



(12) 发明专利

(10) 授权公告号 CN 102597951 B

(45) 授权公告日 2016. 05. 04

(21) 申请号 201080049244. 9

(22) 申请日 2010. 09. 03

(30) 优先权数据

61/239, 730 2009. 09. 03 US

12/616, 636 2009. 11. 11 US

(85) PCT国际申请进入国家阶段日

2012. 04. 28

(86) PCT国际申请的申请数据

PCT/US2010/047784 2010. 09. 03

(87) PCT国际申请的公布数据

W02011/028984 EN 2011. 03. 10

(73) 专利权人 先进微装置公司

地址 美国加利福尼亚州

(72) 发明人 G·萨多斯基 K·伊乌尔查

J·布拉泽尔

(74) 专利代理机构 北京戈程知识产权代理有限公司

公司 11314

代理人 程伟 王锦阳

(51) Int. Cl.

G06F 9/38(2006. 01)

G06F 9/46(2006. 01)

G06F 9/50(2006. 01)

(56) 对比文件

US 2007/0294696 A1, 2007. 12. 20,

EP 1557755 A1, 2005. 07. 27,

US 2004/0160449 A1, 2004. 08. 19,

CN 100489820 C, 2009. 05. 20,

审查员 陈敏

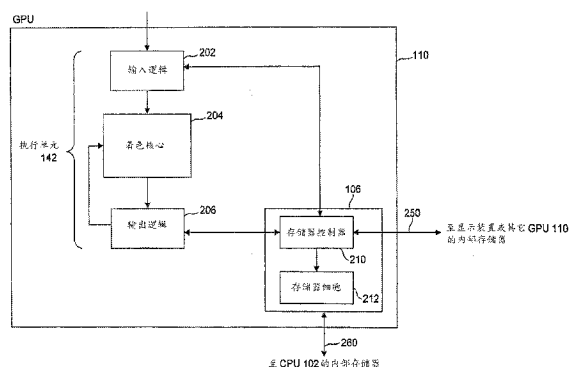
权利要求书2页 说明书8页 附图5页

(54) 发明名称

一般目的使用的内部处理单元的存储器

(57) 摘要

本发明揭露一种具有一般目的使用的内部存储器的图形处理单元 (GPU) 与其应用。此GPU包括第一内部存储器、耦接至该第一内部存储器的执行单元以及配置以耦接该第一内部存储器至其它处理单元的第二内部存储器的接口。该第一内部存储器可包括堆栈动态随机存取存储器 (DRAM) 或嵌入 DRAM。接口可以进一步配置以耦接该第一内部存储器至显示装置。GPU 也可包含另一接口, 配置以耦接该第一内部存储器至中央处理单元。此外, GPU 可体现在软件和 / 或包含在计算机系统中。



1. 一种图形处理单元 GPU, 包括:

第一内部存储器, 包括存储器控制器和存储器细胞;

耦接至该第一内部存储器的第一执行单元, 其中, 该第一执行单元和该第一内部存储器在相同晶粒上; 以及

配置以提供高带宽专用内部存储器存取至额外的处理单元的第二内部存储器的接口, 且进一步配置通过该额外的处理单元以提供高带宽专用内部存储器存取至该第一内部存储器, 该第一内部存储器能与该第二内部存储器结合以供该额外的处理单元使用。

2. 根据权利要求 1 所述的图形处理单元 GPU, 其中, 该额外的处理单元包括 GPU。

3. 根据权利要求 1 所述的图形处理单元 GPU, 其中, 该额外的处理单元包括中央处理单元 CPU, 以及该第二内部存储器为缓存存储器。

4. 根据权利要求 1 所述的图形处理单元 GPU, 其中, 该第一内部存储器包括堆栈动态随机存取存储器。

5. 根据权利要求 1 所述的图形处理单元 GPU, 其中, 该第一内部存储器包括嵌入动态随机存取存储器。

6. 根据权利要求 1 所述的图形处理单元 GPU, 其中, 该接口进一步配置以耦接该第一内部存储器至显示装置。

7. 一种计算系统, 包括:

第一 GPU, 包括有包括第一存储器控制器和第一存储器细胞的第一内部存储器、以及耦接至该第一内部存储器的第一执行单元, 其中, 该第一执行单元和该第一内部存储器在第一相同晶粒上;

第二 GPU, 包括有包括第二存储器控制器和第二存储器细胞的第二内部存储器、以及耦接至该第二内部存储器的第二执行单元, 其中, 该第二执行单元和该第二内部存储器在第二相同晶粒上; 以及

接口, 配置通过该第一 GPU 以提供高带宽专用内部存储器存取至该第二内部存储器, 且进一步配置通过该第二 GPU 以提供高带宽专用内部存储器存取至该第一内部存储器, 该第一内部存储器能与该第二内部存储器结合以供该第二 GPU 使用。

8. 根据权利要求 7 所述的计算系统, 其中, 该第一内部存储器包括堆栈动态随机存取存储器。

9. 根据权利要求 7 所述的计算系统, 其中, 该第一内部存储器包括嵌入动态随机存取存储器。

10. 根据权利要求 7 所述的计算系统, 其中, 该接口进一步配置以耦接该第一内部存储器和该第二内部存储器至显示装置。

11. 根据权利要求 7 所述的计算系统, 进一步包括:

外部存储器;

中央处理单元 CPU, 包括缓存存储器; 以及

耦接在该外部存储器与 CPU 之间的总线。

12. 根据权利要求 11 所述的计算系统, 其中, 该第一 GPU 进一步包括配置以提供高带宽专用内部存储器存取至该 CPU 的该缓存存储器的额外的接口, 以及进一步通过该 CPU 以提供高带宽专用内部存储器存取至该第一 GPU 的该第一内部存储器。

13. 根据权利要求 11 所述的计算系统, 其中, 该第二 GPU 进一步包括配置以提供高带宽专用内部存储器存取至该 CPU 的该缓存存储器的额外的接口, 以及进一步通过该 CPU 以提供高带宽专用内部存储器存取至该第二内部存储器。

