

➤ **AI 芯片赋能算力基石，英伟达垄断全球市场。**根据芯片的类别，AI 算力芯片主要包括 GPU、FPGA，以及以 TPU、VPU 为代表的 ASIC 芯片，其中以 GPU 用量最大，据 IDC 数据，预计到 2025 年 GPU 仍将占据 AI 芯片 8 成市场份额。由于英伟达 GPU 产品线丰富、产品性能顶尖、开发生态成熟，目前全球 AI 算力芯片市场由英伟达的 GPU 垄断，根据中国信通院的数据，2021 年 Q4 英伟达占据了全球 95.7% 的 GPU 算力芯片市场份额。**2023 财年，英伟达数据中心营收达到 150 亿美元，同比增长 41%，FY2017-FY2023 复合增速达 63%，表明全球 AI 芯片市场规模保持高速增长。**

➤ **国产 AI 算力芯片正起星星之火。**全球 AI 芯片市场被英伟达垄断，然而国产 AI 算力芯片正起星星之火，投融资热度高企。根据电子发烧友统计，2022 年多家 AI 芯片公司获得大额融资，其中摩尔线程达 15 亿元、天数智芯超 10 亿元、沐曦达 10 亿元。在国产 AI 算力芯片中，寒武纪推出的云端训练侧产品思元 290 达到 512TOPS INT8 算力，训推一体产品思元 370 达到 256TOPS INT8 算力；海光信息 DCU 产品深算一号部分参数对标英伟达 A100。**国内非上市公司中**，天数智芯的训练侧产品 BI 达到 295TOPS INT8 算力；沐曦的推理侧产品 MXN100 已于 2022 年 8 月回片点亮。**互联网巨头亦强势入局 AI 芯片**，腾讯领投的燧原科技推出推理侧产品云燧 i20，INT8 算力达 256TOPS；百度孵化的昆仑芯推出训推一体 AI 芯片 R200，INT8 算力达 256TOPS；背靠阿里的平头哥亦早在 2019 年就推出推理侧 AI 芯片含光 800。

➤ **自主可控受高度重视，国产 AI 算力芯片迎“芯”机遇。**2 月 27 日，中共中央、国务院印发了《数字中国建设整体布局规划》，《规划》提出要夯实数字基础设施，我们认为，数字中国基础设施的建设有望拉动以数据中心、超算中心、智能计算中心为代表的算力基础设施建设，从而带动服务器与 AI 算力芯片的需求快速增长。**同时，为构筑自立自强的数字技术创新体系，上游 AI 芯片作为算力基础，自主可控需求凸显，数字中国建设对 AI 芯片国产化提出新要求。**

➤ **投资建议：**我们看好以海光信息、寒武纪、天数智芯、沐曦为代表的 AI 算力芯片公司持续突破，以及以芯原股份为代表的 IP 公司在 AI 芯片大潮下加速成长。建议关注 AI 芯片相关赛道投资机遇。

风险提示：疫情反复影响生产经营；下游需求不及预期；研发进展不及预期。

重点公司盈利预测、估值与评级

代码	简称	股价 (元)	EPS (元)			PE (倍)			评级
			2022E	2023E	2024E	2022E	2023E	2024E	
688041	海光信息	57.50	0.38	0.82	1.21	151	70	48	推荐
688256	寒武纪-U	122.93	-2.91	-1.79	-1.00	/	/	/	/
688521	芯原股份-U	77.02	0.15	0.33	0.52	520	230	147	推荐
002156	通富微电	23.38	0.39	0.74	1.06	60	32	22	/
600584	长电科技	30.91	1.85	2.04	2.37	17	15	13	/

资料来源：Wind，民生证券研究院预测；

(注：股价为 2023 年 03 月 17 日收盘价；未覆盖公司数据采用 wind 一致预期；海光信息、寒武纪-U、芯原股份-U 2022E EPS 为业绩快报披露值)

推荐

维持评级



分析师 方竞

执业证书：S0100521120004

邮箱：fangjing@mszq.com

相关研究

1. 电子行业跟踪：高温超导特性优异，行业应用前景广阔-2023/03/13
2. 电子行业周报 20230313：曙光渐显，把握芯片设计底部布局机遇-2023/03/13
3. 电子行业周报 20230228：Chiplet：AI 芯片算力跨越的破局之路-2023/02/28
4. 电子行业周报 20230222：ChatGPT 开启行业变革，Chiplet 引领破局之路-2023/02/22
5. 汽车电子月报：充电政策+禾赛上市催化，电动化智能化加速-2023/02/14

目录

1 国产 AI 算力芯片正起星星之火	3
1.1 AI 芯片赋能算力基石，英伟达垄断全球市场	3
1.2 性能与生态构筑 AI 算力芯片高壁垒	4
1.3 国产 AI 算力芯片全景图	7
2 自主可控受高度重视，国产 AI 算力芯片迎“芯”机遇	15
3 投资建议	16
4 风险提示	17
插图目录	18
表格目录	18

1 国产 AI 算力芯片正起星星之火

1.1 AI 芯片赋能算力基石，英伟达垄断全球市场

AI 算力芯片主要包括 GPU、FPGA, 以及以 VPU、TPU 为代表的 ASIC 芯片。其中以 GPU 用量最大, 据 IDC 数据, 预计到 2025 年 GPU 仍将占据 AI 芯片 8 成市场份额。然而, 相较传统图形 GPU, 通用型算力 GPU 在芯片架构上缩减了图形图像显示、渲染等功能实现, 具有更优的计算能效比, 因而被广泛应用于人工智能模型训练、推理领域。

