



南开大学
Nankai University

南 开 大 学

计 算 机 学 院

并行程序设计

并行体系结构调研

中国自主处理器技术发展调研报告

孟启轩

年级：2022 级

专业：计算机科学与技术

指导教师：王刚

2025 年 3 月 21 日

摘要

通过广泛查阅资料结合课堂内容，我调研了中国自主处理器技术的发展历程，并深入剖析申威神威太湖之光超算的 SW26010 实现的并行体系结构设计以及微架构创新，分析 ARM 授权与自主指令集的风险。

关键字：处理器 CPU 微架构 AI 并行

目录

一、 中国自主处理器技术发展历程	1
(一) 起步阶段 (2000-2010 年): 技术引进与消化吸收	1
(二) 自主化探索 (2011-2020 年): 指令集突围与架构创新	1
(三) 全面自主阶段 (2021 年至今): 生态构建与高端突破	1
1. 龙芯	1
2. 飞腾	2
3. 申威	2
4. 阿里巴巴	2
5. 华为	2
6. AI 芯片	2
二、 中国自主处理器的技术路线创新与微架构突破	3
三、 申威神威太湖之光超算的 SW26010 处理器	3
(一) 并行体系结构设计	4
1. 大规模多核 MIMD 并行架构	4
2. 三维堆叠与片上存储优化	4
3. 超低功耗与能效比优化	4
(二) 微架构创新与关键技术	4
1. SW64 指令集与向量计算单元	4
2. 片上网络与通信优化	5
3. 安全与可靠性设计	5
(三) 技术突破与实际应用	5
1. 突破工艺限制的典范	5
2. 超算的架构基础	5
四、 ARM 授权与自主指令集的风险	6
五、 未来方向	6
六、 总结	7

一、中国自主处理器技术发展历程

(一) 起步阶段 (2000-2010 年): 技术引进与消化吸收

起步阶段主要依赖国外指令集授权，生态碎片化严重，性能落后国际主流 2-3 代。

- 龙芯 1 号 (2002): 中国首款通用 CPU，基于 MIPS 指令集，主频 266MHz，应用于嵌入式设备和教育领域。
- 飞腾 FT-1000 (2006): 基于 SPARC 架构的国产服务器 CPU，首次应用于国防领域，主频 1GHz。

(二) 自主化探索 (2011-2020 年): 指令集突围与架构创新

- 龙芯 3B1500 (2012): 首款 8 核 MIPS 处理器，支持多路互联，峰值性能达 160GFLOPS。
- 申威 1600 (2012): 采用自主 Alpha 指令集扩展，支撑“天河二号”超算在 2013 年登顶 TOP500。
- 飞腾 FT-2000+ (2017): 基于 ARMv8 架构的 64 核服务器 CPU，首次实现单路服务器国产化替代。

自主化探索阶段实现了关键突破。龙芯 LoongArch 指令集 (2020) 彻底摆脱 MIPS 依赖，兼容 x86/ARM 生态。此外，华为昇腾 910 (2019) 是全球首款 7nm AI 训练芯片，算力达 256TFLOPS (INT8)。

(三) 全面自主阶段 (2021 年至今): 生态构建与高端突破

中国自主处理器技术在这一阶段实现了从“可用”到“好用”的跨越，通过指令集自主化、高性能架构创新和生态闭环建设，逐步突破国际技术壁垒，形成多领域协同发展的格局。这一阶段标志着中国从“跟随者”向“创新者”的角色转变，但仍需解决生态碎片化、先进制程依赖等挑战。

1. 龙芯

完成从 MIPS 到自研 LoongArch 指令集的彻底转型，兼容 x86/ARM 生态，实现完全自主。创新使用乱序执行 (ROB 256 条目)、超标量 8 发射，向量扩展 LVZ/LVF 提升计算密度等技术。

- 龙芯 3A5000 (2021): 首款 LoongArch 桌面 CPU，12nm 工艺，IPC 提升 50%，性能对标 Intel i5-7500。
- 龙芯 3A6000 (2023): 4 核 LA664 核心，2.5GHz 主频，单核性能接近 Intel i5-12 代，集成安全可信模块 (EAL4+ 认证)。
- 龙芯 3C6000 (2025): 64 核服务器 CPU，支持多路 CCIX 互连，主频 2.0GHz，能效达 18.3 GFLOPS/W (Green500 榜单)，性能对标鲲鹏 920。