一般目的使用的内部处理单元的存储器

技术领域

[0001] 本发明一般涉及计算装置（例如，计算机、嵌入装置、手持装置以及类似装置）。更特别地，本发明涉及通过此计算装置的处理单元使用的存储器。

背景技术

[0002] 一种计算装置通常包含一个或多个处理单元，例如中央处理单元（central-processing unit；简称 CPU）与图形处理单元（graphics-processing unit；简称 GPU）。CPU 通过遵循指令的精确集而协调计算装置的活动。。该 GPU 通过执行数据并行计算工作，例如图形处理工作和 / 或可能被终端用户应用（如视频游戏应用）所需的物理仿真，而协助 CPU。该 GPU 与该 CPU 可为分离装置和 / 或封装的部分或可以包含在相同装置和 / 或封装中。进一步，每个处理单元可以包含在另一较大的装置中。例如，GPU 经常整合至路由或桥接装置，像是，例如北桥装置。

[0003] 在该终端用户应用与 GPU 之间有几种软件层。该终端用户应用与应用程序接口（application-programming interface；简称 API）通信。该 API 允许该终端用户应用以标准化形式输出图形数据与命令，而非依于 GPU 的形式。许多类型的 API 是商业上可用的，包含华盛顿，雷蒙德的微软公司开发的 DirectX；科纳斯组织维护的 OpenGL 与 OpenCL。该 API 与驱动器（driver）通信。驱动器翻译从 API 所接收的标准码成为 GPU 所了解的指令的本地形式。驱动器通常由 GPU 的制造商所写。接着，GPU 执行从驱动器的指令。

[0004] 在传统的系统中，每个 CPU 与 GPU 通常耦接（couple）至外部存储器。该外部存储器可以包含由 CPU 和 / 或 GPU 执行的指令和 / 或使用的数据。该外部存储器可以为，例如动态随机存取存储器（dynamic random-access memory；简称 DRAM）。该外部存储器可以配置为相当大，从而提供充足的存储容量至其耦接的每个处理单元。不幸地，存取该外部存储器可以花费几百个时钟周期。因此，外部存储器可无法提供存储器足够的带宽或对于高端 GPU 的快速存储器存取。

[0005] 一种用于提供足够的存储器带宽至 GPU 的潜在的解决方案为提供 GPU 内部存储器。该内部存储器可以为，例如嵌入或堆栈 DRAM。相较于该外部存储器，内部存储器提供较高带宽，较快存储器存取，以及消耗较少的功率。然而，该内部存储器的容量不可轻易缩放以达到在高端 GPU 的存储需求。例如，相较可包含在 GPU 的内部存储器，高端 GPU 可能需要较多的存储器。

[0006] 根据前述内容，需要一种提供足够的存储器容量（如外部存储器）与高带宽（如嵌入存储器）的存储器与其应用。

发明内容

[0007] 本发明的实施例通过提供一般目的使用的内部处理单元的存储器与其应用。本发明的实施例的内部处理单元的存储器提供高带宽，因为其嵌入处理单元。其也提供足够的存储容量，因为多个处理单元的存储器可以结合至足够大的存储器池（memory pool）。

[0008] 举例来说,本发明的实施例提供 GPU。该 GPU 包含第一内部存储器、耦接至该第一内部存储器的执行单元以及配置以耦接该第一内部存储器至另一处理单元的第二内部存储器的接口。在一实施例中,该 GPU 体现在软件中。在另一实施例中,该 GPU 包含在系统中。该系统可以包括,例如,超级计算机、桌面计算机、膝上计算机、视频游戏控制台、嵌入装置、手持装置(例如,移动手机、智能型手机、MP3 播放器、相机、GPS 装置、或类似装置),或者包含或配置以包含 GPU 的另一系统。

[0009] 本发明的进一步的特征与优点,以及本发明的各种实施例的结构与操作是伴随着附图而详细描述于下。要注意的是,本发明并非限制于在此描述的特定实施例。在此呈现的实施例仅用于说明的目的。对于相关领域的技术人员而言,基于在此教示的内容,附加的实施例将是明显的。

附图说明

[0010] 在此并入与形成说明书的部分的附图说明本发明,且连同描述,进一步作为解释本发明的原则与使相关领域的技术人员制造与使用本发明。

[0011] 图 1A 与图 1B 说明依据本发明的实施例的包含一般目的使用的内部处理单元的存储器的示范系统;

[0012] 图 2 说明依据本发明的实施例的一般目的使用的具有内部存储器的示范 GPU 的细节;

[0013] 图 3 说明依据本发明的实施例的可包含在处理组件中的示范堆栈存储器;以及

[0014] 图 4 说明依据本发明的实施例的由图 2 的 GPU 实作的示范方法。

[0015] 从以下提出的详细描述并结合图式,本发明的特征与优点将变得更加明显,其中,在全文中,相似的组件字元识别为对应的组件。在图式中,相似的组件符号一般表示相同、功能相似、及/或结构相似的组件。在图式中,第一次出现的组件通过在对应的组件符号中的最左位而表示。

[0016] 主要组件符号说明

[0017]	100	计算系统
[0018]	101、250	第一接口
[0019]	102	CPU
[0020]	103、260	第二接口
[0021]	104	系统存储器
[0022]	105	第三接口
[0023]	106、106A、106B	内部存储器
[0024]	110A	第一 GPU
[0025]	110B	第二 GPU
[0026]	114	总线
[0027]	116	I/O 接口
[0028]	118	外部装置
[0029]	120	次要存储器
[0030]	130	缓存存储器

[0031]	132、142A、142B、142	执行单元
[0032]	140	显示装置
[0033]	164	共享接口
[0034]	202	输入逻辑
[0035]	204	着色核心
[0036]	206	输出逻辑
[0037]	210	存储器控制器
[0038]	212	存储器细胞
[0039]	400	方法
[0040]	402 ~ 414	步骤。