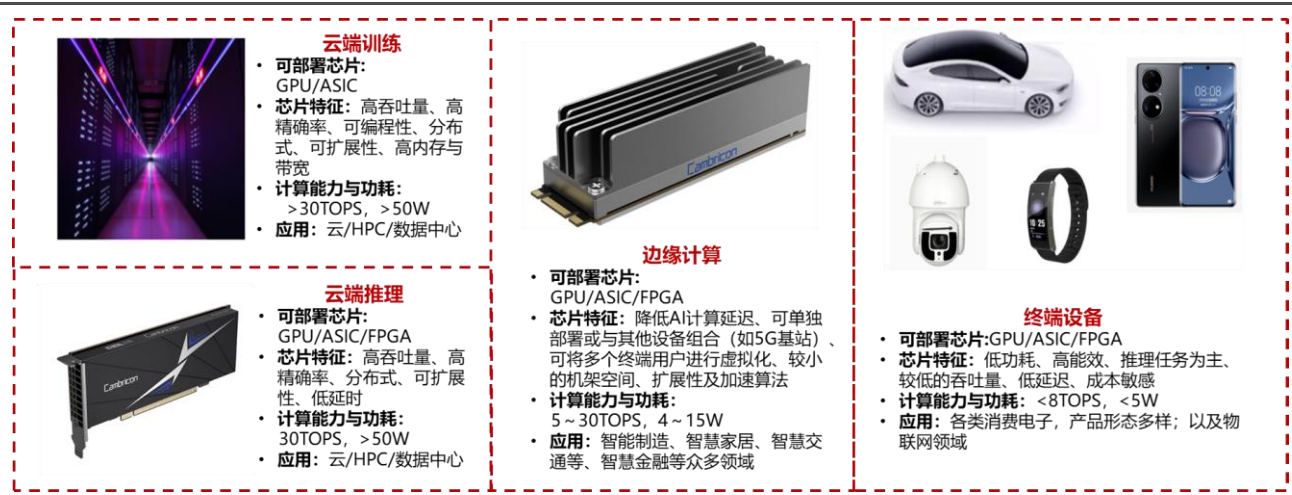
表1: 主流 AI 芯片分类

技术架构种类	定制化程度	可编辑性	算力	价格	优点	缺点	应用场景
图形 GPU	通用型	强	中	高	通用性较强且适合大规模并行运算; 设计和制造工艺成熟	其图形渲染功能无法在推理端完全发挥计算能力	高级复杂算法和通用性人工智能平台
算力 GPU	通用型	强	高	高	相较传统的 GPU, 缩减了图形图像显示、渲染等功能实现, 具有更优的计算能效比	算力性能及设计难度极高	互联网、安防、金融等通用计算领域和人工智能平台
FPGA	半定制化	强	高	中	可通过编程灵活配置芯片架构适应算法迭代, 平均性能较高; 功耗较低; 开发时间较短 (6 个月)	量产单价高; 峰值计算能力较低; 硬件编程困难	适用于各种具体的行业
ASIC	全定制化	弱	高	低	通过算法固化实现极致的性能和能效、平均性很强; 功耗很低; 体积小; 量产后成本最低	前期投入成本高; 研发时间长 (1 年); 技术风险大	当客户处在某个特殊场景, 可以为该独立设计一套专业智能算法软件

资料来源: 甲子光年, 民生证券研究院整理

根据在网络中的位置, AI 芯片可以分为云端 AI 芯片、边缘和终端 AI 芯片; 根据其在实践中的目标, 可分为训练 (training) 芯片和推理 (inference) 芯片。云端主要部署高算力的 AI 训练芯片和推理芯片, 承担训练和推理任务, 具体有智能数据分析、模型训练任务和部分对传输带宽要求比高的推理任务; 边缘和终端主要部署推理芯片, 承担推理任务, 需要独立完成数据收集、环境感知、人机交互及部分推理决策控制任务。

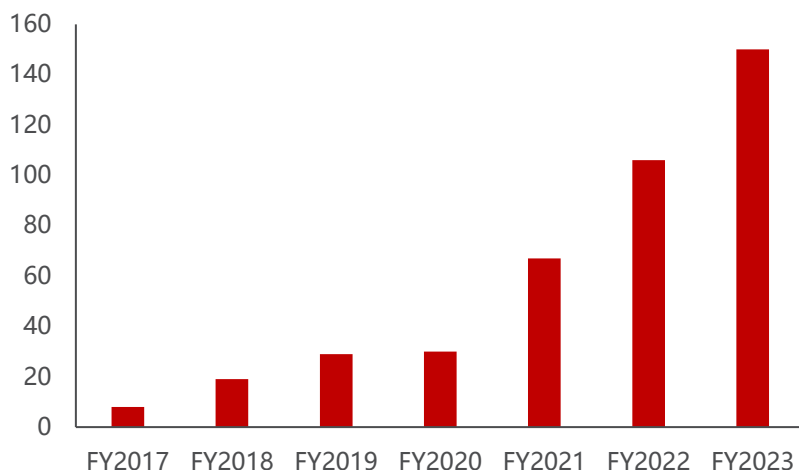
图1：云端、边缘和终端 AI 芯片应用情况



资料来源：亿欧智库，民生证券研究院

由于英伟达 GPU 产品线丰富、产品性能顶尖、开发生态成熟，目前全球 AI 算力芯片市场仍由英伟达垄断。根据中国信通院的数据，2021 年 Q4 英伟达占据了全球 95.7% 的 GPU 算力芯片市场份额，因此，英伟达数据中心业务营收增速可以较好地反应全球 AI 芯片市场增速。2023 财年，英伟达数据中心营收达到 150 亿美元，同比增长 41%，FY2017-FY2023 复合增速达 63%，表明全球 AI 芯片市场规模保持高速增长。

