

AI+汽车智能化系列之三—— 充分重视OEM自研智驾芯片的长期意义

汽车行业证券分析师：黄细里
执业编号：S0600520010001
联系邮箱：huanxl@dwzq.com.cn
联系电话：021-60199790

汽车行业证券分析师：杨惠冰
执业编号：S0600523070004
联系邮箱：yanghb@dwzq.com.cn

2024年4月22日

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

■ 当我们在谈自研智驾芯片时，我们究竟在谈什么？【设计芯片IP核+开发适配底软/工具链】

- 芯片按类可分为计算、存储、信号转换以及片上集成SoC四大类，**AI芯片是指在SoC基础上针对人工智能算法做特殊加速处理的芯片**。智驾领域AI芯片主要用于云端/边缘端两种场景：1) 用于智驾边缘端应用的AI芯片一般涵盖AI计算单元NPU、CPU\GPU\ISP\IO接口等必要组成部分，更强调各IP核之间的综合协调能力；2) 用于云端训练应用的AI芯片则更加强调NPU\GPU的计算能力，对于功耗、各部分间协调等要求较低。
- OEM及三方供应商自研智驾芯片多指：**自身设计SoC系统中NPU/ISP等核心IP核**，外采EDA软件形成逻辑电路，并由其他厂商完成制造以及封装环节；同时为更好调用芯片算力，**玩家需适配性开发底软（计算架构）以及SDK工具链**，便于编辑落地上层应用。
- 为进一步强化智驾“数据闭环”对于软硬件迭代效率的意义，**少部分玩家或将自研云端超算芯片**。

■ OEM自研设计AI智驾芯片必要性以及可行性如何？【边缘端芯片必要性及可行性强】

- **必要性：自研边缘端芯片有足够性价比，云端芯片短期必要性较低**。智能驾驶产品力的竞争**短期看产品体验，中期看迭代效率，长期看降本能力**；边缘端芯片自研有效影响中期软件算法相对成熟后的迭代效率（软件能否充分发挥芯片算力），并直接决定长期智驾全系统降本能力，因此强势OEM当前投资芯片自研在未来3~5年内有足够超额回报，有望形成正循环。云端芯片短期性能要求单一，仅针对AI算力，中长期影响软硬件提升速率，但前期投入较大，当前性价比较低。
- **可行性：OEM玩家自研边缘段智驾芯片可行性较强**。参照地平线、黑芝麻智能发展历程，从团队规模、资金投入以及研发耗时三重角度分析，**千人研发规模；30~50亿研发投入；2~3年耗时**可支持智驾芯片全自研以及配套解决方案落地；特斯拉2016年启动智驾芯片项目，2019年正式搭载上车，国内强势OEM自研芯片以及配套底软具备相当可行性。

■ 第三方Tier玩家自研智驾芯片以及底软，打法及成效如何？【高举高打最强音&自下而上差异化】

➤ **第一类：英伟达/华为，云端&边缘端软硬件全覆盖。**

➤ **1) 英伟达：高举高打，打造硬件算力&软件生态最强音。** 公司依托全球绝对领先GPU芯片&CUDA异构计算架构，软硬件配合构筑高壁垒，汽车为其下游重要终端应用场景。以Hopper架构赋能的DGX高性能芯片布局超算中心，自研DPU芯片支持云端大规模数据传输，配合基于CUDA的高性能算子库和SDK工具包，支持数据训练+图形渲染+仿真模拟等，并通过GPU+Grace CPU组合形成SoC芯片，更好裁剪落地云端算法解决方案。

➤ **2) 华为：全面对标英伟达，赋能车企培育生态。** 硬件端，华为以昇腾310/910为基础分别聚焦推理/训练环节，310系列配合华为自研激光雷达等传感器形成完整车身解决方案，910 NPU配合鲲鹏系列CPU打造Atlas云端服务器，提供最大20PFLOPS的解决方案；软件端，华为对标英伟达CUDA开发CANN计算架构，盘古大模型赋能，MindStudio工具链支持完善第三方应用。软硬件成套配合赋能国内弱势OEM，更好培育自身智驾生态。

➤ **第二类：高通/Mobileye/地平线，聚焦边缘端软硬件，自下而上差异化布局。**

➤ **1) 高通：**边缘端智驾芯片&开发工具链全自研，发挥基盘业务优势自研全芯片IP核，舱驾一体差异化向上突破，国内市场联合创达/毫末/大疆等Tier1迅速入局，补足生态短板；

➤ **2) Mobileye：**依托L2智驾开发积累，由封闭黑盒逐步开放，SDK套件开发完善，聚焦低成本高效能视觉方案，国内联合经纬恒润加速发展；

➤ **3) 地平线：**芯片架构持续优化，征程系列产品以自研BPU AI计算核心，OpenExplorer算法工具链为支撑，以相对“低姿态”赋能国内OEM股东，协同进步。

■ 特斯拉自研智驾云边芯片，国内OEM举旗跟进，布局智驾硬件。

- **特斯拉全栈自研FSD智驾芯片**，底层算法更好适配调用ASIC芯片算力，实现双芯144TOPS算力即可对标英伟达双芯508TOPS算力的智驾功能，同时根据自身软件能力迭代持续优化硬件架构，保障行业领先。另外自研D1芯片支撑云端Dojo超算中心，强化AI计算+传输带宽，AI算力全球领先；并自研训练软件栈，支持通用性计算语言的同时实现对神经网络模型的自动调优和并行化。
- **国内OEM举旗跟进自研**。第一类：以头部新势力为代表，智驾边缘端芯片全栈自研，蔚来对标英伟达Orin智驾芯片已发布；小鹏/理想积极布局，预计2025~2026年亮相；第二类：主流车企以战投合作形式展开，吉利亿咖通以及多OEM战投地平线，进行产业链布局。

■ 投资建议：汽车AI智能化转型大势所趋，硬件为基石，看好布局智驾硬件的OEM长期竞争力。

- **全行业加速智能化转型，产业趋势明确**。下游OEM玩家+中游Tier供应商以及上游原材料厂家均加大对汽车智能化投入，大势所趋；智驾核心环节【软件+硬件+数据】均围绕下游OEM展开，数据催化算法提效进而驱动硬件迭代。因此，以AI芯片为核心的智驾硬件是OEM中长期核心竞争力的重要构成，参考手机行业，核心硬件是玩家【成本控制能力+品牌护城河】的终局竞争要素。
- **国内OEM以软件为先，硬件其次，加速进化**。头部新势力玩家紧随特斯拉引领本轮智驾技术变革，全自研智驾芯片有望于2025~2026年流片量产，构筑品牌核心竞争力以及产品重要卖点。
- **看好智驾头部车企以及智能化增量零部件**：1) 华为系玩家【长安汽车+赛力斯+江淮汽车】，关注【北汽蓝谷】；2) 头部新势力【小鹏汽车+理想汽车】；3) 加速转型【吉利汽车+上汽集团+长城汽车+广汽集团】；4) **智能化核心增量零部件**：域控制器（德赛西威+经纬恒润+华阳集团+均胜电子等）+线控底盘（伯特利+耐世特+拓普集团等）。

■ **风险提示**：智能驾驶相关技术迭代/产业政策出台低于预期；华为/小鹏等车企新车销量低于预期。



■ 一、如何看待OEM自研智驾芯片？

■ 二、第三方玩家自研智驾芯片成效如何？

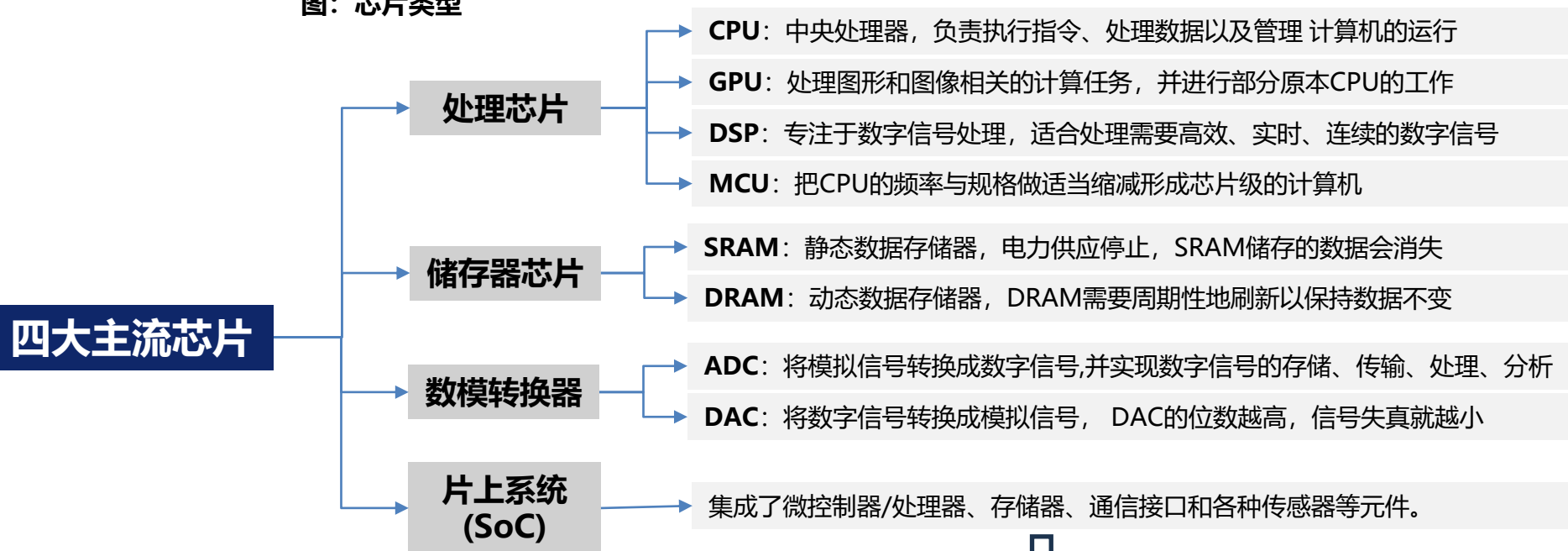
■ 三、下游OEM玩家如何做？

■ 四、投资建议与风险提示

一、如何看待OEM自研智驾芯片？

- 当前市场上流通的主流芯片包括四大类：**1) 处理器芯片**，包括CPU、GPU、DSP、和MCU，负责系统的运算和控制核心，以及信息处理和程序运行的最终执行单元。**2) 存储器芯片**：包括静态（SRAM）以及动态（DRAM）随机存取存储器等，用于数据的存储。**3) 模拟-数字转换器 (ADC) 和 数字-模拟转换器 (DAC)**：这两种芯片分别用于模拟信号和数字信号的互相转换，广泛应用于传感器和测量仪器中。**4) 片上系统(SoC)**：集成微控制器/处理器/存储器/通信接口和传感器等元件，通过简单编程可以实现丰富的功能。
- **AI芯片是属于SoC片上系统芯片的特殊分支**，是指针对人工智能算法做了特殊加速设计的芯片，专门用于处理人工智能应用中的大量计算。

图：芯片类型



■ 为满足行业发展对于芯片处理性质单一但规模庞大的数据计算的需求，产业基于GPU图像处理器的并行计算能力持续升级，开发了以极致性能为代表的**GPU**以及以极致功耗为代表的**ASIC**芯片，以及介于二者之间，兼具灵活性和高性能的**FPGA**等不同类型芯片，应用于包括云端训练以及边缘段推理等不同场景。未来，AI芯片将持续迭代，开发高度模拟人脑计算原理的类脑芯片，围绕人脑的神经元/脉冲等环节，实现计算能力的飞跃提升以及能耗的大幅下降。

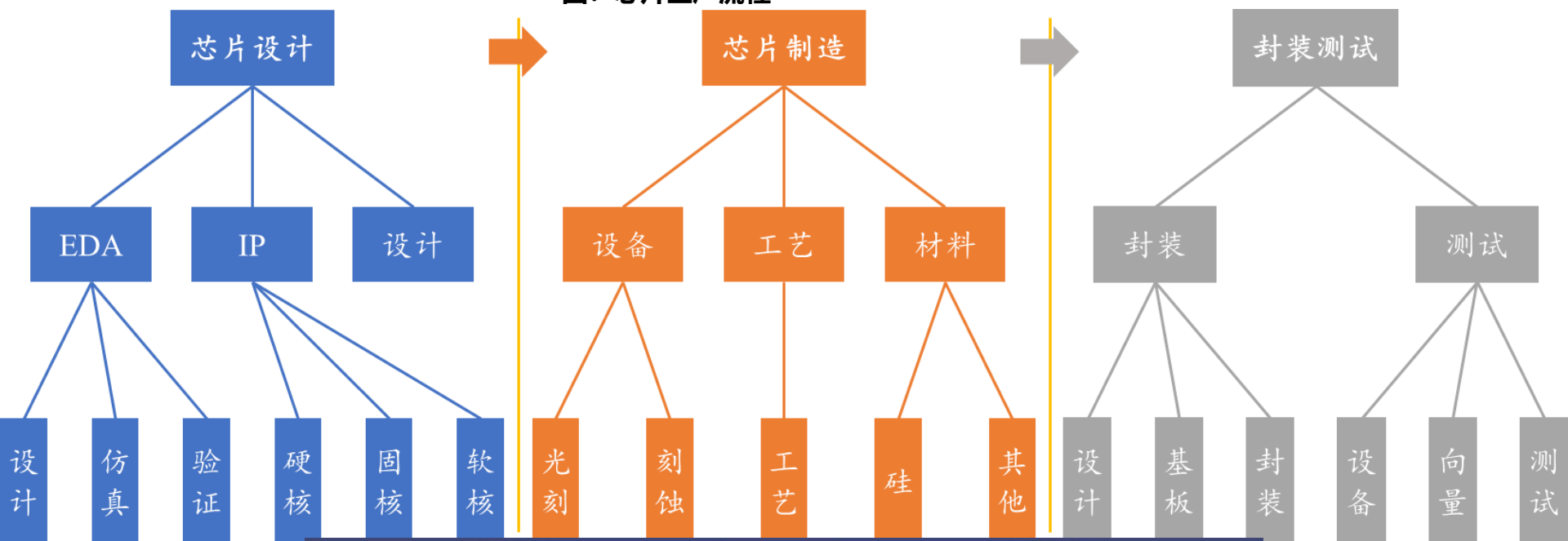
图：芯片按架构和场景分类



■ 芯片制造分为三大步骤，分别是芯片设计、芯片制造、封装测试

- **芯片设计：**在EDA软件工具的支持下，通过购买授权+自主开发获得IP，遵循集成电路设计仿真验证流程，完成芯片设计。首先明确芯片目的（逻辑/储存/功率），编写芯片细节，形成完整HDL代码；其次利用EDA软件（高制程工艺软件市场集中度高）将HDL代码转为逻辑电路图，进一步转为物理电路图，最后制作成光掩模。
- **芯片制造：**壁垒最高！三大关键工序光刻、刻蚀、沉积，在生产过程中不断重复循环三工序，最终制造出合格的芯片。过程中要用到三种关键设备，分别是光刻机、刻蚀机、薄膜沉积设备。
- **封装测试：**测试是指在半导体制造的过程中对芯片进行严格的检测和测试，以确保芯片的质量和稳定性和性能；而封装则是将测试完成的芯片进行封装，以便其被应用在各种设备中。

图：芯片生产流程



- **EDA: (Electronic Design Automation)电子设计自动化**，常指代用于电子设计的软件。目前，Synopsys、Cadence和Mentor (Siemens EDA) 占据着90%以上的市场份额。在10纳米以下的高端芯片设计上，其占有率甚至高达100%。国产EDA工具当前距离海外龙头有较大差距。
- **IP核：指一种事先定义、经过验证的、可以重复使用，能完成特定功能的模块（类似于excel模板）**，物理层面是指构成大规模集成电路的基础单元，SoC甚至可以说是基于IP核的复用技术。其包括处理器IP（CPU/GPU/NPU/VPU/DSP/ISP...）、接口IP（USB/SATA/HDMI...）、存储器IP等等几类。对于当前智驾领域AI芯片而言，常用IP核包括**CPU、GPU、ISP、NPU、内存控制器、对外接口（以太网【用于连接不同车身设备以交换数据】和PCIe接口【用于主板上的设备间通讯】）**等。

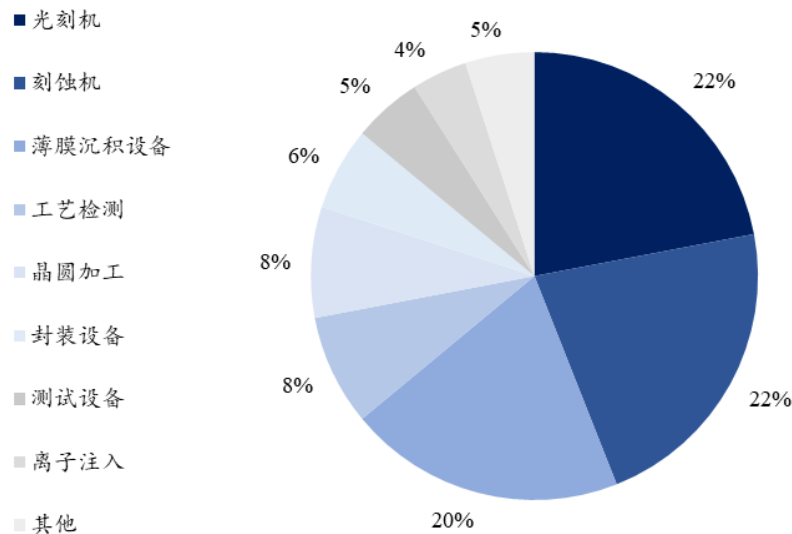
图：全球不同厂家芯片IP销售额以及市场规模/百万美元

Rank	Company	2021	2022	Growth	2022份额
1	ARM (Softbank)	2202.1	2741.9	24.5%	41.1%
2	Synopsys	1076.6	1314.8	22.1%	19.7%
3	Cadence	315.3	357.8	13.5%	5.4%
4	Imagination Technologies	153.0	188.4	23.1%	2.8%
5	Alphawave	89.9	175.0	94.7%	2.6%
6	Ceva	122.7	134.7	9.8%	2.0%
7	Verisilicon	109.4	133.6	22.1%	2.0%
8	SST	102.9	122.0	18.6%	1.8%
9	eMemory Technology	84.8	105.1	23.9%	1.6%
10	Rambus	47.7	87.9	84.3%	1.3%
Top 10 Vendors		4304.4	5361.2	24.6%	80.3%
Others		1217.7	1316.0	8.1%	19.7%
Total		5522.1	6677.2	20.9%	100.0%

