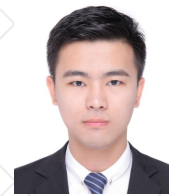


刁海康

15201736882 | diaohaikang@stu.pku.edu.cn | 北京

1997-12 | 男



教育经历

北京大学 2022年09月 - 2026年06月
集成电路科学与技术 博士 集成电路学院 北京

- 导师：王源教授（北京邮电大学副校长）、唐希源研究员（北大博雅青年学者）
- 研究方向：**高效AI芯片与系统设计**、数字存算一体、模型压缩与硬件加速系统设计
- 排名：**1/90 (2024-2025)**，**国家奖学金**，**北京市优秀毕业生**

复旦大学 2019年09月 - 2022年06月
电路与系统 硕士 信息科学与工程学院 上海

- 导师为陈炜教授，国家青年千人计划
- 研究方向：轻量化睡姿感知软硬件系统设计，**模型压缩与加速算法设计**

复旦大学 2015年09月 - 2019年06月
电子信息科学与技术 本科 信息科学与工程学院 上海

- 免试直研至复旦大学电子工程系电路与系统硕士研究生

科研经历

北京大学 PRIME 课题组 2022年07月 - 至今

围绕端侧高效AI芯片与系统设计，于**电路、系统、工具链**三个层面进行探索。主导设计了6款基于数字SRAM/eDRAM存算的AI芯片加速器，其中5颗已流片。全面负责AI算法压缩和部署（CNN、LLM、量化、稀疏、算子近似等）、芯片数字前后端设计、模拟前后端设计、数模混合设计，**成果发表于JSSC, CICC, ICCAD等顶刊顶会**：

1. 电路级——基于算法级近似的数字存算一体电路芯片：

- 基于近似计算的数字存算一体Macro电路设计：采用**无乘法算子来近似乘法算子**，设计专用的动态逻辑比较器电路，功耗降低3倍。
- 浮点数字存算一体电路设计：设计**一步计算浮点数据流**和定制化并行比较器电路、设计**模型无损结构化稀疏算法和架构**、可片上训练的存算一体芯片架构，大幅度提升端侧复杂的环境下的模型推理精度。
- 基于混合精度量化的数字存算一体芯片设计：设计**整套模型压缩框架**（算子重构、无损稀疏和混合精度量化）；设计基于**混合精度计算核电路、混合精度并行同步架构**的加速器芯片，最终实现4-5倍的性能提升。

2. 系统级——基于软硬件协同设计CIM加速器芯片：

- 端侧访存优化、多算子适配**的AI加速器芯片设计：对于端侧芯片面积受限的问题，设计**自适应数据流**降低片外DRAM读写访存，降低2.6倍实际端到端运行功耗；设计行/列共享电路结构支持端侧多种卷积算子；设计**卷积核稀疏算法**实现硬件友好型的稀疏加速。
- 端侧LLM的Prefill-decode统一**的AI加速器芯片设计：设计基于**动态token切分和全定点数据编码、计算的PD统一架构**；设计基于head切分的QKVRA融合数据流，提升带宽利用率；设计基于高密度低功耗免刷新的eDRAM-CIM，提升片上计算能效。

3. 工具链级——基于设计空间探索的CIM自动化工具：

- 针对数字存算一体芯片的手动设计流程复杂的难题，提出了**基于设计空间探索的数字存算一体编译器**（DCIM Compiler）：提出可综合的数字DCIM框架模板和开销模型，并采用遗传法定制DCIM架构和版图实现
- 设计**面向eDRAM CIM的设计空间探索工具**，结合新器件（IGZO）、3D封装等设计，在不同任务场景下的应用进行cell-阵列-架构的多层次设计空间探索，并为硬件设计提供指导。

复旦大学 智慧医疗电子中心课题组（CIME） 2019年02月 - 2022年06月

- 设计**基于压力感应床垫的智能睡姿识别系统**，全面负责硬件、算法、采集实验和数据分析所有流程，成果发表于IEEE TBioCAS。
- 提出了**基于频率选择模型的模型压缩算法**，通过在频率域训练出不同通道对于分类的敏感度，大量减少输入通道，最后实现模型超过100倍的压缩并部署在嵌入式处理器上实现实时的睡姿分类，成果发表于IEEE IoTJ（中科院一区）

