

FPGA和CPU、GPU有什么区别?为什么越来越重要?

2023 年 4 月 17 日 看好/维持 电子 行业报告

—— "FPGA五问五答"系列报告二

分析师 李美贤 电话: 13718969817 邮箱 limx_yjs@dxzq. net. cn

执业证书编号: \$1480521080004

投资摘要:

近年来,诸如 TPU、MPU、DPU 等的"X"PU 们似乎层出不穷,市场经常会对这些新创造出的名词感到困惑:为什么会出现这么多的单元?作为我们 FPGA 五问五答系列报告二,在这篇报告中,我们进一步回答投资人最常问的问题之一:FPGA 和 CPU、GPU有什么区别?为什么越来越重要?我们认为,这个问题不是简单回答他们的区别就可以解决的,因此我们单拎出来解答。

我们认为,众多 "X" PU 出现的原因,本质上是由于 CPU 的算力到达瓶颈了,背后是通用计算时代的终结。从发明以来,CPU 算力的提升主要依靠两大法宝:一是提高时钟频率,二是增加处理器内核数,但这些方法到现在也遇到瓶颈了,人们开始放弃使用一个超强的 CPU 完成所有事情,而是对某些重复的场景,卸载到专用的加速器,以达到新一阶段的降低功耗,提升性能的目的,这就是"XPU"等加速器兴起的原因。同时,自 2010 年 AI 兴起,AI 模型的训练所需的算力是爆发式的增长,且"加&乘"的本质使得算力要求愈发偏向高并行而不是高串行,因此 CPU 越来越难以胜任高算力的场景,通用计算时代终结,数据中心走向加速器时代。未来 10 年,在数据中心高性能计算及 AI 训练中,CPU 这一"主角"的重要性下降,而以往的"配角们",即 GPU、FPGA、TPU、DPU等的加速器的重要性在上升。

FPGA 相比 CPU 的巨大优势在于确定性的低时延,这是架构差异造成的。CPU 的利用率越高,处理时延便越大,而 FPGA 无论利用率大小,其处理时延是稳定的。在汽车和工业这些需要确定低时延的场景,FPGA 具有非常大的优势。此外,FPGA 相比 CPU,具有更高的灵活性。

FPGA 相比 GPU 的优势在于更低的功耗和时延。GPU 无法很好地利用片上内存,需要频繁读取片外的 DRAM, 因此功耗非常高。FPGA 可以灵活运用片上存储,因此功耗远低于 GPU。FPGA"无批次(Batch-less)"的架构,使其在 AI 推理中相比 GPU 具有非常强的时延优势。此外,GPU 的接口单一,而 FPGA 在接口灵活性上具有无可比拟的优势,特别适合工业场景。

为什么 FPGA 是战略芯片?我们认为,未来科技发展有两个领域处于战略地位:一是 AI,二是太空。AI 代表人类更高级别的生产力工具,而太空是可供人类开发探索的广阔而未知领域。FPGA 凭借其架构带来的时延和功耗优势,在 AI 推理中具有非常大的优势。同样,FPGA 独特的优势使其在航空航天领域有非常广泛的应用。

目前,我们看到太空活动发生新变化,背后是太空不断增长的算力需求。变化 1: 地球观测、探火活动在增加;变化 2: 寻求扩大 AI 在太空的应用,以及宽带卫星通信的快速增长,提高了算力要求;变化 3: 航天级器件的代际差在缩小,处理能力越来越接近目前最高水平。当今全球地缘政治紧张的背景下,各国自有卫星星座需求激增,太空活动进入新活跃期。

FPGA 在航天领域为什么更具有优势? 主要有两点原因: 1) FPGA 可以降低项目的时间和金钱成本。航空航天存在着小批量多品种的应用,本质是一个长尾的市场,能够适配的 ASSP 较少,而如果专门设计一颗 ASIC 成本则会非常高,时间也会变得非常长,由此造成的时间成本不是线性的。FPGA"万能"的特点,可以节省 ASIC 的 NRE 成本,设计周期也大幅缩短,避免重复的可靠性认证,加快项目进展; 2) FPGA 动态可重构的特点可以降低在轨错误。太空项目本质是风险厌恶的,一旦发生错误,可能导致数亿美元甚至生命的损失。航天级的 FPGA 可以通过定期刷新回读、动态重构的方式,避免或者减轻宇宙射线对自身造成的破坏,是其它器件所无法做到的。

