



本报记者 许子皓

AMD 董事会主席兼首席执行官苏姿丰在近日举行的 2023 世界人工智能大会上表示, AI 将改变我们所从事的每一个行业, 未来 10 年间会出现一个大型计算机超级周期, 现在正是一个成为技术供应商的好时机。

为了抓住这个好时机, 苏姿丰在此前的“AMD 数据中心与人工智能技术首映会”上, 带来了数据中心 APU（加速处理器）Instinct MI300 和第四代 EPYC 处理器的最新消息。在演讲中, 她将两款产品的性能参数直接对标了英伟达的 H100 芯片和英特尔的至强 8490H。此举被业界认为是 AMD 同时挑战英伟达和英特尔在 GPU 和 CPU 领域的霸主地位。在 CPU 和 GPU 领域里, AMD 始终是个挑战者, 即使在 FPGA、DPU 这些新兴芯片领域, AMD 也没有占据显著优势。这也一直让人们猜想: AMD 何时才能停止“仰望”? 苏姿丰的强势出击能给出答案吗?

CPU: 一直在追赶英特尔

“本次推出的第四代 EPYC, 在云工作负载中的性能是英特尔同类处理器的 1.8 倍, 在企业工作负载中的速度是英特尔处理器的 1.9 倍。”苏姿丰十分自信地在会上表示, 这款代号为 Bergamo 的高密度服务器 CPU 新品, 具有 820 亿个晶体管、128 个 Zen 4c CPU 内核、256 个线程, 可满足“高性能的云端需求”, 是 AMD 打开云原生处理器市场的“钥匙”。AMD 希望 Bergamo 能与基于 Arm 架构的 Ampere、Amazon 等公司竞争, 甚至与英特尔将在 2024 年推出的 144 核心的 Sierra Forest 正面较量。

AMD 和英特尔在 CPU 领域的“红蓝之战”, 可以追溯到 1982 年。那时为了得到 IBM 的 PC 量产订单, AMD 拿到了英特尔的授权, 成为英特尔 CPU 芯片的第二供应商, 并根据英特尔的设计和微代码制造了 80286 处理器的 AMD 版本 Am286, 这也让 AMD 拥有了与英特尔抗衡的技术资本。Am286 比英特尔制造的 80286 主频更高, 前者最高可达 20MHz, 后者最高只能达到 10MHz 左右, 这让 Am286 瞬间成了台式 PC 处理器的首选, 就此拉开了二者竞争的序幕。

AMD 第一次真正靠自己研发获得技术领先是在 1996 年。AMD 发布了 K5 系列处理器, 与英特尔的奔腾系列展开了新一轮较量, 这也是 AMD 第一款自主设计的处理器产品, 此时它的综合性能稍微优于奔腾。1997 年, 英特尔退出 Socket7 架构, 转向使用 Slot1 架构, 希望通过断开兼容性的方式主导市场, 而 AMD 迅速且坚决地在 Socket7 架构上推出 K6。半导体行业专家张先扬表示:“对比英特尔放弃了兼容性接口市场, AMD 此举获得了很多厂商的支持。2003 年, AMD 发布了 K8 的 Athlon 64, 这也是 AMD 在与英特尔 30 多年的技术竞争史上首次真正领先对手。”

但到了 2005 年, 英特尔拿出了杀手锏“Tick-Tock 战略”, 成为英特尔重新领先的关键, AMD 逐渐招架不住, 落于下风。

直到 2014 年, AMD 终于迎来了一位足以改变命运的大救星, 也是

AMD 创办 49 年以来第一位女性 CEO 苏姿丰(Lisa Su), CPU 及显卡消费者熟知的“苏妈”。苏姿丰临危受命, 专注研发出了极具竞争力的 AMD 单独立项的架构——Zen, 并在 2017 年就发布了基于 Zen 架构的从移动到桌面, 再到高端桌面的近 20 款产品, 包括新一代微处理器锐龙 Ryzen、高性能显卡 Radeon RX Vega 和服务处理器 EPYC。其中, Ryzen 处理器采用全新的 Global Foundries 14nm 制造工艺、拥有 8 核 16 线程, IPC 性能提升 40%, 而当时顶级的英特尔 i7 6700K 只有 4 核 8 线程, 更先进的工艺不仅推动性能增长, 还大幅降低了功耗, 终于与当时的英特尔酷睿处理器站到了同一水平线, 再加上高性价比, 这一切都提升了 AMD 在 CPU 市场的口碑, 2017 年年底 AMD 将其 CPU 市场份额从 8% 提升至 22%, 也让连续 5 年亏损的 AMD 终于扭亏为盈。

