2025 年国产 AI 芯片和高性能处理器厂商排名和行业趋势分析

报告概要

作为深芯盟 500 家国产芯片行业分析报告的一部分，2025 年结合高性能 AI 芯片和处理器于一个报告，汇总了 70 余家国产芯片厂商，对于每一家筛选的公司，我们从核心技术、公司发展和应用场景等方面对公司进行全方位画像分析。

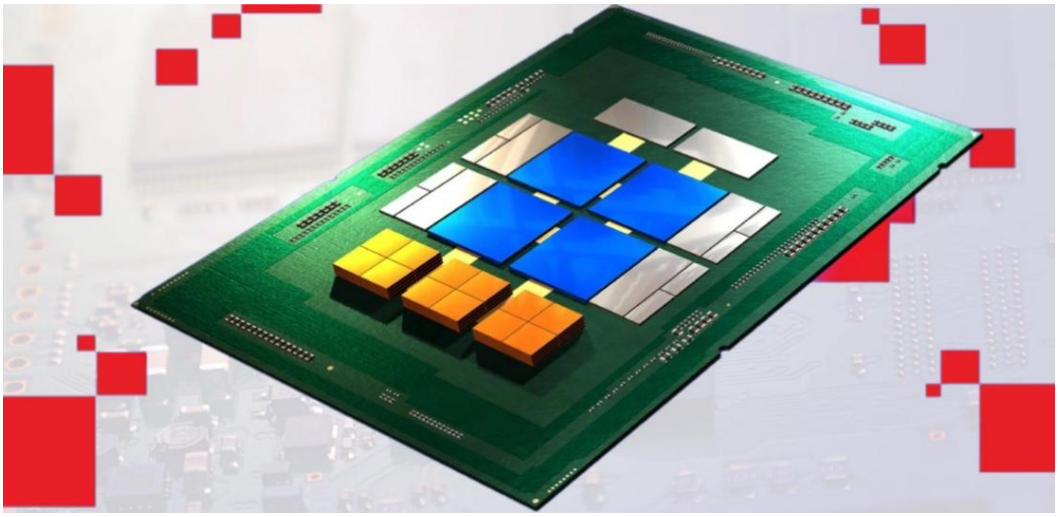
我们首先对 Chiplet、HBM 技术和存算一体技术以及其对 AI 芯片未来发展所带来的影响进行了简要阐述，然后结合应用和潜在需求进行分析，并且针对上市公司的财务数据进行归纳比较。

报告内容目录

一、Chiplet 与高性能计算（HPC）芯片

Chiplet 是最近 AI 芯片和高性能计算领域最火的话题，在芯片设计界有一句话是说，设计一款 3nm 制程的芯片并不困难，但是制造一块 7nm 芯片却让市值千亿的公司花费 4 年时间。而随着摩尔定律进入到 2nm 甚至 1nm 到了近乎原子级别，工艺、设备和材料难度呈几何级上升，而且成本高的吓人，也只有头部的巨头才能玩的起。所以随着芯片技术要求的不断提升，系统级芯片 SoC 开始显得力不从心，Chiplet 技术悄然兴起。

像是大算力 AI 芯片、GPU 和 CPU 芯片，计算单元+存储单元+I/O 接口+电源管理等主要功能模块每个部分都至关重要在一个芯片上设计这么多模块，还要保证制造阶段的良率可以说难度不亚于“登天之道”，而 chiplet 可以说完美契合这一难题，使用模块化的设计方法，通过划分芯片为小块独立的单元来提升芯片的灵活性和可拓展性，使得不同功能晶粒更容易的集成到一个芯片上。



Chiplet 结构示意图（图源：Skyline）

拆分后的芯片甚至可以交给不同的制程去做，各个模块并行开发测试，像是 Intel 和 Nvidia 均采用了 chiplet 开发其产品，既减小了设计难度，又加快了芯片研发进程，实现了更快的产品迭代。并且采用 chiplet 模块化的芯片良率得到的巨大提升，成本也比一整块的芯片低的多。

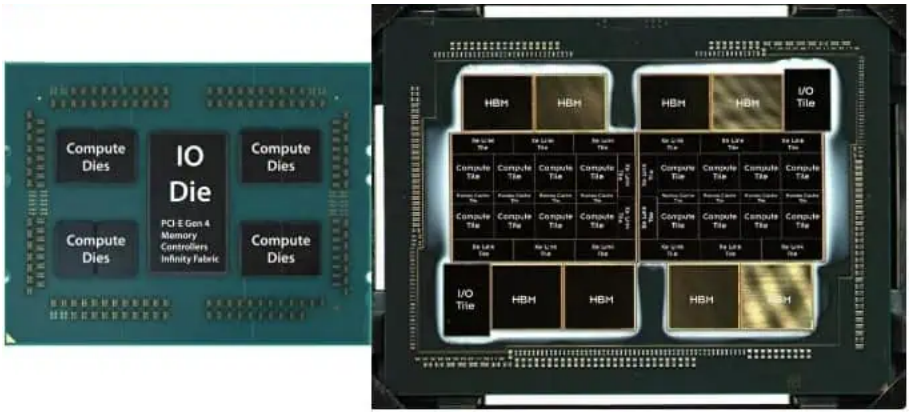
但是新技术就会带来新的挑战，Chiplet 需要在有限的空间内实现芯片的高密度堆叠和信号的高密度互联，不同模块的电信号需要可靠稳定的通信，于是 TSV（硅通孔）、CoWoS 和 InFO 技术等应孕而生；模块多带来的复杂场影响效应也翻倍增长，不同模块的电信号、磁信号、散热、热应力等多物理场互相作用非常复杂，设计工程师和工艺工程师需要紧密配合，不断仿真模拟和改进工艺参数才能保证整个芯片的稳定和可靠。

模块化技术想要推广和发展离不开标准化和兼容性，软件和硬件都绕不开行业的统一标准，UCIe(Universal Chiplet Interconnect Express)就是 Intel、ARM、AMD、TSMC 和三星等十几家 芯片设计和制造巨头联合推出的 Chiplet 标

准，旨在通过统一的接口规范促进 Chiplet 技术的普及和应用。2023 年 9 月 Intel 推出首个遵循 UCIe 连接规范的 Chiplet 测试芯片——Pike Creek，AMD 的 Genoa CPU 和 Instinct MI300 GPU，Nvidia 的 Grace 服务器 CPU 等均是 Chiplet 技术的产物。

Intel、AMD、Nvidia 公司 Chiplet 芯片代表（图源：网络，制表深芯盟）

Chiplet 在高性能计算芯片的设计上显得至关重要，最先进的技术不一定一家公司全都掌握，一块高精尖芯片的诞生就像全球顶级供应链的整合，例如 NVidia 和 AMD 负责设计 GPU 核心，SK 海力士和三星负责 DRAM 和缓存，各大 IP 公司拿出其加速芯片、互联管理芯片和电源芯片等，最终交给半导体芯片制造商台积电或 Intel 负责封装和制造，细看下来其中涉及到的公司可能就有十几家，协调如此庞大的队伍绝非易事，行业联盟的标准化不仅可以降低成本还可以再一个封装体内部实现不同架构不同制程节点的 chip 互联。



UCIe to enable an Open Chiplet Ecosystem delivering Platform on a Package（图源：UCIe）

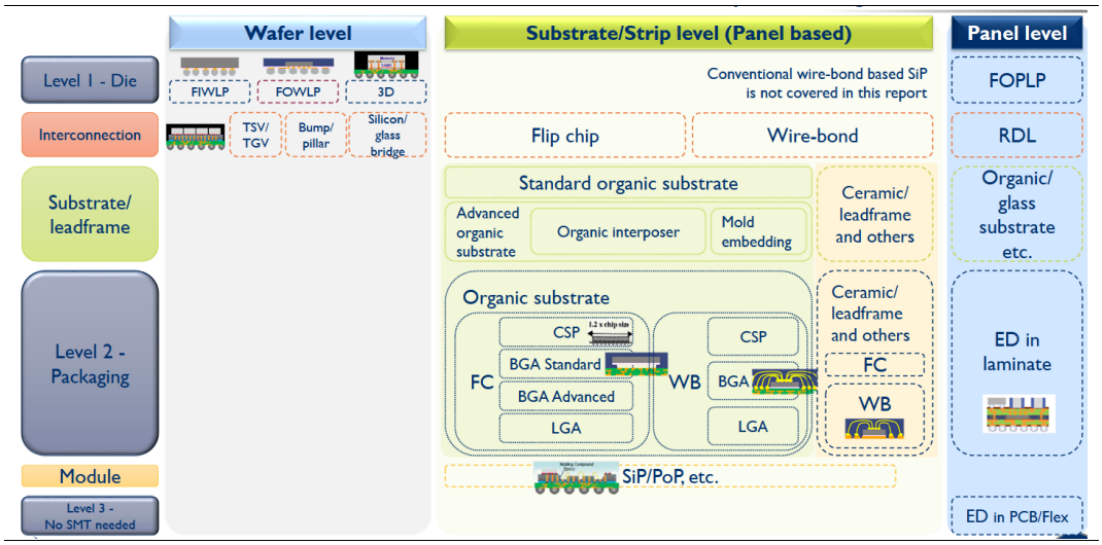
二、CoWoS 与先进封装

CoWoS 初看听陌生，实际上是芯片封装由 2D 向 3D 发展的产物，在芯片不断迭代过程中的一种封装形式。现阶段主流的系统级封装形式：

- 2.5D 封装(Interposer、RDL)
- 3D 封装(TSV)

- 倒装 FC(Flip Chip)
- 凸块(Bumping)
- 晶圆级封装 WLP(Wafer Level Package)
- CoWoS (Chip on Wafer on Substrate)
- InFO (Integrated Fan-Out)
- EMIB (Embedded Multi-die Interconnect Bridge)

CoWoS 正是一种目前台积电主推的 2.5D 封装形式，chip 被放在带有内布线的中介层（Interposer）上，通过芯片上的微小凸块(Bumping)与中介层键合，实现彼此的信号互联。中介层通过硅通孔（TSV）将信号引到另一面，通过锡球或者凸点连接到 PCB 封装基板上，这种设计将原先需要 2D 平面（都放置在基板上）的 die 堆叠起来，极大的提升了芯片集成度，并且 die 与 die 之间的距离大大缩短，高速信号的互联和数据传输的时延降低了几个数量级。此类封装拥有超高布线密度（L/S：0.4/0.4 微米），超高 I/O 密度（大于 400 μbumps/mm² ）和 I/O 间距可扩展性，并且异构芯片和光学、电磁芯片都能完美集成到一个封装体内。

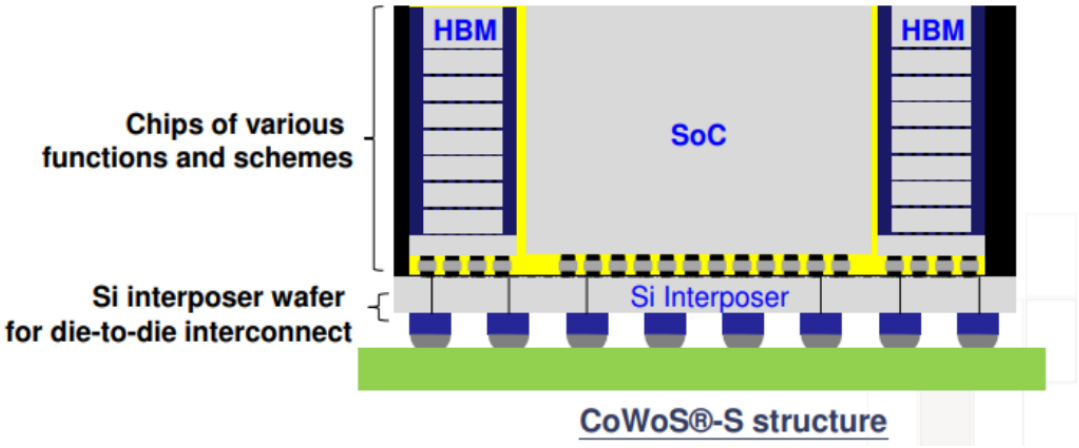


CoWoS 封装技术路线（图源：Yole）

据资料显示，英伟达的算力卡芯片封装就采用了台积电的 CoWoS 方案，单芯片的密度和算力均是之前封装的 4 倍之多，在寸土寸金的 GPU 和 AI 算力芯片领域，CoWoS 不仅节约了空间，还增强了芯片与芯片之间的互联性和降低了传输线损耗。

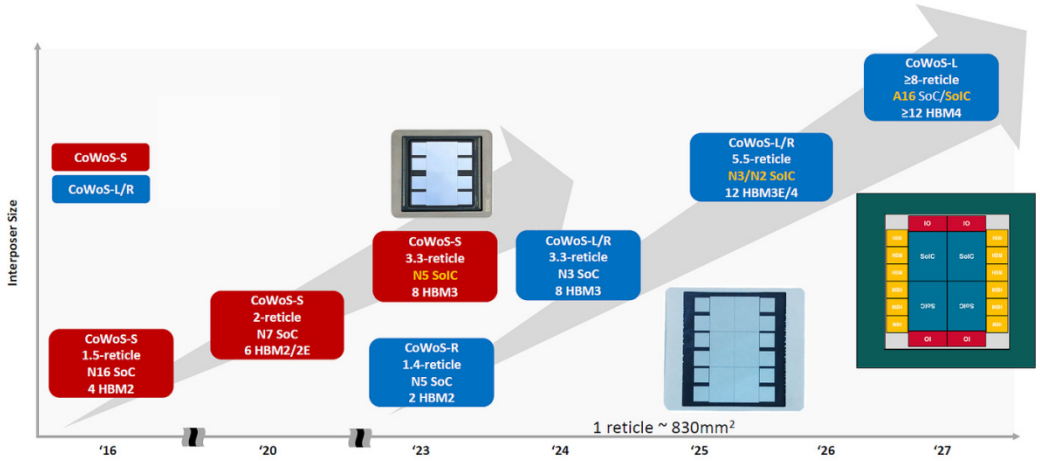
NVIDIA Tesla P100（图源：Nvidia）

这种与 HBM 混合封装技术是获得高速算力和海量数据吞吐的关键技术，目前也是业界最主流的封装方式，CoWoS 为高算力卡的封装提供了其他封装无可比拟的最佳性能和最高集成密度。像是 4xHBM+1SoC 和 2xHBM+1SoC 等等各种中介层尺寸，各个 HBM die 数量自由搭配，可以获得各种性能的算力卡。



HotChips TSMC Packaging Technologies for Chiplets and 3D（图源：TSMC）

TSMC 的 CoWoS 技术，其本质上是 interposer 尺寸的进步，由于 Si Interposer 尺寸的限制，涉及一个词叫 reticle limit，可理解为光刻机可处理的极限尺寸。也就是说即便不考虑良率和成本问题，以现有装置，一片 die 的尺寸再大也是有极限的。其第一代 CoWoS-1，所用的 interposer 尺寸已经达到大约 800mm²，第二代 CoWoS-2，通过使用一种叫 two-mask stitching photolithography 的技术，可以使得 interposer 尺寸可以达到 1200mm²，随后几代 CoWoS 封装的 interposer 尺寸稳步提升到 1700mm²，大约是 2x reticle limit。现在的第五代 CoWoS-5，通过使用一种叫 2-way lithography stitching approach 技术，可以使得 interposer 尺寸可以达到 2500mm²。