图2：2017-2023 财年英伟达数据中心营收（亿美元）



资料来源：英伟达财报，民生证券研究院

1.2 性能与生态构筑 AI 算力芯片高壁垒

评价 AI 芯片的指标主要包括算力、功耗、面积、精度、可扩展性等，其中算力、功耗、面积（PPA）是评价 AI 芯片性能的核心指标：

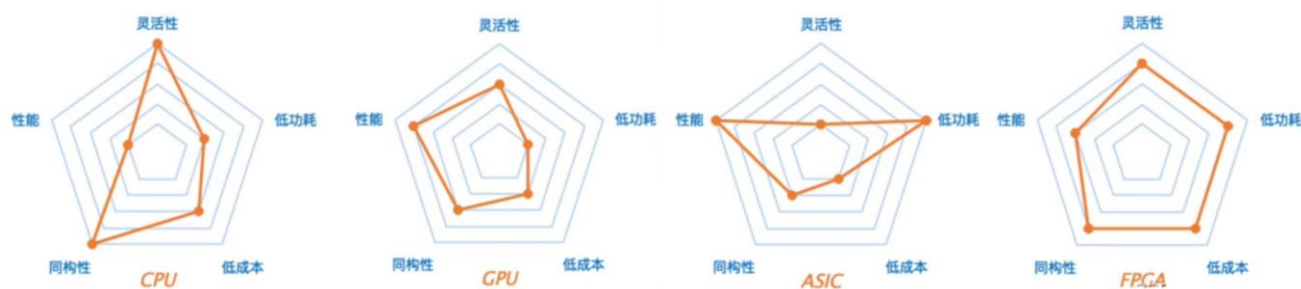
(1) **算力**：衡量 AI 芯片算力大小的常用单位为 TOPS 或者 TFLOS，两者分别代表芯片每秒能进行多少万亿次定点运算和浮点运算，运算数据的类型通常有整型 8 比特 (INT8)、单精度 32 比特 (FP32) 等。AI 芯片的算力越高，代表它的运算速度越快、性能越强。

(2) **功耗**：功耗即芯片运行所需的功率，除了功耗本身，性能功耗比是综合衡量芯片算力和功耗的关键指标，它代表每瓦功耗对应输出算力的大小。

(3) **面积**：芯片的面积是成本的决定性因素之一，通常来讲相同工艺制程之下，芯片面积越小良率越高，则芯片成本越低。此外，单位芯片面积能提供的算力大小亦是衡量 AI 芯片成本的关键指标之一。

除 PPA 之外，运行在 AI 芯片上的算法输出精度、AI 应用部署的可扩展性与灵活性，均为衡量 AI 芯片性能的指标。

图3：常见 AI 芯片评价标准



资料来源：CSDN，民生证券研究院

英伟达的 GPGPU 是全球应用最为广泛的 AI 芯片，决定其性能的硬件参数主要包括：微架构、制程、CUDA 核数、Tensor 核数、频率、显存容量、显存带宽等。其中，微架构即 GPU 的硬件电路设计构造的方式，不同的微架构决定了 GPU 的不同性能，作为英伟达 GPU 的典型代表，V100、A100、H100 GPU 分别采用 Volta、Ampere、Hopper 架构；CUDA 核是 GPU 内部主要的计算单元；Tensor 核是进行张量核加速、卷积和递归神经网络加速的计算单元；显存容量和带宽是决定 GPU 与存储器数据交互速度的重要指标。

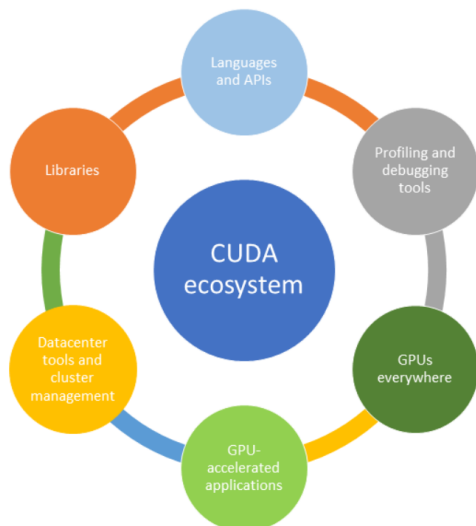
表2：英伟达 V100/A100/A800/H100 对比

性能参数	V100 PCIe	A100 80GB PCIe	A800 80GB PCIe	H100 80GB PCIe
微架构	Volta	Ampere		Hopper
FP64	7TFLOPS	9.7TFLOPS		26 TFLOPS
FP32	14TFLOPS	19.5TFLOPS		51 TFLOPS
FP16 Tensor Core		312TFLOPS		756.5 TFLOPS
INT8 Tensor Core	62 TOPS	624 TOPS		1513 TOPS
GPU 显存	32/16GB HBM2	80GB HBM2e		80GB
GPU 显存带宽	900 GB/s	1935GB/s		2TB/s
最大热设计功耗 (TDP)	250 瓦	300 瓦		300-350W
多实例 GPU		最多 7 个 MIG 每个 10GB		
外形规格		PCIe 双插槽风冷式 或单插槽液冷式		PCIe 双插槽风冷式
互连技术	NVLink: 300 GB/s PCIe: 32 GB/s	搭载 2 个 GPU 的 NVIDIA“ NVLink” 桥接器: 600GB/s PCIe 4.0: 64GB/s	搭载 2 个 GPU 的 NVIDIA“ NVLink” 桥 接器: 400GB/s PCIe 4.0: 64GB/s	NVLink: 600GB/s PCIe 5.0: 128GB/s
服务器选项		搭载 1 至 8 个 GPU 的合作伙伴认证系统和 NVIDIA 认证系统		

