

任务分配方法、装置、电子设备及计算机可读存储介质

技术领域

5 本公开涉及处理器领域，尤其涉及一种任务分配方法、装置、电子设备及计算机可读存储介质。

背景技术

随着科学技术的发展，人类社会正在快速进入智能时代。智能时代的重要特点，就是人们获得数据的种类越来越多，获得数据的量越来越大，而对处理数据的速度要求越来越高。芯片是任务分配的基石，它从根本上决定了人们处理数据的能力。从应用领域来看，芯片主要有两条路线：一条是通用芯片路线，例如 CPU 等，它们能提供极大的灵活性，但是在处理特定领域算法时有效算力比较低；另一条是专用芯片路线，例如 TPU 等，它们在某些特定领域，能发挥较高的有效算力，但是面对灵活多变的比较通用的领域，它们处理能力比较差甚至无法处理。由于智能时代的数据种类繁多且数量巨大，所以要求芯片既具有极高的灵活性，能处理不同领域且日新月异的算法，又具有极强的处理能力，能快速处理极大的且急剧增长的数据量。

在神经网络计算中，经常会用到多核或者众核的芯片。如何让众多的处理核能够高效率的发挥算力，是决定整个芯片性能的关键。各处理核的算力发挥，取决于多种因素，例如任务的调度与分配、芯片的架构、处理核的结构、处理核的电路等。其中任务的调度与分配是一个非常关键的因素，如果任务的调度与分配合理，则能充分发挥各处理核的有效算力。

现有技术中一般使用以下两种方案来进行处理核任务的调度：

如图 1a 所示为使用调度器进行任务分配的方案。在该方案中，调度器接收到来自于指令源的指令，然后将指令按一定的策略，如顺序，传输给各处理核，各处理核执行相同的指令，但是处理不同的数据。各处理核可以是比较简单的结构，例如共用控制电路和寄存器的 SIMD(Single Instruction Multiple Data，单指令多数据结构)，也可以是比较复杂有一定的自主性的结构，例如有独立的控制电路和寄存器等的 SIMT(Single Instruction Multiple Threads，单指令多线程)。但是该方案中，调度器只能被动的从外部指令源接收指令，再分配给各处理核，功能单一，且缺乏灵活性；而无论是 SIMD 结构还是 SIMT 结构，各处理核只能执行相同的指令，各处理核的自由度小，控制简单。

- 如图 1b 所示为处理核独立执行任务的方案。在该方案中，各处理核能够从存储器中独立读取指令，进行运算；各处理核具有完整的控制电路、寄存器组等电路。而对于这种方案，各处理核难以互相配合高效的完成一个完整的任务；电路控制复杂，每一个核都几乎是一个完整的 CPU，设计难度大，且功耗和面积大；多处理核可能频繁访问指令存储区，引起存储访问效率的下降，进而影响芯片算力的发挥。

发明内容

- 提供该发明内容部分以便以简要的形式介绍构思，这些构思将在后面的具体实施方式部分被详细描述。该发明内容部分并不旨在标识要求保护的技术方案的关键特征或必要特征，也不旨在用于限制所要求的保护的技术方案的范围。

为了解决现有技术中处理核的任务分配不灵活、控制复杂的技术问题，本公开实施例提出如下技术方案：

第一方面，本公开实施例提供一种任务分配方法，包括：

- 响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；

响应于需要给所述处理核分配任务段，所述微处理器指示分配任务段至所述处理核；

所述处理核接收所述任务段；

响应于满足第一预设条件，所述处理核发送同步请求信号，其中所述同步请求信号用于触发生成所述同步信号。

- 进一步的，所述响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，包括：

所述微处理执行自身程序；

响应于接收到同步信号，所述微处理器根据接收到的所述同步信号的次数确定需要分配任务段的处理核。

- 进一步的，所述第一预设条件包括：所述处理核接收完所述任务段或者所述处理核执行接收所述任务段的时间达到预设的时间阈值。

进一步的，所述微处理器指示分配任务段至所述处理核，包括：

所述微处理器指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理

核。

进一步的,所述微处理器指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理核,包括:

所述微处理获取所述处理核的存储空间的大小;

- 5 所述微处理器指示所述存储控制器将与所述处理核的存储空间大小相匹配的任务段从所述存储器中读出并发送至所述处理核。

进一步的,所述方法还包括:

响应于不需要给所述处理核分配任务段,所述微处理器执行自身程序;

所述处理核执行前一次接收到的任务段;

- 10 响应于满足第二预设条件,所述处理核发送所述同步请求信号。

进一步的,所述第二预设条件为:所述处理核执行到所述前一次接收到的任务段中的同步请求信号发送指令、或所述处理核执行前一次接收到的所述任务段的时间达到预设的时间阈值。