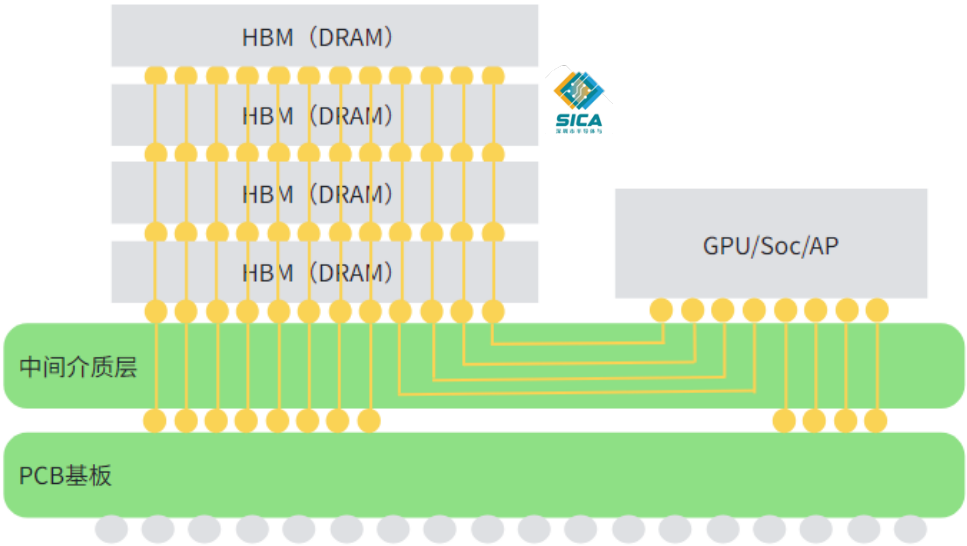


台积电的 CoWoS 路线图（图源：TSMC）

三、HBM 技术

要说将 CoWoS 技术发扬光大的还是 HBM(High Bandwidth Memory)即高带宽存储器这一先进封装，将多个 DRAM 垂直堆叠，超大容量和超高带宽可以让单算力卡的性能直线提升，以满足高性能计算、人工智能等领域对内存的严苛要求。在硅通孔(TSV)和微凸块(Bumping)封装技术，打破了传统内存带宽和功耗瓶颈，内部短距离互联 GPU 和

DRAM，不仅在最大程度上减少封装体面积，还大大缩短了信号数据的传输时间。



典型 HBM 封装示意图（图源：深芯盟）

据各大厂商透露出的数据来看，HBM 拥有比 GDDR4 倍多的带宽，可提供最高位 460GB/s 的带宽，而功耗仅为 GDDR 的二分之一，HBM 提供的现存位宽也来到的 1024bits 是 GDDR5 的 32-bits4 倍大小，虽然时钟频率 HBM 比 GDDR5 慢了不少，但是单次发送数据的 bits 位数翻了 4 倍，实际使用中的显存带宽还是远远高于 GDDR5 的。

晶圆代工尤其是先进制程始终是电子行业的“兵家必争”之地，DRAM 虽不如 CPU 话题度拉满，但是其制程迭代也是不断精进，1x、1y、1z、1α（1-alpha）、1β（1-beta）和 1γ（1-gamma），其中 1β（1-beta）节点是目前量产的最先进制程，从 Trendforce 和各大厂商透露出的资料来看，三星采用的是 1α（1-alpha），而 SK 海力士和美光采用的是 1β（1-beta）制程，大家不相伯仲，随着工艺和材料的进一步发展相信 1γ 会很快就到来。

HBM厂商	SK海力士	三星半导体	美光
2023年底 HBM产能	45k 片/月	45k 片/月	3k 片/月
2024年底 HBM产能	120-125k 片/月	130k 片/月	20k 片/月

三大厂商 HBM 产能（数据源：TrendForce，制表：深芯盟）

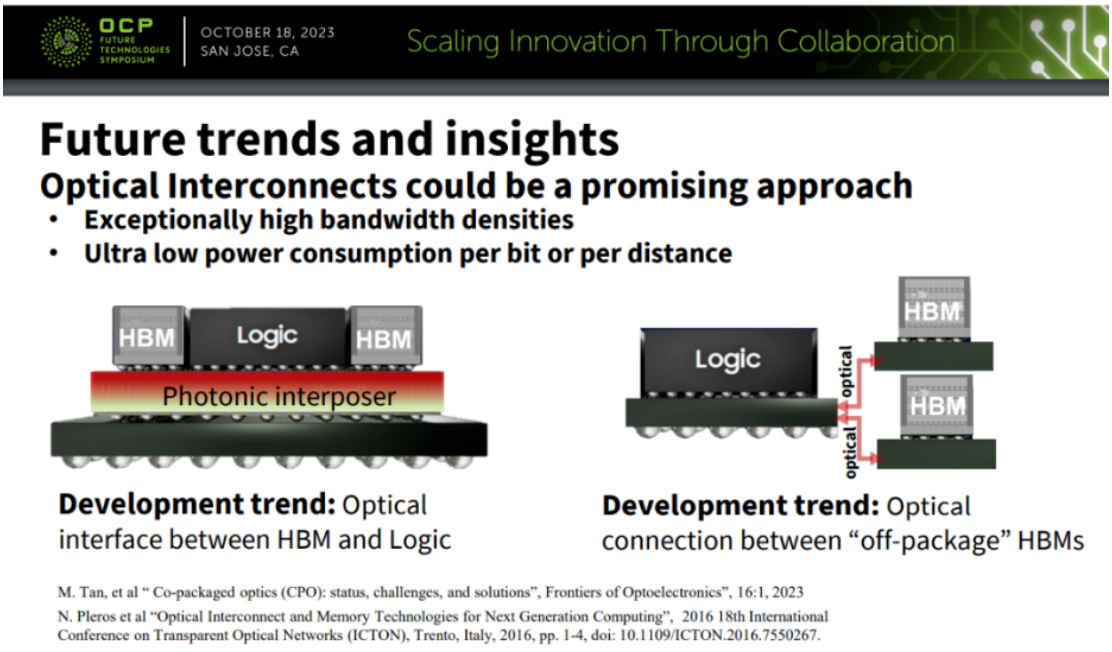
新技术一向是产能紧张，HBM 的供给面也是呈井喷式爆发，根据 TrendForce 分析师给出的产业预测报告，三星和 SK 海力士正在扩充其 HBM 的产能，翻了近 3 倍约为每月 12-13 万片，美光则稍逊一些，每月大约为 2 万片左右，但对比自身产能翻了近 7 倍，实力不容小觑。尤其是 2024 年，HBM 也已经来到了 HBM3e 的超强拓展版本。历经 HBM1、HBM2、HBM2e、HBM3、HBM3e（第五代）各个版本，现在主流量产的版本是 HBM3 的拓展（extension），其带宽、层数、容量和 I/O 速度都有明显的提升。

类别	HBM1	HBM2	HBM2E	HBM3	HBM3E
带宽(GB/s)	128	307	460	819	1125
堆叠高度(层)	1/4	4/8	4/8	8/12	8/12
容量(GB)	1/1	4/8	8/16	16/24	24/36
I/O速率(Gbps)	1	2.4	3.6	6.4	8

HBM 不同代技术参数（数据源：网络信息，制表：深芯盟）

HBM 每更新一次技术，其带宽和 I/O 速率都明显提升，其数据传输速率来到了 8Gbps，相当于 1.18TBps，我们普通家用电脑 1TB 的固态硬盘约有 931GB 的空间，写满整个硬盘花不到 1 秒钟的时间。细看下来，各家最新的 HBM3e 的极限速率还有所不同，SK 海力士提供 8Gbps，美光的提供 9.2Gbps 和 24GB 的显存，而三星的 HBM3e 则更为激进，提供高达 9.8Gbps 的 I/O 速率，整体传输速率可超过 1.2TBps，产品容量达到了 36GB。

HBM4 作为目前最先进的技术，基本也将面世，HBM4 是目前发布的 HBM3 标准的进化版，与 HBM3 相比，HBM4 计划将每个堆栈的通道数增加一倍，物理占用空间也更大，HBM4 会搭载 24Gb 和 32Gb 的内存颗粒，支持 4x-16x 的堆叠高度。其带宽扩展到 2048GB/s,部芯片接口将的微凸块间距缩小到 55μm 以下，堆叠层数也来到了最高 16 层之多，从凸块微小尺度到堆叠层数均是之前技术的瓶颈所在，据三月三星透露的消息看，国际半导体标准组织（JEDEC）同意将 HBM4 产品的标准定为 775 微米 HBM4（HBM3e 为 720 微米（μm）），但从尺寸增加来看现有堆叠也能够做到 16 层，但是复杂度和良率也是各大厂商考量的重要因素，所以新的键合技术应当是接下来几家巨头需要研究的重点课题。



三星 HBM 光互联架构（图源：2025 OCP）

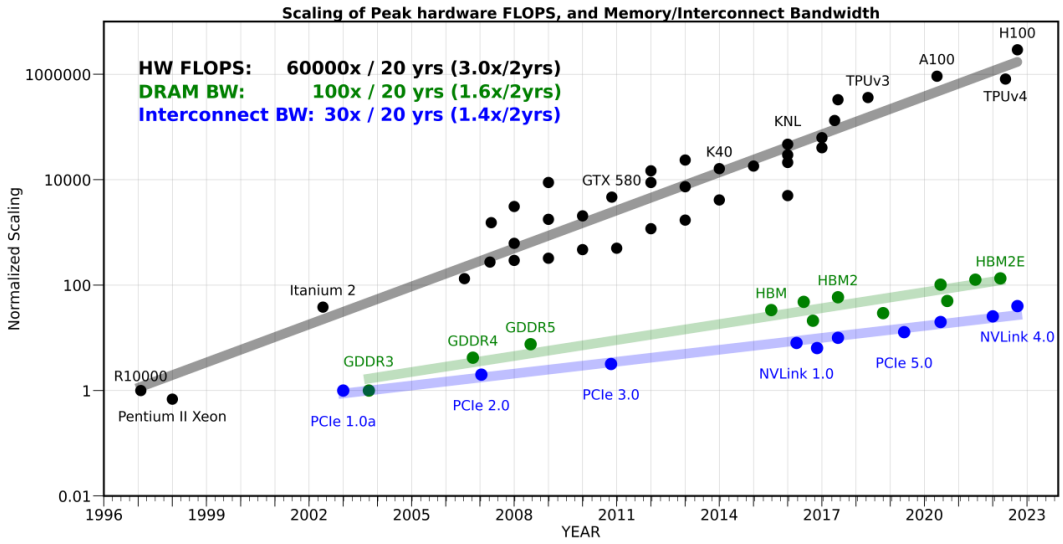
在 OCP 全球峰会上，三星提出了在 HBM 与 Logic 芯片间采用 Optical IO 技术进行数据互联，并给出了两个可能的芯片架构，如上图所示，光子在链路上的流动速度比数字信号的速度更快，而且功耗更低，不得不说 HBM 就是在

比速度，也许未来某一天会见到内部拥有光速传输信号的芯片，我们团队一致认为最先普及的应当就是 HBM 了。

四、存算一体技术

自从 OpenAI 的 ChatGPT 于 2022 年 11 月推出以来，AIGC 迅速在全球掀起一股热潮，大模型成了全球科技公司的座上宾，据统计全球现有超过千万个大模型 24 小时不停运转，其算力总需求预计到 2025 年将达 6.8 ZFLOPS（ 每秒十万京（=10^21）次的浮点运算 ）。而且算力翻倍时间在明显缩短，在大模型横空出世后，全球新的算力增长点井喷式发展，如果按照摩尔定律来进行衡量和估计，平均每十个月算力就将翻一倍，比物理尺度上晶体管翻倍还来得快。当然算力单纯堆砌并不能获得如今质的飞跃，算力中心碰见的「存储墙」和「功耗墙」这两大难题急需解决，而存算一体技术就是目前看来的破解之道。

有必要先介绍下计算机的架构--冯·诺依曼架构，计算单元-内存-存储，通过地址线和数据线相互连接，在计算过程中数据被频繁的搬来搬去，广义上就是我们经常说的 “读” 和 “写” ，这种开销即会消耗很多的能量，又浪费了很多的时间，intel 的一项研究表明，在其 7nm 制程的芯片上，数据搬运功耗高达 35pJ/bit，占总功耗的 63.7%之多，也就是上述提到的「功耗墙」问题。而「存储墙」更是棘手，如下图所示，近二十年 GPU 提升了近 10^6 倍，而内存和接口仅提升了 100 倍左右，存储器的性能越来越跟不上计算核心的性能，导致是说计算核心需要花费大量的时间来等待数据的读写。木桶原理告诉我们说，决定一个木桶能够承载的最大水量是由其最短的木板决定，整个模型的硬件系统最大瓶颈就是数据读取速度太慢，尤其是在面对深度学习和大模型领域时，这一问题是最大障碍。



Hardware FLOPs and Memory/Interconnect BandWidth（数据源： amirgholami）

“存算一体” 技术可以解决传统冯诺伊曼架构处理器所面临的两堵墙：存储墙、能耗墙;剩下的一堵「编译墙」大致可以理解为存储和计算单元之间的调用和数据搬运需要复杂的编程模型，无论是算法还是数据都需要进行一定程度的编译，而存算一体的数据状态都是编译器可以感知的，因此编译效率很高，可以绕开传统架构的「编译墙」。

既然存算一体技术如此卓越，那势必会吸引一大堆公司重金投入开发这一技术，在这条赛道上，最早是美国的 Mythic 公司在 2010 年左右推出了存算一体芯片，中国国内是 2017 年左右出现了存算一体技术的创业团队，经过近 7 年左

右的发展，目前新势力厂商诸如千芯科技、知存科技、九天睿芯和后摩智能等都在追赶第一梯队。身为第一梯队的NV、Intel、AMD 等均在近几年发布了大量新品，例如 2024 年 3 月 NVIDIA 在 GTC 宣布推出 NVIDIA Blackwell 架构 GPU 拥有 2080 亿个晶体管，采用专门定制的台积电 4NP 工艺制造。所有 Blackwell 产品均采用双倍光刻极限尺寸的裸片，通过 10 TB/s 的片间互联技术连接成一块统一的 GPU。Blackwell 架构的 GPU，作为高性能计算和 AI 加速器，参考近存计算的架构高度集成计算单元和存储单元。Intel 则在 4 月为业界带来 Gaudi® 3 AI 加速器，Gaudi 3 拥有 8 个矩阵数学引擎、64 个张量内核、96MB SRAM（每个 Tile 48MB，可提供 12.8 TB/s 的总带宽） 和 128 GB HBM2e 内存，16 个 PCIe 5.0 通道和 24 个 200GbE 链路 。在计算核心的周围，则是八个 HBM2e 内存堆栈，总容量为 128 GB，带宽为 3.7 TBps。

Intel® Gaudi® 3 AI Accelerator White Paper（图源：Intel）

Intel Gaudi® 3 AI 加速器则诠释了近存计算的精髓，内置 128GB HBM2e 内存，96MB SRAM（每个 Tile 48MB），最大程度上削减了访问延迟；内置专用 AI 计算单元针对矩阵与卷积运算进行高效优化，而这一切性能的提升都离不开存内计算技术带来的底层变革。

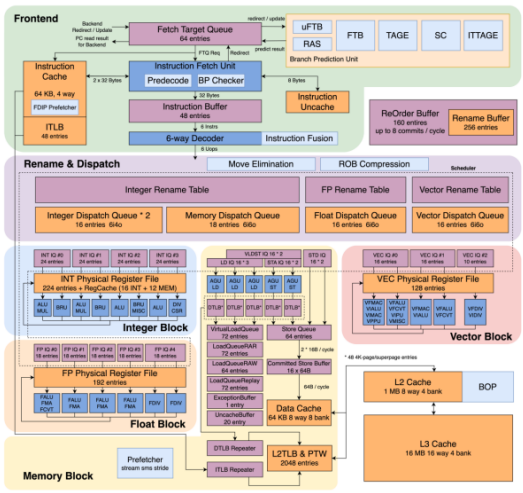
五、RISC-V 架构的高性能处理器

上面聊了很多主流科技巨头的技术，近期在业界也很火热的一个话题就是，**“开源 RISC-V 能不能高性能计算？”** RISC-V 作为发展了近 14 年的“老”架构，一直没有爆火是其生态相对较弱，纯开源可魔改，自主定义无需授权，让众多苦 ARM 久矣的厂商甘之如饴，在指令集和微架构上进行定制优化就能够打造高性能的处理器，诸如 SiFive、Andes、阿里平头哥、中科院计算所等公司发布的 RISC-V 指令架构 CPU，都摒弃了 ARM 指令架构中很多冗余的部分，专注于低功耗和高性能 AI 专用 SoC 部分设计，性能功耗比相较于同级别 ARM 核心均领先不少。