资料来源：英伟达官网，民生证券研究院整理

除 GPU 硬件之外，与之配套的软件开发体系亦是生态的重要组成部分。GPU 的生态包括底层硬件、指令集架构、编译器、API、基础库、顶层算法框架和模型等，英伟达于 2006 年发布的 CUDA 平台是当今全球应用最为广泛的 AI 开发生态系统。通用 GPU 与 CUDA 组成的软硬件底座构成了英伟达引领 AI 计算的根基，当前全球主流深度学习框架均使用 CUDA 平台。

图4：英伟达 CUDA 生态系统



资料来源：英伟达官网，民生证券研究院

生态构建计算壁垒，国产 GPU 厂商初期兼容 CUDA，长期仍需构筑自身软硬件生态。由于当前全球主流深度学习框架均使用 CUDA 平台进行开发，国产 GPU 可以通过兼容 CUDA 的部分功能，快速打开市场，减少开发难度和用户移植成本。然而，CUDA 本身涵盖功能非常广泛，且许多功能与英伟达 GPU 硬件深度耦合，包含了许多英伟达 GPU 的专有特性，这些特性并不能在国产 AI 芯片上全部体现。**因此，长期来看国产 GPU 厂商仍需通过提升自身的软硬件实力，构筑属于自己的软硬件生态。**

1.3 国产 AI 算力芯片全景图

全球 AI 芯片市场被英伟达垄断，然而国产 AI 算力芯片正起星星之火。目前，国内已涌现出了如寒武纪、海光信息等优质的 AI 算力芯片上市公司，非上市 AI 算力芯片公司如沐曦、天数智芯、壁仞科技等亦在产品端有持续突破。

表3：国产 AI 算力芯片公司对比

公司名称	成立时间	创始团队背景	地点	注册资本或市值	融资轮次及金额
寒武纪	2016	中科院	北京	市值 493 亿元	已上市
海光信息	2014	中科曙光、AMD、英特尔	天津	市值 1336 亿元	已上市
景嘉微	2006	国防科技大学	长沙	市值 367 亿元	已上市
沐曦	2020	AMD	上海	671 万元	Pre-B 轮/10 亿元
天数智芯	2015	AMD	上海	1.7 亿元	C++ 轮/超 10 亿元
壁仞	2019	AMD+英伟达	上海	3292 万元	B+ 轮/未披露
燧原	2018	紫光+AMD	上海	394 万元	C+ 轮/未披露（股东腾讯）
昆仑芯	2011	百度	北京	1768 万元	A 轮/未披露（股东百度）
平头哥	2018	中天微+达摩院	杭州	5000 万元	未披露（股东阿里巴巴）
海飞科	2020	英伟达	上海	97 万元	B 轮/未披露
后摩智能	2020	AMD	南京	1637 万元	Pre-A+ 轮/数亿元
登临	2017	图芯+华为	上海	208 万元	B 轮/未披露
摩尔线程	2020	英伟达	北京	2380 万元	B 轮/15 亿元
芯动科技	2006	美光	武汉	1000 万元	未披露
黑芝麻智能	2021	豪威	武汉	1.5 亿美元	未披露
地平线	2015	百度	北京	6 亿美元	未披露

资料来源：各公司官网等，民生证券研究院整理。注：非上市公司统计注册资本。

表4：国产 AI 算力芯片公司主要产品对比

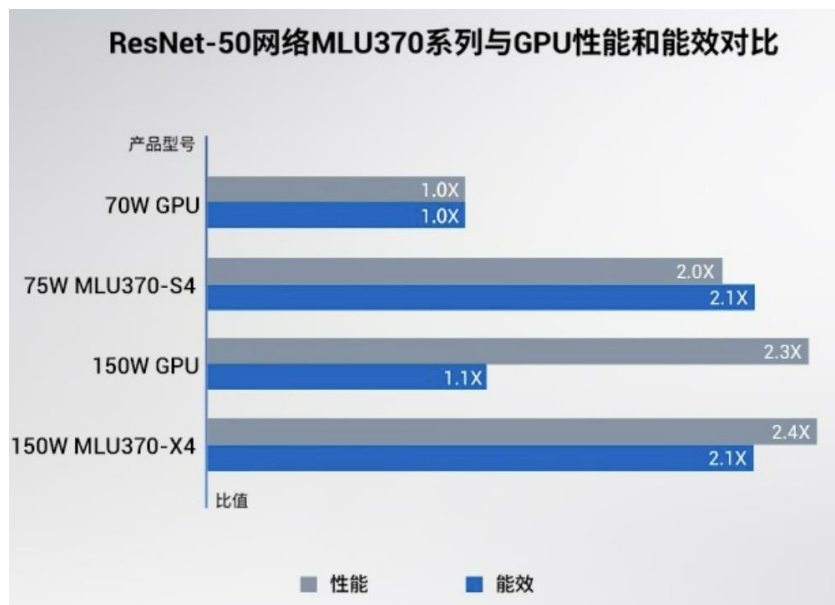
公司名称	产品类型	产品	算力	主频	制程
寒武纪	训练+推理+整机	训推一体 MLU370	256TOPS	/	7nm
		训练侧 MLU290	512TOPS		
海光信息	训练	训练侧 DCU8100	/	1.5GHz	7nm
景嘉微	图形为主	JM9	1.5TFLOPS@FP32	1.5GHz	14nm
沐曦	训练+推理	推理侧 MXN100	/	/	7nm
天数智芯	由训练切入推理	训练侧 BI	295TOPS	/	7nm
壁仞	训练	训练侧 BR100	2000TOPS	/	7nm
燧原	训练+推理+整机	推理侧 i20	256TOPS	1.5GHz	/
		训练侧 T21	256TOPS		
昆仑芯	训练	训推一体昆仑芯 2 代	256TOPS	/	7nm
平头哥	推理	推理侧含光 800	820TOPS	/	12nm
海飞科	训练	云端芯片 Compass	/	/	/
后摩智能	存算一体	边缘域 A1.0	50TOPS	/	/
登临	由推理切入训练	推理侧 Goldwasser UL/	32-64TOPS/	/	/
		训练侧 Goldwasser XL	512TOPS		
摩尔线程	图形为主	MTT S3000	15.2 TFLOPS@FP32	1.9GHz	12nm
芯动科技	图形为主	风华 1 号	25TOPS	/	12nm
黑芝麻	自驾芯片	A1000	58TOPS	/	16nm
地平线	自驾芯片	征程 5	128TOPS	/	16nm