从目前的市场份额来看, 市场研究机构 Counterpoint 的统计数据显 示, 2022 年, 英特尔继续在数据中心 CPU 市场占据主导地位, 但市场份额却下降了 10%~71%, 而 AMD 的市场份额却迎来了增长, 达到了 20%。

这让我们看到了 AMD 打不倒、不服输的精神, 尽管起步较慢, 但始终在追赶, 很多次都在性能和市场份额上反超了英特尔。这也让英特尔不能有丝毫松懈, 必须时刻加紧研发, 因为 AMD 一直紧随其后。

GPU: 难以撼动英伟达

苏姿丰表示:“我们仍处于 AI 生命周期非常早的阶段, 人工智能是塑造下一代计算的决定性技术, 也是 AMD 最大、最具战略意义的长期增长机会。”

当前的 AI 芯片市场可以说是英伟达的天下, 每一位挑战者想要动摇其根基都并非易事。AMD 作为英伟达的老对手, 自然不会放任其独揽如此庞大且增速超快的市场。苏姿丰本次拿出的“王炸产品”就是一款直接对标英伟达 H100 芯片, 专门面向生成式 AI 推出的加速器 AMD Instinct MI 300X。AMD Instinct MI 300X 内部没有集成 CPU 内核, 而是采用了 8 个 GPU chiplet 加 4 个 I/O 内存 chiplet 的设计, 12 个 5nm Chiplets 封装在一起, 使其集成的晶体管数量达到了 1530 亿个, 多于英伟达 H100 的 800 亿个, 可以加速 ChatGPT 等大模型应用。

与英伟达的 H100 芯片相比, AMD Instinct MI 300X 的 HMB 密度是前者的 2.4 倍, 带宽则为前者的 1.6 倍, 理论上可以运行比 H100 更大的模型。

AMD 入局 GPU 领域时态度十分坚定, 但也付出了沉重的代价。在 2006 年, AMD 豪掷自身近一半的市值, 以 54 亿美元买下了英伟达在 GPU 领域的死对头 ATI, 成为当时全球唯一一个同时拥有高性能 CPU 和 GPU 技术的公司, 但这也意味着 AMD 要双线作战, 同时与“双

英”展开厮杀。

尽管英伟达比 AMD 晚成立 24 年, 但作为 GPU 的创立者, 英伟达在 GPU 领域鲜有对手, ATI 也只能在中低端占有部分份额。但 ATI 被 AMD 收购之后, AMD 利用自身 CPU 的研发优势整合 ATI 的 GPU, 推出了新产品——APU。这款新产品既能做 CPU 擅长的通用任务, 也能做游戏、图形处理, 还能凭借 GPU 强大的浮点性能做加速任务, 最终, 高频、小核心、低成本显卡的策略取得了成功, 甚至曾在 GPU 综合市场份额超越了英伟达, 让英伟达开始重新正视这个老对手。

但 AMD 想要撼动如日中天的英伟达, 并非易事。赛迪顾问集成电路产业研究中心研究员邓楚翔表示, 虽然本次 AMD 的 MI 300X 采用了更大的 192GB HBM3, 但英伟达的产品也在迭代, 等未来 MI300X 正式发售时, 英伟达可能已经推出了参数更强的产品。而且, 由于当日未发布价格, 采用 192GB HBM3 的 MI300X 成本可能并不会比预想的低。因此, 正式发售时与 H100 相比可能不会有显著的价格优势。

另外, MI300X 没有 H100 所拥有的用于加速 Transformer 大模型的引擎, 这也意味着用同样数量的 MI300X 将花费更长的训练时间。短期来看, AMD 的 MI300X 可能更多是作为客户买不到 H100 时的“替代品”。