风险提示: 下游需求不及预期, 中美贸易战超预期。



目 录

1.	为什	·么会有这么多的"X"PU? ——"配角们"的时代	3
		i CPU,FPGA 的并行性和灵活性更高,能提供确定性的时延	
3.	相比	i GPU,FPGA 的时延和功耗更有优势	8
4.	FP G	A 的战略意义:Al & Space	11
风	险提,	ភ	14
		插图目录	
图	1:	CPU 面临算力瓶颈的原因	3
图	2:	2010年兴起以来,AI模型对算力的要求呈现爆发式增长,速度远超摩尔定律	4
图	3:	MLP 网络本质是并行的乘法和累加,非常适合在 FPGA 中实现	4
图	4:	微软的 Azure 使用 FPGA 加速 Bing 搜索,网络吞吐量大幅上升,时延下降了 80%	5
图	5:	FPGA 能完成 SIMD、MISD 和 MIMD 的处理,特别适合并行计算	6
图	6:	将 CPU 的核心简化以加快执行速度,是 GPU 设计的思想	6
图	7:	FPGA 的时延远低于 CPU,是因为其架构不需要在获取指令、编译指令、分支预测等方面花费时间	7
图	8:	FPGA 确定性的低时延,使其在工业和汽车上具有非常大的优势	7
图	9:	FPGA 可以灵活利用片上内存,不需要像 CUDA 一样从 DRAM 来回读写数据	8
图	10:	FPGA 仅用 200MHz,就可以实现比 CPU 快 43 倍、比 GPU 快 3 倍的效果,而且功耗仅为 GPU 的 20%	9
图	11:	FP GA 的时延低于 GPU,无批次的结构,使其在 AI 推理特别有优势	
图	12:	FPGA 更适合 AI 推理,在低时延、非标准化的场景非常有优势	10
图	13:	近5年来,在地缘政治紧张的态势下,中美两国太空发射次数迅速增长	11
图	14:	美国太空活动的新变化,反映出航空航天领域不断增长的算力需求	12

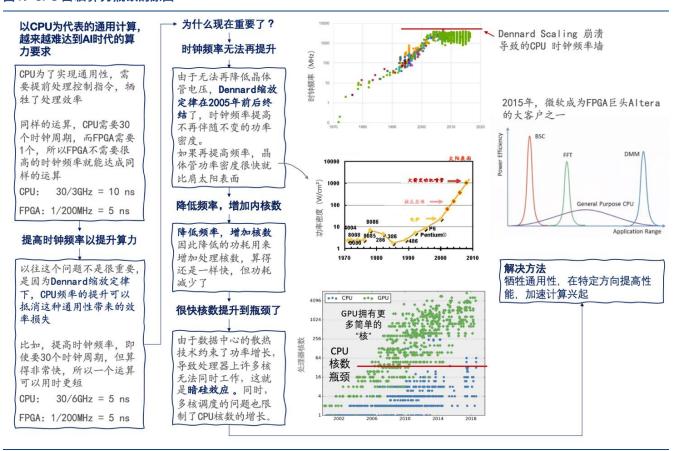
图 15: FPGA 在航天项目的参与度非常高......12



1. 为什么会有这么多的"X"PU? —— "配角们"的时代

近年来,诸如 TPU、MPU、DPU等的"X"PU们似乎层出不穷,市场经常会对这些新创造出的名词感到困惑:为什么会出现这么多的单元? 本质上是由于 CPU 的算力到达瓶颈了,背后是通用计算时代的终结。从发明以来, CPU 算力的提升主要依靠两大法宝:一是提高时钟频率, 但时钟频率提升面临瓶颈了。因为越高的时钟频率, 意味着每秒可执行的运算次数越高, 但随着电压下降到 0.6v 的"底限", Dennard 缩放定律(Dennard Scaling) 在 05 年开始崩溃, 再提高时钟频率就会使得功耗以指数级别增长, 因此我们在 05 年后遇到了频率墙; 二是增加处理器内核数, 换取频率降低带来的功耗预算, 但由于核间调度同样需要时延和功耗花销, 很快, 核数的增长又遇到了瓶颈, 由于数据中心的散热技术约束了功率增长, 导致处理器上许多核无法同时工作, 这就是暗硅效应。因此, 人们开始放弃使用一个超强的 CPU 完成所有事情, 而是对某些重复的场景, 卸载到专用的加速器, 以达到新一阶段的降低功耗, 提升性能的目的, 这就是"XPU"等加速器兴起的原因。

图1: CPU 面临算力瓶颈的原因

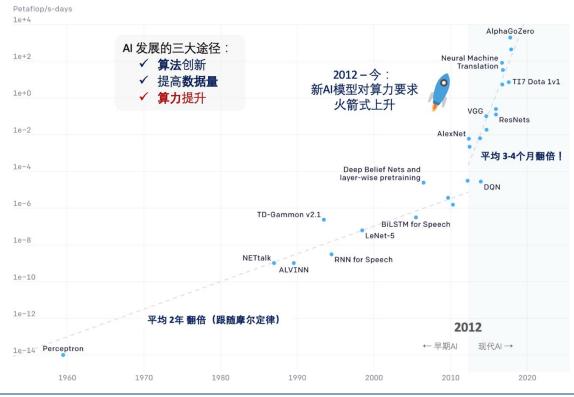


资料来源: Intel, Princeton University, 东兴证券研究所绘制

同时, 自 2010 年 AI 兴起, AI 模型的训练所需的算力是爆发式的增长, 且"加&乘"的本质使得算力要求愈发偏向高并行而不是高串行。CPU 越来越难以胜任高算力的场景,,将需要大规模、高密度的计算任务卸载到在某一方向做了优化的专用处理器,就产生了这些不同的"X"PU,他们之间区别在于在某些场景的专用性。