- 15 进一步的,所述芯片包括至少两个执行所述任务的处理核,所述微处理器分别针对所述至少两个处理核指示分配所述任务段。

进一步的,所述同步请求信号包括所述至少两个执行所述任务的处理核所生成的所述同步请求信号。

进一步的,所述处理核发送同步请求信号,其中所述同步请求信号用于触发生成所述同步信号,包括:

- 20 响应于每个所述执行所述任务的处理核都发送了所述同步请求信号,触发生成所述同步信号。

进一步的,所述微处理器包括计数器,在接收到所述同步信号之后,还包括:

启动所述计数器以记录所述微处理器接收到的所述同步信号的次数。

- 25 进一步的,所述同步信号由同步信号发生单元生成,其中所述同步信号发生单元由硬件或者软件实现。

第二方面,本公开实施例提供一种芯片,包括:

处理核,所述处理核用于执行任务;

微处理器，用于响应于接收到同步信号，根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；响应于需要给所述处理核分配任务段，指示分配任务段至所述处理核；

5 存储控制器，用于基于所述微处理器的指示，从外部存储空间取出与所述指示对应的任务段，并将所述任务段发送至所述处理核。

进一步的，所述处理核还用于：接收所述任务段；响应于满足第一预设条件，发送同步请求信号，其中所述同步请求信号用于触发生成所述同步信号。

进一步的，所述微处理器还用于：执行自身程序；响应于接收到同步信号，根据接收到的所述同步信号的次数确定需要分配任务段的处理核。

10 进一步的，所述第一预设条件包括：所述处理核接收完所述任务段或者所述处理核执行接收所述任务段的时间达到预设的时间阈值。

进一步的，所述微处理器还用于：指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理核。

15 进一步的，所述微处理器还用于：获取所述处理核的存储空间的大小；指示所述存储控制器将与所述处理核的存储空间大小相匹配的任务段从所述存储器中读出并发送至所述处理核。

进一步的，所述微处理器还用于：响应于不需要给所述处理核分配任务段，执行自身程序；所述处理核还用于：执行前一次接收到的任务段；响应于满足第二预设条件，发送所述同步请求信号。

20 进一步的，所述第二预设条件为：所述处理核执行到所述前一次接收到的任务段中的同步请求信号发送指令、或所述处理核执行所述前一次接收到的任务段的时间达到预设的时间阈值。

进一步的，所述芯片包括至少两个执行所述任务的处理核，所述微处理器用于分别针对所述至少两个处理核指示分配所述任务段。

25 进一步的，所述同步请求信号包括所述至少两个执行所述任务的处理核所生成的所述同步请求信号。

进一步的，所述微处理器还包括计数器，在接收到所述同步信号之后，所述微处理器还用于：启动所述计数器以记录所述微处理器接收到的所述同步信号的次数。

30 第三方面，本公开实施例提供一种电子设备，包括：存储器，用于存储计算机可读指令；以及一个或多个处理器，用于运行所述计算机可读指令，使得所述处理器运行时实现

前述第一方面中的任一所述任务分配方法。

第四方面，本公开实施例提供一种非暂态计算机可读存储介质，其特征在于，该非暂态计算机可读存储介质存储计算机指令，该计算机指令用于使计算机执行前述第一方面中的任一所述任务分配方法。

5 第五方面，本公开实施例提供一种计算机程序产品，其特征在于：包括计算机指令，当所述计算机指令被计算设备执行时，所述计算设备可以执行前述第一方面中的任一所述任务分配方法。

第六方面，本公开实施例提供一种计算装置，其特征在于，包括一个或多个所述第二方面所述的芯片。

10 本公开实施例公开了一种任务分配方法、装置、电子设备和计算机可读存储介质。其中该任务分配方法包括：响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；响应于需要给所述处理核分配任务段，所述微处理器指示分配任务段至所述处理核；所述处理核接收所述任务段；响应于满足第一预设条件，所述处理核发送同步请求信号，其中所述同步请求信
15 号用于触发生成所述同步信号。通过上述方法，芯片中的微处理器根据同步信号来分配处理核所执行的任务段，解决了现有技术中的处理核的任务分配不灵活、控制复杂的技术问题技术问题。

上述说明仅是本公开技术方案的概述，为了能更清楚了解本公开的技术手段，而可依照说明书的内容予以实施，并且为了让本公开的上述和其他目的、特征和优点能够更明显易懂，以下特举较佳实施例，并配合附图，详细说明如下。