具体实施方式

[0041] I. 概述

[0042] 本发明提供一种一般目的使用的内部 GPU 存储器与其应用。在以下的详细描述中,参照「一个实施例」、「一实施例」、「示范实施例」等指出该描述的实施例可以包含特别的特征、结构、或特性,但在每个实施例可不必要包含该特别的特征、结构、或特性。此外,此词组不必要参照相同实施例。另外,当描述的特别的特征、结构、或特性连接实施例时,无论是否明确描述,上述内容在本领域的技术人员知识中以影响此特征、结构、或特性连接其它实施例。

[0043] 按照一实施例, GPU 包含内部存储器 (例如, 嵌入或堆栈 DRAM), 其配置以被一个或多个其它处理单元所用。该 GPU 包含接口与实作协议, 允许一个或多个其它 GPU 存取其内部存储器。接口可提供每个其它 GPU 专用存取该内部存储器或可提供该其它 GPU 共享存取该内部存储器。存取该 GPU 的内部存储器可通过 GPU 本身或各个其它 GPU 所控制。

[0044] 在一实施例中, 接口和协议允许该内部存储器与外部存储器结合, 形成可存取 GPU 的较大存储器池。该外部存储器可包含在其它 GPU 中。在一实施例中, 例如, 计算装置包含多个 GPU, 其中, 每个 GPU 包含内部存储器, 配置以与其它 GPU 共享。在此实施例中, 每个 GPU 的内部存储器结合至统一的存储器池。以参与 GPU 的数目缩放存储器池的尺寸。任何参与 GPU 可为了其存储需求使用该存储器池。

[0045] 以下描述依据本发明的实施例的示范 GPU 的进一步细节。然而, 在提供这些细节之前, 其有助于描述示范计算装置, 其中, 可实作此 GPU。

[0046] II. 示范计算系统

[0047] 图 1A 与图 1B 说明依据本发明的实施例的具有多个 GPU 的示范计算系统 100, 各 GPU 包含配置以一般目的使用的内部存储器。相较外部存储器, 该内部存储器提供各 GPU 较高带宽存取数据。此外, 各 GPU 的该内部存储器可通过各 GPU 结合至可存取的较大的存储器池, 藉此提供足够的存储容量至各 GPU。

[0048] 于图 1A 的实施例中, 各 GPU 给予专用存取 (dedicated access) 另一 GPU 的该内部存储器。在图 1B 的实施例中, 各 GPU 有共享存取经由共享接口至该其它 GPU 的内部存储器。在一实施例中, 计算系统 100 可包括超级计算机、桌面计算机、膝上计算机、视频游戏控制台、嵌入装置、手持装置 (例如, 移动手机、智能型手机、MP3 播放器、相机、GPS 装置、或类

似装置),或者包含或配置以包含 CPU 和 / 或 GPU 的一些其它系统。。

[0049] 请参照图 1A 与图 1B, 计算装置 100 包含 CPU 102、第一 GPU110A, 以及第二 GPU 110B。该 CPU 102 执行指令以控制该计算装置 100 的功能。该 GPU 110 通过执行数据并行处理工作 (像是, 例如, 图形处理工作和 / 或一般计算工作) 而协助 CPU 102。基于其设计, 该 GPU 110 通常可执行数据并行处理工作, 较快于 CPU 102 在软件上执行。

[0050] 该第一 GPU 110A 与该第二 GPU 110B 各包含其所有的内部存储器与执行单元。特别是, 第一 GPU 106A 包含内部存储器 106A 与执行单元 142A; 以及第二 GPU 106B 包含内部存储器 106B 与执行单元 142B。同样地, 该 CPU 102 包含缓存存储器 130 与执行单元 132。相较于可能将数据外部存储 (例如, 数据存储在系统存储器 104), 该内部存储器 106 (以及可选缓存存储器 130), 可利用于 GPU 110, 用以提供较快速存取与较高带宽至某些数据。该内部存储器 106 可包括, 例如嵌入或堆栈 DRAM。

[0051] 该内部存储器 106A、106B (与可选缓存存储器 130) 可结合至较大的存储器池, 用以提供实质上的存储容量 (例如超过 4GB), 同时也提供快速, 高带宽存储器的存取。虽然习知外部存储器可提供足够的存储容量 (例如超过 4GB), 对于某些高端的使用, 习知外部存储器提供不足的带宽。同样地, 对于这些高端的使用, 虽然习知嵌入存储器可提供足够的带宽, 对于这些高端的使用, 习知嵌入存储器提供不足的存储容量 (例如少于 4GB)。不像习知外部存储器和 / 或习知嵌入存储器, 本发明的实施例不仅提供足够的存储容量 (例如超过 4GB), 也通过提供包含内部存储器的 GPU 提供高带宽, 所述内部存储器可利用于一般目的使用的其它 GPU。

[0052] 举例来说, 高端 GPU 的帧缓冲 (frame buffer) (也就是存储在显示装置上显示完整数据帧的缓冲) 可需要高带宽存取实质上大的存储器 (例如超过 4 十亿字节 (gigabyte; 简称 GB))。在实施例中, 该第一 GPU 110A 可使用内部存储器 106A、B 与可选 CPU 102 的缓存存储器 130, 用以定义第一 GPU 110A 的帧缓冲。同样地, 该第二 GPU 110B 也可使用内部存储器 106A、B 与可选 CPU 102 的缓存存储器 130, 用以定义第二 GPU 110B 的帧缓冲。以此方法, 不同于习知外部或嵌入存储器, 依据本发明的实施例所定义的帧缓冲提供高带宽存取至实质上大的存储器 (例如超过 4GB)。

