科技专题研究

2023年4月6日

中航证券有限公司
AVIC SECURITIES CO., LTD.

AI大模型开启新一轮大国竞争，半导体战略地位凸显

行业评级：增持

分析师：刘牧野
证券执业证书号：S0640522040001

股市有风险 入市需谨慎

核心观点

- **AI正处史上最长繁荣大周期：**在进入21世纪以来，在大数据和大算力的支持下，归纳统计方法逐渐占据了人工智能领域的主导地位，深度学习的浪潮席卷人工智能，人工智能迎来史上最长的第三次繁荣期，至今仍未有结束的趋势。
- **OpenAI的“暴力美学”：大算力和大数据：**OpenAI 认为，通过独立延长模型训练时间、增加训练数据量或者扩大模型参数规模，预训练模型在测试集上的 Test Loss 都会单调降低，从而使模型效果越来越好。我们认为，在 Scaling Law 的框架下，只要追加数据与算力，大模型的能力就能持续增强。**对于OpenAI 而言，目前大模型的最大限制是数据和算力的总量。**
- **大模型开启新一轮大国竞争，半导体成顶层博弈焦点：**预训练大模型是现阶段人工智能的集大成者，代表了统计学习流派的最高成就。在新一代技术未出现前，它将是人工智能研究和开发的最强武器。围绕大模型的研发和落地，中美之间已经展开了新一轮的竞争，美国已对华限制销售最先进的英伟达A100和H100 GPU 训练芯片。**半导体作为AI算力核心，将受到顶层高度关注，成为大国博弈的焦点之一。**
- **AI模型运算规模增长，算力缺口巨大：**基于大量数据训练、拥有巨量参数的AI预训练模型—GPT-3，引发了AIGC技术的质变，从而诞生ChatGPT。然而，预训练模型参数数量、训练数据规模将按照 300 倍/年的趋势增长，现有算力距离AI应用存巨大鸿沟。运算规模的增长，带动了对AI训练芯片单点算力提升的需求，并对数据传输速度提出了更高的要求。
- **建议关注：**
- GPU：景嘉微、航锦科技，海光信息和未上市的地平线、黑芝麻、摩尔线程
- AI训练芯片：寒武纪、商汤（港股）、燧原科技（未上市）
- AI存力：兆易创新、北京君正、东芯股份
- HBM：雅克科技、深科技
- 半导体大国重器：中芯国际、北方华创、中微公司
- **风险提示：**AI算法、模型存较高不确定性，AI技术发展不及预期；ChatGPT用户付费意愿弱，客户需求不及预期；针对AI的监管政策收紧

目 录

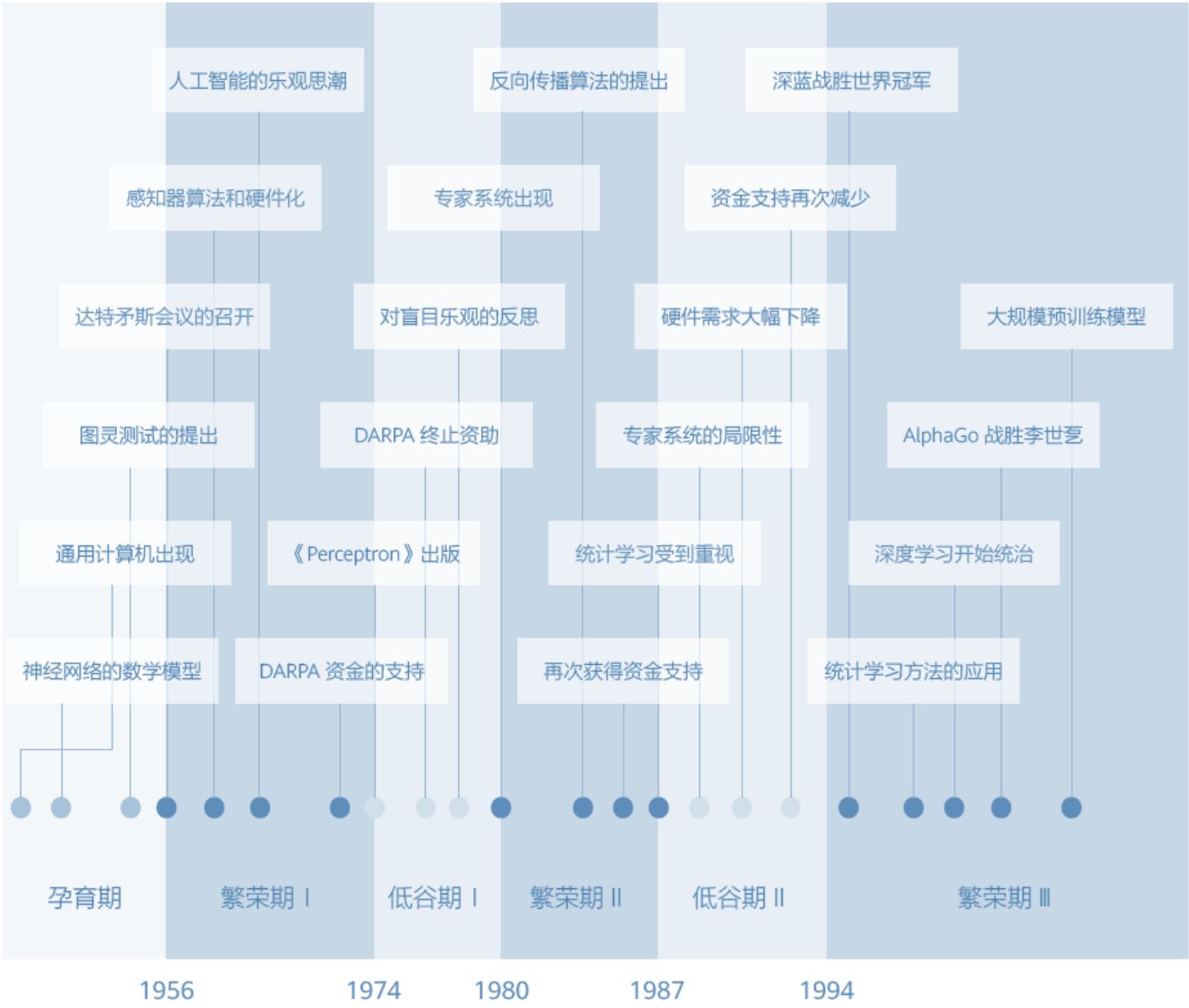
AI正处史上最长繁荣大周期

- 人工智能从1956 年被正式提出以来，经历了数十年的发展历程。人工智能诞生初期，其研究主要分为三个流派，即逻辑演绎、归纳统计和类脑计算。

- 人工智能研究的三大流派各有优劣势。类脑计算流派的目标最为宏远，但在未得到生命科学的支撑之前，难以取得实际应用。归纳演绎流派的思考方式与人类相似，具有较强的可解释性。由于对数据和算力的依赖较少，归纳演绎流派成为人工智能前两次繁荣的主角。随着学界对人工智能困难程度的理解逐渐加深，数理逻辑方法的局限性被不断放大，并最终在第三次繁荣期中，逐渐让位于统计学习的“暴力美学”。

- **在进入21世纪以来，在大数据和大算力的支持下，归纳统计方法逐渐占据了人工智能领域的主导地位，深度学习的浪潮席卷人工智能，人工智能迎来史上最长的第三次繁荣期，至今仍未有结束的趋势。**

图：人工智能发展史

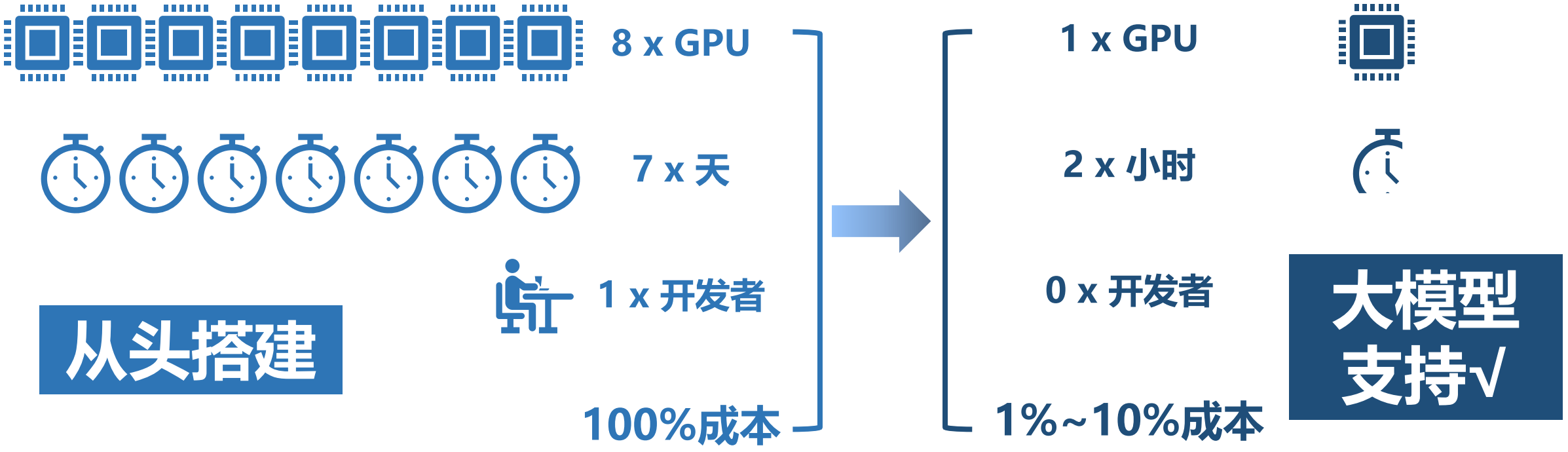


资料来源：华为，中航证券研究所

通用大模型加持，平民化AI普惠千行百业

- **深度学习依然受到统计学习的框架限制：特征抽取和模板匹配。**相比于人类基于知识的推断，这种方式无疑是低效的，因为对于任何新的概念乃至新的实体，算法都需要专门的训练数据来提供相关的信息。**在没有基础模型支撑的情况下，开发者们必须从头开始完成收集数据、训练模型、调试模型、优化部署等一系列操作。**对于人工智能开发者和垂直细分行业应用而言，都是重大的挑战。
- **预训练大模型降本增效，将推动AI普惠千行百业。**预训练大模型加持下的人工智能算法（包括计算机视觉、自然语言处理等），相比于普通开发者从头搭建的算法，精度明显上升、数据和计算成本明显下降，且开发难度大幅降低。

图：在100 张图像上训练基础物体检测算法，从头搭建 vs 大模型支持

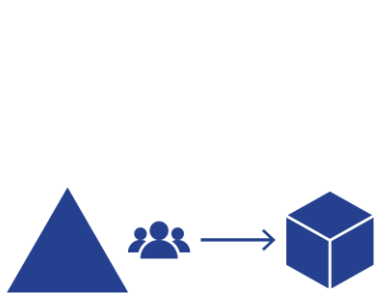


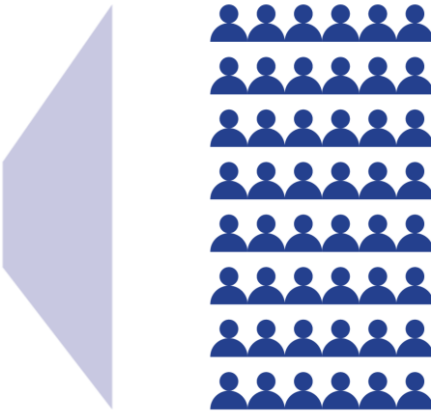
资料来源：华为，中航证券研究所

GPT基础大模型驱动，引发AIGC范式革命

- **以ChatGPT为代表的AIGC应用在 2022 年的爆发，主要是得益于深度学习模型方面的技术创新。**不断创新的生成算法、预训练模型、多模态等技术融合带来了 AIGC （AI Generated Content）技术变革，拥有通用性、基础性多模态、参数多、训练数据量大、生成内容高质稳定等特征的 AIGC 模型成为了自动化内容生产的“工厂”和“流水线”。
- **基础层是核心，GPT-3模型起关键支撑作用。**GPT-3一个大规模的通用语言模型，已经在来自各种来源的大量文本数据上进行了训练。能够产生类似人类的反应，并可用于广泛的语言相关任务。
- ChatGPT基于目前较新的GPT-4模型版本进行研发，专注于自然语言对话，接受了更广泛的语言模式和风格培训，因此，能较GPT-4产生更多样化和微妙的响应。