附图说明

结合附图并参考以下具体实施方式，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解
25 附图是示意性的，原件和元素不一定按照比例绘制。

图 1a-1b 为本公开现有技术的示意图；

图 2 为本公开提供的任务分配方法的实施例的流程示意图；

图 3 为本公开提供的芯片的结构示意图；

图 4 为本公开提供的芯片的一个实例的结构示意图；

图 5 为本公开提供的芯片实例所执行的神经网络计算任务的示意图；

图 6 为本公开提供的芯片实例执行任务分配方法时处理核流水计算过程示意图；

图 7 为本公开提供的芯片实例执行任务分配方法时微处理器和处理核的时序图。

5 具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

- 10 应当理解，本公开的方法实施方式中记载的各个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

- 15 本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

- 20 需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

- 25 图 2 为本公开实施例提供的任务分配方法实施例的流程图，本实施例提供的该任务分配方法用于芯片中，其中所述芯片包括处理核和微处理器。其中。如图 2 所示，该方法包括如下步骤：

步骤 S201，响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；

在该实施例中，所述微处理器除了执行自己程序，完成正常功能，还用于给芯片的处

理核分配任务段，所述任务段为完整任务的一部分，所述完整任务被预先或者动态的划分为多个任务段以便于处理核存储和执行。当微处理器接收到所述同步信号时，其根据接收到的同步信号判断是否需要给所述芯片中的处理核分配所述任务段。

5 可选的，所述步骤 S201 包括：所述微处理器执行自身程序；响应于接收到同步信号，所述微处理器根据接收到的所述同步信号的次数确定需要分配任务段的处理核。在该实施例中，微处理器中存储有预先加载的分配策略，所述分配策略中预先设置了某个同步周期中，哪个或哪些处理核需要更新其所执行的任务段，所述微处理器每次接收到所述同步信号均会对所述同步信号进行计数，并根据计数的结果判断是否需要给处理核分配任务段。

10 在一个实施例中，所述微处理器还包括计数器，在接收到同步信号之后，还包括：启动所述计数器以记录所述微处理器接收到的所述同步信号的次数。在这之后，所述微处理器可以根据所述同步信号的次数确定需要分配任务段的处理核。

步骤 S202，响应于需要给所述处理核分配任务段，所述微处理器指示分配任务段至所述处理核；

15 可选的，当判断需要给所述处理核分配任务段时，所述微处理器指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理核。

在多数时间里，所述微处理器执行其自身的程序以完成计算等其他任务，所述微处理器中存储有任务段分配指令，所述任务段分配指令可以为通用的分配指令，在执行时所述微处理器获取该指令的参数，如任务段分配指令所指示的待分配任务的处理核以及待分配的任务段在外部存储器中的存储位置等。

20 示例性的，所述存储控制器为 DMAC (Direct Memory Access Controller, DMA 控制器)，微处理器执行到任务段分配指令，根据任务段分配指令的参数配置 DMAC，这样 DMAC 直接将任务段分配指令指示的待分配的任务段从外部存储器中读取出来并发送至对应的处理核中，而无需微处理器参与。

25 示例性的，每次接收到所述同步信号，所述微处理器根据同步信号的次数确定待分配任务的处理核，即需要更新任务段的处理核，以及确定待更新的任务段的存储位置等参数，中断当前所执行的自身的程序并调用任务段分配指令，使用上述参数配置所述任务段分配指令。

30 可选的，所述完整任务可以被预先划分为多个任务段，即预先得到每个任务段的大小和各任务段存储的首地址，当执行任务段分配指令时，可以根据任务段的大小和首地址从所述外部存储器中获取所述任务段。

可选的，所述完整任务段可以被动态的划分为多个任务段，即任务段的数量和大小在分配之前是不确定的。此时，所述微处理器指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理核，包括：所述微处理器获取所述处理核的存储空间的大小；所述微处理器指示所述存储控制器将与所述处理核的存储空间大小相匹配的任务段从所述存储器中读出并发送至所述处理核中。每个处理核都有其自身的存储空间，该存储空间的可用大小可能不相同，因此在微处理器指示给处理核分配任务段时，微处理器先获取处理核的存储空间的大小，通过所述存储空间的大小确定任务段的大小，之后通过任务段的首地址和任务段的大小来决定发送给处理核的任务段在整个任务中的范围。在当前的任务段确定之后，下一任务段的首地址也就确定，用于下次任务段的分配。可选的，所述的动态分配可以每次都动态分配，即每次执行任务段分配指令时都获取处理核的存储空间的大小；或者所述动态分配还可以是仅在第一次动态分配，即认为处理核的存储空间大小在执行所述任务时是固定的，在第一次分配任务段时，根据所述处理核的存储空间大小和完整任务的大小将所述完整任务分成多个任务段，之后每次按照此次划分的任务段的大小来给处理核分配任务段。