[0053] 于图 1A 的实施例中, 如上所暗示, 各 GPU 110 给予专用存取另一处理单元的内部存储器 106。特别是, 第一接口 101 提供第一 GPU110A 专用存取第二 GPU 110B 的内部存储器 106B 以及提供第二 GPU110B 专用存取第一 GPU 110A 的内部存储器 106A。基于数据的地址范围, 数据可写入至内部存储器 106A 或内部存储器 106B, 或者数据可从内部存储器 106A 或内部存储器 106B 撷取 (retrieve)。例如, 该内部存储器 106A 可以分配第一地址范围 (例如, 少于第一预定地址 A 与大于或等于第二预定地址 B), 以及该内部存储器 106B 可以分配第二地址范围 (例如, 不是在第一地址范围内的所有地址)。然而, 应领会, 写入数据至内部存储器 106A 和 / 或内部存储器 106B, 或者从内部存储器 106A 和 / 或内部存储器 106B 撷取数据的其它方案可在不偏离本发明的精神与范围下实作, 提供该第一 GPU 110A 与该第二 GPU 110B 可各具有存取第一 GPU 110A 的内部存储器 106A 以及第二 GPU 110B 的内部存储器 106B。

[0054] 在一实施例中, 第一接口 101 包含显示控制器接口。该显示控制器接口提供显示装置 140 以存取 GPU 的帧缓冲。通过并入该显示控制器接口至第一接口 101, 该第一接口

101 可提供在已经包含在习知 GPU 设计标准针 (standard pin) 上。

[0055] 除了第一接口 101, 第二接口 103 提供 CPU 102 专用存取至该第二 GPU 110B 的内部存储器 106B 与提供第二 GPU 110B 专用存取至该 CPU102 的缓存存储器 130。以此方法, 该第二 GPU 110B 与该 CPU 102 可各具有存取至该第二 GPU 110B 的内部存储器 106B 与该 CPU 102 的缓存存储器 130。同样地, 第三接口 105 提供第一 GPU 110A 专用存取至该 CPU 102 的缓存存储器 130 与提供该 CPU102 专用存取至该第一 GPU 110A 的内部存储器 106A。以此方法, 该第一 GPU 110A 与该 CPU102 可各具有存取该第一 GPU110A 的内部存储器 106A 与该 CPU 102 的缓存存储器 130。

[0056] 在图 1B 的实施例中, 各处理单元具有共享存取经由共享接口 164 至其它处理单元的内部存储器。该共享接口 164 提供各处理单元 (例如, 该第一 GPU 110A、该第二 GPU 110B、以及该 CPU102) 高带宽存取至其它处理单元的内部存储器。基于数据的地址范围, 数据可写入至内部存储器 106A、内部存储器 106B 或缓存存储器 130, 或者数据可从内部存储器 106A、内部存储器 106B 或缓存存储器 130 撷取。例如, 该内部存储器 106A 可以分配第一地址范围; 该内部存储器 106B 可以分配第二地址范围; 以及该缓存存储器 130 可以分配第三地址范围。然而, 应领会, 写入数据至内部存储器 106A、内部存储器 106B 和 / 或缓存存储器 130, 或者从内部存储器 106A、内部存储器 106B 和 / 或缓存存储器 130 撷取数据的其它方案可在不偏离本发明的精神与范围下实作, 提供该第一 GPU 110A、该第二 GPU 110B、以及该 CPU 102 可各具有存取第一 GPU 110A 的内部存储器 106A、第二 GPU 110B 的内部存储器 106B、以及 CPU 102 的缓存存储器 130。

[0057] 在实施例中, 该计算装置 100 也包括系统存储器 104、次要存储器 120、输入 - 输出 (I/O) 接口 116、和 / 或显示装置 140。该系统存储器 104 存储经常由 CPU 102 上运行的程序存取的信息。该系统存储器 104 通常包含挥发存储器, 意指当该计算装置 100 的电源关闭时, 失去存储在该系统存储器 104 中的数据。该次要存储器 120 存储由计算装置 100 使用的数据和 / 或应用。相较于系统存储器 104, 该次要存储器 120 通常具有相当大的存储容量及通常包括非挥发 (持续) 的存储器, 意指即使当该计算装置 100 的电源关闭时, 在该次要存储器 120 存储的数据持续。I/O 接口 116 允许该计算装置 100 耦接至外部装置 116 (例如, 外部显示装置、外部存储装置 (例如, 视频游戏盒式磁带 (cartridge)、CD、DVD、闪存驱动器、或类似物)、网络卡、或一些其它形式的外部装置)。该显示装置 140 显示该计算装置 100 的内容。显示装置可包括阴极射线管、液晶显示 (liquid-crystal display; 简称 LCD)、等离子屏幕、或一些不论目前已知或后来发展的其它形式的显示装置。

[0058] GPU 110 与 CPU 102 彼此通信且与系统存储器 104、次要存储器 120、及 I/O 接口 116 在总线 (bus) 114 上通信。该总线 114 可以是在计算装置中使用的任何种类的总线, 包含周围组件接口 (peripheral component interface; 简称 PCI) 的总线、加速图形端口 (accelerated graphics port; 简称 AGP) 的总线, 高速 PCI (PCI Express; 简称 PCIE) 的总线、或另一不论目前可利用或未来发展的其它形式的总线。

[0059] 在实施例中, 该计算装置 100 可包括代替 GPU 110 或除了 GPU 110 以外的视频处理单元 (video processing unit; 简称 VPU)。例如, 在一实施例中, 该计算装置 100 包括该 GPU 110A、该 CPU 102; 及代替说明在图 1A 与图 1B 中的 GPU 110B, 该计算装置 100 包括 VPU。以此方法, 该 CPU 102 可执行一般处理功能、该 GPU 110A 可执行图形处理功能、以及

VPU 可执行视频处理功能。

[0060] III. 示范 GPU

[0061] 图 2 说明具有内部存储器 106 的 GPU 110 的示范细节。依据本发明的实施例,该内部存储器 106 可被另一 GPU、或 CPU 使用,以通过基于增大存储器覆盖区 (footprint) 尺寸结合图形处理功率而增加整体系统的性能。