图：AIGC产业架构





基础层	中间层	应用层
预训练模型	重直化、场景化、个性化模型	图像、语音和文字生成等各种各样的AIGC的应用
以预训练**大规模模型**为基础搭建的 AIGC 技术基础设施层。	在基础层上生成场景化、定制化、个性化的**小模型**，实现不同行业、垂直领域的流水线式部署	面向 **C 端**用户的文字、图片、音视频等内容生成服务

资料来源：腾讯，中航证券研究所

OpenAI的“暴力美学”：大算力和大数据

- **穷尽所有的测试数据和训练材料，AI就会呈现出恐怖的准确率。**OpenAI 意识到了“大”和“规模”的力量，沿着该路径狂飙，阅览了几乎所有互联网数据，并在超级复杂的模型之下进行深度学习。
- 2017-2019年，OpenAI 做出了有别于市场共识的关键决策，公司在Transformer 基础上押注大算力和大数据的“暴力美学”。并在GPT-3后迅速引入了人类反馈，让模型的语言前后逻辑更加明晰、有因果关联。

图：OpenAI决策路径

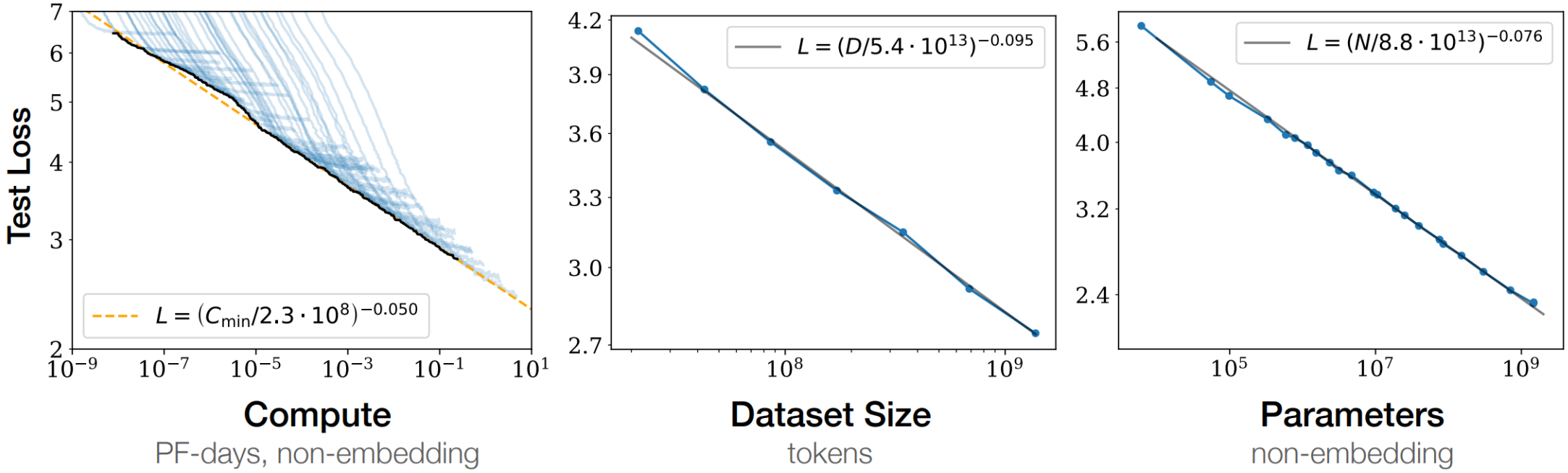
2015-2016	2017-2018	2018-2019	2018-2019	2018-2019	2019-2020	2020-2021
OpenAI 的决策						
早期 ML Engineering 能力和基础设施建设没有落后于行业,甚至目前比 Google 内部的还好用。	从 Unsupervised sentiment neuron 工作开始,逐渐将精力和关注点分配更多给语言模型上。	迅速和深度转向 Transformer，没有在 CNN/RNN 等上一代特征提取器上浪费时间。	在行业对强化学习的效果充满争议的情况下，在 Dota 及之后的项目中坚持探索深度强化学习。	在语言模型中坚持了仅有上文背景的 GPT 式生成式路线，没有追随 BERT 狂潮陷入理解式路线。	团队持续思考 Scaling Law 的问题，在 Transformer 基础上押注大规模数据和算力。	在长期强调安全和使用无监督强化学习的情况下，在 GPT-3 工作完成后迅速引入人类反馈。
当时市场的主流认知						
• AI 的突破是一项研究工作，而非工程问题； • 每个探索 AGI 的公司在工程能力和基建并不会有明显差距。	• OpenAI 的这个工作是优化别的任务时的副作用，歪打正着； • 语言模型不是通往 AGI 的道路。	• Transformer 彻底抛弃了之前 CNN、RNN 等网络结构； • 前几年统治 AI 进展的 CV 圈并不买账 Transformer。	• 深度强化学习的效率非常低； • 强化学习设置奖励函数非常 tricky； • 它会陷入局部最优，并且通常难以稳定复现效果。	• BERT 代表着未来，GPT 只是基于 Transformer 的过渡性技术； • GPT 白白丢掉了下文的信息，在许多自然语言理解任务上都难以和 BERT 竞争。	• AI 的进步来源于算法的创新； • 算力在过去 10 年的进步不一定在未来 10 年持续。	• 随着模型变得更智能，Alignment 问题可以自动解决,人类反馈多此一举； • 人类反馈违反了无监督的原教旨并且缺少可拓展性。

资料来源：拾象科技，中航证券研究所

OpenAI的“暴力美学”：大算力和大数据

- OpenAI 在《Scaling Laws for Neural Language Models》中提出语言大模型所遵循的“规模法则”（Scaling Law）。
- Scaling Law 说明：通过独立延长模型训练时间（Compute）、增加训练数据量（Dataset Size）或者扩大模型参数规模（Parameters），预训练模型在测试集上的 Test Loss 都会单调降低，从而使模型效果越来越好。
- 我们认为，在 Scaling Law 的框架下，只要追加数据与算力，大模型的能力就能持续增强。对于OpenAI 而言，**目前大模型的最大限制是数据和算力的总量。**

图：Scaling Law：规模越大，模型越精确



资料来源：《Scaling Laws for Neural Language Models》，中航证券研究所

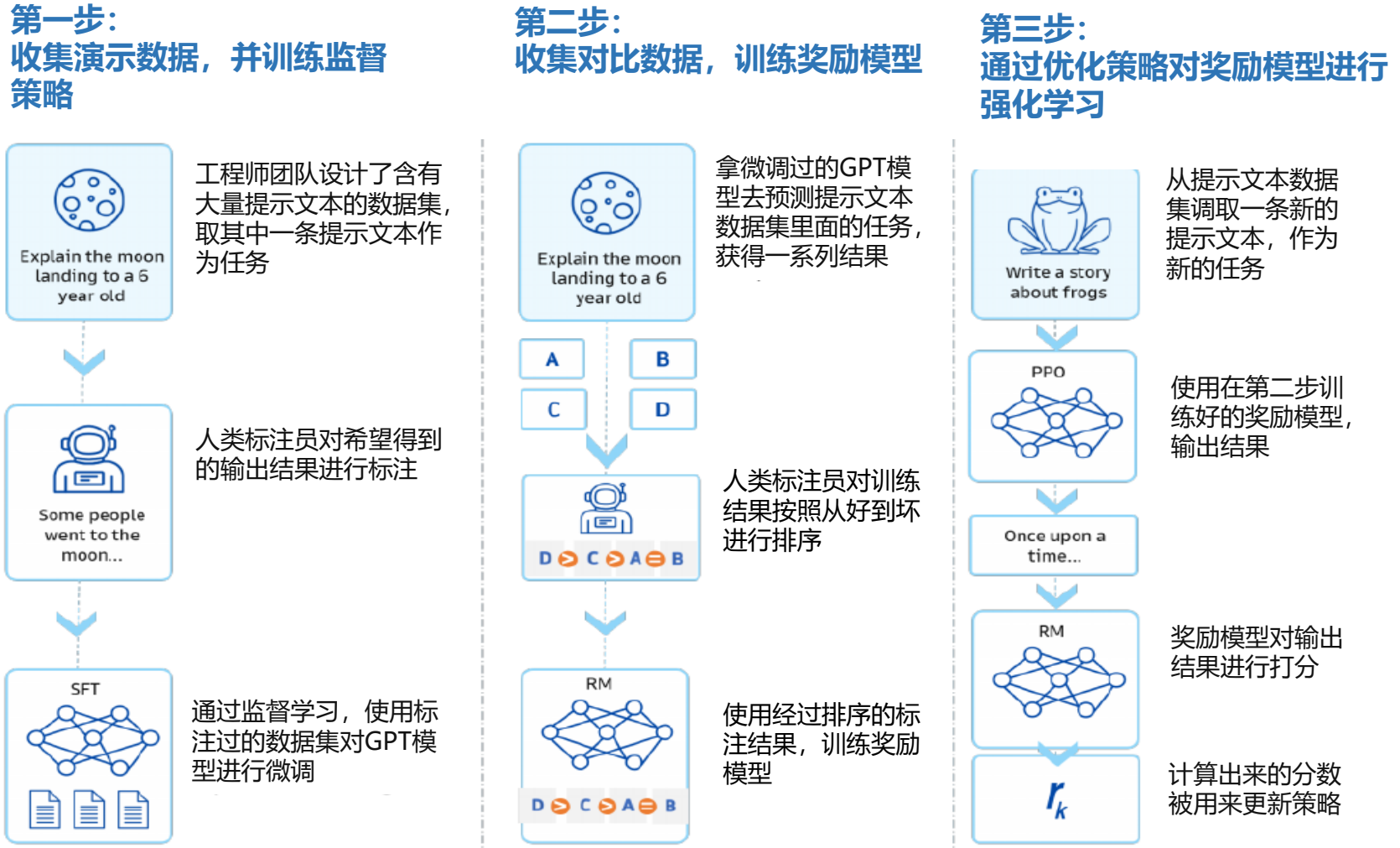
精细化策略+标注提升ChatGPT模型效果

- **预训练大模型分为上游（模型预训练）和下游（模型微调）两个阶段。**上游阶段主要是收集大量数据，并且训练超大规模的神经网络，以高效地存储和理解这些数据；而下游阶段则是在不同场景中，利用相对较少的数据量和计算量，对模型进行微调，以达成特定的目的。

