



独立GPU需求急跌 产业竞争加剧

本报记者 陈炳欣

2021年还在到处缺货的GPU,现在市场却出现急剧下滑。根据 Jon Peddie Research 近日披露的数据,独立显卡的出货量在 2022 年第三季度创下近 20 年来的新低,英伟达、AMD 和英特尔总共出货了约 1400 万颗用于台式机和笔记本电脑的独立图形处理器,同比下降 42%,集成 GPU 与移动 GPU 的市场需求同样不振。市场急跌对于整个 GPU 行业都将产生重要影响,未来企业间的竞争将更加激烈。

消费电子与挖矿需求下降, GPU 市场急跌

近日,市场研究机构 Jon Peddie Research 发布最新 GPU 市场数据统计报告,2022 年第三季度 PC 用 GPU 出货量(包括集成和独立显卡)为 7550 万颗,环比下跌 10.3%,同比下跌 25.1%,其中独立 GPU 同比下降达 42%。集成 GPU 和移动 GPU 市场需求同样不佳,移动 GPU 出货量同比下降 30%。GPU 市场创下了 2009 年以来的最大跌幅。

Jon Peddie Research 认为,之所以在年度转换之间会发生这样的变化,主要源于两点原因:一是全球消费热潮冷却,二是大规

模挖矿行动的终结。新冠病毒感染期间对显卡的需求显著增加,因为大家有更多时间在家玩游戏或者办公,而另一些人则试图开采比特币等虚拟货币以获得收益。但是,现在新冠病毒感染对经济社会的影响逐渐降低,虚拟货币的下跌也导致大规模挖矿行动受到抑制。这些对于 GPU 的需求都造成了极大的影响。

事实上,2021 年年底以来,GPU 市场就开始承压,GPU 价格从 2021 年 12 月就不断下跌。缺货潮期间,显卡价格往往是英伟

2022 年第三季度 PC 用 GPU 出货量 7550 万颗,同比下跌 25.1%,创下了 2009 年以来的最大跌幅。

达、AMD 两家公司市场指导价格的两倍到三倍,但随着下游应用需求疲软,目前多数显卡都可以用市场指导价格或低于市场指导的价格购买到。

这种市况自然对企业业绩形成了一定挤压。财报显示,英伟达 2023 财年第三季度(截至 2022 年 10 月 30 日)收入为 59.31 亿美元,同比下降 17%,环比下降 12%;净利润为 6.8 亿美元,同比下降 72%;其中游戏业务收入为 15.7 亿美元,同比下降 51%,环比下降 23%。

英特尔对计算系统和图形事业部进行拆分合并,希望与原有部门形成更好的协同效应。

将提供 2~3 倍的实际使用性能。而此前,壁仞科技发布的通用 GPU BR100,16 位浮点算力达到 1000T 以上,8 位定点算力达到 2000T 以上。芯动科技发布的“风华 2 号”GPU,像素填充率达 48GPixel/秒,FP32 浮点算力 1.5TFLOPS,性能领跑国产桌面、笔记本电脑和工控机赛道。

对此,芯动科技首席 GPU 科学家杨喜乐指出,现代 GPU 的涉及面广泛开发应用方面有很多技术壁垒,对开发团队能力要求很高;要拿出一款对标行业先进水平且流畅兼容各种应用的图形 GPU,需要具有跨多领域平台能力和大量流片交付经验的复合型技术团队,要不断进行 GPU 底层技术、算法架构和软硬件创新。GPU 行业已经形成稳定的垄断格局,市场下行对本土 GPU 厂商来说,必将承受更多压力,但也将为本土 GPU 厂商提供新的发展机遇。

GPU 部门进行了重新定位。图形芯片部门的业务内容与英伟达大致重叠,包括向消费市场出售 GPU 芯片,用于游戏、渲染等,以及向企业、数据中心客户出售加速计算芯片,用于 AI 计算、云游戏等。此次调整意味着英特尔将分别面向消费、企业级的 GPU 产品线,进一步与英特尔原有部门整合,有望形成更好的协同效应。

国际大厂竞争的同时,本土 GPU 厂商也在积极行动。事实上,本轮市场下行对本土 GPU 厂商来说,既有压力也有机遇。在“东数西算”工程的支持下,近年来,国内涌现出一批 GPU 企业。2022 年 12 月 20 日,天数智芯举行发布通用 GPU 推理产品“智铠 100”。根据天数智芯产品线总裁邹翀介绍,智铠 100 产品卡,可提供最高 384TOPS@int8、96TFlops@FP16、24TFlops@FP32 的峰值算力,相较于市场上现有主流产品,智铠 100

此次出售 HiRel DC-DC 转换器业务,将能够扩大公司在其他高可靠性市场的研发投入。

Micross 是为高可靠性市场提供关键型微电子元件和服务的全球领先供应商。英飞凌科技美洲区总裁 Bob LeFort 谈道:“英飞凌很高兴与 Micross 达成协议,可以为 HiRel DC-DC 转换器业务提供更具战略意义的基地。”