资料来源：各公司官网等，民生证券研究院整理

1.3.1 寒武纪

寒武纪成立于 2016 年，公司研发团队主要来自于中科院，董事长陈天石曾任中科院计算所研究员。寒武纪 AI 芯片思元 290 面向云端训练，思元 370 面向云端训推一体，两款芯片均采用 7nm 制程工艺。此外，思元 370 是寒武纪首款采用 chiplet 技术的 AI 芯片，最大算力高达 256TOPS；训练侧新品思元 590 在研发中。

图5：思元 370 系列板卡与业内主流 GPU 性能对比



资料来源：寒武纪官网，民生证券研究院

2月27日，寒武纪发布2022年业绩快报，公司2022年实现营收7.29亿元，同比增长1.11%；实现归母净利润-11.66亿元，亏损同比扩大41.40%。寒武纪2022年归母净利润亏损同比扩大主要原因系公司研发费用、资产减值损失、信用减值损失有所增长。

2022年，寒武纪云端产品已打入阿里云等头部互联网客户，并与头部银行等金融领域客户进行了深度技术交流，同时亦得到了头部服务器厂商的认可。展望2023年，随着ChatGPT为代表的AI大模型不断涌现，AI算力需求有望加速增长，从而驱动AI算力芯片需求增长。寒武纪作为国内领先的AI芯片公司，产品研发、市场拓展、客户导入均有较强先发优势，有望深度受益AI发展的浪潮与AI算力芯片国产化趋势，在23年取得快速成长。

1.3.2 海光信息

海光信息成立于2014年，公司骨干研发人员多拥有国内外知名芯片公司的就职背景，拥有成功研发x86处理器或ARM处理器的经验。海光信息的DCU芯片深算一号采用7nm工艺，兼容“类CUDA”环境，软硬件生态丰富，已于2021年实现规模化出货，未来将广泛应用于大数据处理、人工智能、商业计算等领域。

表5：海光深算一号与 NVIDIA、AMD 产品参数比较

项目	海光	NVIDIA	AMD
品牌	深算一号	Ampere 100	MI100
生产工艺	7nm FinFET	7nm FinFET	7nm FinFET
核心数量	4096 (64 CUs)	2560 CUDA	120CUs
内核频率	Up to 1.5GHz (FP64)	Up to 1.53Ghz	Up to 1.5GHz(FP64)
显存容量	32GB HBM2	80GB HBM2e	32GB HBM2
显存位宽	4096 bit	5120 bit	4096bit
显存频率	2.0 GHz	3.2 GHz	2.4 GHz
显存带宽	1024 GB/s	2039 GB/s	1228 GB/s
TDP	350 W	400 W	300W
CPU to GPU 互联	PCIe Gen4 x 16	PCIe Gen4 x 16	PCIe GEN4 x 16
GPU to GPU 互联	xGMI x 2, Up to 184 GB/s	NVLink up to 600 GB/s	Infinity Fabric x 3, up

资料来源：海光信息招股说明书，民生证券研究院整理

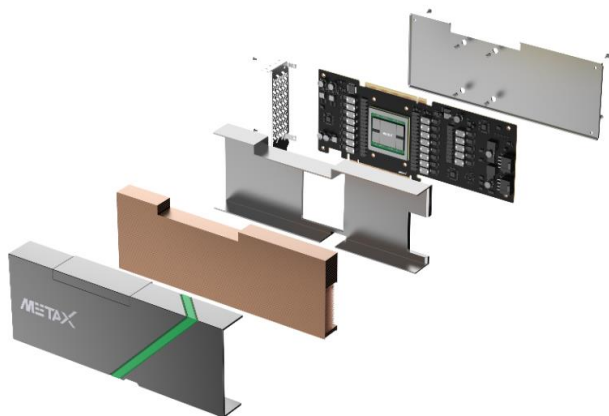
2月23日，海光信息发布2022年业绩快报，公司2022年实现营收51.25亿元，同比增长121.83%；实现归母净利8.02亿元，同比增长145.18%。2022年，海光信息在保持高研发投入的同时，大力拓展国内市场，有效提升了公司在国内高端处理器领域的领先优势和市场地位，业务稳步增长。

风险提示：行业竞争加剧的风险；行业景气度波动的风险；行业技术路线变化的风险。

1.3.3 沐曦集成电路

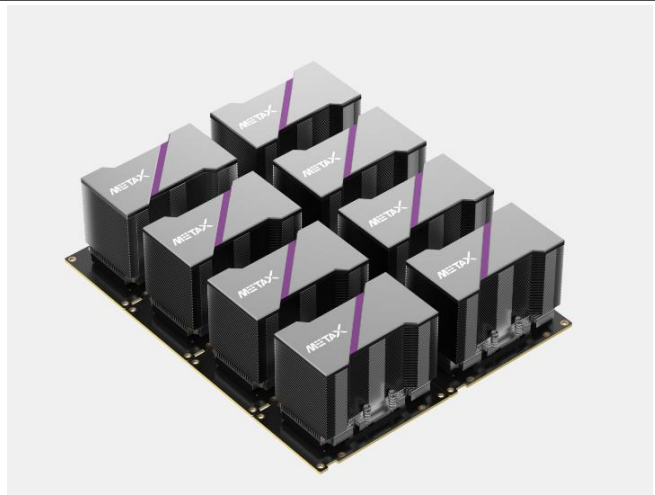
沐曦集成电路成立于2020年，公司创始团队处于国内顶尖行列——创始人陈维良曾任AMD全球GPGPU设计总负责人；两位CTO均为前AMD首席科学家，目前分别负责公司软硬件架构；核心成员平均拥有近20年高性能GPU研发经验。沐曦于2022年7月完成10亿元Pre-B轮融资，由混沌投资领投。沐曦首款异构GPU产品MXN100采用7nm制程，已于2022年8月回片点亮，主要应用于推理侧；应用于AI训练及通用计算的产品MXC500已于2022年12月交付流片，公司计划2024年全面量产。