至顶智库执行主任兼首席分析师孙硕表示, 尽管从 AMD 本次公开的性能参数来看, MI300X 在很多方面都优于英伟达的 H100, 但并不是性能越高, 就越多用人, 这不是一个正向关系。英伟达深耕 GPU 领域多年, 所拥有的市场认可度和产品稳定性都是 AMD 所不具备的。另外, 在软件生态的建立和开发方面, 同样需要不断积累, 而且门槛较高, 需要较长的时间完善。

英伟达的 CUDA 平台经过十几年积累已构建其他竞争对手短时间难以逾越的护城河。AMD 目前已经拥有一套完整的库和工具 ROCm, 也能完全兼容 CUDA, 为 AMD 提供了说服客户迁移的条件和理由, 但兼容只属权宜之计, 进一步完善自己的生态才能形成竞争优势。未来, ROCm 需支持更多的操作系统, 在 AI 领域开拓更广泛的框架, 以此吸引更多的开发者。

GPU 的市场份额方面, 市场调查机构 Jon Peddie Research 发布的最新数据显示, 2024 年第一季度的英伟达继续以 83.7% 的市场占有率持续领导着 GPU 市场。AMD 的市场占有率保持在 12%。虽然差距依旧不小, 但不可否认 AMD 本次带来的产品正式推向市场时, 真正的较量才算开始。

FPGA: 和英特尔旗鼓相当

进入 AI 时代, GPU、FPGA 及 ASIC 被合称为“AI 芯片”, 与 GPU 相比, FPGA 作为可编程逻辑阵列,

具备更低能耗、更强的灵活度和可编辑性, 因而具备较短的设计周期, 更适合算法快速迭代、应用场景不断拓展的 AI 时代, CPU+FPGA 的方案也能够提供 GPU 所擅长的庞大算力支持。

在 FPGA 领域的布局上, 英特尔还是抢先一步, 早在 2015 年就以 167 亿美元的价格收购了当时全球第二 FPGA 厂商的 Altera。AMD 也不甘示弱, 在 2022 年, 以 500 亿美元的天价收购了全球最大的 FPGA 厂商赛灵思, AMD 因此跃升为新晋 FPGA 龙头。

AMD 通过有效整合赛灵思在 FPGA 方面的优势, 提供了具有更广泛高性能的计算产品组合。苏姿丰表示:“借助赛灵思在 5G、通信、自动驾驶和行业领域的资源, AMD 能够将高性能计算能力带入更多领域。”

AMD 通过赛灵思领先的 FPGA、自适应 SoC 技术, 强强联合, 在 6 月 27 日宣布推出 AMD Versal Premium VP1902 自适应片上系统(SoC), 这是一款基于 FPGA 的自适应 SoC, 既是全球最大的自适应 SoC, 也将成为全球最大的 FPGA。

但英特尔由于更早开始研发 FPGA, 长时间的积累也让英特尔在该领域更加成熟。英特尔表示, 计划在今年推出 15 款新 FPGA 产品, 这刷新了英特尔该品类的年度推新纪录。并且最新推出的 Agilex 7 R-Tile FPGA, 相较于 AMD 的产品, 在带宽速度等方面已经取得了领先。

可见, 在 FPGA 这个高速发展的领域, AMD 和英特尔的对擂还将延续。

DPU: 相比“双英”姗姗来迟

随着带宽不断提高, 海量数据涌入, 一些“CPU 做不好, GPU 做不了”的复杂数据处理工作, 开始逐渐转向能够重新分配算力和优化算力资源的 DPU。DPU 因此被列为 CPU、GPU 之外的第三个主力芯片, 成为新一代数据中心创新范式。因此, 英伟达、英特尔、AMD 三大厂商都开始布局 DPU 赛道。

英伟达在 2019 年就发现这一新领域, 以 69 亿美元的价格收购了以色列网络芯片公司 Mellanox, 并于同年推出了 BlueField-2 DPU, 自此拉开 DPU 高速发展的序幕, 打造了全新的芯片路线“GPU+DPU+CPU”。基于 DPU 的结构优势, 英伟达计划在 Bluefield-4 产品实现将 GPU 集成至 DPU 中, 进而实现数据中心的架构整合。