- **ChatGPT的训练过程也遵循预训练大模型的基本原理。**结合了监督学习和强化学习，并且通过人工标注让模型更好地区别回复的好坏。

- 我们认为，ChatGPT在模型和数据等环节进行了大量的细节优化，高质量的海量数据加上充分的训练，人工和算法的有机配合，使ChatGPT在模型层面实现领跑。

图：ChatGPT的训练原理

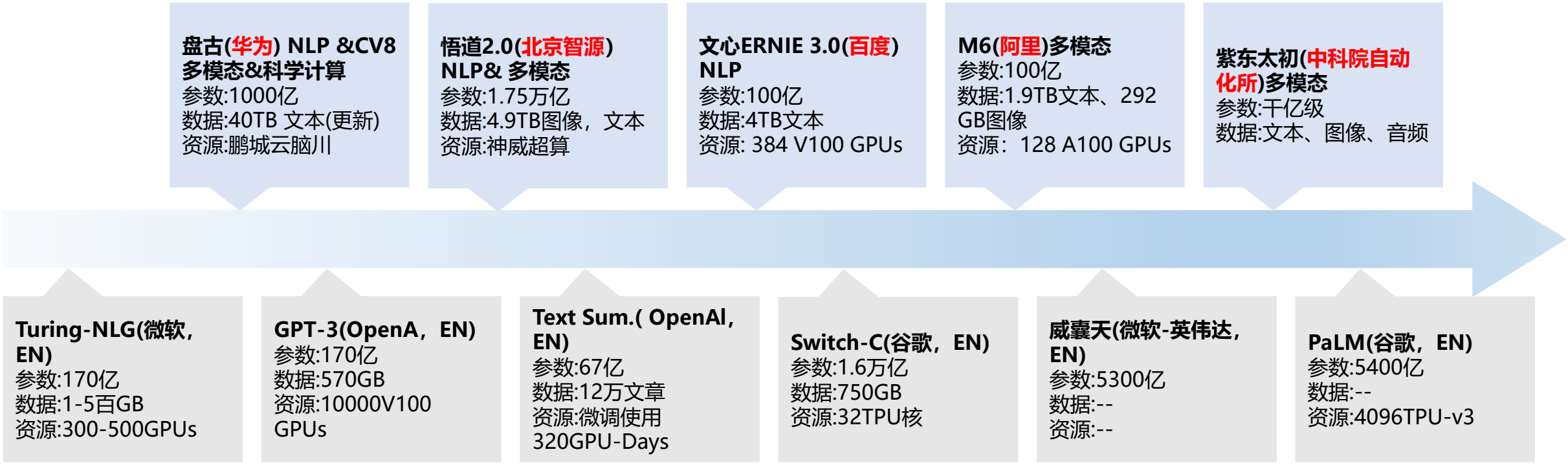


资料来源：高盛，中航证券研究所

大模型开启新一轮大国博弈

- 预训练大模型是现阶段人工智能的集大成者，代表了统计学习流派的最高成就。在新一代技术未出现前，它将是人工智能研究和开发的最强武器。**围绕大模型的研发和落地，中美之间已经展开了新一轮的竞争。**
- 中国科学技术部高新技术司司长陈家昌，于2023年4月3日在国务院新闻办公室新闻发布会上表示，在人工智能方面，**科技部专门加强顶层设计，成立人工智能规划推进办公室，启动实施新一代人工智能重大科技项目。**
- 国内科技企业纷纷对ChatGPT发表看法，百度、华为、腾讯、阿里巴巴等大多数头部企业表示，已经拥有、在研对标ChatGPT相关的模型及产品。

图：中美大模型对比



资料来源：华为，中航证券研究所

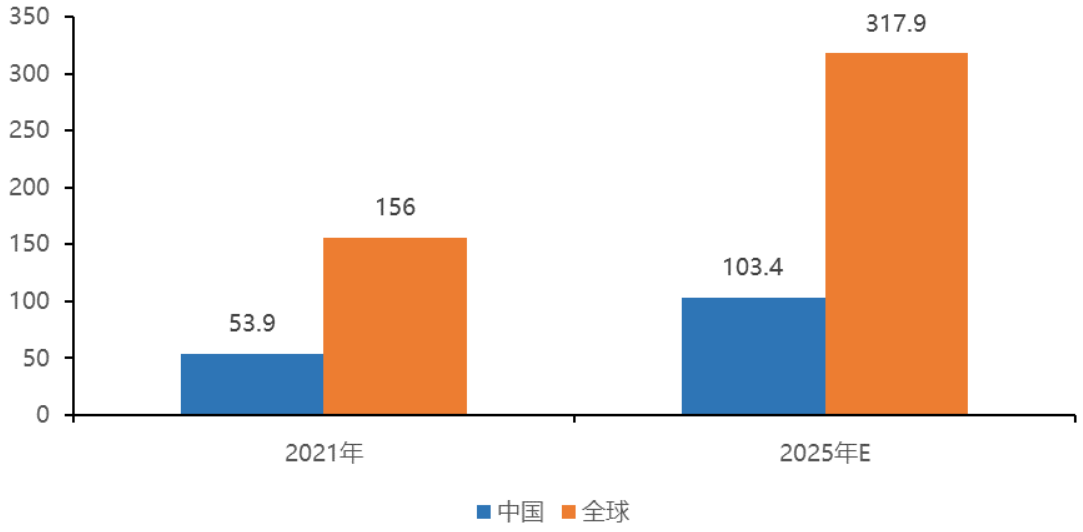
目 录

大国AI竞赛，国内AI支出规模有望高速增长

- 据IDC，中国人工智能（AI）市场支出规模将在2023年增至147.5亿美元，约占全球总规模十分之一。2021年中国加速服务器市场规模达到53.9亿美元（约350.3亿人民币），同比+68.6%；预计到2025年将达到103.4亿美元。年复合增长率为19%，占全球整体服务器市场近三成。
- 我们认为，预训练大模型是现阶段人工智能的集大成者，代表了统计学习流派的最高成就。在新一代技术未出现前，它将是人工智能研究和开发的最强武器。围绕大模型的研发和落地，中美之间已经展开了新一轮的竞争。因此，国内人工智能的支出增速有望超过IDC的预测。

图：中国人工智能市场支出预测（百万美元）

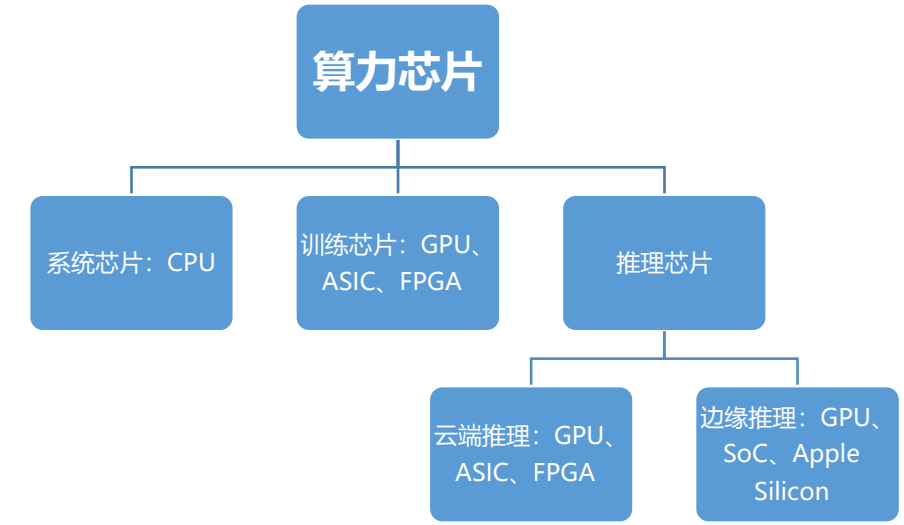
图：全球及中国AI服务器市场规模（亿美元）

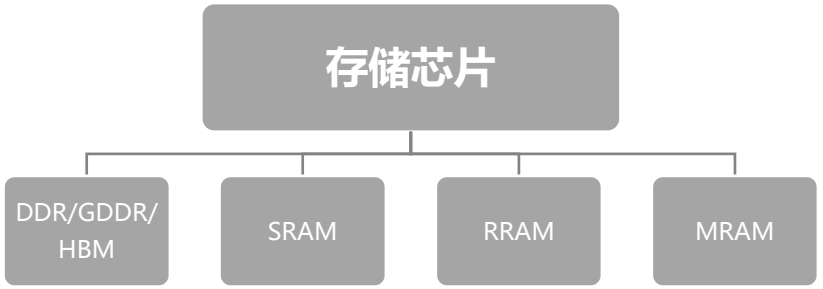


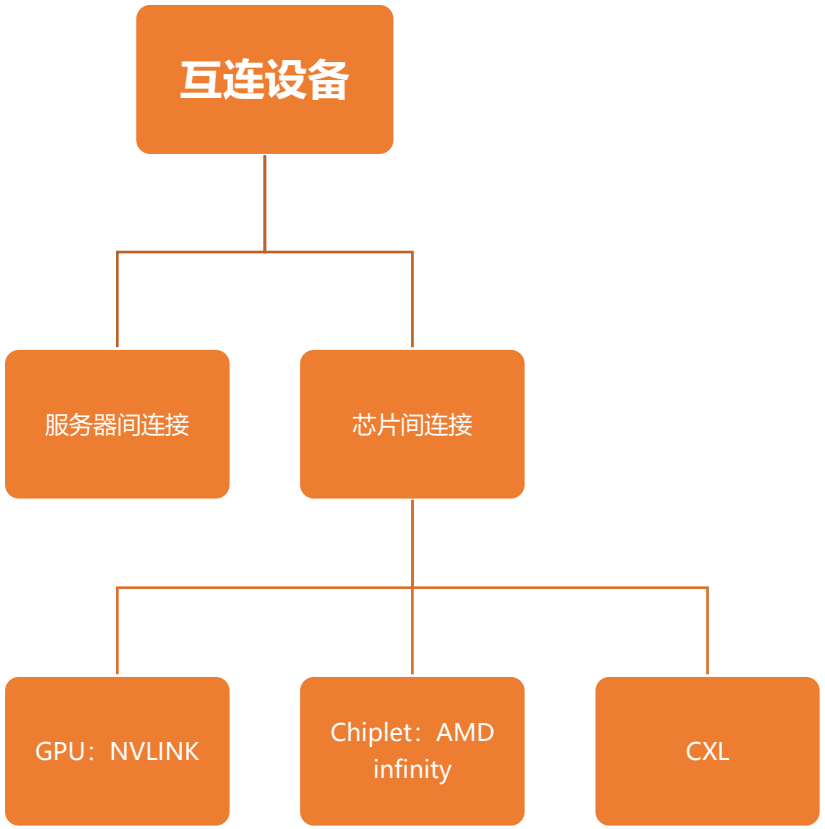
资料来源：IDC，中航证券研究所

算力芯片主导AI计算市场

- AI 分布式计算的市场主要由算力芯片 (55-75%)、内存 (10-20%) 和互联设备（10-20%）三部分组成。美国已限制对华销售最先进、使用最广泛的AI训练GPU—英伟达 A100以及H100，国产算力芯片距离英伟达最新产品存在较大差距，但对信息颗粒度要求较低的推理运算能实现部分替代。
- **我们认为，训练芯片受限进一步强调了高制程芯片设计、代工的国产替代紧迫性。而随着人工智能的应用普及，推理芯片的市场需求将加速增长。**