[0062] 如前述,该 GPU 110 包括执行单元 142 与该内部存储器 106。参照图 2,该执行单元 142 包括输入逻辑 202、着色核心 (shader core) 204、与输出逻辑 206。该内部存储器 106 包括存储器控制器 210 与存储器细胞 (cell) 212。该存储器控制器 210 控制存取至该存储器细胞 212。该存储器细胞 212 存储数据。

[0063] 在一实施例中,内部存储器 106 包含嵌入动态随机存取存储器 (DRAM)。嵌入 DRAM 是封装在具有处理单元的一般封装件的存储器。在另一实施例中,在图 3 中说明该内部存储器 106 包含堆栈 DRAM。堆栈存储器包括多个存储器组件在三维结构中彼此堆栈于顶部。

[0064] 该内部存储器 106 经由该输入逻辑 202 与该输出逻辑 206 两者耦接至该执行单元 142。特别是,该输入逻辑 202 可从该内部存储器 106 撷取数据,及该输出逻辑 206 可传送数据至该内部存储器 106 以存储在该存储器细胞 212 中。

[0065] 该内部存储器 106 也可经由第一接口 250 耦接至另一 GPU 的内部存储器。耦接至另一 GPU 的内部存储器的内部存储器 106 可增加可利用于执行单元 142 的整体存储器池。在一实施例中,如图 1A 的接口 101 所说明,该第一接口 250 提供该 GPU 110 的内部存储器 106 与另一 GPU 的内部存储器之间的专用存取。在此实施例中,该第一接口 250 提供在传统 GPU 的标准针上。例如,该第一接口 250 可包含显示 - 控制器接口,其提供显示装置存取至包含在内部存储器 106 中的本地帧缓冲。在另一实施例中,如图 1B 的接口 164 所说明,该第一接口 250 提供该 GPU 110 的内部存储器 106 与其它处理单元的内部存储器之间的共享存取。

[0066] 该内部存储器 106 也可经由第二接口 260 耦接至该 CPU 102 的缓存存储器 130。因此,该内部存储器 106 与该缓存存储器 130 的结合可增加可利用于该 GPU 110 的存储器池。在一实施例中,该第二接口 260 提供在该 GPU 110 的内部存储器 106 与该 CPU 102 的缓存存储器 130 之间的专用连接 (dedicated connection),例如在图 1 的连接 103 或连接 105。在另一实施例中,该第二接口 260 提供仅由该 GPU 110 与该 CPU 102 共享的连接,例如在图 1B 的连接 164。在进一步的实施例中,第二接口在一般总线上耦接 GPU 110 至 CPU 102,例如图 1A 与图 1B 的总线 114。

[0067] IV. GPU 110 的示范操作

[0068] 图 4 说明依据本发明的实施例的由 GPU 110 实作的示范方法 400。以下参考图 3 与图 4 描述方法 400。

[0069] 方法 400 开始于步骤 402,其中,接收指令。在一实施例中,该输入逻辑 202 接收由该 GPU 110 执行的指令。该指令可包括,例如,由系统 100 的 CPU 102 上运行的终端用户应用所提供的图形处理工作或数据并行处理工作。

[0070] 在步骤 404 中,确认相关于指令的数据的位置。在一范例中,数据可以包含在接收的指令内。此数据通常简称实时数据 (immediate data)。在另一范例中,该指令提供数据的位置。例如,该指令可以包含数据存储的地址。在进一步的范例中,例如,该指令包含信

息,从该信息,该输入逻辑 202 计算数据存储的地址。该数据可以存储在该内部存储器 106、该内部存储器 106 耦接的另一 GPU 的内部存储器、或该 CPU 102 的缓存存储器 130。

[0071] 在步骤 406 中,提取数据。若数据为该实时数据,该输入逻辑 202 仅从指令提取(extract)该实时数据。若数据存储在该内部存储器 106 或该内部存储器 106 耦接的存储器中,该输入逻辑 202 传送请求至存储器控制器 210 用于存取数据。一方面,若数据存储在该存储器细胞 212 中,提取该数据且提供至该输入逻辑 202。另一方面,若数据存储在该内部存储器 106 耦接的另一存储器中,自输入逻辑 202 的请求经由该接口 250 或该接口 260 转交(forward)至其它存储器。然后,该数据从其它存储器提取且提供至该输入逻辑 202。

[0072] 在步骤 408 中,执行指令。基于在步骤 406 中由输入逻辑 202 获得的数据,该着色核心 204 执行指令。

[0073] 在步骤 410 中,提供执行指令的结果至该输出逻辑 206。该输出逻辑 206 根据这些结果判断是否需要进一步的处理,如决定步骤 412 所示。提供至该输出逻辑 206 的结果可具有旗标(flag)或一些其它记号以指出是否需要附加的处理。若在决定步骤 412 中,输出逻辑 206 判断需要进一步的处理,然后输出逻辑 206 转交该结果回着色核心 204,且重复方法 400 的步骤 408 与 410。另一方面,若在决定步骤 412 中,输出逻辑 206 判断不需要进一步的处理,然后输出逻辑 206 提供该结果至内部存储器 106,如步骤 414 所示。

[0074] 依据写入结果的地址,然后该结果可写入至内部存储器 106 或耦接至内部存储器 106 的存储器。若结果为写入至内部存储器 106,该存储器控制器 210 提供存取至在该存储器细胞 212 中适当的地址且存储该结果。另一方面,若结果为写入至耦接至内部存储器 106 的存储器,然后该存储器控制器 210 经由接口 250 或接口 260 转交结果至其它存储器且存储该结果在其它存储器的存储器细胞中。

[0075] V. 示范软件实作

[0076] 除了该 GPU 110 的硬件实作,此 GPU 110 也可体现在置于例如在计算机可读的媒体(computer-readable medium)的软件中,该计算机可读的媒体配置以存储软件(例如,计算机可读的程序码)。计算机可读的程序码允许本发明的实施例,包含下述实施例:(i)在此揭露的系统与技术的功能(例如,提供该 GPU 110 的工作,调度该 GPU 110 的工作,执行该 GPU 110 的工作,或者类似工作);(ii)在此揭露的系统与技术的制造(例如,该 GPU 110 的制造);或(iii)在此揭露的系统与技术的功能与制造的结合。

