

电子行业专题报告

边缘 AI 浪潮已至，AI SoC 赋能终端奋楫争流

方正证券研究所证券研究报告

分析师

郑震湘 登记编号：S1220523080004

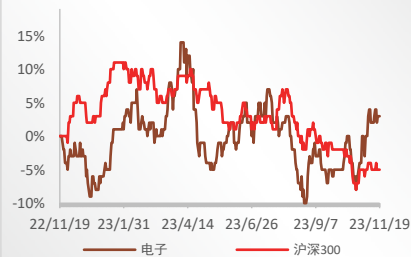
余凌星 登记编号：S1220523070005

行业评级：推荐

公司信息

上市公司总家数	495
总股本(亿股)	4,975.55
销售收入(亿元)	45,290.82
利润总额(亿元)	3,466.42
行业平均 PE	70.27
平均股价(元)	34.36

行业相对指数表现



数据来源：wind 方正证券研究所

相关研究

《先进封装专题三：代工、IDM 厂商先进封装布局各显神通》2023.11.18

《vivo 首款端侧 AI 手机强势发布，拥抱 AI 手机新机遇》2023.11.14

《先进封装专题二：HBM 需求井喷，国产供应链新机遇》2023.11.14

《华为智能汽车专题四：阿维塔 12 大超预期，智能驾驶全面升级》2023.11.13

科技巨头相继发力，边缘 AI 方兴未艾。随着终端应用场景的扩展带来数据即时处理需求的不断增长，云计算在数据传输、即时处理、隐私能耗等方面的不足开始显现，因此更加侧重局部和小规模处理的边缘计算应运而生。我们看到，科技巨头对于边缘计算的关注度正在日益提升：

1) 高通转型“智能边缘计算”，押注 C 端边缘 AI。在 2023.5.30 召开的 COMPUTEX 2023 上，高通宣布自身正在从一家通信公司转型为一家“智能边缘计算”公司。此后在 2023.7.18 宣布和 Meta 达成合作，计划从 2024 年起，在搭载骁龙平台的智能手机等终端上支持基于 Llama 2 的 AI 部署，赋能开发者使用高通 AI 软件栈面向终端侧 AI 进行应用优化，从而推出全新的生成式 AI 应用。我们认为，高通在智能手机 SoC 领域积淀深厚，有利于推动终端 AI 芯片继续向其他 C 端设备延伸。

2) 英伟达 Jetson 平台不断迭代，发力 B 端边缘 AI。英伟达在 2014 年推出第一代 Jetson 平台 TK1，截至目前已迭代至第七代 AGX Orin，该平台使用英伟达自主研发的 GPU 加速技术，支持深度学习框架，能够加速模型的训练和推理，同时还支持各种图像、视频、通信和存储接口，适用于各种嵌入式应用和边缘计算需求，如机器人、自动驾驶等。我们认为，英伟达作为 GPU 传统云端 AI 芯片霸主，更加合适布局 B 端大算力场景。

AI SoC 算力承载，赋能终端百花齐放。我们看到，相比云端 AI 芯片需要兼具训练和推理性能，边缘 AI 芯片则更多承担推理任务，其主要存在形式是作为 AI SoC 嵌入终端设备，因此性能上更加追求在算力、功耗和成本等多方面的综合表现。伴随人工智能在消费电子、智能家居、智能安防、智能驾驶等 IoT 领域的快速渗透，全球边缘 AI 芯片市场规模有望从 2022 年的不足 400 亿美元突破至 2025 年的 600 亿美元。

1) 消费电子：手机市场主要由高通和联发科主导，但在可穿戴设备（TWS 耳机、智能手表、VR/AR 等）SoC 市场，我们认为受益于较多长尾客户的存在，以及更加追求性能和功耗的平衡，国内以恒玄科技、中科蓝讯为代表的厂商仍存在一定的发展机遇。

2) 智能家居：在众多细分品类中，我们判断具备语音/视频交互能力的智能音箱有望成为智能家居全屋互联场景下的 AI 控制入口。百度、小米、天猫精灵等国内智能音箱巨头历经多年的布局，各自的产品生态圈建设已趋于完善，叠加大模型落地驱动产品交互式体验持续升级，我们认为以全志科技、晶晨股份为代表的上游处理器厂商有望步入新的发展阶段。

3) 智能驾驶：自动驾驶方面，ADAS SoC（适用于 L1-L2）和的 AD SoC（适用于 L3-L5）的总体市场规模有望在 2030 年达到 300 亿美元，海外龙头 Mobileye 主打视觉方案以稳固 ADAS SoC 龙头地位，英伟达则聚焦高端 AD SoC 追求极致性能；智能座舱方面，高通领跑市场的同时，我们也看到国内消费级芯片厂商（华为、全志科技、瑞芯微等）、汽车 AI 芯片厂商（地平线、黑芝麻等）、汽车芯片初创企业（杰发科技、芯擎科技、芯驰科技等）三大阵营正在加速追赶，我们持续看好国产供应商的未来发展。

风险提示：研发进展不及预期；产业进度不及预期；中美贸易摩擦加剧。

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

正文目录

1 智能物联新时代，边缘 AI 蓬勃发展	5
1.1 算力需求日益增长，边缘计算应运而生	5
1.2 科技巨头相继发力，边缘 AI 方兴未艾	6
1.3 突破算力及功耗瓶颈，加速边缘 AI 落地	12
2 智能终端 SoC，端侧 AI 算力承载	15
2.1 AI SoC 多模块异构集成，重视性能与功耗平衡	15
2.2 边缘 AI 芯片规模稳健成长，海内外厂商加速布局	16
2.3 自研 IP 积累及先进代工保障，彰显核心竞争力	18
3 AI SoC 赋能下游，终端应用百花齐放	19
3.1 消费电子	20
3.2 智能家居	22
3.3 智能安防	24
3.4 自动驾驶	26
3.5 智能座舱	31
4 相关标的	35
4.1 芯原股份：IP 授权业务短期承压，在手订单保障未来成长	35
4.2 韦尔股份：库存去化大超预期，新品放量加速业绩改善	35
4.3 乐鑫科技：业绩稳健增长，软硬件双轮驱动	36
4.4 恒玄科技：季度营收创历史新高，可穿戴需求复苏启动	37
4.5 中科蓝讯：业绩端表现亮眼，客户端持续渗透	37
4.6 炬芯科技：打造 AIoT 音频芯生态，端侧 AI 赋能长期发展	38
4.7 晶晨股份：海外市场稳步推进，产品矩阵不断拓宽	38
4.8 全志科技：多元化产品布局，拥抱 AI 创新浪潮	39
4.9 泰凌微：无线物联网 SoC 领军者，纵深产品线把握行业红利	39
4.10 瑞芯微：行业回暖带动业绩回升，智能座舱芯片放量在即	40
5 风险提示	41

图表目录

图表 1: 云-边-端视角下的边缘计算	5
图表 2: AI 处理重心向边缘侧转移	6
图表 3: 高通边缘 AI 技术路线图	6
图表 4: 终端侧 AI 的四大关键优势	7
图表 5: Llama 2 模型特点	7
图表 6: Llama-2-chat 模型的训练过程	8
图表 7: Llama 2 (70 亿参数版本) 与其他闭源大语言模型的性能对比	8
图表 8: 英伟达 Jetson 平台产品迭代路线图	9
图表 9: 英伟达 Jetson 平台支持的边缘计算应用场景	9
图表 10: Jetson AGX Orin 工业级模块强化环境性能	10
图表 11: 边缘计算 VS. 云计算	11
图表 12: 云边协同	11
图表 13: 混合式 AI	11
图表 14: 大模型小型化	13
图表 15: 算力 VS. 存储性能提升	13
图表 16: 存算一体芯片架构	14
图表 17: 存算一体存储介质对比	14
图表 18: AI 芯片分类	15
图表 19: 华为麒麟 990 5G SoC 芯片架构	16
图表 20: AI 算法的运行原理	16
图表 21: 全球及中国边缘 AI 芯片市场规模 (亿美元)	17
图表 22: 全球不同功能边缘 AI 芯片出货量 (十亿颗)	17
图表 23: AI SoC 厂商布局	17
图表 24: SoC 芯片制造流程	18
图表 25: 台积电 VS 三星代工芯片参数对比	18
图表 26: 高通 AI 引擎	19
图表 27: 高通传感器中枢	19
图表 28: AI SoC 终端应用	19
图表 29: 苹果 AirPods Pro 第二代搭载的 H2 芯片	20
图表 30: 高通骁龙全新可穿戴平台 W5/W5+	21
图表 31: 恒玄科技 VS. 中科蓝讯 TWS 蓝牙耳机芯片出货量	21
图表 32: 恒玄科技 VS. 中科蓝讯营收及毛利率	21
图表 33: 苹果 Vision Pro 产品效果	21
图表 34: 苹果 Vision Pro 搭载 M2 及 R1 芯片	21
图表 35: 中国智能家居品类 AI 技术应用渗透率预测	22
图表 36: 中国智能家居市场规模及增速	22
图表 37: 中国智能音箱市场竞争格局	22
图表 38: 中国智能音箱头部企业分析	22
图表 39: 中国智能音箱市场销量 (万台)	23
图表 40: 2020 年全球智能音箱处理器出货量份额	23
图表 41: 中国智能音箱处理器厂商客户布局	23
图表 42: 安防芯片分类	24
图表 43: 边缘 AI 在安防领域的应用	25
图表 44: 2021 年全球 IPC SoC 市场份额	25
图表 45: 2021 年全球 NVR SoC 市场份额	25

图表 46: 全球 IPC/NVR SoC 市场规模 (亿美元)	25
图表 47: 自动驾驶域控制器硬件架构.....	26
图表 48: 华为 MDC 计算平台硬件方案.....	27
图表 49: 全球及中国自动驾驶乘用车销量及渗透率.....	27
图表 50: 不同级别自动驾驶 SoC 市场规模预测.....	28
图表 51: Mobileye 自动驾驶 SoC 产品线	28
图表 52: 英伟达历代车载计算平台参数对比.....	29
图表 53: 高通 8155 座舱芯片架构	31
图表 54: 高通座舱芯片迭代历程	31
图表 55: 海外座舱芯片主要厂商产品参数对比.....	32
图表 56: 国内座舱芯片主要厂商产品参数对比.....	34

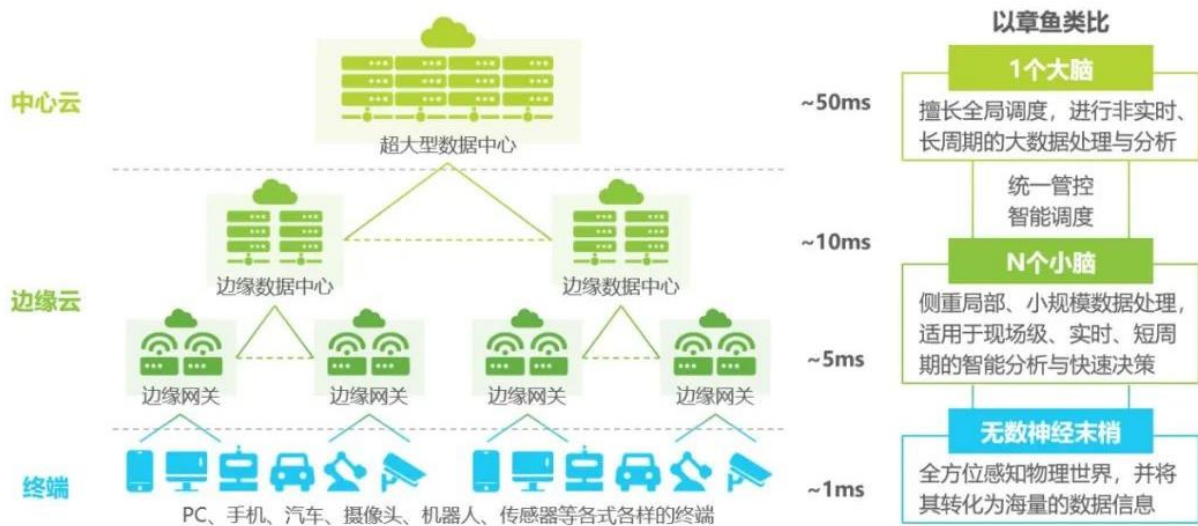
1 智能物联新时代，边缘AI 蓬勃发展

1.1 算力需求日益增长，边缘计算应运而生

云计算作为一种提供强大计算资源的网络服务，主要分为 IaaS（提供云服务器）、PaaS（提供云数据库）和 SaaS（直接帮助用户在云上运行软件）。但我们看到，随着终端应用场景的扩展带来数据即时处理需求的不断增长，云计算的不足开始显现，主要表现在：1）数据传输的限制性：随着越来越多的设备连接到互联网并生成数据，以中心服务器为节点的云计算在数据传输方面开始遇到带宽瓶颈；2）数据处理的即时性：海量物联网设备的接入使得云计算在数据处理的即时响应方面开始有所延迟；3）隐私及能耗问题：下游医疗、工业等设备采集的隐私数据传输到云端数据中心的链路较长，容易导致数据丢失或者信息泄露等风险，此外数据中心的高负载带来的高能耗也是管理规划的核心问题。

边缘计算是云计算向边缘侧分布式拓展的新触角。因此，边缘计算应运而生。根据边缘计算产业联盟的定义，边缘计算是指在靠近物或数据源头的网络边缘侧，融合网络、计算、存储、应用核心能力的开放平台，就近提供边缘智能服务，满足行业数字化在敏捷联接、实时业务、数据优化、应用智能、安全与隐私保护等方面的关键需求，总结下来就是“**在更靠近终端的网络边缘上提供服务**”。可以看到，边缘计算是云计算在面向物联网、大流量等场景下，为了满足更广连接、更低时延、更好控制等需求，向终端和用户侧延伸形成的新解决方案。

图表1：云-边-端视角下的边缘计算



资料来源：艾瑞咨询，方正证券研究所

“云-边-端”形态不断变化，边缘计算重要性日益凸显。我们看到，在早期的电信网时代，“云-边-端”的形态体现为“程控交换中心——程控交换机——电话”；到了互联网时代，这种形态变成了“数据中心——CDN——移动电话/PC”；而在现今的云计算+物联网时代，“云计算中心——小数据中心/网关——传感器”则构成了新的“云-边-端”形态。时代的发展体现在终端上就是海量设备的接入，从而带来数据处理需求的激增，如果仅依赖云端服务器，过大的工作负载不仅会导致处理效率变得低下，也无法适应终端更加多元化的处理需求。因此，更加侧重局部和小规模处理的边缘计算适用于现场级、实时、短周期的智能分析和快速决策，在物联网时代的背景下，其重要性日益凸显。

1.2 科技巨头相继发力，边缘 AI 方兴未艾

我们看到，科技巨头对于边缘计算的关注度正在日益提升。在 2022.12.14 召开的 MEET2023 智能未来大会上，高通产品管理高级副总裁 Ziad Asghar 指出，数据会不断从边缘侧产生，因此 **AI 处理的重心正在持续向边缘转移**。因为消费者希望拥有更好的数据隐私、更可靠的数据来源和即时的处理结果，因此边缘侧是进行 AI 处理的最佳选择，而高通一直在推动这场变革。

图表2:AI 处理重心向边缘侧转移



资料来源：高通，方正证券研究所

高通转型“智能边缘计算”，押注 C 端边缘 AI。目前，高通已经将 AI 算法应用到了智能手机的影像、图形处理等功能中，未来，高通将基于涵盖了边缘 AI、影像技术、图形技术、多媒体效果、极快处理速度，以及 5G 连接能力等特性的“统一的技术路线图”，将这些 AI 技术从耳机等较低复杂度的产品，规模化扩展到汽车等高复杂度的产品，最终实现智能网联边缘的全面覆盖。我们看到，在 2023.5.30 召开的 COMPUTEX 2023 上，高通高级副总裁 Alex Katouzian 表示，高通正在从一家通信公司转型为一家“智能边缘计算”公司。我们认为，此举进一步彰显了高通对于发展边缘 AI 的决心，未来将基于自身在智能手机领域积累的优势，不断研发迭代带有更丰富 AI 技术的芯片，充分赋能可穿戴设备、智能家居、智能汽车等更为广泛的消费者终端。

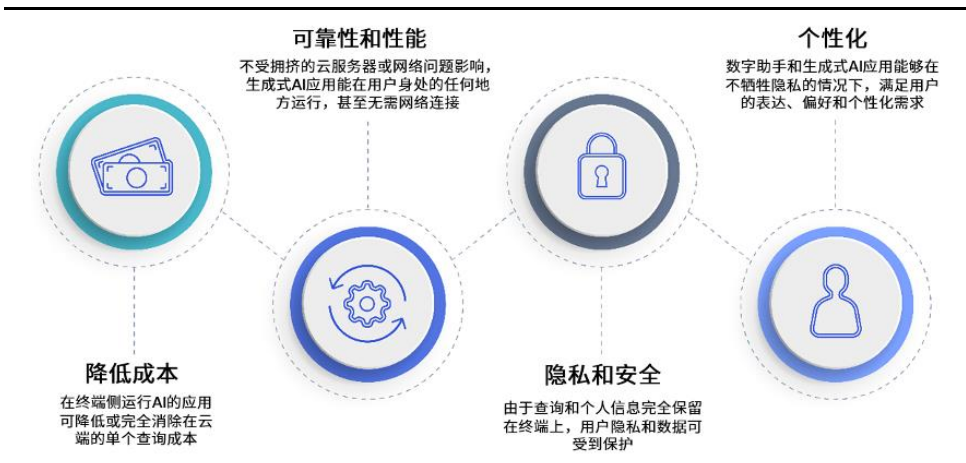
图表3:高通边缘 AI 技术路线图



资料来源：高通，方正证券研究所

借力 Meta Llama 2 大语言模型，高通终端侧 AI 部署有望更进一步。高通的最新动态显示，其在 2023 年 7 月 18 日宣布和 Meta 正在合作优化 Meta Llama 2 大语言模型在智能手机、PC、VR/AR 和智能汽车等终端上的直接运行。我们看到，与仅仅使用云端 AI 部署和服务相比，终端侧 AI 部署能够显著降低开发云成本，同时提升用户隐私保护、满足用户安全偏好、增强应用可靠性，并带来更加个性化的使用体验。高通计划从 2024 年起，在搭载骁龙平台的智能手机等终端上支持基于 Llama 2 的 AI 部署，赋能开发者使用高通 AI 软件栈面向终端侧 AI 进行应用优化，从而推出全新的生成式 AI 应用。