2024 年 4 月中科院计算所带队，联合北京开源芯片研究院、腾讯、阿里、中兴通讯、中科创达、奕斯伟、算能等组成了联合研发团队，发布的第三代“香山”开源高性能 RISC-V 处理器内核，是在国际上首次基于开源模式、使用敏捷开发方法、联合开发的处理器核，性能水平进入全球第一梯队，成为国际开源社区性能最强、最活跃的 RISC-V 处理器核心。

昆明湖架构总览

- 6 宽度重命名
- 多级覆盖分支预测
- 224 INT + 192 FP + 128 VEC 物理寄存器
- 160 ROB + 256 RAB（RenAme Buffer）
- 64 KB ICache / DCache
- 1 MB L2 Cache
- 48-entry ITLB / DTLB + 2048-entry L2TLB
- FDIP 指令预取
- stream / sms / stride / bop / tp 预取器



香山开源处理器昆明湖微架构（图源：OpenXiangShan）

RISC-V 作为最年级的新生代处理器架构，虽目前不能与老牌 CPU 和 GPU 在算力和性能上抗衡，绝对性能也落后于市面上的 ARM 核或 x86 核，但是作为全新和开源架构，本着开源、开放和规范一路走来，从 2021 年第一代雁栖湖架构的质疑 “性能比不过 ARM A76，低主频有何用？” 到南湖架构的流片，（南湖架构-14nm 工艺频率达到 2GHz，SPECCPU 分值达到 10 分/GHz），以及下一代昆明湖架构的紧密研发，无一例外的在彰显着中国科研团队在全球开源架构所作出的贡献。

第三代昆明湖架构拥有较上一代更多的指令集，更低功耗的 ICache，时序更优的 LSU 设计，CHI 总线的 Coupled 2，并且 L2 cache 达到 1024KB，L3 cache 最大可到 16MB，性能位于国际第一梯队。

小结：昆明湖微架构的改进

- 更多的指令集扩展
- 低功耗的 ICache
- 发射后读寄存器堆
- 时序更优的 LSU 设计
- CHI 总线的 Coupled L2

Feature	Kunminghu	Neoverse N2	Nanhu	Cortex A76
Pipeline depth	13	10	13	13
Rename width	6	5	4	4
Rename checkpoint	Y	Y	N	N
ROB size	160 (x6)	160+	192	128
ALUs	4	4	4	3
L1 instruction cache	64KB	64KB	64KB	64KB
L1 data cache	64KB	64KB	64KB	64KB
L2 cache	1024KB	512/1024KB	256KB	256/512KB
L3 cache	Up to 16MB	4MB per slice	Up to 4MB	Up to 4MB
NoC support	Y	Y	N	N
L2 outstanding txns	64	64	32	46
ITLB	48	48	32	48
DTLB	48	44	128 direct mapped	48
L2 TLB	2048	1280	2048	1280
Vector	Y	Y	Y	Y
Virtualization	Y	Y	N	Y
ECC support	Y	Y	Y	Y
PMA/PMP support	Y	Y	Y	Y
Debug support	Y	Y	Y	Y
External interface	AXI4/TL/CHI	AXI4/CHI	AXI4/TL	AXI4/CHI

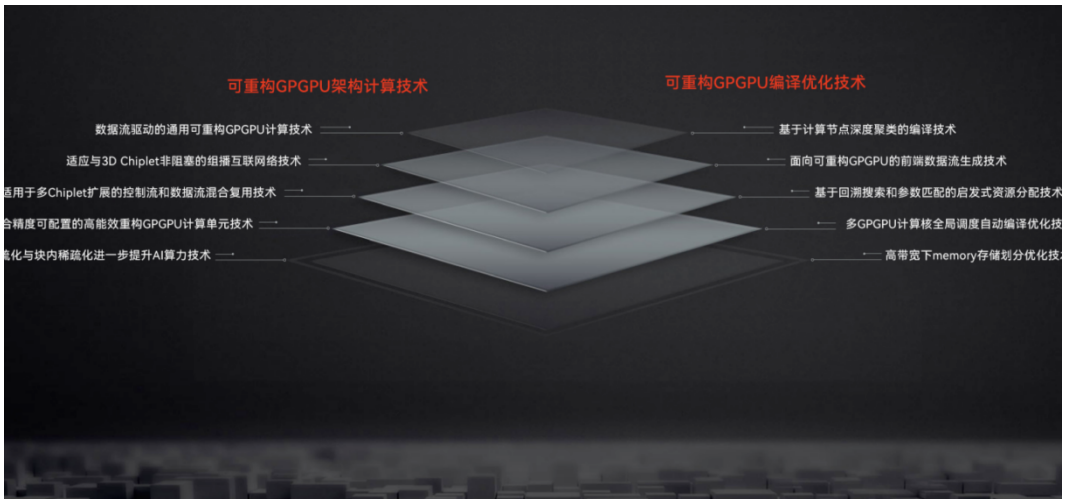
昆明湖微架构的改进（图源：OpenXiangShan）

无独有偶，SiFive 在 2024 RISC-V 欧洲峰会上宣布 SiFive Essential 系列产品的重大升级，其最新的 RISC-V 的嵌入式设备芯片拥有 8 种不同的 32 位/64 位核心配置,改进的 L2 缓存和增强的 L1 缓存,开源可定制的自由处理器 IP，在面积和功耗都绝佳的 RISC-V 架构是目前最有潜力颠覆 AI 芯片行业的种子选手。

六、主流高性能 AI 芯片性能

目前市场主流高性能 AI 算力模组往往会采用多种最新技术，例如 IC 设计上采用 chiplet 多核心、混合异构等；封装形式主要为倒装、CoWos、InFO 等；架构会采用 RISC-V、存算一体等。全球主要算力中心目前以云端算力模组为主，采买算力芯片多数还是欧美公司，诸如 Nvidia、AMD、Intel 等，随着大模型和 DeepSeek 的井喷式发展，适配 DeepSeek 的国产 AI 芯片公司开始崭露头角，目前已经适配和支持 DeepSeek R1 大模型的国产 AI 芯片公司大致分类如下：

- 云端训推：华为昇腾、海光信息、燧原科技、昆仑芯、墨芯、寒武纪
- GPGPU：沐曦、天数智芯、摩尔线程、壁仞科技、芯瞳
- 边缘推理：云天励飞、鲲云科技、瀚博、爱芯元智、江原科技
- 存算一体：灵汐科技、后摩智能
- RISC-V 架构：希姆计算、算能、进迭时空、奕斯伟计算
- 可编程架构：清微智能、芯动力



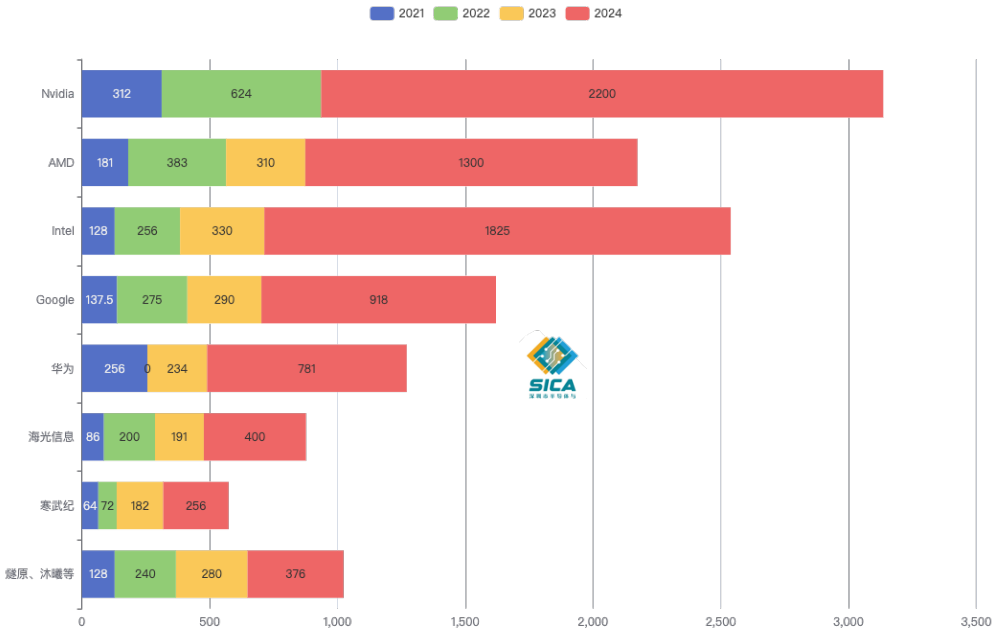
可重构技术体系（图源：清微智能）

以华为、寒武纪、昆仑芯为代表的国产算力芯片厂商开始频频发力，主流云端专用算力芯片约 40 余款，多为 7nm-14nm 制程，例如华为昇腾 910C 芯片发布于 2024 年采用 7nm，拥有 1200 亿晶体管数，其 FPLOPS16 高达 781，可直接对标 Nvidia A800-SXM 系列，由其打造的算力机组 CloudMatrix 384 性能可对标 Nvidia GB200 NVL72 服务器机组，在 PFLOPS、HBM 密度、内存带宽等关键性能指标令人眼前一亮。

Ascend C910C Chip VS GB200 Chip			
Item	Ascend 910C	GB200	Unit
BF16 TFLOPS	780	2,500	TFLOPS
HBM bandwidth	3.2	8	TB/s
HBM capacity	128	192	GB
Scale Up Bandwidth	2,800	7,200	Gb/s Uni-di
Scale Out Bandwidth	400	400	Gb/s Uni-di

Ascend 910C 和 GB200 芯片性能对比（图源：深芯盟）

寒武纪的思元系列一路从 290 的追赶到目前 590 芯片的分庭抗礼，综合性能可达 Nvidia A100 的 70%;平头哥的含光 800 推理专用芯片的 FP16 性能分达到 205，追平 Nvidia V100-PCIe ;海光信息的 K100_AI 系列芯片发布于 2023 年，采用 7nm 制程，其 FP32 高达 98，FP16 也达到 200，功耗仅 350w，能效比为 0.5，接近 V100 水平。更有沐曦科技、壁仞科技、摩尔线程等公司数十款芯片性能均达到主流算力芯片梯队。国芯科技：累积发明专利 138 项。基于 RISC-V 和 PowerPC 指令架构开发系列化高性能 CPU 内核。



主流 AI 芯片公司典型芯片 FP16 性能（图源：深芯盟）

国内厂商相较于一线大厂还是有着不小的差距，2024 年 Nvidia 发布的 GB200 芯片组拥有两个 NVIDIA B200 Tensor Core GPU 和一个 NVIDIA Grace CPU，其 FP16 达到遥遥领先于同行的 5000。

厂商	名称	发布时间	制程	晶体管数量	芯片面积	晶体管密度(百万/mm²)	FP64	FP32	TF32	BF16/FP16	INT8	FP8	INT4	FP4	功耗(W)	能效比(FP16)
第三方芯片																
英伟达	GB200	2024	4NP	4160亿			90	180	2500	5000	10000	10000		20000	2700	2
	B200			2080亿	1600mm²	130	40	80	1100	2200	4500	4500		9000	1000	2.2
	B100			1040亿			30	60	900	1800	3500	3500		7000	700	3
	H20	2023	4nm	800亿	814mm²	98	1	40	74	148	296	296			400	0.4
	H200-SXM						34	67	989	1979	3958	3958			700	2.8
	H200-NVL						30	60	835	1671	3341	3341			600	3
	H100-SXM						34	67	989	1979	3958	3958			700	3
	H100-NVL						30	60	835	1671	3341	3341			400	4.2
	H800-SXM						1	67	989	1979	3958	3958			700	3
	H800-PCIe						0.8	51	756	1513	3026	3026			350	4
	A800-SXM	2022	7nm	542亿	826mm²	66	9.7	19.5	312	624	1248				400	1.6
	A800-PCIe								156	312	624				300	1
	A100-SXM	2020							312	624	1248				400	1.6
	A100-PCIe								156	312	624				300	1
	V100-SXM2	2017	12nm	211亿	815mm²	26	7.8	15.7		125					300	0.4
	V100-PCIe						7	14		112					250	0.4
	P100-SXM2	2016	16nm	153亿	600mm²	26	5.3	10.6		21.2					300	0.1
AMD	MI325X	2024	5nm	1530亿	1017mm²	15	81.7	163.4	650	1300	2600	2600			1000	1
	MI300X	2023					61.3	122.6	653.7	1307.4	2614.9	2614.9			750	2
	MI300A			1460亿			61.3	122.6	490.3	980.6	1961.2	1961.2			550|760	1.78|1.29
	MI250X	2021	6nm	582亿	724mm²	80	47.9	47.9		383	383		383		500|560	0.76|0.68
	MI250						45	45		362	362		362		500|560	0.72|0.65
	MI210						22.6	45.3		181	181		181		300	1
英特尔	Gaudi3	2024	5nm							1835		1835			900	2
	Gaudi2	2023	7nm							432		865			600	0
Groq	LPU	2021	14nm		725mm²					188	750				275	1
大厂自研																
谷歌	TPU v7p*	2025	3nm	2744亿	890mm²	308				2307	4614	4614			959	2.4
	TPU v6e	2024	4nm	867亿	790mm²	110				918	1836				383	2.4
	TPU v5p	2023	5nm	274亿	350mm²	78				196.5	393				225	0.9
	TPU v5e		5nm	549亿	700mm²	78				459	918				537	0.9
	TPU v4	2021	7nm	312亿	780mm²	40				137.5	275				300	0.5
	TPU v4i*	2020	7nm	160亿	400mm²	40				69	138				175	0.4
	TPU v3	2018	16nm	100亿	700mm²	14				123					450	0.3
	TPU v2	2017	16nm	90亿	625mm²	14				46					280	0.2
	TPU v1	2015	28nm	30亿	330mm²	9					92				75	0.0
亚马逊	Trainium3	2024	3nm							1310		2620			728	1.8
	Trainium2	2023	5nm							667		1299			500	1.3
Meta	MTIA v2	2024	5nm	23.5亿	421mm²	6				177	354				90	2
	MTIA v1	2023	7nm	11.2亿	373mm²	3				51.2	102.4				25	2
微软	Maia 100	2023	5nm	1050亿	820mm²	128				800	1600				500	1.6
国产芯片																
华为	昇腾910C	2024	7nm							781						
	昇腾910B	2023		600亿	665mm²	90		128		376	640				310	1.2
	昇腾910	2019		130亿	456mm²	29				256	512				350	0.7
寒武纪	思元590	2024														
	MLU370-S4/S8	2022	7nm	390亿				18		72	192				75	1.0
	MLU370-X4							24		96	256				150	0.6
	MLU370-X8														250	0.4
	MLU290-M5	2021	7nm	460亿						256	512				350	0.7
	MLU270-S4*	2019	16nm							64	128		256		70	0.9
	MLU270-F4*		16nm												160	0.4
昆仑芯	P800	2025	14nm		504mm²					400						
	RG800	2021								128	256				160	0.8
	R200*										256					
	R100*										170					
平头哥	含光800*	2019	12nm	170亿	709mm²	0.2				205	825				276	0.7
海光信息	深算三号	2024	7nm													
	K100_AI	2023						98		200	400				350|400	0.5
	K100						24.5	24.5		100	200				300	0.3
	海光8100	2021		357亿											350	
燧原科技	燧原S60*	2024	12nm						128		256				350	
	邃思2.5*	2021						32	128	128	256					
	邃思2.0				3306mm²			40	160	160	320				300	0.5
	邃思1.0	2019		141亿	480mm²	29		20		80					225	0.4
摩尔线程	MTT S4000	2023	7nm					25	50	100	200				450	0.2
	MTT S3000	2022	12nm	220亿				15							250	
	MTT S2000*							10.6			42.4				150	
沐曦科技	曦云C550	2024	7nm													
	曦云C500-OAM	2023						36	140	280	560				450	0.6
	曦云C500-PCIe							30	120	240	480				350	0.7
	曦思N100*									80	160				70	1.1
壁仞科技	BR106	2023	7nm							170	340				400	0.4
	BR104	2022						128	256	512	1024				300	1.7
	BR100			770亿	1074mm²	72		256	512	1024	2048				550	1.9
天数智芯	天垓150	2024	7nm					48.0		192	384				350	0.5
	智铠100*	2022						24		96	384				150	0.6
	天垓100	2021		240亿				37		147	295				250	0.6