图6: 沐曦 MXN 系列 GPU (曦思) 产品概念图



资料来源: 36 氪, 民生证券研究院

图7: 沐曦 MXC 系列 GPU (曦云) 产品概念图



资料来源: 36 氪, 民生证券研究院

1.3.4 天数智芯

天数智芯成立于 2015 年, 首席科学家郑金山为原 AMD 首席工程师, 首席技术官 Chien-Ping Lu 曾任三星全球副总裁。2022 年 7 月, 公司完成超 10 亿元人民币的 C+ 轮及 C++ 轮融资。天数智芯的 Big Island 云端 GPGPU 是一款具有自主知识产权、自研 IP 架构的 7nm 通用云端训练芯片, 这款芯片达到 295TOPS INT8 算力。

图8: MI100、英伟达 A100、Big Island 对比

GPU Name	AMD Instinct MI100	NVIDIA A100	Big Island
Process Node	TSMC 7nm	TSMC 7nm	TSMC 7nm
Architecture	CDNA 1	Ampere	Unknown
Transistors	50 Billion	54 Billion	24 Billion
Cores	7680	6912	TBC
Memory	32 GB HBM2	40 GB HBM2	32 GB HBM2
Memory Bandwidth	1.2 TB/s	1.6 TB/s	1.2 TB/s
FP32 Compute	23.1 TFLOPs 46.1 TFLOPs (Matrix)	19.5 TFLOPs 156 TFLOPs (Tensor) 312 TFLOPs (sparsity)	37 TFLOPs
BFLOAT16 Compute	92.3 TFLOPs	312 TFLOPs 624 TFLOPs (sparsity)	147 TFLOPs
TDP	300W	400W	300W

资料来源: 天数智芯, 民生证券研究院

1.3.5 壁仞科技

壁仞科技成立于 2019 年，创始人张文曾任商汤科技总裁，CTO 洪洲曾任职于海思的 GPU 自研团队，软件生态环境主要负责人焦国方曾创建高通公司骁龙 GPU 团队、领导了 5 代 Adreno GPU 架构开发。2021 年 3 月，公司完成 B 轮融资，累计融资金额超 47 亿元。2022 年 8 月，壁仞科技发布首款 GPGPU 芯片 BR100，BR100 芯片采用 chiplet 技术，其 16 位浮点算力达到 1000T 以上、8 位定点算力达到 2000T 以上。

图9：BR100 的封装工艺



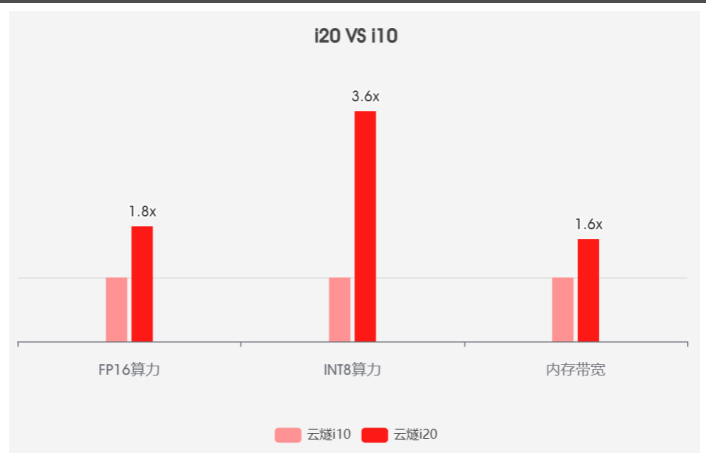
资料来源：壁仞科技，民生证券研究院

1.3.6 燧原科技

对于互联网大厂来说，腾讯、百度、阿里巴巴等均在 AI 芯片领域大力布局。其中，腾讯投资燧原科技、百度投资昆仑芯、阿里巴巴则孵化了平头哥。

燧原科技成立于 2018 年，公司创始人赵立东曾任紫光通信科技集团有限公司副总裁、AMD 计算事业部高级总监；COO 张亚林曾任 AMD 资深芯片经理、技术总监。公司最新发布的第二代推理产品云燧 i20 是面向数据中心应用的第二代人工智能推理加速卡，采用 12nm 工艺，通过架构升级大大提高了单位面积的晶体管效率，算力可媲美 7nm GPU，达到 256TOPS。