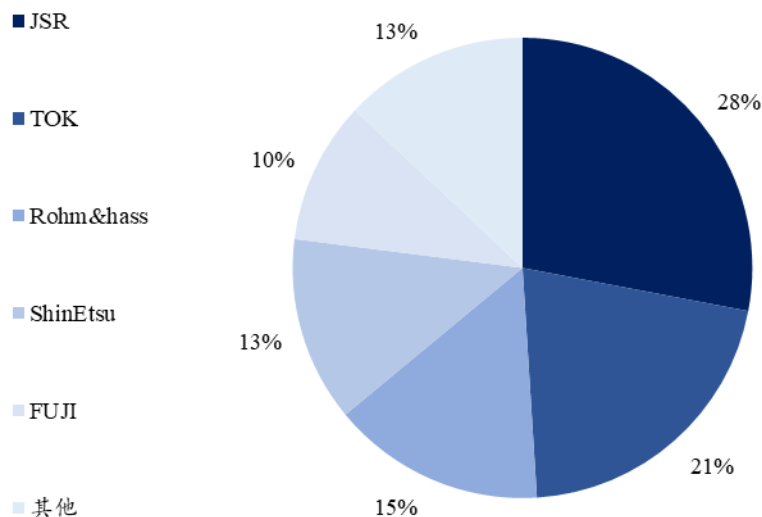
制造环节：设备/工艺/材料多环节，高壁垒高集中度

- 芯片制造三大关键工序：光刻、刻蚀、沉积，三大工序在生产过程中不断循环，最终制造出合格的芯片；其中，设备+工艺+材料等环节尤为关键；**芯片制造以台积电、三星、英特尔寡头垄断。**
- **设备**：三大关键工序要用到光刻机、刻蚀机、薄膜沉积设备三种关键设备，占有所有设备投入的22%、22%、20%左右，是三种难度和壁垒最高的半导体设备。
- **工艺**：芯片制造需要2000道以上工艺制程，主要包括光刻、刻蚀、化学气相沉积、物理气相沉积、离子植入、化学机械研磨、清洗、晶片切割等8道核心工艺。
- **材料**：硅晶圆和光刻胶是最核心的两类材料，90%以上的芯片在硅晶圆上制造，光刻胶是制造过程最重要的耗材，半导体光刻胶壁垒最高，全球CR5接近90%。

图：不同半导体设备占有所有设备投入的比例



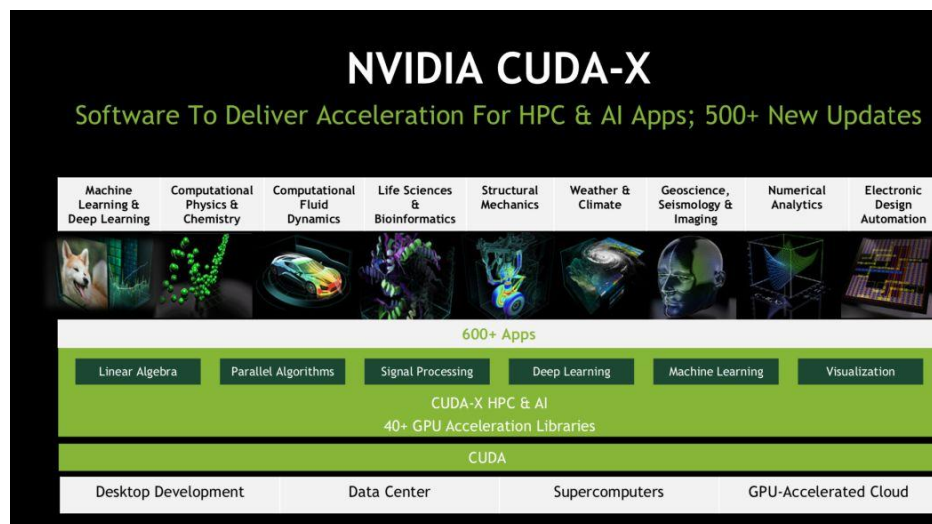
图：2022年全球光刻胶企业市场份额



底软以及工具链开发是自研智驾芯片的后端壁垒

- **异构计算架构/生态开发环境**：以英伟达CUDA和华为CANN为代表的核心软件层，用于调度AI芯片和通用芯片的底层算子，并针对性地进行加速和执行，更好地发挥出芯片的算力，实现效率最大化。
- **SDK软件开发工具包**（Software Development Kit）：是指软件工程师为特定的软件包、软件框架、硬件平台、操作系统等建立应用软件时的开发工具的集合；借助SDK，应用开发者可以迅速基于特定平台开发差异化上层应用。

图：华为和英伟达底层软件架构

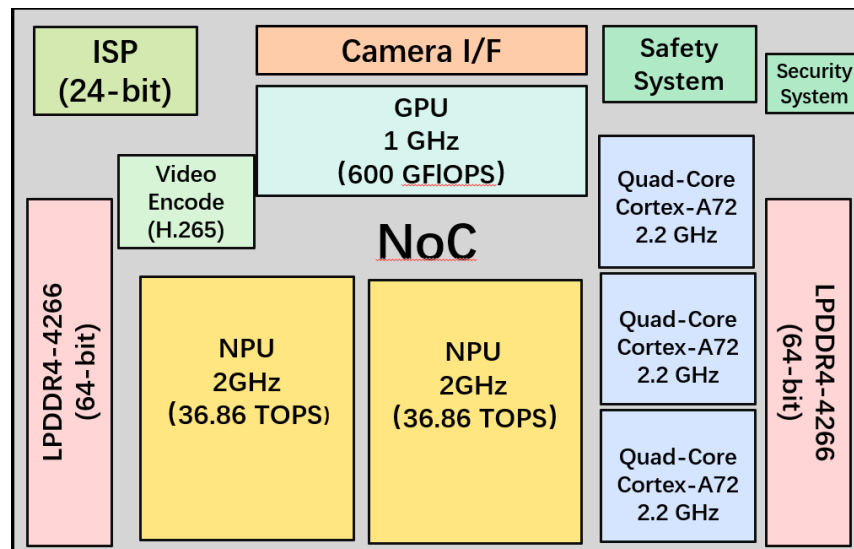


■ 智驾边缘端芯片以自研NPU为主，塑造产品差异化。

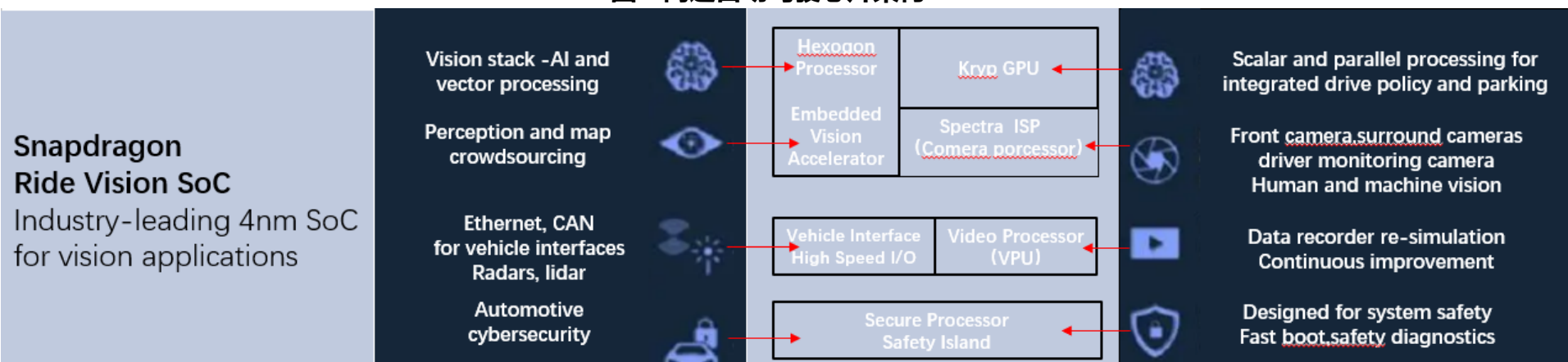
智驾SoC芯片以CPU中央处理器+GPU图形处理器+DSP数字信号处理器+ISP图片处理器+NPU(AI计算单元)以及I/O接口以及存储器等IP核集成组装而成，其中NPU/CPU/ISP等环节对智驾边缘段数据处理更为重要。**产业链玩家自研智驾芯片即指芯片自主设计IP核，尤其是NPU，其次ISP等**，CPU以及GPU多以外采ARM/英伟达等为主，技术相对成熟，其余I/O接口以及存储器同样依赖外部采购。

■ **云端芯片多采用集中外采形式**，主要系云端芯片对于能耗以及CPU/GPU综合能力要求较低，仅对强AI算力也即单一GPU/NPU的计算能力有较高需求，规模效应是核心优势，外部方案更成熟。

图：特斯拉FSD芯片



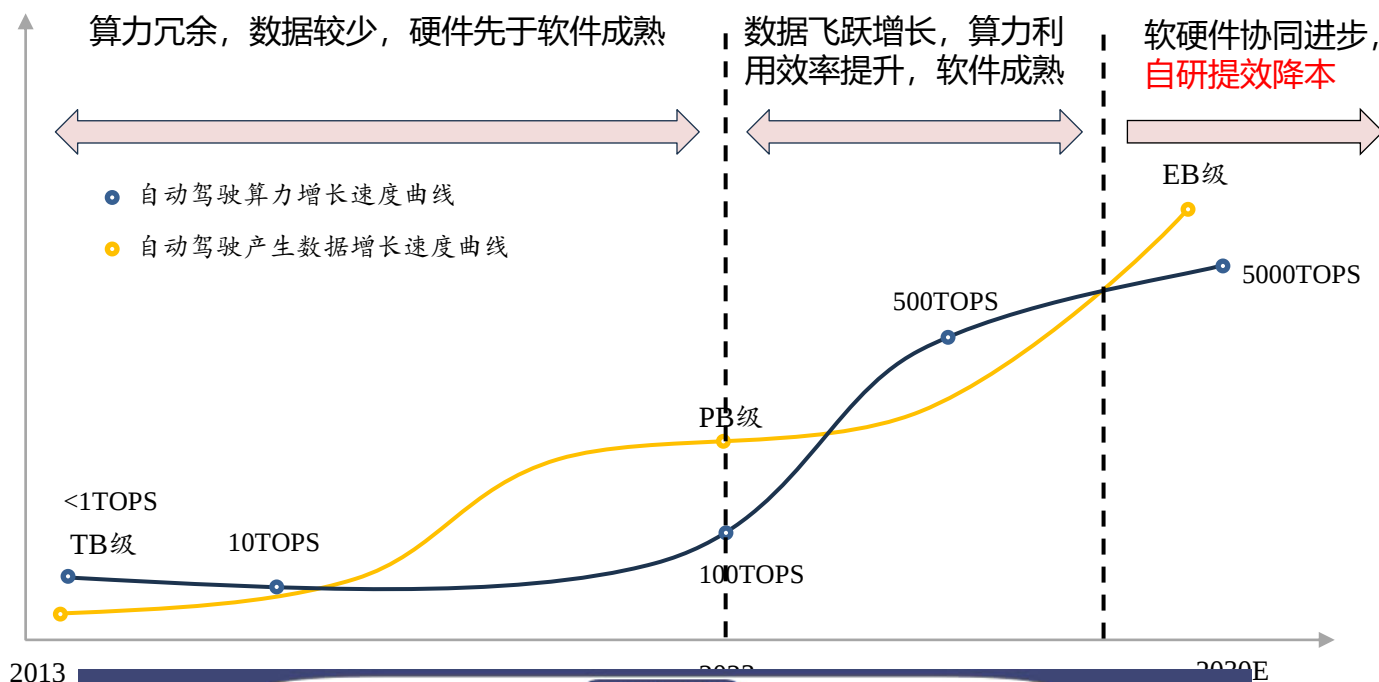
图：高通自动驾驶芯片架构



智能驾驶产品力的竞争短期看产品体验，中期看迭代效率，长期看降本能力。

- 1) 短期——算力冗余阶段：产品体验取决于软件算法成熟度（背后是数据量为支撑），与智驾芯片自研相关性较低，高通/英伟达/华为/地平线等多家第三方供应商产品均可满足。
- 2) 中期——算力提效阶段：在保有量提升带动数据飞跃增长后，前期冗余布局的边缘端硬件的利用效率进一步提升，同时也对底软更好地调用芯片算力提出更高要求，自研芯片NPU/ISP等核心环节的优势显现，迭代速率更快。
- 3) 长期——协同并进阶段：足量数据喂养下软硬件能力协同提升，保障功能体验的同时优化成本结构，要求玩家对底层硬件具备全栈深入了解。

图：随时间推移，智驾所需算力以及数据量持续增加

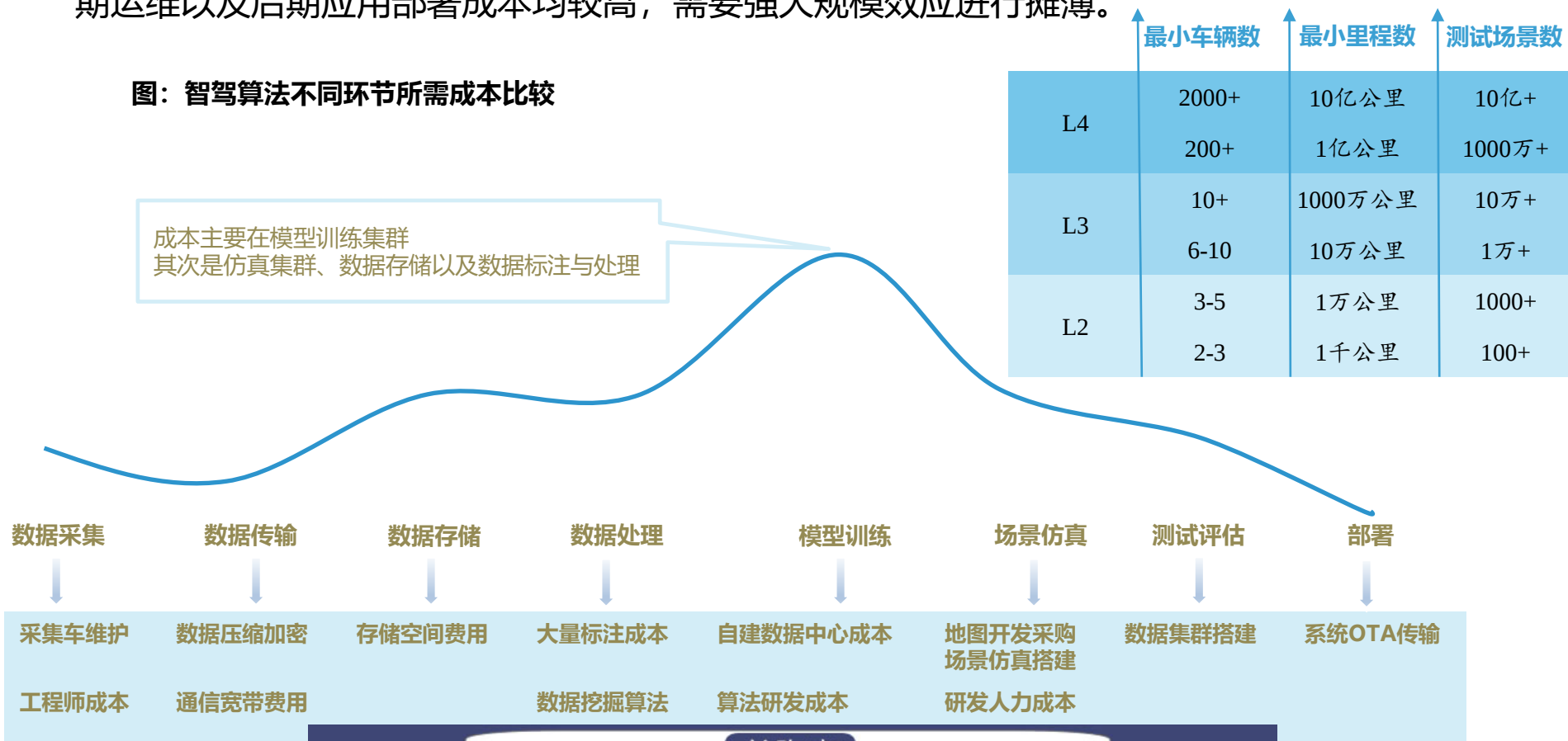


■ 云端芯片自研有利于数据全流程闭环，提升数据利用率和算法迭代速率，但同时成本负担较大。

➤ 智驾数据量指数级增长驱动智驾功能升级，数据的存储、优化、利用、训练等各环节对云端训练/传输等要求较高，“数据驱动”的智驾迭代模式下，数据闭环的模型训练与AI计算平台相互赋能，同时提升多元异构数据的清洗和标注效率，有利于提升算法迭代升级速率。