图2: 2010 年兴起以来, AI 模型对算力的要求呈现爆发式增长,速度远超摩尔定律



资料来源: OpenAI, 东兴证券研究所绘制

通用计算时代终结,数据中心走向加速器时代。未来 10 年,FPGA 的重要性不断上升。随着 CPU 算力逐渐达到瓶颈,越来越无法满足神经网络指数级增长的算力需求。在数据中心这一人类算力需求最高的设施中,算力发展的方向愈发转向专用性,以寻求更高的性能、更低的能耗和成本。我们看到,未来 10 年,在数据中心高性能计算及 AI 训练中,CPU 这一"主角"的重要性下降,而以往的"配角们",即 GPU、FPGA、TPU、DPU等的加速器的重要性在上升。

图3: MLP 网络本质是并行的乘法和累加, 非常适合在 FPGA 中实现



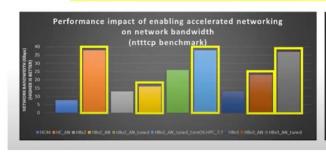
资料来源: FPGA Neurocomputers, 东兴证券研究所

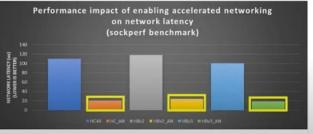


因此,从 2017 年开始,越来越多的大型公有云开始正式使用 FPGA、GPU 作为加速器,以获得数百倍于 CPU 的性能提升。例如,微软的 Catapult 项目就使用了 FPGA 以加速 Bing 的搜索速度,AWS 的 F1 Instances 将 FPGA 的算力作为服务提供给客户, 阿里云使用了 FPGA 为其"双十一"进行了零售交易系统的加速。

图4: 微软的 Azure 使用 FPGA 加速 Bing 搜索, 网络吞吐量大幅上升, 时延下降了 80%

Performance increase and up to 5x latency reduction with Intel FPGA Acceleration in Azure





资料来源:微软,东兴证券研究所

2. 相比 CPU, FPGA 的并行性和灵活性更高, 能提供确定性的时延

处理器负责对外界输入的数据进行处理, CPU、GPU、FPGA等处理器的区别在于处理流程, CPU 的处理流程使其擅长串行计算, 以复杂的控制为特征, GPU 和 FPGA 的则更擅长大规模的并行计算:

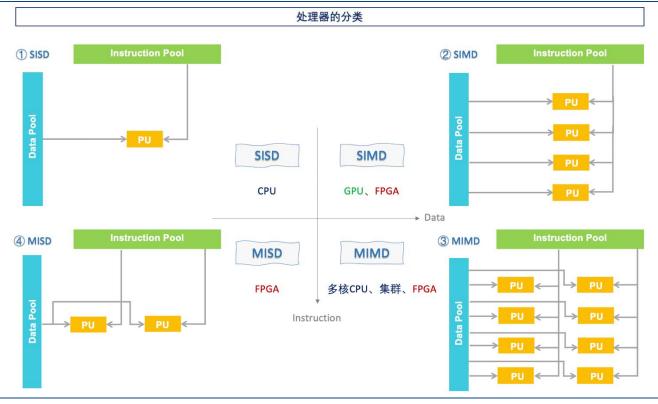
- **CPU** 是冯诺依曼架构下的处理器,遵循"Fetch(取指)-Decode(译码)- Execute(执行)- Memory Access(访存)-Write Back(写回)"的处理流程,数据要先通过控制单元获取存在 RAM 中的指令,再解码得知用户需要对数据做何种运算,然后再将数据送到 ALU 进行对应的处理,结束运算后存回 RAM,再获取下一个指令。这一处理流程,即 SISD(Single Instruction Single Data),决定了 **CPU 擅长决策和控制,但在多数据处理任务中效率较低。**现代的 CPU 可以同时做到 SISD 和 SIMD 的处理,但在并行规模上依然不如 GPU和 FPGA。
- **GPU** 遵循的是 SIMD (Single Instruction Multiple Data) 的处理方式,通过在多个线程上运行统一的处理方式,即 Kernel,来达到将 CPU 发送过来的数据做高并行处理的目的。由于去除了现代 CPU 中分支预测、乱序执行、存储预取等模块,也减少了许多 cache 的空间, GPU 中经过简化后的"核"能实现非常大规模的并行运算,并且节省了大部分 CPU 需要花费在分支预测、重排的时间,但缺点是需要数据适应 GPU 的处理框架,例如需要数据做批次对准,因此依然无法达到最大的实时性。
- **FPGA**则是由用户自定义处理流程,可以直接决定片上的 CLB 是如何相连的,数十万个 CLB 可以独立运算,即 SIMD、MISD (Multiple Instruction Single Data)和 MIMD (Multiple Instruction Multiple Data)的处理都可以在 FPGA实现,由于处理流程已经映射到硬件上,不需要再额外花费时间获取和编译指令,同样不需要像 CPU一样花费时间在乱序执行等步骤,这使得 FPGA 在数据处理中具有非常高的实时性。

