

用于先进视频处理解决方案的现场可编程逻辑门阵列 (FPGA) 产品与技术

(白皮书编号: WP022)

部署先进的网络基础设施不仅可以解决数据传输量激增的问题,而且还能在诸如边缘、核心和云端等网络的不同部分进行数据处理。不足为奇的是大部分数据要么是视频,要么是图像,并且这些数据正以指数级速度增长,并将在未来几年内保持持续增长。因此,需要更多的计算资源来应对数据的大量增长(如图 1 所示)。

由于应用的类型多种多样,因此在数据中心中存在着各种各样的视频或图像处理工作负载。基于专用集成电路(ASIC)的解决方案通常可提供更高的性能,但是无法进行升级以支持未来的算法;基于中央处理器(CPU)的解决方案要比其更加灵活,但其时钟主频已经固定,而且已不再可能大幅提升处理器性能;图形处理器(GPU)是提供视频/图像处理解决方案的另一种候选方案,但其功耗明显高于基于现场可编程逻辑门阵列(FPGA)的解决方案。FPGA 在视频处理和压缩领域内,是一种具有吸引力的选择,因为它们提供了实现创新视频处理算法所需的、平衡的资源。此外,FPGA 提供了一种灵活的解决方案,可以缩短产品上市时间,并能在解决方案的整个生命周期内实现持续升级和部署新的功能。

表 1: 互联网用户和数据流的增长

年份	互联网用户	设备与连接	宽带速度	视频数据流量
2017	34 亿	180 亿	39.0 Mbps	占总量的 75%
2022	48 亿	285 亿	75.4 Mbps	占总量的 83%

来源: 思科 (Cisco) 公司

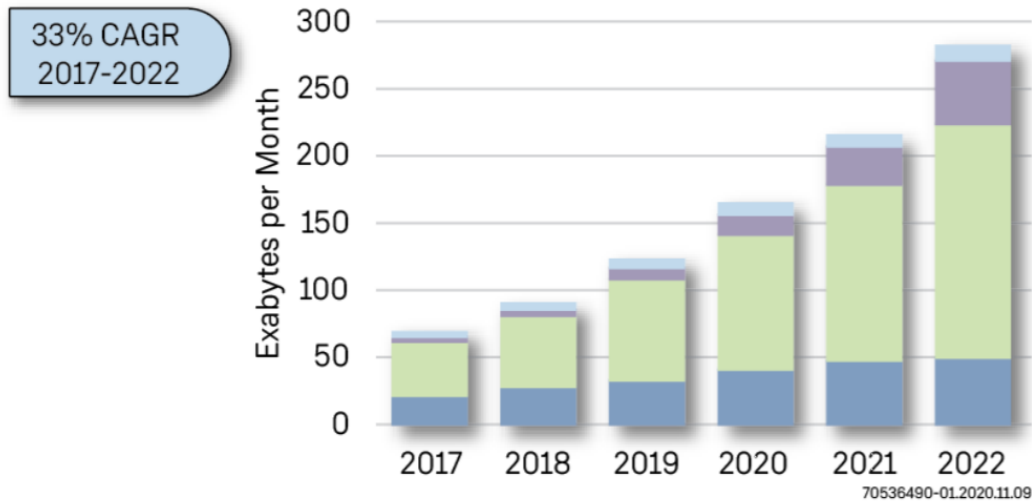


图 1: 全球互联网视频数据流 (来源: 思科)

33% CAGR 2017-2022: 2017-2022 年间的复合年增长率 33%

Exabytes per Month: 每月的 Exabytes 数量

基于 FPGA 的视频解决方案的示例

本白皮书将介绍三种典型的视频应用，以展示基于 FPGA 的解决方案在广播行业中的优势。这些优势包括缩短处理时间、降低功耗，以及为服务提供商和终端用户节省成本。

本白皮书将介绍基于 FPGA 的解决方案在以下三种应用中的优势：

- 视频流
- 使用视频编辑软件来创作视频内容
- 人工智能 (AI) 和深度学习 - 图像识别是该应用的主要部分，其需要高性能的计算资源

视频流传送

为了使媒体流变得快速和高效，对视频进行转码的需求已急剧增加。目前大多数产品都采用了一种基于软件的方法，但该方法无法满足高带宽、广播级视频

流的处理要求。视频流和/或云服务提供商面临着由基于软件的解决方案所带来的低吞吐量、高功耗、长延迟和占用空间大等挑战。根据思科的一份题为《思科可视网络指数：预测与趋势——2017-2022 年白皮书》的报告，视频流数据流量正在增加，并且到 2022 年时将占据整个互联网数据流的 82%。在包括视频点播、流媒体直播和视频监控等所有应用中，视频数据流量将逐年稳步增长。

诸如 Netflix 和 YouTube 等视频流应用的兴起推动了对视频转码的需求。传统广播和视频流媒体之间最显著的区别在于内容量和频道数。为了支持从电脑到智能手机等各种接收设备，内容必须被转码成不同的分辨率和压缩格式。因此，视频流将消耗大量的计算资源。