主流 AI 芯片公司芯片性能统计（图源：半导体综研）

主流云端 AI 算力芯片还大面积采用 HBM 显存，Nvidia 最新发布的 GB200 芯片组，显存已经采用 HBM3e 其带宽达到 16TB/s，容量为 384GB；与之对标的华为 Ascend 910C 芯片组则采用的是 HBM2e，显存带宽为 3.2TB/s，容量为 64GB，多数国产厂商为 HBM2e 和 LPDDR5 系列芯片，相比于 Nvidia 和 AMD 等一线厂商落后一到两代。

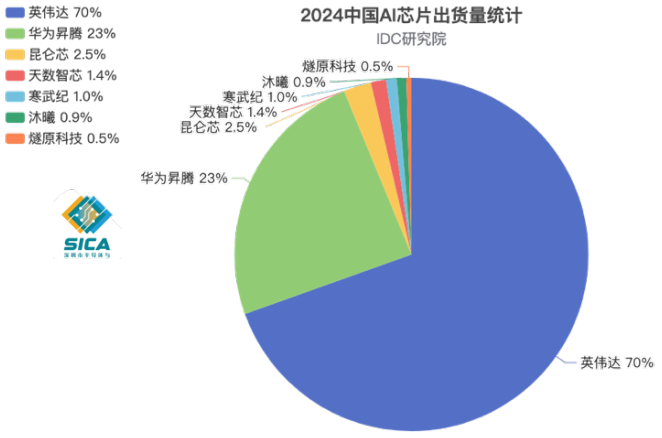
厂商	名称	发布时间	HBM/显存	显存带宽(TB/s)	显存容量(GB)	算术强度(BF16)	BF16/FP16	厂商	名称	发布时间	HBM/显存	显存带宽(TB/s)	显存容量(GB)	算术强度(BF16)	BF16/FP16
英伟达	GB200	2024	HBM3e	16	384	313	5000	华为	昇腾910C	2024	HBM2e	3.2	64	244	781
	B200			8	190	275	2200		昇腾910B	2023	HBM2/HBM2e	1.2	34	313	376
	B100					225	1800		昇腾910	2019	HBM2	1.2	34	213	256
	H20	2023	HBM3e	4.8	141	31	148	寒武纪	思元590	2024					
	H200-SXM					412	1979		MLU370-S4/S8	2022	LPDDR5	0.3	24\48	240	72
	H200-NVL					348	1671		MLU370-X4				24	320	96
	H100-SXM		HBM3	3.35	80	591	1979		MLU370-X8			0.6	48	160	
	H100-NVL			3.9	94	428	1671		MLU290-M5	2021	HBM2	1.2	32	213	256
	H800-SXM			3.35	80	591	1979		MLU270-S4*	2019	DDR4	0.1	16	640	64
	H800-PCIe			2		757	1513		MLU270-F4*					640	
	A800-SXM	2022	HBM2e	2.04	80	306	624	昆仑芯	P800	2025					400
	A800-PCIe			1.94		161	312		RG800	2021		0.5	32	256	128
	A100-SXM	2020		2.04		306	624		R200*						
	A100-PCIe			1.94		161	312		R100*		GDDR6	0.4			
	V100-SXM2	2017	HBM2	0.9	16\32	139	125	平头哥	含光800*	2019					205
	V100-PCIe					124	112		深算三号	2024					
AMD	MI325X	2024	HBM3e	6	256	217	1300	海光信息	K100_AI	2023			64		200
	MI300X	2023	HBM3	5.3	192	247	1307.4		K100						100
	MI300A				128	185	980.6		海光8100	2021	HBM2	1.0	32		
	MI250X	2021	HBM2e	3.2	128	120	383	燧原科技	燧原S60*	2024	HBM2e	0.8	64		
	MI250					113	362		邃思2.5*	2021		0.8	16	160	128
	MI210			1.6	64	113	181		邃思2.0			1.8	64	89	160
英特尔	Gaudi3	2024	HBM2e	3.7	128	496	1835		邃思1.0	2019	HBM2	0.5	16	160	80
	Gaudi2	2023	HBM2e	2.5	96	173	432	摩尔线程	MTT S4000	2023	GDDR6	0.8	48	125	100
谷歌	TPU v7p*	2025	HBM3e	7.3	192	316	2307		MTT S3000	2022		0.4	32		
	TPU v6e	2024	HBM3e	1.6	32	574	918		MTT S2000*				32		
	TPU v5p	2023	HBM3	2.8	16	70	196.5	沐曦科技	曦云C550	2024					
	TPU v5e		HBM2e	0.8	95	574	459		曦云C500	2023	HBM2e	1.8	64	156	280
	TPU v4	2021	HBM2	1.2	32	115	137.5		曦思N100*			0.5	16	160	80
	TPU v4i*	2020	HBM2e	0.3	8	230	69	壁仞科技	BR106	2023	HBM2e	1.6	64	106	170
	TPU v3	2018	HBM2	0.9	32	137	123		BR104	2022		1.6	32	320	512
亚马逊	Trainium3	2024	HBM3e				1310		BR100			2.3	64	445	1024
	Trainium2	2023	HBM3e	2.9	96	230	667	天数智芯	天垓150	2024	HBM2e	1.6	64	120	192
Meta	MTIA v1	2023	LPDDR5	0.1	8	569	51.2		智铠100*	2022	HBM2e	0.8	32	120	96
微软	Maia 100	2023	HBM2e	1.8	64	444	800		天垓100	2021	HBM2	1.2	32	123	147

主流 AI 芯片公司芯片显存统计（图源：半导体综研）

七、全球 AI 芯片出货量排行榜

2024 年，中国加速芯片的市场规模增长迅速，超过 270 万张，其中 Nvidia 以 190 万张占据 70%的市场份额，排名第二的是华为昇腾加速芯片，以 64 万张占比约 23%，昆仑芯、天数智芯、寒武纪分列三四五名。

从技术角度来看，GPU 卡占据 70%的市场份额；从品牌角度来看，中国本土 AI 芯片品牌的出货量已超过 82 万张。通过适配 DeepSeek，中国本土芯片在软件生态领域实现了突破，逐步完善软件生态。这为本土芯片在市场中的竞争力提供了有力支持。同时也促进了本土芯片厂商的技术交流和资源共享，打破了国产芯片生态建设的僵局。许多本土芯片厂商开始围绕 DeepSeek 开展合作，共同打造适配本土芯片的软件栈、工具链等生态组件。

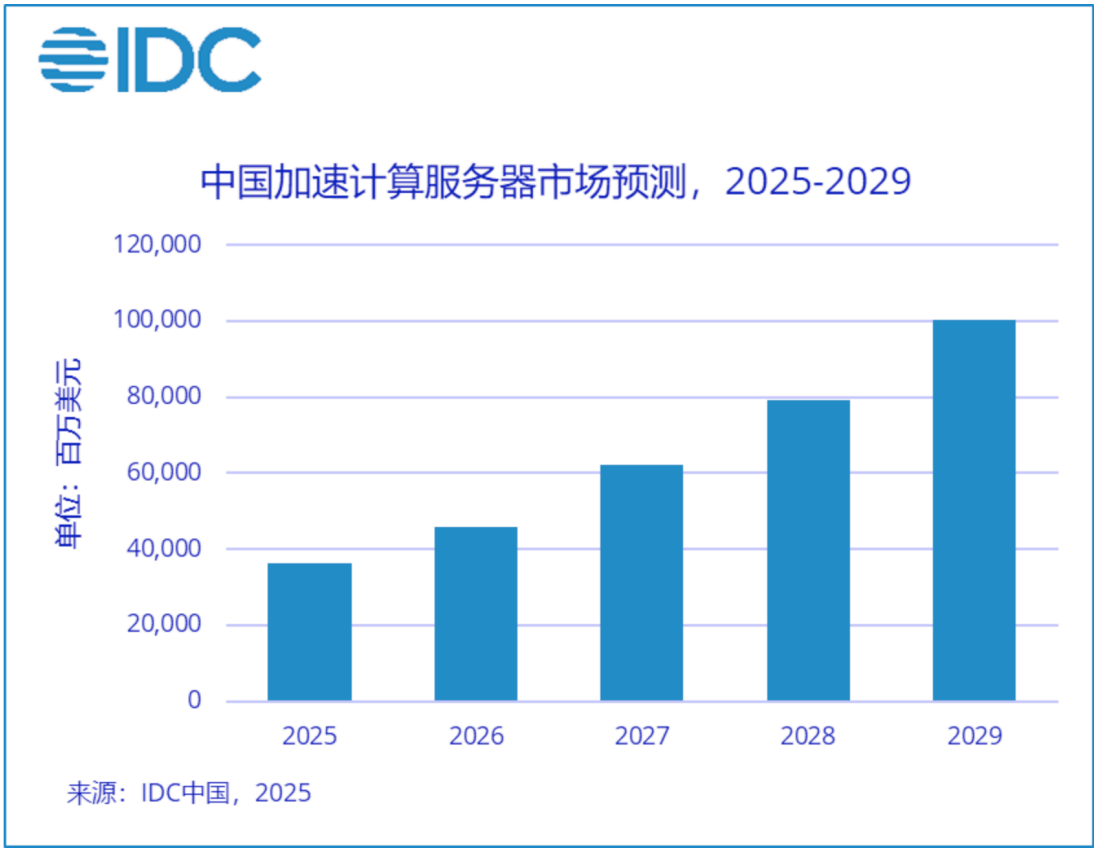


2024 年中国 AI 芯片出货量统计（数据源：IDC，统计未涵盖专供型号、客户定制等算力卡数据，不完全统计仅供参考）

IDC 数据显示，2024 年中国加速服务器市场规模达到 221 亿美元，同比 2023 年增长 134%。其中 GPU 服务器依然是主导地位，占比达到 69%。同时 ASIC 和 FPGA 等非 GPU 加速服务器高速增长，占比超过 30%。

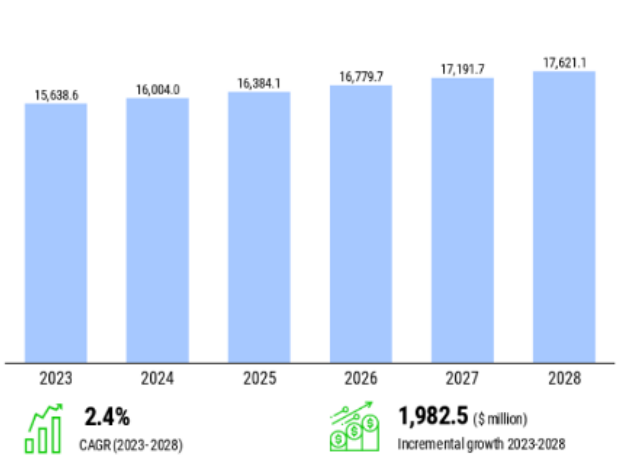
2024 年，从厂商销售额角度看，浪潮、宁畅、新华三位居前三，占据了超过 50%的市场份额；从服务器出货台数角度看，浪潮、宁畅、华为位居前三名，占总体近 55%的市场份额；从行业的角度看，互联网依然是最大的采购行业，占整体加速服务器市场超过 65%的份额，其余行业均有不同幅度的增长。

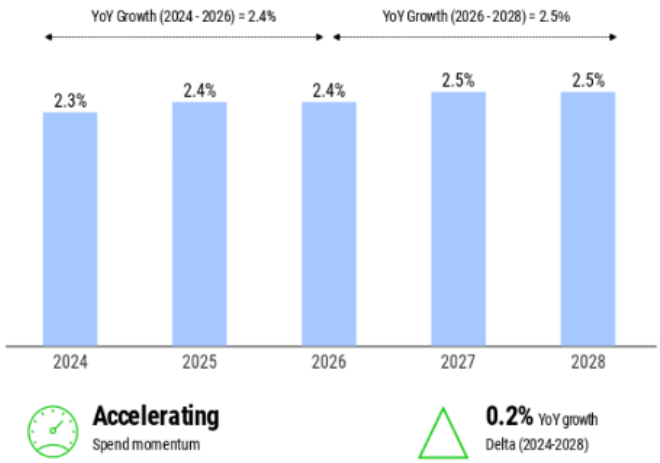
从市场环境来看，DeepSeek 通过算法优化降低了对用户 AI 算力的需求，但同时也大幅度降低了用户使用大模的门槛，创造出更多的新增市场机会。同时，大模型一体机市场有潜力实现快速增长。DeepSeek 等开源算法的推出进一步降低了部署大型模型的门槛，使企业更倾向于在本地部署大型模型，使用模型供应商训练的基本模型通过微调私有域数据来增强业务能力。IDC 预测，到 2029 年中国加速服务器市场规模将超过千亿美元。其中非 GPU 服务器市场规模将接近 50%。



中国加速计算服务器市场预测（图源：IDC）

根据 technavio 预测，全球数据中心加速计算芯片市场 2024 年总规模达到 160 亿美元，年复合增长率为 2.5%，中国本土加速计算芯片市场增长尤其迅猛，目前已经出货超 270 万张，其中 GPU 算力卡约为 190 万张，占比超 70%，值得注意的是，本土品牌厂商 AI 芯片出货量以超 82 万张，虽然目前占比仍旧不到 20%，但是相比于五年前的个位数占比已经实现年增 100%。通过适配 DeepSeek，中国本土芯片在软件生态领域实现了突破，逐步完善软件生态。这为本土芯片在市场中的竞争力提供了有力支持。同时也促进了本土芯片厂商的技术交流和资源共享，打破了国产芯片生态建设的僵局。





Data Table on Global - Market size and forecast 2023-2028 ($ million)

Year	2023	2024	2025	2026	2027	2028
Market size ($ mn)	15,638.6	16,004.0	16,384.1	16,779.7	17,191.7	17,621.1

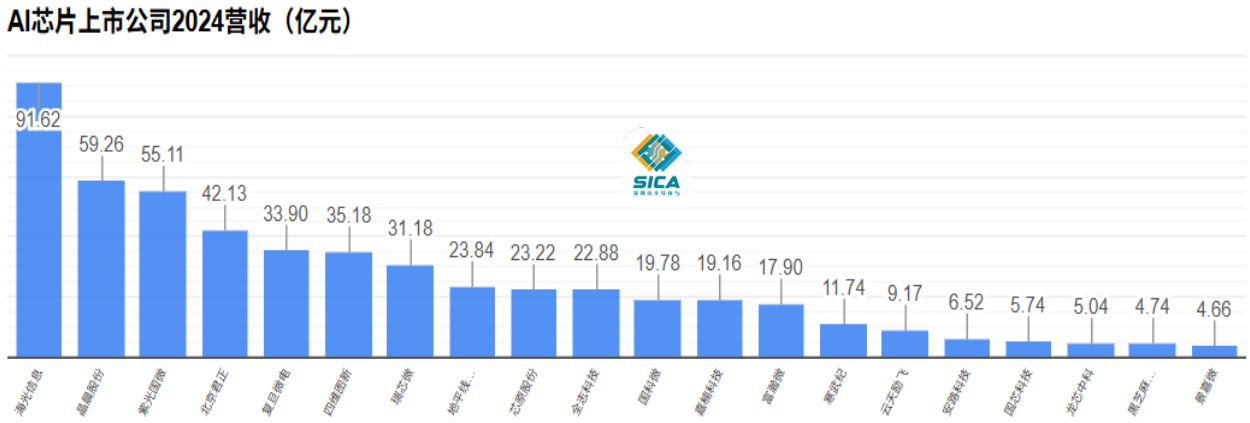
Data Table on Global Market: Year-over-year growth 2023-2028 (%)