图表4:终端侧 AI 的四大关键优势



资料来源：高通中国公众号，方正证券研究所

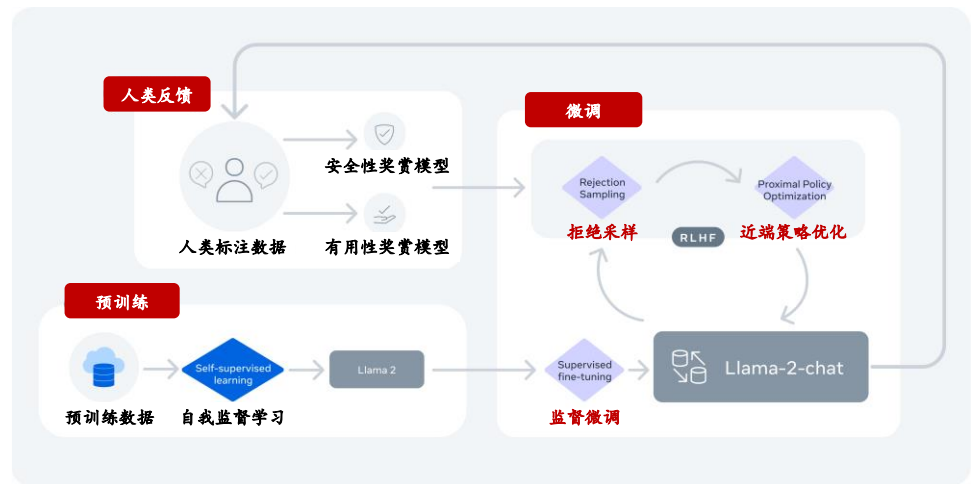
Meta Llama 2 免费商用且开源，性能较初代版本显著提升。具体来看 Meta 在 2023 年 7 月 18 日正式发布的大语言模型 Llama 2，这是 Meta Llama 模型的最新版本，也是 Meta 首个免费商用的开源 AI 模型。该模型包括 Llama 2 和 Llama 2-chat 两种版本，后者针对双向对话进行了微调，是针对类似于 ChatGPT 的聊天应用程序而开发，两者均细分为支持 70 亿、130 亿、700 亿等多种不同参数规模的版本。我们看到，相比于今年 2 月发布的 Llama 1，Llama 2 的 tokens 训练量提升了 40% 达到 2 万亿个，对上下文训练的长度是 Llama 1 的两倍，达到 4096，此外 Llama 2-chat 微调模型还接受了超过 100 万个人工标注数据的训练。

图表5:Llama 2 模型特点

Llama 2 was trained on 40% more data than Llama 1, and has double the context length.		
Llama 2		
MODEL SIZE (PARAMETERS)	PRETRAINED	FINE-TUNED FOR CHAT USE CASES
7B	Model architecture: Pretraining Tokens: 2 Trillion Context Length: 4096	Data collection for helpfulness and safety: Supervised fine-tuning: Over 100,000 Human Preferences: Over 1,000,000
13B		
70B		

资料来源：Meta AI，方正证券研究所

图表6:Llama-2-chat 模型的训练过程



资料来源: Meta AI, 方正证券研究所

Llama 2 目前仍难以媲美 GPT-4, 但有望利用开源生态实现弯道超车。此外, Meta 还选择了 70 亿参数版本的 Llama 2 和其他闭源的大语言模型进行比较, 可以发现其在大规模多任务语言理解 (MMLU) 和高质量小学数学问题 (GSM8K) 上的表现接近 GPT-3.5, 但在编码基准 (HumanEval) 上与 GPT-3.5 及 GPT-4 仍存在显著差距; 相比谷歌 PaLM (540B), Llama 2 (70B) 的所有结果几乎持平, 有些甚至表现更好, 但与 GPT-4 和 PaLM-2-L 相比仍有较大差距。可以看到, 目前 Llama 2 的性能表现还难以撼动 OpenAI 的市场地位, 但我们看到, 开源路线的 Llama 2 可以使得用户可以低成本地在本地服务器上部署开源大模型, 以构建适用于自身业务的专用大模型, 从而不需要将数据放到 OpenAI 等闭源大模型提供商的服务器中; 与此同时, 开源社区中大量用户产生的创意也将提升大模型的迭代速度。

图表7:Llama 2 (70 亿参数版本) 与其他闭源大语言模型的性能对比

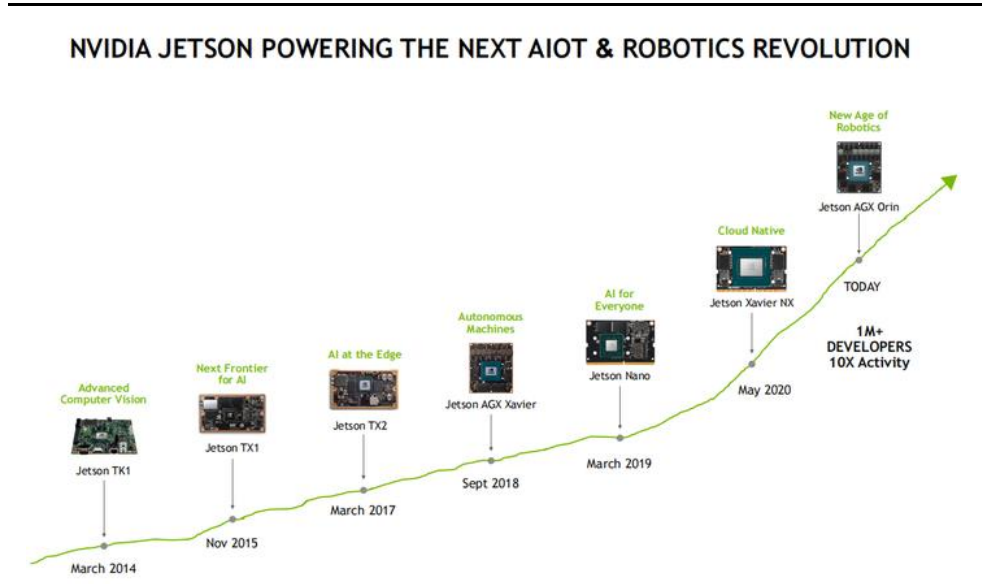
Benchmark (shots)	GPT-3.5	GPT-4	PaLM	PaLM-2-L	LLAMA 2
MMLU (5-shot)	70.0	86.4	69.3	78.3	68.9
TriviaQA (1-shot)	-	-	81.4	86.1	85.0
Natural Questions (1-shot)	-	-	29.3	37.5	33.0
GSM8K (8-shot)	57.1	92.0	56.5	80.7	56.8
HumanEval (0-shot)	48.1	67.0	26.2	-	29.9
BIG-Bench Hard (3-shot)	-	-	52.3	65.7	51.2

资料来源:《Llama 2: Open Foundation and Fine-Tuned Chat Models》, 方正证券研究所

投资 OpenAI 押注闭源路线, 携手 Meta 探索开源方案, 微软两头布局驱动 LLM 竞争日趋激烈, 未来发展机遇与挑战并存。我们看到, Meta 此次选择微软作为 Llama 2 的首选合作伙伴, 微软 Azure 客户将能够在 Azure 上轻松安全地微调 and 部署 Llama 2 模型, 并利用其云原生工具进行内容过滤和安全保护。此外, Llama 2 经过优化后可以在 ows 上本地运行, 为开发人员提供无缝的工作流程, 为跨不同平台的客户带来生成式 AI 体验。可以看到, 微软一方面通过 OpenAI 押注闭源路线, 另一方面又与 Meta 合作探索开源方案, 我们认为此举表明 LLM 领域的技术路线和竞争格局尚未定型, 未来在不同技术路径上的开发将会产生源源不断的创新, 交互式推动 AI 的发展。

英伟达 Jetson 平台不断迭代，发力 B 端边缘 AI。此外，我们同样注意到英伟达在边缘 AI 领域也有长足的积累。英伟达在 2014 年推出第一代 Jetson 平台 TK1，截至目前已迭代至第七代 AGX Orin，作为一套嵌入式系统开发板，主要包括 Jetson 硬件模组（小型高性能计算机）、用于加速软件的 JetPack SDK（软件开发工具包），以及包含传感器、SDK、服务和产品的生态系统。该平台使用英伟达自主研发的 GPU 加速技术，支持深度学习框架，能够加速模型的训练和推理，同时还支持各种图像、视频、通信和存储接口，适用于各种嵌入式应用和边缘计算场景，如机器人、自动驾驶、智能监控、医疗影像处理等。

图表8:英伟达 Jetson 平台产品迭代路线图



资料来源：英伟达，方正证券研究所

图表9:英伟达 Jetson 平台支持的边缘计算应用场景



资料来源：英伟达，方正证券研究所

Jetson AGX Orin 工业级模块全新发布，持续强化 B 端环境适应性能。我们看到，在 COMPUTEX 2023 上，英伟达发布了全新的 Jetson AGX Orin 工业级模块，该模块扩展了上一代 Xavier 工业级模块和商用 Orin 模块的功能，在恶劣环境下可以

提供更高级别的计算能力。具体来看，Jetson AGX Orin 工业级模块可以在 15-75W 功率范围内提供高达 248TOPS 的 AI 性能，通过 Ampere 架构 GPU、新一代深度学习和视觉加速器、高速 I/O 以及超快的内存带宽，可支持多个并行 AI 应用流程，而且可适应更加极端的温度范围，具有更长的工作寿命、更强的耐冲击和振动能力，整体性能相比 AGX Xavier 工业级模块提升了 8 倍以上。

图表10:Jetson AGX Orin 工业级模块强化环境性能

	Jetson AGX Orin 64GB	Jetson AGX Orin工业级模块
AI性能	275 TOPS	248 TOPS
模块	2048核NVIDIA Ampere架构GPU 带有64颗Tensor Core 12核Arm Cortex A78AE CPU 64-GB LPDDR5 64-GB eMMC	2048核NVIDIA Ampere架构GPU 带有64颗Tensor Core 12核Arm Cortex A78AE CPU 64-GB LPDDR5带有内联ECC 64-GB eMMC
工作温度	TTP下-25° C-80° C	TTP下-40° C-85° C
模块功率	15-60W	15-75W
工作寿命	5年	10年(85° C下为87,000小时)
冲击	非运行状态: 140G, 2ms	非运行状态: 140G, 2ms 运行状态: 50G, 11ms
振动	非运行状态: 3G	非运行状态: 3G 运行状态: 5G
湿度	偏置湿度, 85° C, 85%RH, 168 小时	85° C/85%RH, 1000 小时, 开机状态
耐温性能	-20° C下24小时; 45° C下168小时(运行状态)	-40° C下72小时; 85° C下1000小时(运行状态)
物理尺寸	100 mm x 87 mm	100 mm x 87 mm
底部填充	-	SoC角落粘接和元件底部填充
生产生命周期	7年(到2030年)	10年(到2033年)

资料来源：英伟达，方正证券研究所

“高通向左，英伟达向右”，B、C 端共同发力边缘 AI。综上所述，高通和英伟达在边缘 AI 领域的发展路线有所差异，高通更加关注 C 端消费者应用，英伟达则侧重于 B 端应用场景，我们认为这是两家公司基于自身优势而做出的正确战略决策。高通在智能手机 SoC 领域积淀深厚，有利于推动终端 AI 芯片继续向其他 C 端设备延伸；而英伟达是 GPU 传统云端 AI 芯片霸主，更加合适布局 B 端大算力场景。我们看到，边缘计算的快速发展主要受益于其在以下四方面的优势：

- **低时延：**由于消除了将数据从端点转移到云端再返回的需要，大量复杂的数据可以在边缘端进行初筛、分析和计算，从而更快地做出决策，避免了因海量数据涌向云端，带来线路阻塞或响应缓慢等问题。
- **高效率：**因为数据处理的多数任务在靠近数据源的地方完成，分析及反馈的速度得以提升，增强了实时决策能力，同时由于减少了从云端来回传输的数据量，也能够更好地实现降本增效。
- **安全性：**边缘端服务器广泛就近分布，且单个设备存储的数据量较为有限，既降低了因传输距离过长导致数据丢失的风险，也减小了信息窃取的可能性。
- **智能化：**边缘计算系统的可扩展性使得用户可以自由地部署和管理计算资源，其背后拥有的大量自我适应、表达、修复等机制，充分赋予了边缘节点设备的智能化属性，可以在不依赖云端决策的情况下做出响应，单个设备的故障也不会影响生态系统中其他设备的性能，从而提高了整个连接环境的可靠性。

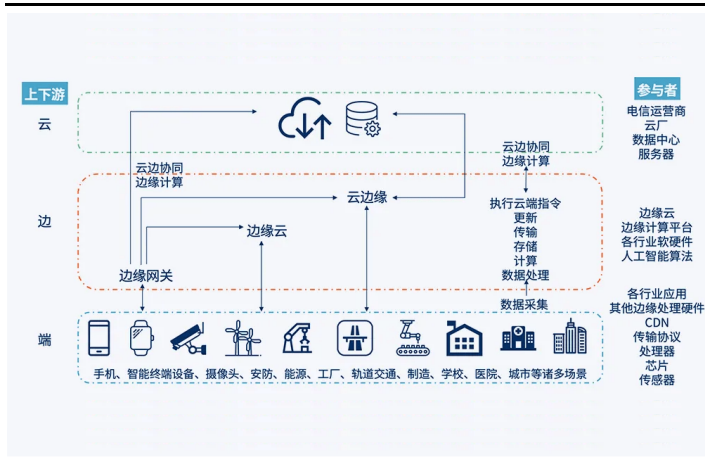
图表11:边缘计算 VS. 云计算

	边缘计算	云计算
目标应用	物联网或移动应用	一般互联网应用
服务器节点位置	边缘网络（网关、WiFi接入点和蜂窝基站等）	数据中心
客户端与服务器的通信网络	无线局域网、4G/5G等	广域网
可服务的设备（用户）数量	数十亿计	数百万计
提供的服务类型	基于本地信息的服务	基于全局信息的服务
地理位置分布	分布式	集中式
访问时延	可达0.5ms	2s左右
移动性支持	支持高速移动性服务	移动性较差
网络带宽	很低	非常高

资料来源：51CTO，方正证券研究所

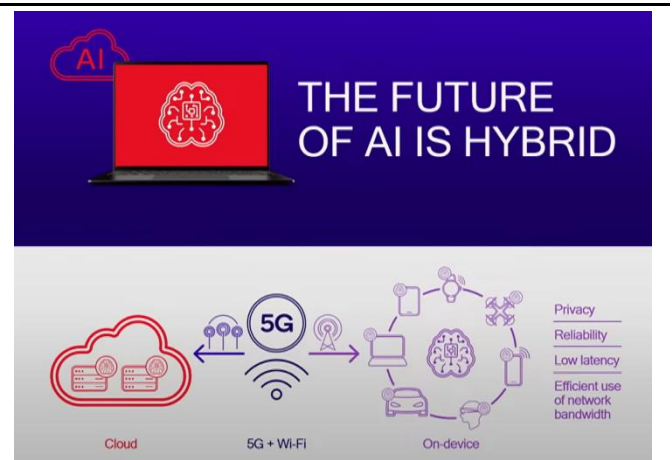
“云-边-端”紧密协同，混合式 AI 时代来临。虽然边缘计算相比云计算存在一定优势，但这并不意味着二者处于对立面，而更多是一种相互依存的关系。一方面，边缘计算通过负责自身范围内的数据计算和存储工作，可以分担云端的计算压力；另一方面，大部分经过边缘设备处理的数据仍需要从边缘节点汇聚到中心云，云端再通过分析完成对算法模型的训练和升级，并将升级后的算法推送到前端，帮助前端设备实现更新和升级，从而达成自主学习的闭环。此外，这些数据也需要上传到云端进行备份，当边缘计算过程中出现意外情况，存储在云端的数据也不会丢失。这与高通在 COMPUTEX 2023 上提出的“混合式 AI”概念一脉相承，即未来云端和边缘/终端需要共同承担日益增长的算力需求。我们认为，云计算与边缘计算需要通过紧密协同才能更好地满足各种需求场景的匹配，从而最大化体现云计算与边缘计算的应用价值。