➤ 云端超算中心芯片与边缘端芯片不同，其能力依赖GPU/NPU等的单一计算能力，前期研发和中期运维以及后期应用部署成本均较高，需要强大规模效应进行摊薄。

图：智驾算法不同环节所需成本比较

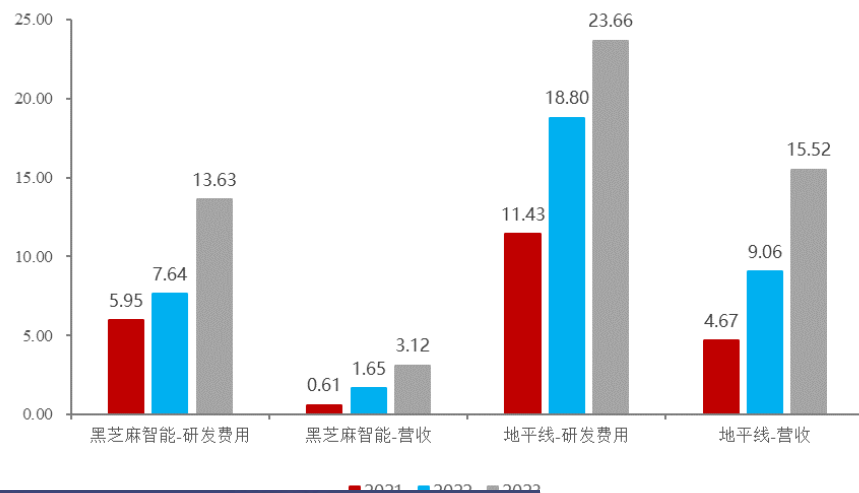


可行性分析：对照地平线/黑芝麻，芯片自研可为

- 对照国内智驾芯片初创企业地平线、黑芝麻智能等公司芯片自研历程，【千人研发规模；30~50亿研发投入；2~3年耗时】可完成智驾芯片全自研以及配套解决方案落地：
 - **地平线**：自2015年成立至2024年，公司累计融资171亿元人民币，创收30亿元以上，截至23年底在手现金114亿元，已完成涵盖L2/L3级别SoC芯片和配套工具链/底软等的开发和规模量产。
 - **黑芝麻智能**：自2017年成立至2024年，累计融资30亿元人民币，创收4.5亿元以上，截至22年底在手现金不足10亿元，同样完成L2级SoC智驾芯片（NPU/ISP）等IP核自研开发和规模量产。
- **研发耗时**：1) 地平线2015年成立，2019年首款智驾芯片落地；2024年预计落地征程6系列支持L3级别芯片；2) 黑芝麻智能2017年成立，2019年首款智驾芯片落地，2024年大算力落地。
- **团队规模**：截至2023年底，地平线/黑芝麻智能研发团队人数分别有1478/950人。
- **资金投入**：地平线2021年至2023年，研发费用累计投入54亿元，黑芝麻智能2020年至2023年研发费用累计投入30亿元，大额研发投入保证智驾芯片持续迭代升级。

图：地平线以及黑芝麻智能发展历史梳理以及财务比较/亿元

企业	时间	事件
地平线	2015	公司成立
	2019.8	发布征程2，支持L2
	2020.9	发布征程3，支持L2+
	2021.7	发布征程5，支持L2++
	2024	发布征程6，支持L3
	研发团队/人	1478
黑芝麻智能	2017	公司成立
	2019	华山一号A500
	2020	华山二号A1000
	2024	A2000
	研发团队/人	950



二、第三方玩家自研智驾芯片成效如何？

厂商布局比较：英伟达/特斯拉最全，其余快速跟进

■ 综合OEM主机厂以及Tier环节供应商，我们梳理自研智驾芯片并已有或即将有成熟产品量产出货的玩家进行横向对比：英伟达/特斯拉目前**云端&边缘端芯片硬件**以及**对应底软&工具链布局**最为完善，高通聚焦**边缘端自研&Tier1落地**模式迅速落地，地平线/黑芝麻智能由低到高布局。

图：行业智驾硬件各玩家对比

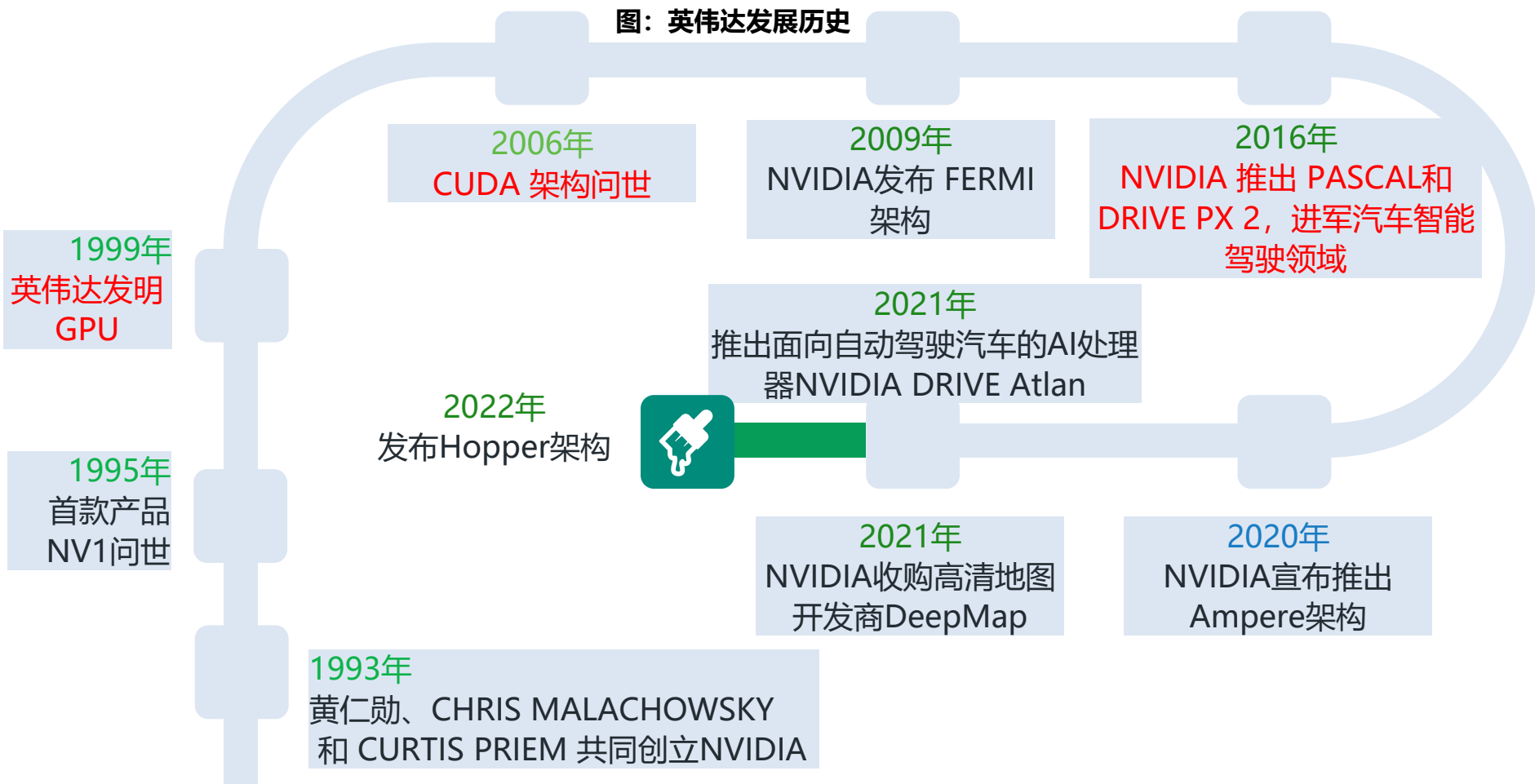
	边缘端芯片					云端		底层软件		策略总结
	NPU	ISP	CPU	GPU	其他（接口/传输类IP）	超算芯片	虚拟仿真环境	计算架构平台	SDK工具链	
英伟达	√	√	√	√	√	GB200(单颗4PFLOPS, 集群1.44EFLOPS)	NVIDIA DRIVE Sim	CUDA	DRIVE SDK	高举高打，GPU+CUDA构筑高壁垒，算力/生态最强音
华为	√	√	√	√	√	昇腾Atlas（昇腾910, 集群算力14-20PFLOPS)	华为云	CANN	MindStudio	全面对标英伟达，绑定部分主机厂定义整车
高通	√	√	√	√	√	-	-	联合谷歌/英特尔开发	Snapdragon Ride SDK	由座舱切入舱驾一体，边缘端芯片发力；国内绑定Tier1环节，快速入局
地平线	√	×	×	×	×	-	-	-	Horizon OpenExplorer	L2中低端产品线入局逐步向上突破，吸引产业战投赋能合作
黑芝麻智能	√	√	×	×	×	-	-	-	-	布局中低端产品
特斯拉	√	√	×	×	×	D1芯片（单颗0.36PFLOPS, 集群1.1EFLOPS)	自研	自研	自研	智驾软硬件全栈自研整合，加速能力迭代
Mobileye	√	√	×	-	-	-	-	-	EyeQ Kit SDK	黑盒转开放，高算力利用效率极致降本

2.1、英伟达：高举高打，算力+生态最强音

发展历程：由GPU起构建软硬件壁垒，拓展全行业

- 英伟达成立于1993年，由黄仁勋联合Sun公司两位年轻工程师共同创立。最初致力于GPU的研发，1999年成功上市。随着GPU在图形和高性能计算领域的成功，英伟达逐渐扩展至人工智能、深度学习、自动驾驶和医疗等领域。公司的GPU技术在科学计算、游戏和专业工作站等领域取得巨大成功，成为全球领先的半导体公司之一。

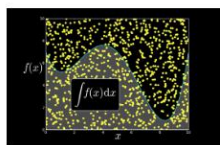
图：英伟达发展历史



CUDA：更好加速GPU计算，构建英伟达生态壁垒

- **CUDA 是 NVIDIA 发明的一种并行计算平台和编程模型**，全称Compute Unified Device Architecture。它通过**更好地调用图形处理器 (GPU) 的处理能力**，对算法运行进行加速，可大幅提升计算性能，并构建英伟达自身的软件生态。CUDA的优势在于：**1) 并行计算**：CUDA允许开发者使用GPU的大量核心进行并行计算，以加速各种计算密集型任务；**2) 高效内存管理**：CUDA提供了高效的内存管理机制，包括全局内存、共享内存、常量内存等，可以最大限度地利用GPU的内存资源；**3) 强大的工具支持**：CUDA提供了一系列强大的工具支持，包括CUDA编译器、CUDA调试器、CUDA性能分析器等，可以帮助开发者更加高效地开发和调试CUDA程序。

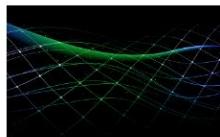
Libraries



cuRAND



NPP



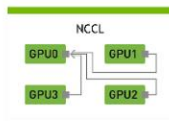
Math Library



cuFFT

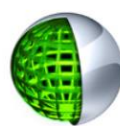


nvGRAPH

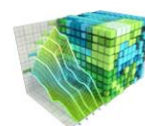


NCCL

Tools and Integrations



Nsight



Visual Profiler



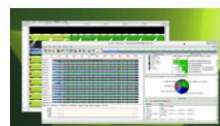
CUDA GDB



CUDA MemCheck

OpenACC

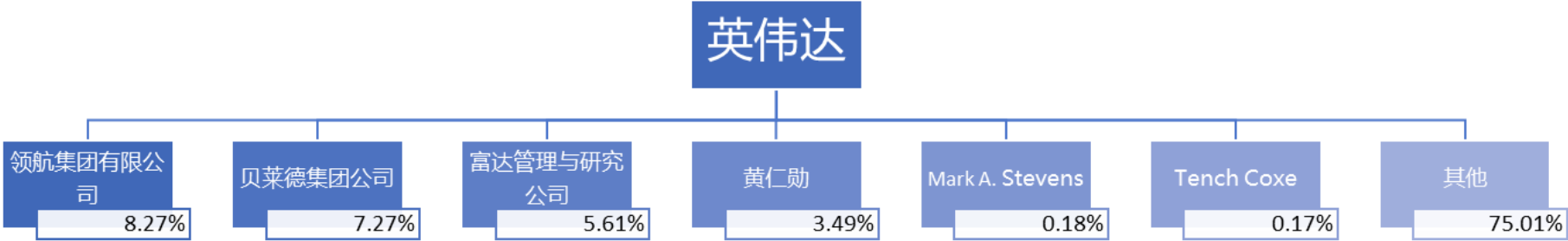
OpenACC



CUDA Profiling Tools Interface

组织架构以及股权关系：黄仁勋为最大个人股东

➤ 英伟达的股权结构呈现多元化，包括机构和个人持股。根据2023年年报数据显示，领航集团有限公司、贝莱德集团公司和FMR LLC等机构股东持有相对较大的股份，分别为8.27%、7.27%和5.61%，公司创始人黄仁勋本人持有3.49%的股份。



图：英伟达公司主要高管

公司高管	职位	工作历史	职能
黄仁勋	英伟达总裁、首席执行官和董事会成员	曾在AMD（美国微处理器制造商Advanced Micro Devices）工作，后在LSI Logic 继续从事芯片设计	大规模集成电路芯片系统以及策略，人工智能与高效能计算领域
Chris A. Malachowsky	英伟达创始人，担任公司管理人员和高级技术主管	在 HP 和 Sun Microsystems 担任工程和技术领导职务	公司技术和架构发展
Colette Kress	英伟达执行副总裁兼首席财务官	曾在Cisco, Microsoft, Texas Instruments等公司担任高级财务职位	财务战略、规划、报告和业务开发
Jay Puri	英伟达运营执行副总裁	曾为Sun Microsystems, Hewlett-Packard Company, Booz Allen & Hamilton 和Texas Instruments等公司工作	销售、营销和综合管理
Debora Shoquist	英伟达运营执行副总裁	曾为JDS Uniphase, Coherent 以及 Quantum工作	公司的运营和供应链职能

■ 英伟达主要系列芯片包括GeForce、Quadro、Tesla、Tegra、Jetson和DXG，算力&架构持续迭代。1999年，英伟达推出GeForce系列芯片，主要应用于游戏娱乐；此后，专业级GPU Quadro系列芯片凭借强大的计算能力和大容量显存，广泛应用于专业可视化领域；2008年推出的Tesla系列芯片可提供快速运算和推理，作为深度学习加速器运用于数据中心；2015、2016年分别推出Jetson系列和DGX系列，计算推理能力进一步提高，应用于数据中心、汽车、医疗等领域。英伟达凭借其算法架构的迭代升级，不断开拓产品线，专业化高算力芯片提高整体竞争力。

图：英伟达主要芯片系列

系列	推出时间	产品线	简介	特点	主要应用领域	代表芯片			
						名称	推出时间	FP32算力 (TFLOPS)	微架构
GeForce	1999	游戏	消费级GPU	具备实时光线追踪和DLSS等先进技术	游戏娱乐、图形设计、科学计算、工业	RTX 4090	2022	39.69	Ada Lovelace
Quadro	1999	专业可视化	专业级GPU	计算能力强大、大容量显存、专业	CAD、动画制作、科学计算、虚拟现实	RTX 6000	2022	31	Ada Lovelace
Tesla	2008	数据中心	深度学习加速器	提供快速的矩阵运算和神经网络推理	科学计算、数据分析、深度学习	P100	2016	10.6	Pascal
Tegra	2008	游戏、汽车	移动处理器	高性能图形和计算能力，低功耗、高度集成	嵌入式系统、智能手机、平板电脑、汽车电子	Tegra 3	2011	-	ARM
Jetson	2015	数据中心、汽车、医疗	嵌入式开发平台	计算和推理能力强大	边缘计算、人工智能、机器人	TX 2	2017	6.2	Pascal
DGX	2016	数据中心、汽车、医疗	HPC服务器	计算和训练能力强大，大规模学习	深度学习、人工智能研究和开发	H100	2022	60	Hopper

产品线：游戏以及数据中心为当前出货主力

■ 英伟达产品线以游戏、数据中心、专业可视化、汽车、医疗为主。

- **游戏方面，以GeForce系列芯片为主。**1999年，英伟达即推出首款GPU GeForce256，随着算法架构的迭代升级，GeForce系列芯片性能也不断提高。英伟达凭借GeForce系列芯片久远的迭代历史，不断提升产品竞争力和品牌影响力，巩固其开拓其他产品线的基础。
- **数据中心方面，高算力芯片助力实现AI高性能计算。**2011年起，英伟达开始发布应用于数据中心的高算力芯片，至2022年，基于Hopper架构的H100芯片单精度浮点算力可达51TFLOPS，算力实现飞跃，保障高性能计算。目前，英伟达数据中心Volta系列芯片和A100为AI训练加速器，以Tesla T4和Jetson Xavier NX为AI推理加速器，以Tesla系列芯片为高性能计算加速器。
- **专业可视化方面，Quadra+RTX实现可视化。**基于RTX和Quadro系列芯片，Omniverse搭建实时图形仿真平台，用于数字内容创作、医疗和建筑设计等领域的CloudXR提升创作速度质量。

图：英伟达游戏芯片迭代

年份	发布芯片	简介
1999	GeForce256	首款真正的GPU
2006	GeForce 8800 GTX	首款支持DirectX11
2010	GeForce GTX660	首款基于Kepler架构
2014	GeForce GTX980	首款基于Maxwell架构
2016	GeForce GTX1080	首款基于Pascal架构
2018	GeForce RTX2080	首款基于Turing架构
2020	GeForce RTX3080	首款基于Ampere架构
2022	GeForce RTX4080	首款基于AdaLovelace架构

图：英伟达数据中心芯片迭代

年份	发布芯片	架构	算力/TFLOPS
2011	Tesla M2090	Fermi 2.0	1.3
2013	Tesla K40	Kepler	4.2
2015	Tesla M40	Pascal	7
2016	Tesla P100	Pascal	9.3
2017	Tesla V100	Volta	14
2020	A100	Ampere	19.5
2022	H100	Hopper	51

■ 英伟达产品线以游戏、数据中心、专业可视化、汽车、医疗为主。

- **汽车方面，高算力芯片助力智驾功能突破升级。** 1) 硬件方面，自动驾驶平台经历了DRIVE PX、DRIVE PX2、DRIVE Xavier、DRIVE Pegasus、DRIVE Orin、DRIVE Thor的迭代。最新一代自动驾驶平台DRIVE Thor支持L4/L5级别智驾，算力可达2000TOPS，同时，自动驾驶开发平台Hyperion也将搭载Thor实现性能升级；2) 软件方面，CUDA+TensorRT持续优化DRIVE OS，进而提升DRIVE SDK整体性能。
- **医疗方面，**2016年英伟达开始布局医疗领域；2017年合作医疗保健解决方案提供商，将AI带入医学影像；2018年发布Clara平台；2021年合作Schrödinger，利用DGX A100扩大计算药物发现平台的速度和准确性；2022年发布IGX平台，改善人机协同。