Year	2024	2025	2026	2027	2028
Year-over-year growth	2.34%	2.38%	2.41%	2.46%	2.50%

全球加速芯片市场和年复合增长率预测（图源：technavio）

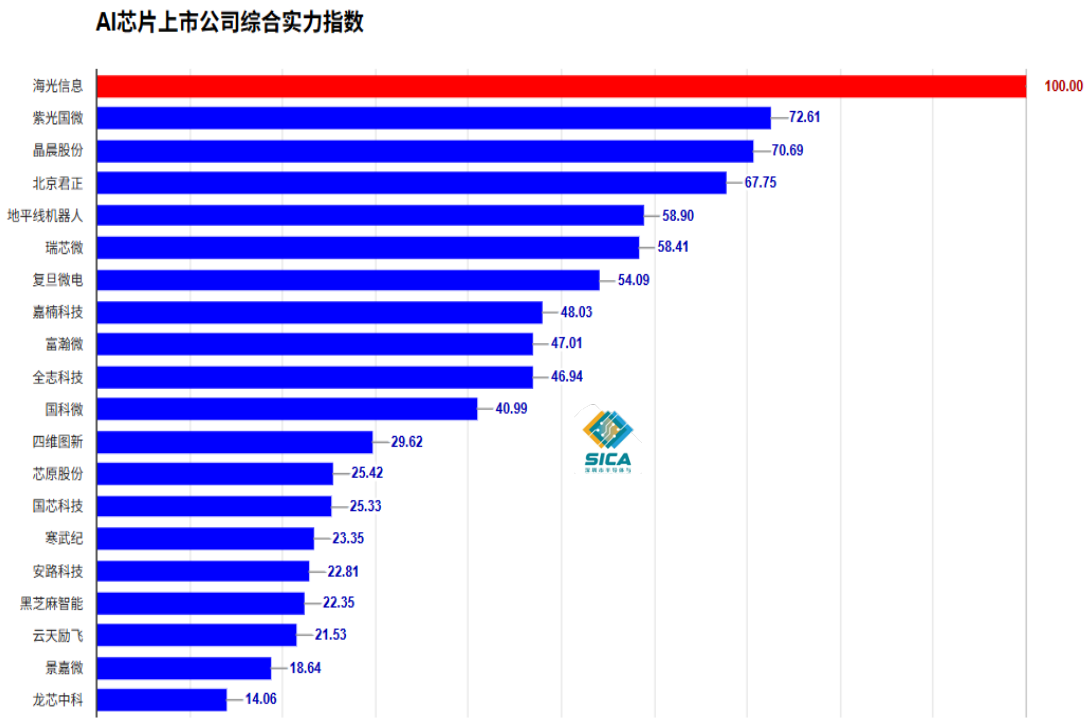
八、国产 AI 芯片和处理器上市公司综合排名

根据公开资料显示，国产 AI 芯片和处理器公司 2024 年营收排名，海光信息、晶晨股份和紫光国微为前三甲，紧随其后的是北京君正和复旦微电，此后不再赘述见下图。



20 家上市公司的 2024 财年营收对比（在港交所或海外上市的公司营收金额统一换算为人民币元）（来源：深芯盟）

根据深芯盟自研量化模型结合上市公司财报和未来发展潜力，独家发布国产 AI 芯片和处理器上市公司综合实力指数。



AI 芯片上市公司综合实力指数（图源：深芯盟）

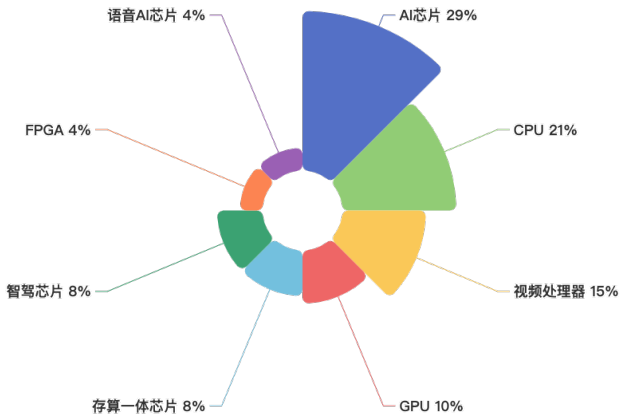
综合实力指数是多个指标加权而成的，包括营收、利润，以及研发人员人均创收等。其中，营收的权重占比最大，因此综合实力指数跟营收的相关性比较强。但对于营收相差不大的公司来说，利润和人均创收的指标就起到决定性作用了。

九、74 家国产 AI 芯片和处理器厂商信息汇总

深芯盟分析师团队汇总 74 家国产 AI 芯片和高性能处理器厂商分为技术亮点，企业简介和应用场景概括性介绍厂商信息。

按芯片类别划分的厂商数量对比：

芯片类别	厂商数量	厂商占比
AI芯片	23	31%
CPU	15	20%
视频处理器	11	15%
GPU	7	9%
存算一体芯片	6	8%
智驾芯片	6	8%
FPGA	3	4%
语音AI 芯片	3	4%
总计	74	100%



深芯盟统计国内 74 家 AI 芯片分类(图源：深芯盟)

专用 AI 芯片和 CPU 的研发和生产占据国内 74 家厂商的五成产品类别，细分视频处理专用芯片、GPU、存算一体芯片也是近五年的后起之秀，且多数厂商主营产品涵盖多种类别，现取其主流产品和厂商业务侧重进行分类。

企业总部	企业数量	企业占比
上海	20	27.0%
北京	16	21.6%
深圳	8	10.8%
杭州	7	9.5%
成都	4	5.4%
珠海	4	5.4%
福建	2	2.7%
合肥	2	2.7%
苏州	2	2.7%
天津	2	2.7%
长沙	2	2.7%
安徽	1	1.4%
广州	1	1.4%
湖北	1	1.4%
南京	1	1.4%
重庆	1	1.4%

按厂商总部所在地划分的厂商数量对比（图源：深芯盟）

上海、北京和深圳作为国内一线城市吸引了大量的资金和人才，近六成厂商总部选择与此，芯片生态链也有地区之分，上海和北京高校众多，更加容易诞生初创和有创新活力的公司，所以近二十年来主要是高科技企业的设计研发中心，深圳依托于制造业和经济活性也诞生了不少芯片公司，其制造材料厂商众多和物流成本低廉以及产业工人人力资源雄厚都是得天独厚的存在。剩下近四成厂商则如天女散花一般分落于全国次一线和二线城市，依托于各级省市补贴和劳动力相对成本低，也获得了不错的发展。

公司简称	公司全称	企业总部	细分类别	公司简称	公司全称	企业总部	细分类别
埃瓦智能	上海埃瓦智能科技有限公司	上海	AI芯片	千芯科技	千芯科技（北京）有限公司	北京	存算一体芯片
爱芯元智	爱芯元智半导体股份有限公司	上海	AI芯片	清微智能	北京清微智能科技有限公司	北京	GPU
安路科技	上海安路信息科技股份有限公司	上海	FPGA	全志科技	珠海全志科技股份有限公司	珠海	CPU
北京君正	北京君正集成电路股份有限公司	北京	视频处理器	锐思智芯	北京锐思智芯科技有限公司	北京	视频处理器
比特大陆	北京比特大陆科技有限公司	北京	存算一体芯片	瑞芯微	瑞芯微电子股份有限公司	福建	AI芯片
壁仞科技	上海壁仞智能科技有限公司	上海	GPU	睿思芯科	睿思芯科（深圳）技术有限公司	深圳	CPU
登临科技	上海登临科技有限公司	上海	AI芯片	申威科技	成都申威科技有限责任公司	成都	CPU
地平线	北京地平线信息技术有限公司	北京	智驾芯片	时擎科技	时擎智能科技(上海)有限公司	上海	CPU
飞腾信息	飞腾信息技术有限公司	天津	CPU	时识科技	成都时识科技有限公司	成都	CPU
复旦微电	上海复旦微电子集团股份有限公司	上海	CPU	视海芯图	成都视海芯图微电子有限公司	成都	视频处理器
富瀚微电子	上海富瀚微电子股份有限公司	上海	视频处理器	四维图新	北京四维图新科技股份有限公司	北京	智驾芯片
国科微电子	国科微电子股份有限公司	长沙	AI芯片	算能科技	厦门算能科技有限公司	福建	AI芯片
国芯科技	苏州国芯科技股份有限公司	苏州	CPU	燧原科技	上海燧原科技有限公司	上海	AI芯片
海光信息	海光信息技术股份有限公司	天津	CPU	探境科技	北京探境科技有限公司	北京	AI芯片
寒武纪	中科寒武纪科技股份有限公司	北京	AI芯片	天数智芯	上海天数智芯半导体有限公司	上海	GPU
瀚博半导体	瀚博半导体（上海）有限公司	上海	GPU	微纳核芯	杭州微纳核芯电子科技有限公司	杭州	AI芯片
杭州国芯	杭州国芯微电子股份有限公司	杭州	视频处理器	物奇微	重庆物奇微电子股份有限公司	重庆	CPU
黑芝麻智能	黑芝麻智能科技（上海）有限公司	上海	智驾芯片	曦智科技	上海曦智科技有限公司	上海	AI芯片
后摩智能	南京后摩智能科技有限公司	南京	存算一体芯片	芯驰科技	北京芯驰半导体科技股份有限公司	北京	智驾芯片
华为海思	深圳市海思半导体有限公司	深圳	AI芯片	芯动力科技	珠海市芯动力科技有限公司	珠海	CPU
嘉楠科技	北京嘉楠捷思信息技术有限公司	杭州	视频处理器	芯砺智能	芯砺智能科技（上海）有限公司	上海	CPU
晶晨半导体	晶晨半导体（上海）有限公司	上海	CPU	芯明智能	合肥芯明智能科技有限公司	合肥	AI芯片
景嘉微电子	长沙景嘉微电子股份有限公司	长沙	GPU	芯擎科技	湖北芯擎科技有限公司	湖北	智驾芯片
九天睿芯	深圳市九天睿芯科技有限公司	深圳	存算一体芯片	芯原股份	芯原微电子（上海）股份有限公司	上海	AI芯片
酷芯微	合肥酷芯微电子有限公司	合肥	AI芯片	依图科技	上海依图网络科技有限公司	上海	视频处理器
昆仑芯科技	昆仑芯（北京）科技有限公司	北京	FPGA	亿智电子	珠海亿智电子科技有限公司	珠海	AI芯片
鲲云科技	深圳鲲云信息科技有限公司	深圳	AI芯片	亿铸科技	苏州亿铸智能科技有限公司	苏州	存算一体芯片
蓝芯算力	蓝芯算力(深圳)科技有限公司	深圳	AI芯片	奕行智能	奕行智能科技（广州）有限公司	广州	智驾芯片
灵汐科技	北京灵汐科技有限公司	北京	AI芯片	云豹智能	深圳云豹智能有限公司	深圳	AI芯片
聆思智能	安徽聆思智能科技有限公司	安徽	语音AI芯片	云天励飞	深圳云天励飞技术股份有限公司	深圳	视频处理器
龙芯中科	龙芯中科技术股份有限公司	北京	CPU	肇观电子	上海肇观电子科技有限公司	上海	视频处理器
每刻深思	每刻深思智能科技（北京）有限责任公司	北京	视频处理器	知存科技	杭州知存算力科技有限公司	杭州	存算一体芯片
摩尔线程	摩尔线程智能科技（北京）有限责任公司	北京	GPU	智芯科微	杭州智芯科微电子科技有限公司	杭州	语音AI芯片
墨芯人工智能	墨芯人工智能科技(深圳)有限公司	深圳	AI芯片	中昊芯英	中昊芯英(杭州)科技有限公司	杭州	AI芯片
沐曦集成电路	沐曦集成电路（上海）有限公司	上海	GPU	中星微	中星微技术股份有限公司	珠海	视频处理器
平头哥半导体	平头哥半导体有限公司	杭州	AI芯片	紫光国微	紫光国芯微电子股份有限公司	北京	FPGA
启英泰伦	成都启英泰伦科技有限公司	成都	语音AI芯片	紫光展锐	紫光展锐（上海）科技有限公司	上海	CPU

74 家国产芯片厂商信息纵览（图源：深芯盟）

十、74 家国产 AI 芯片和处理器厂商信息汇总

埃瓦智能

技术亮点：埃瓦科技是一家专注于 AIOT 智能计算芯片、AI 算法及 3D 智能视觉应用基础技术研发的高新技术企业，累计申请知识产权百余件

企业简介：公司成立于 2018 年总部位于上海机器人谷，历经多次融资，目前已经在深圳、西安设立分支机构，产品涵盖 3D AI 芯片、机器视觉模组等各类图像识别芯片。

应用场景：智能车载视觉、智能机器人、无人机、门锁/门禁、VR/AR、智能安防、扫地机、智能家居等人工智能落地场景

爱芯元智

技术亮点：公司专注于人工智能感知与边缘计算芯片的设计、研发与服务，爱芯元智自研两大核心技术——爱芯智眸 AI-ISP 和爱芯通元混合精度 NPU

企业简介：公司成立于 2019 年 5 月，并且自研两大先进技术，业内领先的 AI-ISP 自研 IP（爱芯智眸 AI-ISP）和爱芯通元混合精度 NPU 采用多线程异构多核设计，实现了算子、网络微结构、数据流和内存访问优化，公司四年时间发布四代芯片产品的量产，发展势头十分迅猛。

应用场景：智慧城市、智能驾驶、机器人以及 AR/VR 等巨大的边缘和端侧设备市场。

安路科技

技术亮点：公司具备 FPGA 芯片硬件和 FPGA 编译软件的自主研发能力，专注于研发通用可编程逻辑芯片技术及系

统解决方案，目前拥有 376 项知识产权申请。

企业简介：创立于 2011 年 11 月，是国内领先的集成电路设计企业，于 2021 年在上交所科创板成功上市，成为 A 股首家专注于 FPGA 业务的上市公司。

应用场景：工业控制、消费电子、医疗设备、网络通信、汽车电子等领域

北京君正

技术亮点：君正在处理器技术、多媒体技术和 AI 技术等计算技术领域持续投入，自研了多核异构跨界处理器—X2000、T41：普惠 AI 视频处理器，搭配自研 Ingenic AIE 技术专为边缘设备而设计，旨在提供高效的深度学习推理能力；还自研了 Magik 平台，专注于端侧 AI 全栈式开发，平台集模型量化训练等任务场景

企业简介：君正集成电路成立于 2005 年，基于创始团队创新的 CPU 设计技术，迅速在消费电子市场实现 SoC 芯片产业化，2020 年，君正完成对美国 ISSI 的收购，将整合其积累十几年的计算技术，及 ISSI 三十余年的存储、模拟和互联技术成为国内首屈一指的 CPU、DRAM 和 NAND 设计公司，并于 2011 年 5 月公司在深圳创业板上市（300223）

应用场景：智能视频监控、智能音频、驾驶舱内 dms aiot 设备、Roomba 机器人、AIoT、工业和消费、生物识别及教育电子领域

比特大陆

技术亮点：比特大陆拥有独特的算力能效比技术，先后发布几十款算力产品，其产品 BM 系列 TPU(Tensor Processing Unit 张量计算单元)芯片拥有高度定制的 BMDNN Chip Link Subsystem，集成海量 NPU 单元，功耗极低算力超强。

企业简介：比特大陆成立于 2013 年，是全球领先的区块链服务器厂商；目前主要产品市场占有率全球第一，在全球 14 个国家和地区设立研发中心，客户遍及 100 多个国家和地区。

应用场景：区块链、大算力机房、超算集群、专门用于图像/视频处理方向、人工智能中的深度学习加速

壁仞科技

技术亮点：壁仞科技致力于开发原创性的通用计算体系，建立高效的软硬件平台，同时在智能计算领域提供一体化的解决方案，实现国产高端通用智能计算芯片的突破

企业简介：壁仞科技创立于 2019 年，团队由国内外芯片和云计算领域核心专业人员、研发人员组成，在 GPU、DSA（专用加速器）和计算机体系结构等领域具有深厚的技术积累和独到的行业洞见

