

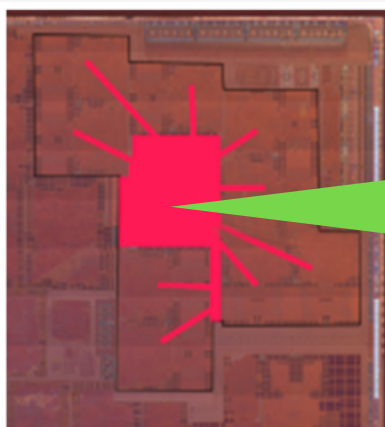
IMG B 系列

多核 技术

IMG B 系列引入了一种去中心化、
松耦合的多核性能可扩展性方法。

传统的GPU可扩展方法

传统的GPU 可扩展性方法受到集中共享模块和着色器内核之间的连接限制。通常，共享逻辑包括集中式内存数据通路、作业管理器和几何分片引擎。这种中心化依赖性产生了一种星形网状结构，其中所有内核都需要连接到单一的中央单元。但是，这会导致拥塞和布局灵活性问题，如图1所示。



复杂中心模块，作业管理器
分片，缓存

图1，传统的多着色器核心设计，带中心模块和星形网络



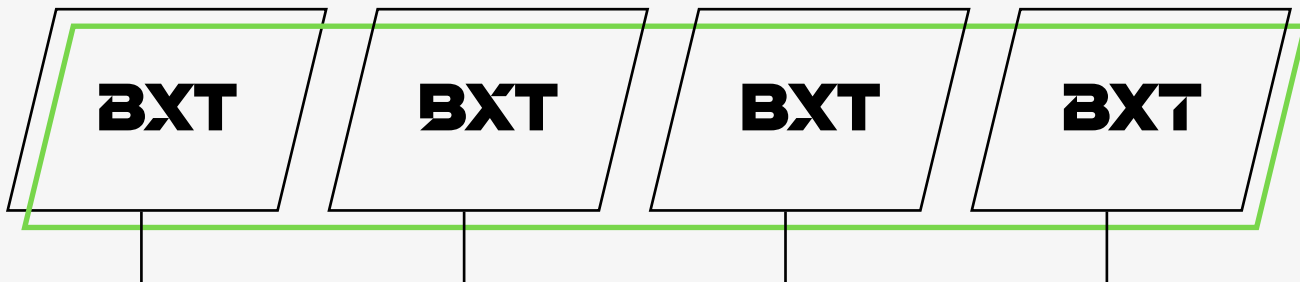
扫一扫关注官方微信号
► 点击中文社区查看更多

Imagination的去中心化多核

设计理念

Imagination多核方案可灵活调整GPU内核数量，无需依赖所连接的中央单元。简单来说，该方案可被看作是SoC设计中的多个GPU，具有多内核共同处理计算和图形的能力。

Imagination的GPU是基于分块延迟渲染，通过不同分块上的各个内核工作，使多个GPU协同工作，以完成整个渲染目标。



Imagination去中心化多核的四核配置

实际上，内核是松耦合的，仅在内存中共享命令和分块缓冲区列表，并在它们之间共享工作负载。由于每个内核都设计为完全独立的GPU，因此它包含基于优先级的自我管理和执行工作负载所需的所有功能。其中主要变化为，以前是直接处理工作负载的单一内核式布局，而现在是多个GPU共享命令流并尽快共同完成工作。

当每个GPU内核持续渲染屏幕连贯区域时，通过处理渲染目标中的不同区域，确保最大命中率，以保证带宽效率。类似的方法也适用于其它处理类型，包括几何和计算处理，它们可以分配给不同的GPU内核来处理。

更灵活的物理设计

每个GPU内核都直接位于SoC交换结构上，提供更好的布局 and 更灵活的设计。数据被缓存于每个GPU内核的特定工作中，从而避免了不必要的数据迁移，以及各内核间的数据重复，以提升效率。

通过使用如AXI这样的总线结构标准，可使多核GPU适用于不同的设计，可将GPU内核分布到多个芯片小片、芯片甚至是开发板上。这让客户可以更灵活地选择设计方案，并大大降低成本，因为客户可以设计单个芯片小片（或芯片），作为随后构建多性能点的基础模块，只需封装一个或多个芯片小片来扩展GPU性能。