东兴证券深度报告

FPGA 和 CPU、GPU 有什么区别?为什么越来越重要?——"FPGA 五问五答"系列报告二



图5: FPGA 能完成 SIMD、MISD 和 MIMD 的处理, 特别适合并行计算



资料来源: Flynn, 东兴证券研究所绘制

因此, GPU 和 FPGA 都是作为 CPU 的任务卸载单元, 在并行计算的效率都高于 CPU。在数据中心高性能计算的场景中, GPU 和 FPGA 往往以分立的加速卡形式存在,即 CPU 将部分密集计算的任务"卸载"到 GPU 或者 FPGA,这些"器件"通过 PCle 和 CPU 互联,以完成高并行的计算加速。

图6:将 CPU 的核心简化以加快执行速度,是 GPU 设计的思想

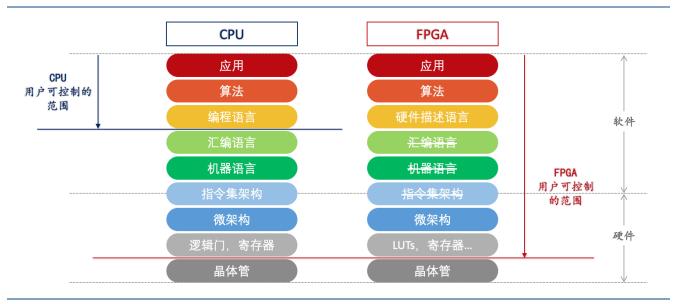


资料来源: Virginia Tech, 东兴证券研究所绘制



FPGA相比 CPU 的巨大优势在于确定性的低时延,这是架构差异造成的。CPU 的时延是不确定的,当利用率升高时,CPU 需要处理更多的任务,这就需要 CPU 进行任务调度重排,因此造成处理时延往往是不可控制地变大,即任务越多算得越慢。而 FPGA 的时延之所以是确定的,是因为在布局布线阶段,设计工具就已经确保能够让最差路径满足时序要求,不需要再花费时间在获取指令、解码指令等通用处理器需要的步骤,也避免了随之而来的重排执行顺序、指令调度等待的问题。

图7: FPGA 的时延远低于 CPU, 是因为其架构不需要在获取指令、编译指令、分支预测等方面花费时间



资料来源: Parker, 东兴证券研究所绘制

CPU 的利用率越高,处理时延便越大,而 FPGA 无论利用率大小,其处理时延是稳定的。FPGA 可以提供 纳秒级的处理时延,而 CPU 通常在毫秒级。例如,在自动驾驶系统中,将摄像头的数据直接传输到 FPGA 的 MIPI 接口中,其最好和最差情况下的处理时延差距仅为 22ns,而在有 CPU 参与数据传输的情况下,这一差距在 23ms 以上,相当于 CPU 在繁忙情况下时延翻倍。此外,当利用率上升到 90%时,CPU 的处理时

图8: FPGA 确定性的低时延, 使其在工业和汽车上具有非常大的优势



资料来源: Intel, 东兴证券研究所绘制

东兴证券深度报告

FPGA 和 CPU、GPU 有什么区别? 为什么越来越重要? —— "FPGA 五问五答"系列报告二



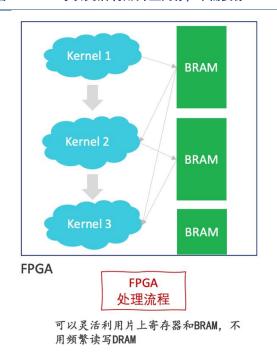
间长达 46ms,对于以 100km/h 的速度行驶的汽车,46ms 意味着摄像头从看到障碍物,到汽车系统采取制动措施时,车子已经开出了至少 1.28 米的距离,而 FPGA 仅有 3⁻⁶米,即可以等同为瞬间就能反应,省去的这 1.28m 的距离,就可能减少许多碰撞事故的发生概率。因此,在汽车和工业这些需要确定低时延的场景,FPGA 具有非常大的优势。