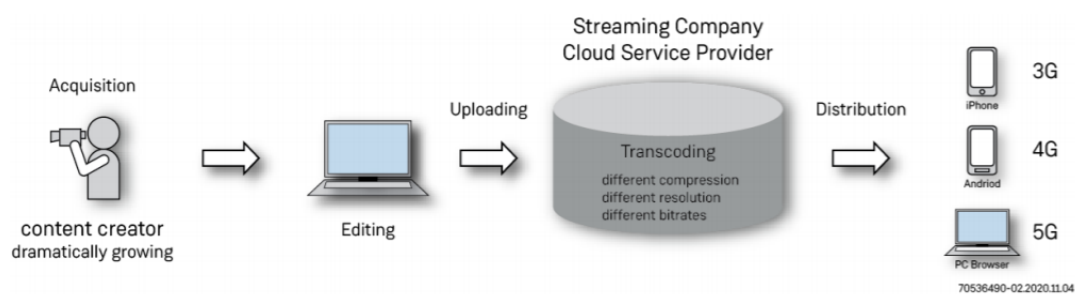


图 2：视频转码工作流程

Acquisition: 获取

content creator dramatically growing: 内容创作者的数量在急剧增加

Editing: 编辑

Uploading: 上传

Streaming Company: 流媒体公司

Cloud Service Provider: 云服务提供商

Transcoding: 转码

different compression: 不同的压缩率

different resolution: 不同的分辨率

different bitrates: 不同的比特率

Distribution: 发布

iPhone: iPhone 手机

Andriod: 安卓手机

PC Browser: 电脑浏览器

流媒体和云服务提供商需要一种解决方案来缓解对计算需求的压力。Achronix Speedster®7t 系列 FPGA 器件中搭载了 IBEX 这种最先进的视频处理半导体知识产权 (IP) 能够解决这一重大问题。这种基于 FPGA 的解决方案可以提供高吞吐量的、低功耗的和占用空间小的系统，而且无需牺牲灵活性。尽管基于 ASIC 的解决方案功能强大，但只能支持在设计时定义的功能集，而不能支持现场更新。

视频内容创作

在过去，高清分辨率 (HD) 格式在视频内容创作中占据主导地位。最近，标准分辨率已被提升至 4K，甚至到 8K，这使得视频编码或解码面临挑战。用于这些较高分辨率的压缩格式主要有 Apple ProRes、Avid DNx 和 SONY XAVC。由于这些压缩格式是专有的，因此 ASIC 或 GPU 并不能原生支持这些格式，而且 CPU 提供的性能也不佳。因此，在较高分辨率下创作视频内容时，FPGA 是最佳的解决方案。



图 3: 视频编辑工作流程

Import: 导入

Editing Software: 编辑软件

Import (Decode): 导入 (解码)

Export (Encode): 导出 (编码)

Remote Edit: 远程编辑

Export: 导出

在新的趋势下，远程后期制作的概念正变得越来越普遍。然而，现有的电脑并没有足够的能力来实时处理高分辨率的内容（例如 8K）。因此，编辑人员开始借助云基础设施来获得更好的计算性能。此外，由于需要保持社交距离，新冠肺炎疫情也加速了这一趋势。基于云和 FPGA 的解决方案为编辑人员提供了巨大的好处。Achronix Speedster7t 系列 FPGA 器件进行架构创新，例如二维片上网络（NoC），使其特别适合于加速编码和解码算法。

人工智能与深度学习

人工智能、机器学习和深度学习是众所周知的领域，它们在过去几年中得到了迅速的发展。除了这些领域，图像识别也逐渐成为一个全新的重要领域，这得益于人工智能/机器学习（AI / ML）的创新。例如，先进驾驶员辅助系统（ADAS）使用深度学习算法来处理捕获的图像。安装在车上的行车记录仪使用 H.264 压缩技术记录视频，然后将视频流转码为诸如 JPEG 或 PNG 等合适的图像格式，以用于深度学习图像识别。根据应用场景，可以同时完成丢帧、更改分辨率或其他图像处理任务。

在零售业的安全摄像头或物流业的包裹分拣中也有类似的应用案例，其数据流与上述示例相同——这些应用中的摄像头使用 H.264 或 H.265 等压缩比相对较高的压缩格式记录视频，然后将编码的视频流传输到云端或数据中心。在云端，视频流由原始格式转码为适合深度学习的格式，将视频文件转换为图像资料库。