实现成本降低主要在于，只需制造一个芯片小片，通过封装即可满足多个细分市场的产品需求。而在以前，这需要极高成本制造多个芯片。



扫一扫关注官方微信号

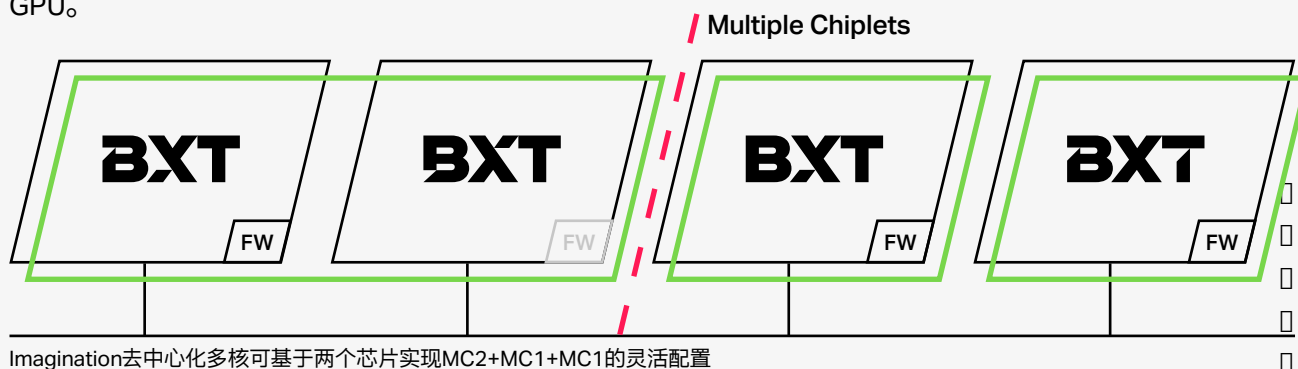
► 点击中文社区查看更多



imaginationtech.com

更灵活的应用

由于每个内核都是一个完整的GPU，能够独立工作或者多个协同工作，可实现最大的使用灵活性。在单独GPU使用时，内核可由驱动程序直接控制，执行特定工作负载，也可以协同工作以形成性能更强的GPU。

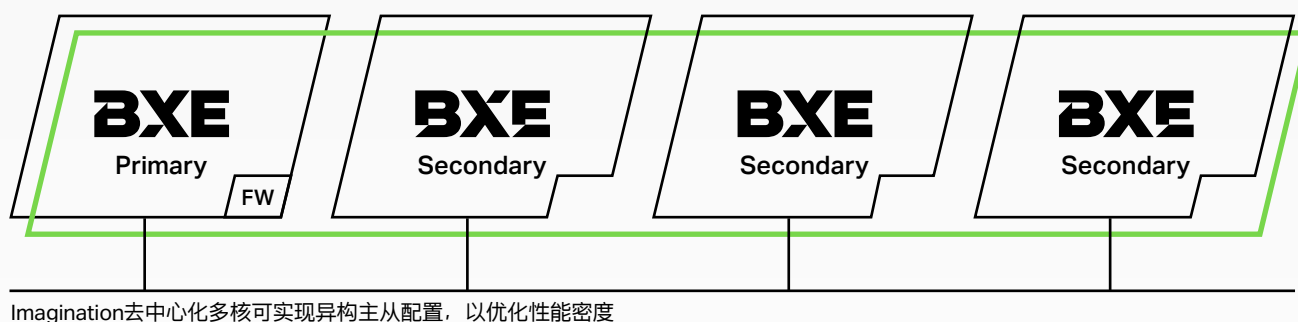


Imagination的HyperLane技术结合使用，基于优先级的调度机制（单GPU模式）实现虚拟化，也可以使用更简单的隔离方法，其中特定内核可以分配给特定虚拟机（多个单GPU）。当然，这实际上等同于冗余性处理，也用在我们的BXS系列中，我们用它来确保ISO 26262功能安全性，其中GPU可以相互检查彼此的操作。

成本优化

对于低成本细分市场，带有重复功能的GPU设计不合理也没有必要。可通过创建一个主GPU内核和一个辅助GPU内核来解决。主GPU内容包括所有功能，而辅助GPU内核删减了不必要的功能，如固件处理器和几何处理管线。

这意味着主/从配置明显小于多主核配置，同时仍具有足够的性能来处理计算，并优先进行几何处理，同时实现多GPU内核基于固件的事件处理。实际上，这是一种异构扩展形式，其中具有不同功能集的GPU可以高效协同工作。



扫一扫关注官方微信号

► 点击中文社区查看更多

 Imagination

imaginationtech.com

去中心化多核技术如何让市场获益？

手机/平板电脑

通过避免GPU设计中到单个中心点的复杂数据通道，去中心化多核技术允许更大程度的模块重用和更轻松的布局，否则必须重新设计每个性能点。单个芯片小片设计还可以实现多个性能点，其中单个芯片小片面向高端移动市场，双芯片小片瞄准高端移动市场，四芯片小片面向移动超便携/平板电脑设计。

汽车

多核方法固有的冗余性允许低成本的 ISO 26262 合规性，其中内核可以相互检查彼此的正确操作。但是，在不需要功能安全性的情况下，可以根据使用情况动态重新配置内核，以最大限度地提升多 GPU 内核的性能。当然，通过按需封装芯片小片，可以以更低成本支持从入门级到高级汽车系统的性能扩展。

数字电视 (DTV) 和机顶盒 (STB)

对成本敏感的市场，可以通过使用具有多个辅助 GPU 内核的单个主内核来优化性能扩展。这可实现优质的 1080p GUI 渲染解决方案，该方案可与芯片小片技术进行组合，只需封装多个芯片小片即可提供 4K GUI 渲染解决方案，从而扩展性能。

服务器

服务器需要灵活的性能扩展，这与云游戏密切相关，其中 Imagination 灵活的多核方法可使用 HyperLane 技术来动态重配置 GPU 网格。一个 BXT MC4 设计，使用 7nm 工艺以 1.5GHz 运行时，可提供高达 6 TFLOPS 的性能，这比下一代中等水平的主机性能超出 50%。与多个视频编码器结合时，可将多个 B 系列 GPU 用作单云游戏或远程桌面芯片组。使用 HyperLane 技术，可为多个用户实现性能保证。例如，一个 MC4 可以高效的方式为多达 36 个用户提供灵活的性能分配，并使用硬件虚拟化和隔离来确保隐私安全。



扫一扫关注官方微信号

► 点击中文社区查看更多



imaginationtech.com