2. 飞腾

飞腾腾云 S5000C-E 通过 64 核 ARMv9 架构和超标量乱序执行技术显著提升了单核性能（主频 2.8GHz），结合 8 通道 DDR5-4800 内存实现高带宽数据传输，同时采用优化的片上网络拓扑将通信延迟降至 <100ns，解决了大规模核间协同瓶颈；其内生安全架构 PSPA2.0 与硬件虚拟化技术保障了高安全场景需求，并通过全栈优化支持 DeepSeek 大模型等 AI 负载，在性能与生态适配性上实现突破，成功服务全国两会等国家级智算项目，推动国产服务器芯片在高性能计算与 AI 领域的自主化落地。

3. 申威

申威威鑫 H8000 (2024) 通过自主 SW64 指令集和 256 核大规模并行架构，实现了 1PFLOPS 的峰值性能，支撑“神威”超算迭代。其核心突破包括 6D Torus 拓扑优化了片间通信延迟至 <100ns，大幅提升多核协同效率；结合申威智能服务器与 DeepSeek 大模型一体机，构建了从硬件到算法的全栈国产 AI 生态，适配语音转写、代码生成等场景，并通过与统信 UOS 深度适配，强化了在高性能计算、AI 推理和行业数字化转型中的自主可控能力。

4. 阿里巴巴

玄铁 C930 (2025) 作为阿里巴巴首款服务器级 RISC-V 芯片，以开源 RISC-V 指令集为基座，采用 64 核设计（12nm 工艺），同时支持 256 核 Chiplet 扩展，解决了 RISC-V 在数据中心规模化应用的瓶颈。其突破性在于通过模块化设计和生态开源策略，推动 RISC-V 架构在边缘计算与数据中心的落地，与云天励飞芯片协同支持 LLM 推理，并兼容 DeepSeek 等国产大模型，为构建去中心化、灵活扩展的计算生态提供了技术路径。

5. 华为

华为通过自主架构创新（TaiShan V120）、异构计算整合（CPU+NPU 协同）和全栈生态闭环（芯片 + OS + 云），在鲲鹏 930 和麒麟 X90 上实现了众多突破。性能与能效方面，通过 7nm 工艺、超线程、SVE2 指令等技术，缩小与 x86 的差距。并且与鸿蒙 OS 深度整合，推动 ARM 在 PC 和服务器的国产化替代，逐步实现生态闭环。

- 鲲鹏 930 (2024): 7nm 工艺, 64 核 ARMv9 架构, 集成 128TOPS NPU, 支持 PCIe 5.0 和 CXL 2.0, 面向云原生与 AI 推理。
- 麒麟 X90 (2025): 首款国产 PC 级 ARM 处理器, 3.0GHz 主频, 7nm 工艺, 支持 LPDDR5X 与 PCIe 4.0, 与鸿蒙 OS 深度绑定, 构建“芯片 + 系统”垂直生态。

6. AI 芯片

实现了高算力密度、低功耗设计与生态兼容性的突破，加速 AI 芯片从云端训练到边缘推理的全场景国产化替代进程。

- 寒武纪思元 370 作为首款 Chiplet 设计 AI 芯片，通过模块化封装实现 256TOPS@INT8 算力，结合 LPDDR5X 内存接口提升带宽，突破传统单芯片算力瓶颈。
- 昇腾 910 采用 7nm 工艺，FP16 算力达 512TFLOPS，支持液冷散热，首次在国产芯片中实现训练与推理场景全覆盖，满足大模型高性能需求。

- 飞腾 FT-2000AI 创新性地采用 CPU+AI 加速器混合架构，以 10TOPS/W 的能效比平衡算力与功耗，同时兼容 OpenVINO 等主流框架，推动国产 AI 芯片在边缘计算与通用场景的落地。

二、 中国自主处理器的技术路线创新与微架构突破

中国自主处理器技术路线的创新主要体现在多路径并行发展的战略选择与系统级架构的颠覆性突破。在技术路径上，形成了三大核心模式：“高铁模式”，海光、飞腾通过引进-吸收-再创新，基于 X86、ARM 架构实现快速迭代；“北斗模式”，龙芯、申威完全自主设计 LoongArch、SW64 等指令集，构建独立生态；“5G 模式”，参与国际标准制定并融合自研技术。在此基础上，进一步提出“双模融合”理念，通过双指令集实现技术自主性与市场兼容性的平衡。此外，以 SDSoW 为代表的系统级创新开辟了新赛道，通过晶圆级异构集成与软件定义硬件重构，突破制程工艺限制，实现性价比与能效的指数级提升，形成“功能解构-晶上重组”的超限创新范式，为算力架构与 AI 模型的协同进化提供物理载体支撑。

