

GP100 率最新 Quadro Pascal 家族亮相，NVIDIA 高端显卡剑指制造业应用

撰文 / 齐健

2017 年 4 月 12 ~ 14 日，NVIDIA 携全新发布的 Pascal 架构 Quadro 新品，亮相 InfoComm China 2017 展会。在展会首日，NVIDIA 在国家会议中心的大会现场召开媒体发布会，公开展示了 GP100、P6000、P5000 等全新 Quadro Pascal 架构高端专业级显卡。会上，NVIDIA 专业可视化业务高级总监 Sandeep Gupte 及英伟达中国区销售总监何犹卿向记者展示了全新 Quadro 显卡家族的强大性能和多种应用场景。

Pascal 架构的 Quadro 显卡主要定位于专业级工作站，完成包括仿真分析、VR 模拟以及视频渲染等工作。与上一代相比，全新 Quadro 产品的运算速度提升了两倍（基于 SPECviewperf 12 基准的计算结果），且具有超快速内存。新 NVIDIA 专业可视化业务高级总监 Sandeep Gupte 说：“NVIDIA 的愿景是希望用户凭借 Quadro 产品家族的这些新产品，简化设计的工作流程，从传统的工作站与工作方式中解脱出来，把工作站打造成个人的超级计算机。”

全新 Quadro 旗舰显卡 GP100

Sandeep Gupte 对 Quadro 家族的旗舰产品 GP100 进行了重点介绍。GP100 基于 Pascal 核心，单精度 CUDA 核心数 3584 个，双精度 CUDA 核心数 1792 个，具有 16 GB HBM2 高带宽显存，半精浮点运算性能达到 20.7 TFLOPS、单精浮点运算性能为 10.3 TFLOPS、双精浮点

运算性能为 5.2 TFLOPS。同时，支持 4K、5K 以及 VR 显示。

GP100 具有以下特点：

◎ 统一仿真、HPC、渲染和设计：GP100 将双精度性能（5 TFLOPS 以上，比 Quadro K6000 快将近 3 倍）与 16GB 的高带宽内存（HBM2）相结合，因此用户可以在设计过程中进行仿真，并以前所未有的速度采集逼真的多物理场仿真。客户能够将两个 GP100 GPU 与 NVLink™ 技术相结合，并扩展到 32GB 的 HBM2，以在单一工作站上创建庞大的可视化计算解决方案。

◎ 探索深度学习：GP100 提供超过 20 TFLOPS 的半精度计算（单精度（FP32）性能为 10 TFLOPS，以半精度（FP16）模式运行可使吞吐量翻倍至 20 TFLOPS），可谓在 Windows 和 Linux 环境中进行深度学习的理想开发平台。





何犹卿

◎ 将 VR 集成到设计和仿真工作流程中：VR 认证的 Quadro GP100 和 P4000 具有创建详细、逼真、沉浸式环境的能力，能够大规模实现更加宏大、更加复杂的设计体验。

◎ 尽享照片级真实感设计的优势：基于 Pascal™ 架构的 Quadro GPU 渲染照片级真实感图像的速度为 CPU 的 18 倍（基于采用 Iray 渲染技术的内部测试，在 720p 场景中，比 Intel Xeon E5 2697 V3 2.6 GHz 14 核 CPU 快 18 倍）。

◎ 创建广阔的可视化工作区：可在多达四个 5K 显示器上以高分辨率和 HDR 颜色显示数据。

◎ 经济高效地构建超高分辨率及多显示屏配置：通过将达 8 个 P4000 GPU 和两个 Quadro Sync II 显卡（Quadro 驱动程序目前支持使用单个 Sync II 显卡配置多达 16 个 4K 显示器。未来驱动程序更新中还将添加配置多达 32 个 4K 显示器的功能）相结合，可通过单一机箱实现多达 32 个 4K 显示器的配置。

GP100 在仿真与渲染领域的应用

GP100 在 CAE 方面表现突出，可以帮助工程师快速完成包括结构力学、流体动力学和电磁学等方面的双精度计算。同时，GP100 可以兼顾设计与仿真工作的需要，在同一个工作站中一卡两用。Sandeep Gupte 介绍说，在 Simulia 和 ANSYS 的 CAE 分析过程中，利用 GP100 加速比之只使用 CPU 进行运算仿真效率得到了大幅提高，Simulia 的仿真速度提高了 3 倍，ANSYS 则提高了 2 倍。除了性能提升之外，GPU 加速还可以在软件许可采购方面帮助用户节省开支，用户可以利用 GPU 取代部分 CPU 运算，从而减少在软件多核节点方面的开支。

在软件应用方面，NVIDIA 与达索系统、西门子 PLM、PTC 等多家软件公司合作，在软件中开发专门针对 GPU 加速的功能。例如，NVIDIA 与达索系统的深度合作，在 CATIA 的渲染模块中，采用了 Iray 技术，在 SOLIDWORKS Visualize 中，也可以应用 GPU 加速，把渲染时间从 5 小时缩短到 30 分钟。

Sandeep Gupte 表示，NVIDIA 在全球组建了一个由软件工程师组成的技术团队叫做 Def Tech，他们的工作就是与本地企业的软件工程师开展技术合作，确保用户的应用软件可以在硬件上流畅运行。

Quadro 显卡在娱乐影音渲染方面的应用更为广泛，英伟达中国区销售总监何犹卿表示，GPU 加速对电影渲染来说非常重要。他以《变形金刚 3》的一帧 4K 画面为例比较，只用双核 CPU 渲染需要 288 个小时的画面，GPU 加速大概只需要 3 个小时。

QUEST 计划

在本次媒体会上，Sandeep Gupte 还提到了 NVIDIA 的 QUEST (Quadro Embedded System Program) 计划，即基于 Quadro 的嵌入式系统计划。目前，NVIDIA 已经面向全球开放了针对高性能嵌入式系统的开发者计划，开发企业可以申请加入此项计划，与 NVIDIA 共同设计适用于不同软件解决方案的 GPU。除此之外，NVIDIA 还为这些开发者提供设计包 (Design Kit)、SDK、样本代码以及培训课程等。

Sandeep Gupte 强调说：“要打造嵌入式的解决方案，对品质、可靠性和性能的要求都非常高。因此我们关注的焦点是找到能够在不同行业中实现这三点的合作伙伴，不拘数量。QUEST 计划不收取费用，但是我们在遴选合作伙伴的时候也会认真考虑，因为我们在技术团队方面要与这些企业开展密切地合作。因此，我们要针对目标市场选对合作伙伴。” 