应用场景：GPU 集群、AI 算力中心、人工智能训练、机器推理、智能计算等多种需要大算力的场景。

登临科技

技术亮点：登临科技是国内首家完全凭借自主创新，构建 GPGPU 核心技术的云端 AI 计算平台公司。登临科技的 GPU + 系列产品开创了新一代 AI 通用处理器/加速器的先河

企业简介：登临科技，成立于 2017 年底，专注于芯片研发与技术创新，致力于打造云边端一体、软硬件协同的前沿芯片产品和平台化基础系统软件，作为国内首个实现规模化商业落地的 GPU 企业，登临首款基于 GPU+的创新 AI 计算加速器-Goldwasser 已规模化运用在各个应用场景

应用场景：数据中心、互联网、计算机视觉、自然语言处理、人工智能、算力中心等大算力复杂场景

地平线

技术亮点：地平线自主研发兼具极致效能与高效灵活的智能计算方案，开发出了创新性的智能计算架构 BPU（Brain Processing Unit），明星产品征程系列、旭日系列等采用自研的 BPU® 贝叶斯、伯努利架构设计，兼具高性能、低功耗等特点，算力强大，安全可靠

企业简介：公司成立于 2015 年，次年发布自研 BPU 架构，目前已经拥有高斯、贝叶斯、伯努利等多种架构，智能计算方案性能遥遥领先，是市场领先的乘用车高级辅助驾驶（ADAS）和高阶自动驾驶（AD）解决方案供应商，目前智驾方案已经在多家整车厂交付量产。

应用场景：高级别自动驾驶及智能座舱量产、边缘计算和智能前视市场、智能座舱多模人机交互 HMI

飞腾信息

技术亮点：公司致力于通算处理器、智算处理器等高端芯片的研发设计和产业化推广，始终围绕国家战略需求和重大工程，不断推出高性能服务器 CPU、高效能桌面 CPU、高端嵌入式 CPU 和飞腾 XPU 系列四大系列

企业简介：是 CPU 研发设计的国家队、国内领先的自主核心芯片提供商，由中国电子信息产业集团、天津市滨海新区政府和天津先进技术研究院 2014 年联合支持成立。目前公司总部设在天津，在北京、长沙、成都、广州和深圳设有子公司

应用场景：应用覆盖各种类型的终端、服务器、存储、网络、安全和嵌入式等设备

复旦微电

技术亮点：致力于超大规模集成电路的设计、开发和提供系统解决方案;是国内从事超大规模集成电路的设计、开发、生产（测试）和提供系统解决方案的专业公司,现已建立健全安全与识别芯片、非挥发存储器、智能电表芯片、FPGA 芯片和集成电路测试服务等产品线

企业简介：司于 1998 年 7 月创办，并于 2000 年在香港上市，2014 年转香港主板，是国内成立最早、首家上市的股份制集成电路设计企业。2021 年登陆上交所科创板，形成“A+H”资本格局。

应用场景：金融、社保、汽车电子、城市公共交通、电子证照、移动支付、防伪溯源、智能手机、安防监控、工业控制、信号处理、智能计算等众多领域

富瀚微电子

技术亮点：专注于以视频为核心的智慧视频、智能家居、汽车电子领域芯片的设计开发，为客户提供高性能视频编解码 SOC 芯片、图像信号处理器 ISP 芯片及完整的产品解决方案，以及提供技术开发、IC 设计等专业技术服务。

企业简介：公司成立于 2004 年 4 月，并于 2027 年深圳创业板上市，控股子公司眸芯科技、仰歌电子，并在成都、深圳、杭州等地设立分支机构，芯片累计出货量达 3.5 亿颗，视频编码芯片、图像 ISP 芯片市占率领先。

应用场景：智慧视频、智能家居、汽车电子领域芯片

国科微电子

技术亮点：公司拥有大规模集成电路及解决方案开发；在先进制程工艺的芯片及其终端产品上积累了大量知识产权，具备了快速研发及量产 SoC 芯片能力，并且凭借在底层算力和工具链等方面的深厚技术积累，国科微自主研发并成功推出神经网络处理器（NPU），实现全场景算力布局

企业简介：成立于 2008 年，总部位于长沙，并在北京、上海、深圳、杭州、成都、济南等地设有分子公司及研发中心。公司是国内重点集成电路设计企业，国家知识产权示范企业

应用场景：智慧超高清、智慧视觉、人工智能、车载电子等领域

国芯科技

技术亮点：致力于服务安全自主可控的国家战略，为国家重大需求和市场需求领域客户提供 IP 授权、芯片定制服务和自主芯片及模组产品；自主芯片及模组产品现阶段以信息安全类为主，聚焦于“云”到“端”的安全应用

企业简介：公司成立于 2001 年，是一家聚焦于国产自主可控嵌入式 CPU 技术研发和产业化应用的芯片设计公司；IP 授权与芯片定制服务基于自主研发的嵌入式 CPU 技术，为实现三大应用领域芯片的安全自主可控和国产化替代提供关键技术支撑

应用场景：信息安全、汽车电子和工业控制、边缘计算和网络通信三大关键领域

海光信息

技术亮点：海光处理器兼容市场主流的 x86 指令集，具有成熟而丰富的应用生态环境。 海光处理器内置专用安全硬件，支持多种先进的漏洞防御技术，内置高性能的国密协处理器和密码指令集

企业简介：公司成立于 2014 年，主要从事高端处理器、加速器等计算芯片产品和系统的研究、开发，拥有海光深度计算处理器，x86 中央处理器等明星产品，支持主流通用并行计算架构。

应用场景：互联网、电信、金融、交通、能源、中小企业等行业的广泛应用需求。

寒武纪

技术亮点：公司自研全新 MLUv02 架构，基于信用卡大小的模组上可以实现 16TOPS AI 算力的单系统解决方案，打造出 MLU370-S4 智能加速卡、MLU220-SOM 模组等系列产品，功耗极低，性能强大；搭配自研推理加速引擎 MagicMind 可实现模型快速高效部署

企业简介：寒武纪成立于 2016 年，专注于人工智能芯片产品的研发与技术创新，致力于打造人工智能领域的核心处理器芯片，提供云边端一体、软硬件协同、训练推理融合、具备统一生态的系列化智能芯片产品和平台化基础系统软件。

应用场景：互联网、金融、交通、能源、电力和制造等领域的复杂 AI 应用场景提供充裕算力，推动人工智能赋能产业升级

瀚博半导体

技术亮点：瀚博半导体为人工智能核心算力和图形渲染、内容生成、AIGC 提供全栈式芯片解决方案；目前拥有自主研发的核心 IP 以及两代 GPU 芯片，提供适用于通用 AI 计算和图形渲染的 GPU 产品。

企业简介：瀚博半导体是一家高端 GPU 芯片提供商，成立于 2018 年 12 月，注册地在中国上海；瀚博凭借前沿的自主原创架构、强大的软硬件融合开发能力以及丰富的设计经验研发出高质量的 GPU 产品，瀚博两代芯片现已量产并商业化落地

应用场景：人工智能与渲染产业，助力大模型与生成式人工智能、智算中心、智慧工业、智慧交通、数字孪生、工业软件、云渲染等应用落地。

杭州国芯

技术亮点：拥有自主研发的神经网络处理器、指令集及编译器等核心技术，GX 系列 AI 芯片搭载自研的第二代神经网络处理器 gxNPU V200 和自研的硬件 VAD，专为人工智能和物联网应用设计的嵌入式设计，独特地设计为多核异构架构，集成自主产权的 NPU 神经网络处理器

企业简介：公司成立于 2001 年，总部位于杭州，专注于数字电视及物联网人工智能领域的芯片设计和系统方案开发。

公司开发的数字电视芯片产品已遍布全球，同时公司深耕人工智能领域，率先推出多款面向物联网的人工智能芯片，目前公司进入上市辅导阶段。

应用场景：TWS 耳机、智能手表、智能眼镜、运动手环、网红风扇、语音电视、白色家电、智能车载、麦克风阵列

黑芝麻智能

技术亮点：黑芝麻智能是领先的车规级智能汽车计算芯片及基于芯片的解决方案供应商，拥有用于自动驾驶的华山系列高算力芯片和武当系列跨域计算芯片，自行研发的 IP 核、算法和支持软件驱动的 SoC 和基于 SoC 的解决方案，提供全栈式服务自动驾驶应用场景。

企业简介：公司成立于 2016 年，是国内首家集齐了功能安全专家认证的企业+功能安全流程认证+产品认证的自动驾驶芯片公司；目前在武汉、硅谷、上海等地成立研发和销售中心，自动驾驶芯片已经出货几十家整车厂商，是国内领先的智驾和域控制器设计公司。

应用场景：为 L3/L4 级别自动驾驶提供多场景解决方案

大算力架构支持 L3/L4 高级别自动驾驶功能，实现从停车场泊车， 城市内部，到高速道路等多场景的完美无缝衔接

后摩智能

技术亮点：基于先进的存算一体技术和存储工艺，后摩智能致力于突破芯片的性能与功耗瓶颈，提供的大算力、高能效比芯片及解决方案

企业简介：后摩智能创立于 2020 年，是全球存算一体 AI 芯片的先行者，为了加速人工智能技术的普惠落地，自研了，大算力存算一体技术架构，解决了多种工程难题，计算效率大幅提升，同时可靠性达到了车规级标准。

应用场景：应用于 AI PC 等边端大模型以及智能驾驶、智能工业等场景。

华为海思

技术亮点：海思技术有限公司是一家全球领先的半导体与器件设计公司，致力于打造安全可靠、性能领先的芯片与板级解决方案，其手机 SOC、移动通信芯片等领域为华为公司提供多种解决方案。

企业简介：出身于华为集团，现海思独立运营，在全球设有 12 个能力中心，自有核心技术涵盖全场景联接、全域感知、超高清视音频处理、智能计算、芯片架构和工艺、高性能电路设计及安全等，目前拥有 200 余项自主知识产权芯片和 8000 余项专利技术。

应用场景：联接、智慧视觉、智慧媒体、显示交互、MCU、智能感知、模拟、光模块、激光显示、消费电子、智慧家庭、汽车电子等行业智能终端

嘉楠科技

技术亮点：嘉楠科技是一家领先的 ASIC 芯片设计公司，以“区块链+AI”为多元化经营战略，业务范围涵盖高性能 ASIC 计算芯片及设备研发、AI 芯片及产品开发，其勘智 K 系列采用 RISC-V 架构算力强悍。累计拥有 191 项专利知识产权、117 项 IC 布图设计权和 71 项软件著作权

企业简介：作为一家纳斯达克上市公司，嘉楠科技是全球“区块链第一股”，也是第一家在美上市的中国自主知识产权 AI 芯片公司。

应用场景：在人体骨骼识别，动态手势多维摄像头处理有着强大的算力优势，广泛应用于智能家居、智能园区、智能能耗和智能农业等场景

晶晨半导体

技术亮点：晶晨半导体是全球布局、国内领先的无晶圆半导体系统设计厂商，拥有丰富的 SoC 全流程设计经验，坚持超高清多媒体编解码和显示处理、内容安全保护、系统 IP 等核心软硬件技术开发，整合业界领先的 CPU/GPU 技术和先进制程工艺，提供各类多媒体 SoC 芯片和系统级解决方案，产品技术先进性和市场覆盖率位居行业前列

企业简介：晶晨半导体成立于 1995 年，总部位于美国硅谷核心区山景城，并在全球多个科技与产业枢纽设有分支机构，遍及美国加州 Santa Clara、上海、深圳、中国台北及中国香港，并于 2019 年登陆科创板，名称为晶晨股份。

应用场景：智能机顶盒、智能电视、音视频系统终端、无线连接及车载信息娱乐系统等

景嘉微电子

技术亮点：致力于信息探测、处理与传递领域的高新技术开发和综合应用，拥有国内领先的自主研发高性能 GPU 图形显示芯片能力，主要从事高可靠电子产品的研发、生产和销售，产品涵盖芯片、模块、整机、系统等多种形态。

企业简介：景嘉微是国家重点高新技术企业、国家重点集成电路设计企业、国家技术创新示范企业、国家专精特新“小巨人”企业，公司于 2016 年 3 月在深交所创业板成功上市，股票代码 300474。

应用场景：GPU 芯片、图形显控模块、无线通信系统、小型化雷达、计算存储设备

九天睿芯

技术亮点：专注于研究神经拟态感存算一体架构，九天睿芯芯片基于类脑计算，以模数混和形式，实现感存算一体芯片的研发落地，量产销售。

企业简介：九天睿芯科技 2018 年成立，总部成立于深圳，先后成立深圳，成都，上海，瑞士四个办公点， 成立三年完成融资近亿，销售额近 5000 万。

应用场景：应用于低功耗无线摄像头/ARVR/手机平板、86 开关等智能家居产品

酷芯微

技术亮点：依托智能感知、智能计算、智能传输三大核心技术，通过自主研发芯片核心架构、核心 IP，提供专用于人工智能的高性能低功耗芯片及相关工具链解决方案。

企业简介：公司成立于 2011 年 7 月，致力于成为全球智能芯片领导者，总部位于合肥高新区，并在上海、成都、深圳等多地开设分、子公司。

应用场景：主要应用于智能安防、智能硬件、智能车载等多个领域。

昆仑芯科技

技术亮点：早在 2011 年公司跨入 AI 加速芯片行业，专注于自研 FPGA AI 加速器，自演了昆仑芯 XPU 架构，陆续发布昆仑芯 1 代、2 代 AI 芯片；昆仑芯科技拥有 400 余项发明专利申请，授权近百项，软件著作权多项

企业简介：昆仑芯前身为百度智能芯片及架构部，成立与 2021 年四月，团队在国内最早布局 AI 加速领域，深耕十余年，是一家在体系结构、芯片实现、软件系统和场景应用均有深厚积累的 AI 芯片企业。

应用场景：智能计算、互联网、智慧工业、智慧金融、智慧交通、智慧物流、智慧园区

鲲云科技

技术亮点：鲲云科技是全球领先的人工智能算力供应商，以开创性数据流 AI 芯片技术为核心，致力于提供高性能、低延时、高算力性价比的下一代人工智能计算平台，100 余项重磅奖项，400 余项专利申请。

企业简介：鲲云科技成立于 2017 年，提供算力、算法、平台一体化的智能解决方案，助力 2000+终端用户完成智能化转型升级，加速人工智能技术落地。

应用场景：能源、化工、电力、城市等行业领域

蓝芯算力

技术亮点：专注于为数据中心应用场景设计高性能芯片的公司，致力于构建创新的产品、与潜在数据中心客户一起设计高效、可靠和安全的服务器解决方案

企业简介：公司成立于 2023 年，在上海、和北京设立分公司，专注于高性能服务器芯片设计，采用先进的优化技术和高效的架构，并提供定制化的芯片设计服务，全方位服务于客户。

应用场景：AI 芯片、算力中心、数据中心、服务器集群、智能计算等多个领域

灵汐科技

技术亮点：北京灵汐科技是一家全球领先的类脑计算技术公司，致力于创造持续自主进化的新智能体。灵汐科技产品包括类脑芯片、计算模组、边缘智能计算盒子、类脑服务器等以及相关算法和软件

企业简介：公司成立于 2018 年，总部位于北京，次年成果登上 Nature 封面，陆续设立多家武汉、深圳、无锡、成都分支机构。

应用场景：应用覆盖智算中心、各种边缘智能计算场景和消费市场

聆思智能

技术亮点：聆思科技是一家专注提供智能终端系统级（SoC）芯片的高科技企业；持续致力于将全球前沿的人工智能算法与自主领先的芯片设计深度耦合，打造极致性价比的单芯片解决方案

企业简介：公司 2020 年成立，地址位于安徽省合肥市，公司拥有专利和著作权 80 余项，明星产品多核异构 AI SoC 芯片性能强大，已和多家企业形成紧密合作。