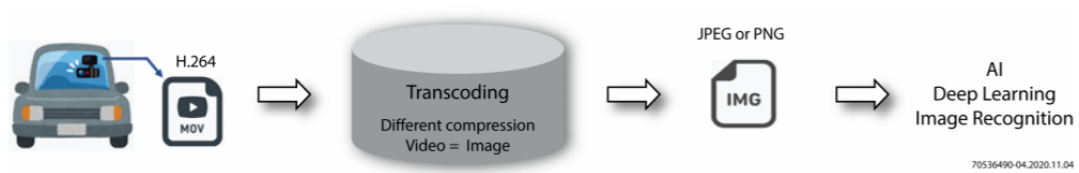


图 4：典型的深度学习图像数据流

Transcoding: 转码

Different compression: 不同的压缩率

Video=Image: 视频=图像

AI: 人工智能

Deep Learning: 深度学习

Image Recognition: 图像识别

从历史来看，FPGA 一直擅长将电影转码为图像。此外，使用 FPGA 中的深度学习算法对图像预先进行预处理，不仅可以提高吞吐量，而且还能减少系统级的数据事务量。Achronix Speedster7t 的创新架构及其带有的专用机器学习处理器（MLP），使之成为实现定制的和既定的深度学习算法的理想选择。

FPGA 代表性视频用例的性能

我们分别使用 FPGA 和 CPU 来实现上述三个典型应用案例，并对一些关键指标进行对比，如下表所示。

表 2: FPGA 与 CPU 的性能比较

压缩格式		编码器	解码器	备忘录
H.264, H.265	典型参数	↓	↔	典型参数表示 8 位、420、2K。英特尔 QSV、GPU 硬核功能仍然强大，尤其是在编码器端。
	次要参数	↔	↑	次要参数表示 10 位、422、3K 或 5K。英特尔 QSV、GPU 硬核不支持它。
帧内		↑	↑	专有编解码器，诸如 Apple ProRes、Avid DNx、SONY XAVC、Panasonic AVC-Intra。图像格式，诸如 JPEG 和 PNG 等。

表注

↑ FPGA 提供更佳的性能。

↔ FPGA 和 CPU 提供同等的性能，但 FPGA 是卸载 CPU 负担的首选解决方案。

↓ FPGA 和 CPU 提供同等的性能，但 CPU 是首选解决方案。

视频流传输

在视频流传输应用中，常用的压缩格式是 H.264 或 H.265，因为终端（接收端）设备原生支持这些格式。诸如位深或色度和分辨率等参数通常为 8 位、4: 2: 0 和 1920×1080 或 1280×720。在解码器方面，基于 FPGA 的实现比基于 CPU 的系统提供更高的吞吐量。在数据层面，FPGA 效率更高，因为如果将 CPU 用于纯数据处理之外的其他任何与数据相关的任务时，它通常都没有得到充分的利用。然而在编码器方面，硬化的 CPU 编码器内核是专门针对这些典

型参数而设计的，并提供了足够的性能。

为了获得两全其美的效果，将 FPGA 和 CPU 解决方案相结合，并由 FPGA 来处理繁重的工作负载是理想的解决方案。FPGA 上的高效功能可以被移植到可重新配置的硬件上去运行。例如，运动估计算法是一种适合 FPGA 的工作负载。另一方面，CPU 更适合处理比特率控制算法。

一些服务提供商要求在软件解决方案中实现与 x264 相同的视频质量和流媒体格式。FPGA 和 CPU 的组合解决方案可以有效地满足这些要求。使用这种方法，每种功能都被合理地分配，较繁重的处理负载被转移到 FPGA，与纯软件解决方案相比，这种方法能提供类似或更好的视频质量和流媒体格式，而且编码时间显著减少。

下表列出了使用这种方法的 x264 评测结果，第一行显示了在 FPGA 上的运动估计函数 (x264_8_me_search_eref) 的结果。运动估计是 CPU 最繁重的工作负载之一，占据总处理时间的 21.2278%。

表 3: x264 评测结果 (通过评测软件获得)

样本	百分比	符号名称
3679706	21.2278	x264_8_me_search_ref
2078100	11.9883	x264_8_pixel_ads_mvs_ssse3
1453998	8.3880	x264_8_pixel_sad_x3_8x16_sse2
1176121	6.7849	x264_8_pixel_sad_x3_16x16_avx2
1156301	6.6706	x264_8_pixel_sad_x3_8x8_sse2
1095731	6.3211	x264_8_pixel_ads2_avx2
868943	5.0128	x264_8_pixel_sad_x3_16x8_avx2
779812	4.4986	x264_8_pixel_ads1_avx2
318990	1.8402	x264_8_pixel_ads_avx2
275943	1.5919	x264_8_quant_4x4_trellis
255712	1.4752	x264_8_trellis_cabc_4x4_psy_ssse3
231397	1.3349	x264_8_pixel_satd_8x8_interval_avx2
187422	1.0812	x264_8_mc_chroma_avx2

168559	0.9724	x264_8_pixel_satd_16x8_interval_avx2
168484	0.9720	x264_8_pixel_sad_8x8_mmx2

视频内容创作

用于内容创作的视频编辑软件支持多种压缩格式，其中包括 **Apple ProRes**、**Avid DNx**、**Sony XAVC** 和 **Panasonic AVC-Intra**，这些格式都带有基于内帧结构的专有压缩方案。此外，还有一些支持 **RAW** 模式的格式，诸如 **Apple ProRes RAW**、**RED RAW**、**ARRI RAW** 和 **Blackmagic RAW**，这些格式都得到了摄像机制造商的支持。由于这些格式（以及新型的和不断出现的格式）具有不断变化的特性，因此基于 **ASIC** 的解决方案并不实用，而需要基于 **FPGA** 的解决方案。