在微架构方面，国产处理器通过深度自研与技术迭代实现了性能与安全的双重跨越。龙芯基于 LoongArch 指令集，推出 3A6000 处理器，其性能逼近国际主流 10 代酷睿 i3 水平，微架构设计覆盖 PCIe、DDR5 等关键 IP 的自主重写，并构建“loongISA+LCC+GS464E”软硬件协同体系，提升安全性与生态可控性；海光则在 X86 架构基础上发展 C86 指令集，通过硬件级国密算法支持与主动防御机制，实现政务、金融等核心场景的虚拟化性能提升，同时兼容性与安全性显著增强。申威依托 SW64 指令集，在高性能计算领域持续深耕，其神威超算系统通过液冷技术与全栈国产化设计，展现了超算领域的自主可控能力。此外，SDSoW 架构通过生成式变结构计算，实现晶圆级硬件动态重构，可针对 AI、超算等场景按需优化，使“二流工艺”实现“一流性能”，为国产芯片在先进制程受限下提供了非对称技术博弈的路径。这些突破不仅缩小了与国际主流产品的差距，更在安全可控、生态构建与系统级创新层面形成了差异化优势。

三、 申威神威太湖之光超算的 SW26010 处理器

申威神威太湖之光超算系统曾以每秒 9.3 亿亿次的浮点运算能力位居 2016 世界超算 TOP500 榜首，其核心处理器为申威 SW26010。该处理器完全基于中国自主设计的 SW64 指令集架构，采用 28nm 工艺，在制程落后国际主流的情况下，通过架构创新实现了性能突破，在 2024TOP500 榜单中排第 15 位（如图 1），是国产超算并行体系结构的标杆。其技术路径为后续国产处理器发展提供了重要参考，尤其在异构计算优化、能效平衡、自主生态构建方面，为中国超算技术突破提供了非对称竞争策略。

#	A	B	C	D	E	F
1	Ran	Previ	First	Firs	Name	Computer
2	1		64	1	El Capitan	HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slir
3	2	1	59	1	Frontier	HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instir
4	3	2	62	2	Aurora	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz
5	4	3	62	3	Eagle	Microsoft Ndv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infir
6	5		64	5	HPC6	HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instir
7	6	4	55	1	Supercomputer Fugaku	Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D
8	7	6	63	6	Alps	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Sling
9	8	5	59	3	LUMI	HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instir
10	9	7	60	4	Leonardo	BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64
11	10		64	10	Tuolumne	HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slir
12	11	8	62	8	MareNostrum 5 ACC	BullSequana XH3000, Xeon Platinum 8460Y+ 32C 2.3GHz, NVIDIA H100 64GB,
13	12	10	62	9	Eos NVIDIA DGX SuperPOD	NVIDIA DGX H100, Xeon Platinum 8480C 56C 3.8GHz, NVIDIA H100, Infinibar
14	13	11	63	11	Venado	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Sling
15	14	12	51	3	Sierra	IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual
16	15	13	47	1	Sunway TaihuLight	Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway
17	16		64	16	CHIE-3	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband
18	17		64	17	CHIE-2	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband
19	18		64	18	JETI - JUPITER Exascale	BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Super
20	19	14	57	5	Perlmutter	HPE Cray EX 235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, SI
21	20		64	20	El Dorado	HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slir
22	21		64	21	Gefion	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100 SXM5 80GB, (
23	22	17	63	17	CEA-HE	BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Super
24	23	15	55	7	Selene	NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR 1

图 1: 申威神威太湖之光在 2024TOP500 榜单中的排名

(一) 并行体系结构设计

1. 大规模多核 MIMD 并行架构

SW26010 采用 28nm 工艺集成 260 个计算核心（256 个通用核心 +4 个管理核心），通过二维 Mesh 片上网络实现核心间互联，形成 MIMD（多指令多数据）并行架构。每个核心可独立执行异构任务，通用核心以 1.45GHz 主频提供 1.23 TFLOPS/核（单精度）的峰值性能，而 4 个管理核心专负责任务调度、错误处理及系统监控，通过硬件级容错机制保障超大规模计算的稳定性，兼顾高性能计算与系统可靠性。