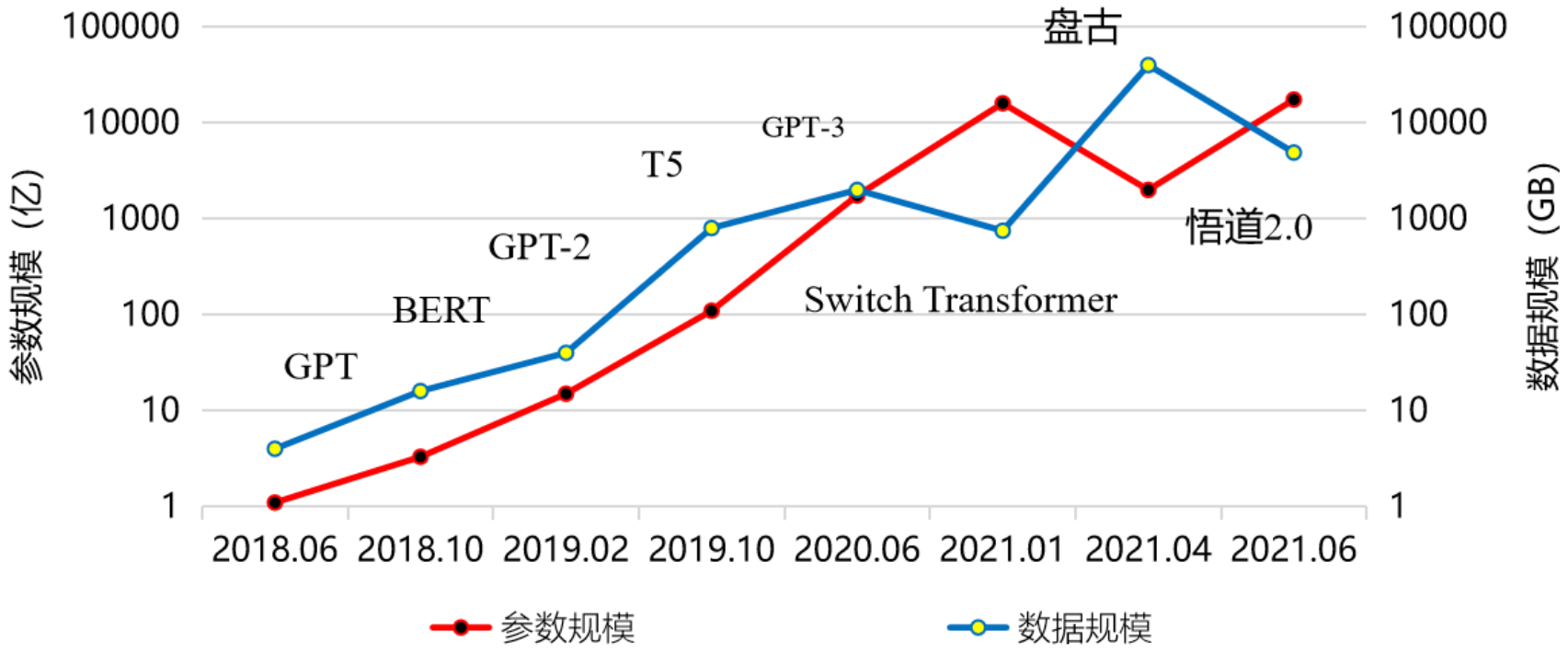


资料来源：拾象科技，中航证券研究所

AI模型数据规模增长，AI算力需求井喷

- 当前，预训练模型参数数量、训练数据规模按照 300 倍/年的趋势增长，继续通过增大模型和增加训练数据仍是短期内演进方向。未来使用更多种图像编码、更多种语言、以及更多类型数据的预训练模型将会涌现。
- **当前算力距离AI应用存巨大鸿沟。根据 Open AI 数据，模型计算量增长速度远超人工智能硬件算力增长速度，存在万倍差距。**英特尔表示，目前的计算、存储和网络基础设施远不足以实现元宇宙愿景，而要想实现真正的元宇宙，目前的计算能力需量要再提高1000倍。

图：大模型参数量和训练数据规模增长迅速

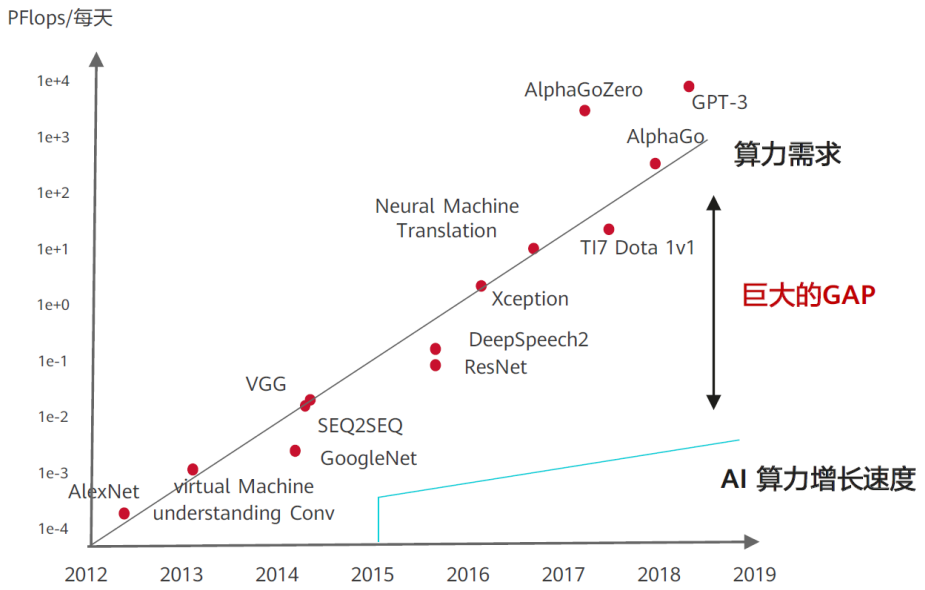


资料来源：中国信通院，中航证券研究所

AI模型数据规模增长，AI算力需求井喷

- 据IDC预计， 2021-2026年期间，中国智能算力规模年复合增长率达52.3%。2022年智能算力规模将达到268.0 EFLOPS，预计到2026年智能算力规模将进入每秒十万亿亿次浮点计算（ZFLOPS）级别，达到1,271.4 EFLOPS。
- **运算数据规模的增长，带动了对AI训练芯片单点算力提升的需求，并对数据传输速度提出了更高的要求。**

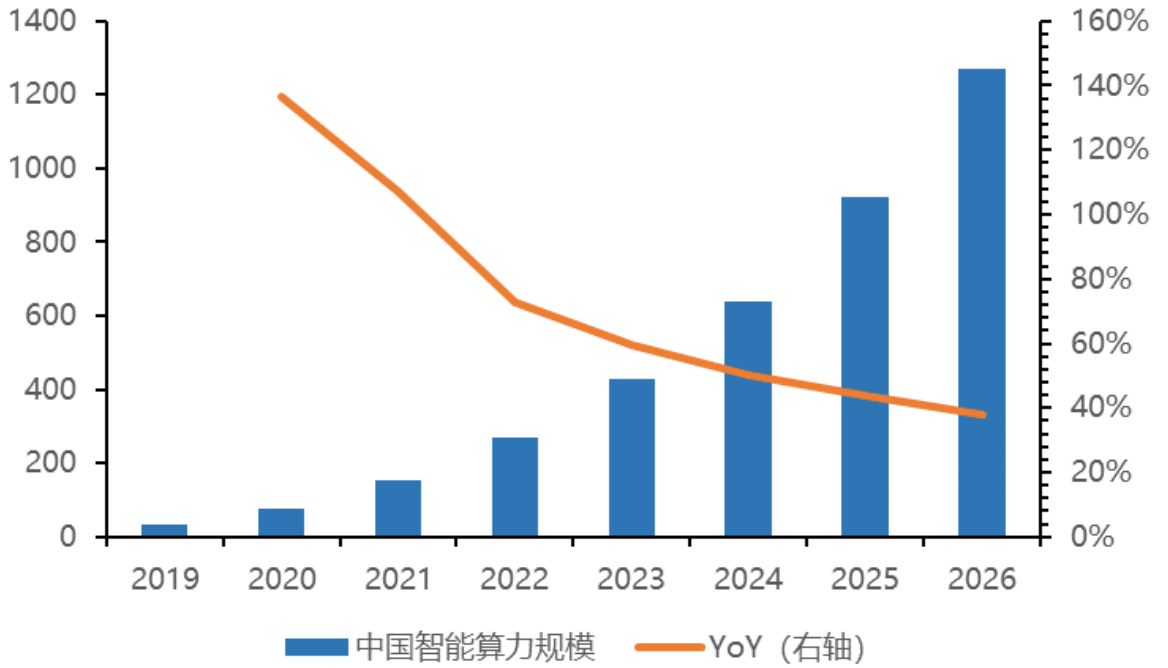
图：2012至2019年算力需求增长近30万倍



AI模型训练花费不菲

GPT-3	460万美金	10000块GPU * 13天

图：中国智能算力规模百亿亿次浮点运算/秒（EFLOPS）



资料来源：华为，IDC，中航证券研究所

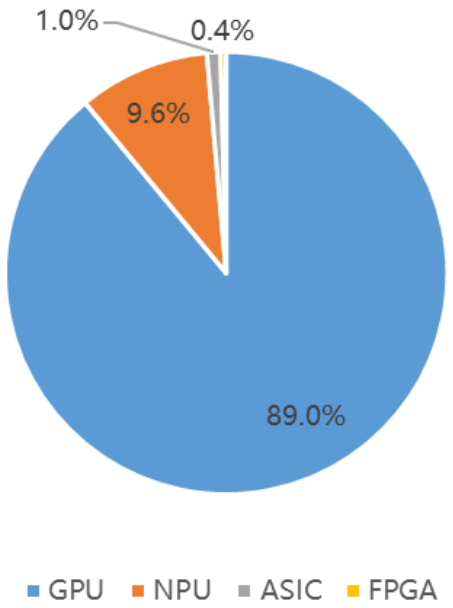
算力升级：AI训练芯片空间广阔

- IDC预计，到2025年人工智能芯片市场规模将达726亿美元。IDC全球范围调研显示，人工智能芯片搭载率将持续增高。目前每台人工智能服务器上普遍多配置2个GPU，未来18个月，GPU、ASIC和FPGA的搭载率均会上升。**通用性递减，专用性增强，为AI芯片的主要发展方向。**
- 2021年中国以GPU为主实现数据中心计算加速，GPU在算力芯片的市场占有率接近90%。ASIC，FPGA，NPU等非GPU芯片市场占有率超过10%。国际科技网络巨头公司谷歌、脸书，亚马逊等等在AI芯片领域从云端训练到终端产品应用，在开源框架赋能产业行业上有一定的领先优势。国内企业也在打造从AI芯片注重云端训练+AI芯片终端响应+AI算法框架开源的生态体系。**建议关注面向 GPU 的创新企业，包括景嘉微、航锦科技，和未上市的地平线、黑芝麻、摩尔线程等。以及面向基于ASIC架构、感知识别等AI训练芯片公司，如寒武纪、商汤（港股）、燧原科技（未上市）等。**

表：AI芯片架构及发展方向

	芯片架构	芯片特点	代表公司	专用性（L1到L5依次增强）
发展方向一：从通用到专用	CPU	CPU的通用架构设计使运行效率受限。当前CPU虽然在机器学习领域的计算大大减少,但是不会被完全取代。	英特尔	L1
	GPU	目前商用最广泛的AI芯片, 可以执行深度学习和神经网络任务。GPU主要从事大规模并行计算,比CPU运行速度快,并且比其他专用AI处理器芯片价格低。	英伟达、AMD	L2
	DSP	仅作为处理器IP核使用。目前基于DSP的设计有一定的局限性,一般都是针对图像和计算机视觉的处理器IP核芯片,速度较快,成本不高。	新思科技、Cadence	L3
	FPGA	FPGA具有三大优点:单位能耗比低、硬件配置灵活、架构可调整。但是,FPGA的 使用有一定门槛,要求使用者具备硬件知识。	赛灵思、微软	L4
	TPU /ASIC	当前为谷歌公司专用,还不是市场化产品。ASIC芯片不能像FPGA很快改变架构,适应变化,对企业而言成本较昂贵。	谷歌	L5
发展方向二：颠覆经典冯氏架构，采用人脑神经元的结构来提升计算能力	TrueNorth	模仿人脑神经元和神经突触的结构,功耗非常低。有可能实现人工智能领域的通用化路径,但从短期来看,离大规模商业生产还有很远的距离。	IBM	