应用场景：可广泛应用于家居家电、教育办公、消费电子、智能车载等领域。

龙芯中科

技术亮点：全面掌握 CPU 指令系统、处理器 IP 核、操作系统等计算机核心技术，打造自主开放的软硬件生态和信息产业体系，为国家战略需求提供自主、安全、可靠的处理器，为信息产业的创新发展提供高性能、低成本的处理器和基础软硬件解决方案。

企业简介：2001 年，中国科学院计算技术研究所开始研制龙芯处理器，得到了中科院知识创新工程、863、973、核高基等项目大力支持，完成了十年的技术积累。主营业务为处理器及配套芯片的研制、销售及服务，主要产品与服务包括处理器及配套芯片产品与基础软硬件解决方案业务。

应用场景：电子政务、能源、交通、金融、电信、教育等行业领域

每刻深思

技术亮点：每刻深思是一家拥有自研核心技术的“感算共融” 智能芯片设计公司，目前是国内首家专注于模拟领域智能感知芯片设计的公司，拥有全球首颗感算一体芯片，自研“感算共融”架构，μW 级视觉 AI 芯片等技术。

企业简介：公司前身团队来自清华大学电子系，历经多个全球首创存算一体等芯片流片后于 2020 年正式成立，致力于为客户提供边缘端可持续智能感知的超低功耗芯片，向全球提供 AI、AR/VR、高性能、高扩展、边缘计算以及半导体芯片设备六大品类。

应用场景：智能穿戴、智能家居、机器人、智能视觉

摩尔线程

技术亮点：摩尔线程专注于研发设计全功能 GPU 芯片及相关产品，支持 AI 计算加速、 3D 图形渲染、超高清视频编解码、物理仿真与科学计算等多种组合工作负载。

企业简介：摩尔线程成立于 2020 年 10 月，以全功能 GPU 为核心，致力于向全球提供加速计算的基础设施和一站式解决方案，为各行各业的数智化转型提供强大的 AI 计算支持。

应用场景：大模型、AIGC、科学计算、数字孪生、物理仿真、元宇宙等应用

墨芯人工智能

技术亮点：墨芯提供云端和终端 AI 芯片加速方案。墨芯通过优化计算模式，支持全面稀疏化神经网络开发，提供超高算力、超低功耗的通用 AI 计算平台。

企业简介：墨芯是一家致力于颠覆式创新的 AI 芯片设计商，采用领先于世界的稀疏化算法，旨在打造世界下一代人工智能芯片。2018 年，墨芯人工智能在硅谷创立，目前总部位于深圳。创始团队来自于卡内基梅隆大学顶尖 AI 科学家、世界顶尖半导体公司核心量产芯片研发团队。

应用场景：数据中心、互联网、运营商、安防等领域。可高效加速计算机视觉、自然语言处理、智能推荐、语音识别与合成、知识图谱等诸多云端推理场景

沐曦集成电路

技术亮点：沐曦致力于为异构计算提供全栈 GPU 芯片及解决方案；公司拥有丰富 GPU 量产经验，完整的软件生态能力，大量创新专利打造出打造全栈 GPU 芯片产品，推出曦思®N 系列、曦云®C 系列以及曦彩®G 系列 GPU 产品。

企业简介：沐曦集成电路（上海）有限公司，于 2020 年 9 月成立于上海，并在北京、南京、成都、杭州、深圳、武汉和长沙等地建立了全资子公司暨研发中心。

应用场景：广泛应用于智算、智慧城市、云计算、自动驾驶、数字孪生、元宇宙等前沿领域

平头哥半导体

技术亮点：平头哥拥有端云一体全栈产品系列，涵盖数据中心芯片、IoT 芯片等，实现芯片端到端设计链路全覆盖。

企业简介：公司于 2018 年 9 月宣布成立，是阿里巴巴集团的全资半导体芯片业务主体，总部位于杭州在上海、北京、深圳等地均设有分支机构，公司倚天、含光、羽阵系列芯片均量产出货数百家企业，部署到海量终端。

应用场景：云原生、视频编解码、高性能计算、基于 CPU 的机器学习和游戏服务等场景

启英泰伦

技术亮点：专注于语音芯片，涵盖智能算法、解决方案和开发平台的完整生态，全链条智能语音技术,完全自主知识产权，成立至今，启英泰伦芯片产品已历经三次大迭代，四次小迭代，共计推出 15 款型号的智能语音芯片，涵盖 AI 语音芯片，AI 语音 Wi-Fi 芯片，AI 语音 BLE 芯片，形成系列化的芯片产品布局。

企业简介：公司于 2015 年 11 月在成都高新区注册成立，是集语音芯片、语音算法、应用方案、开发平台于一体的行业领导型语音解决方案供应商，致力于成为离线语音芯片开创者和引领者。

应用场景：智慧汽车、语音 AI 平台、智慧教育、娱乐、智慧玩具、智慧安防、智慧家居家电。

千芯科技

技术亮点：核心产品是先进的存算一体芯片技术，可在高于同期 ASIC 芯片性价比的前提下，兼具 ASIC 的算力与 GPU 的灵活性，核心产品包括大算力的计算板卡和计算 IP 核，可为客户提供灵活易用的计算加速及一站式解决方案

企业简介：千芯科技成立于 2019 年总部位于北京，致力于为人工智能领域的客户提供最先进的存算一体算力产品与计算解决方案，背靠国家顶级院校与研究院所的协作优势，在存算一体芯片及 AI 计算加速领域具备深厚的技术积累。

应用场景：广泛应用于云计算、自动驾驶、智慧安防等领域。在云计算方向。

清微智能

技术亮点：专注于可重构计算芯片的创新研发和产业应用，提供高性能算力支持，致力于打造自主可控的可重构通用计算生态。

企业简介：可重构计算(CGRA)领导企业，清微智能已为全球众多知名企业提供芯片产品及服务。

应用场景：智算中心、智能安防、智慧办公、机器人、智能家居/家电、面向云端训推一体，视频分析，大模型微调、安防监控等智能计算场景

全志科技

技术亮点：全志科技是卓越的智能应用处理器 SoC、高性能模拟器件和无线互联芯片设计厂商，在超高清视频编解码、高性能 CPU/GPU/AI 多核整合、先进工艺的高集成度、超低功耗、全栈集成平台等方面提供具有市场突出竞争力，T 系列和 V 系列 AI 算力芯片涵盖图形解码，ISP 影像、车载中控等多个复杂场景。

企业简介：全志科技（AllwinnerTechnology）成立于 2007 年，总部位于中国珠海，在深圳、西安、上海、成都、横琴、广州、香港等地设有研发中心或分支机构，2015 年于深交所创业板上市，公司 AI 芯片产品种类繁多，已经广泛出货搭载到上百家客户的终端产品中。

应用场景：广泛适用于工业控制、智慧汽车、智慧家电、机器人、智慧安防、网络机顶盒、智能硬件、平板电脑、虚拟现实以及电源模拟器件、无线通信模组、智能物联网等多个产品领域。

锐思智芯

技术亮点：基于独创的 Hybrid Vision® 融合视觉技术，我们突破性地研发了 ALPIX™ 系列融合视觉传感芯片，独创专利化的融合视觉技术，通过创新的芯片架构和像素设计，辅以先进的 3D 堆叠和 BSI 背照式工艺，将事件传感技术与传统图像传感技术完美融合至芯片同一像素内。

企业简介：公司创立于 2019 年，拥有来自全球 6 个国家的超过 150 位成员，以深圳为总部，在苏黎世、北京和南京设立了研发团队，致力于在芯片架构、像素设计、数据处理及算法应用层实现了经典图像传感技术和新型事件感知技术的全面融合。

应用场景：智能手机、消费电子、智能安防、智能汽车、智能家居、机器人领域

瑞芯微

技术亮点：专注于集成电路设计与研发，目前已发展为领先的物联网（IoT）及人工智能物联网（AIoT）处理器芯片企业，在处理器和数模混合芯片设计、多媒体处理、影像算法、系统软件开发上具有丰富的经验和技术储备；代表产品为边缘计算加速库 RK 系列芯片多核 Arm 芯片，大小核心，超低功耗

企业简介：瑞芯微成立于 2001 年，总部位于福州，在深圳、上海、北京、杭州、香港设有分/子公司，致力于为客户提供多层次、多平台、多场景的专业解决方案

应用场景：涵盖高清显示、智能视觉、智能视频处理、机器视觉、人工智能加速、多模型 NPU 训练、智能视频处理应用等多元领域

睿思芯科

技术亮点：公司专注于研发高性能 RISC-V 边缘计算处理器、64 位视频/图像 DSP IP V9+芯片，拥有业界领先架

构，强悍计算性能保证，完备的生态体系：方便开发、移植和调试；还在全球范围内首次将向量处理器用于音频 DSP 领域

企业简介：公司于 2018 年成立，总部位于深圳，是一家提供 RISC-V 高端核心处理器解决方案的公司,创始团队来自于加州大学伯克利分校 RISC-V 原创项目组。

应用场景：数据中心、5G 通讯、存储控制器、机器学习、Wi-Fi 6/7、音频编解码、AI 语音识别

申威科技

技术亮点：公司专注于申威处理器芯片封装设计、技术支持服务及销售；基于申威异构众核处理器的小型超级计算机研发以及各类计算终端和高性能服务器的研发

企业简介：公司成立于 2016 年，自有 4200 ㎡研发、生产场地，致力于申威处理器的产业化推广和市场销售；公司自有知识产权 50 余项，种类涵盖了集成电路布图权、发明专利、实用新型专利、外观专利、软件著作权等

应用场景：神威·太湖之光超级计算机、嵌入式 CPU、电力、能源、服务器等场景

时擎科技

技术亮点：时擎科技围绕各类端侧智能应用场景，基于领域专用架构(DSA)的方法学，打造了处理器、部署开发工具、端侧 AI 算法三位一体、紧密融合的关键技术，为端侧智能处理和交互方案提供核心竞争力。

企业简介：时擎科技 Timesintelli 成立于 2018 年，总部位于上海张江，并在无锡、济南、深圳、香港等地设有分支机构。时擎科技是国家级专精特新小巨人、上海市科技小巨人企业，成立以来先后完成 SIG 海纳亚洲、上海科创投、海望资本、新尚资本等知名投资机构的多轮投资。

应用场景：智能家居、智慧家电、泛安防、智能 MCU 等不同细分领域的需要

时识科技

技术亮点：时识科技专注类脑智能的研究与开发，聚焦边缘计算应用场景，提供超低功耗、超低延时的全栈式解决方案与服务，是全球首个同时拥有类脑智能领域感知与计算技术，并掌握该领域大量核心底层专利的类脑智能公司。

企业简介：SynSense 时识科技（原名 aiCTX）创立于 2017 年，是全球领先的类脑智能与应用解决方案提供商。以苏黎世大学和苏黎世联邦理工学院 20+年全球领先的研发经验和技术积累为基石，时识科技率先实现了类脑芯片商业化应用零的突破

应用场景：手势控制、人体识别、人脸检测、高速避障、物体追踪、自动驾驶、智能安防等

视海芯图

技术亮点：视海芯图创新性使用 DRAM 存算技术进行神经网络运算和图像处理加速，解决其中的存储墙问题，实现超低功耗的算力芯片。

企业简介：公司成立于 2020 年 12 月，已在北京、成都、杭州和新加坡建立研发中心，已经与股东合作围绕 IoT、元宇宙和车载方面的核心图像处理算法进行存算一体加速，研发领域通用芯片。

应用场景：智能教育、自动驾驶、智能机器人、虚拟现实、3D 人脸识别

四维图新

技术亮点：四维图新专注于提供极致性价比的软硬一体组合产品解决方案，服务于各类智能出行设备及应用场景，并在智能座舱、汽车电子芯片领域积极拓展，智能 AI 芯片 AC 系列，8 核高性能 CPU，内置 NPU 加速 AI 应用计算

企业简介：四维图新成立于 2002 年，历经二十年技术积累在智能驾驶技术方面取得了显著突破，已为全球多个汽车品牌提供安全、可靠的智能驾驶量产解决方案，满足各类车规产品认证

应用场景：主要应用于高端智能座舱、人机交互多场景应用需求解决方案

算能科技

技术亮点：算能专注于 RISC-V、TPU 处理器等算力产品的研发和推广应用；致力于引领智算技术创新，打造覆盖“云、边、端”全场景产品矩阵

企业简介：公司成立于 2020 年，在北京、上海、深圳、青岛等国内 10 多个城市及美国、新加坡等国家设有研发中心，聚焦于 RISC-V 高性能核心产业领域，提供高性能服务器和整体云服务解决方案，在操作系统、应用、算法、编译器、产品硬件上与伙伴厂商紧密合作

应用场景：数据中心、城市运营、智能制造、大模型应用、智能终端等多元场景

燧原科技

技术亮点：燧原科技专注人工智能领域云端算力产品，致力为通用人工智能打造算力底座，提供原始创新、具备自主知识产权的 AI 加速卡、系统集群和软硬件解决方案。

企业简介：燧原科技成立于 2018 年 3 月，陆续经历多轮融资，先后发布了数款算力加速 AI 芯片，客户遍及海内外几十个国家和地区。

应用场景：广泛应用于泛互联网、智算中心、智慧城市，智慧金融、科学计算、自动驾驶等多个行业和场景

探境科技

技术亮点：探境科技是一家具有高创新能力的 AI 边缘芯企业，高水平的研发团队，拥有软件、硬件、算法、系统等全链条研发能力，公司自研了 SFA（存储优先）架构芯片和音旋风系列语音芯片

企业简介：公司成立于 2017 年，在北京、上海、深圳、合肥及美国硅谷设立研发基地，针对 AI 计算“高差异、高并发、高耦合”特性公司自研了 SFA 架构和通用性 AI 芯片架构符合大规模商业化需求

应用场景：边缘计算、工业视觉、智能安防、新零售、辅助驾驶

天数智芯

技术亮点：是中国领先的通用 GPU 高端芯片及超级算力系统提供商。天数智芯致力于开发自主可控、国际领先的高性能通用 GPU 产品，为全产业提供高端算力解决方案

企业简介：上海天数智芯成立于 2015 年， 2018 年正式启动 7 纳米通用并行（GPGPU）云端计算芯片设计，拥有一支全球顶尖的数字集成电路设计与基础软件设计科学家团队

应用场景：智慧医疗、互联网、智慧教育、智能语音、智能制造、内容生成

微纳核芯

技术亮点：微纳核芯专注于 AIoT SoC 芯片领域，依托世界领先的芯片科研团队和业界一流的芯片工程化队伍，打造 AIoT 芯片技术“科研成果”到“产业落地”的持续性“产学研循环”，参与承担国家重点研发计划等项目，拥有 64 项
专利技术。

企业简介：总部位于杭州，拥有无锡、北京子公司和上海、深圳分公司

应用场景：物联网、新能源、智能终端、无人智慧系统等未来重要的战略性应用领域

物奇微

技术亮点：是国内领先的高性能短距通信与边缘计算领域 SoC 芯片设计厂商，依托领先的通信连接技术，为万物互联提供一流的 SoC 芯片和软件解决方案，边缘计算芯片具有 4 个 32 位 RISC-V CPU、采用专用 NNU 处理器

企业简介：物奇成立于 2016 年，在重庆、上海、长沙、香港、深圳等地设有研发中心和客户支持中心；产品性能和品质处于业内领先地位，为 TPLINK、OPPO、哈曼、荣耀、安克创新、商汤科技、吉利汽车、小米等国内外众多知名客户提供一流的芯片方案

应用场景：车载智能语音、智能家居、语音机器人、智能人机交互

曦智科技

技术亮点：是全球领先的光电混合算力提供商。公司凭借在集成光子领域的开创性技术；以光子矩阵计算(oMAC)、片上光网络(oNOC)和片间光网络(oNET)三大核心技术出发，打造光子计算和光子网络两大产品线

企业简介：公司成立于 2017 年，致力于成为全球光电混合计算创领者，公司拥有专利和软件著作权约 60 余项。

应用场景：大数据、云计算、金融、自动驾驶、生物医药、材料研究等领域

芯驰科技

技术亮点：面向中央计算+区域控制电子电气架构提供高性能、高可靠的车规芯片产品和解决方案，芯驰是国内首个完成车规芯片领域五大安全认证的企业，目前，芯驰全系列产品已完成超百万片规模化量产