在过去，主要的分辨率为 **HD/2K**，**CPU** 具有足够的速度来处理这些视频流。但是，随着 **4K** 或 **8K** 分辨率变得越来越普遍，仅靠 **CPU** 加软件的解决方案不能够提供实时处理。另一方面，基于 **FPGA** 的解决方案可以轻松地实时处理 **4K** 和 **8k** 分辨率视频。

内部基准测试表明，即使与中级 **FPGA** 芯片相比，基于 **FPGA** 解决方案的处理速度也比最新的 **CPU** 加软件解决方案快五倍。虽然 **GPU** 可以提供与 **FPGA** 类似的性能，但其功耗更高、解决方案占用空间更大。

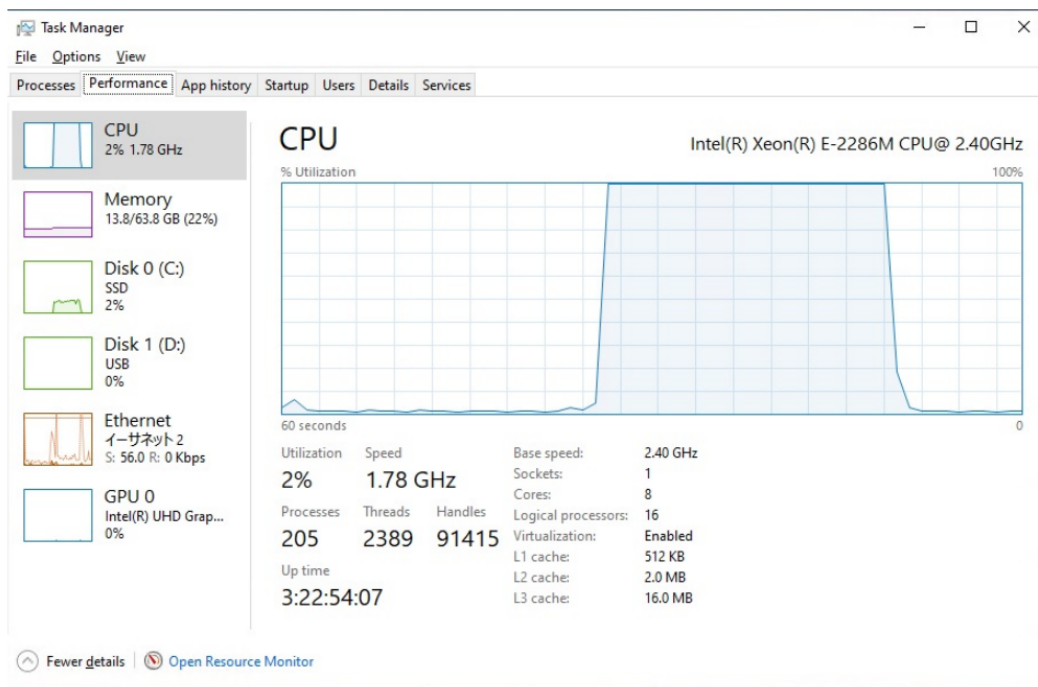


图 5: 仅使用 CPU (无 FPGA 卸载) 的处理方案性能

FPGA 解决方案的好处不仅仅在于加速，而且还能降低 CPU 的繁忙程度。在只有 CPU 的解决方案中，所有 CPU 周期都被 4K 或 8K 内容的编码所消耗，而使用 FPGA 来卸载编码任务可以释放 CPU 周期。因此，FPGA 加速器为该应用提供了最佳的解决方案，通过减少 4K 和 8K 视频制作所需的处理时间，来提高视频编辑人员的创作效率。