15 步骤 S203，所述处理核接收所述任务段；

示例性的，在该步骤中，所述处理核接收通过所述存储控制器发送来的任务段。

可选的，在该步骤中，所述处理核在接收到所述步骤 S201 中所述的同步信号时，开始接收所述任务段，此时所述同步信号在发送给所述微处理器的同时，还发送给所述处理核以指示所述处理核准备接收所述任务段。

20 可选的，在该步骤中，所述处理核也可以不接收所述同步信号，此时可以通过其他方式指示接收任务段，如通过存储控制器的发送信号提示所述处理核接收任务段等，在此不再赘述。

步骤 S204，响应于满足第一预设条件，所述处理核发送同步请求信号，其中所述同步请求信号用于触发生成所述同步信号。

25 在执行所述任务段时，所述处理核在执行任务段的期间和/或任务段执行结束时，发送同步请求信号，其中所述同步请求信号用于触发同步信号的产生。

其中，所述的第一预设条件包括：所述处理核接收完所述任务段或者所述处理核执行接收所述任务段的时间达到预设的时间阈值。

30 在该实施例的一个实现方式中，在所述处理核接收完所述任务段时，调用同步请求信号发送指令，表示可以产生新的同步信号以指示所述微处理器执行下一同步周期的任务段

分配。

- 在该实施例的另一个实现方式中，所述处理核接收任务段的时间可能较长，在一个同步周期中无法接收完，因此在该实现方式中，在开始接收任务的时候启动一个计时器，所述计时器的超时时间为预设的时间阈值，当所述计时器超时，表示接收所述任务段的时间已经超
- 5 已经超过所述时间阈值，此时处理核调用所述同步请求信号发送指令以发送同步请求信号，并重置所述计时器。

可以理解的，所述同步信号可以由芯片中的同步信号发生单元产生，此时同步信号发生单元为专门的部件，其在启动时触发产生第一个同步信号，之后在接收到所有处理核发送来的同步信号请求之后，触发产生同步信号。

- 10 可选的，所述同步信号发生单元可以是程序单元，此时所述同步信号由所述程序单元控制产生，此时需要一个程序，在启动之后产生第一个同步信号，之后在接收到所有处理核发送来的同步信号请求的消息包之后，触发产生同步信号。

- 在步骤 S204 之后，当微处理器接收到所述同步信号，其调用任务段分配指令，并在任务段分配指令中加入包括处理核、任务段地址或大小等的下一任务段的参数，该任务段分配指令用于指示发送下一任务段至根据所述同步信号所确定出的处理核。其中，所述任务段分配指令中的下一任务的参数由预先设置于所述微处理器中的分配策略所确定。
- 15

在一个实施例中，上述任务分配方法还包括：

响应于不需要给所述处理核分配任务段，所述微处理器执行自身程序；

所述处理核执行前一次接收到的任务段；

- 20 响应于满足第二预设条件，所述处理核发送所述同步请求信号。

在该实施例中，在步骤 S201 中判断不需要给所述处理核分配任务段的情况下，所述微处理器执行其自身的程序；此时所述处理核继续执行之前未执行完的任务段，即执行上一次接收到的任务段。

- 其中所述第二预设条件为：所述处理核执行到所述前一次接收到的任务段中的同步请求信号发送指令、或所述处理核执行所述前一次接收到的任务段的时间达到预设的时间阈值。
- 25

可选的，在该实施例的一个实现方式中，所述任务段中被预先插入至少一个同步请求信号发送指令。对于在任务段的中间设置的所述同步信号请求发送指令，处理核执行到所述同步信号请求发送指令时，所述处理核发送同步请求信号，并继续执行所述任务段；对

于在任务段的最后设置的如果所述同步信号请求发送指令,处理核执行到所述同步信号请求发送指令时,所述处理核发送同步请求信号,并等待接收新的任务段。

5 可选的,在该实施例中的一个实现方式中,所述任务段不插入所述同步请求信号发送指令,而是在执行所述任务段时启动一个计时器,所述计时器的超时时间为预设的时间阈值,当所述计时器超时,表示所述任务段的执行时间已经超过所述时间阈值,此时处理核调用所述同步请求信号发送指令以发送同步请求信号,并重置所述计时器。可以理解的,每个任务段的时间阈值可以相同也可以不同,可以在分配任务段时分配每个任务段的时间阈值,这样可以更加灵活的触发同步信号。