图表12:云边协同



资料来源：CB Insights China，方正证券研究所

图表13:混合式 AI



资料来源：高通，方正证券研究所

1.3 突破算力及功耗瓶颈，加速边缘 AI 落地

解决算力和功耗限制，赋能边缘 AI 落地。我们看到，随着 ChatGPT 的横空出世，生成式人工智能迅猛发展。但以 GPT-4 为代表基于 Transformer 架构的生成式预训练大模型由于具备较大规模的参数，其训练与推理仍然需要在云端借助算力强大的 GPU 等 AI 芯片完成。但我们同样看到，随着机器学习、神经网络训练等网络架构和工具不断适配、兼容到嵌入式系统上，越来越多的 AI 应用也可以直接在边缘设备上运行，因此“边缘 AI”悄然而生。我们认为，在“云-边-端”的架构下，边缘 AI 的主要存在形式仍为以智能手机为代表的终端嵌入式设备。但由于终端设备的算力和功耗有限，因此如何解决这两大问题便成为了边缘 AI 落地的关键所在。目前在算力方面，主要是通过压缩模型并采用联网方式来降低算力需求，而功耗方面则是采用存算一体的方式来突破功耗瓶颈。

一、降低算力需求——压缩模型+联网

一般而言，神经网络模型包含的网络层数和参数越多，其结构就越复杂，但实际运行的效果也会越好。所谓模型压缩，是指通过算法将一个复杂的预训练大模型转化成精简的小模型。按照压缩过程对网络结构的破坏程度，模型压缩技术可以分为“前端压缩”和“后端压缩”，前端压缩几乎不改变原有的网络结构，仅仅是在原模型的基础上减少了网络的层数，但后端压缩会对网络结构造成不可逆的改变，后续的维护成本也会很高。因此，业内一般采用前端压缩，主要包含以下三种细分方法：

1) 知识蒸馏

属于迁移学习的一种，主要思想是将学习能力强的复杂教师模型中的“知识”迁移到简单的学生模型中，本质是让小模型去拟合大模型，从而让小模型学到与大模型相似的函数映射，使小模型保持其快速计算速度的前提下，同时拥有复杂模型的性能，达到模型压缩的目的。

2) 模型剪枝

大模型虽然参数很多，但也存在着大量冗余的参数，将这些不太重要的训练参数剔除，可以减少计算资源的消耗并提高实时性，这就是模型剪枝算法。训练步骤为“正常训练模型——模型剪枝——重新训练模型”，三个步骤反复迭代进行，直到模型精度达到目标，则停止训练。

3) 模型量化

是指将神经网络的浮点运算（FP32、FP16）转换为定点运算（INT8），从而减少内存占用并提高计算效率。目前的低精度模型一般采用的数值格式为 INT8，但也有些混合精度模型中会使用 FP16 的数值格式来保证某些参数和操作符的准确度。

可以看到，以上三种压缩算法可以降低模型对计算能力的要求，使得 AI 模型在终端设备上的运行成为可能。此外我们也注意到，模型压缩不可避免会造成性能损失，有可能导致小模型在某些应用场景下的使用出现困难，因此终端设备可通过联网访问云端的大模型，可进一步保障设备 AI 性能的落地。

图表14: 大模型小型化

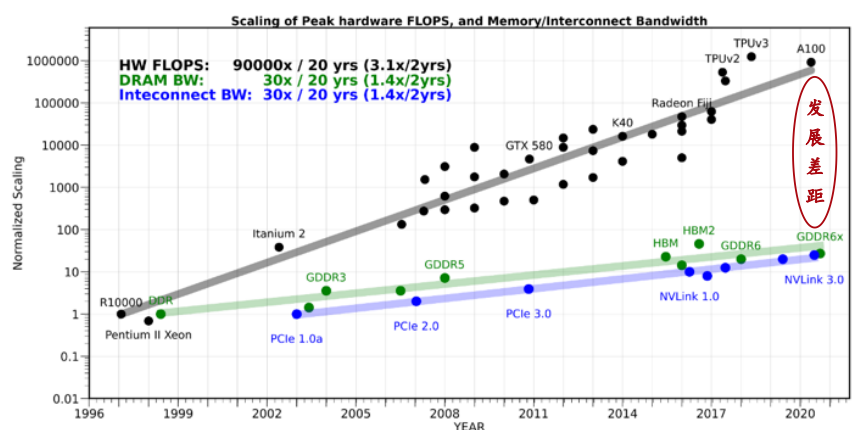


资料来源：WAVE SUMMIT，方正证券研究所

二、突破功耗限制——存算一体

AI 时代算力需求猛增，“存储墙”亟待解决。我们知道，在传统的冯诺依曼架构下，芯片的计算单元在计算之前需要先从存储器中读取数据，但数据的搬运时间往往是计算时间的成百上千倍，这背后是处理器和存储长久以来性能发展不均衡的结果。在过去二十年，处理器性能大概每年以 55% 的速度提升，但内存性能每年的提升速度只有 10% 左右。而内存性能的落后直接导致的问题就是数据搬运的功耗过高，由此带来的无用功耗能占到整个 AI 计算功耗的 60-90%，造成非常低的能效比，这就是我们熟知的“存储墙”问题。现如今随着 AI 的快速发展，算力需求大幅提升，“存储墙”带来的计算效率和功耗问题日益突出。

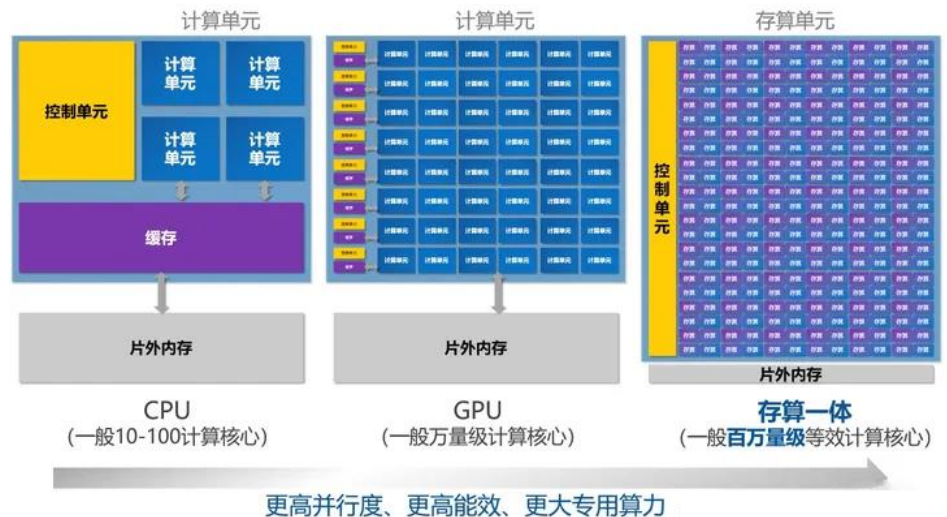
图表15: 算力 VS. 存储性能提升



资料来源：51CTO，方正证券研究所

存算一体突破“存储墙”限制，大幅提升计算效率并降低功耗。在此背景下，存算一体技术（Computing in Memory, CIM）应运而生，该技术是在存储器中嵌入了计算能力，直接利用存储器进行数据处理，从而把数据存储与计算融合在芯片的同一片区，从本质上消除了不必要的搬运，因此可以大幅提升计算效率并降低功耗，适用于深度学习、人工智能等大规模并行计算的应用场景，是在冯诺依曼架构之外的一种全新芯片计算架构。

图表16: 存算一体芯片架构



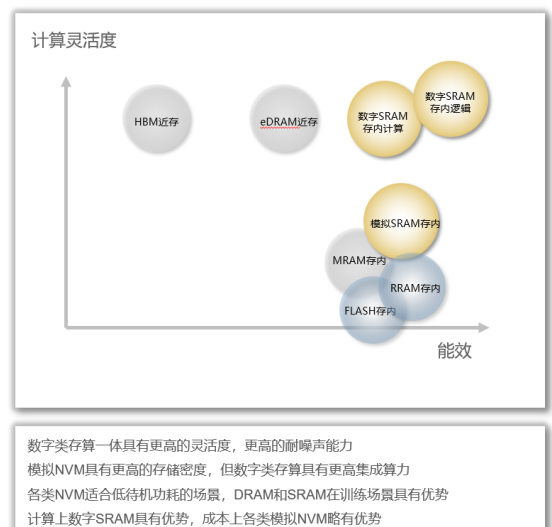
资料来源：知乎@陈巍谈芯，方正证券研究所

在计算方式上，存算一体芯片分为数字计算和模拟计算，其中数字计算更偏向于“算”，采用的是先进逻辑工艺，拥有高性能、高精度的优势，且具备很好的抗噪声能力和可靠性，适用于大算力、高能效的商用场景；而模拟计算更偏向于“存”，具有更高的存储密度，但对环境噪声和温度非常敏感，适用于小算力、不需要太强可靠性的民用场景。

在存储介质上，主要可分为成熟工艺存储器（DRAM、Flash 和 SRAM）和新型存储器（MRAM、RRAM 和 FRAM 等）。细分为来看，DRAM成本低、速度快，但容量小；Flash属于非易失性存储介质，具有低成本、高可靠性的优势，但在工艺制程上有明显瓶颈，一般适用于小算力场景；SRAM在速度和能效比方面具有很大优势，但容量密度略小，一般适用于云计算等大算力场景；新型存储器中，RRAM在神经网络计算中具有特别的优势，是下一代存算一体介质的主流研究方向，但目前工艺良率还在爬坡，而且还需要解决非易失存储器固有的可靠性问题，因此目前仍主要用于边缘 AI 计算和端侧小算力场景。

图表17: 存算一体存储介质对比

存储器类型	优势	不足	适合场景
DRAM	高存储密度，整合方案成熟	只能做近存计算，速度略低，工艺迭代慢	适合现有冯氏架构向存算过渡
Flash	高密度低成本，非易失，低漏电	对PVT变化敏感，精度不高，工艺迭代时间长	小算力、端侧、低成本、待机时间长的场景
SRAM（数字模式）	能效比高，高速高精度，对噪声不敏感，工艺成熟先进，适合IP化	存储密度略低	大算力、云计算、边缘计算，也适合训练
SRAM（模拟模式）	能效比高，工艺成熟先进	对PVT变化敏感，对信噪比敏感，存储密度略低	小算力、端侧、不要求待机功耗
各类NVRAM（包括RRAM/MRAM等）	能效比高，高密度，非易失，低漏电	对PVT变化敏感，有限写次数，相对低速，工艺良率尚在爬坡中	小算力、端侧/边缘Inference、待机时间长的场景



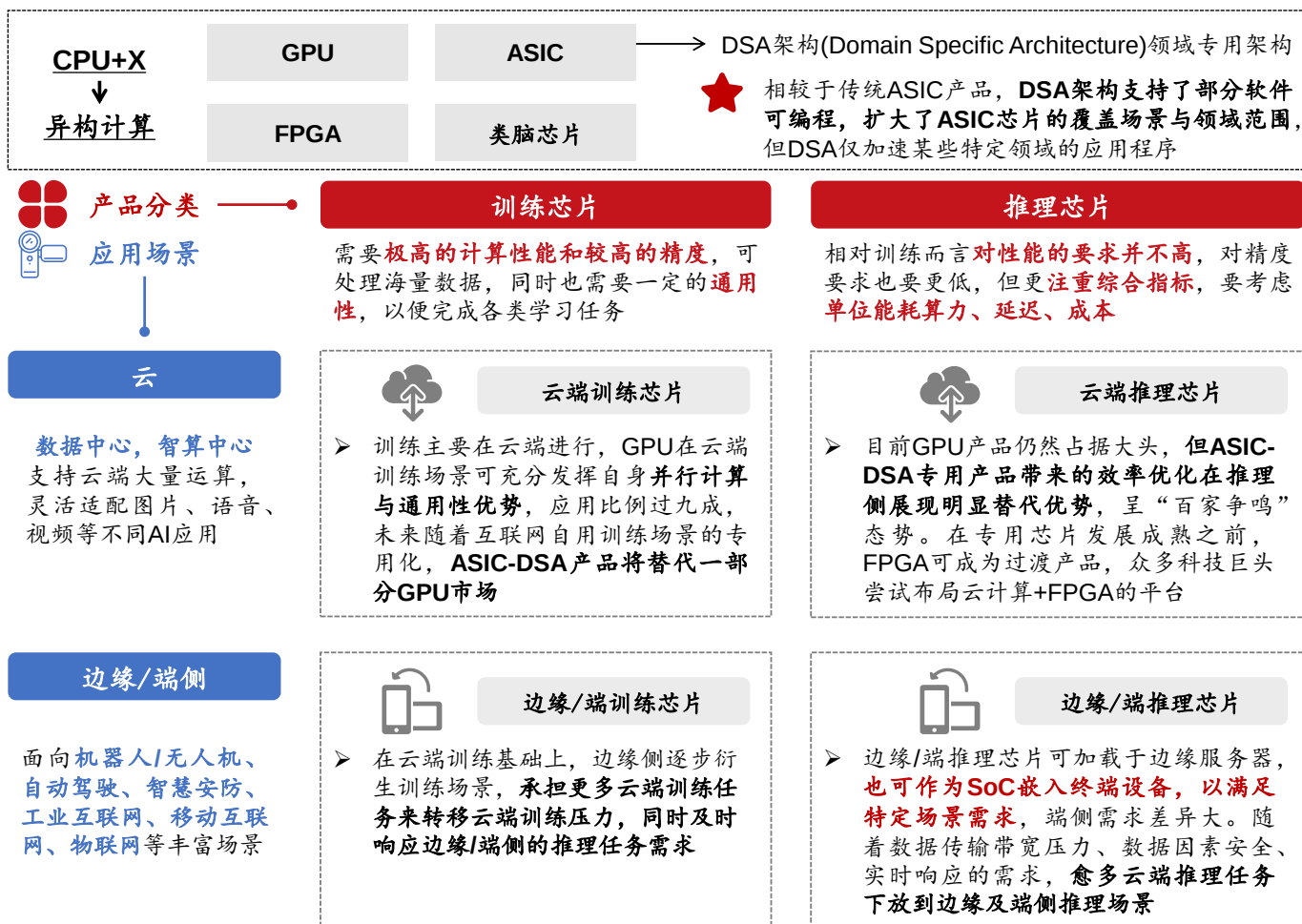
资料来源：电子发烧友，方正证券研究所

2 智能终端 SoC，端侧 AI 算力承载

2.1 AI SoC 多模块异构集成，重视性能与功耗平衡

边缘 AI 算力承载者——智能终端 SoC。所谓 AI 芯片，其实是在 CPU 等传统芯片的基础上，针对 AI 算法（以深度学习为代表的机器学习算法）做了特殊加速设计的芯片，换言之，是牺牲了一定的通用性，换取了芯片在海量数据并行计算方面的效率提升。在类型上，AI 芯片主要分为 GPU、FPGA、ASIC 和类脑芯片；在应用场景上，主要分为云端（训练+推理）和边缘端（边缘计算+终端），其中应用在云端和边缘计算服务器上的 AI 芯片是以 GPU 为代表的强通用性芯片，而终端 AI 芯片更多是将 AI 功能模块结合 CPU、GPU、ISP 等模块融合进 SoC 中，并针对语音识别、人脸识别等不同的应用场景完成定制化设计，同时需要做好性能与功耗的平衡。因此我们认为，海量的物联网终端设备是边缘 AI 的主要落地场景，而 AI SoC 作为边缘算力的承载体，有望跟随 IoT 设备的智能化浪潮实现快速成长。

图表18: AI 芯片分类



资料来源：艾瑞咨询，方正证券研究所

SoC (System on a Chip, 系统级芯片)是在 CPU 的基础上扩展了音视频功能和专用接口的超大规模集成电路，一般集成了 CPU、GPU、NPU、ISP、存储、WiFi、蓝牙、总线、接口等 IP 模块，具体来看其中重要模块的功能：

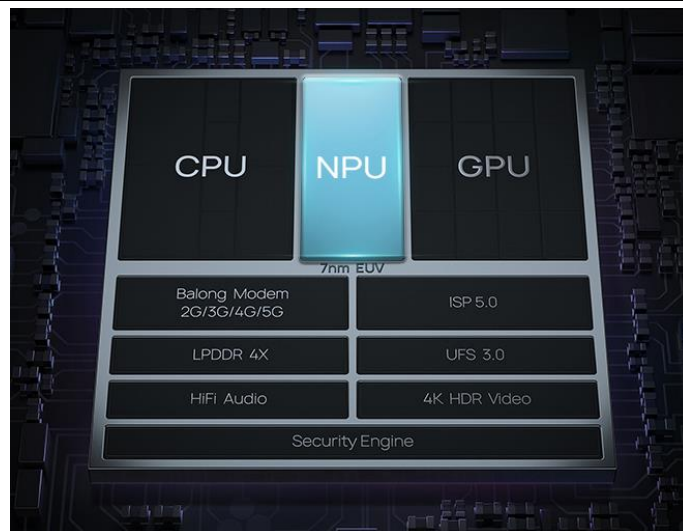
1) **CPU (Central Processing Unit, 中央处理器)**：主要执行通用计算，用来解释计算机指令、处理计算机软件中的数据。

2) GPU (Graphics Processing Unit, 图形处理器): 主要完成高清视频加速解码、游戏图形处理、图像处理 and 辅助运算等功能。

3) ISP (Image Signal Process, 图像处理器): 主要用来处理前端图像信号, 能够将光信号变成电信号, 其性能会直接影响到拍照和摄像性能。