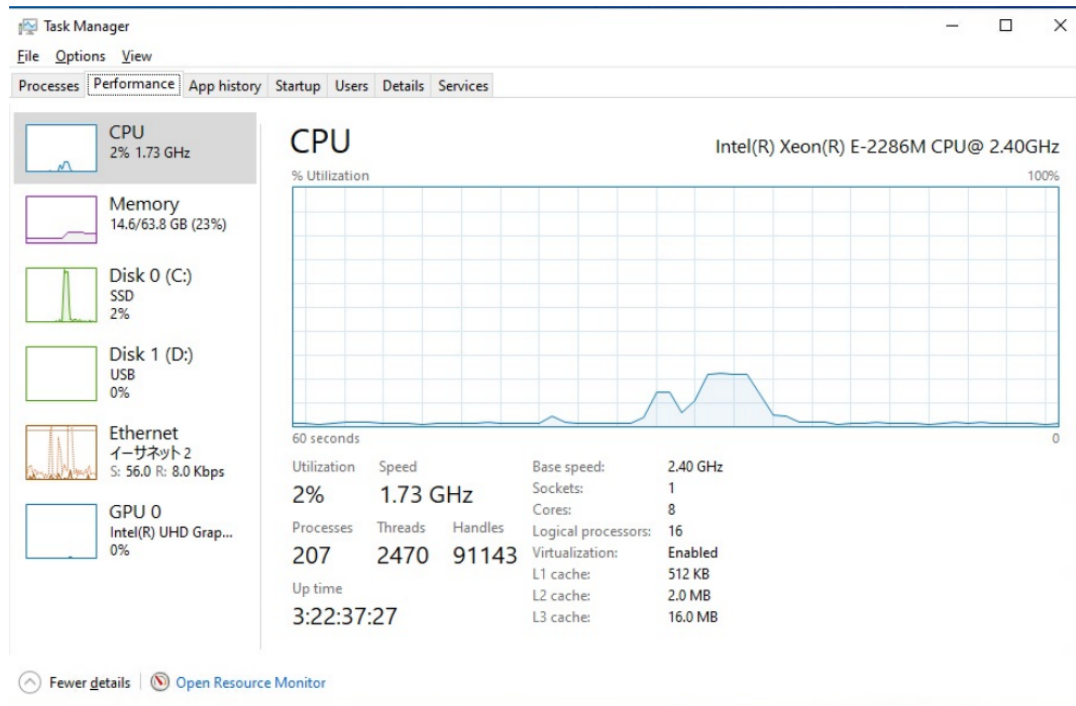


图 6: 使用FPGA 卸载的CPU 利用率

人工智能与深度学习

如前所述，在处理 H.264/H.265 解码方面，FPGA 提供了与 CPU 相当或更高的性能。如果解码器和内帧编码器（例如 JPEG 或 PNG）都位于同一 FPGA 中，那么基于 FPGA 的解决方案将提供比 CPU 更佳的性能。此外，在深度学习应用中，在将图像数据发送到深度学习处理之前，通常会进行一些图像预处理。在同一个 FPGA 上可以执行所有的处理，包括解码、图像处理和编码等（如图 7 所示），并且与 CPU 相比，FPGA 可以提供高吞吐量、低延迟和更少的数据事务。深度学习技术在现在和未来都将被广泛应用于各个行业或领域，而基于 FPGA 的解决方案将助力这一发展。



图 7：使用深度学习进行视频和图像处理的典型数据流

Decoding: 解码

Image Processing: 图像处理

Encoding: 编码

针对性能而优化的 Speedster7t 架构

Speedster7t FPGA 是专为满足最高性能的数据加速应用而设计的，该架构非常适合解决本白皮书中提到的所有应用挑战。具体而言，Achronix 开发了一种全新的创新型二维片上网络，它力助在 I/O 带宽、外部存储带宽和片上性能之间提供一种平衡架构，以确保总体最高的吞吐量。在传统的 FPGA 架构中，用户需要设计电路来连接加速器，从而导致并不理想的布局和布线。现在更新的 FPGA 架构使用一种网络，在逻辑阵列内的处理单元与各种片上高速接口和存储器端口之间传输数据流（如图 8 所示）。