[0077] 该软件可,例如,透过一般编程语言(如 C 或 C++)、硬件描述语言(hardware-description language;简称 HDL),其包含 Verilog HDL、VHDL、Altera HDL(AHDL)等等,或其它可利用的编程和/或示意捕捉工具(schematic-capture tool)(譬如电路捕捉工具)的使用而完成。该计算机可读的程序码可置于任何已知的计算机可读的媒体,包含半导体,磁盘,或光盘(例如,CD-ROM、DVD-ROM)。因此,该计算机可读的程序码可在通讯网络(包含互连网与网际网络)上传播。应了解由上述系统与技术的功能与/或所提供的结构可由核心(如着色核心)所代表,该核心体现在计算机可读的程序码,且可转换成硬件作为集成电路产品的部分。

[0078] VI. 结论

[0079] 以上描述一般目的使用的内部 GPU 存储器与其应用。应领会,具体实施方式,而非发明内容与摘要,意图用于解释权利要求书。发明内容与摘要可提出由发明人所思及的本

发明的一个或多个但非所有的实施例,且因此,并非以任何方式意图限制本发明与附加的权利要求书。

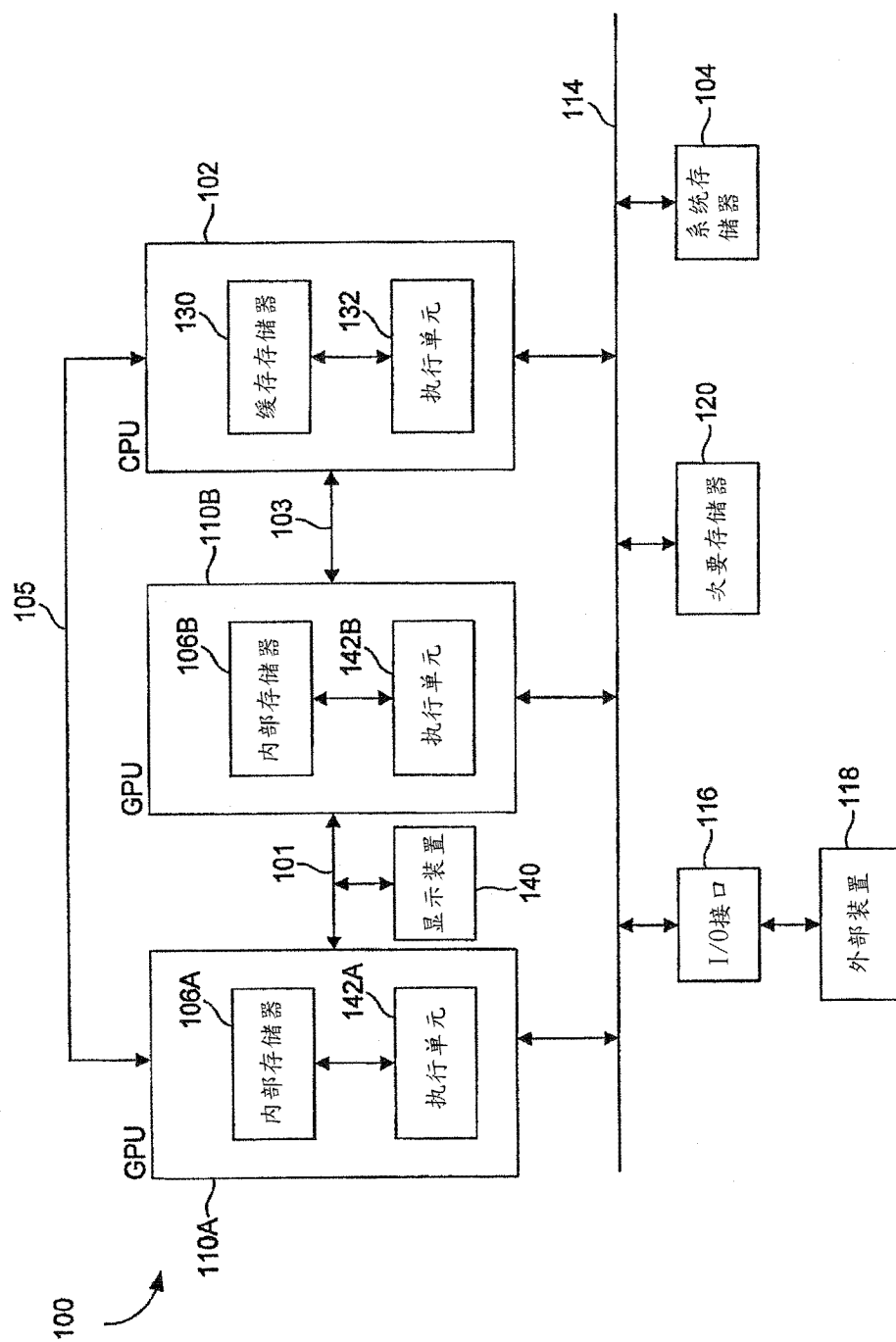


图 1A

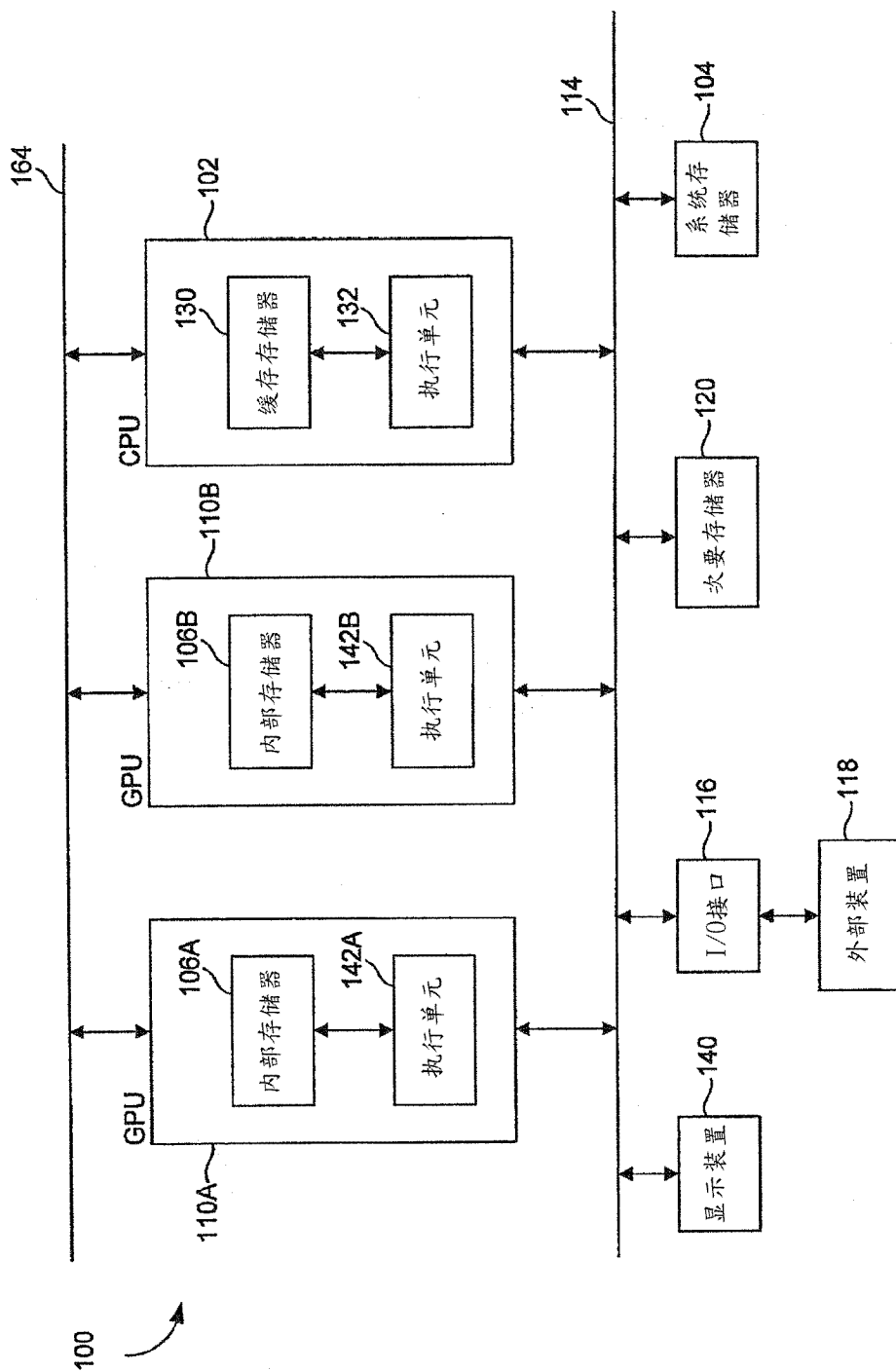


图 1B

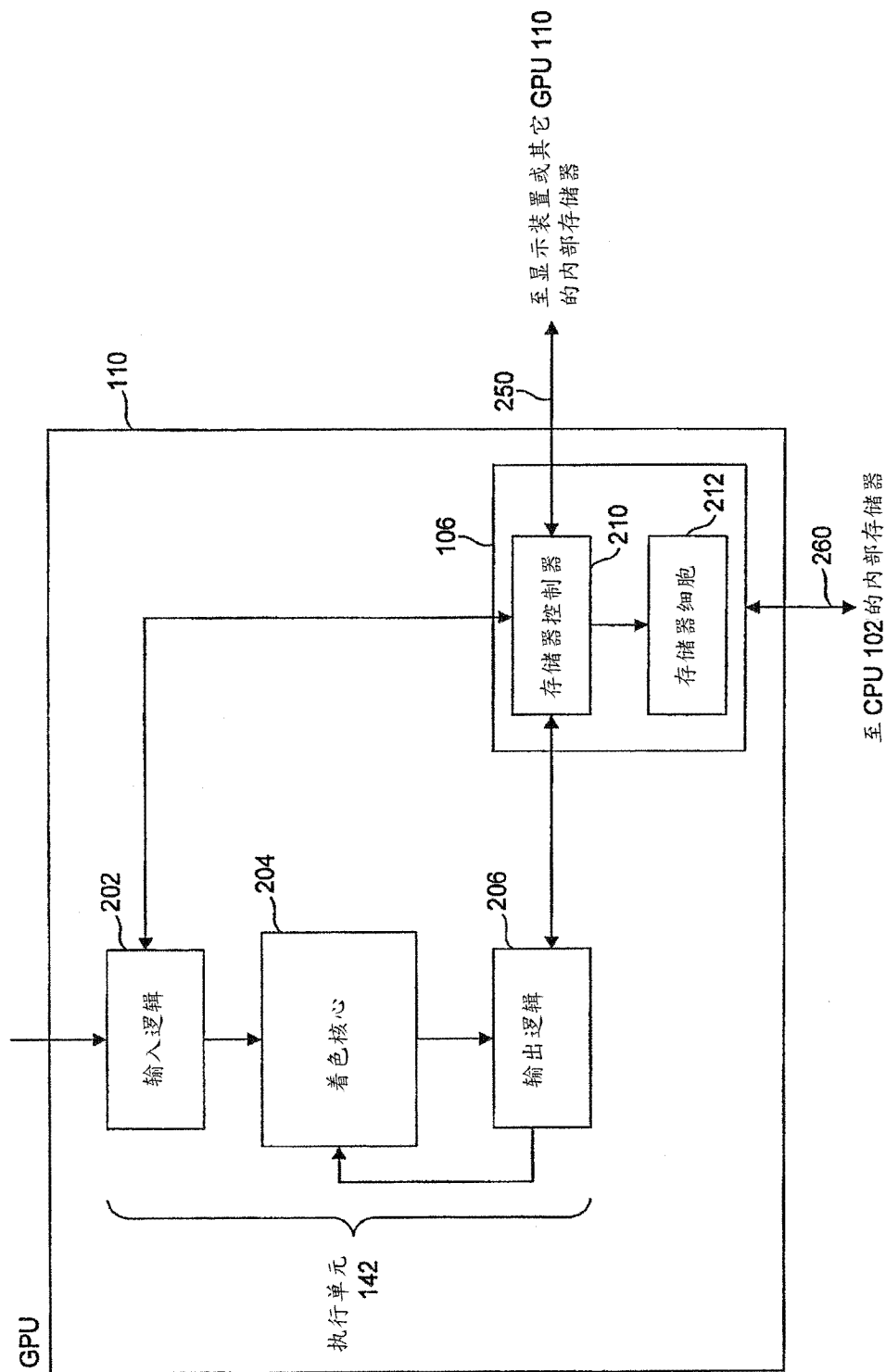


图 2

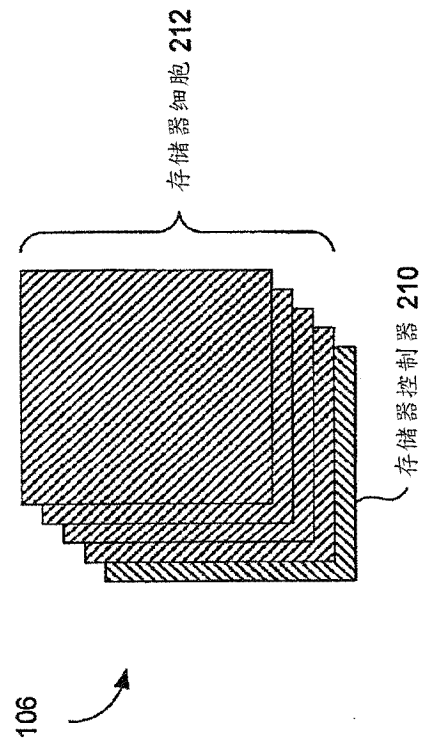


图 3

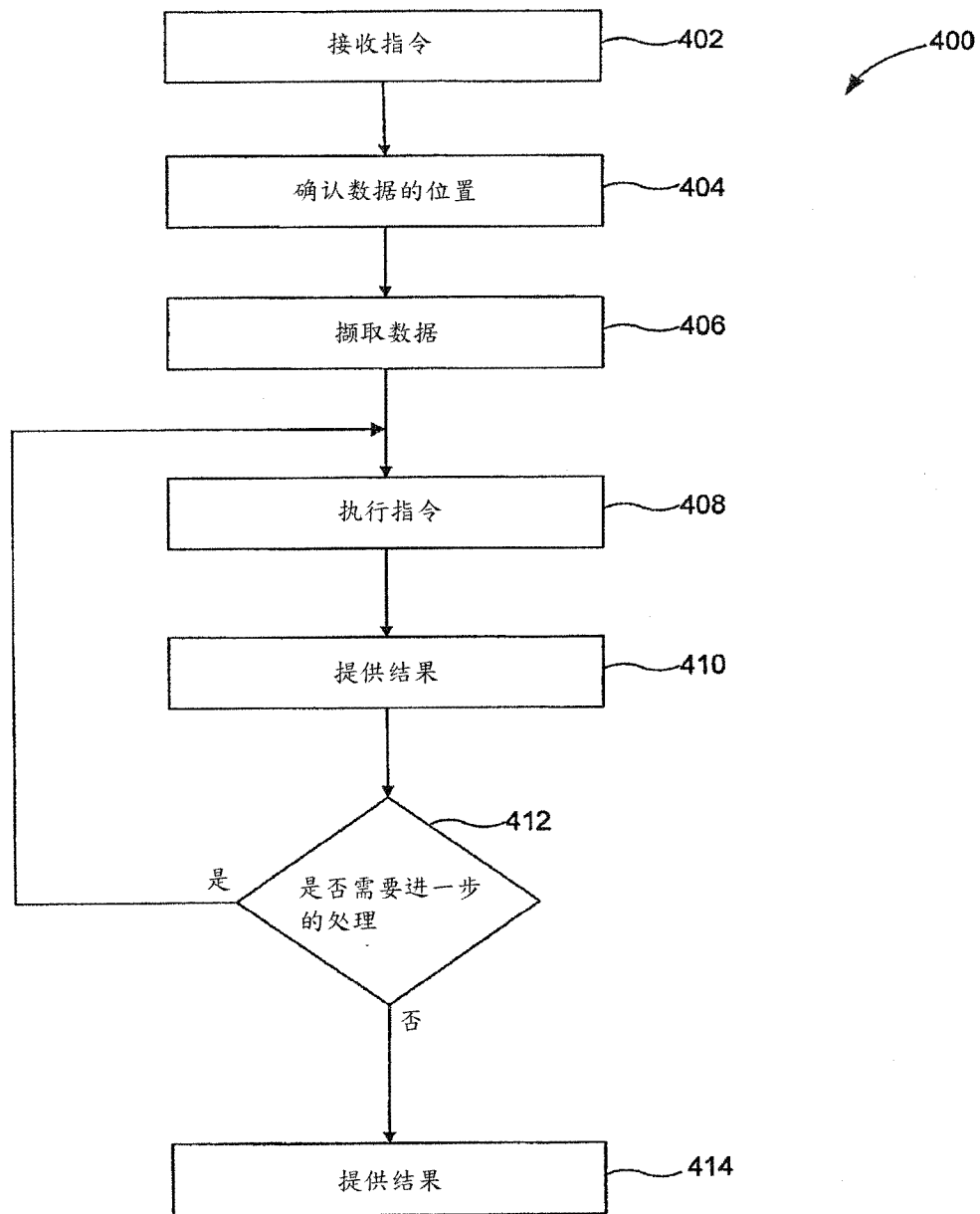


图 4