图：英伟达汽车芯片迭代历程

平台	发布时间	GPU	智驾级别	功耗 (W)	算力 (TOPS)	制程 (nm)	搭载车型
PX	2015	Tegra X1	L2/L3	150	2	28	/
PX2 (Auto Cruise)	2016	Tegra X2	L2/L3	125	4	16	ZF ProAI
PX2 (Auto Chauffeur)	2016	Tegra X2 Pascal GPU	L3/L4	250	24	16	Model S/X/3
AGX Xavier	2017	Tegra Xavier	L3/L4	30	30	12	小鹏P5/P7
AGX Pegasus	2017	Tegra Xavier	L5	500	320	12	戴姆勒&博世Robotaxi
AGX Orin	2019	Turning GPU	L4/L5	75	254	7	理想/小鹏/蔚来
AGX Thor	2022	Hopper GPU Ada Lovelace GPU	L4/L5	/	2000	/	极氪

- 算法平台方面，英伟达六大不同算法平台匹配高性能计算（云端数据中心）、边缘端以及虚拟仿真、智驾等多个领域。
- DGX和HGX为AI高性能计算平台，配备H100/A100，均用于大规模学习和计算，后者相对更加灵活；
- EGX和IGX为边缘计算平台，均配备Ampere系列GPU，EGX因其灵活性，适用于视频分析、机器视觉等领域，IGX专为工业医疗等领域设计；
- AGX为自动驾驶领域的可扩展式开放平台，根据自动驾驶需求配备不同架构GPU；
- OVX为虚拟化平台，配备L40S，主要用于数字孪生模拟。

图：英伟达算法平台

平台	介绍	配备GPU	特点	适用范围
DGX	AI高性能计算平台	H100/A100	标准化	大规模深度学习/人工智能应用
HGX	高性能计算和AI平台	H100/A100	灵活定制化	大规模数据中心/云计算
EGX	边缘计算平台	Ampere系列	高度灵活	视频分析/物联网数据处理/机器视觉
IGX	边缘AI平台	Ampere系列	工业级/安全/可靠	工业/医疗
AGX	可扩展式开放平台	Tegra/Pascal/Turning/Hopper/AdaLovelace系列	低能耗/高性能/安全/灵活	自动驾驶
OVX	虚拟化平台	L40S	可靠稳定/高性能	数字孪生模拟（建筑/工厂/城市）

GPU微架构持续迭代，制程升级，覆盖更多领域

- 英伟达GPU微架构持续迭代升级，Fermi、Kepler、Maxwell、Pascal、Volta、Turing、Ampere、Ada Lovelace和Hopper，每一代都在性能、能效和特定任务方面取得不断进步：
- 2010年引入CUDA架构，2012年进入深度学习领域，2016年拓展HPC，2017年加速数据传输。

架构	时间	介绍	核心参数	代表产品	纳米制程	应用领域
Fermi	2010	引入 CUDA架构 、ECC内存、NVIDIA Parallel DataCache、和GPU直接支持C++等	Fermi架构共含4个GPC，16个SM，512个CUDA Core。每32个CUDA Core组成1个SM每个SM为垂直矩形条带。	GeForce 400/500系列，Tesla M2050/M2070/M2090	40/28nm 30 亿晶体管	科学计算、图形处理和高性能计算。
Kepler	2012	引入了GPU Boost技术， 增加了对动态并行计算的支持	15 个 SMX，每个 SMX 包括 192 个 FP32+64 个 FP64+CUDA Cores	GeForce 600/700系列，Quadro K/M系列	28nm 71 亿晶体管	科学计算、深度学习和游戏等领域
Maxwell	2014	引入多层次的内存系统、动态超分辨率技术和VR Direct技术等	16 个 SM，每个 SM 包括 4 个处理块，每个处理块包括 32 个 CUDA Cores +8 个 LD/ST Unit +8 SFU	GeForce 900系列，Quadro M系列	28nm 80 亿晶体管	游戏、深度学习和移动设备。
Pascal	2016	引入了16nm FinFET制程技术，提出 NVIDIA Tensor Cores	GP100 有 60 个 SM，每个 SM 包括 64 个 CUDA Cores，32 个 DP Cores	GeForce 10系列，Quadro P系列	16nm 153 亿晶体管	深度学习和高性能计算领域。
Volta	2017	Nvlink2.0 Tensor Core 1.0	80个SM，每个SM包括32个FP64 +64个Int32+64个FP32 + 8个Tensor Cores	Titan V，Quadro GV100	12nm 211 亿晶体管	深度学习、科学计算和高性能计算。

■ 英伟达GPU微架构持续迭代升级，Fermi、Kepler、Maxwell、Pascal、Volta、Turing、Ampere、Ada Lovelace和Hopper，每一代都在性能、能效和特定任务方面取得不断进步：

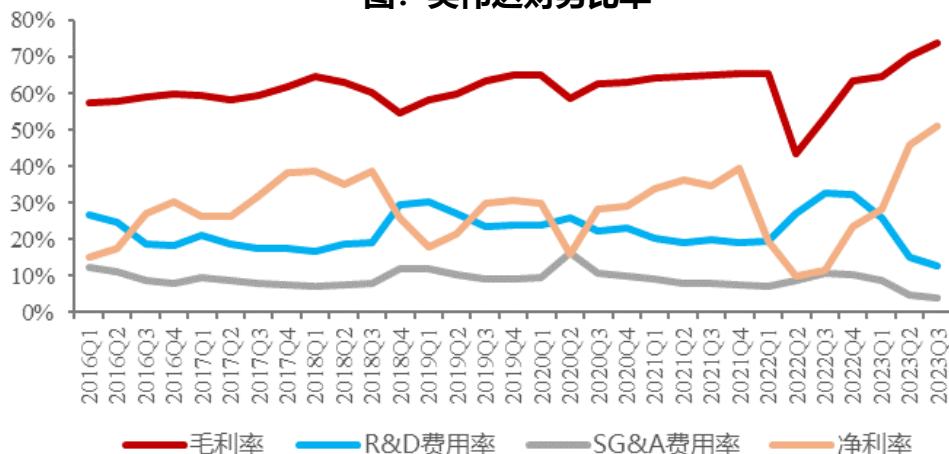
➤ 2017年后引入Tensor Core，减少乘加操作时间，提供更快计算速度，成为企业级AI首选，年拓展专业可视化市场，2020年之后正式引入支持AI神经图形以及算力稀疏化的微架构方案。

架构	时间	介绍	核心参数	代表产品	纳米制程	应用领域
Turing	2018	引入了 实时光线追踪技术 、深度学习技术（如RT Cores和Tensor Cores）以及新的流程图渲染技术	102核心92个SM，SM重新设计，每个SM包含64个Int32+64个FP32+8个Tensor Cores	GeForce 16/20系列, Quadro RTX系列	12nm 186 亿晶体管	游戏、深度学习和专业可视化等领域
Ampere	2020	引入了更多的Tensor Cores、第三代NVLink以及改进的Ray Tracing技术	108个SM，每个SM包含64个FP32+64个INT32+32个FP64+4个Tensor Cores	GeForce 30系列, A100,A40,A30,A10	7nm 283 亿 晶体管	深度学习、科学计算和高性能计算领域
Ada Lovelace	2022	引入了第四代 Tensor Core和第三代 RT Core	144 个 SM，每个 SM 包含 128 CUDA Cores, 1 个第三代 RT Core, 4 个第四代 Tensor Core, 四个纹理单元、一个256 KB 的寄存器文件 和 128 KB 的 L1/共享内存	GeForce RTX40系列	4nm 763 亿 晶体管	光线追踪和 基于AI的神经图形
Hopper	2022	Tensor Core 4.0 Nvlink 4.0 结构稀疏性矩阵 MIG 2.0	132个SM，每个SM包含128个FP32+64个INT32+64个FP64+4个Tensor Cores	Telsa H100	4nm 800 亿 晶体管	深度学习、科学计算和高性能计算

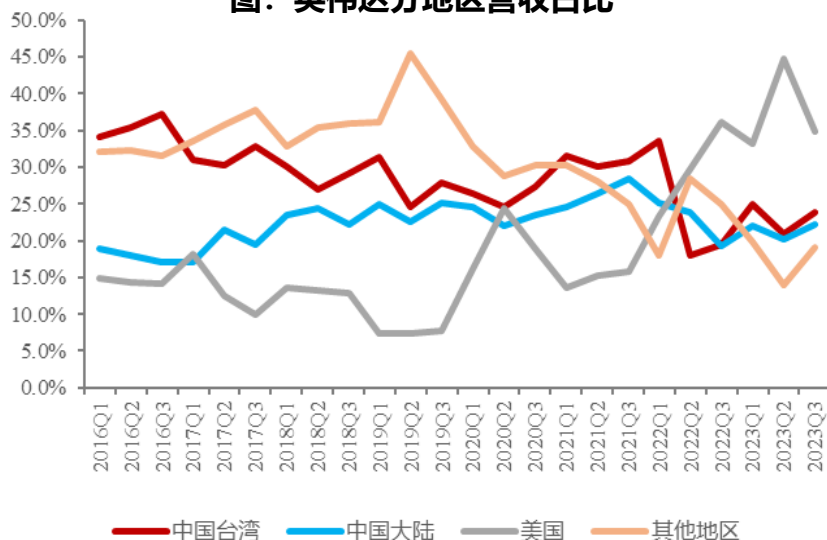
财务：游戏业务贡献营收增量，净利率靓丽

- **营收维度**，游戏业务/数据中心业务接力，先后成为公司主力业务，2023Q2以来数据中心业务出货量迅速爆发，支撑营收保持高速增长，主要集中于北美市场
- **盈利能力领先**。毛利率持续高位，规模效应提升驱动2023Q3毛利率提升至70%以上水平，带动净利率突破50%，

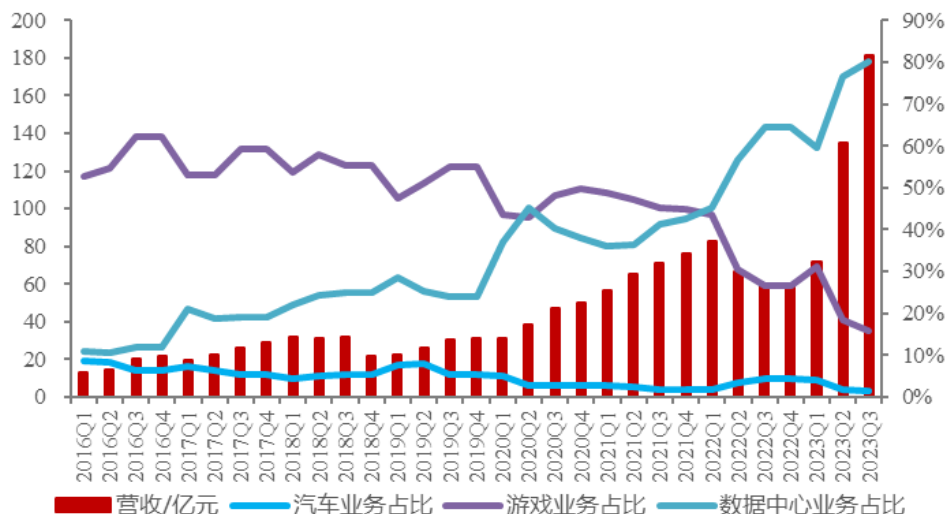
图：英伟达财务比率



图：英伟达分地区营收占比



图：英伟达营收以及分业务板块占比



2.1.1 英伟达——汽车智驾业务布局

英伟达：GPU硬件+CUDA软件构建壁垒，衍生AI

➤ 以GPU++CUDA为自身AI智能领域核心壁垒，英伟达辐射拓展汽车智驾业务，具备领先技术优势

GPU以及衍生产品：游戏显卡为基础，Hopper架构+Transformer加持，加速AI训练

芯片Soc异构方案：GPU配合Grace CPU形成大算力超级AI芯片

BlueField数据传输：自研DPU芯片，支持数据中心超大规模AI数据安全+快速传输

云数据中心：融合CPU/GPU/DPU形成HPU超算中心，支持云端/本地大模型计算以及仿真

硬件：由GPU拓展CPU/DPU，部署云端超算以及边缘端解决方案

软件：CUDA构建高生态壁垒，配合硬件形成各类解决方案

深度学习：CUDA编程高性能库，支持块API，便于利用GPU进行基于大数据的推理以及训练

图形渲染：分为图像研究、图像处理、渲染性能以及光线追踪等四部分，应用于科研以及游戏等

仿真模拟：动力学以及医学场景模拟，加速数据搜集以及模型训练进程

数据网络加速：通过创建DPU加速服务，对数据中心基础架构进行编程，满足数据传输需求

行业主流

开发操作系统及中间件，初步形成包括感知-实时构图-规控的软件栈平台

智驾硬件解决方案上车，搜集场景化数据

数据处理（清洗/标注），基于实测数据调参，进行仿真模拟

云端超算中心利用实测+仿真数据训练算法

云端算法裁剪落地边缘端，形成实时解决方案

英伟达

DRIVE Chauffeur平台

DRIVE SDK 软件工具包
DRIVE OS 操作系统
DRIVE Works 中间件
DRIVE Map 智驾地图
DRIVE AV 智驾感知

DRIVE Hyperion

15摄像头
9毫米波雷达
12超声波雷达
1激光雷达
完整软件栈

Omniverse Cloud

OVX硬件服务器：
GPU+高速网卡

Omn-Replicator：
生成3D数据

DGX-SuperPOD

DGX高性能服务器+InfiniBand网卡驱动，提供卓越性能训练算法

智驾完整软件栈

云端算法裁剪落地边缘车端，并搜集数据持续OTA升级，形成闭环迭代。

1、AI基础设施一：硬件——DGX Super POD

- NVIDIA DRIVE 基础架构包括开发自动驾驶技术（从原始数据采集到验证）所需的**完整数据中心硬件、软件和工作流**。该基础架构为神经网络开发、训练和验证以及仿真测试提供了所需的端到端构建模块。其包括：**DGX云、NVIDIA DGX Super POD以及AI Enterprise软件套件三大核心**。
- NVIDIA DGX SuperPOD：一站式AI基础架构。多个 DGX 服务器组成的先进 AI 计算基础架构，可提供卓越的性能。这使得 OEM 能够更快、更高效地训练和优化深度学习模型，从而缩短开发安全自动驾驶系统所需的时间。

图：英伟达DGX SuperPOD服务



H100/A100等大算力GPU满足DGX云平台计算需求

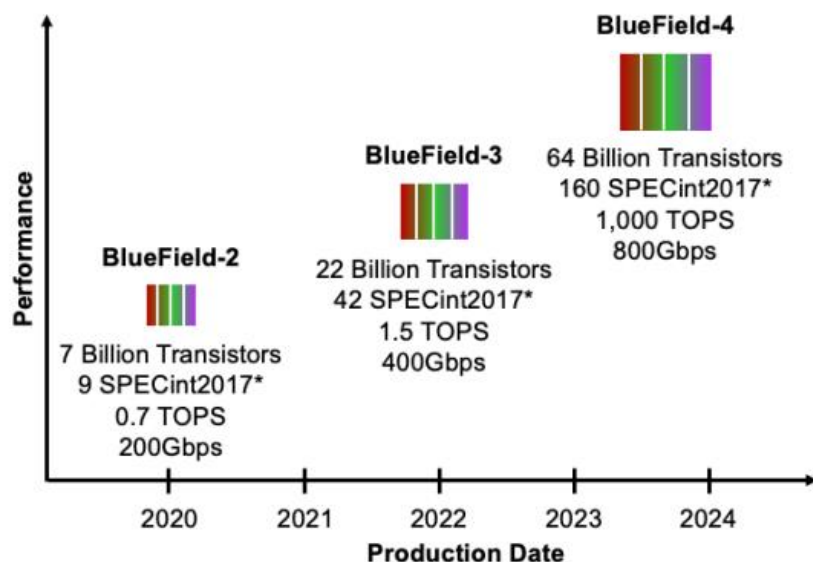
- 超大算力芯片主要用于AI领域的图形和计算，产品矩阵不断丰富。目前，英伟达主流GPU产品均基于Ampere、AdaLovelace和Hopper架构构建，应用于图形和计算领域，能力覆盖深度学习训练、数据分析、推理、高性能计算、AI等。