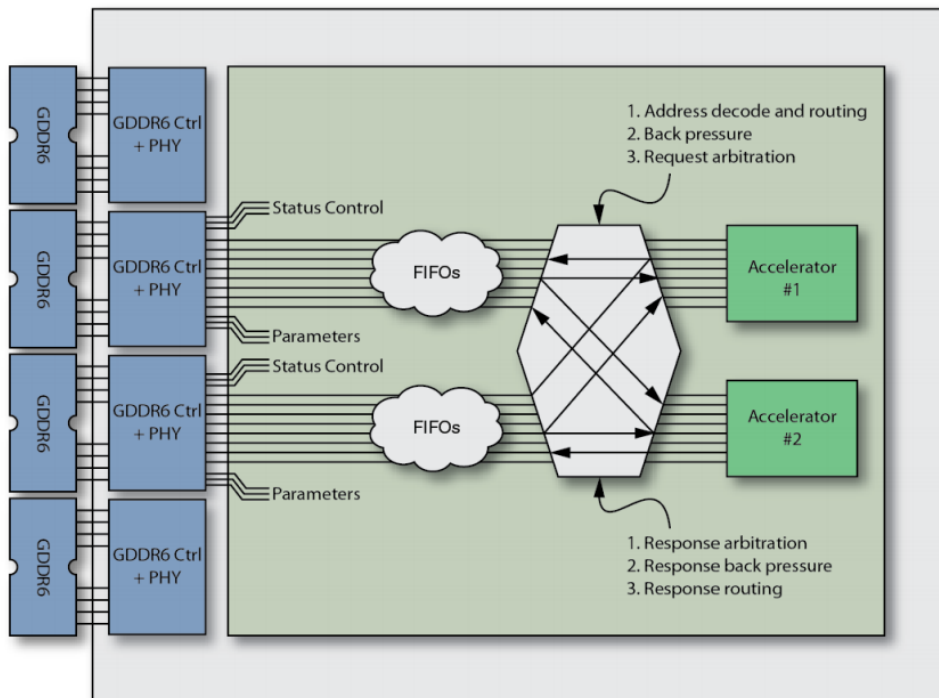


图 8：在传统的 FPGA 架构中连接加速器

Status Control: 状态控制

Parameters: 参数

Address decode and routing: 地址解码和布线

Back pressure: 背压

Request arbitration: 请求仲裁

Response arbitration: 响应仲裁

Response back pressure: 响应背压

Response routing: 响应布线

Accelerator: 加速器

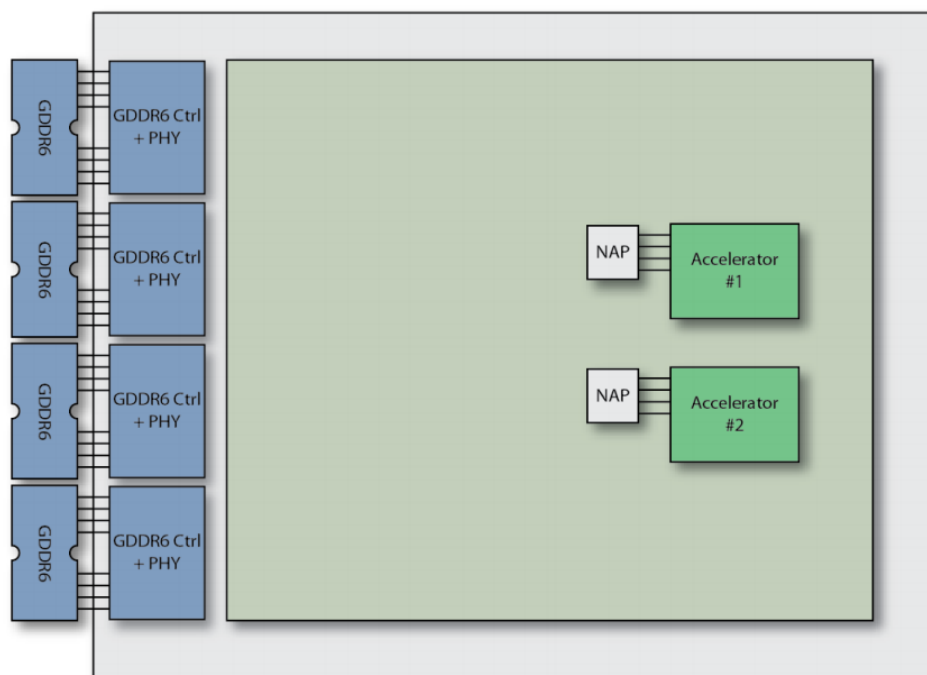


图 9: 先进的 FPGA 减少了所需的电路数量

硬连线架构极大地改善了处理的延迟和能效，但是缺乏应对需求变化的灵活性。Speedster7t 系列 FPGA 器件中的第一款芯片 AC7t1500 提供了一系列高速接口，包括可分配的（fracturable）以太网控制器（支持高达 400G 的速率）、PCI Gen 5 端口和多达 32 个 SerDes 通道，速率高达 112 Gbps。AC7t1500 器件是首款部署多通道 GDDR6 存储器接口的 FPGA，它满足了需要高速缓存海量数据的编码器的需求。除了在可编程逻辑阵列中采用的面向位的布线结构外，这些外围设备还通过一个智能二维片上网络进行互连。因此，Speedster7t FPGA 是第一款能够实现上述视频处理用例的器件，该 FPGA 器件利用一种平

衡架构，在计算密度和数据传输能力方面带来重大改进。

Speedster7t 架构通过提供总带宽超过 **20 Tbps** 的多级片上网络（NoC）层级化结构，消除了由于需要将高速 I/O 通道直接连接到以较低时钟速率运行的可编程逻辑所造成的瓶颈。与采用 **FPGA** 逻辑阵列实现互连方式相比，NoC 不仅在速率上有了大幅的提升，而且 NoC 还能在不消耗任何 **FPGA** 可编程资源的情况下传输大量数据。内部 NoC 不仅提供了更高的带宽，而且 **Speedster7t** **FPGA** 中的智能连接机制也简化了将数据从 NoC 端口传输到逻辑阵列中的任务。

这种架构可支持进一步的设计创新，例如支持上述机器学习用例的面向矩阵的算术单元。通过使用诸如深度学习或较为简单的统计方法等技术，设备可以分析数据流的模式，以观察和增强数据包在网络中的传输，并对不断变化的情况做出快速反应。概括而言，以下三项 **Speedster7t** 的架构创新为上述用例提供了更好的 **FPGA** 设计：

高速存储接口

Speedster7t 架构师对存储接口的选择反映出了以太网和 NoC 连接可提供的巨大带宽。一种可能的方法是在一系列产品设计中采用即将推出的 **HBM2** 接口。尽管这样的接口可以提供所需的性能等级，但 **HBM2** 是一种价格昂贵的选择，这将迫使客户去等待必要的组件和集成技术进入市场。