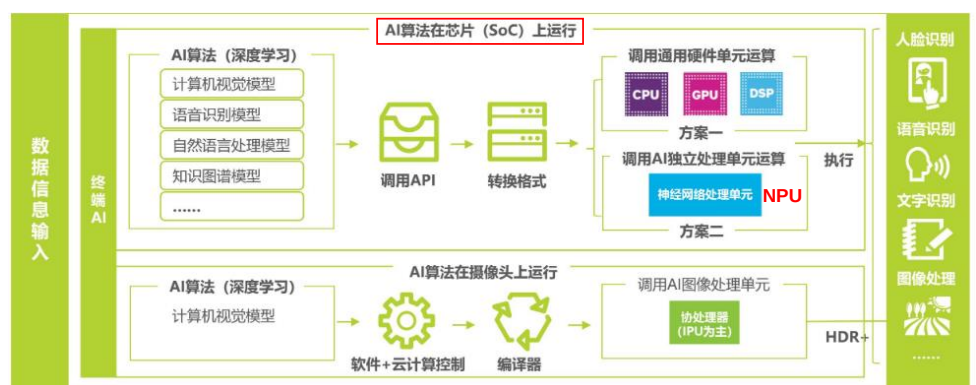
4) NPU (Neural-network Processing Unit, 神经网络处理器): 专为物联网人工智能而设计, 用于加速神经网络的运算, 解决传统芯片在神经网络运算时效率低下的问题, 是目前主流的 AI 算力载体。例如麒麟 990 5G SoC 采用华为自研架构 NPU, 创新设计 NPU 双大核+NPU 微核架构, 其中 NPU 大核针对大算力场景 (人像虚化/超级夜景等复杂 AI 任务), NPU 微核赋能超低功耗 AI 应用 (隔空手势/AI 信息保护等轻量任务), 充分发挥全新 NPU 架构的智慧算力。

图表19: 华为麒麟 990 5G SoC 芯片架构



资料来源: 华为官网, 方正证券研究所

图表20: AI 算法的运行原理



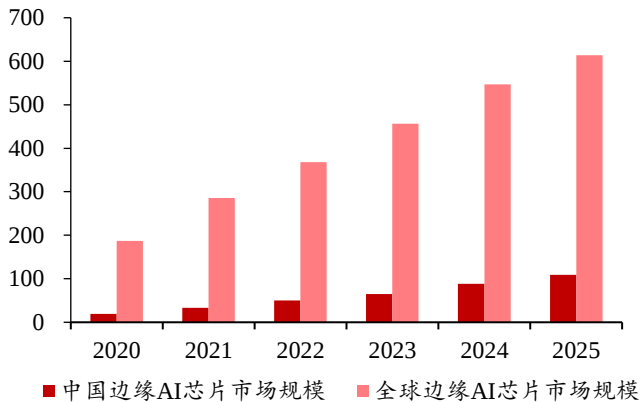
资料来源: 艾瑞咨询, 方正证券研究所

2.2 边缘 AI 芯片规模稳健成长, 海内外厂商加速布局

AIoT 智能化浪潮来临, 边缘 AI 芯片赋能下游迎来蓬勃发展。我们看到, 相比云端 AI 芯片需要兼具训练和推理性能, 边缘 AI 芯片则更多承担推理任务, 其主要存在形式是作为 AI SoC 嵌入终端设备, 因此性能上更加追求在算力、功耗和成本等多方面的综合表现。伴随人工智能在消费电子、智能家居、智能安防、智能驾驶等 IoT 领域的快速渗透, 边缘 AI 芯片也将迎来蓬勃发展。

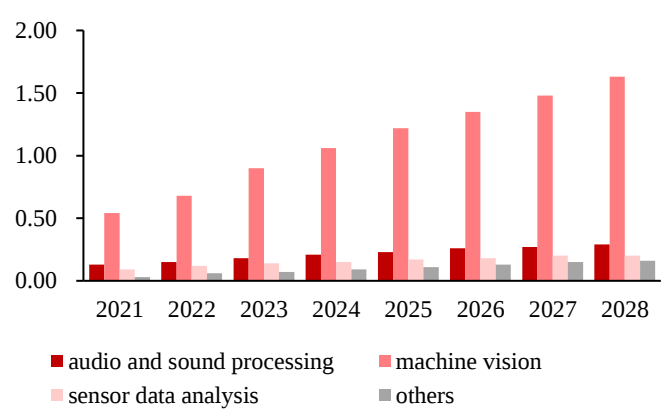
根据 Gartner 数据，全球边缘 AI 芯片市场规模在 2022 年不到 400 亿美元，但有望在 2025 年突破 600 亿美元，其中国内市场空间将从 2022 年的 50 亿美元增长至 2025 年的 110 亿美元，CAGR 达 30%。就出货量而言，ABI Research 发布的数据显示，2023 年全球边缘 AI 芯片出货预计达 13.12 亿颗（其中语音处理 1.93 亿颗，机器视觉 9.01 亿颗，传感器数据分析 1.40 亿颗，其他种类 0.78 亿颗），并预计出货量到 2028 年将提升至 22.86 亿颗，其中语音处理芯片、机器视觉芯片、传感器数据分析芯片出货量 CAGR 分别达到 9%、12%、8%。

图表21:全球及中国边缘 AI 芯片市场规模（亿美元）



资料来源：半导体行业观察，方正证券研究所

图表22:全球不同功能边缘 AI 芯片出货量（十亿颗）



资料来源：半导体行业观察，方正证券研究所

回顾 AI SoC 领域的竞争格局，由于下游应用场景较多，每一细分领域的竞争情况也有所差异，主要可分为消费电子（智能手机、可穿戴设备）、智能家居、智能安防、智能驾驶（自动驾驶、智能座舱）四大下游市场，其中智能手机、智能驾驶和智能安防市场的进入门槛较高，但可穿戴设备和智能家居长尾市场的存在仍为国内厂商提供了切入契机。

图表23:AI SoC 厂商布局

下游应用	国内厂商	海外厂商
智能手机	HISILICON 紫光展锐™ UNISOC	Qualcomm MEDIATEK
可穿戴设备	中科蓝讯 bluetrum BES 恒玄科技	REALTEK Apple
智能家居	ESPRESSIF 乐鑫科技 @mlogic 晶晨 Rockchip 瑞芯微电子 ALLWINNER 全志科技	Qualcomm MEDIATEK
智能安防	HISILICON Fulhan	Ambarella® AI envisioned™ NOVATEK 联咏科技
自动驾驶	HUAWEI 地平线 Horizon Robotics 黑芝麻智能 BLACK SESAME TECHNOLOGIES	NVIDIA® mobileye™
智能座舱	HUAWEI siengine 芯擎科技 Rockchip 瑞芯微电子 ALLWINNER 全志科技	Qualcomm RENESAS

资料来源：各公司官网，方正证券研究所整理

2.3 自研 IP 积累及先进代工保障，彰显核心竞争力

设计环节自研 IP，追求差异化竞争。具体来看 SoC 芯片的制造流程，从最初借助 EDA 软件，利用 IP 核复用等技术进行芯片设计，到后续的晶圆制造和封装测试，每一环节都面临着不同的技术挑战。在设计环节，IP 复用性虽然帮助设计厂商提高了设计效率，但是如果完全依赖第三方的 IP 授权就不能做到差异化的竞争，因此自研 IP 便成为了头部厂商的战略选择。我们认为，自研 IP 不仅可以帮助节省授权费用，更重要的是能深入地学习理解底层技术，从而精简不同 IP 核之间的电路设计，提高产品质量和迭代速度，切实增强自身竞争力。

图表24: SoC 芯片制造流程



资料来源：亿欧智库，各公司官网，方正证券研究所

制造环节绑定先进制程代工产能，高端路线徐图进取。再来看制造环节，我们认为拥有稳定的代工产能是 SoC 设计厂商保障芯片落地和迭代的重要基础，尤其是与台积电、三星世界上唯二两家拥有 10nm 以下先进制程量产能力代工厂的合作，更是设计厂高端路线的坚实依靠。2020 年发布的麒麟 9000 卓越性能的背后，虽然离不开海思数十年设计能力的积累，但台积电 5nm FinFET 工艺的保驾护航同样发挥了关键作用。相比之下，三星的功耗控制能力仍有待提高。

图表25: 台积电 VS 三星代工芯片参数对比



资料来源：搜狐，电子工程专辑，安兔兔，新浪科技，方正证券研究所

3 AI SoC 赋能下游，终端应用百花齐放

AI 技术加速渗透 SoC，大幅革新终端设备智能化体验。如前所述，我们认为海量的物联网终端设备是边缘 AI 的主要落地场景，而融合了不同 AI 功能模块的 SoC 作为终端推理算力的承担者，将为不同应用的嵌入式设备带来更加智能化的体验。例如高通将模仿神经网络模型的 Hexagon 处理器融合进 SoC 中，打造专属的 AI 引擎并持续进行迭代，搭载最新一代 AI 引擎的骁龙 8 Gen2 芯片在自然语言处理、目标检测等方面的性能均实现了大幅提升。此外，高通还在 SoC 中开发了一个称之为“高通传感器中枢 (Sensing Hub)”的芯片子系统，能同时处理多路数据流，实现关键词识别、异常检测、图像分类等 AI 性能。可以看到，高通在 AI 领域的全面进军预示着终端 IoT 设备的智能化时代已经来临。

图表26:高通 AI 引擎



资料来源：高通，方正证券研究所

图表27:高通传感器中枢



资料来源：新浪，方正证券研究所

图表28:AI SoC 终端应用

	消费电子	智能家居	智能安防	智能驾驶
功能需求	● 图像/场景识别 ● 拍照美化 ● 语音助手等	● 图像识别 ● 语义识别与理解 ● 语音助手等	● 图像/视频识别 ● 图像/视频检测等	● 图像识别 ● 数据融合 ● 路径规划等
算力要求	1-8TOPS	<1TOPS	4-20TOPS	20-4000TOPS
功耗要求	高	较高	较高	中等
可靠性需求	高	较高	较高	高
成本敏感性	高	较高	较高	较低

资料来源：亿欧智库，方正证券研究所

3.1 消费电子

在消费电子产品中,SoC 中的 AI 模块主要用于实现语音识别和拍照美化等功能,例如苹果手机的 Siri 语音助手,近些年流行的计算摄影,都是 AI 技术的体现。就智能手机 SoC 而言,除了苹果使用自家的 A 系列芯片,安卓高端市场基本被高通垄断,中低端市场则由联发科主导。不管是资金实力、研发能力,还是产业链配套的先进代工、终端迭代等因素,高耸的行业进入壁垒都让新兴厂商望而却步。

但在可穿戴设备(TWS 耳机、智能手表、VR/AR 等) SoC 市场,虽然高通和苹果也有着领先优势,但由于存在着较多的长尾客户,叠加此类 SoC 对性能要求并不是很高,更多追求的是低功耗带来的续航时间,因此给到了国内的恒玄科技、中科蓝讯等公司一定的生存空间。

先看国外厂商的布局。苹果方面,其在 2022 年 9 月发布的 AirPods Pro 第二代搭载全新的 H2 芯片,运用计算音频算法,带来更智能的降噪表现(主动降噪效果相较上一代最高提升至 2 倍)、卓越的三维空间音效(全新自适应算法更快速地处理声音)和更高效的电池续航(单次充电最长能听 6 小时)。可以看到,此类 TWS 耳机的智能蓝牙音频 SoC 芯片更多追求的是降噪和音质表现。

图表29:苹果 AirPods Pro 第二代搭载的 H2 芯片



资料来源:苹果,方正证券研究所

高通方面,其在 2022 年 7 月发布了全新的骁龙可穿戴平台 W5/W5+,回顾过去,高通分别于 2016、2018、2020 年发布了骁龙 2100、3100、4100+可穿戴设备平台。命名方式的变化彰显着新一代平台相比前一代 4100+有着全面提升,其中整体功耗降低超 50%,性能提升达两倍以上,芯片尺寸缩小 30%以上,同时从蓝牙 4.2 升级为蓝牙 5.3,并新增集成式扬声器功率放大器等特性。

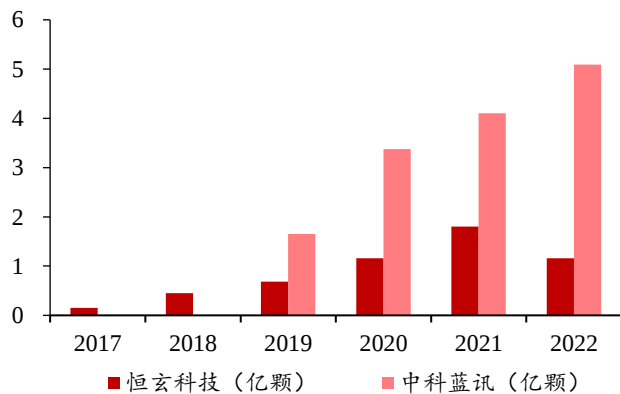
再来看国内厂商的布局。自从苹果在 2016 年发布第一代 AirPods 推动了 TWS 耳机的热潮,国内 TWS 芯片企业迅速跟进,也在这个市场中分得了一杯羹。国内的 TWS 芯片厂商主要分为两类,一类是以恒玄科技为代表的品牌芯片厂商,产品供货给华为、小米、OPPO 等终端耳机厂商;另一类是以中科蓝讯为代表的白牌芯片厂商,产品供货给品牌知名度较低的众多第三方商家,主要通过低价走量取胜。

图表30:高通骁龙全新可穿戴平台 W5/W5+



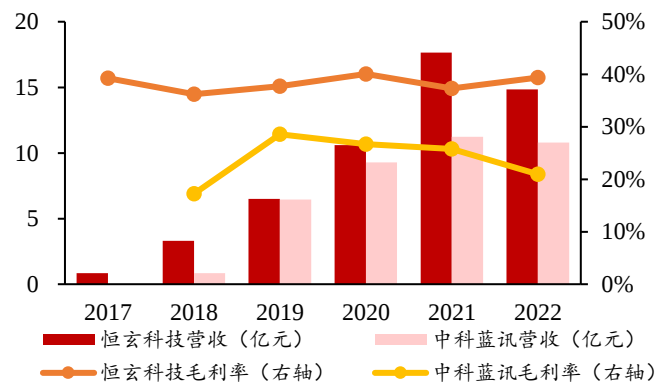
资料来源: cnBeta, 方正证券研究所

图表31:恒玄科技 VS. 中科蓝讯 TWS 蓝牙耳机芯片出货量



资料来源: 公司公告, 方正证券研究所

图表32:恒玄科技 VS. 中科蓝讯营收及毛利率



资料来源: 公司公告, 方正证券研究所

Apple Vision Pro 正式发布, 开启新一轮革命性创新。长期来看, 智能手机、TWS 耳机、智能手表等消费电子产品的成长空间已接近天花板, 我们认为 AR/VR 将是下一代现象级单品。苹果最新发布的 Vision Pro 已经为我们探明了该类产品的初步形态, 全新的交互体验也对内置的 SoC 芯片提出了更高要求。

图表33:苹果 Vision Pro 产品效果



资料来源: 苹果, 方正证券研究所

图表34:苹果 Vision Pro 搭载 M2 及 R1 芯片

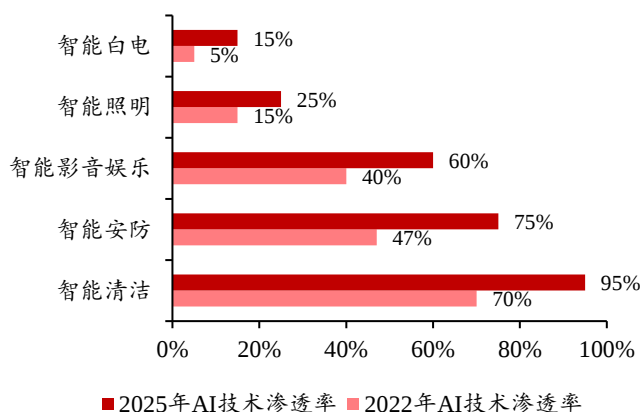


资料来源: 苹果, 方正证券研究所

3.2 智能家居

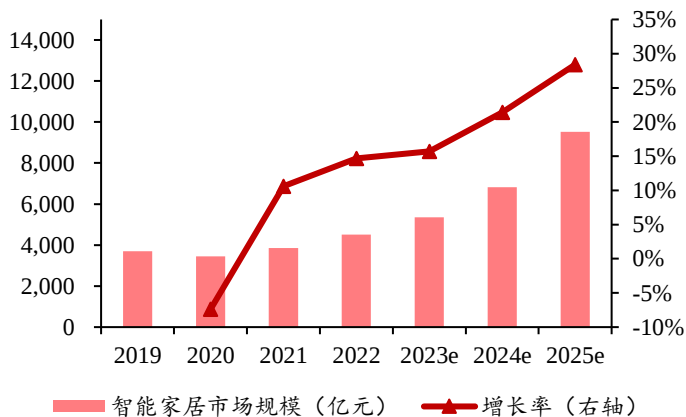
AI 技术持续渗透，智能家居市场蓬勃发展。我们看到，在智能家居领域，AI 技术的渗透主要是通过智能语音/视觉模组的方式应用于智能扫地机、家庭安防摄像机、智能音箱、智能门锁等各类智能家居产品中。根据艾瑞咨询的数据，2022 年 AI 技术在智能家居行业的整体渗透率约为 25%，预测到 2025 年将提升至 50% 左右，其中智能清洁（95%）、家庭安防（75%）、智能影音娱乐（60%）、智能照明（25%）、智能白电（15%）是 AI 渗透率排名前五的细分品类。伴随设备智能化体验的提升和消费者对居家舒适度的追求，艾瑞咨询预测中国的智能家居市场规模将从 2022 年的 4517 亿元增长至 2025 年的 9523 亿元，CAGR 达 28%。