从英伟达最新商业路线来看, 英伟达正在布局“CPU+GPU+DPU”的三芯战略, 未来极有可能将构建自己的 XPU 体系, 并催生新的解决方案。

英特尔没有选择收购 DPU 厂商, 而是推出了与 DPU 功能相似的 IPU 产品。英特尔数据平台事业部首席技术官 Guido Appenzeller 表示:“IPU 是一种全新的技术类别, 是英特尔云战略的重要支柱之一。它扩展了英特尔的智能网卡功能, 旨在应对当下复杂的数据中心, 并提升效率。”

IPU 的发布, 让英特尔成为当时业界唯一拥有 CPU、独立 GPU、IPU、ASIC、FPGA 和各种加速器的企业。在同年 的英特尔架构日上, 英特尔向业界提出了 XPU 异构愿景: 一个由标量、矢量、矩阵、空间组成的 SVMs 架构——分别对应了 CPU、GPU、IPU 和 FPGA, 可进行多种异构组合。这也是英特尔首次在业界明确将单一架构之后的发展方向定义为 XPU。

AMD 则一直到 2022 年才姗姗来迟, 宣布以 19 亿美元收购 DPU 芯片厂商 Pensando。AMD 基于 Pensando 的 DPU 技术, 发布了代号为“Giglio”的下一代 DPU 路线图, 与当前一代产品相比, 该路线图旨在为客户带来更高的性能和能效, 预计将于 2023 年年底上市。

至此, AMD 的“CPU+GPU+FP-GA+DPU”的芯片版图搭建完成, 未来将以此继续扩大数据中心和人工智能业务版图, 丰富产品类型。

可以看出, AMD 想要在 CPU 和 GPU 领域战胜“双英”, 不再仰望, 还有很长的一段路要走, 但在 FPGA 和 DPU 赛道上, 和“双英”已然并驾齐驱。“AMD 目前的势头很好, 在产品性能和技术上已经可以匹敌‘双英’, 但仍需要加强与业界的合作, 增强市场产品认可度。我们希望 AMD 能继续斗志昂扬地与英特尔、英伟达展开竞争, 高强度的激烈竞争势必会加速技术创新, 让消费者有更多的选择, 让市场上出现性价比更高的产品。”张先扬说道。

英特尔推出面向中国市场的第二代 Gaudi 深度学习加速器

本报讯 记者沈丛报道: 7 月 11 日, 英特尔推出针对中国市场的第二代 Gaudi 深度学习加速器——Habana Gaudi 2, 该加速器专用于高性能深度学习 AI 训练, 能够帮助中国客户以更低的成本进行 AI 训练, 并更高效地进行大规模 AI 应用部署。

据了解, 此次发布的第二代 Gaudi AI 深度学习夹层卡 HL-225B 专为数据中心实现大规模横向扩展而设计。该训练处理器基于第一代 Gaudi 的高效架构打造。

英特尔公司执行副总裁、数据中心与人工智能事业部总经理 Sandra L.Rivera 表示, 第二代 Gaudi 是为数不多的能替代英伟达 H100 进行大模型语言训练的可靠方案。在性价比方面, 第二代 Gaudi 预计在 2023 年 9 月份将超过英伟达 H100, 成为更具成本优势的产品。

此外, 记者了解到, 与其他代工厂进行合作, 也是英特尔深化 IDM 2.0 的关键步骤之一, 而此次发布的第二代 Gaudi 搭载的便是台积电 7nm 工艺。对此, 英特尔数据中心与人工智能集团副总裁兼中国区总经理陈葆立向《中国电子报》记

显示产业向“绿”而行

(上接第 1 版)

TCL 华星首席执行官赵军向记者表示, 下一步, TCL 华星将持续发掘供应链减排潜力, 并加强上下游协作, 助力每年碳排放强度下降 3%~5%。并承诺不晚于 2030 年实现“碳达峰”, 不晚于 2050 年实现“碳中和”。