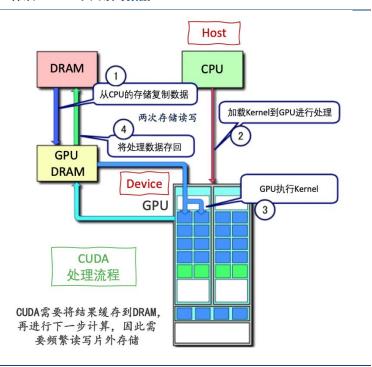
FPGA 相比 CPU, 具有更高的灵活性。在工业现场往往有许多需要细微调整,例如,根据传送带磨损情况对马达进行细微的控制调整,为设备更新新的协议等等,CPU 往往难以做到,由于 FPGA 是动态可重构的,可以在使用现场调整,随时适应新的变化。此外,FPGA 还可以同时融合工业现场的 PLC、网关、传感器、马达、HMI 等设备,实现不同设备的实时控制和通信。

3. 相比 GPU, FPGA 的时延和功耗更有优势

GPU 的功耗非常高,因为其无法很好地利用片上内存,需要频繁读取片外的 DRAM。尽管在吞吐量上的优势使得 GPU 几乎垄断了深度学习领域,但 GPU 依赖片外存储的处理流程,使其在功耗和时延上对比 FPGA 有非常大的弱势。以英伟达的 GPU 为例,使用 CUDA 进行训练,主要有四个步骤: 1)将数据从 CPU 的外部存储 (DRAM) 复制到 GPU 的存储中; 2) CPU 加载 (Lauch)需要进行的计算,即 Kernel 到 GPU 中; 3) GPU 执行 CPU 发送过来的指令; 4) GPU 将结果最终存回 CPU 的 DRAM 中,再进行下一个 Kernel 的计算。因此,CUDA 涉及了两次存储读写。而 FPGA 可以将第一个 Kernel 的结果缓存到片上星罗棋布的 BRAM中,完全可以不需要读写外部存储就能完成整个算法。由于读取 DRAM 所消耗的能量是 SRAM 的 100 倍以上,是加法的 6400 倍,GPU 这一需要频繁读取 DRAM 的处理,使其功耗远高于 FPGA,而且 DRAM 的带宽往往成为了性能的瓶颈。一片 FPGA 的典型功耗通常是 30W~50W,而单片 GPU 功耗就可以高达

图9: FPGA 可以灵活利用片上内存, 不需要像 CUDA 一样从 DRAM 来回读写数据





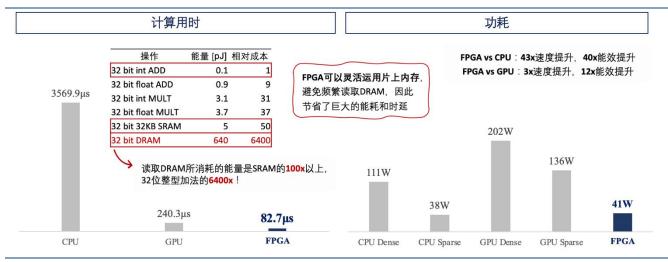
资料来源: NVIDA, 东兴证券研究所绘制



250W~400W,使得单机柜的功率密度可能高达28kw,这对数据中心的现有的散热造成了巨大压力,往往需要专门更改制冷和供电系统,以适应单柜15kw以上的功率密度,而FPGA数十瓦的功耗可以和现有数据中心散热兼容.不需要额外改造。

FPGA 可以灵活运用片上存储,因此功耗远低于 GPU。FPGA 完全可以不需要读 DRAM,整个算法在片上完成。例如,深鉴科技利用 FPGA 做出了 ESE 的模型并在不同的处理器(CPU/GPU/FPGA)上运行,发现 FPGA 上训练时长最短,能耗最小。在能耗上,CPU Dense 耗能 11W、CPU Sparse 耗能 38W、GPU Dense 耗能 202W,这是耗能最大的一种情况、GPU Spare 耗能 136W,相比之下 FPGA 仅需 41W;在训练时延上,FPGA 用时 82.7µs,远小于 CPU 的 6017.3µs,也仅为 GPU 训练时长的三分之一。

图10: FPGA 仅用 200MHz, 就可以实现比 CPU 快 43 倍、比 GPU 快 3 倍的效果, 而且功耗仅为 GPU 的 20%



资料来源: ESE: Efficient Speech Recognition Engine with Sparse LSTM on FPGA, Energy table for 45nm process, 东兴证券研究所绘制