图表35: 中国智能家居品类 AI 技术应用渗透率预测



资料来源：艾瑞咨询，方正证券研究所

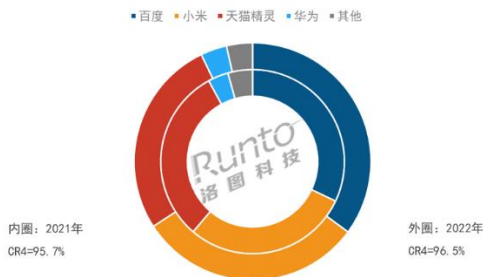
图表36: 中国智能家居市场规模及增速



资料来源：艾瑞咨询，方正证券研究所

在众多细分品类中，我们判断具备语音/视频交互能力的智能音箱有望成为智能家居全屋互联场景下的 AI 控制入口。智能音箱集成了人工智能处理能力，能够通过语音识别、语音合成、语义理解等技术完成语音交互功能，同时还可以提供智能家居控制、音乐内容服务、互联网服务等功能，目前主要分为无屏和带屏两大类，其中无屏音箱只具备最基础的语音交互功能，价格主要分布在 50-200 元；带屏音箱则在语音交互的基础上进一步增加了视觉交互体验，可通过屏幕进行文字/图片展示和视频通话，价格主要分布在 200-1500 元。

图表37: 中国智能音箱市场竞争格局



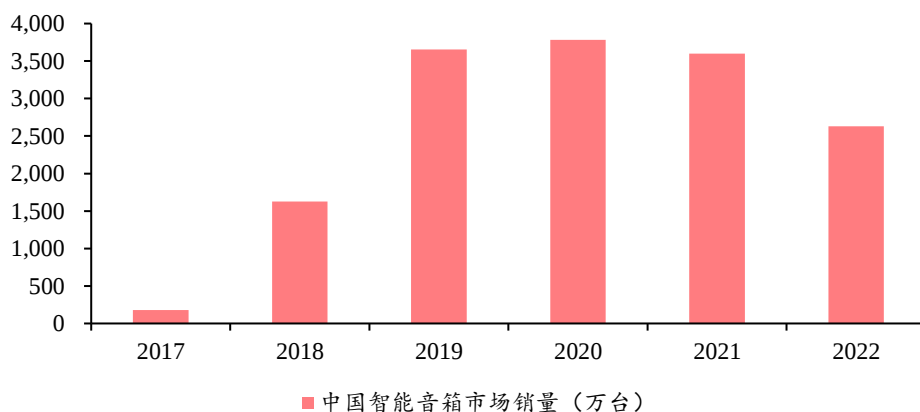
资料来源：洛图科技，方正证券研究所

图表38: 中国智能音箱头部企业分析

品牌名称	品牌优势
小度	百度智能音响的特色为人工智能技术、自身的 对话式人工智能系统DuerOS 和良好的人机交互体验，在产品战略方面更倾向于C端消费群体。通过大量的优质内容资源来吸引用户，再通过用户的使用数据来完善智能音响技术
小爱同学	小米的 智能家居生态链 带动了小米智能音响的销量。小米智能音响的 性价比高 ，智能家居产品通过智能音响连接，升级为AIoT，实现了生态闭环
天猫精灵	阿里巴巴为旗下的智能音响产品提供高质量的销售渠道，同时依靠 阿里云 的强数据处理，提高人工智能技术。阿里巴巴在产品战略方面是 B端和C端消费群体兼顾 ，不断拓展产品的应用场景，也积极寻找内容供应商，努力打造良好的生态链
HUAWEI	华为进入智能音响行业的时间较短，华为主打的是 高端无屏智能音响 ，和 鸿蒙系统 进行配合以及提供良好的音频体验，打造系统生态圈

资料来源：头豹研究院，方正证券研究所

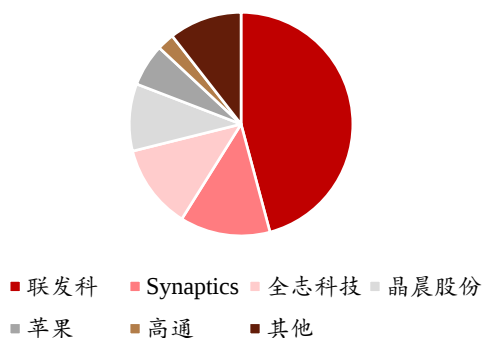
图表39:中国智能音箱市场销量(万台)



资料来源:洛图科技,方正证券研究所

AI 赋能叠加生态完善,智能音箱有望重拾增长。我们看到,智能音箱销量的爆发主要集中在 2018 和 2019 年,但在完成首轮的用户普及之后便陷入了增长瓶颈。究其原因,一是由于智能化升级缓慢,各大品牌推出的新品创新力度不够甚至没有创新,导致同质化现象严重;二是产品生态建立不完善,智能音箱与其它家居设备的互联互通能力受到限制,消费者体验不佳。但现如今随着 ChatGPT 的推出,生成式 AI 加速发展,我们认为以 GPT4 为代表的大模型所具备的多模态能力将大幅提升终端应用的智能化体验,智能音箱的语音/视频交互能力有望得到充分升级。此外,百度、小米、天猫精灵等国内智能音箱巨头历经多年的布局,各自的产品生态圈建设已趋于完善。因此,站在当前视角,阻碍智能音箱增长的两大关键因素均得到较大程度的缓解,我们看好该品类在未来的 AI 浪潮中重拾增长。

图表40:2020 年全球智能音箱处理器出货量份额



资料来源:Strategy Analytics,方正证券研究所

图表41:中国智能音箱处理器厂商客户布局

厂商名称	客户布局
 全志科技	天猫精灵、小米、小度等
 晶晨	小米、阿里巴巴、Google、Amazon等
 瑞芯微电子	阿里巴巴、百度、创维、安克创新、小米等
 恒玄科技	阿里巴巴、百度、安克创新等
 中科蓝讯	创维、天猫精灵等

资料来源:各公司年报,方正证券研究所

行业贝塔向上,头部 SoC 厂商有望充分受益。具体到智能音箱上游的 SoC 处理器,目前以全志、晶晨、瑞芯微为代表的国内厂商已经在该领域占据了较大份额,产品广泛应用于小米、小度、天猫精灵等下游终端品牌。我们认为,在 ChatGPT 引领的 AI 浪潮下,交互式体验升级的智能音箱有望突破销量增长瓶颈,实现渗透率的进一步提升,从而拉动上游处理器厂商的出货。此外,智能化升级也对芯片性能提出了更高要求,例如需要提升 SoC 中的 CPU 算力、附加专属的 AI 内核和 DSP、提供更加丰富的音视频扩展接口,以及降低功耗增加续航时间。因此,我们看好头部 SoC 厂商在行业正贝塔的环境下,通过提供性能更优的产品实现自身的阿尔法提升,巩固并扩大领先优势。

3.3 智能安防

安防芯片主要分为两类四种芯片。安防视频监控系统主要分为前端和后端两类设备，前端设备负责采集和处理原始的图像视频信号，将其转化为模拟/数字信号后再传输至后端设备进行分析、显示和存储等。回顾摄像机的发展历程：

1) 模拟监控时代：最初为模拟摄像机，其前端设备搭载的 ISP 芯片将电信号转换为模拟信号后，再通过后端的 DVR (Digital Video Recorder, 数字视频录像机) 转换为数字信号进行编码压缩和存储。

2) 网络监控时代：网络摄像机开始普及，其前端设备搭载的 IPC SoC 包含 ISP 模块和视频编码模块，原始的视频信号经过 ISP 模块处理后，需要通过视频编码模块进行压缩，而后端 NVR (Network Video Recorder, 网络视频录像机) 可以通过网络接收 IPC(网络摄像机) 传输的数字视频码流，并进行存储和管理。

前端走向主动识别，后端强化计算分析。我们看到，AI 技术对安防的赋能也是体现在前端和后端两方面：

1) 前端：通过在前端摄像机嵌入 AI 芯片，摄像机可以实现对视频数据的结构化处理，促使视频监控设备从被动监控走向主动识别。

2) 后端：通过在后端图像存储设备添加人工智能加速芯片和应用处理软件，智能网络视频录像机可以实现图像识别，强化后端设备的计算分析功能。

图表42: 安防芯片分类

	前端	后端
模拟摄像机	➤ ISP SoC芯片：处理CMOS传感器原始图像信号的降噪和曝光调整	➤ DVR SoC芯片：将CPU处理器、内存、ASIC芯片、外设接口等进行整合，集合录像机等，其核心功能是模拟音频信号数字化、编码压缩与存储
网络摄像机	➤ IPC SoC芯片：集成ISP技术和视频编解码技术，同时集成视频分析功能	➤ NVR SoC芯片：基于网络的全IP视频监控解决方案，其接收IPC的数字码流，进行集中录像存储、管理和转发
智能化摄像机	➤ IPC SoC+AI-IP/独立AI芯片：在现有IPC上集成算法实现识别任务；在SoC中集成协处理或增加独立AI芯片进行结构化分析	➤ GPU/ASIC智能服务器：将推理功能集成在边缘的服务器产品中，实现大规模的人工智能应用，如GPU或ASIC服务器方案

资料来源：头豹研究院，方正证券研究所

算力前移趋势明显，前后端共同推动智能安防落地。我们注意到，在安防发展的早期，前端的摄像机作为数据采集设备，其面临的数据量较少，对数据精度的要求也不高，因此用于图像或视频处理的算力主要部署在后端。但是随着物联网的不断发展，前端设备采集的数据量激增，数据精度也从 1080P 提高到 4K/8K，使得网络传输带宽与后端的分析处理面临很大的压力。因此：1) 一方面，通过在前端设备芯片中嵌入 AI 功能，使其可以对视频图像进行预处理，将过滤冗余信息后的关键性信息再上传至边缘侧或中心侧，由此分摊后端的计算和存储压力并提高视频分析的速度。目前适用于前端设备的 AI 芯片主要为低功耗、低成本的 ASIC 芯片，在指定的识别场景下具备高效的性能表现；2) 另一方面，后端设备芯片也需要对应进行 AI 功能的升级，前后端协同发展共同推动智能安防的落地。

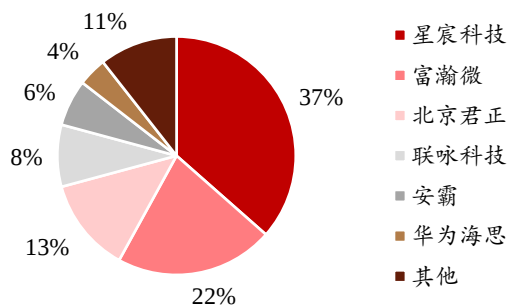
图表43:边缘 AI 在安防领域的应用



资料来源：头豹研究院，方正证券研究所

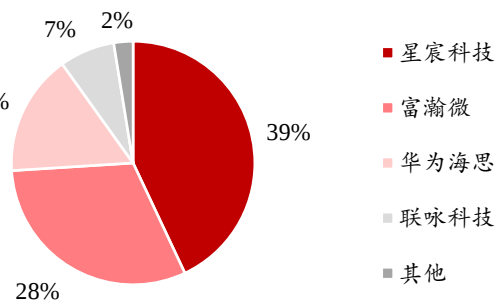
海思高端市场根基深厚，期待未来王者归来。回顾安防芯片市场的竞争格局，我们看到在 2020 年以前的绝大部分时间里，海思一直占据着主导地位，其 IPC/NVR SoC 的市场份额一度高达七成以上，这得益于海思在高清成像和音视频编解码算法上的大力投入，成功与海康、大华等头部终端客户绑定共同实现高速增长。但在海思逐渐退出后，台企联咏科技和背靠联发科的星宸科技承接了一部分中高端市场，而国内厂商富瀚微也成功抓住机遇拿下了中低端市场的大多数份额。我们认为，海思在高端安防芯片领域积累的技术和客户优势仍然有着很强的根基，未来如果回归有望延续其王者地位。

图表44:2021 年全球 IPC SoC 市场份额



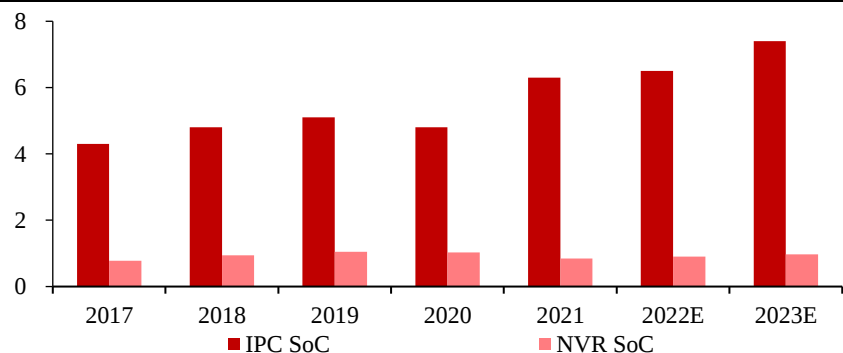
资料来源：中商产业研究院，方正证券研究所

图表45:2021 年全球 NVR SoC 市场份额



资料来源：中商产业研究院，方正证券研究所

图表46:全球 IPC/NVR SoC 市场规模（亿美元）



资料来源：中商产业研究院，方正证券研究所

3.4 自动驾驶

汽车智能化浪潮迭起，电子电气架构从分布式向域集中式演变。我们看到，在汽车智能化的发展趋势下，传统的 ECU 分布式 EEA（Electrical/Electronic Architecture，电子电气架构）已无法满足日益丰富的功能需求，因此很多功能相似但却分离的 ECU 被集成整合到一个性能更强的处理器硬件平台上，这便是汽车域控制器（Domain Control Unit，DCU）的诞生。在域集中式 EE 架构之下，主机厂会根据自身设计理念的差异对功能域做出具体的划分，其中博世的五域集中式 EEA 便被分为动力域、底盘域、车身域、座舱域和自动驾驶域，而大众的 MEB 平台以及华为的 CC 架构则在此基础上把原本的动力域、底盘域和车身域融合为整车控制域，从而形成了三域集中式 EEA，即车控域控制器、自动驾驶域控制器和智能座舱域控制器。

自动驾驶 SoC 负责决策环节，异构集成主导多种数据处理。域控制器作为每个功能域运算和决策的中心，其功能的实现依赖于主控 SoC 芯片、操作系统和应用算法等多层次软硬件的结合。在三大功能域中，驾驶域和座舱域是实现智能驾驶的技术核心，其中驾驶域负责集成传感器、定位、通讯和控制等功能，通过外接车载摄像头、激光雷达等多种传感器感知外部信息，再输入到域控制器中进行运算处理，最后输出执行命令。我们看到，在自动驾驶域控制器中的运算决策环节，核心在于主控 SoC 芯片的数据处理能力，其异构集成结构主要分为：1) 逻辑处理：由 CPU 完成通用计算；2) AI 处理：由 GPU/FPGA/ASIC 等 AI 芯片完成加速计算；3) 图像处理：由 ISP 完成图片、视频等数据的处理；4) 存储处理：由 DRAM、NAND 完成缓存并存储数据。

图表47: 自动驾驶域控制器硬件架构



资料来源：Vehicle 公众号，方正证券研究所

我们看到，华为推出的 MDC（Mobile Data Center，移动数据中心）计算平台即为自动驾驶域控制器，以 MDC 300 为例，其主控 SoC 中的 CPU 处理器为 1 颗自研的鲲鹏 920 芯片，采用 7nm 工艺，支持 12 个内核，主频 2.0GHz，算力达到 150K DMIPS；而 AI 处理器为 4 颗自研的昇腾 310 芯片，采用 12nm 工艺，基于自研的达芬奇架构，总算力达到 64 TOPS。存储方面，内存为 24GB，硬盘空间为 128GB。同时外接 15 个摄像头接口，4 个前向摄像头支持 4K 视频分辨率。

