

刘勇

Architecture Mode...

内推门户(人才推荐)

社招

内推

已入职

★

男 33岁 · 5年工作经验

+86 158 0197 8636 (上海)

1337752508@qq.com

211

双一流大学

湖北杰之行体育产业发展股份有限公司

北京希姆计算科技有限公司

紫光展锐(上海)科技有限公司

南京林业大学 · 本科

上海大学

原始简历

附加信息 1



## 简历

刘勇

基本情况

名 族：汉族

学 历：工学学术型硕士

户 籍：上海

政治面貌：中共党员

邮 箱：1337752508@qq.com

手机号码：15801978636

教育经历

2015/09-2018/04

上海大学 · 控制理论与控制工程（均分：85.75/100）

2011/09-2015/06

南京林业大学 · 自动化（均分：90.33/100，保研）

专业技能

● 编程能力：掌握 C 和 Python，熟悉汇编和 C++

获 证 书

● 证书：英语：CET-6 计算机：(全国)三级网络技术+四级网络工程师 (江苏)三级偏软+ 三级偏硬

● 荣誉：国家奖学金，国家励志奖学金，2017 年全国大学生数学建模挑战赛三等奖

工作经验

之行无界

系统工程师

2023/08 - 至今

● RISC-V-Matrix 指令集的完备性：  
1) 指令抽象：根据相关算子或网络的分析，抽象出具体的矩阵指令  
2) 指令验证：用 spike 仿真器建模矩阵指令，验证其功能  
● NPU 算子的开发：  
在 spike 上使用 rvm 指令，实现：  
1) 算子：76/120 个具体算子  
2) 网络：yolov5s、bert 和 vit 等具体网络  
3) 量化：yolov5s 网络的量化(PTQ)版本  
● llm 相关进行中)：  
1) 使用 tensorRT\_llm、vllm、tgi 和 triton 等平台测试 llama 等模型的性能数据和相关分析  
2) llm 性能的相关优化

希姆计算

系统架构工程师

2021/06 - 2023/08

● 指令集-验证类：  
1) 测试框架：RISC-V Vector 和公司定制的 Vector(如 bf16)、矩阵和卷积类等指令的 python 实现件 golden  
2) 设计测试：根据指令定义设计不同的验证场景，如基础、特殊值、对齐、全覆盖、随机和系统参数等类型  
3) 统一 defaultNan 值：浮点指令(如计算、比较和转换等)和 softfloat 库与 Nan 值相关的问题的分析、验证和解决，最终推动软硬件可进行 bit 级结果比对，加速验证效率  
4) 联合验证：与硬件联合验证，发现和解决软、硬件大量 bugs，加快问题的最终收敛  
● NPUv2-建模类：  
1) 指令级别：用 spike 仿真器建模，支持定制指令的建模：bf16 和 fp8 等浮点接口的开发和验证；定制指令的 commitlog 机制开发便于 debug  
2) cycle 级别：用 gem5 仿真器建模，支持定制 MME 指令的建模(准确度约 95%)，加速和相关问题的解决  
● NPUv3-探索类：  
1) 精度(spike)：基础累加算法、精度位位数的精度分析，多输入浮点加法器设计(和使用方式的)精度分析  
2) 架构(gem5)：(1)分布式共享内存系统的架构探索：共享内存，集中式->分布式；noc 拓扑，mesh->ring。(2)并行编程模型(如矩阵的 m、k 和 n 方向的拆分的实现和对应的 NPUv2 与 NPUv3 性能分析  
3) 带宽：无损压缩方案的调研

Thanks for your attention!

