



Aupera 视频与 AI 加速平台颠覆数据中心流媒体市场

与 x86 系统相比，专用分布式架构可提供高 33 倍的性能，低 90% 的能源成本与机架空间占用

概述：

Aupera Technologies 是数据中心视频处理系统领域的新兴企业。Aup2600 是一种专用分布式视频处理系统，内置 48 个赛灵思 Zynq® UltraScale+™ MPSoC。此外，Aup2600 还提供基于赛灵思 Vivado 环境的完整视频 +AI 软件框架和用于神经网络处理的深度学习处理器单元 (DPU)。

客户：Aupera Technologies

行业：数据中心

www.aupерatech.com

挑战：

替代数据中心内用于视频处理、转码和 AI 分析应用的 x86 服务器



解决方案：

采用赛灵思工具、视频和机器学习 IP 实现基于 Zynq MPSoC 的分布式处理架构与软件框架。

成效：

单个 Aup2600 替代 30 个 x86 E5 服务器，将性能提高 33 倍，并降低 90% 的功耗与空间占用 (图 1)。

图 1：Aup2600 采用 48 片赛灵思 Zynq MPSoC 器件

挑战：**改变数据中心内视频处理的现状**

成立于 2014 年的 Aupera Technologies 是数据中心视频处理系统领域的新兴企业，以实现流媒体应用的“实况视频”为使命。该公司的业务重点是在系统级实现异构计算架构，构建高效率视频处理平台。到 2021 年，80% 以上的互联网流量将被视频占据（来源：思科），在这样的大背景下，传统的 x86 系统将变成计算瓶颈。公司 CEO 廖玉峰 (Roy Liao) 博士表示：“我们的目标是用全新架构与软件框架解决当前实时流媒体业务的痛点，改变数据中心视频处理的现状。”

编解码等视频处理功能计算强度极大。使用通用型 CPU，在软件中完成全部处理，在流媒体普及化并大幅增长的背景下，已难再有突破。即便堆叠 CPU 也难以进一步提高效率，尤其是对实时流视频应用而言。为了解决 CPU 数据中心的瓶颈，Aupera 与中国最大的实时流媒体视频公司、活跃用户数量上亿的欢聚时代公司 (YY, Inc) 合作，为大规模实时视频转码与内容分析设计 Aup2600 系统。

Aup2600 是一种专用分布式视频处理系统，内置 48 片赛灵思 Zynq® UltraScale+™ MPSoC，能同时对 380 个高清 1080p 视频流（兼容 H.264/265）的并行转码进行处理。除了它独特的架构，Aup2600 还采用基于赛灵思 Vivado 环境的完整视频 +AI 软件框架，以及用于神经网络处理的深度学习处理器单元 (DPU) 引擎。这种软件框架包含面向 xfOpenCV 计算机视觉库定制的引导加载程序和加速器、视频编解码器，以及用于对象检测与特征提取的深度学习算法。此外，Aupera 也成功地将 FFmpeg 流媒体平台迁移到 ARM 应用处理核上，借助感兴趣区域 (ROI) 优化，建立完整的视频转码应用。

解决方案：**通过 MPSoC 架构，使用基于 Vivado 和 AI 环境构建的软件框架，实现快速部署**

Aup2600 项目于 2018 年 4 月正式启动。在简要考虑 ASIC 替代方案（在视频处理灵活性上存在问题）后，Aupera 工程团队选择了 Zynq UltraScale+ MPSoC。团队中的工程师都曾长期为 MILCOM Telecom 和 NASA 工作，积累了深厚的 FPGA 开发专业能力。仅用六个月的时间，Aup2600 项目就完成了最初的实验室测试、系统集成开发与测试、实地测试以及商业测试，赢得首个产品订单并与 YY 共同完成部署工作。