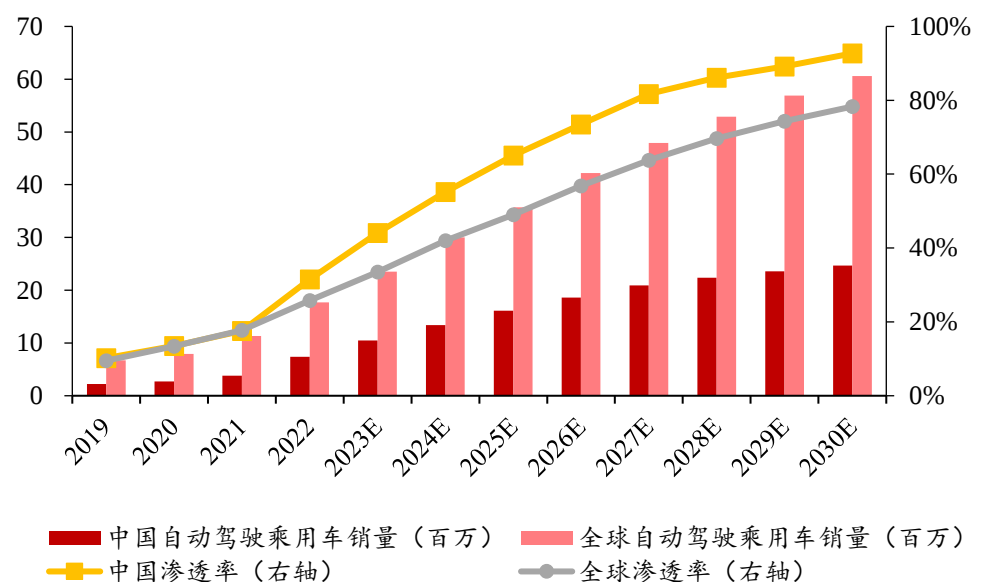
图表48: 华为 MDC 计算平台硬件方案



资料来源: Vehicle 公众号, 方正证券研究所

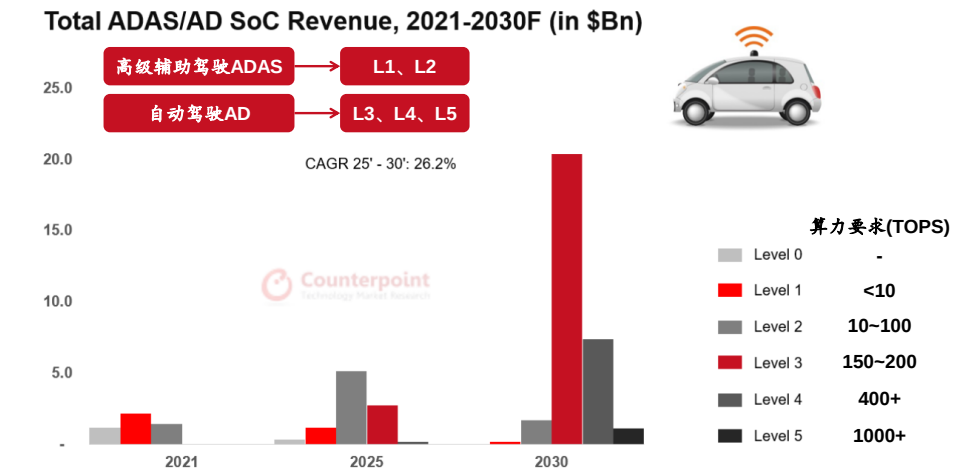
ADAS/AD SoC 市场规模或在 2030 年达到 300 亿美元。根据 Frost&Sullivan 数据, 2022 年全球及中国的自动驾驶乘用车 (L2 及以上) 销量分别达到 1770、740 万辆, 渗透率分别为 25.8%、31.5%; 预计到 2030 年销量将分别增长至 6060、2470 万辆, 渗透率分别提升至 78.3%、92.7%。我们看到, 终端销量的增长也带动了上游 SoC 芯片的需求。一般而言, 自动驾驶 SoC 主要分为适用于 L1-L2 的 ADAS SoC 和适用于 L3-L5 的 AD SoC, ADAS SoC 的技术门槛较低, 算力要求基本在 100TOPS 以下, 但 AD SoC 的技术门槛更高, L5 级别的算力要求高达 1000TOPS 以上。市场规模方面, Counterpoint 预计全球 ADAS/AD SoC 市场空间将在 2030 年达到 300 亿美元, 2022-2027 年的 CAGR 为 26.3%。此外, 在 2025 年之前, 适用于 L2 的 ADAS SoC 仍将占据主要份额, 但此后受益于自动驾驶技术的不断成熟, 适用于 L3 及以上的 AD SoC 将主导市场。

图表49: 全球及中国自动驾驶乘用车销量及渗透率



资料来源: Frost&Sullivan, 方正证券研究所; 注: 自动驾驶乘用车指具有 L2-L5 级自动化功能的车辆

图表50:不同级别自动驾驶 SoC 市场规模预测



资料来源: Counterpoint, 方正证券研究所

Mobileye 主打视觉方案, ADAS SoC 龙头地位稳固。回顾汽车芯片市场的竞争格局, 我们看到, 以前负责整车控制域的 MCU 芯片市场主要由瑞萨、恩智浦等传统厂商占据主导地位, 但随着汽车电动化、智能化的发展, 汽车架构中新涌现出的智能座舱域和智能驾驶域需要更高算力的 SoC 芯片来支持其功能的实现, 这也给后来者提供了弯道超车的机遇。就自动驾驶而言, 在早期 L1、L2 的 ADAS 发展阶段, ADAS SoC 市场基本由主打视觉方案的 Mobileye 主导, 其于 1999 年成立, 2004 年发布了第一代基于 180nm 工艺的芯片 EyeQ1, 2007 年开始量产上车, 随后分别于 2010、2014、2015 年推出了 EyeQ2、EyeQ3、EyeQ4, 其中 EyeQ4 芯片于 2018 年量产, 陆续被蔚小理、大众、宝马等厂商采用, 成为公司出货量最高的一款芯片。2017 年 3 月, Mobileye 被英特尔以 153 亿美元的价格收入麾下; 到 2021 年末, 公司芯片出货量突破 1 亿颗; 2022 年 10 月, Mobileye 正式登陆纳斯达克市场, 最新市值已超 300 亿美元。截至目前, Mobileye 的芯片已迭代至 EyeQ6 Light、EyeQ6 High 和 EyeQ Ultra, 其中 EyeQ Ultra 将于 2025 年量产, 采用 5nm 工艺, 算力提升至 176TOPS, 主打 L4/L5 高阶自动驾驶应用场景。

图表51:Mobileye 自动驾驶 SoC 产品线

芯片型号	适用功能	DL算力 (INT8, TOPS)	制程工艺	自动驾驶等级	发布时间
EyeQ4 H	入门级L2	2	28nm	L2+	2018年
EyeQ5 M	智能摄像头一体机	7	28nm	L1/L2/L2+	2021年
EyeQ5 H	360全景ADAS	16	7nm	L2+/L4	2021年
EyeQ6 L	智能摄像头一体机	5	7nm	L1/L2	2021年
EyeQ6 H	行泊一体	34	7nm	L2+/L4	2022年
EyeQ Ultra	360全景+激光雷达+毫米波雷达	176	5nm	L4/L5	2023年

资料来源: 高工智能汽车研究院, 方正证券研究所

但我们同样看到, Mobileye 软硬件捆绑的销售模式虽然在早期能够帮助大量客户实现 ADAS 功能的快速普及,但随着自动驾驶朝着 L2+ 发展,软硬件逐步解耦,整车厂开始重视在硬件芯片和软件算法层面自研能力的开发,由于 Mobileye 捆绑销售的方案导致客户既不能自行开发算法,也无法接触核心的行车数据,因此英伟达等具备完整工具开发链且足够开放的芯片平台开始受到整车厂的青睐。此外,由于 Mobileye 主打视觉方案,因此芯片算力并不高,但自动驾驶感知层面的多传感器融合方案已渐成主流,这意味着芯片需要具备更强的算力才能支持多种数据的运算处理工作,例如英伟达 2022 年量产的 Orin 芯片算力高达 254TOPS,但 Mobileye 2021 年上车的 EyeQ5 H 算力也仅有 16TOPS,仍存在较大差距。

虽然 Mobileye 在当下面临一些困难,但我们并不认为其已经失去了竞争力。首先,公司针对自身软硬件捆绑销售的模式做出了一定的改变,于 2022 年 7 月发布了首个面向 EyeQ 系统芯片的软件开发工具包 (SDK)——EyeQ Kit,该工具包将充分利用 EyeQ6 High 和 EyeQ Ultra 处理器强大的高能效架构,让车企在充分利用 Mobileye 已被验证的核心技术的同时,也能在 EyeQ 平台上部署差异化的代码和人机接口工具,这预示着公司在高阶的自动驾驶领域 (L3+) 也在尝试着走向开放式道路。其次,在低阶的 ADAS 领域 (L1/L2), Mobileye 有着深厚的技术积累,非头部车企由于自身研发能力的不足且面临着成本压力,仍倾向于选择 Mobileye 的整套解决方案。**因此我们认为, Mobileye 在中低端的 ADAS SoC 领域已基本站稳脚跟,至于能否在高端的 AD SoC 领域继续分一杯羹仍取决于自身的战略转变和研发跟进。**

图表52: 英伟达历代车载计算平台参数对比

	第一代Drive PX	第二代Drive PX2	第三代Drive PX Xavier	第四代Drive AGX Orin	第五代Drive Thor
发布时间	2015	2016	2017	2019	2022
SoC芯片	2×Tegra X1	2×Tegra X2(Parker)	1×Xavier	2×Orin	1×Thor
SoC架构	CPU+GPU	CPU+GPU	CPU+GPU+DLA+PVA	CPU+GPU+DLA+PVA	CPU+GPU
SoC制程	20nm	16nm	12nm	7nm	-
SoC算力	1 TFLOPS	4 TOPS	30 TOPS	254 TOPS	2000 TOPS
平台算力	2 TFLOPS	8 TOPS	30 TOPS	2000 TOPS	-
功耗	-	250W	30W	800W	-
搭载车型	奥迪A7	特斯拉、沃尔沃等	小鹏P7、IM智己等	蔚来ET7、理想L9、小鹏G9、威马M7、比亚迪等	极氪

资料来源: 方正证券研究所整理

聚焦到英伟达在高端 AD SoC 领域的发展历程,我们发现其从 2015 年发布首代车载计算平台 Drive PX,再到 2022 年最新推出的 Drive Thor 平台,英伟达一直在追求着极致的性能表现,引领着自动驾驶高端芯片的发展,具体来看:

1) 第一代 Drive PX

在 2015 年的 CES 消费电子展上发布,初代车载计算平台基于 Tegra X1 SoC 芯片开发,单颗芯片拥有 8 核 CPU 和 256 核 Maxwell 架构的 GPU,采用 20nm 工艺,算力在 1TFLOPS 左右,其中搭载 1 颗 Tegra X1 的 DRIVE CX 面向数字座舱,搭载 2 颗 Tegra X1 的 DRIVE PX 面向自动驾驶。此外, Tegra X1 其实是一颗移动处理器,并不是专门为汽车而设计。

2) 第二代 Drive PX2

在 2016 年的 CES 上发布，基于 2 颗 Parker SoC 芯片开发，单颗芯片拥有 6 核 CPU 和 256 核 Pascal 架构的 GPU，采用 16nm FinFET 工艺，算力达 4TOPS，成功打动了和 Mobileye 分手的特斯拉，2016 年 HW2.0 即搭载了定制版 DRIVE PX2 AutoCruise 单芯片版本，2017 年 HW2.5 升级为 AutoChauffeur 双芯片版本。

3) 第三代 Drive PX Xavier/Pegasus

Xavier 平台在 2017 年的 CES 上发布，基于 1 颗 Xavier SoC 芯片开发，单颗芯片拥有 8 核 CPU 和 512 核 Volta 架构的 GPU，采用 12nm 工艺，算力达 30TOPS。此外，Xavier 芯片还搭载了深度学习加速模块 DLA 和向量处理单元 PVA 两个专用 ASIC，DLA 用于推理，PVA 用于加速传统视觉算法，这是英伟达首次采用 CPU+GPU+ASIC 的技术路线。可以看到，Xavier 平台相比于 PX2 平台，虽然算力只有小幅提升，但面积缩小了一半，功耗仅为 PX2 的 1/8。同年，英伟达推出了性能更强的 DRIVE PX Pegasus 平台，由 2 颗 Xavier 芯片和 2 颗单独的 Turing 架构的 GPU 组成，可以实现 320TOPS 的算力。

4) 第四代 Drive AGX Orin

在 2019 年的 GTC 大会上发布，基于 2 颗 Orin SoC 芯片开发，单颗芯片拥有 8 核 CPU 和 2048 核 Ampere 架构的 GPU，采用 7nm 工艺，算力达 254TOPS，与 Xavier 芯片同为“CPU+GPU+DLA+PVA”的整体架构。此外，该平台也配备了 2 颗单独的 Ampere 架构的 GPU，平台最高算力可达 2000TOPS。

5) 第五代 Drive Thor

原本英伟达的第五代车载计算平台为 2021 年推出的 Drive Atlan，单颗 SoC 的算力能够达到 1000TOPS，相比上一代 Orin SoC 算力提升接近 4 倍，预计将在 2025 年量产上车。但 Atlan 还未量产便被淘汰，随后英伟达在 2022 年的秋季 GTC 大会上发布了新一代计算平台 Drive Thor，正式取代了 Atlan。我们看到，Thor 芯片已经不再局限于自动驾驶 SoC，而是可以在同一计算平台上实现全车自动驾驶和智能座舱功能的新一代系统级芯片，采用基于 Grace 架构的 CPU、Ada Lovelace 架构的 GPU 和 Hopper 架构的处理 Transformer 模型的引擎，由两种不同架构的 GPU 分摊 AI 运算与影像处理需求，相对 Atlan 性能翻倍，不需搭配额外的 GPU 即可达到 2000TOPS 的算力性能，预计将在 2024 年量产。极氪已宣布成为 Thor 芯片的首发搭载车企，并于 2025 年初开始生产。

回顾英伟达的产品迭代历程，我们认为 Thor 芯片的发布彰显出其在自动驾驶领域战略规划的两重要方向：**1) 使用集中式车载计算平台的单一 SoC 整合多个功能域**，因此 Thor 大幅领先市场已有芯片的算力水平凸显出其合理性，通过对众多计算资源的整合，不仅可以降低功耗和成本，也能实现功能的飞跃；**2) 主推 FP8 格式标准以打通训练与推理之间的鸿沟**，一般来说，AI 训练端的数据精度要求比较高，通常采用 FP64 或 FP32 的浮点数运算，而推理端采用 INT8 的整数定点运算，虽然精度有所下降，但成本和效率表现更好。我们看到，Thor 芯片在训练和推理端均采用 FP8 格式的浮点运算，虽然训练端的精度略有下降，但可以简化从训练到推理的过渡，可以在不牺牲精度的情况下进行数据传输，做到训练和推理的无缝衔接，大幅提升整体的计算效率和准确度。更重要的是，此举会打破专注训练或者推理单一领域的 AI 芯片厂商的竞争格局。

3.5 智能座舱

座舱主控 SoC 芯片作为座舱域的算力提供单元，主要负责座舱内海量数据的运算处理工作，主要包括多个摄像头的视频接入、多个显示屏的图像渲染和输出（GPU、DPU）、神经网络加速器 NPU、车内音频处理、蓝牙 WiFi 互联以及其他主要 ECU 的以太网数据交互等。芯片架构方面，我们以高通第三代骁龙汽车数字座舱平台的代表芯片 Qualcomm SA8155P 为例进行说明。高通 8155 芯片于 2019 年 1 月发布，作为全球首款 7nm 制程的车机芯片，首发搭载于 2021 年 1 月推出的长城 WEY 摩卡，此后便开启了上车潮，至今在国内新势力和传统车企的 30 余款新车型上都能看到 8155 的身影，主要得益于 8155 先进的异构计算架构带来的强劲性能，有力支撑了座舱的智能化演变。

图表53: 高通 8155 座舱芯片架构



资料来源：高通中国公众号，方正证券研究所

图表54: 高通座舱芯片迭代历程

	第一代620A	第二代820A	第三代SA8155P	第四代SA8295P
发布时间	2014.01	2016.01	2019.01	2021.01
制程工艺	28nm	14nm	7nm	5nm
CPU算力(K DMIPS)	-	45	105	200
GPU算力(GFLOPS)	-	320	1142	3000
NPU算力(TOPS)	-	-	8	30

资料来源：九章智驾，方正证券研究所

消费级芯片厂商立足高算力移动芯片具备性能与成本优势。回顾座舱芯片的竞争格局，在 2015 年以前，车载系统的运算和控制主要由 MCU 和低算力的 SoC 为主，瑞萨、恩智浦、德州仪器等传统汽车芯片厂商凭借自身在车规领域的深厚积累，一直在车机芯片市场占据主导地位。根据 Strategy Analytics 数据，瑞萨、恩智浦、TI 在 2016 年全球汽车 MCU/SoC 芯片市场中的份额分别达到 31%、27%、10%，CR3 高达 68%。然而在 2015 年之后，随着汽车电动化、智能化的持续推进，以高通、三星为代表的消费级芯片厂商强势入局，其优势主要体现在两方面：

- 1) 已有消费级芯片的高算力基础足以支撑未来座舱芯片的算力需求。例如目前主流的高通 8155 座舱芯片便脱胎于 2018 年发布的移动处理器骁龙 855；
- 2) 在消费电子领域积累的资金和规模优势可帮助自身在汽车芯片领域实现低成本开发。一般来说，高性能座舱芯片的开发成本动辄数亿美元，而手机等移动端芯片数千万的生命周期销量足以帮助消费电子厂商轻易收回开发成本，但汽车芯片的出货量最高也只能达到数百万片。因此，传统汽车芯片厂商在开发能力和开发成本上均处于不利地位。

传统汽车芯片厂商主打安全与可靠，目前仍占据较高份额。不过我们同样看到，安全性、可靠性和长效性是消费级芯片与汽车芯片的本质区别，因此传统汽车芯片厂商倾向于使用成熟技术来保证安全与稳定性，这也是最终的芯片算力表现不佳的原因。但凭借着安全可靠的性能，瑞萨、恩智浦和 TI 的产品仍在销量占比较大的中低端车型中广泛应用，因此传统厂商在汽车芯片市场仍占有较高的份额。不过在高端市场，高通的 8155 芯片已经占据了座舱芯片 80% 以上的份额。我们认为，虽然传统厂商凭借目前的汽车销售结构仍拥有较高的市占率，但在智能化趋势下如果不能及时推出高性能产品，其与消费级芯片厂商的差距将会被逐渐拉开。