图：中国数据中心AI芯片市场规模占比

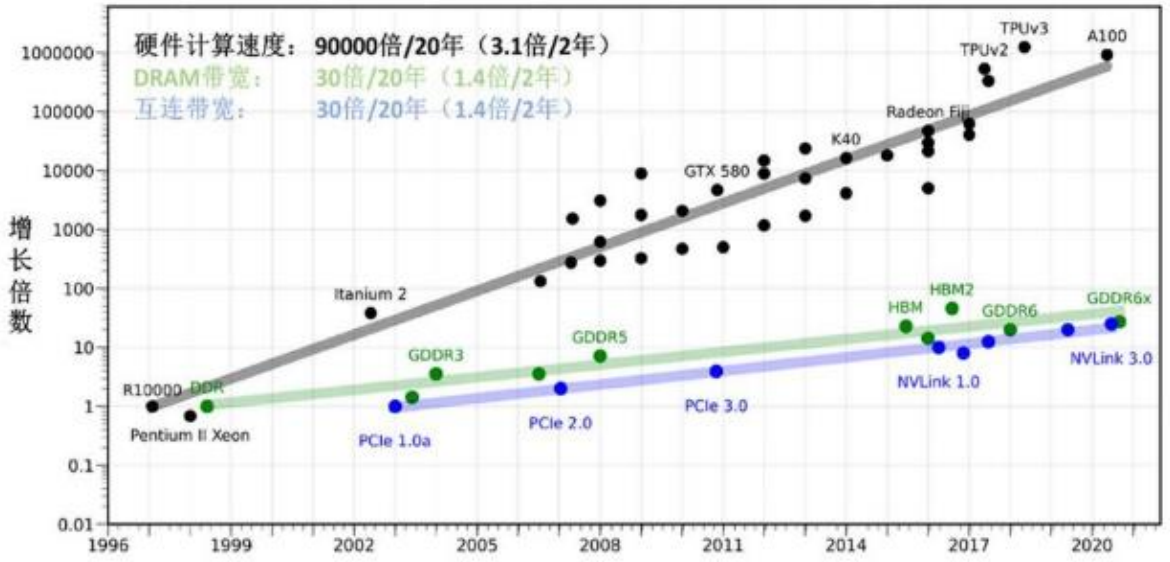


资料来源：IDC，松鼠厂，中航证券研究所

算力升级：冯氏架构“破壁者”，存算一体突破瓶颈

- 冯氏架构以计算为中心，计算和存储分离，二者配合完成数据的存取与运算。然而，由于处理器的设计以提升计算速度为主，存储则更注重容量提升和成本优化，“存”“算”之间性能失配，从而导致了访存带宽低、时延长、功耗高等问题，即通常所说的“存储墙”和“功耗墙”。
- **存算一体作为一种新的计算架构，被认为是具有潜力的革命性技术。**核心是将存储与计算完全融合，有效克服冯·诺依曼架构瓶颈，并结合后摩尔时代先进封装、新型存储器件等技术，减少数据的无效搬移，从而提升计算效率。中国移动已将存算一体纳入算力网络的十大关键技术。

图：存储计算性能存在“剪刀差”



表：存算一体化应用场景广泛

场景	重点需求	存算一体优势
端侧	低延时、低功耗、低成本、隐私性	当前存内计算产品已成功在端侧初步商用，提供语音、视频等AI处理能力，并获得十倍以上的能效提升，有效降低了端侧成本。
边侧	低延时、低功耗、低成本、通用性	存算一体在深度学习等领域有独特优势，可以提供比传统设备高几十倍的算效比，此外存内计算芯片通过架构创新可以提供综合性能全面兼顾的芯片及板卡，预计将在边侧推理场景中有着广泛的应用，为广泛的边缘AI业务提供服务。
云侧	大算力、高宽带、低功耗	存内计算可通过多核协同集成大算力芯片，结合可重构设计打造通用计算架构，存内计算作为智算中心下一代关键AI芯片技术，正面向大算力、通用性、高计算精度等方面持续演进，有望为智算中心提供绿色节能的大规模AI算力。

资料来源：中国移动研究院，中航证券研究所

算力升级：冯氏架构“破壁者”，存算一体突破瓶颈

- 当前NOR Flash、SRAM等传统器件相对成熟可率先开展存内计算产品化落地推动。新型器件中RRAM各指标综合表现较好，MRAM寿命和读写性能较好，均有各自独特优势与发展潜力可持续推动器件成熟，同步进行存内计算探索。
- 三星电子、SK海力士、台积电、美光、IBM、英特尔等都在进行存算一体技术的研究。国内公司中，亿铸科技、千芯科技、后摩智能专注于大算力存算一体芯片，闪易半导体、苹芯科技、知存科技、智芯科、九天睿芯专注于小算力存算一体芯片。**上市公司中，推荐关注打造存算生态的头部公司兆易创新，研发布局NOR Flash的恒烁股份，以及拥有存算一体研发项目的东芯股份。**

图：存内计算器件对比分析

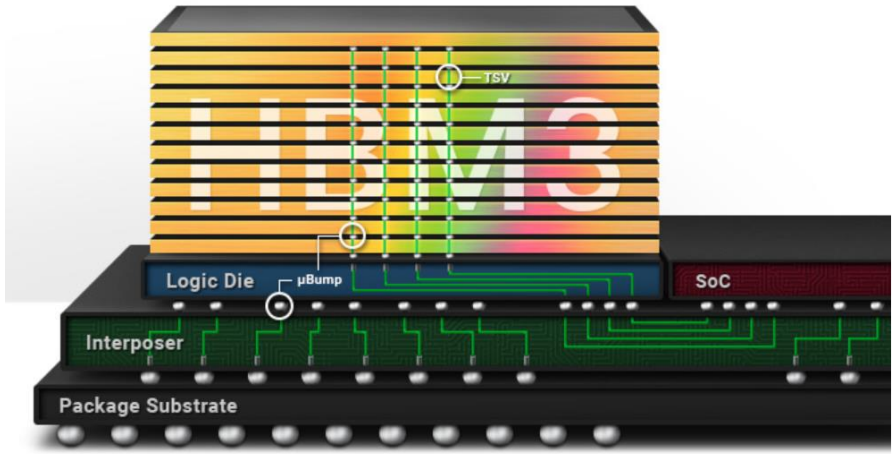
器件	SRAM	NOR FLASH	RRAM	MRAM	PCM
易失特性	易失	非易失	非易失	非易失	非易失
多值存储	否	是	是	否	是
现有工艺节点	5nm	28nm	28nm	16nm	28nm
理论工艺极限	2nm	14nm	5nm	5nm	5nm
单比特存储面积 (F^2/bit)	~300	~7.5	20~40	~30	~24
读写次数	无限	10^6	10^8	~10^15	10^8
应用场景	云侧和边侧的推理和训练	边侧和端侧的推理	云侧、边侧和端侧的推理	云侧和边测的推理和训练	云侧、边侧和端侧的推理

资料来源：中国移动研究院，中航证券研究所

存力升级：HBM提升存储带宽

- 以ChatGPT为代表的生成类模型需要在海量数据中训练，对存储容量和带宽提出新要求，HBM（High Bandwidth Memory，高带宽存储器）成为减小内存墙的优选项。HBM将多个DDR芯片堆叠并与GPU封装在一起，是一种基于3D堆叠工艺的高附加值DRAM产品。通过增加带宽，扩展内存容量，让更大模型、更多参数留在离计算核心区更近的地方，从而减少内存和存储解决方案带来的延迟。据Omdia预测，到2025年，HBM市场的总收入将达到25亿美元。
- 由于ChatGPT的爆火，GPU需求明显，英伟达也加大对三星和SK海力士HBM3的订单。**建议关注有HBM技术布局的A股相关标的，如：深科技、雅克科技、国芯科技、通富微电。**

图：HBM3产品结构



图：海力士HBM产品性能演进

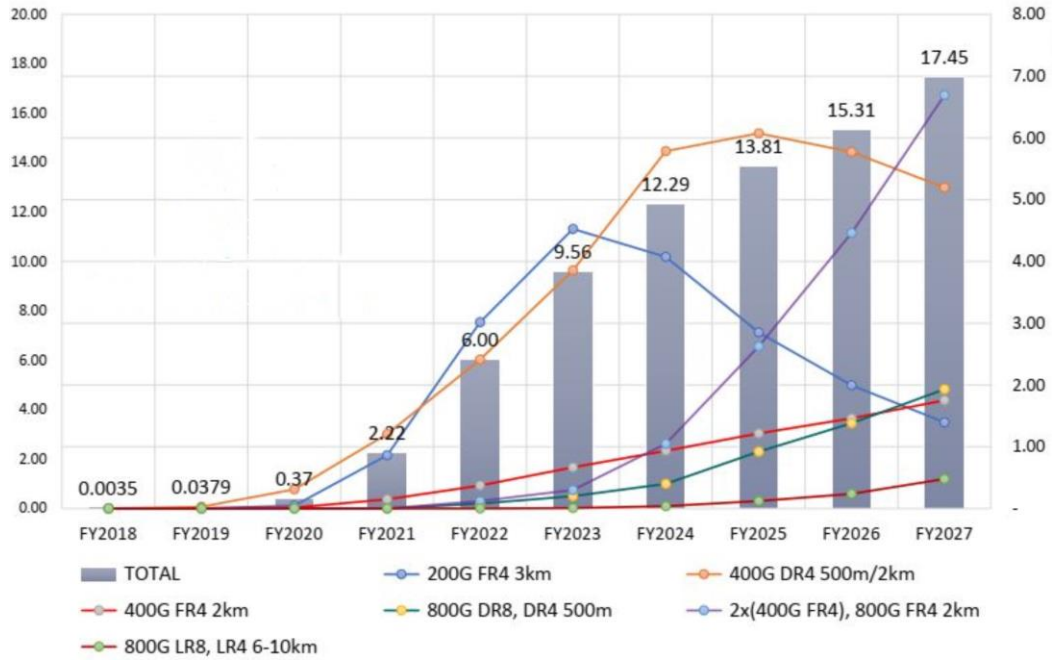
	HBM1	HBM2	HBM2E	HBM3
发布年份	2014年	2018年	2020年	2022年
芯片密度	2Gb	8Gb	16Gb	16Gb
堆叠高度	4层	4层/8层	4层/8层	8层/12层
容量	1GB	4GB/8GB	8GB/16GB	16GB/24GB
带宽	128GB/s	307GB/s	460GB/s	819GB/s
I/O速率	1Gbps	2.4Gbps	3.6Gbps	6.4Gbps

资料来源：SK海力士，中航证券研究所

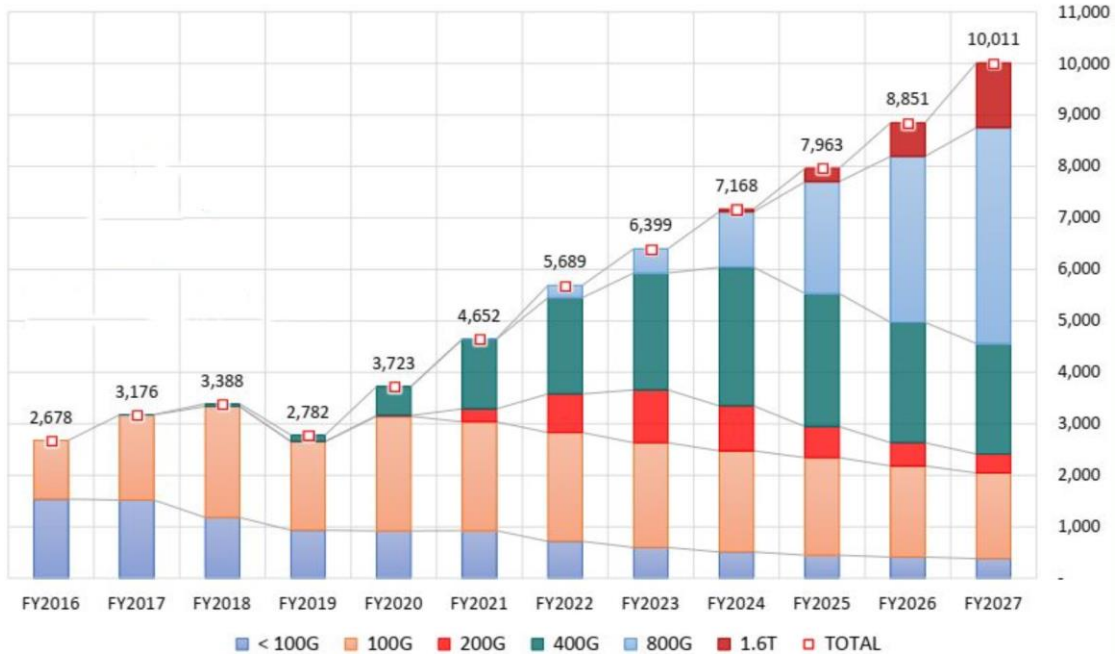
传输升级：高速光模块放量

- **传输速度迭代不止，高速光模块出货预计大幅增长。**据lightCounting统计，2021年，200G、400G和800G的高速以太网光模块发货量达222万只，2022年预计将达600万只，同比170%以上，800G的产品有望在2022年开始逐步放量。
- 据lightcounting2022年3月预测，未来随着AI、元宇宙等新技术不断发展，以及网络流量长期保持持续增长，以太网光模块销售额也将保持较快增长并不断迭代升级。预计到2027年，以太网光模块市场将达到100.11亿美元。