廖博士表示：“FPGA 可同时提供硬件计算速度与软件灵活性。值得一提的是，虽然 ZU7EV MPSoC 是一种复杂的异构器件，但也是极为先进的设计。它内置 4 个 ARM 处理器内核、视频编解码单元和充裕的 FPGA 逻辑资源。我们比较过各种器件，发现这种 MPSoC 最适合用来设计我们的创新系统，从而实现视频处理的优化。”（图 2）。

在与 YY 合作的最初阶段，Aupera 需要解决低时延、高效率视频转码的快速部署，并支持运行在 FPGA 系统上的 AI 功能的未来无缝升级。此外，为了支持 YY 的实况直播和广播，Aupera 还需要对不适当的内容进行监控与过滤。Aupera 还要知道应推送什么样的标准内容才能符合客户的口味。Aupera 即将在下季度部署 ROI 应用，利用 FPGA 做人脸检测，视频编码和特效，提高直播图像质量。

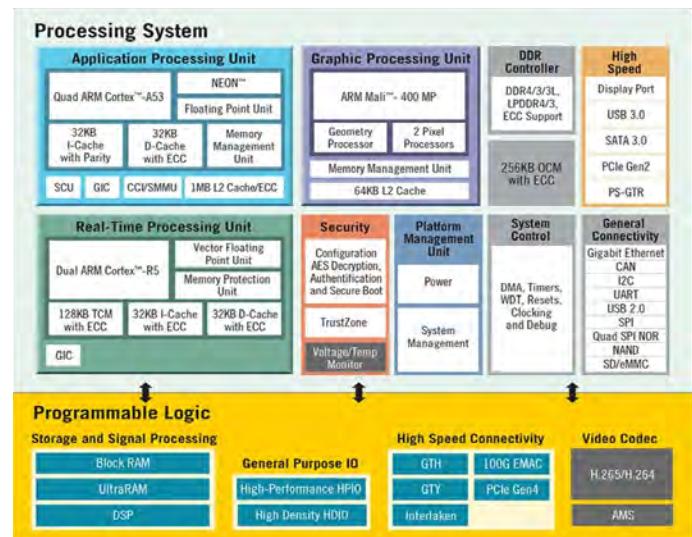


图 2: MPSoC 原理图

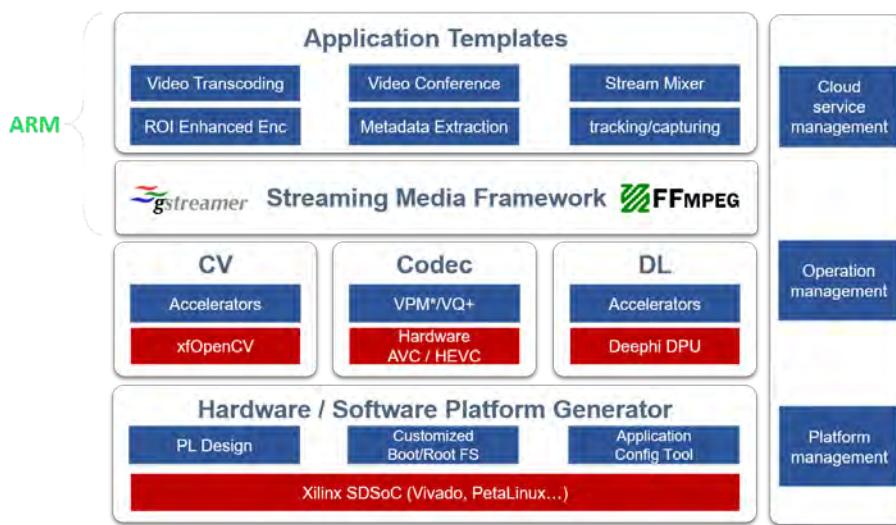


图 3: Aupera 的视频+AI 软件框架

Aupera 在赛灵思完成的工作基础之上，为 Caffe 等常用模型框架提供综合的 AI 环境。廖博士说：“通过我们的视频 +AI 软件策略，我们能够帮助客户简化部署工作，以最快速度完成部署。”他补充道：“在这个应用的基础上，我们还提供模板，方便客户使用这些模板继续开发和设计自己的应用。这种做法非常高效，客户能更容易地采用我们的新架构。我们把它称为‘视频天才’计算架构，并将它视为数据中心的未来。”（图 3）。