图表55: 海外座舱芯片主要厂商产品参数对比

芯片厂商	产品型号	制程	CPU算力 (K DMIPS)	GPU算力 (GFLOPS)	NPU算力 (TOPS)	量产时间	搭载车企
消费级芯片厂商							
Qualcomm	602A	28nm	-	-	-	-	奥迪、本田、比亚迪
	820A	14nm	42	320	3	2018	蔚来、小鹏、理想、奥迪等
	SA6155P	11nm	40	430	-	-	奇瑞
	SA8155P	7nm	105	1142	4	2020	蔚来、小鹏、理想等
	SA8195P	7nm	106	1842.5	15	2021	凯迪拉克
	SA8295P	5nm	220	3100	30	2023	集度
SAMSUNG	Exynos Auto 8890	14nm	63	398	-	2018	奥迪
	Exynos Auto V910	8nm	111	1205	1.9	2021	奥迪
	Exynos Auto V920	5nm	-	-	30	2025	现代
MEDIATEK	MT2712	28nm	23	133	-	-	大众
	MT8666	12nm	56.8	113	2	-	长安、吉利
	MT8675	7nm	-	-	-	2022	-
NVIDIA	Tegra 3	40nm	-	12.4	-	2012	特斯拉
	Tegra Parker	16nm	59	500	1	-	奔驰
	Xavier NX	12nm	96	691	21	2020	奔驰
intel	A3950	14nm	42	187	-	-	特斯拉
	A3960	14nm	42	216	-	-	宝马、凯迪拉克
AMD	Ryzen	12nm	-	10	-	2021	特斯拉定制
传统汽车芯片厂商							
RENESAS	R-CAR M3	-	-	-	-	-	大众
	R-CAR H3	16nm	40	288	-	-	长城
	R-CAR H3e	-	40	280	-	-	丰田、大众、长城等
NXP	i.MX 8QM	28nm	26	128	-	2021	福特
TEXAS INSTRUMENTS	Jacinto 7	28nm	22	166.4	-	-	-

资料来源：方正证券研究所整理

海外供应商实力强劲，高通、三星领跑市场。可以看到，海外座舱芯片供应商主要分为两大阵营：

1) 以高通、三星为代表的消费级芯片厂商

a) **高通**：2014 年开始布局，同年 1 月便推出了第一代智能座舱主控 SoC 602A，但该款产品在当时并未掀起多大波澜；此后在 2016 年 1 月便推出了第二代产品 820A，凭借较大的技术进步和生态能力，从 2018 年开始陆续在小鹏、理想、蔚来、奥迪等多款车型上量产装车；而 2019 年 1 月推出的第三代旗舰产品 8155 芯片，作为全球首款 7nm 制程的座舱 SoC，几乎横扫了中高端市场；此外高通于 2021 年 1 月推出了全球首款 5nm 制程的 8295 芯片，将于 2023 年量产并首发搭载于集度汽车。可以看到，目前高通在中高端智能座舱 SoC 市场已经牢牢占据了头把交椅的地位。

b) **三星**：近几年才开始进入市场，不过其 Exynos Auto V910 芯片的整体性能已经可以与高通 8155 相媲美，成功从高通手中抢下了奥迪的订单，并且规划在 2025 年前后量产 5nm 制程的 Exynos Auto V920 芯片，其 AI 算力与 8295 相当均为 30TOPS，对高通的攻势已愈演愈烈。

c) **联发科**：分别于 2018、2019 年推出 28nm/12nm 的 MT2712/MT8666，虽然性能不及高通和三星的产品，但凭借性价比优势成功打入中低端市场，获得了大众、吉利、长安等车企的认可。此外，联发科已于 2022 年推出 7nm 制程的 MT8675 芯片，旨在进军高端市场。

“英特尔+宝马”&“英伟达+奔驰”深入绑定，AMD 成功切入特斯拉。其他消费级厂商方面，一直以来，英伟达基本与奔驰绑定，英特尔与宝马绑定，但这两家目前已基本不再迭代智能座舱芯片产品，而是主攻于自动驾驶芯片市场，后续不排除宝马、奔驰更换供应商的可能。此外我们可以看到，特斯拉座舱芯片的供应商在近十年内曾多次易主，其 2012 年推出的 Model S 的第一代智能座舱采用的是英伟达 40nm 制程的 Tegra 3 芯片，这在当时可谓是特斯拉的最好选择；随着智能座舱功能的日益增加，Tegra 3 的算力已经跟不上时代的脚步，2017 年特斯拉与英伟达分道扬镳，选择投入英特尔的怀抱，其第二代智能座舱选用了英特尔 14nm 制程的 A3950 芯片；但英特尔在此后没有继续进行算力迭代，于是特斯拉的 2022 款新车便换装了 AMD 为其定制的 12nm Ryzen 芯片，主打车载游戏性能。此外，AMD 于 2022 年 8 月 5 日宣布与吉利旗下亿咖通科技战略合作，共同打造下一代高性能智能座舱车载计算平台，将首发搭载于 Smart 品牌于 2024 年规划推出的纯电量产车型之上。

2) 以瑞萨、恩智浦、德州仪器为代表的传统汽车芯片厂商。瑞萨在 2016 年推出的 R-CAR Gen3 系列是其经典的智能座舱产品，其中 16nm 制程的 R-CAR H3 芯片专为高端汽车打造，收获了长城等客户，并在 2021 年又推出了新的 R-CAR Gen3e 系列（以 R-CAR H3e 为代表）；恩智浦目前的主流产品 i.MX 8QM 还是发布于 2013 年，量产于 2018 年，采用 28nm 制程，主要应用于一些入门级车型；德州仪器的主打产品是其 2020 年推出的 Jacinto 7 芯片系列，同样基于 28nm 制程，主要客户是日系车和德系车。整体来看，传统汽车芯片厂商的产品在芯片算力和迭代速度上均落后于消费级厂商。

图表56: 国内座舱芯片主要厂商产品参数对比

芯片厂商	产品型号	制程	CPU算力 (K DMIPS)	GPU算力 (GFLOPS)	NPU算力 (TOPS)	发布时间	搭载车企
 HUAWEI	Kirin 710A	12nm	-	-	-	-	比亚迪
	Kirin 990A	7nm	-	-	3.5	2021	北汽
 地平线 Horizon Robotics	征程2	28nm	-	-	4	2019	埃安、智己、长安等
	征程3	16nm	-	-	5	2020	理想、上汽等
	征程5	16nm	-	-	128	2021	理想、比亚迪、蔚来等
 ALLWINNER 全志科技	T3	-	-	-	-	-	一汽、小鹏
	T7	-	-	-	-	2018	长安、上汽、一汽
 AutoChips 杰发科技	AC8015	-	17	30	-	-	广汽、上汽、长安
	AC8025	28nm	60	120	1.2	2023	-
 siengine 芯擎科技	龍鷹一号	7nm	90	900	8	2021	吉利、一汽
 Rockchip 瑞芯微电子	RK3588	8nm	100	512	6	2022	-
 紫光展锐 UNISOC	A7862	12nm	54	88	-	2020	-
 芯驰 SemiDrive	X9U	16nm	100	300	1.2	2021	-

资料来源：方正证券研究所整理

国产供应商加速追赶，竞争格局尚未定型。我们看到，国内座舱芯片供应商起步较晚，目前仍处发展初期，参与厂商主要分为：

1) 消费级芯片厂商（华为、全志科技、瑞芯微等）：华为的麒麟 990A 基于 7nm 制程，已具备较高的算力表现，目前已经应用在北极狐阿尔法 S 和比亚迪的部分车型中；瑞芯微的 8nm RK3588M 芯片可支持实现 360°环视和“一芯多屏”，目前已经实现量产上车；全志科技的 T7 芯片早在 2017 年便完成开发，是国内第一颗通过车规认证的自主平台型 SoC 芯片。

2) 汽车 AI 芯片厂商（地平线、黑芝麻等）：产品可同时用于智能驾驶和智能座舱领域，其中地平线的征程 5 已迭代至 16nm，AI 算力高达 128TOPS，已应用于理想 L8 Pro。

3) 汽车芯片初创企业（杰发科技、芯擎科技、芯驰科技等）：杰发科技作为四维图新旗下成员，成立于 2013 年，专注于车规级芯片设计，其首颗 AC8015 座舱芯片已于 2021 年实现正式装车量产，高阶的 AC8025H 也即将推出；芯擎科技由吉利旗下亿咖通科技和 ARM 中国等在 2018 年共同出资成立，2021 年 12 月发布了国内首颗 7nm 智能座舱 SoC “龍鷹一号”，填补了国内 7nm 车规芯片的空白，并且已经被吉利、一汽等车企采用。此外，芯擎科技在 2022 年 7 月完成了由红杉中国领投的近十亿元的 A 轮融资，未来将继续致力于高算力汽车芯片的研发；芯驰科技成立于 2018 年，自研 X9U 芯片的 CPU 算力高达 100K DMIPS，可支持多达 10 个独立全高清显示屏。展望未来，我们看好国产座舱芯片供应商在华为等巨头的入局带领下，借助国内新能源汽车高速发展和产业链自主化的双重东风，不断增强自身实力以实现海外供应商的加速追赶乃至超越。

4 相关标的

4.1 芯原股份：IP 授权业务短期承压，在手订单保障未来成长

半导体行业景气下行渐进尾声，短期 IP 授权业务波动不改长期趋势。截止 2023Q3，公司营收 17.65 亿元，同比降低 6.34%，实现毛利 7.66 亿元，同比上升 0.19%，得益于公司一站式芯片定制业务毛利率同比提升，综合毛利率达 43.42%，同比提升 2.83pcts。此外，公司实现归母净利润为-1.34 亿元，同比下降-509.53%。整体而言，受知识产权授权业务收入波动等因素影响营业收入同比有所下降，随着公司坚持研发投入，研发费用等期间费用增加，导致净利润有所波动。

芯原是一家依托自主半导体 IP，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。公司至今已拥有高清视频、高清音频及语音、车载娱乐系统处理器、视频监控、物联网连接、智慧可穿戴、高端应用处理器、视频转码加速等多种一站式芯片定制解决方案，以及自主可控的 GPU IP、NPU IP、VPU IP、DSP IP、ISP IP 和 Display Processor IP 六类处理器 IP、1500 多个数模混合 IP 和射频 IP。2022 年实现营收 26.79 亿元，同比增长 25.23%，下游应用领域包括物联网（34%）、消费电子（22%）、工业（13%）、计算机及周边（12%）、汽车电子（10%）和数据处理（9%），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，截至 2022 年底累计半导体 IP 授权服务客户总数量近 380 家，一站式芯片定制服务客户总数量超 300 家。

在先进半导体工艺节点方面，公司已拥有 14nm/10nm/7nm/5nm FinFET 和 28nm/22nm FD-SOI 工艺节点芯片的成功流片经验。此外，根据 IPnest 在 2022 年的统计，从半导体 IP 销售收入角度，芯原是 2021 年中国大陆排名第一、全球排名第七的半导体 IP 授权服务提供商，在全球排名前七的企业中，芯原的 IP 种类排名前二。2020 和 2021 年，芯原的知识产权授权使用费收入均为全球第四。

4.2 韦尔股份：库存去化大超预期，新品放量加速业绩改善

三季度业绩环比显著改善，产品结构优化助推盈利提升。韦尔股份发布三季度业绩公告，23Q1-Q3 实现营收 150.81 亿元(yoy-2.0%)，归母净利 3.68 亿元(yoy-82.9%)，扣非归母净利 1.30 亿元(yoy-89.7%)，其中 Q3 单季度实现营收 62.23 亿元(yoy+44.4%，qoq+37.6%)，归母净利为 2.15 亿元(yoy+279.6%，环比扭亏，qoq+467.4%)，扣非归母净利 2.09 亿元(yoy+206.3%，环比扭亏，qoq+306.9%)。此外，公司半导体设计业务在三季度实现营收 54.46 亿元，占比进一步提升至 87.5%，下游需求逐步回暖。

盈利能力方面，23Q1-Q3 毛利率为 21.28%，同比-11.4pcts，其中 Q3 单季度毛利率为 21.78%，同比-6.7pcts，环比+4.5pcts，表明高成本老产品库存对利润的拖累逐渐消除，以 OV50H 为代表的低成本高价新产品陆续放量，产品结构大幅改善。23Q1-Q3 净利率为 2.39%，同比-11.5pcts，其中 Q3 单季度净利率同环比双双转正达到 3.41%，同比+6.4pcts，环比+4.5pcts。

库存去化大超预期，利润拖累逐步消除。韦尔公告显示，截至 2023Q3 末，公司库存水平已经下降至 75.52 亿元，较 2022 年末的 123.56 亿元减少了 38.9%，较 2023Q2 末的 98.28 亿元减少了 23.2%，库存去化进度大超预期。我们判断主因华

为凭借 Mate 60 系列新品强势复出，叠加下半年其他安卓品牌新机频出，消费电子产业链拉货动能充足，公司以 OV64B 为代表的老库存产品在副摄、长焦上的应用具备较强的性价比优势，需求大幅回暖带动库存持续出清，后续库存规模有望继续下降至 2023Q4 单季度营收水平。

OV50H 率先登陆小米 14 全系列主摄，光影猎人开辟移动光学新时代。我们看到，韦尔在 2023 年 1 月发布的 OV50H 正式打响了公司在智能手机主摄 50MP 大像素产品市场的第一枪，其拥有 1/1.28 英寸大底，单像素尺寸达到 1.2um，标示着公司正式进军高端市场。此外，OV50H 已应用于 2023 年 10 月 26 日发布的小米 14 全系列 50MP 主摄，同时采用全新的影像传感器品牌“光影猎人”，相比索尼 IMX989，光影猎人 900 的原生低噪声降低了 50%，动态范围提升了 8 倍，相比同尺寸的其他感光元件，功耗也降低了 42%，产品性能全方位升级，引领移动光学新时代。

高端新品陆续放量，汽车业务成长可期。我们认为，后续伴随以 OV50H 为代表，包括年底即将发布的 OV50K 在内的高端新品陆续放量，公司手机 CIS 产品结构有望大幅改善。此外，汽车 CIS 方面，公司凭借晶圆级封装具备较为明显的成本优势，产品矩阵逐步拓宽，在华为问界等多款热销车型占据较高的供应份额，同时公司积极开发海外市场，汽车业务成长动能不容忽视。我们认为，手机和汽车业务的共振有望带动公司业绩步入新一轮上行周期。

4.3 乐鑫科技：业绩稳健增长，软硬件双轮驱动

拓展产品矩阵，营收稳定增长。2023Q1-Q3 公司实现营业收入 10.27 亿元，同比增长 8.18%；实现归母净利润 0.87 亿元，同比增长 3.96%；实现扣非净利润 0.68 亿元，同比增长 6.57%。23Q3 单季度，公司实现营业收入 3.60 亿元，同比增长 7.29%；实现归母净利润 0.23 亿元，同比增长 9.85%；实现扣非净利润 0.14 亿元，同比下降 0.86%。归母净利润增长主要受营收增长、毛利率稳定、研发费用增长，且与研发费用加计扣除相关的所得税费用冲减金额增加。我们预计，随着公司近年来不断拓展产品矩阵，次新类的高性价比产品线 ESP32-C3 和高性能产品线 ESP32-S3 已于今年顺利进入了快速增长阶段，能够满足更广泛的客户应用需求，助力公司拓展新的客户与业务，最终实现整体营收的持续性增长。

乐鑫是物联网领域的专业芯片设计企业及整体解决方案供应商。公司产品以“处理+连接”为方向，“处理”以 MCU 为核心，包括 AI 计算；“连接”以无线通信为核心，目前已包括 Wi-Fi、蓝牙和 Thread、Zigbee 技术，产品边界扩大至无线通信 SoC 领域。公司以 AIoT 领域为核心，通过自研软件工具链和芯片硬件形成研发闭环，致力于成为一家物联网平台型公司，向全球所有的企业和开发者们提供一站式的 AIoT 产品和服务。

软硬件协同发展，产品矩阵逐步成型。硬件方面，ESP32-S3 芯片增加了用于加速神经网络计算和信号处理等工作的向量指令，可以实现高性能的图像识别、语音唤醒和识别等应用；ESP32-C5 是公司第一款 2.4&5GHz 双频 Wi-Fi6 产品线，是自研高频 Wi-Fi 技术上的重大突破；ESP32-H2 芯片的发布，标志着公司在 Wi-Fi 和蓝牙技术领域之外又新增了对 IEEE 802.15.4 技术的支持，进入 Thread/Zigbee 市场，进一步拓展了公司的无线通信 SoC 的产品线和技术边界。软件方面，云产品 ESP RainMaker 已形成一个完整的 AIoT 平台，集成芯片硬件、第三方语音助

手、手机 App 和云后台等一站式产品服务，可帮助客户快速实现产品智能化，缩短开发周期。

4.4 恒玄科技：季度营收创历史新高，可穿戴需求复苏启动

Q3 营收超预期，去库存效果显著。2023Q1-Q3，公司实现营业收入 15.64 亿元，同比增长 33.75%；实现归母净利润 1.18 亿元，同比下降 21.57%。23Q3 单季度来看，受益于智能可穿戴和智能家居终端市场的恢复，Q3 单季度营收创历史新高，实现营业收入 6.54 亿元，同比增长 35.67%，环比增长 24.21%；实现归母净利润 0.69 亿元，同比下降 0.79%，环比增长 37%，归母净利润增幅不及收入增幅，主要为研发投入同比增长且毛利率同比下降双重因素所致。此外，毛利率方面，由于上游成本有一定上涨，叠加芯片去库存压力，2023Q1-Q3 公司毛利率 34.84%，同比下降 4.93 个 pcts；Q3 单季度毛利率 34.47%，同比下降 5.83 个 pcts，环比下降 0.22 个 pct，环比 Q2 下降幅度收窄，已经进入相对稳定的状态。