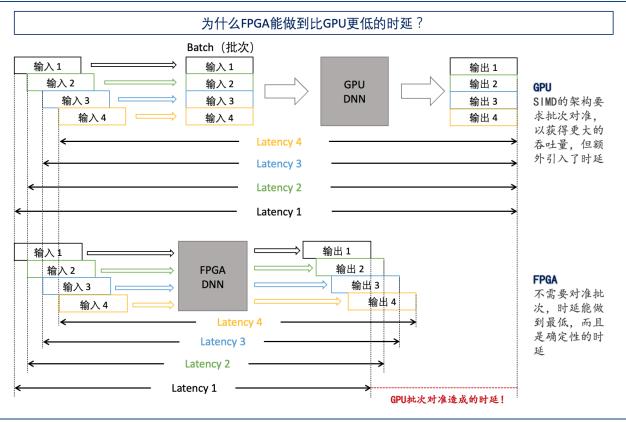
FPGA"无批次(Batch-less)"的架构,使其在 AI 推理中具有非常强的时延优势。受限于网络条件和时延,许多决策来不及上传云端,只能本地执行,这就是边缘计算。边缘计算通常面临时延和功耗两大约束。GPU需要等待批次的特点,使其时延要高于 FPGA。GPU通常需要将不同的训练样本划分成固定大小的"Batch(批次)",为了最大化达到并行性,需要将数个 Batch 都集齐,再统一进行处理,每个 Batch 的数据一般有近百个。这使得 GPU 在训练大型样本时非常有优势,但在做涉及小样本的推理时,这一优点成为了劣势,因为推理通常只需要很小的输入数据,而 GPU 的架构额外引入了时延。FPGA 的架构是无批次(Batch-less)的,可以根据数据特点确定处理方式,不需要像 GPU 一样将输入的数据划分成 Batch,因此可以做到最低的时延,使得 FPGA 在进行 AI 推理时具有非常大的优势。

FPGA在接口灵活性上具有无可比拟的优势,特别适合工业场景。工业实质是高度分散的小批量场景,存在大量的非标准的接口,例如,工业的图像传感器的 LVDS 编码格式往往没有统一的标准,工程师很难找到对应的专用芯片去对接。GPU的接口单一,只有 PCIe 一种,而 FPGA 的可编程性使其能与任何的器件进行通信,能够适应任何的标准和非标准的接口,这种硬件可编程带来的高度灵活性是 FPGA 在工业场景的优势。

目前,阻碍 FPGA 市场进一步扩大的原因是其较高的使用门槛,正通过 HLS 等工具解决。CPU 使用人员更多是软件工程师,语言基本为 C/C++等编程语言,GPU 亦有 CUDA 等非常完善的开发框架,而 FPGA 的使用者更多像是硬件工程师,需要自行定义电路功能、进行时序优化等步骤,语言基本为 Verilog/VHDL 这两种



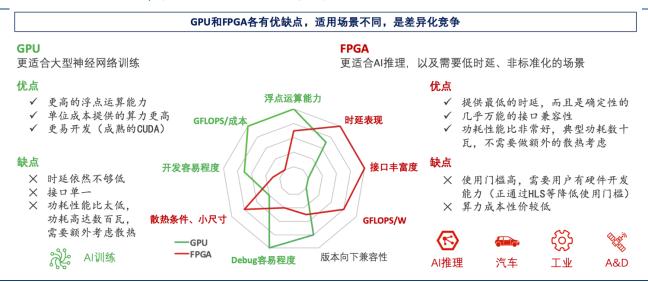
图11: FPGA 的时延低于 GPU, 无批次的结构, 使其在 AI 推理特别有优势



资料来源: Xilinx, 东兴证券研究所绘制

硬件描述语言,需要使用者精通软件和硬件,难度因此更大。因此,为了降低使用门槛,FPGA 学界和业内合作推出了 HLS (High-level Synthesis,高层次综合)的工具,可以通过 C/C++语言直接生成能供 FPGA使用的 RTL 网表,跳过中间的硬件描述环节,让工程师更加专注 AI 算法的开发和迭代。

图12: FPGA 更适合 AI 推理, 在低时延、非标准化的场景非常有优势



资料来源: BERTEN, 东兴证券研究所绘制

P11



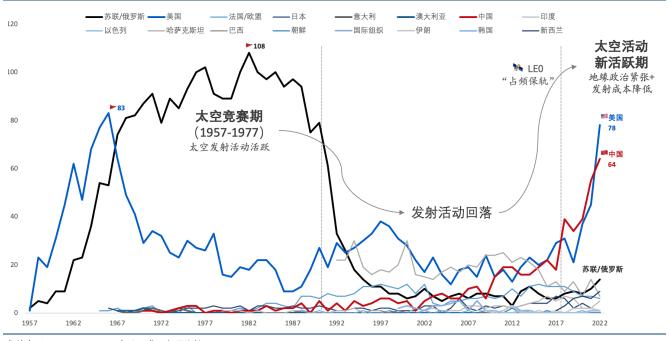
FPGA和 CPU、GPU有什么区别?为什么越来越重要?-"FPGA 五问五答"系列报告二

4. FPGA 的战略意义: AI & Space

为什么 FPGA 是战略芯片?我们认为,未来科技发展有两个领域处于战略地位:一是 AI,二是太空。AI 代 表人类更高级别的生产力工具,而太空是可供人类开发探索的广阔而未知领域。FPGA 凭借其架构带来的时 延和功耗优势, 在 AI 推理中具有非常大的优势。同样, FPGA 独特的优势使其在航空航天领域有非常广泛的 应用。