图：英伟达大算力芯片产品矩阵

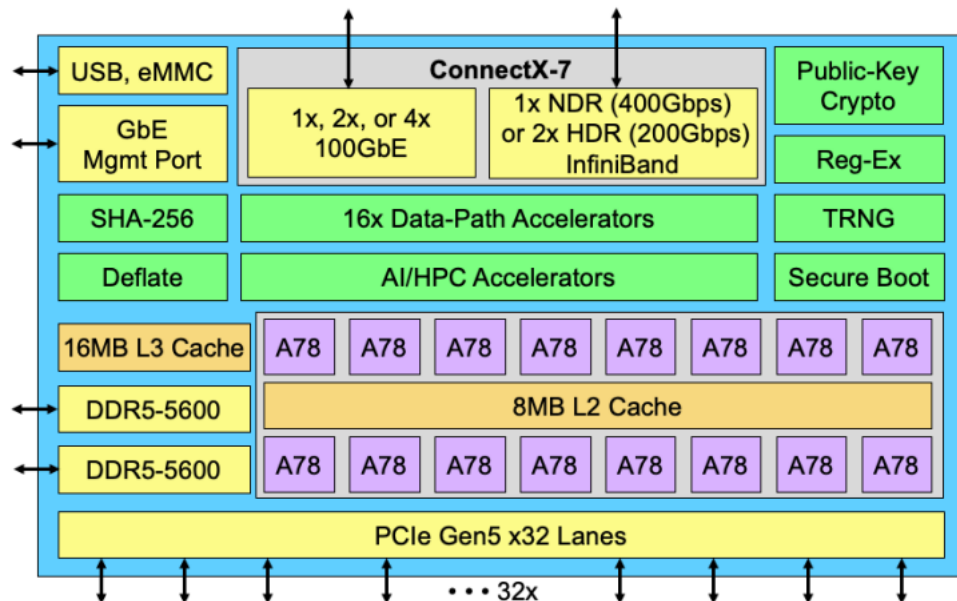
解决方案	GPU	推出时间	架构	FP32算力 (TFLOPS)	网络解决方案	深度学习训练和数据分析	深度学习推断	HPC/AI	NVIDIA Omniverse	虚拟工作站	虚拟桌面	AI视频	远端加速	AI-on-5G
计算	AX800/A100X	2023	Ampere	19.5	QTM1 SPTM3	HGX PCIe A100X	HGX PCIe	HGX PCIe A100X						AX800/A100X
	A30	2021	Ampere	10.3	SPTM3		Pcie	Pcie						A30x
	GH200	2023	Hopper	-	QTM2 SPTM4									
	H100	2022	Hopper	51	QTM2 SPTM4	HGX PCIe NVL	HGX PCIe NVL	HGX PCIe NVL						
图形和计算	A40	2023	Ampere	37.4	SPTM3									
	A10	2021	Ampere	31.2	SPTNE									
	A16	2021	Ampere	18	SPTM3									
	L40S	2023	Ada Lovelace	91.6	SPTM4									
	L40	2022	Ada Lovelace	90.5	SPTM4									
小型计算和图形	A2	2021	Ampere	4.5	SPTM3									
每个解决方案类别（计算、图形和计算、SFF 计算和图形）和工作负载列的性价比比较						■ NVIDIA HGXTM或DGX™with SXM form factor			■ NVIDIA Quantum-1 IB switch plus BlueField-2 DPUs或 NVIDIA ConnectX®-6/6 Dx SmartNICs					
						■ PCIe form factor			■ NVIDIA Spectrum™-3 Ethernet switch plus BlueField-2 DPUs或 ConnectX-6/6 Dx SmartNICs					
						■ A100X/A30X 融合加速器: A100 or A30 Tensor Core GPU			■ NVIDIA Spectrum-4 Ethernet switch plus BlueField3 DPUs或 ConnectX-7 SmartNICs					
						■ H100 NVL form factor								

- DPU（数据处理器，Data Processing Unit），是数据中心第三颗主力芯片。2020 年，NVIDIA 推出 BlueField-2 DPU，将其定义为继 CPU 和 GPU 之后“第三颗主力芯片”，正式拉开 DPU 发展的序幕
- **DPU有望提高数据中心的效率，为异构处理组合增添了新的元素。**DPU 对于数据中心的分解非常重要，它允许服务器处理器只执行计算任务，而 DPU 则处理网络计算和存储之间的数据移动。通过使用基于 DPU 的智能网络接口卡 (NIC)，云服务提供商可以节省服务器处理器的计算周期，用于创收服务。DPU 还能比服务器处理器更有效地处理网络流量，从而降低数据中心的能耗。在存储系统中，DPU 可以取代标准处理器，处理 SSD 阵列的巨大吞吐量，同时降低功耗。

图：BlueField DPU 迭代图



图：BlueField-3 DPU



1、AI基础设施二：软件——AI Enterprise

- NVIDIA AI Enterprise 是 NVIDIA AI 平台的软件层，可访问数百个 AI 框架。其中包括 TensorFlow、PyTorch 和 NVIDIA® CUDA-X™，可让 AI 公司创建、测试、训练和部署复杂的 AI 算法。

MLOps

AI Applications

NVIDIA AI Enterprise

Infrastructure Management

Cloud Native Management and Orchestration
GPU Operator, Network Operator

Cluster Management
Base Command Manager Essentials

Infra Acceleration Libraries
Magnum IO, vGPU, CUDA

Application Frameworks



LLM
NeMo



Speech AI
Riva



Cybersecurity
Morpheus



Medical Imaging
Clara

...

More

AI Development

Data Science/Prep
RAPIDS, RAPIDS Accelerator
for Apache Spark

Model Training and Customization
NeMo, TAO, PyTorch, TensorFlow

Deploy at Scale
Triton Inference Server

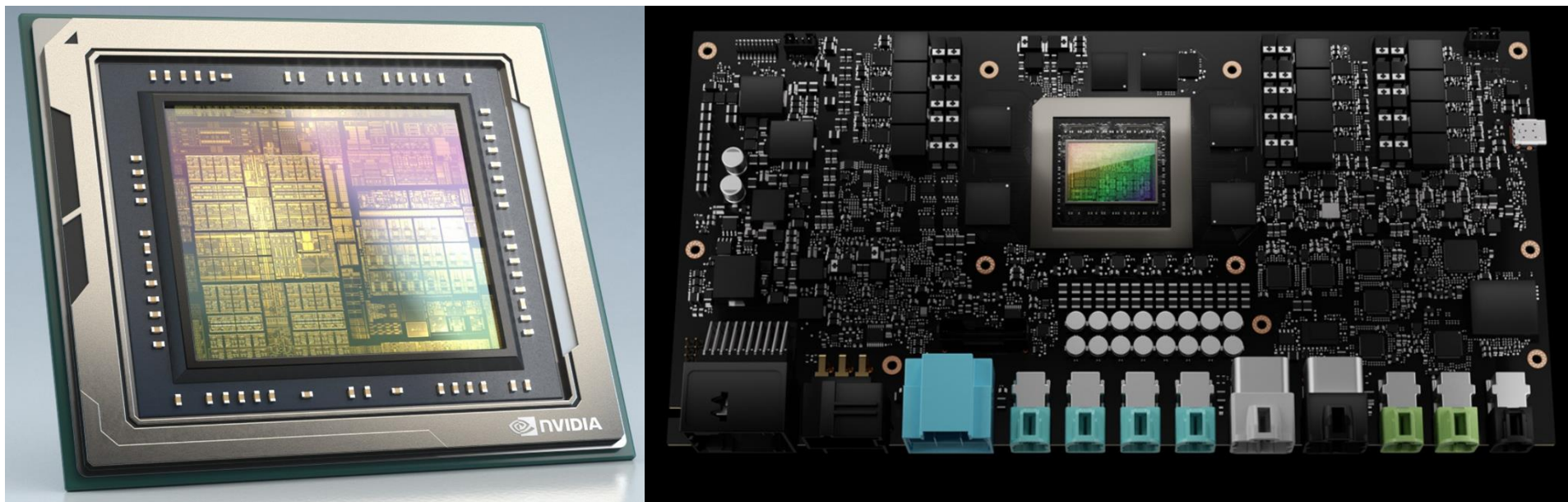
Optimize for Inference
TensorRT, TensorRT-LLM

Cloud | Data Center | Workstations | Edge

2、智驾硬件：Hyperion感知与规控硬件套件

- NVIDIA DRIVE Hyperion™ 是用于量产自动驾驶汽车的平台。此自动驾驶汽车参考架构通过将基于 DRIVE Orin™ 的 AI 计算与完整传感器套件（包含 12 个外部摄像头、3 个内部摄像头、9 个雷达、12 个超声波、1 个前置激光雷达和 1 个用于真值数据收集的激光雷达）相集成，能够加速开发、测试和验证。
- ORIN可提供每秒 254 TOPS（万亿次运算），包括87TOPS的DLA算力以及157TOPS的Ampere架构下的GPU算力，专用于智驾行业。下一代THOR芯片预计于2025年正式量产，支持Soc多域计算，可同时聚焦智驾与智舱多领域，运行Linux、QNX以及安卓多系统，提供1000TOPS算力，同时有效降低成本。

图：英伟达智驾边缘端硬件芯片



- 英伟达2019年推出 DRIVE AGX Orin，是适用于自动驾驶车辆和机器人的高度先进的软件定义平台，由名为 Orin 的新型片上系统 (SoC) 提供支持，该系统由 170 亿个晶体管组成。Orin SoC 集成了 NVIDIA 的下一代 GPU 架构和 Arm Hercules CPU 内核，以及新的深度学习和计算机视觉加速器，每秒可实现 254万亿次运算，几乎是 NVIDIA 上一代 Xavier SoC 性能的 7 倍。借助可扩展的 DRIVE Orin 产品系列，开发者只需在整个车队中构建、扩展和利用一次开发投资，便可从 L2+ 级系统一路升级至 L5 级全自动驾驶汽车系统。
- 2022年，DRIVE Thor问世。汽车级片上系统 (SoC) 基于最新的 CPU 和 GPU 技术而构建，**可提供 1000teraflops 的性能，同时降低总体系统成本**。DRIVE Thor 统一了车辆中传统的分布式功能，包括数字集群、信息娱乐、停车和辅助驾驶，以提高开发效率和加快软件迭代速度。

图：英伟达智驾边缘端硬件迭代发展历史

	DRIVE AGX Xavier	DRIVE AGX Orin	DRIVE AGX Thor
发布时间	2017	2019.12	2022.9
CPU	8*ARM 64	16*ARM 64	Grace CPU
GPU	1*Tegra Xavier	2*Turning GPU	Hopper GPU Ada Lovelace GPU
自动驾驶级别	L3/L4	L4/L5	L4/L5
功耗W	30	75	/
算力TOPS	30	254	2000
制程nm	12	7	/
搭载车型	小鹏P5/P7	理想/小鹏/蔚来/智己/高合/广汽/长安	极氪

3、智驾软件：DRIVE SDK多样化工具覆盖

- 开放式 NVIDIA DRIVE SDK 为开发者提供了**自动驾驶所需的所有构建块和算法堆栈**。该软件有助于开发者更高效地构建和部署各种先进的自动驾驶应用程序，**包括感知、定位和地图绘制、计划和控制、驾驶员监控和自然语言处理**。

图：英伟达底软架构

软件名称	软件内核
DRIVE OS操作系统	DRIVE Software 堆栈的基础所在，这是针对车载加速计算率先推出的安全操作系统。它包括用于传感器输入处理的 NvMedia、用于实现高效并行计算的 NVIDIA CUDA® 库、用于实时 AI 推理的 NVIDIA TensorRT™，以及可访问硬件引擎的其他开发者工具和模块
DriveWorks中间件	在 DRIVE OS 之上提供对自动驾驶汽车开发至关重要的中间件功能。这些功能包括传感器抽象层 (SAL) 与传感器插件、数据记录器、车辆 I/O 支持和深度神经网络 (DNN) 框架。该工具拥有模块化和开放的特点，在设计上符合汽车行业软件标准
DRIVE AV智驾汽车算法	软件栈包含感知、地图构建和规划层，以及各种经过高质量真实驾驶数据训练的深度神经网络 (DNN)。这些丰富的感知输出可以用于自动驾驶和地图构建，使 DRIVE AV 成为您的“司机”。在规划和控制层中，NVIDIA Safety Force Field™ (NVIDIA 安全力场) 计算模组使车辆远离伤害，并确保它不会造成或引发不安全的情况
DRIVE IX智舱感知平台	开放软件平台，可为 AI 驾驶舱创新解决方案提供舱内感知。可提供用于访问各项功能的感知应用程序，还可提供 DNN 以实现高级驾驶员和乘客监控功能、AR/VR 可视化以及车辆与乘客之间的自然语言交互
DRIVE Map智驾地图	使用准确的真值地图和可扩展的车队来源地图来创建和更新自动驾驶汽车地图。它利用数百万消费者车队的集体记忆，以及来自 DRIVE Hyperion™ 车辆的丰富传感器数据。NVIDIA DGX SuperPOD 基础架构使其能够处理 TB 级数据并在全球范围内维护地图

- **CUDA® 是 NVIDIA 开发的并行计算平台和编程模型，用于 GPU 上的通用计算。**
- **NVIDIA TensorRT™ 是一个高性能深度学习推理平台。它包括硬件感知的深度学习推理优化器和运行时，可为深度学习推理应用程序提供低延迟和高吞吐量（DLA）。**
- **NvStreams 是一种高效的 API，可提供对高速数据传输的访问，从而实现自动驾驶车辆所需的复杂处理 workflow。**
- **NvMedia 是一组高度优化的 API，可直接访问硬件加速的计算引擎和传感器，包括编码器/解码器、传感器输入处理、图像处理等。**



- DRIVE Sim 是一个开放式模组化可扩展平台，可让用户根据自己的需求自定义仿真器，可以使用随附的 SDK，为传感器模型、车辆动力学、交通模型或自定义硬件的界面轻松构建扩展程序。其包括：
- **硬件端：** NVIDIA OVX 系统均由 NVIDIA 认证的合作伙伴制造和销售，最多可将八个最新的 **NVIDIA Ada Lovelace L40S GPU** 与高性能 ConnectX 和 Bluefield 网络技术相结合，满足企业组织对加速性能的大规模需求；
- **软件端：** 借助 NVIDIA OMNIVERSE Replicator™，开发者可以为罕见和复杂场景创建多样化的合成数据集，包括基于物理性质的传感器数据和像素准确的真值标签。这些标签包括深度、速度、遮挡和其他难以标记的参数。



4、智驾解决方案客户持续开拓，朋友圈持续扩容

- 在 2022GTC 大会上，比亚迪和 Lucid 集团宣布他们的新一代车型将采用 NVIDIA DRIVE。蔚来、理想汽车、小鹏、上汽集团旗下的智己汽车和飞凡汽车、集度、华人运通、VinFast、威马汽车等新能源汽车初创企业也都在 DRIVE 平台上开发软件定义汽车。
- 世界前沿的无人驾驶通用解决方案公司轻舟智航（QCraft）宣布推出搭载NVIDIA DRIVE Orin的最新一代车规级前装量产自动驾驶解决方案，并实现了L4级乘用车车队在国内的率先落地。
- 英伟达助力蔚来 ES7 重新定义智能电动 SUV；路特斯Eletre搭载NVIDIA DRIVE Orin，专为极致驾驶体验和高速AI计算而生
- 小鹏汽车超快充全智能SUV-G9配备NVIDIA DRIVE 集中式计算平台以及DRIVE Orin系统级芯片，以每秒508万亿次算力，可实现车辆从起点停车位到终点停车位全程使用辅助驾驶
- 哪吒汽车基于NVIDIA DRIVE Orin打造软件定义的电动汽车，除了在车辆设计工作中采用DRIVE Orin，还在使用NVIDIA解决方案来开发先进的自动驾驶功能。双方正在通力合作，设计和开发用于L4级自动驾驶的集中式跨域融合计算平台系统。