恒玄科技主要从事智能音视频 SoC 芯片的研发、设计与销售。公司主要产品为蓝牙音频芯片、智能手表芯片和智能家居主控芯片，并基于在无线连接领域的技术积累，逐步延伸至 Wi-Fi/BT 连接芯片，产品广泛应用于智能蓝牙耳机、Wi-Fi 智能音箱、智能手表等低功耗智能音视频终端产品。2022 年实现营收 14.85 亿元，同比下降 15.89%，其中蓝牙音频类芯片营收 10.9 亿元，占比 74%；手表类芯片营收 2.9 亿元，占比 19%，营收结构逐步多元化；Wi-Fi SoC 芯片已应用于多家品牌客户的智能家电产品，Wi-Fi 4 连接芯片也开始量产落地，已应用于翻译笔、智能家电等终端产品，近一步向 AIoT 平台型公司迈进。

高强度投入研发，核心竞争力稳步提升。公司 2023 年 Q1-Q3 研发费用达到 3.67 亿元，较上年同期增长 12.23%，研发费用率高达 20.21%。此外，根据 2022 年报数据，研发人员总数 521 人，较上年同期增加 183 人，研发人员占比 85%。研发成果方面，公司基于 12nm FinFET 工艺研发的新一代 BES2700 系列可穿戴主控芯片成功实现量产上市，已应用于多家品牌客户的旗舰 TWS 耳机和智能手表产品；同时在集成 NPU 领域进行拓展，带有图形 NPU 功能的芯片研发进展顺利，极大的增强了低功耗下可穿戴设备的算力，以及对更大的神经网络算法的支持；此外，公司实现了 Wi-Fi 4 连接芯片的量产出货，支持最新 WiFi 6 的连接芯片已经顺利完成认证，已进入客户推广送样阶段；为了满足领先客户对于高品质音频和低功耗平台的双重要求，公司研发了新一代的低功耗音频功放技术，在达到优异的信噪比（SNR>120dB）的同时，能够有效的降低 50%以上的动态功耗。

4.5 中科蓝讯：业绩端表现亮眼，客户端持续渗透

需求稳步增加叠加传统旺季，Q3 单季利润同比超预期。2023Q1-Q3，公司实现营业收入 10.50 亿元，同比增长 35.19%；实现归母净利润 1.97 亿元，同比增加 66.97%；毛利率 22.98%，净利率达 18.79%。23Q3 单季度，受益于讯龙系列占比提升、现有产品更新迭代、产品种类日益丰富、供应商保障供给等因素，实现营业收入 3.96 亿元，同比增长 69.11%；实现归母净利润 0.85 亿元，同比增长 242.42%；实现扣非归母净利润 0.55 亿元，同比增长 245.29%；单季度毛利率达 24.43%，环比上升 0.09pct；Q3 净利率达 21.41%，环比上升 3.22pcts。

中科蓝讯是无线音频 SoC 芯片主要供应商。公司主要产品包括 TWS 蓝牙耳机芯片、非 TWS 蓝牙耳机芯片、蓝牙音箱芯片、智能穿戴芯片、无线麦克风芯片、数字音频芯片等，产品可广泛运用于 TWS 蓝牙耳机、颈挂式耳机、头戴式耳机、商务单边蓝牙耳机、蓝牙音箱、车载蓝牙音响、电视音响、智能可穿戴设备、物联网设备等无线互联终端。

原有产品保持迭代，研发创新拓宽应用场景。一方面，公司针对原有的无线音频 SoC 芯片保持快速迭代，前期已成功推出“蓝讯讯龙”系列高端蓝牙芯片，凭借出色的性能表现和性价比优势，目前已进入 TCL、传音、魅蓝、联想、铁三角、天猫精灵、魔声 Monster、Sudio 等终端品牌供应体系；另一方面，公司在深耕无线音频芯片领域的基础上，持续推动技术升级以优化产品结构。目前已经完成了高集成低功耗 22nm 工艺蓝牙耳机 SoC 芯片工程样片、高性能 TWS 蓝牙耳机 SoC 芯片工程样片、第一代语音控制 SoC 芯片工程样片、第一代蓝牙控制 SoC 芯片工程样片的设计工作并进入流片阶段。公司部分芯片产品已应用至智能可穿戴设备、智能家居等物联网终端产品中，进一步丰富了公司产品的应用场景。

4.6 炬芯科技：打造 AIoT 音频芯生态，端侧 AI 赋能长期发展

Q3 业绩环比改善，业务有望逐季向好。公司 Q3 单季度实现营收 1.57 亿元，同比增长 64.92%；归母净利润 0.22 亿元，同比增长 76.22%。Q3 毛利率为 45.48%，同比增加 5.49%；净利率为 14.16%，同比增加 0.90%。伴随消费电子下游市场呈现向好态势，公司在索尼、哈曼等头部品牌客户的渗透率持续提升，订单量逐渐增加，带来第三季度业绩快速度增长。

深耕高品质音质产品，水平业界领先。公司产品布局主要为蓝牙音频 SoC 芯片、便携式音视频 SoC 芯片和端侧 AI 处理器芯片等，公司头部品牌渗透率不断提升。公司自主掌握低延迟的 2.4G 无线通信私有协议设计，全链路 48K 24bit 高清音频处理，音质指标 SNR 高达 120dB，底噪低于 2uV，处于业界先进水平，端到端延迟最低低至 10ms 以下，高保真低延迟降噪技术延迟小于 3ms，产品技术处于业界领先水平。

产品性能优异，已进入众多知名终端品牌供应链。公司历经多年已积累众多知名客户资源，已进入众多终端品牌的供应链，此外，还进入三诺、奋达、通力等业界知名的 ODM、OEM 厂商的供应链体系。首批品牌客户西伯利亚、倍思、猛玛等多款产品已上市规模销售，将持续推出多款终端产品。

4.7 晶晨股份：海外市场稳步推进，产品矩阵不断拓宽

积极推进海外市场拓展，三季度盈利能力回升。2023Q1-Q3，公司实现营收 38.58 亿元，同比下降 12.32%；归母净利润为 3.14 亿元，同比下降 53.88%；扣非归母净利润为 2.67 亿元，同比下降 58.37%；毛利率 35.36%，同比-2.10pcts。单季度来看，受益于公司各产品线在海外市场均取得积极进展，营收实现 15.07 亿元，同比增长 16.60%；归母净利润为 1.29 亿元，同比增长 35.23%；扣非归母净利润为 1.09 亿元，同比增长 37.70%；实现毛利率 35.98%，同比+3.33pcts。

晶晨主要从事系统级 SoC 芯片及周边芯片的研发、设计与销售。目前主要产品包括多媒体智能终端 SoC 芯片、无线连接芯片、汽车电子芯片等，为众多消费类电

子领域提供 SoC 主控芯片和系统级解决方案。公司拥有丰富的 SoC 全流程设计经验，致力于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，实现成本、性能和功耗优化。

手握多系列芯片，客户基础牢固。1) S 系列：主要包括全高清系列芯片和超高清系列芯片，广泛应用于中兴通讯、创维、小米、阿里巴巴、Google、Amazon 等客户的机顶盒中；2) T 系列：智能显示终端核心关键部件，成功斩获小米、海尔、TCL、海信、长虹、Seewo（希沃）等客户；3) A 系列：内置最高算力为 5TOPS 的 NPU，支持最高 1600 万像素高动态范围影像输入和超高清编码，广泛应用于小米、TCL、阿里巴巴、Harman Kardon、Zoom 等客户的智能终端产品；4) W 系列：作为公司自主研发的高速数传 Wi-Fi 蓝牙二合一集成芯片，可应用于高吞吐视频传输，第二代产品已于 2022 年 12 月预量产，即将进入商业化阶段，将进一步驱动公司无线连接芯片业务进入新的增长通道。

4.8 全志科技：多元化产品布局，拥抱 AI 创新浪潮

伴随下游需求复苏，公司 Q3 业绩回暖。2023Q1-Q3，公司实现收入 1.21 亿元，同比-4.29%；归母净利润-0.21 亿元，同比-109.26%；毛利率 32.29%，同比-6.75pcts。单季度看，23Q3 公司实现收入 4.45 亿元，同比增长 31.32%，环比+1.89%；实现归母净利润-0.04 亿元，同比-118.78%，环比-114.58%；毛利率 33.46%，同比-0.74pct。我们认为，伴随终端客户库存逐步去化，行业陆续回暖叠加低价晶圆成本逐步体现，公司业绩有望企稳回升。

全志主要从事智能应用处理器 SoC、高性能模拟器件和无线互联芯片的研发与设计。公司通过以 SoC、PMU、WIFI、ADC 等芯片产品组成的套片组合为基础，结合智能技术服务平台的支持，为客户提供优质低成本智能芯片及解决方案。2022 年实现营收 15.14 亿元，同比减少 26.69%，主要产品包括智能终端应用处理器芯片（79%）、智能电源管理芯片（10%）、无线通信产品（9%）等。

多元化产品布局，全面 AI 赋能落地。公司通过 AI 全面赋能，聚焦 AI 语音、AI 视觉应用的完整链条，实现智能音箱、智能家电、智能安防、智能座舱、智能工控等细分 AI 产品量产落地：1) 智能音箱市场，公司与行业头部一线标杆客户保持产业深度合作，R 系列芯片产品已实现带屏、无屏音箱全面量产；2) 智能清洁机器人市场，面向扫地机产品呈现的无感清洁/功能复合/智能升级/体验升级/场景多元等趋势，公司推出了面向中高端扫地机的 MR 系列芯片新品；3) 智能视觉市场，公司已成功在多家行业头部客户中完成新一代视觉芯片 V853 的全面落地量产；4) 智能汽车电子市场，公司针对智能车载人机交互需求，发布 T113 芯片产品及解决方案，已在车载人机交互和仪表类应用落地；5) 智能显示市场，推出智慧屏芯片 TV303，后续将逐步在智能电视、智能投影、智能商显领域投入量产。

4.9 泰凌微：无线物联网 SoC 领军者，纵深产品线把握行业红利

Q3 营收大幅增长，景气度向上拐点已显现。受益于持续开拓市场，公司实现出货量、销售收入的大幅增长，于前三季度实现营业收入 4.76 亿元，同比增长 9.29%；归母净利润同比大幅增长 72.85%至 3758.44 万元，第三季度公司的营业收入大幅

增长 44.67%。同时由于产品结构改善以及成本优化，前三季度毛利率 43.78%，同比增长 3.17%，整体来看，出货数量与盈利水平均呈现良好弹性。

重视自主创新，低功耗蓝牙连接芯片类产品性能国际领先。凭借在蓝牙领域的突出贡献及行业地位，公司 2019 年 7 月获选为国际蓝牙技术联盟（SIG）董事会成员公司，与同为成员公司的国际知名科技公司苹果、爱立信、英特尔、微软、摩托罗拉移动、诺基亚和东芝一起负责蓝牙技术联盟的管理和运营决策；公司副总经理、核心技术人员金海鹏博士被聘请为 SIG 董事会联盟成员董事，深度参与国际蓝牙标准的制定与规范，积极推动蓝牙技术的发展。

终端客户资源丰富，与大品牌深度绑定。公司与多家行业领先的手机及周边、电脑及周边、遥控器、家居照明等厂商或其代工厂商形成了稳定的合作关系，产品广泛应用于汉朔、小米、罗技、欧之、涂鸦智能、朗德万、瑞萨、科大讯飞、创维、夏普、松下、英伟达、哈曼等多家主流终端知名品牌。

4.10 瑞芯微：行业回暖带动业绩回升，智能座舱芯片放量在即

三季度行业边际回暖，公司收入利润同环比双增。2023 年以来全球电子行业需求萎靡，整体呈现负增长态势，2023Q1-Q3 公司实现收入 14.55 亿元，同比-7.37%；实现归母净利润 0.77 亿元，同比-71.99%；实现毛利率 34.81%，同比-3.66pcts；净利率为 5.32%，同比-12.26pcts。三季度以来行业边际回暖，公司长期布局的 AIoT 业务迎来高增长；同时，汽车电子智能座舱进入量产阶段，导致三季度业绩有所改善，Q3 单季度实现营收 6.02 亿元，同比增加 83.26%，环比增加 15.0%；毛利率为 36.08%，同比-0.96pct，环比+1.86pcts；公司净利率为 8.73%，同比+7.60pcts，环比+0.47pcts。

瑞芯微是领先的物联网（IoT）及人工智能物联网（AIoT）处理器芯片企业。公司产品为智能应用处理器芯片，按功能侧重方向可以分为通用处理器、机器视觉处理器、车载处理器、工业控制处理器等。公司以不同算力层次的智能应用处理器芯片和不同性能层次的传统通用芯片，充分契合不同终端产品的市场定位，提供更具针对性和性价比的芯片产品和解决方案，广泛应用于日益增长的 AIoT 市场。此外，公司产品还包括数模混合芯片、接口转换芯片、无线连接芯片及与自研芯片相关的模组产品等。

公司围绕 RK3588 软硬件方案持续进行优化，此外还成功推出了 AIoT 通用算力平台 RK3562、流媒体处理器 RK3528、机器视觉处理器 RV1106/RV1103 等三个新一代 SoC：1) **RK3588**：公司最新一代高制程、高性能、通用性的旗舰 SoC 芯片，是公司技术的集大成者，也是目前国内市场同类产品的最高水平之一，目标应用场景涵盖智能座舱、大屏设备、边缘计算、多目摄像头、NVR（网络视频录像机）、高性能平板、ARM PC 及 AR/VR 等领域，目前已经在智能座舱领域成功实现量产，从产品发布到上车量产仅用了一年半时间；2) **RK3562**：搭载了最新一代 NPU，神经网络计算效率大幅提升，尤其是对 Transformer 的支持，主要应用于平板电脑、智能家居、教育电子、工业应用等领域，契合客户对中等神经网络算力、高性价比 AIoT 芯片的需求；3) **RK3528**：针对智能 IPTV/OTT 和中高端多媒体应用研发的新一代流媒体处理器，重点优化了视频编解码器、图像显示、图像后处理等关键技术；4) **RV1106/RV1103**：针对轻量级智能需求推出的机器视觉处理器，适用于 IPC、智能家居、汽车电子、会议设备等领域的应用。

5 风险提示

- 1) **研发进展不及预期。**边缘 AI 的发展对终端 AI SoC 的性能提出了更高要求，相关厂商也在加大新品的研发力度，但如果研发进展不及预期，可能会影响企业的产业布局和业绩增长。
- 2) **产业进度不及预期。**当前生成式 AI 快速发展，国内大厂纷纷推出各自的大模型，但就现阶段而言，每家大模型的应用效果大同小异，都是基于已有的开源模型进行训练，后续如果无法得到更深层次的迭代优化，将会影响下游的产业应用进度，进一步抑制 AI SoC 的需求增长。
- 3) **中美贸易摩擦加剧：**若中美关系趋于紧张，相关制裁风险将不利于行业内公司的发展。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，保证报告所采用的数据和信息均来自公开合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法存在局限性。特此声明。

免责声明

本研究报告由方正证券制作及在中国（香港和澳门特别行政区、台湾省除外）发布。根据《证券期货投资者适当性管理办法》，本报告内容仅供我公司适当性评级为 C3 及以上等级的投资者使用，本公司不会因接收人收到本报告而视其为本公司的当然客户。若您并非前述等级的投资者，为保证服务质量、控制风险，请勿订阅本报告中的信息，本资料难以设置访问权限，若给您造成不便，敬请谅解。

在任何情况下，本报告的内容不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求，方正证券不对任何人因使用本报告所载任何内容所引致的任何损失负任何责任，投资者需自行承担风险。

本报告版权仅为方正证券所有，本公司对本报告保留一切法律权利。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。如需引用、刊发或转载本报告，需注明出处且不得进行任何有悖原意的引用、删节和修改。

评级说明：

类别	评级	说明
公司评级	强烈推荐	分析师预测未来12个月内相对同期基准指数有20%以上的涨幅。
	推荐	分析师预测未来12个月内相对同期基准指数有10%以上的涨幅。
	中性	分析师预测未来12个月内相对同期基准指数在-10%和10%之间波动。
	减持	分析师预测未来12个月内相对同期基准指数有10%以上的跌幅。
行业评级	推荐	分析师预测未来12个月内行业表现强于同期基准指数。
	中性	分析师预测未来12个月内行业表现与同期基准指数持平。
	减持	分析师预测未来12个月内行业表现弱于同期基准指数。
基准指数说明		A股市场以沪深300 指数为基准；香港市场以恒生指数为基准，美股市场以标普500指数为基准。

方正证券研究所联系方式：

北京：西城区展览馆路 48 号新联写字楼 6 层

上海：静安区延平路71号延平大厦2楼

深圳：福田区竹子林紫竹七道光银行大厦31层

广州：天河区兴盛路12号楼隼峰苑2期3层方正证券

长沙：天心区湘江中路二段36号华远国际中心37层

网址：<https://www.foundersc.com>

E-mail：yjzx@foundersc.com