厦门天马微电子有限公司、武汉天马微电子有限公司入选国家级绿色工厂名单。天马微电子在 2023 年 6 月发布的“碳达峰碳中和”行动方案中明确, 计划 2030 年, 万元工业产值综合能耗较 2020 年下降 25%, 万元工业产值二氧化碳排放量较 2010 年下降 60%。

记者从维信诺获悉, 该公司合肥 G6 全柔 AMOLED 生产线规划屋顶光伏项目, 分布式光伏发电项目于 2021 年 5 月并网成功, 2022 年发电量 1393 万千瓦时, 每年可节约标准煤约 1713 吨, 年减少二氧化碳及二氧化硫等废气排放 11202 吨, 绿色低碳效应显著。

显示产业越来越多的“绿色工厂”, 从能源管理、水资源节约、资源化利用等方面发挥绿色制造示范带动作用。

技术研发 渗透绿色理念

“以 3 寸电子纸标签计算, 在过去 7 年间, 全球已安装约 6 亿个。若以每天更换 4 次价格信息计算, 使用一次性纸质价格标签所产生的二氧化碳排放量是电子纸标签的 3.2 万倍。”E Ink 元太科技工业股份有限公司董事长李政昊指出。

随着绿色经济和数字经济的协同融合发展, 电子纸作为一种绿色低碳的显示技术有着广阔的发展前景。其实, 显示产业正通过全方位研发绿色显示技术和开发低能耗显示产品, 提高生产能效和原料利用率, 推动实现终端产品的节能降碳。

京东方总裁高文宝在“京东方全球创新伙伴大会·2023”上指出, 公司推出的“全方位低碳方案”已应用于低 碳显示器、低功耗笔记本电脑等多领域的节能环保产品上, 向终端消费者倡导更加绿色的生活方式。

据悉, 京东方 27 英寸低碳环保显示器产品, 采用可回收材

料降碳生产, 实现低功耗的同时保证 QHD 分辨率显示效果。京东方 Tandem 柔性屏通过集成 LTPO 和 Tandem 两种技术, 不仅实现了降低能耗超 20%, 还将器件使用寿命提高了 3~4 倍。

TCL 华星通过芯片、玻璃、包材的减料减量, 实现显示模组轻量化, 例如, 进一步开发金属背板和导光板减薄技术, 使其减薄 25% 以上, 生产轻量化显示模组; 为显示面板生产导入 0.4T 超薄玻璃基板。

在绿色技术研发方面, TCL 华星加大 LCD HFS/HVA 高效技术迭代开发, 并着重攻克 FMM-OLED 超低频背板技术, 提升背光光效来生产低功耗产品。赵军表示, 该公司规划 Oxide 技术平台, 通过新型高迁高稳氧化物半导体材料及器件开发, 建立下一代显示技术最重要的平台性共用技术, 以研发高规格、低功耗产品。

天马着力低功耗车载屏幕技术的研发, 相较于常规 LED 背光, AM Mini LED (主动式发光) 在可靠性和使用寿命上更具优势, 同时可以实现对单独光源的主动控制, 通过减少 LED 的点亮数量以降低屏幕的功耗与发热, 在促进“双碳政策”推动实施的同时, 助力新能源汽车超长续航, 实现“里程无忧”。

维信诺通过优化工艺、减少材料、降低功耗等方面积极推动工艺端绿色低碳发展。记者从维信诺获悉, 其通过对模组工艺进行分析, 并协同模切合作伙伴开发出新型产品, 缩短加工耗时约 75%, 这意味着同样生产一块 AMOLED 屏幕, 采用此工艺可比常规工艺降低 75% 的能耗。

此外, 维信诺开发出 Hybrid-TFT 低功耗技术、EnV ALT 低频 LTPS 技术等 AMOLED 屏幕低功耗整体解决方案, 为消费者打造低功耗、高画质、更环保的屏幕, 助力绿色低碳理念在消费端落地。

显示产业领军企业正在发挥引领示范作用, 带动上下游全产业链共同推进绿色制造理念, 实施绿色制造工程, 以转型升级推动绿色发展。中国电子信息产业发展研究院发布的《中国新型显示产业发展现状与趋势洞察》指出, 低碳环保成为新的显示技术的核心驱动力。