10 可选的,在所述任务分配方法中,芯片包括至少两个执行所述任务的处理核,所述微处理器针对所述至少两个处理核分别指示分配所述任务段。对于每个执行所述任务的处理核,其与所述微处理器之间都可以执行如上述步骤 S201-步骤 S204 中的步骤,微处理器可以根据同步信号判断当前的同步周期需要分配任务段的处理核以及任务段的存储地址等。

15 在该实施例中,微处理器接收到第一个同步信号时,该同步信号不需要同步请求信号的触发,此时微处理器调用执行任务段分配指令。任务段分配指令可以将不同任务分成不同的段数,并分给不同的处理核并行处理。示例性的,所述芯片中执行所述任务的处理核为两个,任务段分配指令可以将完整的任务分成四段,依次交给一个处理核执行;还可以首先将完整的任务分成两份,每一份由一个处理核执行,将每一份分为四个任务段,也就是说每个处理核处理四个任务段,当两个处理核所执行的任务段之间没有数据相关性时,
20 两个处理核可以并行处理所述任务段。

在该可选实施例中,所述同步请求信号包括所述至少两个执所述任务的处理核所生成的所述同步请求信号。也就是说,当有多个处理核时,每个处理核都会发送自己的同步请求信号。

25 进一步的,所述同步请求信号用于触发生成所述同步信号,包括:响应于每个所述执行所述任务的处理核都发送了所述同步请求信号,触发生成所述同步信号。在该实施例中,当每个执行所述任务的处理核都发送了同步请求信号之后,才生成所述同步信号。

30 由于任务被分段处理,因此微处理器需要知道任务何时被执行完成,以给处理核分配其他任务。为了确定任务是否被完成,可选的,所述任务分配方法中,所述微处理器还包括任务计数器,所述处理核对应至少一个所述任务计数器,其中在所述微处理器指示分配任务段之后,还包括:所述微处理器启动所述任务计数器以记录已经分配给执行所述任务的所述处理核的任务段的个数;响应于所述处理核的已分配任务段的个数达到预设阈值,重置所述任务计数器。示例性的,一个任务被分为四个任务段,由一个处理核来执行,

则当给所述处理核分配第一个任务段时，启动该处理核所对应的任务计数器，并将任务计数器的数值置为 1，之后每当分配任务段给所述处理核之后，将所述任务计数器的数值累加 1，直至所述任务计数器的值达到 4，重置所述任务计数器，以表示该处理核的任务已经执行完毕。在多个处理核一起完成一个任务时，每个处理核所对应的任务计数器的值都要达到其预设的阈值，才能表示整个任务的完成。某个处理核对应的任务计数器达到其预设的阈值之后，表示该处理核所执行的部分任务已经完成，该处理核可以被分配新的任务。微处理器根据任务计数器的次数确定哪个处理核完成了任务，需要分配新的任务段。

5 可选的，还可以使用预先设置的同步信号次数来判断任务是否完成，由于分配策略中已经预先设置了某个同步周期中需要分配任务段的处理核以及任务段，则可以预先判断任务在哪个同步周期结束，则当同步信号的次数达到所述同步周期时，判断任务结束。

10 上述实施例公开了一种任务分配方法，用于芯片中，所述芯片包括处理核和微处理器，其中该任务分配方法包括：响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；响应于需要给所述处理核分配任务段，所述微处理器指示分配任务段至所述处理核；所述处理核接收所述任务段；响应于满足第一预设条件，所述处理核发送同步请求信号，其中所述同步请求信号用于触发生成所述同步信号。通过上述方法中芯片中的微处理器根据同步信号来分配处理核所执行的任务段，解决了现有技术中的处理核的任务分配不灵活、控制复杂的技术问题技术问题。

20 在上文中，虽然按照上述的顺序描述了上述方法实施例中的各个步骤，本领域技术人员应清楚，本公开实施例中的步骤并不必然按照上述顺序执行，其也可以倒序、并行、交叉等其他顺序执行，而且，在上述步骤的基础上，本领域技术人员也可以再加入其他步骤，这些明显变型或等同替换的方式也应包含在本公开的保护范围之内，在此不再赘述。

25 图 3 为本公开实施例所提供的芯片的结构示意图。如图 3 所示所述芯片 300，包括：处理核 301，微处理器 302 以及存储控制器 303。所述处理核用于执行任务；所述微处理器用于响应于接收到同步信号，根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；响应于需要给所述处理核分配任务段，指示分配任务段至所述处理核；所述存储控制器用于基于所述微处理器的指示，从外部存储空间取出与所述指示对应的任务段，并将所述任务段发送至所述处理核。其中所述处理核还包括内部存储空间 PRAM，所述内部存储空间用于存储所述任务段。