紫光展锐

系统工程师

2018/05 - 2021/06

● 满足 LTE/NR 等通信系统功能需求的 VDK 平台的设计、开发和维护(ESL 建模)  
1) 模块建模：虚拟 RFT、DFE 等加速器模块的建模，Uniclk、中断机制等的设计  
2) 平台搭建：根据建 model、系统的处理器和 AMBA 总线等，在 PCT 软件上完成平台集成  
3) 维护升级：release 满足 PHY 对该平台的需求，解决使用中的问题，升级平台满足新需求和功能  
4) 使用验证：使用该平台完成对 performance 上需要的验证和测试，给出优化建议  
● 利用 coresight(Arm CortexR5/R8)的 PMU 进行 performance 分析和功能开发，利用 ETM 和 ETB 进行问题的分析和定位  
● performance 的分析和研究，时序的分析和时序图的功能开发  
1) 对 ARM 的 Performance Mointoring Unit(PMU)模块的维护和功能开发  
2) 据测试开发的 case，定期检测代码的 Profile，查找版本性能变差的原因，达到预警  
3) 主要针对 LTE/NR 制式下时序图的开发和维护，精确到 slot/subframe 的 symbol 级的 task/irq 的时序开销 (122M88hz)，极大提高对时序问题的分析，也能为早期预言等项目提供重要的基础数据  
4) 对 assert 时 bug 问题的分析和工具的开发，综合 assert 时存储的 PMU 和 Timing 的相关信息，利用 Python 对数据的处理和分析并开发成工具，极大提高对 bug 问题的分析，相关工具在部门间推广  
5) 支持在 task 级和 function 级的相关模块的开销定位，利于定位开销大头并对其进行优化  
● 优化和探索：如 L2 命中率与 codeSize 和 dataSize 的关系，log 机制的优化等  
● 定制 benchmark 代码的开发和测试，利用其对 core 和 code 进行性能评估

智能照明控制系统设计及应用(上海市科委项目)

项目负责人

2016/03 - 2017/12

● 简介：采集板将微波、红外、声音和照度传感器感知的信息处理后无线传给触摸屏，触摸屏由模糊算法将计算的亮度和色温值由无线传给驱动板，控制 3 路 PWM 比例和大小进行调光，实现筒灯的亮度和色温随环境的变化而变化，亦可根据预先设定的典型室内情景的照明模式。直接点击触摸屏图标实现对筒灯的无线控制。  
● 责任：触摸屏（STM32F03ve）的编程；采集板、驱动板（Atmega48）的 PCB 绘制、焊接和编程；光源板的 PCB 绘制和焊接；传感器的选型和编程；BOM 表的制作和元器件的采购。  
● 扩展：控制平台由触摸屏换成 Android 手机，通信由 Zigbee 改成 WIFI 和 BlueTooth，实现类似功能。  
● 成果：基于触摸屏控制的无线筒灯；基于 Android 手机控制的无线筒灯；核心论文 2 篇，软件著作权 2 份。

自我评价

自学能力较强，喜欢研发和技术，有较强的责任心与团队意识，善于总结经验和优化方案；喜爱阅读，大学有四百多本书的阅读量；喜爱诗词，偶写诗自娱；喜欢游和摄影，爱爬上和跑步。

面试

修改反馈

备注

### 面试

#### 复试 ~

● 现场面试 2024年3月6日 星期三 17:00 ~ 18:30

负责人 Gloria Shen

面试地址 上海市 闵行区 田林路1016号科技绿洲三期11号楼

面试会议室 San Ya-1F

候选人状态 已接受

#### 面试官

N Ninos Kang 4.Recommend to hire 03-07 17:12

Q1、Technical Evaluation  
Please input your comments :  
对vector的指令有实际经验，具备软硬件调试经验，具有时序模型开发和通用模型框架的使用经验  
Q2、Culture Fit  
Please input your comments :  
踏实细致，有责任心  
Q3、Overall Comments

查看详情

### 简历筛选

T Travis.Zhang... 通过了该简历 2024-02-29

阶段：HM Review

备注

4

>

G Gloria Shen

2024-03-05

@Travis zhang hi travis，需要安排刘勇F2F二面吗？

T Travis.Zhang@armchin...

2024-03-05

请安排二面，谢谢。

C Chengkun.Sun@armchi...

2024-02-29

@Travis zhang Please have a look at this candidate, Thanks.

?