与此不同，**Speedster7t** 系列则采用了 **GDDR6** 标准，该标准为当今片外存储器提供了最高的性能。**Speedster7t** **FPGA** 是市场上首款支持该接口的器件，每个片上 **GDDR6** 存储控制器可维持 **512 Gbps** 的带宽。在单个 **AC7t1500** 器件中最多可带有八个 **GDDR6** 控制器，因此一个 **Speedster7t** **FPGA** 器件可提供高达 **4 Tbps** 的总存储带宽。

对 **PCIe Gen 5** 的支持

除了以太网和存储控制器，Speedster7t FPGA 上提供的对 PCIe Gen 5 的支持还能够与主机处理器紧密集成，以支持高性能加速器应用。PCI Gen 5 控制器使其能够读取和写入存储在 FPGA 存储层级结构中的数据，包括许多位于逻辑阵列内的块 RAM，以及连接到 FPGA 存储控制器的外部 GDDR6 和 DDR4 SRAM。在 FPGA 逻辑阵列中实例化的数据传输控制器（例如 DMA 引擎），可以类似地通过 PCIe Gen 5 总线访问与主机处理器共享的内存。无需消耗 FPGA 逻辑阵列内的任何资源即可实现这种高带宽连接，并且设计时间几乎为零。用户只需启用 PCIe 和 GDDR6 接口，即可通过 NoC 发送事务数据。

PCIe 子系统与任何 GDDR6 或 DDR4 存储接口之间的直接连接如下图 10 所示。

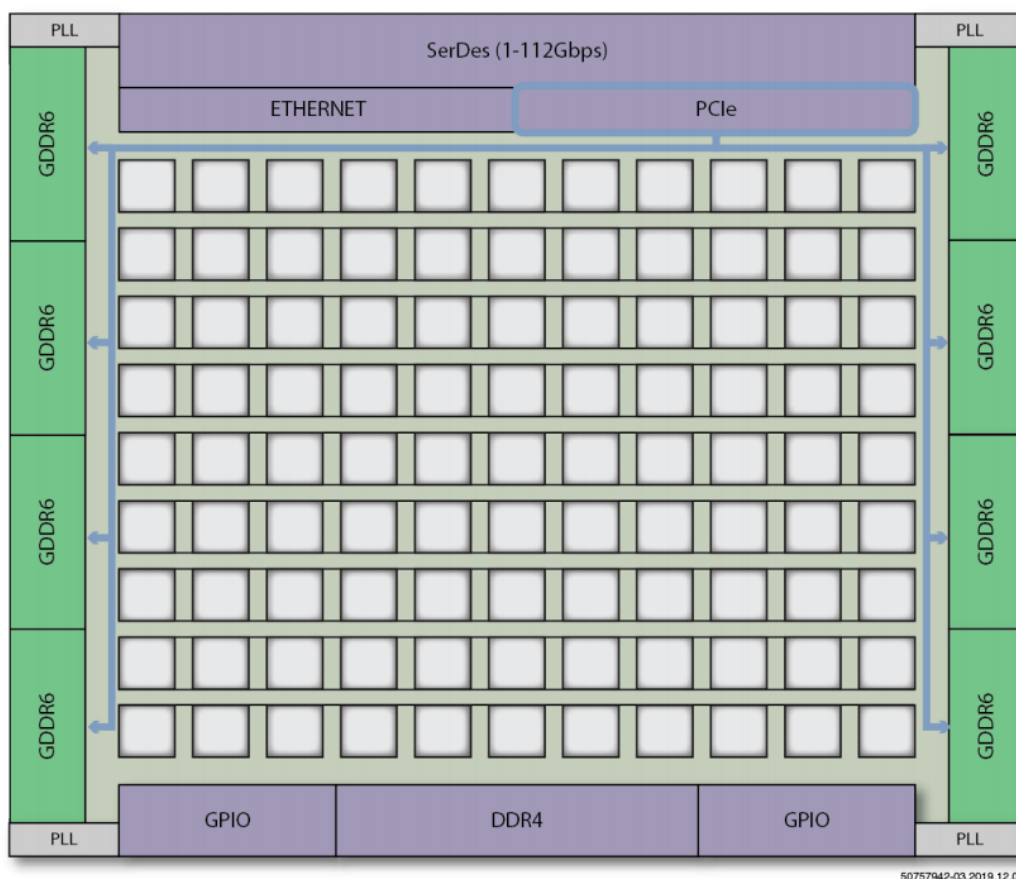


图 10：无需消耗 FPGA 逻辑阵列即可实现 PCIe 和 GDDR6 之间的数据传输

机器学习处理器（MLP）

对于计算密集型任务，在 Speedster7t FPGA 上部署的 Speedster7t 机器学习处理器（MLP）是灵活的且可分配的算术单元。机器学习处理器是高密度乘法器阵列，带有支持多种数字格式的浮点和整数 MAC 模块。机器学习处理器带有集成的存储模块，可以在不使用任何 FPGA 资源的情况下执行操作数和存储级联功能。机器学习处理器适用于一系列矩阵数学运算，从 5G 无线电控制器的波束成形计算到加速深度学习应用，如视频处理系统所需的数据流模式和数据包内容分析。

