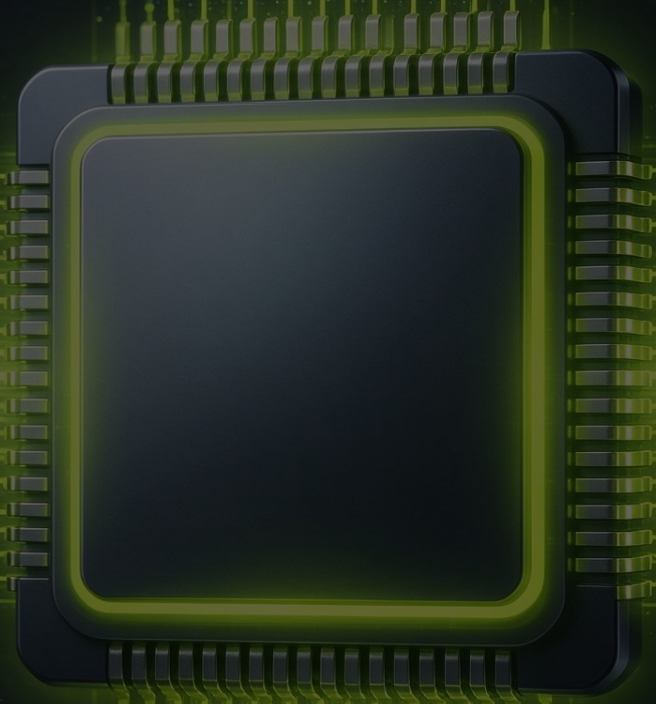
RISC-V 开放软硬件生态

如何支撑 AI 硬件产品快速落地



SPACEMIT
进迭时空

关于进迭时空

进迭时空成立于 2021 年 11 月，是一家 RISC-V 架构 AI CPU 芯片企业，致力于推动和完善 RISC-V+AI+AI 编程语言（Triton/Tilelang）新一代智算生态，助力 AI Agent、具身智能等未来产业的发展



办公地点

- 杭州、珠海、上海、北京、深圳、香港、西安

核心团队

- 现有全职员工超过 350 人，80% 研发人员，62% 毕业于 985/211 院校，56% 硕博
- 创始团队和核心骨干来源于阿里平头哥、华为海思、全志科技等知名企业

研发成果

- 已经迭代量产了两代自研芯片
- 芯片产品：K1、K3 AI CPU芯片、V100 CPU服务器芯片、P1电源管理芯片

01 | 趋势

算力下沉家庭，NAS 从「存东西」变成「想东西」

AI 算力正复刻十年前存储下沉的路径，数据与智能在本地开始融合

传统 NAS

存储 · 备份 · 共享

智能 NAS

App 管理 · 远程访问

AI NAS

本地大模型 · 智能检索 · 知识库

主流NAS厂商，已在 2024—2026 年间陆续把「本地 AI」写入产品路线图；
AI 越强，越需要私有数据——照片、视频、合同、录音里真正有价值的东西，不该只是躺在硬盘里，也不该无脑交给云端。

AI 硬件落地，真正的门槛在生态与软件

把 AI 做进硬件，
比做一款 AI 产品更难

芯片之外，生态、软件、模型、人才，每一项都是隐性成本。
绝大多数团队的时间，没有花在差异化上，而是耗在了底层重复建设。

生态碎片化

架构与平台林立，软件跨平台难以复用

软件栈封闭

依赖专用 NPU 工具链，闭源库多、上手门槛高

模型适配成本高

开源模型需转译、等待原厂适配，迭代慢

重复造轮子

缺乏开放可复用底座，每个项目从零起步

PART 03

破局的起点，是开放

为什么是 RISC-V 开放生态

03

开放，是把生态做厚、把成本做薄的唯一方式——
不被单一厂商锁定，才能让全球的开发者一起，把底座共建到足够强。

RISC-V 是全球共建的开放公共指令集

对比维度	闭源指令集（X86 / ARM）	RISC-V
授权模式	商业授权，需支付授权费	**开放标准，免授权费**
主导方式	单一商业公司主导	**基金会治理，全球共建**
软件互通	受授权与厂商限制	**遵循规范即可互通**
扩展能力	定制扩展受限	**模块化，可自由扩展**

「基础指令 + 标准扩展」的模块化设计，让 RISC-V 可按需裁剪；RVV 向量、Vector Crypto 等扩展，天然适配 AI 与高并行计算——这是它成为端侧 AI 硬件底座的根本原因。

进迭时空全栈开放支持体系

从芯片到开发者平台，四层全部开放，可直接复用、按需裁剪

开放开发者平台	SpacemiT Studio · 文档中心 · AI 问答 · 云端体验集群
开放 AI 栈	llama.cpp / ONNX Runtime · 模型仓库 · 类 CUDA 通用 AI · 模型免转译
开放软件层	Bianbu OS · BSP / SDK · Linux 主线 · 多 OS 原生适配
开放硬件层	K1 / K3 AI CPU · 处理器核 · 全量原理图 / BOM · 参考设计