全球地缘政治紧张下,各国自有卫星星座需求激增,太空活动进入新活跃期。上一个太空发射活跃期在 1957-1977 年,美苏在太空领域展开激烈竞争,两国年平均发射活动均超 40 次。随着美苏太空竞赛结束, 20 年间太空发射数大幅回落。而在近年地缘政治紧张态势下,各国部署自有通信卫星星座需求激增。中美两 国在近3年的太空发射活动剧升,仅去年全年,中美发射次数合计占全球76%。由于频段和低轨空间是不可 再生资源,各国的低轨卫星计划实际承担"占频保轨"的任务。随着大批的低轨卫星计划在未来 4-5 年内完成 发射组网, 太空活动实际已进入新的活跃期。

图13: 近5年来,在地缘政治紧张的态势下,中美两国太空发射次数迅速增长



资料来源: Launch Logs, 东兴证券研究所绘制

目前,我们看到太空活动发生了三大新变化,背后反映的是太空不断增长的算力需求。美国 JPL (Jet Propulsion Laboratory, 喷气推进实验室) 是美国国家航空航天局(NASA)负责无人太空探测的机构,我 们统计了 JPL 目前的所有任务目标,发现了以下三大变化: 1) 地球观测、探火活动在增加。以地球为目标 的太空活动占比 35%,目的主要有气象和环境观测,构建与外太空交流的深太空网络(DSN),利用合成孔 径雷达对地面进行高精度观测等,主要是出于军事及科研目的;而火星相关的活动占比高达 15%,是因为火 星是与地球最相似的行星,了解火星表面的岩石、气候,目的是了解火星在过去是否有生命存在,可以为人 类探索和开发火星做准备。



代际差在缩小

2) 寻求扩大 AI 在太空的应用,以及宽带卫星通信的快速增长,提高了算力要求。以观测卫星为例,地球60%

图14: 美国太空活动的新变化,反映出航空航天领域不断增长的算力需求

美国太空活动的新变化:背后是航空航天领域不断增长的算力需求

➤ 变化1: 地球、火星、月球为JPL三 ➤ 变化2: 寻求扩大AI在太空中的 大任务目标,火星任务近年增多 应用,包括观测卫星、通信卫星

| 19% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10% | 10

➤ 变化3: 龙头的航天级器件与商业级的代际差正在缩小,处理能力越来越接近目前最高水平

一般航天级晚于商业级3-5年

宇航级FPGA	推出时间	商业级推出时间	代际差
Virtex II QPro (0.15μm)	2003	2001	2年
Virtex-4QV (90nm)	2008	2004	4年
Virtex-5QV (65nm)	2010	2006	4年
RT Kintex U (20nm)	2020	2013	7年
Versal XQR (7nm)	2021	2019	2年

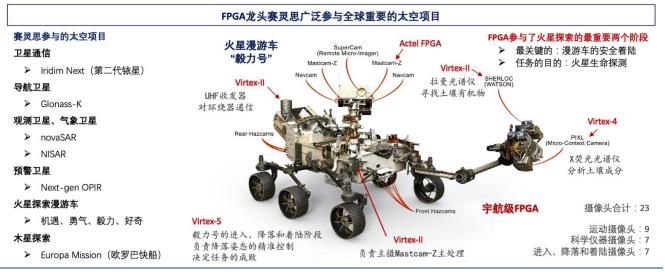
资料来源: JPL, Xilinx, 东兴证券研究所绘制

以上的面积常年被云层覆盖,只有10%的区域是晴空无云的状态,以往观察卫星都是不加甄别,将拍摄的照片全部回传地面处理。如今,在观察卫星上使用AI识别出含有云层的照片并丢弃,只回传清晰的照片,可以节省本就有限的星地通信的带宽。除此之外,宽带卫星通信要求卫星具备星上处理和转发数据的能力,以降低时延.减少对地面站的依赖:

3) 航天级器件的代际差在缩小,处理能力越来越接近目前最高水平。过去,航天级 FPGA 的推出时间一般晚于对应商业级器件 3-5 年,长期落后于当时最领先的器件 1-2 个代际,10-15 年前的 FPGA 依然在航天器上广泛使用。然而,近两年来,我们看到 FPGA 龙头赛灵思加快了宇航级 FPGA 的推出。目前,赛灵思最先进的 FPGA 产品是 19 年推出的 Versal (7nm),而赛灵思在 21 年初就推出了宇航级的 Versal XQR,做到了和商业级同代际。Versal XQR 不仅逻辑单元数相比往代大幅增加,还嵌入了 AI 处理单元、高速的收发器等,大幅提高了低轨卫星的处理能力和反应时间。