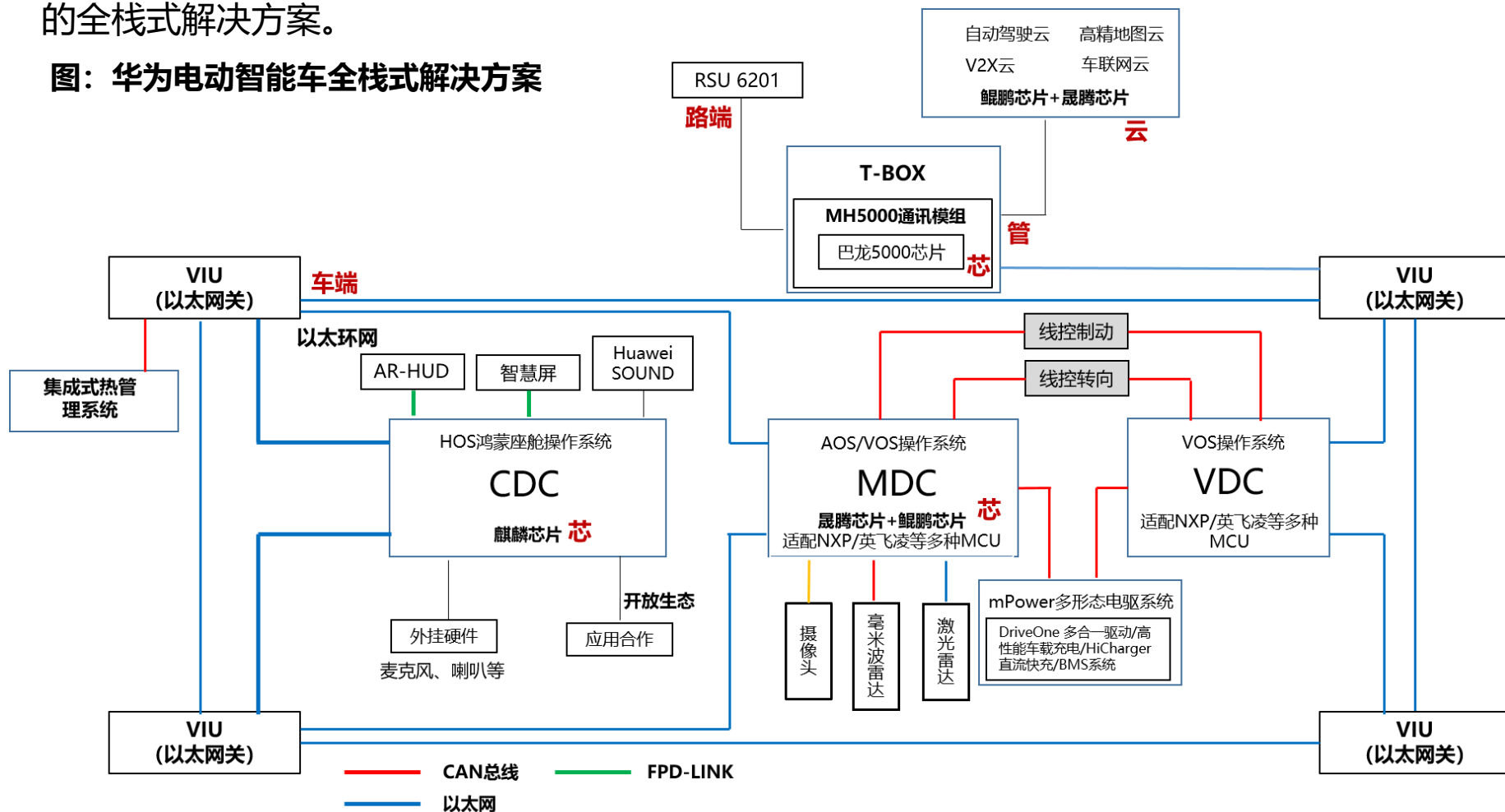


2.2 华为：技术对标英伟达，联合车企培育生态

以“计算+通信”为核心—CCA架构+Vehicle Stack跨域集成软件框架

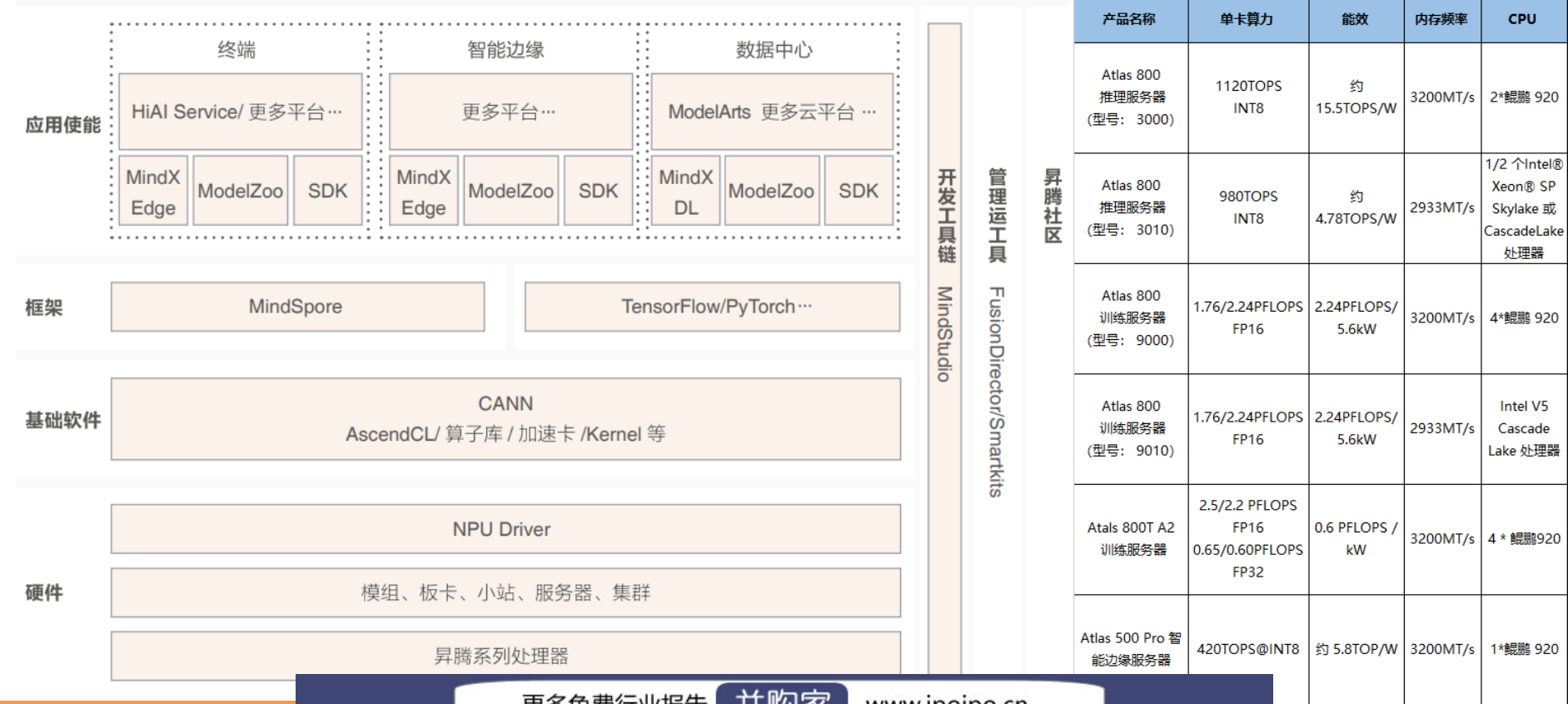
- 以ICT技术为基础，建立以一个架构（CCA）、五大智能系统（智能驾驶/智能座舱/智能电动/智能车云/智能网联）、全套智能化部件（智慧屏+AR-HUD+集成式热管理+感知铁三角等）组成的全栈式解决方案。

图：华为电动智能车全栈式解决方案



- **基础硬件是核心**：基于华为达芬奇架构，Atlas训练集群可提供 256P~1024P FLOPS FP16的总算力，并可提供能效比小于2TOPS/W的边缘端算力，满足效率与能耗的双重需求。
- **适配不同类型需求，华为提供异构计算架构CANN/AI框架/应用使能等不同类型的开发工具**：面向上层应用开发者以及专业AI模型开发者，华为分别提供MindX/MindSpore完整开发工具包；面向底层算子开发者，华为提供CANN以及MindStudio支持底层开发。

图：华为昇腾计算软硬件架构以及产品比较



- **华为昇腾芯片**是华为发布的两款人工智能处理器，包含**昇腾 310 用于推理**和**910 用于训练业务**，均采用自研达芬奇架构。昇腾 310 整数精度（INT8）**算力可达 16TOPS**，主要应用于边缘计算产品和移动端设备等低功耗的领域。昇腾 910 整数精度（INT8）**算力可达640TOPS**，在业界其算力处于领先水平，性能水平接近于英伟达 A100，支持全场景人工智能应用。
- **昇腾310是一款高效能、灵活可编程的人工智能处理器**，在典型配置下可以输出16TOPS@INT8, 8TOPS@FP16，功耗仅为8W。采用自研华为达芬奇架构，集成丰富的计算单元，提高AI计算完备度和效率，进而扩展该芯片的适用性。全AI业务流程加速，大幅提高AI全系统的性能，有效降低部署成本。

图：昇腾 310 性能

自研华为达芬奇架构NPU	
在8W数据精度下算力可达16TOPS	
高性能3D Cube	
Architecture	• HUAWEI Da Vinci
Computing Engine	• 3D Cube
Performance	• 16 TOPS@INT8 and 8 TOPS@FP16
Max Power	• 8W
Process	• 12nm FFC

图：昇腾 910 性能

自研华为达芬奇架构NPU	
640 TOPS@INT8, 320TFLOPS@FP16	
最大功耗310W	
Architecture	• HUAWEI Da Vinci
Computing Engine	• 3D Cube
Performance	• 320 TFLOPS @FP16 and 640 TOPS @INT8
Max Power	• 310W
Process	• N7+

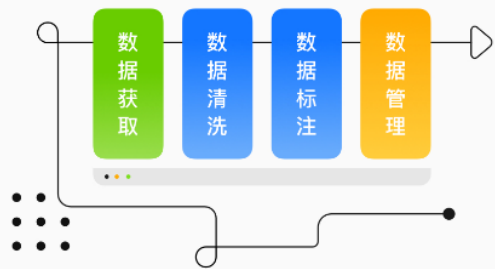
➤ Atlas 系列硬件产品基于昇腾处理器和业界主流异构 计算部件，通过模组、板卡、小站、服务器、集群等丰富的产品形态，打造面向“云、边、端”的全场景 AI 基础设施方案，包括 Atlas 200 AI 加速模块、Atlas 300 AI 加速卡、Atlas 500 智能小站、Atlas 800 AI 服务器、Atlas 900 AI 集群等产品，覆盖深度学习领域推理和训练全流程；以鲲鹏系列CPU+昇腾系列NPU结合，在人工智能计算中心、城市智能人工中枢、通用训练服务器以及视频图像分析等领域，由大到小细节全覆盖

图：昇腾系列AI产品以及应用

类别	产品	AI处理器/加速卡	AI算力	功耗	应用场景
AI模块	Atlas 200 DK开发者套件	昇腾310	11/8/4TFLOPS FP16	典型功耗20W	在端侧实现目标识别、图像分类等，广泛用于智能摄像机、机器人、无人机等端侧AI场景
	Atlas 200 AI加速模块	昇腾310	11/8/4TFLOPS FP16	4 GB:6.5W/ 8GB: 9.5W	
AI加速卡	Atlas 300I推理卡	昇腾310	70/140 TFLOPS FP16	最大72 W/ 150W	提供AI推理、视频分析等功能，支持检索聚类、OCR识别、语音分析、视频分析等场景
	Atlas 300T训练卡	昇腾910	280 TFLOPS FP16	最大300 W	
智能边缘	Atlas 500智能小站	昇腾310	11/8 TFLOPS FP16	有盘配置25W/ 有盘配置40W	具有超强计算性能、体积小、环境适应性强、易于维护等特点，可以在边缘场景广泛部署
	Atlas 500 Pro 智能边缘服务器	最大支持4个Atlas 300I/V Pro	最大420TFLOPS INT8	-	
AI服务器	Atlas 800 推理服务器	最大支持8个Atlas 301I/V Pro	最大1120TFLOPS INT8	-	具有超强计算性能，可广泛应用于中心侧AI推理、深度学习模型开发和训练场景
	Atlas 800 推理服务器	最大支持7个Atlas 302I/V Pro	最大980TFLOPS INT8	-	
	Atlas 800 训练服务器	8*昇腾910	1.76-2.24PFLOPS FP16	最大功耗5.6kW	
	Atlas 800 训练服务器	8*昇腾910	1.76-2.24PFLOPS FP16	最大功耗5.6kW	
AI集群	Atlas 900 PoD	64*昇腾910	14.14-20.40PFLOPS FP16	最大功耗46kW	具有超强AI算力、更优AI能效、极佳AI拓展等特点，可广泛应用于深度学习模型开发和训练
	Atlas 900 AI集群				

- **盘古大模型支持汽车智驾，高效数据闭环及大模型赋能，软件算法加速训练成熟。**盘古大模型涵盖自然语言、多模态、视觉、预测、科学计算等五大领域，支持包括汽车、政务、气象、医学等在内的多个行业；通过全面接入大模型，搭建底层AI算力，华为云搭载大模型深度赋能数据闭环各个环节，如clip内容理解、自动标注、加速模型训练、生成高质量训练场景库与仿真场景库等。

图：大模型工程套件，加速行业落地



数据工程套件

构建自动化数据清洗模型，持续提升数据质量

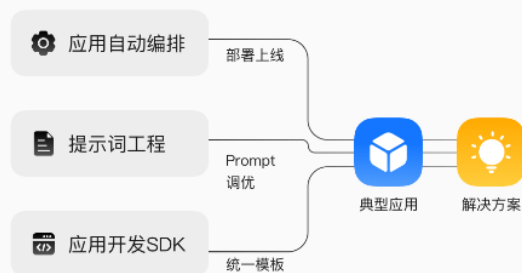
- ✓ 多渠道数据获取
- ✓ 自动化数据清洗
- ✓ Prompt撰写及Rank标注工具
- ✓ 可视化数据管理平台



模型开发套件

一站式开发套件，覆盖模型训练、部署、评测到上线全流程

- ✓ 自监督学习/有监督微调/人类反馈强化学习
- ✓ 在线部署/边缘部署
- ✓ 模型自动化评测
- ✓ 模型管理，持续迭代



应用开发套件

打通大模型行业应用的“最后一公里”

- ✓ LLM驱动，应用自动编排
- ✓ 大模型提示词工程平台
- ✓ 灵活的应用开发SDK

底层软件：CANN异构计算架构

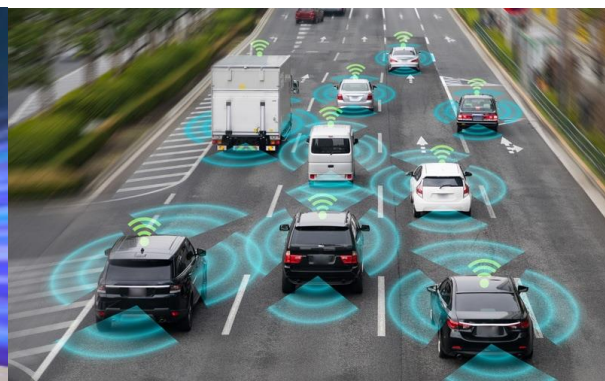
- CANN异构计算架构类似英伟达CUDA，为华为AI计算平台提供开发者工具，便于完善华为软件开发环境和漏洞。



- MindStudio是华为面向昇腾AI开发者提供的一站式开发环境和工具集，致力于提供端到端的昇腾AI应用开发解决方案，使开发者能够在一个工具上高效完成算子开发、训练开发和推理开发。其提供了应用开发、调试、模型转换、网络移植、优化和分析功能。开发者能够高效地完成端到端开发，支持训练和推理业务，可使训练速度提升 25%，推理提速 47%。



- 2022年12月，宁德时代与华为签署合作，将为华为智选车提供动力电池。
- 2023年8月，深蓝汽车携手华为签订合作框架协议，协力探索，加速智能电动时代
- 2023年11月，华为深度合作长安汽车，加速智驾核心技术突破。
- 2024年1月，一汽解放宣布与华为合作的自动驾驶产品将应用于L4低速场景。该产品正处于概念设计阶段，预计将于2025年年底实现量产应用。同月，华为智能座舱与百度地图签署生态合作协议，共创导航出行新体验。
- 2024年2月，东风旗下猛士科技和高端智慧电动汽车品牌岚图与华为正式签署战略合作协议，推动实现各自领域的产业资源共享，围绕用户需求共同打造极致的智能出行体验，共建智能汽车产业生态，通过合作车型在多领域创新探索，加速智能化技术大规模商业化落地。同月，广汽传祺8宗师先锋版上市，搭载首款基于HarmonyOS开发的MPV座舱。