2. 三维堆叠与片上存储优化

SW26010 通过 TSV 硅通孔技术实现 HBM 高带宽内存三维堆叠，将 4GB HBM 直接集成于芯片封装，提供 1TB/s 带宽和 10ns 低延迟，显著降低内存访问能耗。同时，每个核心配备 64KB I-Cache 与 64KB D-Cache，并通过分布式片上 NoC 共享 16MB 二级缓存，减少对外部 DRAM 的依赖，形成“核心-缓存-HBM”三级存储层次，大幅提升数据吞吐效率与能效比。

3. 超低功耗与能效比优化

SW26010 采用动态电压频率调节 DVFS 技术，根据负载实时调整核心电压与频率，并支持空闲核心休眠，整机在 Linpack 测试中实现 5.93 GFLOPS/W 的能效比，远超同期国际同类产品。其二维 Mesh NoC 通过低功耗分组交换与信用流控协议，优化通信能耗，结合全局地址空间管理，进一步降低系统级功耗，为超算长期稳定运行提供高效节能保障。

(二) 微架构创新与关键技术

1. SW64 指令集与向量计算单元

SW64 指令集基于 RISC 精简指令集原则设计，支持 128 位向量运算（SIMD），通过双发射超标量流水线（14-16 级流水线）实现单周期双指令执行，显著提升浮点密集型任务的运算效率。针对超算典型场景（如流体力学、气候模拟），其向量扩展引入专用融合指令，例如 FMA（乘加

融合) 减少中间结果存储开销, 提升浮点吞吐量; 跨核心数据广播功能则通过片上网络直接传输向量数据, 绕过内存带宽瓶颈, 优化大规模并行计算的通信效率。

2. 片上网络与通信优化

SW26010 采用二维 Mesh NoC 架构, 每个核心通过 4 方向链路互联, 支持数据分组交换与路由仲裁, 通信延迟低至 0.5ns/跳。其流量控制机制基于信用流控协议, 动态调节数据传输速率以避免网络拥塞; 异步非阻塞通信模式允许任务间数据交换无需 CPU 干预, 提升系统吞吐量。此外, 全局地址空间通过软件层统一管理超算系统的物理内存, 实现 Peta 级内存的全局透明访问, 显著简化并行编程复杂度。

3. 安全与可靠性设计

SW26010 集成硬件级安全机制, 包括国密算法加速引擎, 实现加密通信与数据保护; 片上存储和内存通道采用双纠错、四检测的 ECC 技术, 保障长期运行稳定性。在可靠性层面, 系统通过动态任务迁移和冗余核心切换, 可在单个核心或内存故障时快速隔离故障单元并恢复任务执行, 确保超算系统的高可用性, 同时满足政务、国防等关键领域对安全性和容错性的严苛要求。

(三) 技术突破与实际应用

1. 突破工艺限制的典范

SW26010 在 28nm 工艺下实现了 1.45GHz 主频与 3.16 TFLOPS (双精度) 的单芯片峰值性能, 通过架构创新显著弥补了制程劣势。首先, 通过提升向量单元与指令并行度, 单位面积性能密度超越同期 Intel Xeon Phi 等竞品, 例如 128 位向量运算与双发射超标量流水线设计, 大幅增强了浮点密集型任务的吞吐效率; 其次, 采用三维堆叠技术集成 HBM 内存, 借助 TSV 硅通孔工艺, 以 28nm 节点实现了接近 14nm HBM 的 1TB/s 带宽, 显著降低对外部先进制程工艺的依赖, 同时将内存延迟压缩至 10ns 级别, 为超算算力释放提供了关键支撑。

2. 超算的架构基础

SW26010 在超算领域的实际应用充分验证了其技术价值。在高性能计算场景中, 搭载该处理器的“神威·太湖之光”系统支持的“千万核可扩展大气动力学全隐式模拟”项目, 凭借其大规模并行能力与高效能特点, 成功斩获 2016 年戈登·贝尔奖, 展现了在气候模拟、核聚变研究等复杂科学计算领域的卓越性能; 此外, 其向量指令的灵活重构能力, 使其能够支持轻量级 AI 计算任务, 通过向量单元与全局地址空间的协同优化, 为超算与 AI 融合提供了架构基础, 进一步拓展了其在智能计算领域的应用潜力。