赛迪顾问集成电路高级分析师杨俊刚表示,英飞凌出售高可靠性 DC-DC 转换

器业务主要是为了集中精力,加快核心半导体产品的研发。目前,市场对电源管理、功率器件等半导体产品的需求依旧旺盛,卖掉高可靠性 DC-DC 业务可以减轻英飞凌在研发、生产方面的压力,向通用型产品转移。随着新型基础设施建设、轨道交通建设、新能源汽车等新型领域的快速发展,市场对于功率半导体产品的需求量还将持续增加。Micross 收购 DC-DC 业务将增加其业务范围,配合原有产品,可以提供更全面的组合方案。

台积电乐观看待下半年市况

本报记者 姬晓婷 沈丛

1 月 12 日,台积电公布了 2022 年第四季度财务报告,总营收折合人民币约 1379 亿元,同比增长 42.8%,与第三季度相比增长 2%;净利润约 655.8 亿元,同比增长 78%,比第三季度增长 5.4%。2022 年第四季度毛利率达到 62.2%,5 纳米与 7 纳米制程芯片销售金额分别占到台积电当季销售总额的 32%和 22%,先进制程销售额超过台积电当季总营收的一半。

台积电下调第一季度预期

1 月 12 日的法说会上,台积电总裁魏哲家预测,2023 年上半年,全球半导体库存水位将大幅降低,并逐渐平衡到健康水平。他预计,2023 年半导体产业市场产出将下滑 4%,晶圆代工产业则减少 3%。台积电 2023 年上半年收入也将同比出现个位数的下降,但公司仍会继续扩大产品组合种类及目标市场。至于库存调整何时结束,魏哲家预测称,全球半导体市场有望在 2023 年下半年实现复苏,届时台积电营收也将有所增长。

关于台积电 2023 年第一季度的业务,台积电 CFO 黄仁昭下调了其预期,预计第一季度收入将在 167 亿美元至 175 亿美元之间,毛利率约为 53.5%至 55.5%。黄仁昭表示,调整预期的原因主要是当前工厂产能利用率较低,客户进一步调整库存水平和不利的汇率。此外,研发费用占到台积电 2022 年净收入的 7.2%,而随着公司持续加大技术投入,2023 年研发费用将同比增长约 20%,约占 2023 年总营收的 8%至 8.5%。

关于 2023 年库存水位的调整,当前业界的声音基本一致,认为当前正处于各厂商去库存阶段,晶圆厂产能利用率不足。芯谋研究企业服务部总监王笑龙在接受记者采访时表达了相似的观点:“2023 年下半年市场将会回暖,虽然很难判断明确的拐点,但能够确定 2023 年半导体市场走势将相对平缓。”

2nm 工艺将更换晶体管架构

关于台积电先进制程情况,魏哲家表示,台积电 7 纳米、6 纳米制程的产能利用率低于 3 个月前的预期。他认为,7 纳米、6 纳米的低产能利用率将持续到 2023 年上半年,因为半导体供应链库存需要几个季度才能重新回归正常水平。魏哲家表示,当前台积电正在与客户密切合作,开发差异化技术,以推动来自射频连接等领域的消费。针对投资者提出的许多模拟产品并不需要先进工艺制程的疑问,魏哲家表示,越来越多的计算功能将被应用到产品中,包括 WiFi、射频芯片等,都需要非常高的计算性能,同时要求更低的功耗,而只有领先的工艺节点才能满足这些要求。

台积电 3 纳米产品于 2022 年第四季度投产。魏哲家预计,在高性能计算和智能

手机等应用的推动下,3 纳米产品将在 2023 年实现市场平稳增长,并表示客户将在 2024 年、2025 年及之后对 3 纳米产品有强劲的需求。结合台积电在 7 纳米、6 纳米等制程产品出现的产能利用率降低的现状,有投资者对台积电如何避免类似现象出现在 3 纳米节点上提出了疑问。针对这一问题,魏哲家回应称,他认为当前芯片需求减弱主要是受到新冠病毒感染的影响,而这种情况很大程度上不会再生。

在此次法说会上,台积电透露,2 纳米工艺也正在积极研发中,预计 2025 年量产。与此同时,台积电将在 2 纳米工艺中首次更换晶体管架构,从 FinFET 转至 GAA。虽然这项变动台积电“落后”了三星整整三年,但也不难看出,台积电为了保证芯片良率,在更换晶体管架构方面比较保守。据了解,三星首批搭载 GAA 晶体管架构的 3 纳米芯片良率只有 10%~20%,过低的良率导致三星 2023 年的大部分订单都被台积电抢走,竞争相当激烈。而台积电计划在 2025 年的 2 纳米制程工艺上用上 GAA 晶体管架构,也有较大的机率遭遇这样的瓶颈期。