图：高速光模块发货量预测（百万只）



图：以太网光模块营收预测（百万美元）

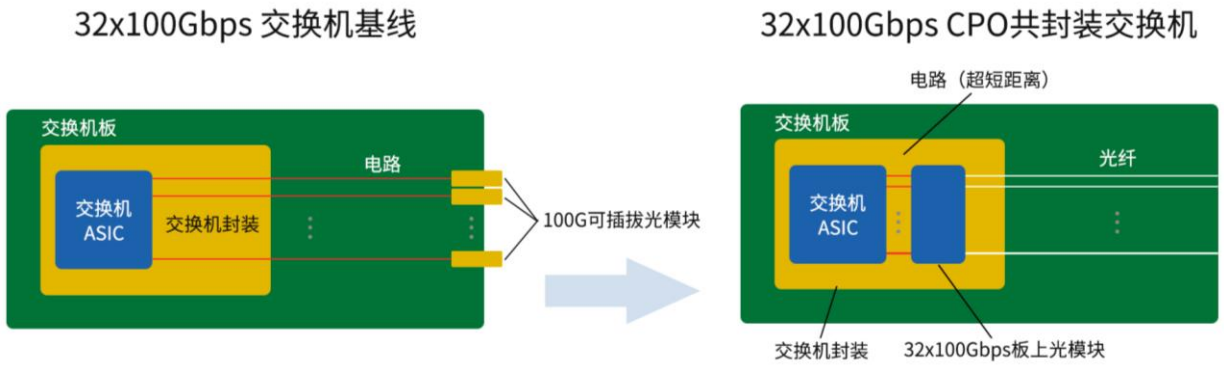


资料来源：光通信之家，lightcounting，中航证券研究所

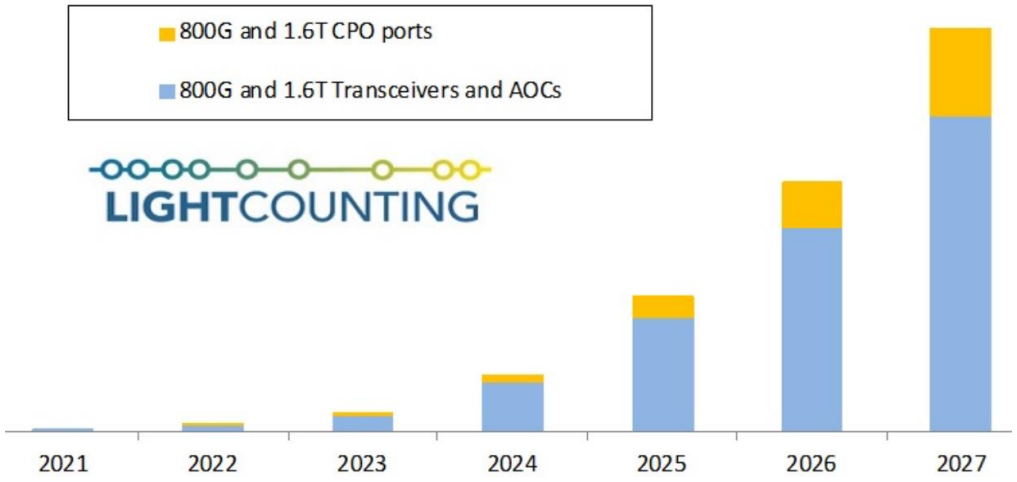
传输升级：CPO与硅光技术降本增效

- **CPO（协同封装光子技术）提升数据中心应用中的光互连技术。**CPO将光学器件和ASIC紧密结合在一起，通过 Co-packaging 的封装方式，大体积的可插拔模块被简单的光纤配线架所取代，因此前面板的物理拥塞得以缓解。而交换机和光学器件之间的电气通道大大缩短，因此CPO将增加带宽和缩小收发器尺寸，提升系统集成度，同时降低功耗和封装成本。
- 据lightcounting预测，数据中心将率先使用CPO封装技术。同时，随着AI集群和HPC的架构正在不断演进发展，可能会看到CPO部署在GPU、TPU以及以太网、InfiniBand或NVLink交换机上，另外有许多基于FPGA的加速器也可能受益于CPO。预测在2027年，CPO端口将占总800G和1.6T端口的近30%。据机构CIR预测，CPO市场规模将在2025年超过13亿美元，2027年达到27亿美元。**建议关注中际旭创、光迅科技、华工科技、天孚通信、德科立、源杰科技等光模块产业相关标的。**

图：CPO交换机



图：CPO端口、可插拔以太网光模块和AOC出货占比预测

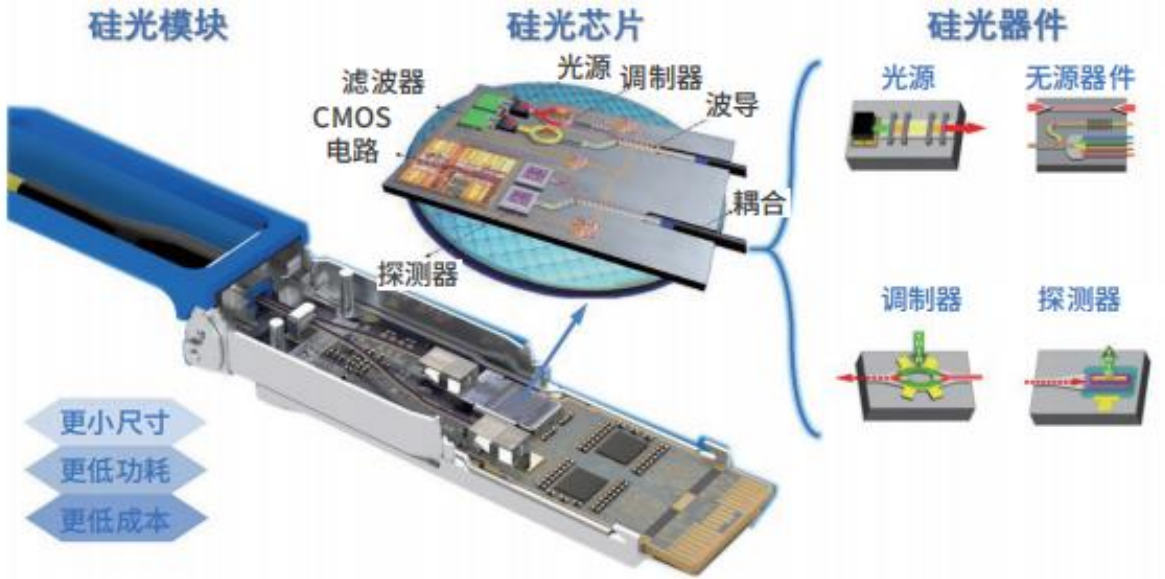


资料来源：易飞通信，lightcounting，CSDN，中航证券研究所

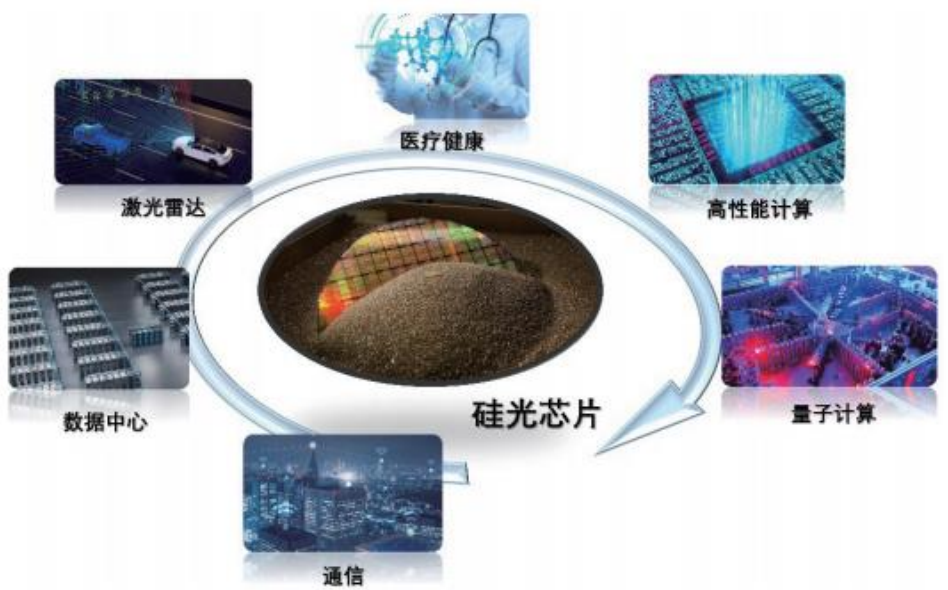
传输升级：CPO与硅光技术降本增效

- 硅光芯片基于绝缘衬底上硅（Silicon-On-Insulator，SOI）平台， 兼容互补金属氧化物半导体 (Complementary Metal OxideSemiconductor，CMOS） 微电子制备工艺，同时具备了 CMOS 技术超大规模逻辑、超高精度制造的特性和光子技术超高速率、超低功耗的优势。硅光芯片商业化至今较为成熟的领域为数据中心、通信基础设施等光连接领域。目前，硅光技术在第一代4x25G光模块中主要应用于500m内的100G QSFP28 PSM4；在第二代1x100G产品中，应用有100G QSFP28 DR1/FR1和LR1，作用于500m-10km场景中；在400G产品中，主要聚焦在2km以内的中短距离传输应用场景，产品有400G DR4。未来随着技术逐渐成熟，激光雷达、光子计算等领域的应用有望实现突破。
- **建议关注光库科技、声光电科、赛微电子等硅光制造产业链相关标的。**

硅光芯片在光模块中的应用



硅光芯片的未来主要应用场景展望



资料来源：华中科技大学武汉光电国家研究中心，易飞扬通信，中航证券研究所

目 录

AI大模型催化新一轮半导体制裁

- 围绕大模型的研发和落地，中美之间已经展开了新一轮的竞争。半导体作为AI算力核心，美国已在2022年9月限制中国采购最先进的AI训练芯片。**我们认为半导体将受到顶层高度关注，成为大国博弈的焦点之一。**

图：制裁加剧，顶层高度重视自主可控

2022年9月1日，美国芯片巨头英伟达收到美国官方通知，若对中国和俄罗斯的客户出口两款高端GPU芯片——A100和H100，需要新的出口许可。

2023年3月2日，国务院副总理刘鹤调研北京集成电路企业发展并主持召开相关座谈会。会上提及发展集成电路产业必须发挥新型举国体制优势。

2023年3月31日，日本政府宣布拟对23种半导体制造设备实施出口管制，并就有关措施征求公众意见。

资料来源：新华网、智能网联汽车网、新京报、中航证券研究所

晶圆制造现状：资本开支回落，大国竞争鼓励本土建厂

- **Capex回落符合预期规律，国内代工龙头逆势上修。**终端需求疲软，使得以存储为代表的厂商率先大幅削减资本开支，其中美光FY23预计下调3成，SK海力士预计下调5成。根据IC Insights，2023年全球半导体资本开支1466亿美元，同比下滑19%，但仍处于历史第三高位。大陆代工龙头中芯国际大幅上调资本开支并扩建天津西青工厂，“举国体制”下，国内IC制造的景气度无需过度忧虑。
- **大国竞争愈演愈烈，“竞赛式”补贴层出不穷。**半导体产业发展历经多次重心转移，国家变迁，如今日本、欧洲半导体产业逐渐式微，各国危机意识强烈。中美欧日韩纷纷出台补贴政策刺激，重点补贴IC制造。美国《芯片与科学法案》中补贴390亿美元投入IC制造，美光、Intel、TI纷纷宣布扩产。我们判断，随着周期回暖及各国补贴政策的逐步实施，对晶圆厂投资会有所刺激和拉动。