企业简介：芯驰科技成立于 2018 年在北京、上海、南京、深圳、大连设有研发中心，同时在长春和武汉设有办事处，拥有国内为数不多的具备车规芯片产品定义、技术研发及大规模量产落地的国际化整建制团队。

应用场景：覆盖智能座舱和智能车控等领域。

芯动力科技

技术亮点： 芯动力科技的核心技术为可重构并行处理器架构（简称 RPP），它是自主研发专为并行计算设计的处理器架构，具有良好的生态兼容和超高能效的并行计算能力，能够打破高性能芯片和通用芯片的鸿沟，广泛应用于各个场景，为各行业在计算领域提供一体化的解决方案。

企业简介：珠海市芯动力科技有限公司成立于 2017 年，总部位于广东省珠海市，目前在深圳、西安、美国均设有研发中心；是一家专注于高性能及高通用性芯片设计研发为主的科技公司。

应用场景：AI PC、工业自动化、泛安防、工业自动化、智能驾驶、安防监控等多个领域。

芯砺智能

技术亮点：专注于 Chiplet 异构集成技术，自研通用 NPU 及工具链，自研差异化核心自主 IP，同时开发核心自主软件

企业简介：芯砺智能成立于 2021 年 11 月，在全球拥有多个研发中心。芯砺智能是全球首家利用芯粒（Chiplet）技术研发车载大算力芯片的高科技初创企业，致力于成为智能汽车平台芯片的全球领导者，目前拥有五大研发中心。

应用场景：车载信息娱乐系统、车载导航、智能识别、域控制器、智能汽车等大算力场景。

芯明智能

技术亮点：是一家专注 3D 空间计算及人工智能芯片及产品设计的高科技企业，其自研系列芯片拥有全球领先的 3D 视觉感知处理引擎，且是全球唯一单芯片集成芯片化 3D 视觉感知、AI 人工智能、SLAM 实时定位建图的系统级芯片

企业简介：芯明智能原名银牛微电子，成立于 2020 年，公司全球总部坐落于合肥，并在上海、以色列、北京和深圳设有子公司和分公司，在 3D 视觉、复杂 SoC 芯片设计、低功耗设计、光学、嵌入式系统软件、边缘 AI 计算等方面具有深厚的经验。。

应用场景：泛机器人、元宇宙 XR、消费电子、物流无人机、3D 扫描、虚拟数字人等多个前沿应用领域

芯擎科技

技术亮点：专注于设计、开发并销售先进的汽车电子芯片，致力于成为世界领先的汽车电子芯片整体方案提供商。公司自研了智能座舱芯片、舱泊一体芯片等均实现大规模落地。

企业简介：湖北芯擎科技有限公司于 2018 年在武汉经济技术开发区成立，在武汉、北京、上海、深圳、沈阳和重庆设有研发和销售分支机构，

应用场景：智能座舱、自动泊车、智能驾驶、机器视觉、行车安全监控等领域

芯原股份

技术亮点：基于芯原独有的芯片设计平台即服务（Silicon Platform as a Service, SiPaaS）经营模式，为客户提供平台化、全方位、一站式芯片定制服务和半导体 IP 授权服务的企业；已拥有丰富的面向人工智能（AI）应用的软硬件芯片定制平台解决方案

企业简介：芯原成立于 2001 年，总部位于中国上海，在全球设有 8 个设计研发中心，拥有六大核心处理器 IP 和 1,600 多个数模混合 IP 和射频 IP

应用场景：消费电子、汽车电子、计算机及周边、工业、数据处理、物联网等，主要客户包括芯片设计公司、IDM、系统厂商、大型互联网公司、云服务提供商等

依图科技

技术亮点：自研求索 QuestCore™机器视觉芯片，计算密度高：单芯片 高达 50 路视频解析，支持主流十几种算法和模型，是目前国内唯一具有提供超大规模、复杂环境下亿级规模城市的智能化运营管理技术能力的人工智能公司

企业简介：依图科技公司成立于 2012 年，专注于人工智能创新型研究，致力于将先进的人工智能技术与行业应用相结合，拥有自研的 care.ai 全链路产品、依图语音开放平台、求索 AI 芯片等产品。

应用场景：掌上智能数据中心、万路智能视频解析、城市大脑、AI 医疗、智慧城市

亿智电子

技术亮点：亿智电子科技有限公司是以 AI 机器视觉算法和 SoC 芯片设计为核心的系统方案供应商，专注于边缘侧/端侧通用算力 AI SoC 芯片的研发，致力于为客户提供覆盖多场景的系统级解决方案

企业简介：公司成立于 2016 年，目前在珠海设立研发总部，在北京、深圳、上海、香港均设有分支。2019 年量产搭载自研 NPU 的边缘侧/端侧 AI SoC 芯片，已助力众多合作伙伴实现 AI 产品的规模量产

应用场景：产品线涵盖智能车载、智能硬件、智慧安防三大应用领域。

亿铸科技

技术亮点：亿铸科技将新型存储器 ReRAM 及存算一体计算架构相结合，通过全数字化的芯片设计思路，致力于实现数倍性价比、更高能效比、更大算力发展空间的新一代 AI 大算力芯片。

企业简介：亿铸科技成立于 2020 年 6 月，是一家基于存算一体这一创新架构自研 AI 芯片的公司，目前，亿铸科技点亮了基于忆阻器 ReRAM 的高精度、低功耗存算一体 AI 大算力 POC 芯片

应用场景：数据中心、云计算、中心侧服务器、自动驾驶及边缘计算等场景

奕行智能

技术亮点：是一家自动驾驶芯片研发商，聚焦车用 AI 算力等核心技术，自研通用的 DSA 架构芯片产品及软件栈，目前已经已流片芯片并为大规模量产

企业简介：奕行智能成立于 2022 年是一个自动驾驶芯片研发商，奕行智能拥有整建制的芯片软硬件团队，并在上海、成都、南京、北京等地设立研发中心。

应用场景：智能驾驶、智慧汽车、机器视觉、域控制器、辅助驾驶、自动驾驶等复杂算力领域

云豹智能

技术亮点：是一家专注于云计算和数据中心数据处理器芯片（DPU）和解决方案的领先半导体公司，量产全功能云霄 DPU 产品，全面支持裸金属、虚拟机和容器服务资源一体化和性能加速。

企业简介：公公司成立于 2020 年，核心团队来自 Broadcom、Intel、Arm、华为海思、阿里巴巴等，拥有中国最有经验的 DPU 芯片和软件研发团队。旨在成为引领数据中心和云计算最前沿技术，并建立“软件定义芯片”行业标准的高科技公司。

应用场景：数据中心、智能超算、弹性网络、AI 算力

云天励飞

技术亮点：公司自研从 22nm 到 14nm 不同芯片制程，构建算法+芯片+大数据构建全栈 AI，明星产品 NNP100-400 系列 DeepEye(Edge)芯片，算法精度高，活体检测性能 99%以上，

企业简介：公司成立于 2014 年，发布自主知识产权 AI 芯片和公共安全系统--“深目”目前已经在 20 多国家和地区落地，政企、公安、消防、博览会安保等多场景服务百余次，是国家发改委、科技部重点发展的芯片技术领域

应用场景：智慧交通、测温防疫大数据、公共安全数据挖掘、社区治理、个性化推荐系统

肇观电子

技术亮点：上海肇观电子科技有限公司是一家从事计算机视觉人工智能芯片设计的科技公司，公司拥有各类针对超高清 AI 智能摄像头产品开发的低功耗高性能 SoC 芯片，超 200 人研发专业团队，累计申请国内外专利两百余件

企业简介：总部位于上海张江科学城集成电路产业园，拥有员工逾 300 人，研发占比 80%，公司已成功发布刷新世界记录的视觉处理能力的 N 系列、D 系列、V 系列芯片

应用场景：可广泛应用于安防、门禁、家用摄像、3D 视觉、车载

知存科技

技术亮点：致力于全球领先的存内计算技术及芯片，自主研发的边缘侧算力芯片，针对 AI 应用场景，在全球率先商业化量产基于存内计算技术的神经网络芯片

企业简介：知存科技 2017 年成立，是全球领先的存内计算芯片企业，89 项专利和 61 荣誉，2022 年，知存科技推出全球首颗大规模量产的存内计算芯片 WTM2101，已经广泛应用于百余家客户终端产品。

应用场景：智能语音、AI 健康监测、边缘计算、算力中心、AR\VR、智能手机、医疗健康等场景

智芯科微

技术亮点：专注于具备高算力、高能效优势的存内计算 A!芯片开发，自主研发的核心技术-基于精度无损 SRAM 存内计算(CIM)超低功耗神经网路处理器芯片

企业简介：公司成立于 2019 年 9 月，是存内计算芯片开拓者，基于多元场景已实现产品与技术快速落地。现有杭州、深圳两地分支机构。

应用场景：TWS 耳机、智能语音、麦克风、智能家居(如遥控器、小夜灯) 、智能手表、扫地机等场景

中昊芯英

技术亮点：致力于研发可支撑超大规模 AI 大模型计算的高性能 AI 芯片与计算集群，中昊芯英以自研的专为 AI 大模型而生的高性能 TPU AI 芯片 “刹那®” 为基石，打造支持 1024 片芯片片间高效互联、可支撑超千亿参数大模型的大规模 AI 计算集群 “泰则®” ，同时自研 AIGC 预训练大模型并携手行业合作伙伴进行金融、教育、医疗等垂直领域专业大模型的探索落地。通过 “自研 AI 芯片 + 超算集群 + AIGC 预训练大模型” 的三位一体化方案，为全球客户提供具备生产变革能力的 AI 创新技术方案，加速 AI 工程落地与产业化进程。

企业简介：中昊芯英创始于 2018 年，2020 年落户杭州，先后完成多轮融资，顺利完成 TPU-刹那，推理专用 SOC-祝融设计和量产验证，核心团队由来自谷歌、微软、甲骨文、三星、英伟达、亚马逊、Facebook 等顶尖科技公司的 AI 软硬件设计专家组成，掌握从 28nm 到 7nm 各代制程工艺下大芯片设计与优化完整方法论，全栈式的技术梯队覆盖芯片设计、电路设计、软件栈研发、系统架构、大模型算法等各类技术领域，公司研发人员占比 70%以上。

应用场景：AI 算力、人工智能、推理逻辑、万卡集群、AI 大模型

中星微

技术亮点：中星微技术是在数字感知领域拥有国际领先的芯片设计技术和新一代机器视觉编解码技术的高科技企业；拥有国际领先自主知识产权的 XPU 多核异构处理器架构和人工智能算法，低功耗、超高清、NPU、加密等独特优势的新一代高算力双模 SVAC2.0+H.265+NPU 芯片已经发布

企业简介：公司成立于 1999 年，由工信部牵头成立，历时二十年发展多次参与制定国家、国际标准体系，提供自主可控核心知识产权、芯片、产品、方案及承担国家重大战略工程，2005 年在纳斯达克上市，成为第一家在纳斯达克上市的中国芯片设计企业

应用场景：公共安全、数字信创、智慧能源、智慧交通、智慧金融、智慧水利、工业物联网、车联网及家庭等领域提供数智化行 业应用

紫光国微

技术亮点：聚焦特种集成电路、智能安全芯片两大主业，并涵盖石英晶体频率器件、功率半导体等重要业务，拥有接触卡 IC 芯片、智能终端安全芯片等全流程设计研发能力，在主要业务领域拥有相关核心技术，曾获国家科技进步

一等奖、二等奖，国家技术发明二等奖等权威奖项

企业简介：公司隶属新紫光集团，是国内领先的综合性半导体上市企业，2005 年深交所挂牌上市，后经多次收购和股权融资，并入紫光集团，控股子公司遍布西安、成都、西藏、深圳等地。

应用场景：移动通信、金融、政务、汽车、工业、物联网等领域

紫光展锐

技术亮点：紫光展锐是世界领先的平台型芯片设计企业，是全球少数全面掌握 2G/3G/4G/5G、Wi-Fi、蓝牙、电视调频、卫星通信等全场景通信技术的企业之一，累计申请专利超 11000 项，拥有 3G/4G/5G、多卡多待、多模等核心专利。

企业简介：紫光展锐具备大型芯片集成及套片能力，产品包括移动通信中央处理器，基带芯片，AI 芯片，射频前端芯片，射频芯片等各类通信、计算及控制芯片等，场测覆盖全球 133+国家和地区，通过全球 260+运营商的出货认证

应用场景：智能手机、平板电脑、手机 SoC、移动设备

本报告由深芯盟半导体产业研究部首席分析师顾正书主笔撰写，报告中若涉及公司信息或专业知识方面的错误，

欢迎指正。

顾正书（Steve Gu）| 深芯盟首席分析师

Email: steve.gu@semibay.cn

WeChat：gusteve

现任深芯盟半导体产业研究部首席分析师，主要负责半导体产业分析报告、排行榜和会议论坛筹划。曾在 Aspencore、Global Sources 和 CapitalOne 等国际半导体/电子行业媒体及高科技企业任职，拥有多年美国及中国高科技行业数据分析和市场营销管理经验。获得美国德州大学(UT-Austin)商学院 MBA 和南京理工大学电子工程学士学位。

雷 阳（Ray Lei） | 深芯盟产业分析师

上海理工大学电气工程硕士，曾担任芯片封装和摄像头工艺工程师，专注于芯片设计、半导体材料等上下游垂直领域研究。

关于深芯盟

深圳市半导体与集成电路产业联盟（深芯盟）是深圳市委、市政府部署支持，市发展改革委指导设立的开放性和公益性联盟组织，由深重投集团会同 20 余家半导体产业链各环节的龙头单位发起设立。深芯盟将围绕“12345”发展战略，肩负“有为政府”、“有效市场”两项使命，赋能半导体制造类、设计类、服务类三大集群跃升，推动创新、产业、人才、资本四链融合发展，铸造生态展会、会议论坛、产业报告、招引品牌、资源对接平台五大驰名品牌，着力打造具有全球影响力的全过程、创新型产业生态联盟。

关于湾芯展

SEMiBAY/湾芯展旨在贯彻落实深圳“20+8”产业“一集群、一展会”决策部署，由深圳市人民政府指导、深圳市发展与改革委员会主办、深圳市半导体与集成电路产业联盟（深芯盟）承办。湾芯展定于 10 月 15–17 日在深圳会展中心（福田）举行，将充分依托深圳及大湾区的广阔应用市场，以及深重投主导的重大产业项目集群等优质资源，聚焦半导体设备、材料、晶圆制造、封测、EDA/IP、IC 设计和应用等重点领域。

湾芯展展览区域分为五大专区：晶圆制造、先进封装与测试、化合物半导体、EDA/IP 与 IC 设计。与展览同期举行的还有湾区半导体大会，包括半导体高峰论坛、集成电路院长论坛，以及 20 多场细分领域的技术论坛，涵盖晶

圆制造工艺、先进封装与测试、化合物半导体产业发展、汽车半导体和智能网联、EDA/IP/Chiplet、AI 芯片与高性能计算、RISC-V 开源生态、HBM 与存储、AIoT 与智能终端，以及半导体产业投资和集成电路人才培养和招聘主题。

晶圆制造论坛

1、国际半导体设备技术与工艺论坛
2、晶圆工艺与管理论坛
3、集成电路材料论坛
4、核心零部件论坛

化合物半导体论坛

1、大湾区第三代半导体产业及应用发展论坛
- 氮化镓功率器件材料及应用发展论坛
- 碳化硅与汽车半导体技术及应用论坛

IC设计论坛

1、EDA/IP与IC设计服务论坛
2、AI芯片与高性能计算论坛
3、RISC-V生态发展论坛
4、无线通信芯片技术与应用论坛
5、电源管理和功率半导体论坛
6、模拟信号链与传感器论坛

先进封装论坛

1、先进封装工艺与材料论坛
2、TGV玻璃基板关键工艺
3、Chiplet设计与异构集成
4、先进封装与测试论坛