30 其中，所述芯片 300 还可以包括同步信号发生单元 304，其用于产生同步信号 Sync，同步信号发送到每一个处理核和微处理器。

可选的，所述同步信号单元为专门的硬件部件，其在启动时触发产生第一个同步信号，之后在接收到所有处理核发送来的同步信号请求之后，触发产生同步信号。

可选的，所述同步信号发生单元为程序单元，此时所述同步信号由所述程序单元控

制产生，此时需要一个程序，在启动之后产生第一个同步信号，之后在接收到所有处理核发送来的同步信号请求的消息包之后，触发产生同步信号。

在可选的实施方式中，同步信号发生单元还可以仅与微处理器相连，处理核的同步通过微处理器指示 DMAC 来实现，在此不做赘述。下面以附图所示的方式进行描述和说明。

- 5 进一步的，所述处理核还用于：接收所述任务段；响应于满足第一预设条件，所述处理核发送同步请求信号，其中所述同步请求信号用于触发所述同步信号发生单元产生所述同步信号。该实施例中的所述第一预设条件与方法实施例中的相同，产生同步信号的过程也相同，在此不再赘述。如图 3 所示，当满足第一预设条件时，各个所述处理核分别发送同步请求信号 Sync_C1、Sync_C2、……Sync_CN，当所述同步信号发生单元接收到所有的处理核发送的所述同步请求信号之后，生成所述同步信号 Sync，该同步信号同时发送给各个处理核和微处理器。
- 10

进一步的，所述微处理器还用于：执行自身程序；响应于接收到同步信号，根据接收到的所述同步信号的次数确定需要分配任务段的处理核。

- 进一步的，所述微处理器还用于：响应于不需要给所述处理核分配任务段，执行自身程序；所述处理核还用于：执行前一次接收到的任务段；响应于满足第二预设条件，发送所述同步请求信号。
- 15

进一步的，所述第二预设条件为：所述处理核执行到所述前一次接收到的任务段中的同步请求信号发送指令、或所述处理核执行所述前一次接收到的任务段的时间达到预设的时间阈值。

- 20 进一步的，所述芯片包括至少两个执行所述任务的处理核，所述微处理器用于分别针对所述至少两个处理核指示分配所述任务段。

进一步的，所述同步请求信号包括所述至少两个执行所述任务的处理核所生成的所述同步请求信号。

- 进一步的，所述微处理器还包括计数器，在接收到所述同步信号之后，所述微处理器还用于：启动所述计数器以记录所述微处理器接收到的所述同步信号的次数。
- 25

- 如图 3 所示，处理核所执行的任务段由微处理器 302 来调度；微处理器按照一定的策略给各个处理核分配任务，任务的分配策略可以预先设置，典型的可以设置在外存储器中或者其他主机中，初始化时，微处理器从外部读取分配策略以为各个处理核分配对应的任务段。通常微处理器执行自身的程序，一个完整任务的分配策略从第一个同步周期开始。
- 30 在第一个同步周期，微处理器根据分配策略得到第一个同步周期需要分配任务段的处理核以及任务段的存储地址，之后微处理器执行任务段分配指令 PSend，根据该 PSend 指令指示存储控制器给对应的处理核分配任务段，此时微处理器配置存储控制器从外部存储器中读取相应的任务段并发送给对应的处理核。

处理核包括内部存储空间 PRAM，用于存储所述任务段；任务段由一系列指令组成，所述处理核从所述 PRAM 中读取指令、解码指令再执行指令，不再需要从外部存储器中读取指令。每个处理核的 PRAM 的存储空间有限，因此一个完整的任务需要分成多个任务段来执行，处理核执行任务段的过程中或者执行完任务段的时候会执行同步信号请求指令以

5 发送同步信号请求给所述同步信号发生单元。

同步信号发生单元在收到每个处理核的同步信号请求之后，产生新的同步信号；产生的同步信号发送给微处理器和各个处理核；当微处理器接收所述同步信号之后，判断收到的同步信号的次数，根据同步信号的次数判断是否需要给处理核分配任务段，并根据分配策略获取需要分配任务段的处理核以及对应的任务段的存储地址等任务段分配指令所需的

10 参数；在获取到参数之后，根据所述参数配置 PSend 指令以配置存储控制器将 PSend 参数中所指示的任务段从外部存储器中取出发送至 PSend 所指示的处理核中。在不执行 PSend 指令的其他时间里，所述微处理可以执行其他处理，而无需像专用调度器那样空闲，提高了微处理器的执行效率。