表：全球部分大厂资本支出调整计划（除三星外，均为亿美元）

	2020年	2021年	2022E	2022/2021增长率	2021/2020增长率	最新调整措施
台积电	172.4	300.4	360	20%	74%	减少40亿美元
联电	9.5	17.6	30	71%	84%	减少6亿美元
中芯国际	/	45.0	66	47%	/	增加16亿美元
格芯	5.9	17.7	30-33	70%-87%	198%	下修12-15亿美元
英特尔	142.6	187.3	250	33%	31%	减少20亿美元
德州仪器	6.5	24.6	35	42%	279%	不变
三星电子	32.9万亿韩元	43.6万亿韩元	47.4万亿韩元	8.7%	33%	不变

表：全球半导体产业刺激政策

国家/地区	出台时间	半导体产业振兴措施
中国	2020年8月-至今	《新时期促进集成电路产业和软件产业高质量发展的若干政策》,出台产业专项引导政策；此后陆续出台税收优惠政策、“十四五”战略规划等。计划2025年国产芯片自给率达70%。
美国	2022年8月	《芯片与科学法案》，拨款527亿美元扶持半导体产业，其中390亿美元投入半导体制造。
欧盟	2022年2月	《欧洲芯片法案》，投入430亿欧元，提振欧洲芯片产业，计划2030年将欧洲芯片产能从不足10%提升到20%以上。
日本	2021年底	批准7740亿日元（68亿美元）的半导体投资预算，54亿美元用于支持IC生产，包括支持台积电熊本工厂。
韩国	2021年5月	实施“K半导体战略”，携手三星电子、SK海力士到2030年投资超过510万亿韩元；2023年1万亿韩元投资半导体产业。

资料来源：各国政府官网，IC Insights，各公司公告，中航证券研究所

全球晶圆扩产脚步不止，内资产能有望提升

- **国内晶圆产能将以远超全球增速的态势增长。**根据我们对全球63家主流IDM/Foundry企业的产能统计，当前全球晶圆月产能2125万片（折合8英寸），未来三年以7%左右的增速持续增长。且扩产以12英寸为主，预计2024年全球12英寸达到808万片/月。值得注意的是，国内12英寸产能将达到155万片/月，保持30%以上的CAGR，中国大陆内资总产能有望从当前的15%增长至2024年的24%。
- **目前国内主要在建项目以12英寸28nm及以上的成熟制程为主。**28nm是成熟的性价比的工艺节点，可以用在中低端手机、平板等绝大多数电子设备，且能覆盖增速最快的汽车电子。SMIC一边突破先进制程，一边不断巩固自己在28nm的地位。

表：全球及中国大陆晶圆产能概览

晶圆产能（万片/月）				
	2021年	2022E	2023E	2024E
全球总产能：等效8英寸	2125	2295	2459	2613
全球产能增速		8.0%	7.2%	6.3%
国内厂商产能：等效8英寸	326	427	532	625
国内产能增速		31.2%	24.6%	17.4%
国内厂商产能占比	15.3%	18.6%	21.6%	23.9%
国内厂商产能：分尺寸统计				
8英寸	92	111	121	126
8英寸产能增速		20.4%	9.2%	4.0%
12英寸：非等效	64	94	124	155
12英寸产能增速		47.1%	32.6%	24.8%

表：国内代工厂部分主要在建项目

公司	尺寸	产线	产线地址	规划产能（万片/月）	预计建成时间	制程
中芯国际	12英寸	上海临港基地	上海	10	2023年	28nm及以上
	12英寸	中芯京城（1期）	北京	10	2022年底	28nm及以上
	12英寸	中芯深圳	深圳	4	2022年底	28nm及以上
	12英寸	中芯天津西青	天津	10	2024年	28-180mm
华虹半导体	12英寸	华虹七厂一期扩产	无锡	新增3	2022Q4	90-65/55nm
长江存储	12英寸	国家存储器基地2期	武汉	20	2022年	/
长鑫存储	12英寸	长鑫二期	合肥	12	爬坡中	17nm
上海积塔半导体	8英寸	特色工艺生产线Fab1	上海	6	爬坡中	0.11/0.13/0.18μm
	12英寸	特色工艺生产线Fab2	上海	0.3	爬坡中	55/65nm
晶合集成	12英寸	晶合集成N2厂	合肥	4	2022年	55nm

资料来源：各公司官网，产业链调研，Omdia2020报告，中航证券研究所整理（注：完整数据统计表，可以联系团队/对口销售）

半导体设备：大国重器，玉汝于成

- **半导体设备是晶圆制造的投资核心。**设备投资占IC制造资本开支的70%-80%，且以前道晶圆制造设备为主，占设备总投资的85%以上。
- **上游基石环节，撬动千亿美元市场。**根据SEMI数据，2022-2023年全球半导体设备市场规模将达到1175、1208亿美元，同比增长15%、3%。

硅片制造

多晶硅
- 多晶硅还原炉

拉晶
- 全自动单晶炉（直拉法/区熔法）

截断+边缘研磨
- 半导体单晶硅截断机
- 滚磨机

切割
- 半导体金刚线切片机

倒角+研磨
- 倒角机
- 激光刻字机
- 晶圆减薄机

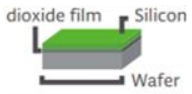
清洗+刻蚀
- 腐蚀机
- 清洗机

CMP抛光
- 双面抛光机
- CMP抛光机
- 边缘抛光机

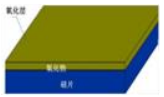
检测
- 表面检测仪
- 几何参数检测仪

前道晶圆制造工艺

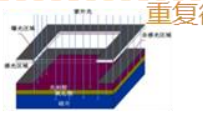
重复循环



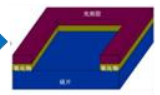
氧化/扩散
- RTP设备
- 激光退火炉
- 扩散炉
- 氧化炉



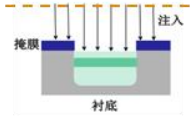
薄膜沉积
- CVD设备
- PVD设备
- ALD设备
- 气相外延炉



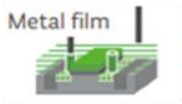
光刻
- 涂胶/显影设备
- 光刻机



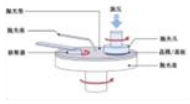
刻蚀
- 刻蚀机
- 去胶设备



离子注入
- 离子注入机



金属化
- PVD设备
- 电镀设备



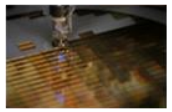
抛光
- CMP设备

晶圆检测
- 量测设备
- 探针台

清洗设备贯穿其间

后封装测试工艺

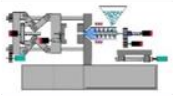
背面减薄
- 检测设备
- 贴膜机
- 减薄机



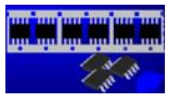
晶圆切割
- 晶圆安装机
- 划片机
- AOI

贴片
- 贴片机
- 烤箱

引线键合
- 引线键合机
- AOI



模塑
- 注塑机
- 激光打标机
- 烤炉
- X-ray



切筋/成型
- 切筋/成型设备
- AOI

成品测试
- 测试机
- 分选机
- 探针台

清洗设备贯穿其间

资料来源：SUMCO，TEL，华海清科，华峰测控，中航证券研究所整理

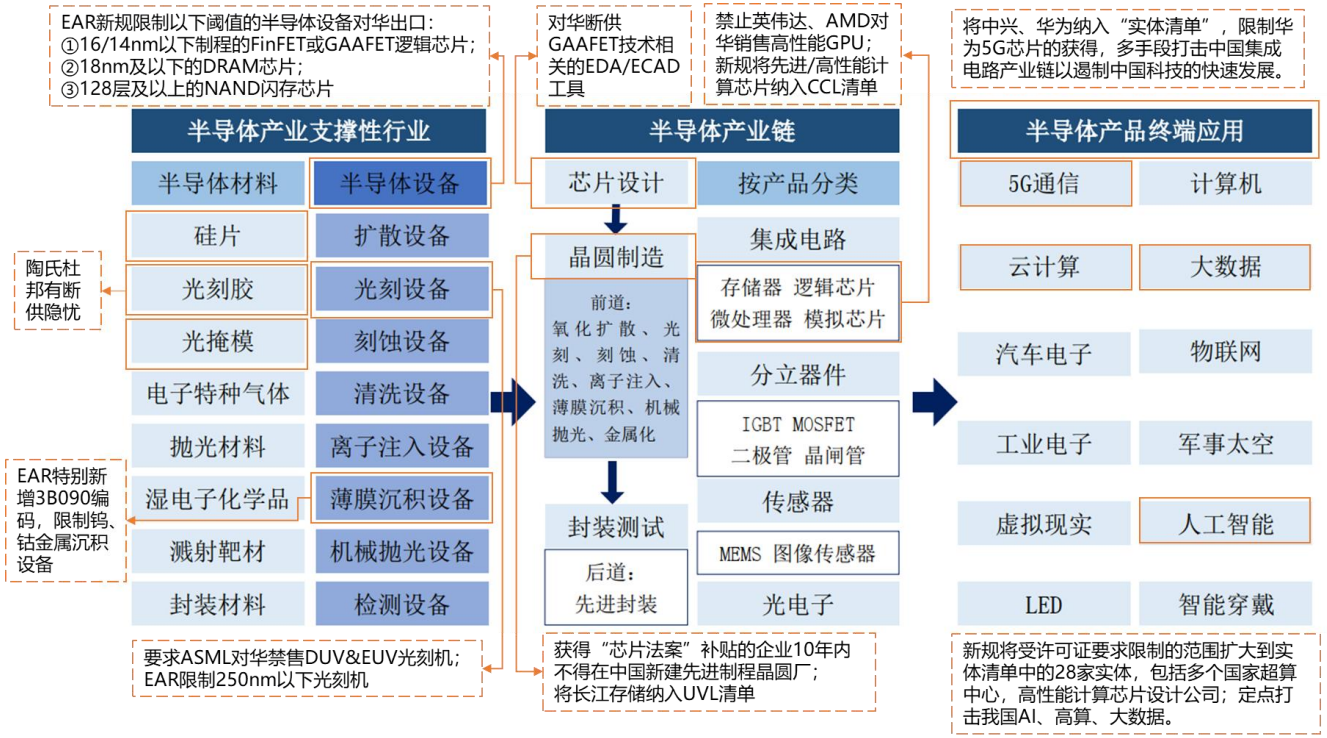
中国是全球最大的设备市场，制裁强化替代逻辑

- **美对华上下游封锁形势已成，国产替代势在必行。**美国10月7日《出口管制条例》（EAR）对半导体设备出口设定了明确的阈值：①16/14nm以下制程的FinFET或GAAFET逻辑芯片；②18nm及以下的DRAM芯片；③128层及以上的NAND闪存芯片。我们判断，制裁短期内会给晶圆厂扩产带来阵痛，可能会延缓长存、长鑫等高端存储芯片厂的扩产节奏，对逻辑芯片制造影响有限。
- 从海外大厂披露的情况来看，预计EAR 2023年将影响三家美系设备厂（AMAT+LAM+KLA）51~59亿美元的收入，且考虑到海外较大的订单积压，部分收入/订单有望转移至国内设备大厂。

表：海外龙头对EAR影响的判断及订单积压情况

公司	2022 Q3收入	Q3中国大陆收入占比	EAR对Q4的影响	EAR对2023年的影响	EAR影响的相关表述
ASML	57.8亿欧元	15%	未披露	5%积压订单（5%*380亿欧元）	ASML为欧洲公司，只有有限的美国技术，EAR影响有限，考虑供应链，预计将间接影响5%的积压订单。
AMAT	67.5亿美元	20%	4.9亿美元	25亿美元	公司预计FY23Q1的影响约4.9亿美元，全年或影响25亿美元，对公司Non-GAAP毛利率影响约1pct。
LAM Reasearch	50.7亿美元	30%	未披露	20-25亿美元	预计CY2023年出口限制对总收入的影响在20亿-25亿美元。
KLA	27.2亿美元	31%	1亿美元	6-9亿美元	设备多为定制化，若中国fab厂缺乏服务或备件，即使获得设备也很难正常运行。EAR预计影响KLA 2023年6-9亿美元收入。