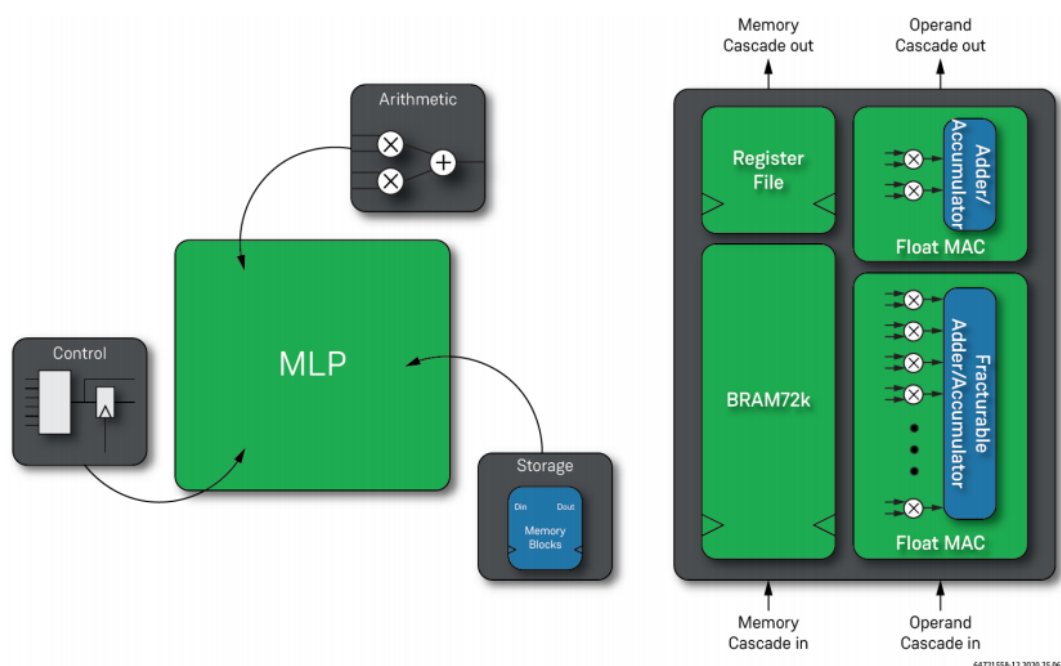


图 11: 机器学习处理器原理框图

结论

虽然 ASIC 的性能通常很高，但它只支持设计时设想的功能集，不能进行现场升级；CPU 是最灵活且最易于设计的，但是其时钟频率已经难以提升，其性能大幅提升的时代已经结束；随着工作负载逐年增加，CPU 已无法满足需求。FPGA 在性能和灵活性之间提供了良好的平衡。由于需要大量的并行处理，因此视频编码、解码和图像处理算法都更适合于用 FPGA 来实现。总之，基于 FPGA 的解决方案可以缩短上市时间，具有高度的可定制性，并且可以有效地

用于实现不断发展的算法。

IBEX Technology 公司简介

IBEX Technology 公司总部位于日本，可提供视频编解码 IP、IC 设计服务和咨询，诸如 ASIC 和大规模 FPGA 设计、PCB 等。IBEX 开发了多种视频编解码器，诸如 MPEG-2、H.264 / AVC、Apple ProRes、Avid DNxHD、SONY XAVC、Panasonic AVC-Intra 和 H.265 / HEVC。IBEX Technology 公司主要在广播设备行业有着良好的业绩记录，这得益于其产品的高质量、稳定性和可靠性。

IBEX 于 1985 年在日本神奈川县（Kanagawa）厚木市（Atsugi）成立，是公认的大规模集成电路（LSI）和视频系统设计领域的领导者。过去，IBEX 的主要用户包括日本电视机制造商和广播公司，并为 ASIC 厂商提供了视频编解码器 IP。在过去的十年中，由于 FPGA 的性能和成本竞争力，该公司的重点已转向提供基于 FPGA 的解决方案。IBEX 面向全球用户提供 FPGA IP。

版权所有©2020 Achronix 半导体公司保留所有权利。

Achronix、Speedcore、Speedster 和 ACE 是 Achronix 半导体公司在美国和/或其他国家/地区的商标。所有其他商标均为其各自所有者的财产。所有规格如有更改，恕不另行通知。

免责声明

本文件中所提供的信息被认为是准确和可靠的。但是，Achronix 半导体公司不在此类信息的完整性或准确性作出任何声明或担保，并且对于使用本文包含的信息不承担任何责任。Achronix 半导体公司保留随时更改本文件及其所含信息

的权利，恕不另行通知。所有 Achronix 商标、注册商标、免责声明和专利均在网站上列出 <http://www.achronix.com/legal>。