

# 夏立斌

libin.xia@139.com | 13716959661

## EDUCATION

**硕博. 中国科学院大学** 计算机应用技术

北京 | 2018 - 2023

课题: 面向高能物理计算的并行内存计算技术研究 (分布式计算, 大数据技术)

课程: 高级计算机系统结构; 操作系统高级教程; 算法设计与分析; 形式化方法; 高级人工智能; 随机过程; 最优控制

**本科. 哈尔滨工业大学** 测控技术与仪器

哈尔滨 | 2014 - 2018

课题: 火星车低重力模拟中的三维位姿测量系统设计 (机器视觉)

课程: 模拟电子技术; 数字电子技术; 数字信号处理; 自动控制原理; 数字图像处理; 嵌入式系统

## WORK EXPERIENCE

**中国科学院大学** | 研究生/研究助理

北京 | 2019.9 - 2023.6

- 北京正负电子对撞机国家实验室 - 参与基于机器学习的插入件姿态多目标优化算法研究。
- 高能物理研究所计算中心 - 参与面向高能物理分波分析方法的内存计算关键技术研究。
- 高能物理研究所计算中心 - 参与基于内存计算的机器学习与 LQCD 并行计算关键技术研究。
- 高能同步辐射光源 - 参与下一代高能同步辐射光源的数据处理系统设计与研究。
- 中物院北京九所 - 参与高性能计算机运行时优化方法研究。

**哈尔滨工业大学** | 本科生

哈尔滨 | 2014.9 - 2018.6

- 超精密光电仪器工程研究所 - 作为组长完成了基于激光干涉的远程窃听器设计的项目。
- 机器智能与翻译研究室 - 学习机器翻译与人工智能相关理论技术。
- 自动化测试与控制系 - 参与中国空间技术研究院的火星车位姿运动视觉测量系统的研制部分工作。

## PROJECTS

**分布式内存计算系统 BLAZE (博士课题)**

SPARK, MPI, SCALA, C, JAVA

针对高能物理中高性能计算和大数据任务, 提供了统一的内存计算系统。通过 Spark 内存计算与 MPI 消息传递模型相结合, 解决了大数据计算 MapReduce 编程模型表达能力的缺陷, 并为 MPI 提供了面向数据的 DAG 计算方式。适用于大规模数据量环境下的科学计算与数据处理任务, 如迭代密集型计算或具有任务通信特征的算法与应用等。

**分布式内存管理系统 LATTICE (博士课题)**

DISTRIBUTED SYSTEM, NETWORKS, CONCURRENCY, C++

Lattice 为计算系统 Blaze 中的一个可独立运行的子系统, 提供分布式数据集在内存中的缓存功能, 及跨系统/语言间的 zero-copy 数据共享。利用上层计算系统的 data-locality 特性, 在节点内通过共享内存的方式实现, 并使用自定义的内存分配器对 MPI 进程通信进行了优化。该系统的具体实现基于 ASIO, GRPC。

**高性能计算机性能诊断方法研究 (九所实习)**

C, C++, LINUX

为满足高性能计算机上数值模拟程序的运行时效率提升需求, 提出了访存不连续问题、伪共享问题、冗余计算问题的诊断方法 FsDt、McDt, 利用性能问题监控、诊断及优化平台 JPerfCT 中。

**分布式异构存储系统 HTFS (合作参与)**

HDFS, JAVA, TAPE, LINUX

HTFS 被设计用于高能物理海量数据存储系统中。通过在分布式文件存储系统 HDFS 的基础上, 增加磁带层的存储支持, 通过统一的内存级元数据管理方式, 使得用户可以对异构存储系统中的文件进行透明访问。

**SUPER-LEVEL SET ESTIMATION 算法优化 (算法研究)**

PYTHON, MATLAB

研究了一类从含有噪声的数据中提取函数水平集的信号估计问题, 通过加入基于贝叶斯方法的高斯建模优化, 得到了更高的分类精度与更快的收敛速度。

**OPENROAD 芯片设计 (课程项目)**

OPENROAD, DOCKER, VERILOG

基于 OpenRoad 实现了 RTL-to-GDSII 的全流程芯片设计。在该项目中使用 Verilog 自主实现了一个具有五级流水线的 CPU, 同时对一个单核 32 位的小型 RISC-V 处理器核 (tinyriscv) 进行了从 RTL 到 GDS 的测试。

**基于单目视觉的位姿测量系统设计 (本科课题)**

C++, HALCON, LABVIEW, OPENCV

基于单目相机成像理论, 完成了图像特征点提取与位姿解算的算法设计及软件开发, 并最终部署于实时嵌入式系统中。项目重点研究了视觉采集与处理实时性的提高, 以及计算精度的优化以及稳定性的增强。