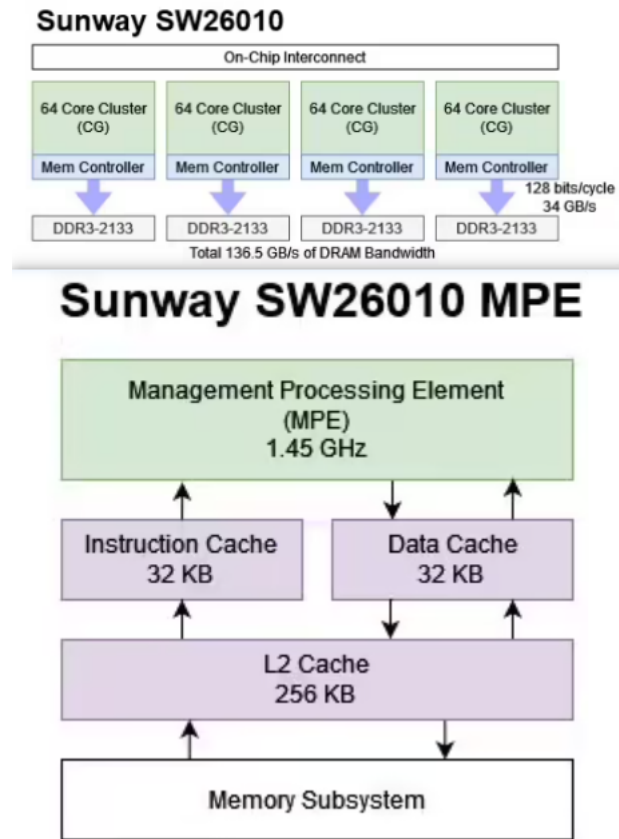


图 2: SW26010

四、 ARM 授权与自主指令集的风险

ARM 授权路线虽能快速接入成熟生态，但存在技术主权风险，包括受国际政治影响的断供威胁、授权费用随架构升级攀升导致的成本压力，以及国防、超算等高端领域可能受 ARM 条款限制；而自主指令集路线虽实现完全可控，却面临生态构建的长期挑战，要从零搭建软件工具链、操作系统和应用兼容性，而且市场接受度受限于生态成熟度。所以需要在技术自主性与生态兼容性间权衡，ARM 路线短期可以抢占市场但隐含“卡脖子”风险，自主路线长期可控但需要跨越生态鸿沟，并且要持续投入架构创新以缩小与 x86/ARM 的性能差距。

五、 未来方向

中国自主处理器未来的发展方向应当聚焦于技术自主性、生态体系化和场景多元化。

首先，加速先进制程与架构创新，通过产学研协同攻关突破先进工艺瓶颈，同时深化异构计算架构，如 CPU+GPU+AI 加速器的 Chiplet 集成、存算一体及软件定义硬件等超前架构设计，提升能效比与算力密度；其次，构建自主可控的生态，在延续 ARM 授权路线市场优势的同时，加速 LoongArch、SW64 等自主指令集的开源生态建设，完善编译器、操作系统及行业应用的兼容性，通过“双模兼容”策略平衡技术主权与生态兼容；再者，在超算、AI 训练推理、边缘计算等高价值领域打造“非对称优势”，不要一味地盲目追赶，也可以开创并提升自身的优势；最后，抓住 RISC-V 的国产化机遇，有望在 AI 时代重塑全球芯片格局。

六、 总结

我国的自主处理器技术经过重重考验，克服技术封锁，逐步实现指令集自主化，高性能架构创新，以及生态构建，已经从“跟跑”实现“并跑”和“局部领跑”的跨越。当然，我们从未停下脚步，相信未来我国能在全球处理器产业格局中形成以技术主权为基础、以生态活力为支撑的第三极竞争力。

NIKU

参考文献

飞腾系列自主核心芯片高性能服务器 CPU [飞腾腾云 S5000C](#)

2024 年 TOP500 榜单 [TOP500List](#)

The Sunway TaihuLight supercomputer: system and applications

我国最新申威 SW26010-Pro 处理器展示, 新一代超算同步亮相 [SW26010](#)

NIJU