图：中国集成电路产业链及各环节所受制裁情况梳理



资料来源：各公司法说会，BIS《出口管制条例》，盛美上海招股书，中航证券研究所（其中AMAT Q4 指的是2022年11月-2023年1月）

锻造内功，国产设备成长可期

- **“国家安全”的定调坚定芯片自主可控决心，关键在于国内厂商持续锻造内功。**当前设备厂商在某些细分环节已经具备完全替代的能力，且“去A化”倒逼晶圆厂验证、采购国产设备，并促进设备厂加快技术攻坚。从国内几家半导体设备厂商的现有能力来看：除光刻机以外，其他主流环节28nm及以上工艺基本已经能实现国产替代，部分厂商正在向14nm及以下拓展。当前时点国内晶圆厂扩产仍以28nm及以上成熟制程为主，也为国产化设备的制程突破创造了一定的时间窗。
- 我们测算，2022H1中国半导体设备市场规模141.3亿美元，但国产化率仅15%左右，美国三巨头2022H1在中国的合计收入约80亿美元，假定国产替代能实现50%/80%的去美国化，则国产设备厂还有2.1/3.3倍的成长空间。

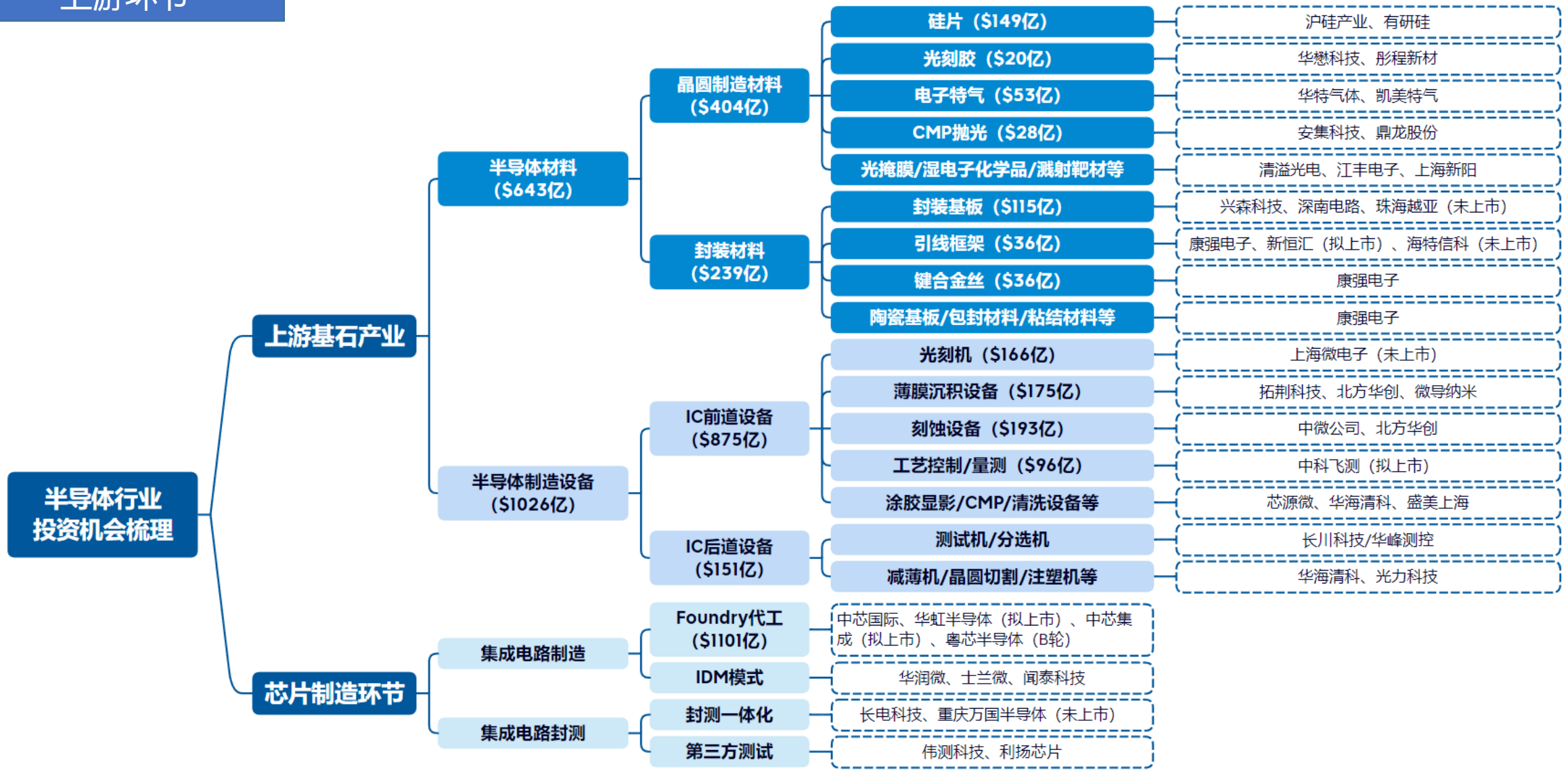
表：我国主要前道设备上市公司产品制程能力

公司名称	主要设备	制程能力
北方华创	平台型公司：刻蚀、薄膜沉积（PVD为主）、清洗设备等	刻蚀机和薄膜沉积设备突破14nm，产业化应用
中微公司	刻蚀机/MOCVD	CCP刻蚀机突破14nm及以下已实现产业化应用，进入5nm及以下晶圆生产线
华海清科	CMP设备	28nm已实现所有工艺全覆盖，14nm几个关键工艺CMP设备处于验证中
拓荆科技	PECVD/ALD/SACVD	主力PECVD产品应用于28nm及以上逻辑芯片，28nm以下产业化验证中；部分产品可以用于14-28nm逻辑芯片
芯源微	涂胶显影设备/清洗设备	涂胶显影设备28nm及以上产业化应用，并继续关键技术的突破；前道清洗机28nm产业化应用
盛美上海	清洗设备（向平台化转型）	SAPS兆声波清洗技术已实现28nm产业化应用，14nm及以下正在开发

资料来源：各公司公告，中航证券研究所

附表：半导体行业投资机会梳理

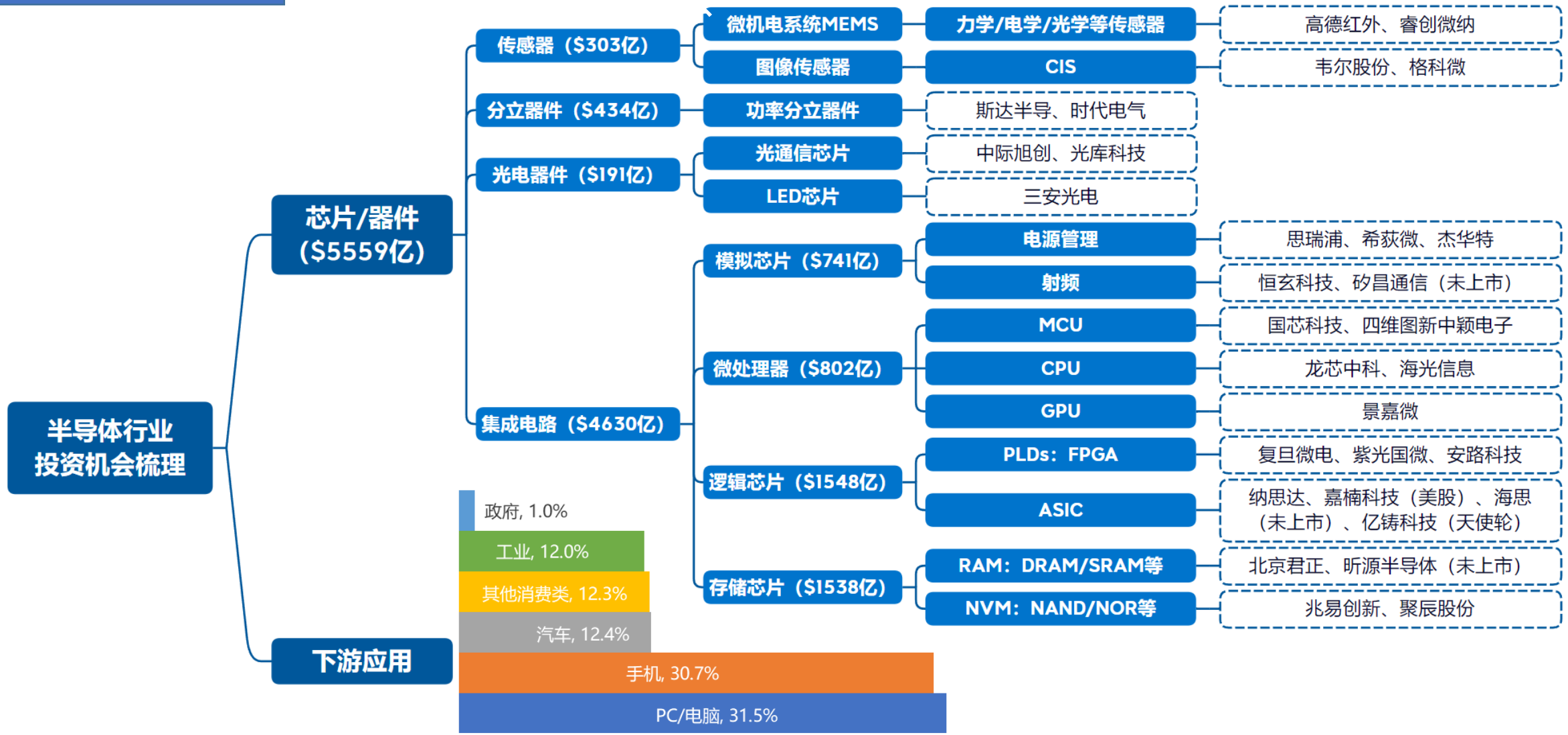
上游环节



资料来源：SEMI、Gartner、IC Insights、中航证券研究所（市场规模取2021年数据）

附表：半导体行业投资机会梳理（接上页）

中下游环节



资料来源：WTST、SIA、中航证券研究所（市场规模取2021年数据）

目 录

风险提示

- AI算法、模型存较高不确定性，AI技术发展不及预期
- ChatGPT用户付费意愿弱，客户需求不及预期
- 针对AI的监管政策收紧

我们设定的上市公司投资评级如下：

买入 ：未来六个月的投资收益相对沪深300指数涨幅10%以上。
持有 ：未来六个月的投资收益相对沪深300指数涨幅-10%-10%之间
卖出 ：未来六个月的投资收益相对沪深300指数跌幅10%以上。

我们设定的行业投资评级如下：

增持 ：未来六个月行业增长水平高于同期沪深300指数。
中性 ：未来六个月行业增长水平与同期沪深300指数相若。
减持 ：未来六个月行业增长水平低于同期沪深300指数。

中航科技电子团队介绍：

首席：赵晓琨 SAC执业证书：S0640122030028
十六年消费电子及通讯行业工作经验，曾在华为、阿里巴巴、摩托罗拉、富士康等多家国际级头部品牌终端企业，负责过研发、工程、供应链采购等多岗位工作。曾任职华为终端半导体芯片采购总监，阿里巴巴人工智能实验室供应链采购总监。

分析师：刘牧野 SAC执业证书：S0640522040001
约翰霍普金斯大学机械系硕士，2022年1月加入中航证券。拥有高端制造、硬科技领域的投研经验，从事科技、电子行业研究。

研究助理 刘一楠 SAC执业证书：S0640122080006
西南财经大学金融硕士，2022年7月加入中航证券，覆盖半导体设备、半导体材料板块。

研究助理 苏弘宇 SAC执业证书：S0640122040021
俄亥俄州立大学金融数学学士，约翰霍普金斯大学金融学硕士。 2022年加入中航证券。

分析师承诺

负责本研究报告全部或部分内容的每一位证券分析师，再次申明，本报告清晰、准确地反映了分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与，未来也将不会与本报告中的具体推荐或观点直接或间接相关。

风险提示：投资者自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券证券投资损失的书面或口头承诺均为无效。

免责声明