以下以一个实际应用场景来说明本公开中的调度过程。如图 4 所示为本公开实施例中

15 所述的芯片的一个实例，在该实例中，所述芯片包括两个处理核 C1 和 C2，一个微处理器，一个存储控制器，其中存储器与外部存储器连接。其中，所述芯片还可以包括同步信号发生单元，可以理解的，所述同步信号发生单元可以实现为一个芯片内部的硬件也可以实现为一个同步信号发生的程序，在此不再赘述。

所述微处理器用于任务的调度，所述任务为一个两层的神经网络的计算任务，如图 5

20 所示，该示例中的神经网络包括两层，layer1 和 layer2，其中 layer1 接收输入数据并计算，将计算结果输入 layer2，layer2 对从 layer1 输入的数据进行计算得到最终的输出。第一层的计算程序 P1 的大小是 64KB，第二层的计算程序 P2 的大小是 128KB，那么整个网络可以按照任务分配策略分配给两个处理核进行流水计算，即 C1 和 C2 各负责一层网络的计算，各自运行相应层的计算程序。输入数据被输入 C1，C1 会将输入数据根据第一层的

25 计算程序进行处理，并将处理结果发送给 C2，C2 将第一层的处理结果作为输入，根据第二层的计算程序进行处理，得到最终的结果并输出。

其流水计算的过程如图 6 所示，在 t1 时刻，C1 接收第一段输入数据，并按照第一层的计算程序对所述输入数据进行处理，将处理结果输出到 C2，需要注意的是，此时 C1 所执行的是所述计算程序的一部分，即上述的任务段，因此其输出的也只是部分处理结果，

30 在 t2 时刻，C1 接着接收输入数据，根据其所执行的任务段对输入数据进行处理，将处理结构输出到 C2；而 C2 在 t2 时刻，根据 C1 在 t1 时刻的输出结果和 C2 上所执行的任务段对所述 C1 在 t1 时刻的输出结果进行处理得到部分输出结果；以此类推，直至整个任务计

算完毕，得到最终的输出结果。

在该示例中，每个核的 PRAM 的大小为 32KB，而第一层的计算程序是 64KB，第二层的计算程序是 128K，因此每一层的计算程序需要按照一定的策略来分配两个核所执行的程序。示例性的，在该示例中，将第一层计算程序分为 2 份，将第二层计算程序分为 4 份，即处理核 C1 处理 2 个任务段，处理核 C2 处理 4 个任务段，整个两层的计算程序被分为 6 个任务段，每个任务段的大小为 32KB。两个处理核在执行任务段的时候是相互独立的，其在数据上有相关性，如 C2 需要等待 C1 的输出数据作为其输入数据，但是其在接收任务段、执行任务段和更新任务段的时候是相互独立的。

图 7 为处理核任务分配和执行的时序图。如图 7 所示，初始时，同步信号发生单元产生第一个同步信号，此时微处理器中断其执行的自身程序，启动计数器准备对同步信号 Sync 进行计数，根据 Sync 的次数以及分配策略确定在第一个同步周期需要分配任务段的处理核 C1 和 C2，以及每个处理核所需要分配的任务段的存储地址等参数，微处理器根据上述参数调用任务段分配指令 PSend 以配置 DMAC；DMAC 根据 PSend 指令从外部存储器中读取 C1 和 C2 的第一个任务段并写入 C1 和 C2 的 PRAM 中，即图 7 中 t0 周期中，C1 和 C2 处于任务段分配状态(U)，以接收所述任务段；分配任务段需要一个同步周期，当 C1 和 C2 完成任务段的接收后，分别执行同步请求信号发送指令并发送同步请求至所述同步信号发生单元；同步信号发生单元在收到 C1 和 C2 都发送了同步信号请求之后，产生 t1 同步周期的同步信号，发送给 C1、C2 和微处理器；在该同步周期中，C1 和 C2 均处于任务段的执行状态(E)，以执行所述任务段；由于 C2 的执行速度较快，在该实施例中，C2 在 t1 的同步周期内执行完了分配的任务段，此时 C2 执行同步请求信号发送指令以产生同步请求信号并发送至同步信号发生单元，而 C1 执行时间较长，还未执行完，但是 C1 的任务段中预先插入了同步请求信号发送指令或者根据 C2 的执行时间在预先设置的时间阈值处执行同步请求信号发送指令，从而，在 C2 执行完毕的时候 C1 也会执行同步请求信号发送指令以产生同步请求信号并发送至同步信号发生单元；当同步信号发生单元接收到 C1 和 C2 均发送了同步请求信号之后，产生 t2 同步周期的同步信号，微处理器此时根据接收到的同步信号的次数，确定在 t2 同步周期内，只有 C2 需要分配任务段，则根据分配策略配置 DMAC 为 C2 分配任务段，任务段分配的步骤与前述的任务段分配步骤相同，在此不再赘述。而 C1 虽然发送了同步请求信号，但是 C1 中的任务段需要执行 3 个同步周期，因此 C1 在 t1, t2, t3 周期执行完时均发送同步请求信号，但是在该几个周期并不接受新任务段，而是继续执行 t0 时间段接收到的任务；当微处理器接收到 C1 在 t3 周期执行完毕发送的同步请求信号时，即微处理器根据接收到的同步信号的次数为 4 时，判断 C1 的任务执行完毕，确定在 t4 同步周期内，C1 和 C2 均需要分配任务段，则配置 DMAC 为 C1 和 C2 分配任务段，任务段分配的步骤与前述任务段分配的方式相同，在此不再赘述。