图15: FPGA 在航天项目的参与度非常高





资料来源: Xilinx, NASA, 东兴证券研究所绘制

FPGA 在航天领域为什么更具有优势? 主要有两点原因: 1) FPGA 可以降低项目的时间和金钱成本。 航空 航天存在着小批量多品种的应用,本质是一个长尾的市场,因此能够适配的 ASSP 较少,而如果专门设计一颗 ASIC 成本则会非常高,时间也会变得非常长,由此造成的时间成本不是线性的。例如,地球和火星大约 每两年(780 天)距离最短,此时发射所需能量最小,窗口期大多只有 2 个月,如果没有来得及在窗口期完成发射,即使只差一个月,下一次发射就要再等两年,而 FPGA"万能"的特点,可以节省 ASIC 的 NRE 成本、设计周期也大幅缩短、避免重复的可靠性认证、加快项目进展:

2) FPGA 动态可重构的特点可以降低在轨错误。太空项目本质是风险厌恶的,因为一旦发生错误,可能导致数亿美元甚至生命损失。航天器运行在充满高能粒子的太空环境中,宇宙射线随时可能对电子器件造成损坏,而航天级的 FPGA 可以定期刷新回读,避免 SEU 等事件造成的故障。当在轨航天器发生错误时,地面可以直接发送指令让 FPGA 进行重构,即在轨动态重构;如果部分 LUT 被宇宙射线造成物理损坏,完全可以绕开损坏部分进行重构;对于已经运行了多年的卫星,可以重构更新其功能,不需要发射新的卫星。FPGA 这一在轨可重构是其它器件所无法做到的。

东兴证券深度报告 FPGA和CPU、GPU有什么区别?为什么越来越重要?——"FPGA五问五答"系列报告二



风险提示

下游需求不及预期, 中美贸易战超预期。



FPGA和 CPU、GPU有什么区别?为什么越来越重要?

分析师简介

分析师: 李美贤

中国人民大学硕士,2019年7月加入东兴证券研究所,从事通信及电子研究,关注新兴科技领域,重点 覆盖AI、云计算、工业互联网等产业链。

分析师承诺

负责本研究报告全部或部分内容的每一位证券分析师,在此申明,本报告的观点、逻辑和论据均为分析师本 人研究成果,引用的相关信息和文字均已注明出处。本报告依据公开的信息来源,力求清晰、准确地反映分 析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与,未来也将不会与本报告中的具体推荐或 观点直接或间接相关。

风险提示

本证券研究报告所载的信息、观点、结论等内容仅供投资者决策参考。在任何情况下,本公司证券研究报告 均不构成对任何机构和个人的投资建议,市场有风险、投资者在决定投资前、务必要审慎。投资者应自主作 出投资决策, 自行承担投资风险。

P16

东兴证券深度报告

FPGA和 CPU、GPU有什么区别?为什么越来越重要?——"FPGA 五问五答"系列报告二



免责声明

本研究报告由东兴证券股份有限公司研究所撰写,东兴证券股份有限公司是具有合法证券投资咨询业务资格的机构。本研究报告中所引用信息均来源于公开资料,我公司对这些信息的准确性和完整性不作任何保证,也不保证所包含的信息和建议不会发生任何变更。我们已力求报告内容的客观、公正,但文中的观点、结论和建议仅供参考,报告中的信息或意见并不构成所述证券的买卖出价或征价,投资者据此做出的任何投资决策与本公司和作者无关。

我公司及报告作者在自身所知情的范围内,与本报告所评价或推荐的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下,我公司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头寸并进行交易,也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。本报告版权仅为我公司所有,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。如引用、刊发,需注明出处为东兴证券研究所,且不得对本报告进行有悖原意的引用、删节和修改。

本研究报告仅供东兴证券股份有限公司客户和经本公司授权刊载机构的客户使用,未经授权私自刊载研究报告的机构以及其阅读和使用者应慎重使用报告、防止被误导,本公司不承担由于非授权机构私自刊发和非授权客户使用该报告所产生的相关风险和责任。

行业评级体系

公司投资评级(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):

以报告日后的6个月内,公司股价相对于同期市场基准指数的表现为标准定义:

强烈推荐:相对强于市场基准指数收益率 15%以上:

推荐:相对强于市场基准指数收益率5%~15%之间;

中性:相对于市场基准指数收益率介于-5%~+5%之间;

回避:相对弱于市场基准指数收益率5%以上。

行业投资评级(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):

以报告日后的6个月内, 行业指数相对于同期市场基准指数的表现为标准定义:

看好:相对强于市场基准指数收益率5%以上;

中性:相对于市场基准指数收益率介于-5%~+5%之间;

看淡:相对弱于市场基准指数收益率5%以上。

东兴证券研究所

北京 上海 深圳

西城区金融大街 5 号新盛大厦 B 虹口区杨树浦路 248 号瑞丰国际 福田区益田路 6009 号新世界中心

座 16 层 大厦 5 层 46F

邮编: 100033 邮编: 200082 邮编: 518038

电话: 010-66554070 电话: 021-25102800 电话: 0755-83239601 传真: 010-66554008 传真: 021-25102881 传真: 0755-23824526