问题，研发分布式并行计算环境下的通信加速方案，进一步提升系统性能。

二、面向大规模深度学习的并行存储系统性能优化研究

随着模型和数据集的增长，大规模深度学习应用面临大量随机输入访问、中间数据倍增、检查点开销高等问题，使得基于磁盘和固态盘的存储系统成为其性能瓶颈。为此，研究在计算节点引入非易失性内存 NVM，并设计高效的存储管理方法，以提高存储系统性能。首先，研究基于 NVM 的深度学习数据输入方法，将深度学习访问可预知性引入 NVM 缓存管理，解决传统缓存方法在随机负载下的低效问题；其次，研究基于 NVM 的深度学习混合内存管理技术，融合多种方法协调管理内存，解决单一方法在克服内存不足时不能最大化系统性能问题；最后，研究基于 NVM 的深度学习检查点机制，引入选择性检查点策略，降低现有周期性检查点方法带来的 NVM 写开销。

三、内存存储层次结构的智能 I/O 缓冲系统

由于磁盘存储系统的性能提升速度要远落后于内存性能提升速度，

I/O 系统性能问题凸显起来。为了缩小这种差距，存储子系统飞速发展，即在内存和存储层次结构中出现了很多新型介质层。但是，随着内存层次结构越来越深，内存存储系统的数据移动越来越复杂，内存存储层次结构的潜力难以充分发挥。针对上述问题，研究一种智能 I/O 缓冲系统来提高整个内存存储系统性能，并扩展高性能 I/O 库与 HDF5 软件库，从而促进大数据和高性能计算应用的发展。

四、基于存算一体技术的人工智能模型与算法研究

针对当前冯·诺依曼计算架构对智能计算的性能制约问题，研究存储与计算深度融合的存内人工智能算法设计。基于新型存储器，如 3D 内存和忆阻器，设计并实现人工智能算法到存算一体芯片的映射和部署方案，并完善相应的系统支持。通过存储支持和多芯片的互联与并行化分解，解决大规模算法在存算一体芯片上的可扩展性问题。针对不同人工智能算法的特点，实现算法在传统体系结构和存算一体结构的自适应部署，进一步提高系统的能效比。

五、深度学习场景下异构池化资源管理和任务调度研究

深度学习技术的革新大大推动了计算机视觉，自然语言处理等领域的的发展，并在学术界和工业界都受到了极大的关注。随着技术的发展，深度学习应用对设备的计算、存储以及通信提出了更高的要求。然而，当前深度学习模型选择在现有数据中心单一服务器或分布式集群上进行训练或推理任务，应用只对本地资源可见，从而带来如下问题：(1) 资源无法被充分利用。(2) 资源的灵活性、容错性无法保证。(3) 异构设备如何部署成为难题。为此，研究深度学习场景下异构池化资源管理和任务调度算

法，以加速应用运行流程，更高效地使用池化资源。