实习经历

知存科技	2025年05月 - 2026年02月
架构实习生 架构部	北京
<ul style="list-style-type: none">设计面向eDRAM CIM的设计空间探索工具，结合新器件（IGZO）等设计进行cell-阵列-架构的多层次设计空间探索面向实际训推一体的demo任务，结合算法团队和电路团队，设计eDRAM CIM的芯片架构、模型量化算法以及误差补偿方式	
华为诺亚方舟实验室	2021年07月 - 2021年12月
AI算法实习生 计算机视觉实验室	北京
<ul style="list-style-type: none">加法神经网络的低精度量化：针对加法网络的特性使用聚类分组、权重截断等方法实现低精度量化，4bit量化仅存在1%的损失使用加法算子替代乘法算子，实现Adder-MLP，并采用知识蒸馏提升精度	
OPPO AI研究院	2021年03月 - 2021年06月
模型压缩与部署实习生 智能感知中心	上海
<ul style="list-style-type: none">基于MobileNet、EfficientNet的量化感知训练和静态量化基于高通SNPE、阿里巴巴MNN等边缘端部署平台，进行模型量化、部署和性能测试	

荣誉奖项

北京市优秀毕业生、北京大学优秀毕业生	2026.04
北京大学博士生国家奖学金	2025.09
北京大学优秀科研奖	2023.09
复旦大学优秀毕业生	2022.05

一作论文发表/投稿

- Haikang Diao** et al., "A Computing-in-Memory Engine Supporting One-Shot Floating-Point NN Inference and On-Device Fine-Tuning for Edge AI," in **IEEE Journal of Solid-State Circuits**, doi: 10.1109/JSSC.2024.3522304.
- Haikang Diao**, et al., "A Multiply-Less Approximate SRAM Compute-In-Memory Macro for Neural-Network Inference," in **IEEE Journal of Solid-State Circuits**, doi: 10.1109/JSSC.2024.3433417
- Haikang Diao**, et al., "AdapCIM: A 28nm 73TOPS/W Adaptive Stationary Digital CIM Accelerator with Inter-Column/Row Sharing and Kernel-Wise Sparsity," in **IEEE Transactions on Circuits and Systems for Artificial Intelligence**.
- Haikang Diao**, et al., "Adder-DCIM: A Parallel Bit-Flexible Digital CIM Joint Model Compression Framework for AdderNet Inference" in The 2025 IEEE/ACM International Conference on Computer-Aided Design (**ICCAD**)
- Haikang Diao**, et al., "A 28nm 128TFLOPS/W Computing-In-Memory Engine Supporting One-Shot Floating-Point NN Inference and On-Device Fine-Tuning for Edge AI," 2024 IEEE Custom Integrated Circuits Conference (**CICC**), Denver, CO, USA, 2024, pp. 1-2.
- Haikang Diao***, Haoyi Zhang*, et al., "SEGA-DCIM: Design Space Exploration-Guided Automatic Digital CIM Compiler with Multiple Precision Support," in Design, Automation and Test in Europe Conference (**DATE**), March. 2025.
- Yizhe Chen*, **Haikang Diao***, et al., "DEAL-CIM: Cross-Layer Design Space Exploration and Automated Layout Compiler Framework for Hybrid eFlash CIM Accelerators" in **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (IEEE TCAD)** (Major revision).
- Haikang Diao**, et al., "Real-Time and Cost-Effective Smart Mat System Based on Frequency Channel Selection for Sleep Posture Recognition in IoMT," in IEEE Internet of Things Journal, vol. 9, no. 21, pp. 21421-21431
- Haikang Diao**, et al., "Deep Residual Networks for Sleep Posture Recognition With Unobtrusive Miniature Scale Smart Mat System," in IEEE Transactions on Biomedical Circuits and Systems, vol. 15, no. 1, pp. 111-121, Feb. 2021.
- Haikang Diao**, et al., "Unobtrusive Smart Mat System for Sleep Posture Recognition," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), Daegu, Korea, 2021, pp. 1-5
- Haikang Diao**, et al., "Edge-Computing System based on Smart Mat for Sleep Posture Recognition in IoMT" In 10th EAI International Conference on Wireless Mobile Communication and Healthcare (EAI MobiHealth), 2021.