图10：云燧 i20 VS i10 性能对比

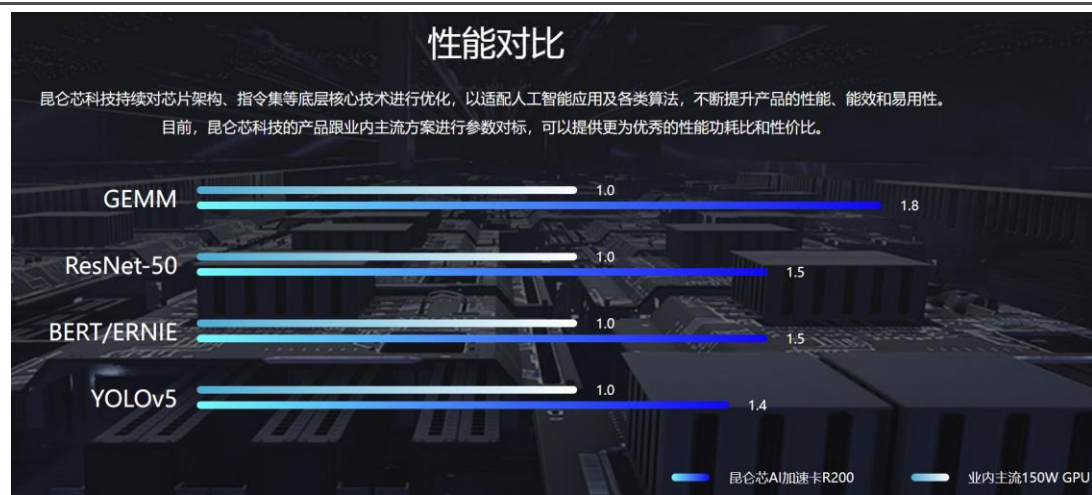


资料来源：燧原科技官网，民生证券研究院

1.3.7 昆仑芯

昆仑芯成立于 2011 年，前身为百度智能芯片及架构部，于 2021 年 4 月完成独立融资，首轮估值约 130 亿元。公司成员多数成员来自百度、高通、Marvell、Tesla 等行业头部公司，22 年完成 A 轮融资。昆仑芯新品 AI 芯片 R200 于 2022 智算峰会上正式发布，基于新一代昆仑芯自研架构 XPU-R，通用性和性能显著提升，采用 7nm 先进工艺，算力可达 256TOPS。配合百度飞桨平台，获得更友好开发的环境。

图11：昆仑芯 R200 对比业内主流 GPU



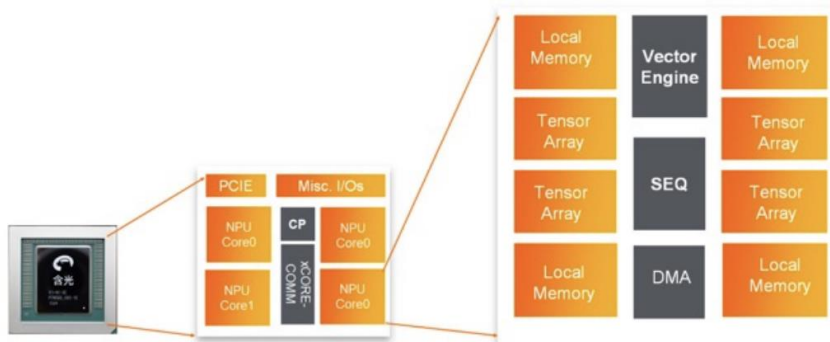
资料来源：昆仑芯官网，民生证券研究院

1.3.8 平头哥

平头哥成立于 2018 年，由阿里全资收购的中天微与达摩院芯片研发团队合并

而来。公司有两条研发主线，一是利用 ARM 的 IP 为阿里云数据中心研发芯片，在云端提供普惠算力，即倚天系列和含光系列；另一边集中在 RISC-V 处理器架构的研发，如玄铁系列，主要应用是在 AIoT 领域。公司技术团队由原中天微、高通、AMD、华为海思等拥有丰富研发经验的人员构成。平头哥于 2019 年 9 月发布首颗数据中心芯片含光 800，采用 12nm 工艺，性能峰值算力达 820 TOPS。

图12：含光 800NPU 架构图



资料来源：平头哥官网，民生证券研究院

2 自主可控受高度重视, 国产 AI 算力芯片迎“芯”机遇

3月5日, 中共中央总书记、国家主席、中央军委主席习近平在参加十四届全国人大一次会议江苏代表团审议时指出, **加快实现高水平科技自立自强, 是推动高质量发展的必由之路**。此外, 在“两会”期间, 国务院组建国家数据局、重组科技部, 结合前期数字中国建设规划落地、浪潮集团被美国商务部列入“实体清单”等事件, **我们认为, 国产 AI 算力芯片迎来发展“芯”机遇, 自主可控亟待加速**。

自主可控受高度重视, 高水平科技亟需自立自强。根据新华网, 3月5日, 中共中央总书记、国家主席、中央军委主席习近平在参加他所在的十四届全国人大一次会议江苏代表团审议时指出, 加快实现高水平科技自立自强, 是推动高质量发展的必由之路。此外, 国务院在“两会”期间组建国家数据局, 负责统筹推进数字经济发展、推进数据基础设施布局建设等任务, 有望加速推进数字中国建设落地; 国务院重组科技部, 并组建中央科技委员会, 亦有望加速推动国内高水平科技实现自立自强。AI 算力芯片作为数字中国的算力基础, 国产突破势在必行, 国产 AI 算力芯片迎来发展“芯”机遇。

数字中国建设对 AI 芯片国产化提出新要求。2月27日, 中共中央、国务院印发了《数字中国建设整体布局规划》。《规划》提出要夯实数字基础设施和数据资源体系“两大基础”, 我们认为, 数字中国基础设施的建设有望拉动以数据中心、超算中心、智能计算中心为代表的算力基础设施建设, 从而带动服务器与 AI 算力芯片的需求快速增长。同时, 《规划》提出要构筑自立自强的数字技术创新体系, 上游 AI 芯片作为算力基础, 自主可控需求凸显, 数字中国建设对 AI 芯片国产化提出新要求。

浪潮集团被列入“实体清单”, AI 芯片国产化进程有望加速。当地时间3月2日, 美国商务部将浪潮集团、龙芯中科等37个实体列入“实体清单”, 扩大对浪潮等中国企业的出口禁令。浪潮集团等公司被列入“实体清单”, 再度敲响了中国 AI 产业发展的警钟, 我们不但要加大数字基础设施建设, AI 算力芯片的自主可控推进也势在必行。