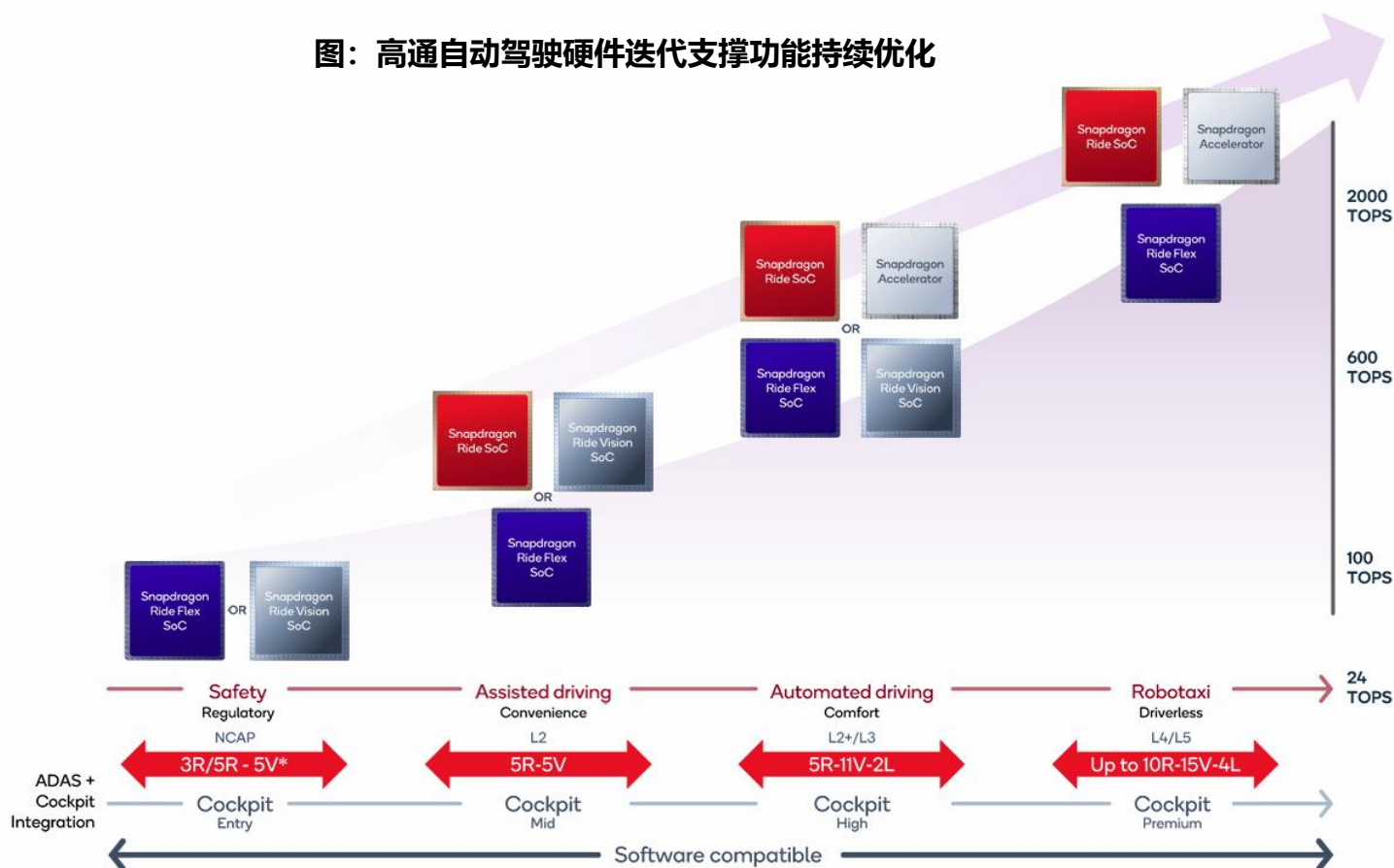


2.3高通：边缘端全自研，由座舱切入快速发力

Snapdragon Ride智驾系统高性能，全开放/可升级

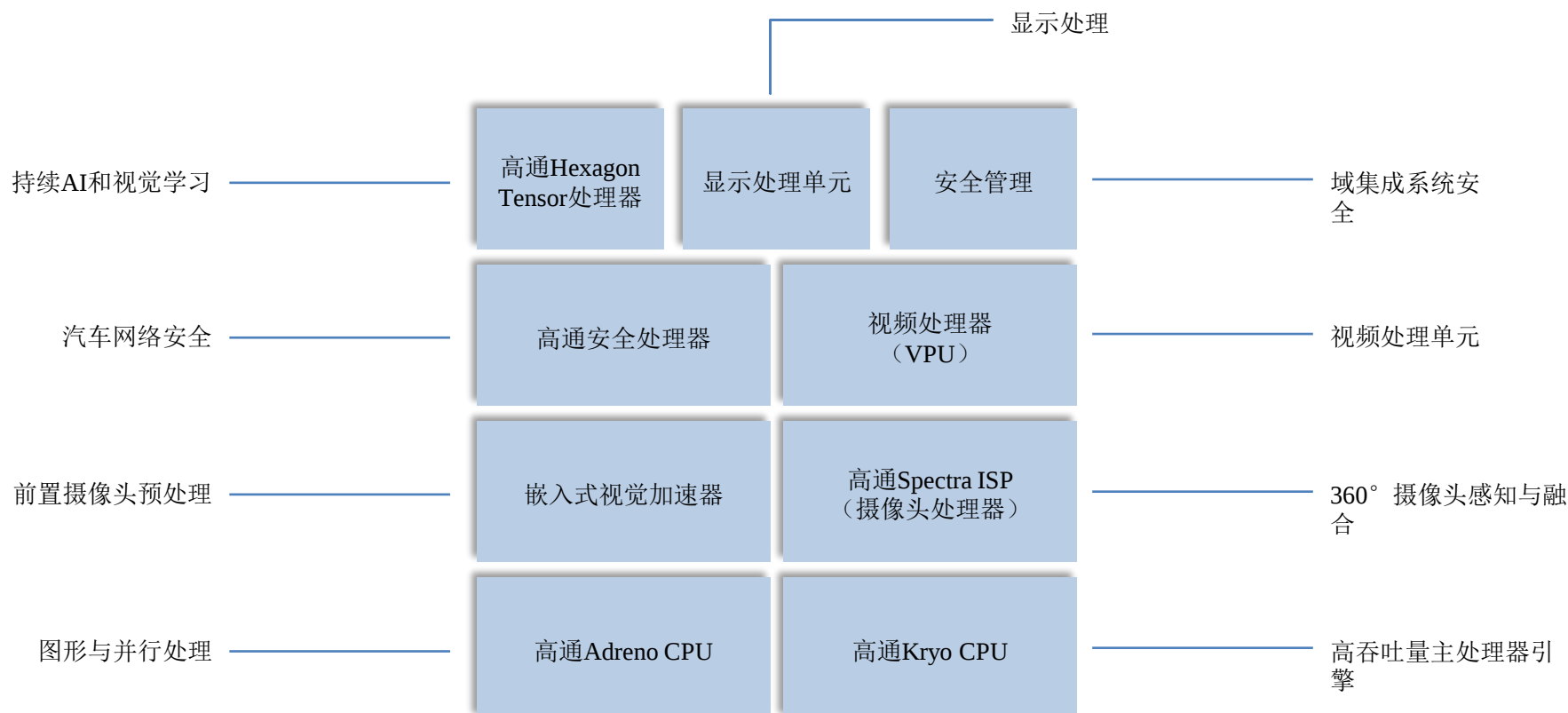
- 2022年初，高通推出Snapdragon Ride™平台最新视觉系统产品，集成了4nm制程专用高性能系统级芯片Snapdragon Ride SoC和Arriver下一代视觉感知软件栈。Snapdragon Ride平台重点关注客户广泛的自动驾驶需求，覆盖可扩展SoC、集成式AD软件栈和开发平台、工具，从而提供面向L2-L3级别自动驾驶的全面解决方案。该系统预计于2024年首搭上市，未来仍将持续迭代升级。

图：高通自动驾驶硬件迭代支撑功能持续优化



- 智驾芯片硬件领域，高通专注于边缘端舱驾一体芯片方案自研，依托于自身在手机等移动设备芯片领域的积累，自研包括Kryo系列CPU，Spectra系列ISP、Adreno系列DPU以及GPU、Hexagon系列DSP，后统一集成以8155/8295等座舱芯片和Ride SoC系列舱驾一体SoC，最大程度发挥汽车芯片自研过程中与手机、通信领域芯片的协同效应。

图：高通智驾芯片IP核单元构成



- **Snapdragon Ride SDK是一个支持平台**，可在 Qualcomm Snapdragon ADAS SoC 和 Snapdragon Ride 硬件平台上为 ADAS 系统构建经过安全认证的应用程序，可最大限度地提高延迟、内存和带宽使用等 KPI，**使得应用程序开发人员能够专注于自定义应用程序逻辑，并使用针对性能和延迟进行预先优化的模块快速开发和部署应用程序。**



深度绑定Tier1，创达/大疆/毫末国内开放式合作

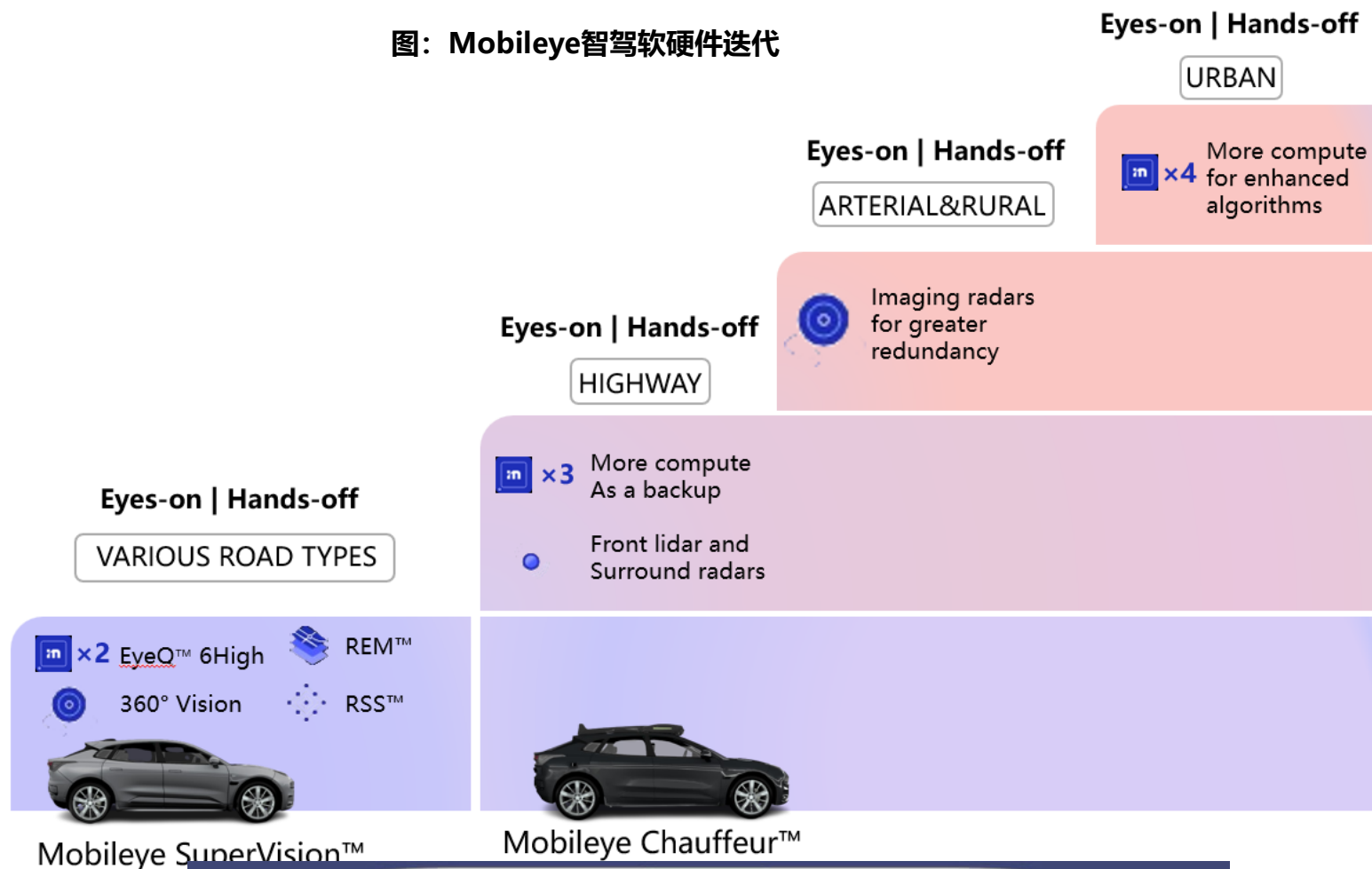
- 国内市场，高通主要以**绑定合作Tier1模式**落地自身智驾软硬件解决方案，边缘端软硬件全配套。
- **中科创达**：高通投资入股创达子公司畅行智驾，目前已推出高通8540的首款自动驾驶域控制器产品，主打L2++级市场；还将在2024年推出基于高通QC8650平台打造的中算力智驾域控产品以及基于QC8795平台打造的首款高性能计算平台产品，并于2025年前完成多平台、全覆盖的产品布局。
- **大疆**：成行平台基于高通骁龙Ride（SA8650P），不依赖高精地图和激光雷达，实现了城区道路点对点 and 快速路段的领航辅助驾驶、跨层记忆泊车。目前已拿到国内车企以及大众的陆续定点。
- **毫末智行**：2021年12月，毫末智行获得了高通创投的战略投资，毫末HPilot3.0和小魔驼3.0均基于高通Snapdragon Ride平台，毫末HPilot3.0所具备的城市NOH功能，单板算力360TOPS。



2.4 Mobileye：聚焦边缘低成本方案，逐步开放

- Mobileye自动驾驶围绕自身全自研Eye Q系列SoC片上系统，由Q1逐步迭代至Q6/Qultra，覆盖L2级别升级至高阶自动驾驶，自研除CPU之外包括MPC\VMP\PMA\XNN等IP核，以极致ASIC芯片配合自身全栈自研底软算法实现L2+级别全球龙头，并快速向高阶智驾领域迭代。

图：Mobileye智驾软硬件迭代



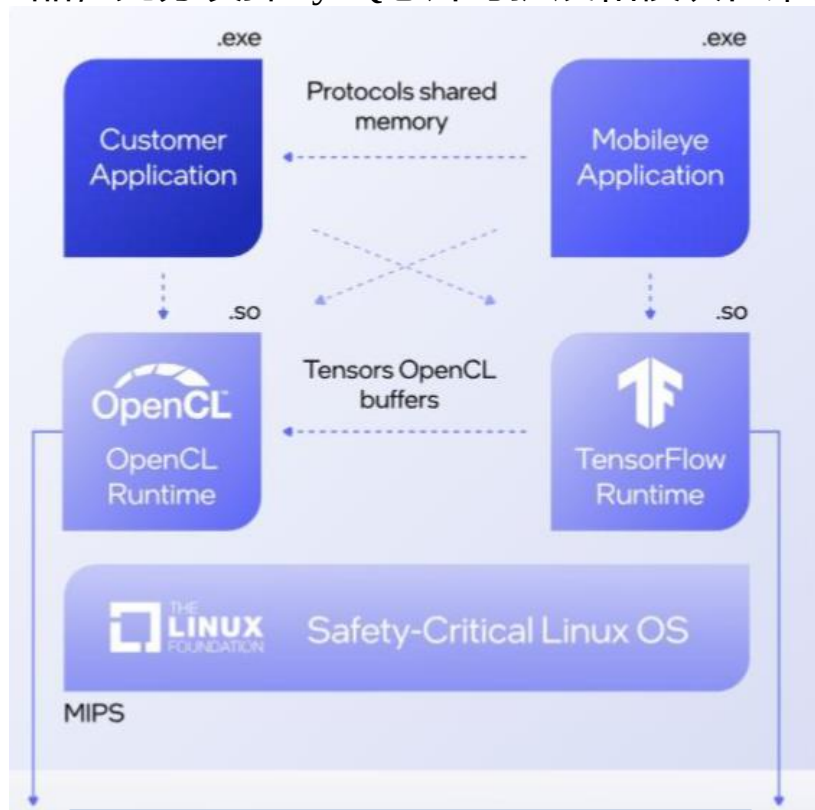
自研IP核驱动EyeQ系列芯片迭代，功能持续优化

- 在高性能计算领域，Mobileye坚持在通用CPU基础上，自研向高密度专用计算领域升级的ASIC芯片MPC\VMP\PMA\XNN等，为不同领域适配开发最适合的IP核以及底软，保证EyeQ芯片能够以更少的算力实现更高更好的智驾体验。

图：Mobileye芯片硬件迭代

	EyeQ1	EyeQ2	EyeQ3	EyeQ4		EyeQ5		EyeQ6		EyeQ Ultra
Configuration				Mid	High	Mid	High	Lite	High	
On market	2008	2010	2014	2018		2021	2021	2023	2024	2023
Claimed autonomous level	Driver Assistance		2	2	2+	2+	4	2	2+	5
Performance (int8 TOPS)	0.0044	0.026	0.256	1.1	2	4.6	16	5	34	176
Power consumption	2.5watt	2.5watt	2.5watt							
CPU								2 core,8 thread	8 core,32 thread	12 core,24 thread
Memory								LPDDR4(X)	LPDDR5	LPDDR5X
Semiconductor node	180 nm CMOS	90 nm CMOS	40 nm CMOS	28 nm FD-SOI		7 nm FinFET		7 nm FinFET		5 nm
Algorithms & neural networks			•Vehicle's Rear •Pedestrians •Lane Markings •Semantic Free Space •Traffic signs	•3D Vehicles •Next-Gen Lane Markings •Road Markings •Traffic Lights •Relevance of Traffic Lights •Next-Gen Semantic Free Space •Road Profile •General Objects •Hazards •Animals •Path Prediction •Road Edges		•3D Vidar (PseudoLidar) •Pixel-Level Scene •Segmentation •Full Image Detection •Surface Segmentation •Lane Semantics •Road Users Trajectory Prediction •Parallax Net •Vector Field, •Multi-Camera Bird's Eye View Network •Road Users Understanding				

- **EyeQ Kit**使 OEM 能够在 Mobileye 核心技术（包括计算机视觉功能、REM™ 众包地图和基于 RSS 的驾驶策略）的基础上实现其市场产品的差异化，基于 Mobileye 成熟的系统构建自己的应用程序（而不是从头开始或使用通用处理器），以降低集成复杂性和成本，加速上市时间。
- **EyeQ Kit SDK**使OEM能够在 Mobileye 高效且可扩展的 EyeQ SoC 之上开发和部署自己的差异化市场产品，充分发挥EyeQ芯片可扩展和模块化架构的优势，提供不同集成度的解决方案。



图：Mobileye软件开发环境

- OpenCL 运行环境
- TensorFlow 支持
- 标准Linux，支持第三方中间件和库
- 基于X86的开发平台
- 支持完整的开发周期：从功能启动到部署和性能调整

EyeQ Kit所基于的成熟底软系统

EyeQ Accelerators

- Mobileye早期在全球L2级别及以下市场占绝对主导地位，市场份额一度超70%，主要系其产品解决方案性价比高，同时对原有整车底盘架构变化要求较小，易被OEM接受；后因黑盒供应阻碍OEM自身研发和迭代效率，陆续被英伟达、地平线等供应商取代，当前Mobileye合作OEM主要为海外主机厂以及国内极氪等品牌，未来随公司启动开放式战略合作，有望进一步拓展朋友圈。
- 国内Mobileye积极与经纬恒润展开配套合作，公司预计于2024Q2落地搭载Eye Q6芯片的ADAS解决方案。

图：Mobileye车企合作伙伴

	EyeQ3	EyeQ4		EyeQ5	
Configuration		Mid	High	Mid	High
On market	2014	2018		2021	2021
Claimed autonomous level	2	2	2+	2+	4
OEM Implementations	•Audi's Traffic Jam Pilot •2019 AudiA8 •Cadillac CT6(2017-2019) •GM's Supercruise •Ford F-150(2018-2020) •Nissan'sProPilot •2020 Nissan Skyline •2022 Nissan Ariya •Tesla Autopilot Hardware 1 •Model S & X (Sep-2014-Oct-2016) •Volvo's PilotAssist 1, 2 & 3	•BMW Driving Assistant Professional •X5 (2019-2020) •3 Series(2019-2020) •Ford Mustang Mach-E(2021) •Ford F-150(2021) •Li Xiang One •NIO Pilot •Nio ES8 •Nio ES6 •Nissan's ProPilot2.0 •VW TravelAssist •Volkswagen Passat (2019-2020) •Volkswagen Golf 8 (2020)		•BMW's Personal Copilot •BMW iX (2022) •Zeekr Copilot(SuperVision™) •Zeekr 001(2021)	

