

秦善天

17863581726 | qinshantian23s@ict.ac.cn
shantianqin.github.io
2002-09 | 山东省聊城市



教育经历

中国科学院计算技术研究所

2023年09月 - 2026年06月

计算机科学与技术 硕士 处理器芯片全国重点实验室

北京

- 推免学硕，导师为范东睿研究员和李文明研究员
- 硕士期间一作发表CCF-A类论文1篇，CCF-C类论文2篇（其中1篇为TH-CPL B类），在投CCF-B类会议1篇
- 研究兴趣：数据流架构，软硬件协同设计，AI加速器设计，数字存算一体架构
- 平均GPA：3.94/4，曾获海光硕士生奖学金（2/200+，2024）

同济大学

2019年09月 - 2023年06月

电子科学与技术 本科 国家示范性微电子学院

上海

- 通过本科生导师制加入童美松教授（微电子学院副院长）课题组
- 本科期间一作论文被 ACES-China 2022 会议收录（清华大学重要国际学术会议 B类）
- 平均GPA：90/100，同济大学本科优秀学生一等奖学金（前5%，2021-2022）

学术成果

面向数据流驱动执行模式的软硬件协同设计方向论文：发表一作论文3篇，其中CCF A类 1篇，另有在审论文1篇

- S. Qin, Z. Fan, W. Li et al., "PANDA: Adaptive Prefetching and Decentralized Scheduling for Dataflow Architectures," in **ACM TACO, 2025**, 27 pages. (CCF A)
- S. Qin et al., "StreamDCIM: A Tile-Based Streaming Digital CIM Accelerator with Mixed-Stationary Cross-Forwarding Dataflow for Multimodal Transformer," in **IEEE ISCAS 2025**, 5 pages. (CCF C, TH-CPL B)
- S. Qin, W. Li, Z. Fan et al., "ROMA: A Reconfigurable On-Chip Memory Architecture for Multi-Core Accelerators," in **IEEE HPCC 2023**, Melbourne, Australia, 9 pages. (CCF C)
- S. Qin et al., "Towards Semi-Centralized Scheduling in Dataflow Architectures." (Euro-Par 在审, CCF B)

发明专利：

- 秦善天，李文明，范志华等. 一种数据流众核处理器的数据预取方法及处理器. CN2024102010162

科研经历

面向众核加速器的可重构片上存储架构

2022年12月 - 2023年12月

本科毕设

北京

- 课题概述：为了解决如何面向跨领域应用（规则与不规则访存占比差异大）实现高效数据访存的问题，同时不引入过多的额外开销，我们提出了一种应用自适应的数据预取机制和动态可重构的片上存储架构，显著降低了访存延迟和芯片面积开销
- 项目依托：国家重点研发计划（青年科学家项目），多层次融合的软件定义数据流关键技术与系统
- 负责工作：便签式存储器SPM与高速缓存Cache可复用架构的模拟器与RTL实现，VCS仿真，DC综合，PTPX分析功耗
- 本科毕业设计，相关成果已被应用于DPU系列12nm工艺异构融合处理器芯片DPU-F

基于相关成果已发表1篇CCF-C类会议论文（IEEE HPCC 2023），申请发明专利1项（已公开）

面向数据流众核架构的去中心化动态调度策略

2023年12月 - 2024年12月

主要负责人

北京

- 课题概述：针对传统的集中式调度在面向多并发任务时调度延迟高和资源利用率低的问题，我们提出了一种分布式的动态任务调度策略和去中心化的数据流架构，提高了硬件灵活性，实现了更好的负载均衡
- 项目依托：国家重点研发计划课题，异构数据流处理器系统的研制
- 负责工作：支持多APP运行时动态调度的模拟器优化设计与功能验证，case开发与性能测试，去中心化PE的RTL实现

基于相关成果已发表1篇CCF-A类期刊论文（ACM TACO），在投1篇CCF-B类会议论文（Euro-Par 2025）

面向数字存算架构的多模态Transformer数据流设计2024年02月 - 至今

主要负责人（合作者：何银涛，杜克大学陈怡然教授课题组博士后，曾获中科院院长特别奖）北京

- 课题概述：针对基于SRAM的数字CIM在面向多模态Transformer时重写延迟开销高的问题，我们提出了一种细粒度的Attention计算数据流与流水线，有效掩盖了CIM的高重写延迟
- 项目依托：北京市科技新星计划，高通量多模态数据流处理器芯片研究；华为合作项目，基带高通量运算阵列架构研究
- 负责工作：细粒度数据流与流水线设计的模拟器实现与性能测试；与何银涛老师展开合作，为进一步解决基于SRAM的数字CIM面向基于内容相关性的动态稀疏场景时利用率低的问题，我们提出了一种稀疏感知的动态调度优化设计

基于相关成果已发表1篇TH-CPL B类会议论文（IEEE ISCAS 2025），拓展工作拟投稿CCF-A类期刊论文（IEEE TCAD）

实习经历

AMD（超威半导体）2022年12月 - 2023年02月

FPGA软件开发北京

- 参与开发和维护面向Xilinx FPGA和Versal ACAP的开源加速库
- 使用Vitis HLS工具完成AES、ECDSA等算子的开发

中科睿芯2023年02月 - 至今

AI加速器芯片设计北京

- SPM与DMA总体维护及功能优化**：DPU的片上存储SPM模块的总体维护及功能优化，对原有数据传输模式进行维护，并结合对DMA模块的优化，实现DDR与SPM之间更高效的数据传输。**个人职责**：对原有普通/SIMD传输模式进行维护；在模拟器上新增INT8和FP16类型64×64的矩阵正常传输TRM和转置传输TRS的数据传输功能，并进行case验证与功能调试。
- 多APP动态调度**：为DPU添加动态任务调度的功能—实现不同APP可以根据Runtime的调度在PE阵列的不同区域执行，且能满足多种模式组合实时调度需求。**个人职责**：Runtime原型设计；合作完成模拟器API修改；单粒度case开发（4PE，8PE，16PE）。
- Tensor Unit计算部件优化**：对DPU的Tensor Unit进行优化—为了降低功耗（后端需求），对Tensor Unit的FP16计算部件进行了砍半，所以原有的指令与功能模块需要重新进行设计。**个人职责**：HMMA与HMMAL指令拓展（涵盖FP16类型的4×4、8×8至64×64矩阵运算）；汇编器指令添加；模拟器PE模块修改；模拟器case验证调试；RTL测试trace生成。

专业技能

- ASIC设计：了解Linux系统的基本使用；可以熟练使用VCS/DC等前端设计工具；有Innovus/PT等后端工具的使用经验
- 编码能力：Verilog/SystemVerilog HDL；C语言；了解Python的基本使用，可以编写简单脚本
- 外语水平：CET-6，具有良好的英语读写能力，能快速浏览英文文献、文档以及Debug信息
- 科研写作：可以熟练使用LaTex进行论文写作；可以熟练使用Visio进行科研绘图

荣誉/获奖

- 海光硕士生奖学金（2/200+，企业冠名奖学金），中科院计算所，2024
- 研究生学业奖学金一等奖，中国科学院大学，2023-2024
- 全国总决赛金奖（一等奖），第十三届“挑战杯”中国大学生创业计划竞赛，2022
- 全国总决赛银奖（二等奖），第七届中国国际“互联网+”大学生创新创业大赛，2021
- 本科优秀学生一等奖学金（前5%），同济大学，2021-2022

学生工作

- 学生科协，科普部部长，中科院计算所，2024至今
- 校学生会，执行委员兼任办公室部门负责人，同济大学，2021-2022