3 投资建议

数字中国建设对 AI 芯片国产化提出新要求，美国商务部将浪潮集团、龙芯中科等列入“实体清单”亦有望加速 AI 算力芯片国产化进程。我们看好以海光信息、寒武纪、天数智芯、沐曦为代表的 AI 算力芯片公司持续突破，以及以芯原股份为代表的 IP 公司在 AI 芯片大潮下加速成长。建议关注 AI 芯片相关赛道投资机遇。

表6：电子行业重点关注个股

代码	简称	股价 (元)	EPS (元)			PE (倍)			评级
			2022E	2023E	2024E	2022E	2023E	2024E	
688041	海光信息	57.50	0.38	0.82	1.21	151	70	48	推荐
688256	寒武纪-U	122.93	-2.91	-1.79	-1.00	/	/	/	/
688521	芯原股份-U	77.02	0.15	0.33	0.52	520	230	147	推荐
002156	通富微电	23.38	0.39	0.74	1.06	60	32	22	/
600584	长电科技	30.91	1.85	2.04	2.37	17	15	13	/

资料来源：Wind，民生证券研究院预测；

(注：股价为 2023 年 03 月 17 日收盘价；未覆盖公司数据采用 wind 一致预期；海光信息、寒武纪-U、芯原股份-U 2022E EPS 为业绩快报披露值)

4 风险提示

疫情反复影响生产经营：疫情影响因素仍未完全消除，若疫情反复或政策趋严或影响企业正常生产经营节奏；

下游需求不及预期：如果 AI 应用渗透不及预期，则将对上游需求带来不利影响；

研发进展不及预期：AI 算力芯片设计难度较大，具有较高技术壁垒，若国内厂商研发进度不及预期，将对相关公司业绩带来不利影响。

插图目录

图 1: 云端、边缘和终端 AI 芯片应用情况	4
图 2: 2017-2023 财年英伟达数据中心营收 (亿美元)	4
图 3: 常见 AI 芯片评价标准	5
图 4: 英伟达 CUDA 生态系统	6
图 5: 思元 370 系列板卡与业内主流 GPU 性能对比	9
图 6: 沐曦 MXN 系列 GPU (曦思) 产品概念图	11
图 7: 沐曦 MXC 系列 GPU (曦云) 产品概念图	11
图 8: MI100、英伟达 A100、Big Island 对比	11
图 9: BR100 的封装工艺	12
图 10: 云燧 i20 VS i10 性能对比	13
图 11: 昆仑芯 R200 对比业内主流 GPU	13
图 12: 含光 800NPU 架构图	14

表格目录

重点公司盈利预测、估值与评级	1
表 1: 主流 AI 芯片分类	3
表 2: 英伟达 V100/A100/A800/H100 对比	6
表 3: 国产 AI 算力芯片公司对比	7
表 4: 国产 AI 算力芯片公司产品对比	8
表 5: 海光深算一号与 NVIDIA、AMD 产品参数比较	10
表 6: 电子行业重点关注个股	16

分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并登记为注册分析师，基于认真审慎的工作态度、专业严谨的研究方法与分析逻辑得出研究结论，独立、客观地出具本报告，并对本报告的内容和观点负责。本报告清晰地反映了研究人员的研究观点，结论不受任何第三方的授意、影响，研究人员不曾因、不因、也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

评级说明

投资建议评级标准	评级	说明
以报告发布日后的 12 个月内公司股价（或行业指数）相对同期基准指数的涨跌幅为基准。其中：A 股以沪深 300 指数为基准；新三板以三板成指或三板做市指数为基准；港股以恒生指数为基准；美股以纳斯达克综合指数或标普 500 指数为基准。	推荐	相对基准指数涨幅 15%以上
	谨慎推荐	相对基准指数涨幅 5% ~ 15%之间
	中性	相对基准指数涨幅-5% ~ 5%之间
	回避	相对基准指数跌幅 5%以上
	推荐	相对基准指数涨幅 5%以上
	中性	相对基准指数涨幅-5% ~ 5%之间
	回避	相对基准指数跌幅 5%以上

免责声明

民生证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。

本报告仅供本公司境内客户使用。本公司不会因接收人收到本报告而视其为客户。本报告仅为参考之用，并不构成对客户的投资建议，不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，客户应当充分考虑自身特定状况，不应单纯依靠本报告所载的内容而取代个人的独立判断。在任何情况下，本公司不对任何人因使用本报告中的任何内容而导致的任何可能的损失负任何责任。

本报告是基于已公开信息撰写，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、意见及预测仅反映本公司于发布本报告当日的判断，且预测方法及结果存在一定程度局限性。在不同时期，本公司可发出与本报告所刊载的意见、预测不一致的报告，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问、咨询服务等相关服务，本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

若本公司以外的金融机构发送本报告，则由该金融机构独自为此发送行为负责。该机构的客户应联系该机构以交易本报告提及的证券或要求获悉更详细的信息。本报告不构成本公司向发送本报告金融机构之客户提供的投资建议。本公司不会因任何机构或个人从其他机构获得本报告而将其视为本公司客户。

本报告的版权仅归本公司所有，未经书面许可，任何机构或个人不得以任何形式、任何目的进行翻版、转载、发表、篡改或引用。所有在本报告中使用的商标、服务标识及标记，除非另有说明，均为本公司的商标、服务标识及标记。本公司版权所有并保留一切权利。

民生证券研究院：

上海：上海市浦东新区浦明路 8 号财富金融广场 1 幢 5F； 200120

北京：北京市东城区建国门内大街 28 号民生金融中心 A 座 18 层； 100005

深圳：广东省深圳市福田区益田路 6001 号太平金融大厦 32 层 05 单元； 518026