2.5、地平线：征程+天工开物，自下向上突破

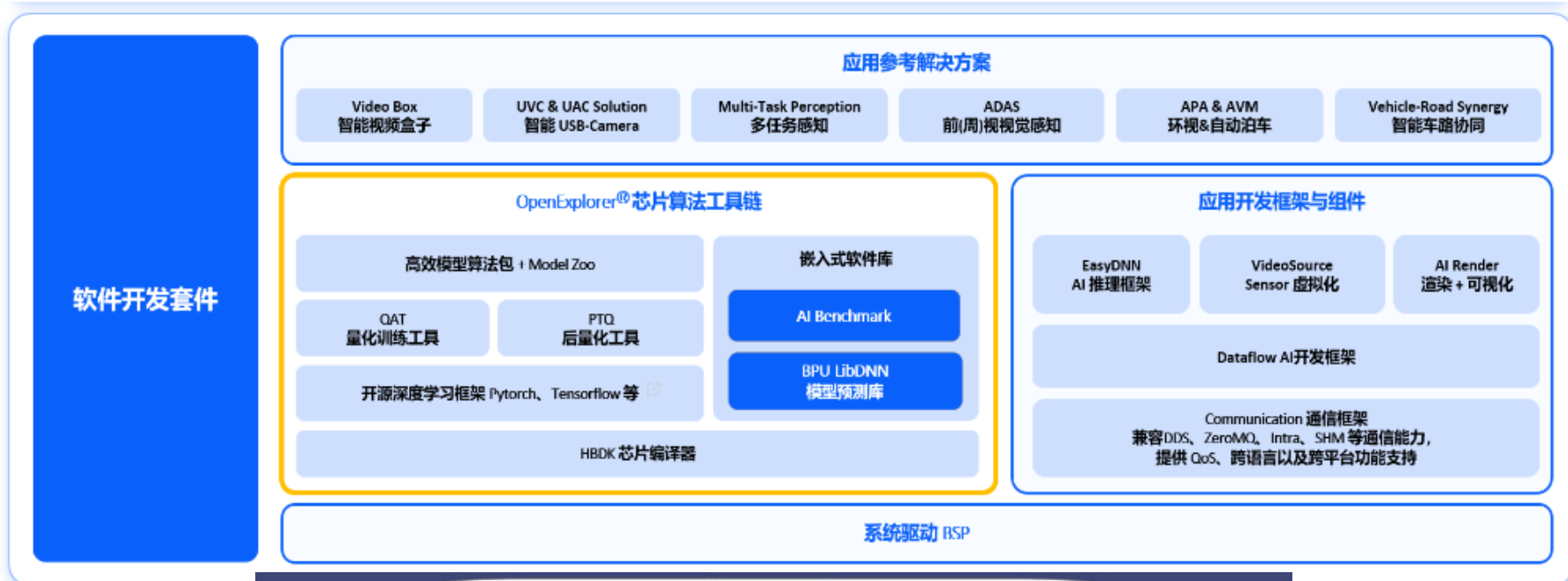
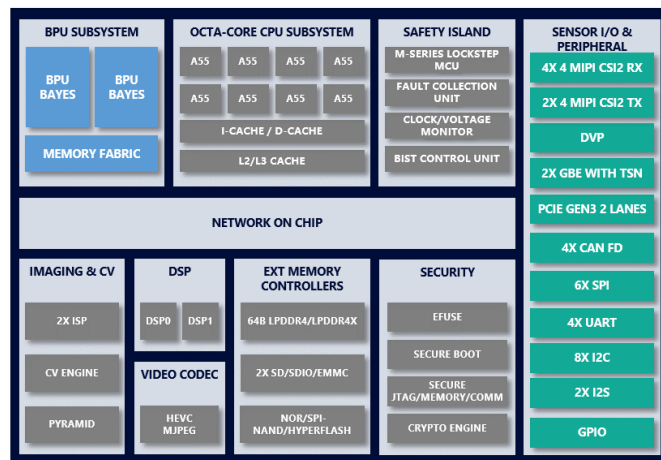
- 得益于前瞻性的**软硬结合理念**，地平线自主研发兼具极致效能与高效灵活的智能计算方案，可面向智能驾驶以及更广泛的智能物联网领域，提供包括**高性能智能计算芯片**、**开放工具链**、**丰富算法样例**等在内的全面赋能服务。

图：地平线之家底层软硬件以及上层应用梳理



征程J5高性能、低功耗芯片奠基，完善工具链配合

- J5是公司2021年面向自动驾驶AUTO领域推出的低功耗、高性能芯片算法处理器。计算能力方面，J5内置了八核Cortex-A55 CPU, 双核Vision P6 DSP及双核BPU 芯片算法加速器，其中BPU是地平线自研的加速核，可提供128TOPS的BPU (AI) 算力。
- 同时，地平线还提供适配J5的完整的边缘芯片算法落地解决方案，可以把浮点模型量化为定点模型，并在芯片上快速部署自研算法。



以征程系列芯片为核心，自研全套工具链赋能开发

地平线智驾软硬件平台开放赋能下游外部客户系统开发：

硬件领域强大异构能力：

依托于公司强大的异构能力，以自研AI加速计算单元BPU以及CPU/DSP和ISP等IP，并采用PCIe高速信号接口以及双路千兆以太网保障数据信号传输，提供有力硬件支持。

软件领域完善工具链：

- 1) 提供完整的软件 API 和开发套件，适配主流的训练框架Caffe、MXNet、TensorFlow和PyTorch可通过基础算法和参考算法赋能客户，最终实现产品级算法。
- 2) 开放 ISP Tuning 工具赋能客户，使其可自行调试 Camera。



朋友圈持续扩容，主机厂批量参股投资强化合作

- 地平线自成立以来，已获得上汽集团、广汽资本、长城汽车、比亚迪、一汽集团等中国汽车制造企业以及英特尔、SK海力士、宁德时代、立讯精密等供应链上下游企业的投资。
- 地平线与安波福 (Aptiv PLC) 及其子公司风河 (River) 结成战略合作伙伴关系，开发专为中国民用汽车制造商量身定制的全集成硬件和软件解决方案。
- 地平线机器人宣布与比亚迪达成最新合作，为比亚迪汽车配备地平线最新一代汽车级机器人计算解决方案 - Journey™ 5。比亚迪参与了地平线此前的融资，是首批表达征途5生产意向的顶级车企之一。
- 大众汽车通过CARIAD与地平线机器人公司成立合资公司加强在中国的开发能力，其中CARIAD将持有60%的多数股权，双方计划共同为中国开发尖端、高度优化的全栈 ADAS/AD 解决方案。
- 地平线获得奇瑞汽车战略投资，并已完成融资交割。奇瑞全新高端智能电动汽车平台E0X将采用地平线Journey™3产品作为算力基石。



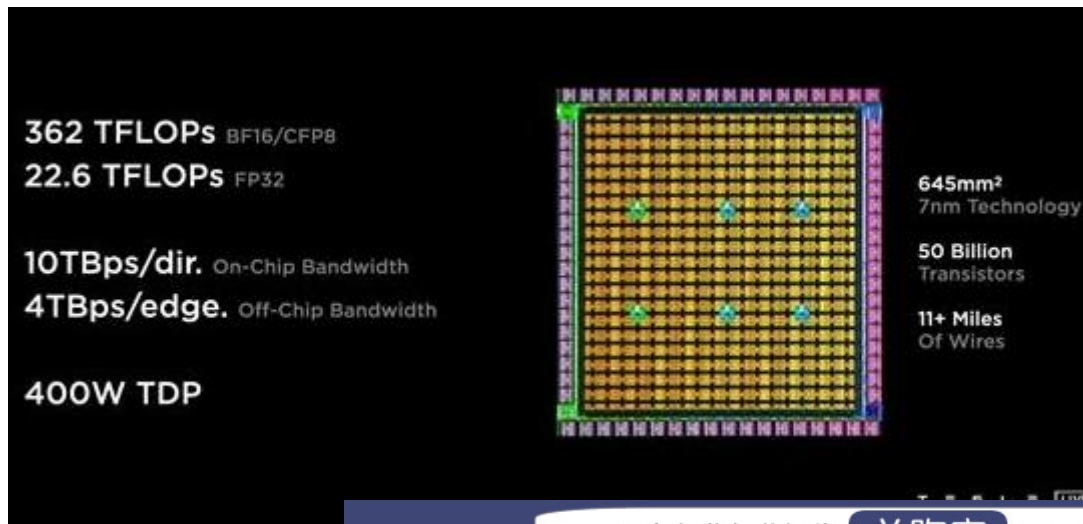
三、下游OEM玩家如何做？

- 特斯拉自动驾驶硬件解决方案持续升级，2016年10月由Mobileye黑盒转为英伟达开放，逐步掌握智驾算法后，2019年全自研FSD芯片上车，2024年新一代方案预计上车，芯片再升级。
- **自研智驾芯片保障成本与性能双领先。** 特斯拉自2016年2月组建智驾芯片研发团队，2019年4月FSD芯片正式搭载上车，单车搭载2颗FSD芯片；每颗配置 4 个三星 2GB内存颗粒，单 FSD总计8GB，同时每颗 FSD配备一片东芝的 32GB闪存以及一颗 Spansion的 64MB NOR flash 用于启动。凭借NPU针对AI计算更好的专业适用性，**3.0时代FSD芯片以14nm制程+260mm²面积实现144TOPS算力，相比英伟达12nm制程+350mm²支持30TOPS AI算力更为领先。**

图：特斯拉硬件解决方案迭代

版本		HW1.0	HW2.0	HW2.5	HW3.0	HW4.0
推出时间		2014 年 9 月	2016 年 10 月	2017 年 7 月	2019 年 4 月	2023Q2
计算平台		Mobileye EyeQ3	Nvidia drive PX2	Nvidia drive PX2+	Tesla FSD	Tesla FSD
核心处理器		1-Mobileye EyeQ3 1-Nvidia Tegra3	1-Nvidia Parker SoC (1-Nvidia Pascal GPU+1-Infineon TriCore CPU)	2-Nvidia Parker SoC (1-Nvidia Pascal GPU+1-Infineon TriCore CPU)	2-FSD 1.0 (14nm制程, 12个CPU核+2 个 NPU)	2-FSD2.0 芯片 (7nm制程, 20 个 CPU核+3 个 NPU)
算力 (TOPS)		0.256Tops	12Tops		144Tops	720Tops
传感器配置	摄像头	1 颗 EQ3 系列前视摄像头	8颗摄像头，3颗前置+2颗侧置前视+2颗侧置后视+1颗后置（分辨率为120万像素）			12个摄像头，其中1个为备用（分辨率升级至500万像素）
	毫米波雷达	1个 (160m)		1个 (170m)		1个
	超声波雷达	12个 (5m)		12个 (8m)		

- **可扩展+强计算，特斯拉D1性能表现业内领先。** 1) **基础性能方面**，特斯拉D1由台积电代工，采用7nm制程工艺，芯片面积为645mm²，小于英伟达A100（826 mm²）；D1芯片拥有多达354个训练节点，是特斯拉专门设计的特别用于AI训练相关的8×8乘法的芯片，**浮点计算性能FP32算力22.6TFLOPS(英伟达A100为19.5)**，对应热功耗仅为400W；D1芯片集成四个64位超标量CPU核心，支持完整向量以及矩阵计算，灵活性远超众核架构的GPU。2) **高带宽+低延迟保障强可扩展性**：D1芯片采用带宽最高可达10TB/s的“延迟交换结构”进行互连，加速数据传输。D1芯片运行频率2GHz，拥有440MB SRAM，是存算一体架构，降低过程数据缓存压力。Tile角度，每个D1训练模块由5x5的D1芯片阵列排布而成，以二维Mesh结构互连，片上跨内核SRAM达11GB，每个训练模块外部边缘的40个I/O芯片达36/10TB/s的聚合/横跨带宽，保障信息传输过程的低损耗。
- 图：特斯拉D1 Chip



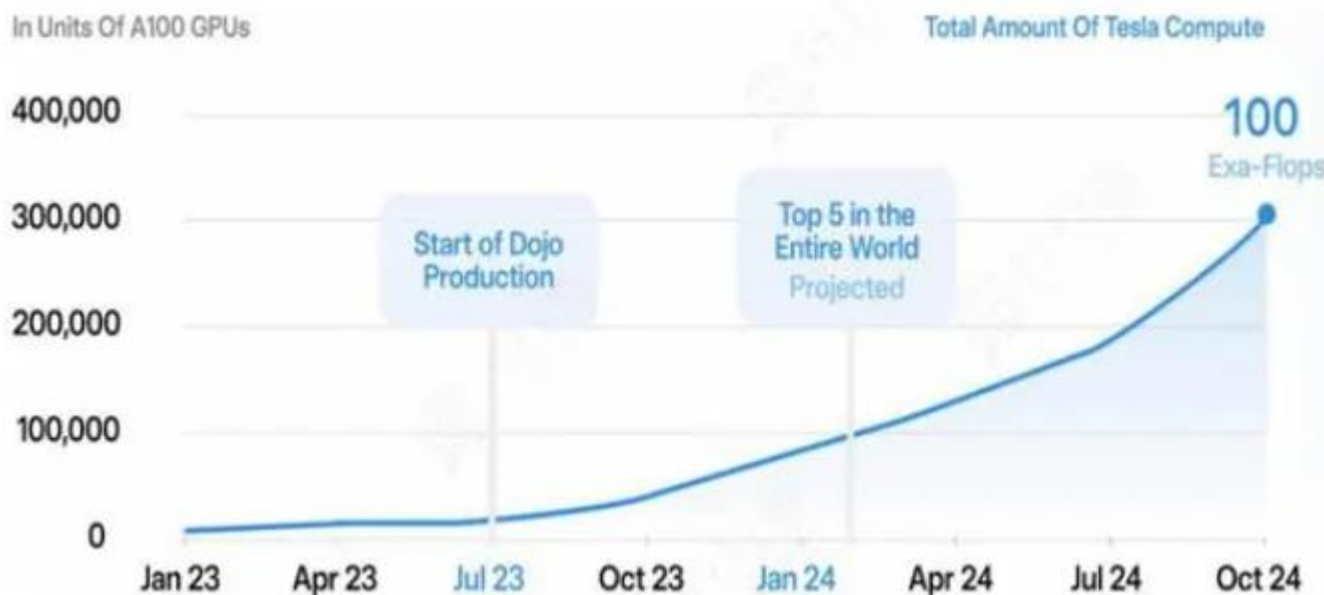
图：特斯拉D1 Tile



Dojo超算中心高算力+低功耗+高带宽，行业领先

- Dojo从底层开始由Core、D1、Tile、Tray、Cabinet、ExaPOD构成：**D1芯片**由354个Core构成，强算力+高带宽；**Tile**由25个D1芯片打造，**通过InFO_SoW封测技术实现低延迟、高带宽**，4边对外传输带宽均为9TB/s；**ExaPOD**由3000个D1芯片构成，在BF16精度下算力高达1.1 EFLOPs。特斯拉预计将Dojo打造成全球五台最先进的超级计算机之一，总算力达到100 Exa-Flops。
- 特斯拉为追求Dojo极致算力性能，在算法通用性和生态角度做了较大牺牲，为此，特斯拉为Dojo全新编译一套完整软件栈，包括Dojo编译器、Dojo Ingest Pipeline、Dojo Runtime和Dojo Library，**实现对神经网络模型的自动调优和并行化，并使Dojo支持ensorFlow、PyTorch等流行的深度学习框架。**

图：特斯拉Dojo超算中心算力以及数据量



软件栈	用处
Dojo编译器	将Dojo大型分布式系统视为加速器，实现对神经网络模型的自动优化和并行化
Dojo Ingest Pipeline	快速将视频数据转换为神经网络的输入格式，提高数据加载效率
Dojo Runtime	管理Dojo系统的资源分配和调度，实现多个任务的并发执行
Dojo Library	提供了常用的神经网络算法和操作，支持TensorFlow、PyTorch等流行的深度学习框架

- 国内OEM厂商跟进特斯拉，自研边缘端智驾芯片，基于自身需求和算法架构深度定义芯片规格，挖掘芯片潜力，并根据车企自身节奏快速迭代，降低冗余成本，构建软硬件一体化整合高壁垒。
- **第一类：头部新势力玩家选择全自研AI智驾芯片。**蔚小理先后从Mobileye黑盒供应切换至英伟达（高端）/地平线（低端）开放式供应体系，未来三家车企均规划采用自研芯片。蔚来于2023年发布神玢芯片NX9031，采用5nm先进工艺，包含32核CPU，对标4*OrinX；小鹏自2020年开始在中美两地同步布局芯片，理想也招募团队聚焦智驾NPU芯片开发。
- **第二类：股权合作合资公司投入，开发智驾芯片。**以吉利为代表，吉利汽车关联公司亿咖通与安谋中国成立芯擎科技，并于23年底量产7nm车规级智能座舱芯片「龍鷹一号」，并预计于**24H2推出对标英伟达Orin的智驾芯片，最大算力256TOPS**。其余车企方面，车企股权战投芯片公司案例较多，比亚迪、长城、上汽、广汽、长安等投资地平线，东风、上汽投资黑芝麻智能。

图：亿咖通双芯域控支持L3级智驾



图：蔚来自研神玢芯片，强调领先NPU\CPU能力



四、投资建议与风险提示

■ **投资建议：汽车AI智能化转型大势所趋，硬件为基石，看好布局智驾硬件的OEM长期竞争力。**

- **全行业加速智能化转型，产业趋势明确。** 下游OEM玩家+中游Tier供应商以及上游原材料厂家均加大对汽车智能化投入，大势所趋；智驾核心环节【软件+硬件+数据】均围绕下游OEM展开，数据催化算法提效进而驱动硬件迭代。因此，以AI芯片为核心的智驾硬件是OEM中长期核心竞争力的重要构成，参考手机行业，核心硬件是玩家【成本控制能力+品牌护城河】的终局竞争要素。
- **国内OEM以软件为先，硬件其次，加速进化。** 头部新势力玩家紧随特斯拉引领本轮智驾技术变革，全自研智驾芯片有望于2025~2026年流片量产，构筑品牌核心竞争力以及产品重要卖点。
- **看好智驾头部车企以及智能化增量零部件：** 1) 华为系玩家【长安汽车+赛力斯+江淮汽车】，关注【北汽蓝谷】； 2) 头部新势力【小鹏汽车+理想汽车】； 3) 加速转型【吉利汽车+上汽集团+长城汽车+广汽集团】； 4) 智能化核心增量零部件：域控制器（德赛西威+经纬恒润+华阳集团+均胜电子等）+线控底盘（伯特利+耐世特+拓普集团等）。

- **智能驾驶相关技术迭代/产业政策出台低于预期。**若智能驾驶相关技术迭代节奏低于预期，可能会对消费者对智驾的认知和接受度产生影响，政策出台节奏低于预期也可能影响节奏。
- **华为/小鹏等头部车企新车销量低于预期。**头部车企智驾新车销量表现低于预期，可能拖累智驾渗透率提升，对板块产生负面影响。

免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后6至12个月内行业或公司回报潜力相对基准表现的预期（A股市场基准为沪深300指数，香港市场基准为恒生指数，美国市场基准为标普500指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证50指数），具体如下：

公司投资评级：

买入：预期未来6个月个股涨跌幅相对基准在15%以上；

增持：预期未来6个月个股涨跌幅相对基准介于5%与15%之间；

中性：预期未来6个月个股涨跌幅相对基准介于-5%与5%之间；

减持：预期未来6个月个股涨跌幅相对基准介于-15%与-5%之间；

卖出：预期未来6个月个股涨跌幅相对基准在-15%以下。

行业投资评级：

增持：预期未来6个月内，行业指数相对强于基准5%以上；

中性：预期未来6个月内，行业指数相对基准-5%与5%；

减持：预期未来6个月内，行业指数相对弱于基准5%以上。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所
苏州工业园区星阳街5号
邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>

东吴证券 财富家园