拟在日欧建厂

关于此前备受关注的台积电资本预算情况,黄仁昭介绍称,鉴于近期市场的不确定性,台积电将适当收紧资本支出。2022 年,台积电的资本支出在 363 亿美元;而 2023 年,台积电资本预算预计在 320 亿美元至 360 亿美元之间,其中约有 70%分配给先进工艺技术,20%用于专业技术,10%用于封装、掩膜等方面。而在资本支出方面,黄仁昭重点介绍了对先进制程的投资,他表示,用于 3 纳米、2 纳米的研发支出将是先进工艺技术方面投资的重点,设计新的晶体管架构、购买新的设备等都是投资的主要目标。

关于台积电未来的全球扩张进程,魏哲家表示,台积电在美国的产能将实现增长。台积电正在推进建设亚利桑那州的半导体工厂, Fab1 计划在 2024 年进行 N4 制程生产, Fab2 计划在 2026 年生产 N3 制程。台积电也在考虑在日本建立新工厂,发展 12 纳米/16 纳米和 20 纳米/28 纳米工艺,预计在 2024 年底实现批量生产。中国台湾地区则将布局台积电最先进的产能, N3 产品已经在台南实现量产,计划 2025 年在新竹厂和台中厂量产 N2 制程产品。黄仁昭表示,28 纳米及以下制程的海外产能占比将在未来 5 年或更长的时间内占到台积电总体产能的 20%或更多。除此之外,台积电称正在与欧洲客户合作,评估建厂的可能性,将聚焦在汽车特殊工艺。

至于在不同地区建厂的成本,黄仁昭表示,在美国建厂的建设成本约比在中国台湾地区建厂高出 4~5 倍。高昂的成本包括根据职业安全和健康法规获得许可的人工成本、近年来的通货膨胀成本,以及人员和学习成本。

边缘人工智能:MCU 发展新方向

本报讯 记者陈炳欣报道:随着云计算与物联网的发展,两者相结合的边缘人工智能技术近年来开始受到越来越多的重视。近日,亚德诺半导体(ADI)公司发布了一款带有神经网络加速器的超低功耗人工智能微控制器 MAX78000。该产品面向边缘人工智能应用,可以支持采取电池供电的嵌入式物联网设备在边缘设备中实现快速且低功耗的人工智能推理。ADI 表示,与软件方案相比,采用这种方案可使复杂的 AI 推理能耗降低到前期方案的 1/100 以内。

简单来说,边缘人工智能是指在边缘设备上实现的 AI 运算。以往大多数尖端的人工智能流程都是在云上执行的,因为它们需要大量的计算能力。但随着数据的增长速度对本地数据计算和本地数据存储的需求变得更加迫切,大量数据操作逐渐转向本地进行,这在边缘广泛采用人工智能提供了理由。

“各种温度传感器、加速度传感器、声学传感器等,这些传感器联系了我们所处的真实的物理世界和看不见摸不着的数字世界,构成一组组边缘节点。那么,智能边缘就是这些边缘节点也具备智能,可以提供高级数据分析、场景感知、实时决策、自组织与协同等服务。”ADI 中国技术支持中心高级工程师辛毅表示。相比于云端 AI,由于不需要将数据上传至云端,边缘

AI 具备实时性好、带宽资源要求低、隐私性高等特点,特别适合物联网应用。和云端 AI 相同,边缘 AI 也具备 AI 的共性特征。

根据辛毅的介绍,本次发布的 MAX78000 具备特殊的架构,由两个微控制器内核 (ARM Cortex M4F 和 RISC-V) 加上一个卷积神经网络 (CNN) 加速器构成。这一架构针对边缘进行了高度优化,数据的加载和启动由微控制器内核负责,而 AI 推理由卷积神经网络加速器专门负责。借助在芯片内部集成的卷积神经网络加速器,可以完成 AI 推理,不需要再通过互联网上传。微控制器内核负责完成加载和启动之后,不需要任何操作,因此功耗也很低,只要使用电池就可以给整个系统供电。借助于这两个硬件的分工合作,MAX7800X 系列非常适合边缘 AI 的应用。

在现实中,边缘人工智能的应用越来越广泛。工厂操作人员可以将信息技术整合到操作流程中,以提高能源效率、产量和正常运行时间;汽车制造商可以推动自动驾驶辅助系统由相对基础的级别向高级自动驾驶系统发展,提高安全性和生产力;医疗机构甚至能够预测健康问题,预防潜在风险,而不是等患者发病或者发展到危重症之后再进行治疗。此外,在森林火灾应用、公路、铁路或者是大坝坝方监测等应用场景,边缘 AI 都可以得到很好的应用。

聚焦核心产业 英飞凌出售 DC-DC 转换器业务

本报讯 记者许子皓报道:1 月 10 日,英飞凌宣布与 Micross 签订协议,出售旗下 HiRel DC-DC 转换器业务,包括混合和定制板载电源产品。该交易预计将于 2023 年第一季度完成。

DC-DC 转换器是转变输入电压并有效输出固定电压的电压转换器,广泛应用于手机、MP3、数码相机、便携式媒体播放器等产品中。HiRel DC-DC 转换器业务可为外太空等最恶劣环境提供高可靠性的 DC-DC 电源转换解决方案。英飞凌表示,