## SKILLS

Languages: C++, Java, Python, Scala, Go, Rust

Technology: MPI, Spark, HDFS, Git, Docker, Ray

## SELF INTRODUCTION

### 个人自述

本人出生于 1995 年 9 月，目前在中国科学院大学攻读计算机博士学位，本科毕业于哈尔滨工业大学自动化测试与控制系。

曾于北京正负电子对撞机国家实验室担任研究助理，参与基于机器学习的插入件姿态多目标优化算法研究，深入学习研究了光束线物理设计及人工智能算法在插入件控制中的应用。后进入高能物理研究所计算中心，进行计算技术及系统架构的研究，参与了“高能物理数据分析的 Hadoop/HBASE 平台研究”、“面向高能物理分波分析方法的内存计算关键技术研究”等国家自然科学基金项目。博士研究课题为面向高能物理的并行内存计算技术研究，主要贡献为：将高性能计算技术与互联网大数据技术结合，提出了一种基于 Spark 内存计算模型的消息传递方法，以及针对分布式计算框架的内存优化方法，为高能物理数值计算应用提供了一个新型的内存计算平台。目前以第一作者发表 EI/核心期刊两篇，在投一篇。

2022 年 7-8 月于中物院北京九所进行了为期两个月的暑假实习，研究课题为“计算、访存低效问题的细粒度诊断方法研究”，经过数十篇的文献调研，以及软件测试实验，总结提出了“FsDt、McDt”两种访存低效性诊断方法，并形成了 20 页的技术报告。在九所的实习过程中，经由武林平和景翠萍两位老师的悉心指导，使得本人从零开始学习到了高性能计算机的运行时优化方法，收获颇丰，并对该方向产生了浓厚的兴趣，希望可以作为一生奋斗奉献的方向。九所的两弹精神、国防使命、历史传承无不打动着这颗青年人的澎湃的内心，面对当前复杂的国际局势，作为国家花费十余年培养出来的博士研究生，更应该有担当、有能力、有信心解决难题。在实习过程中，深刻体会到了高性能计算机在国防领域数值模拟中拥有着无可替代的重要作用，并且九所有着强大的技术积累，和优秀的科研技术团队，这都吸引着我加入九所，渴望贡献出自己的力量。

## RESEARCH PROPOSAL

### 研究计划 1：高性能计算机运行时优化

在高性能计算机利用数值模拟方法进行物理实验及武器研制工作，已经是当前环境下的重要手段。但在实际的使用过程中，计算、存储、通信等诸多部件的利用率都处于很低的水平，不仅会拖慢实验的研究进展，同时也对硬件设备造成极大的资源浪费、经济损失以及能源消耗。因此，从系统和应用两个行为层面，对硬件利用率瓶颈进行分析诊断，找出限制运行效率的根本问题，并给出针对性的优化方案，这对于提升科学研究效率和节约经济和能源资源都具有重要意义。

在九所暑期实习的过程中，针对应用行为中的访存问题，对国内外的研究现状进行了调研，发现了诸多在技术上需要解决的问题。虽然在实习过程中提出了两种诊断方法，并在开源平台上进行了初步的实验，但仍然缺乏充分完备的实验验证，同时下一步还需要在自主研发的 JPerfCT 上进行实现，并为用户提供直观的诊断报告。同时对于系统行为的诊断，如系统噪声等问题也是运行时优化的重点，这些都需要进行更加深入的研究。与技术研究同样重要的是，JPerfCT、JSysCT 这样的工具产品，需要健壮稳定的工程实现，这些都是我在未来计划工作的方向。

### 研究计划 2：科学数据管理平台研究

随着高性能计算机技术的快速发展，以及对模拟实验要求的不断精细化，数据量规模的正面临着爆炸式增长的挑战。在九所的科研环境下，该问题也在不断凸显，海量数据的存储、管理、检索等都是亟需解决的问题。具体包含：多种存储介质及文件系统如何进行统一的视图化，文件元数据的管理方法，以及分级存储的调度策略等。当前开源社区虽然有诸多存储系统的解决方案，但往往都是针对于互联网数据、云计算或特定场景下的实现与优化。面对九所科学数据的特殊性，同样需要设计一套适用于九所的科学数据管理平台。

在高能物理研究所的计算中心，管理着 PB 级别的高能物理数据，并在未来会增长至 EB 级别。在我于计算中心学习的过程中，接触到了诸多并行/分布式文件系统，如 Lustre、HDFS、EOS 等。并参与了基于 HDFS 的分布式异构存储系统 HTFS 的研究工作。该工作包含了对 HDFS 增加了对磁带库的支持，并提供统一的元数据管理功能。相信在高能物理领域的存储系统研究经验，也会对九所科学数据管理平台的研究与建设有所帮助，未来也计划参与到这一方面的研究和开发工作中。