“对于像 YY 这样有着数百万用户的实时流媒体应用来说，控制时延也至关重要。”

通道成本的情况下可提供极高画质的视频服务。采用单个 Aup2600 系统运行统一的视频 +AI 功能，YY 不仅可以淘汰专门用于视频转码的传统服务器，还能将其中的部分服务器用于视频内容分析。与此同时，在 FPGA 上加速对象和特征检测神经网络算法，还能实现实时视频分析。（图 4）。

廖博士评述道：“虽然赛灵思的视频编解码单元 (VCU) 是硬编码 IP，但它具备足够的灵活性，能够支持数量不断增加的视频工作负载。我们的综合性解决方案不仅能够满足要求高密度的应用的需求，还能提供低时延。“对于像 YY 这样有着数百万用户的实时流媒体应用来说，时延控制至关重要。”

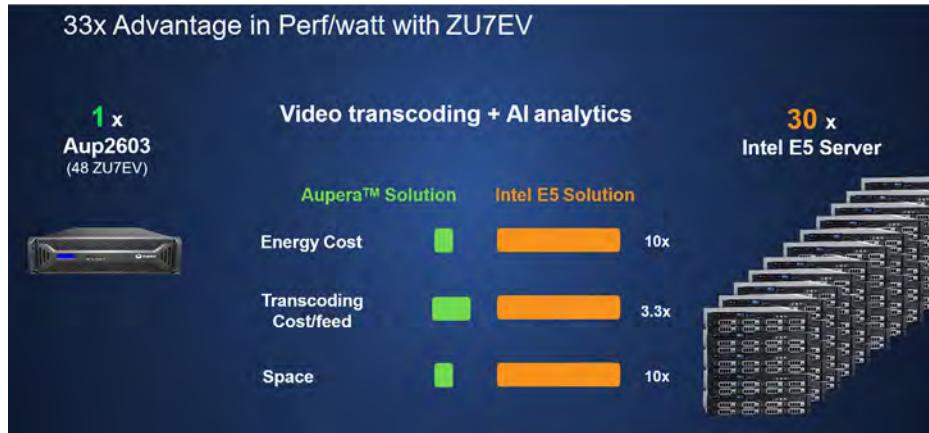


图 4: Aup2600 替换了数据中心内的 30 部视频处理 x86 服务器。

结论：

总之，Aupera 对与赛灵思的合作伙伴关系非常满意。廖博士总结说：“赛灵思对自己的合作伙伴非常开放。他们的生态系统对我们很有帮助，其“数据中心优先”也是非常优秀的策略。随着 5G 市场与 IoT 市场的兴起，我们正面临着一场视频数据的庞大洪流。这就要求数据中心具备强大的处理能力。基于 FPGA 的视频处理将成为未来数据中心最重要的计算能力。”

企业总部

Xilinx, Inc.
2100 Logic Drive
San Jose, CA 95124
USA
电话: 408-559-7778
www.xilinx.com

赛灵思欧洲

One Logic Drive
Citywest Business Campus
Saggart, County Dublin
Ireland
电话: +353-1-464-0311
www.xilinx.com

日本

Xilinx K.K.
Art Village Osaki Central Tower 4F
1-2-2 Osaki, Shinagawa-ku
Tokyo 141-0032 Japan
电话: +81-3-6744-7777
japan.xilinx.com

Asia Pacific Pte. Ltd.

Xilinx, Asia Pacific
5 Changi Business Park
Singapore 486040
电话: +65-6407-3000
www.xilinx.com

印度

Meenakshi Tech Park
Block A, B, C, 8th & 13th floors,
Meenakshi Tech Park, Survey No. 39
Gachibowli(V), Seri Lingampally (M),
Hyderabad -500 084
电话: +91-40-6721-4747
www.xilinx.com