芯片、处理器核、设计资料，全部公开



60 TOPS · INT4 **130** KDMIPS

K3：8 个 X100 大核 + 8 个 A100 AI 核

硬件资产，全部开放

自研 RISC-V AI CPU

K1 规模量产、K3 可流畅运行 30B 大模型

处理器核开放

X100 支持完整虚拟化，A100支持1024bit RVV1.0并行计算

全量设计资料

原理图、BOM、芯片手册公开，无需签署 NDA

参考设计

Pico-ITX、CoM260 核心板、Cluster、Shelf 服务器

一套 BSP 主线，原生支撑全球主流 OS

完整 BSP，开源开放

- **Bianbu OS（GNOME / LXQt / Minimal / NAS）**
- 根文件系统 Rootfs（可定制）
- Linux Kernel（持续 Upstream）
- U-Boot / UEFI
- OpenSBI（S 模式固件接口）

原生适配的操作系统

- Ubuntu
- openEuler
- Debian
- Fedora
- OpenCloudOS
- Armbian
- openKylin
- Deepin
- OpenHarmony
- Arch Linux

K3 是 Linux 主线官方指定的 RISC-V 适配样板，并获 Ubuntu 首个 RVA23 官方支持；Buildroot 与 Bianbu 共享同一 BSP。

开源模型免转译，像写 CPU 一样做 AI

传统专用 NPU 路径

开源模型

↓

专用编译器转译、量化

↓

绑定专用 NPU 工具链

模型适配滞后，生态封闭

进迭时空 CPU AI 路径

开源模型，直接获取

↓

llama.cpp / ONNX Runtime 直接运行

↓

1024bit RVV 标准向量指令

模型生态同步，无需专用工具链

底层能力支撑

1024bit RVV 1.0 + Vector Crypto

BF16 / FP16 / FP8 / INT8 / INT4

兼容 DeepSeek / Qwen / Llama

模型仓库 + 微调工具

从文档到云端，开发者全程在线



文档中心

developer.spacemit.com
技术文档与示例

SpacemiT Studio

线上一站式开发
在线编码、调试

AI 问答机器人

7×24 智能答疑
社区与知识库联动

云端体验集群

珠海 K1/K3 集群
远程测试模型性能

没有开发板也能先上手：通过云端 K3 集群即可远程部署、运行模型、验证性能，
确认方案后再采购硬件，把评估与试错成本降到最低。

AI NAS 落地架构：四层结构，软硬一体

应用层

- WebUI
- Mobile App
- AI Agent

服务层

- **多媒体处理** — FFmpeg · GStreamer
- **AI 推理** — llama.cpp · ONNX Runtime
- **存储服务** — File Sync · Backup
- **访问控制** — User Access

OS 层

Linux Kernel · Bianbu OS（V4L2 / Docker / 标准驱动）

硬件层

- **计算** — CPU · GPU · VPU · DRAM
- **网络** — ETH · Wi-Fi · BT
- **存储** — SSD · SATA · eMMC

K3 把 AI 做进主控：多模型并发，还留得住升级空间

关键维度	外挂式 AI（主控 + 独立 AI 芯片）	K3 单芯片主控 AI
AI 可用内存	外挂内存，独立焊死，不与主控共享	**32GB 统一内存，全部模型共享**
常驻模型	单模型串行，切换需卸载重载	**3+ 模型并发常驻（Embedding/Rerank/LLM）**
最大模型	7B 封顶	**30B 流畅推理，>15 tok/s**
扩展与升级	AI 占满 PCIe，0 卡位，只能换整机	**AI 不占通道，PCIe3.0×8，可加算力卡、存储**
系统与工具链	双芯片、两套工具栈，BOM 与维护复杂	**单芯片、标准 CPU 工具链，模型免转译**

AI NAS 是一个要不断加场景、多模型并行的品类。选外挂式 AI，等于在选型当天就把天花板钉死在「5GB / 7B / 单模型」，且没有补救插槽；K3 用统一内存 + 通道余量，把 AI NAS 做成一个可成长、可演进的平台，软件栈也能持续沉淀、不必随芯片换代而迁移。

06 | 共建开放生态

三步上手，把 AI 快速做进你的硬件

01

云端体验

无需硬件，远程访问珠海 K3 集群，先跑模型、验证性能。

02

开发板评估

Pico-ITX / CoM260，开放 BSP、SDK 与全量设计资料。

03

量产落地

芯片供应 + 认证伙伴，联合营销打单。

资源与合作入口

官网 spacemit.com

模型仓库 archive.spacemit.com

开发者社区 spacemit.com/community

商务合作 business@spacemit.com

开放 RISC-V 生态，让 AI 硬件落地更快、更省、更可持续

进迭时空期待与各位伙伴共建 RISC-V 普惠智算生态