整个任务的任务段的分配是预先在分配策略中设置好的，在该实施例中，微处理器如果发现当前 Sync 的次数是 4 的倍数，那么会配置 DMAC，分别给 C1 和 C2 分配对应的任务段，如图 7 中的 t0，t4，t8 等时刻，图中对 C1 和 C2 的任务段分配分别用虚线箭头标出；如果发现当前 Sync 的次数是偶数但不是 4 的倍数则只给 C2 分配任务段，如图 7 中 t2，t6 等时刻，如图中对 C2 的任务段分配用虚线箭头标出。

在上述示例中，在第一个同步周期中，根据分配策略确定需要向 C1 和 C2 分配任务段；可以理解的，在实际应用中，在第一个同步周期也可能仅向一个处理核分配任务段，比如 C2 需要等待 C1 的第一个任务段的执行结果作为 C2 的第一个任务段时的输入数据，此时 C2 在第一个同步周期可以执行其他任务或者空闲等待等，更普遍的说，当一个处理核需要等待另一个处理核的输出结果作为其输入时，其在得到另一个处理核的输出结果之前均可以执行其他任务或者空闲等待，这与预先设置好的分配策略相关。

本公开实施例提供一种电子设备，包括：存储器，用于存储计算机可读指令；以及一个或多个处理器，用于运行所述计算机可读指令，使得所述处理器运行时实现实施例中的任一所述任务分配方法。

本公开实施例还提供一种非暂态计算机可读存储介质，其特征在于，该非暂态计算机可读存储介质存储计算机指令，该计算机指令用于使计算机执行前述实施例中的任一所述任务分配方法。

本公开实施例还提供一种计算机程序产品，其中，其特征在于：包括计算机指令，当所述计算机指令被计算设备执行时，所述计算设备可以执行前述实施例中的任一所述任务分配方法。

本公开实施例还提供一种计算装置，其特征在于，包括所述实施例中的任一所述的芯片。

本公开附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、任务段、或代码的一部分，该模块、任务段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开实施例中所涉及到的单元可以通过软件的方式实现,也可以通过硬件的方式来实现。其中,单元的名称在某种情况下并不构成对该单元本身的限定。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如,非限制性地,可以使用的示范类型的硬件逻辑部件包括:现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、片上系统(SOC)、复杂可编程逻辑设备(CPLD)等等。

在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

15

20

25

30

权 利 要 求 书

1、一种任务分配方法，用于芯片中，所述芯片包括处理核和微处理器，其特征在于，所述方法包括：

5 响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；

响应于需要给所述处理核分配任务段，所述微处理器指示分配任务段至所述处理核；

所述处理核接收所述任务段；

10 响应于满足第一预设条件，所述处理核发送同步请求信号，其中所述同步请求信号用于触发生成所述同步信号。

2、如权利要求 1 所述的任务分配方法，其中，所述响应于接收到同步信号，所述微处理器根据所述同步信号判断是否需要给所述处理核分配任务段，包括：

所述微处理执行自身程序；

15 响应于接收到同步信号，所述微处理器根据接收到的所述同步信号的次数确定需要分配任务段的处理核。

3、如权利要求 1 所述的任务分配方法，其中所述第一预设条件包括：所述处理核接收完所述任务段或者所述处理核执行接收所述任务段的时间达到预设的时间阈值。

20 4、如权利要求 1-3 任一项所述的任务分配方法，其中所述微处理器指示分配任务段至所述处理核，包括：

所述微处理器指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理核。

5、如权利要求 4 所述的任务分配方法，其中，所述微处理器指示存储控制器将所述任务段从外部存储器中读出并发送至所述处理核，包括：

25 所述微处理获取所述处理核的存储空间的大小；

所述微处理器指示所述存储控制器将与所述处理核的存储空间大小相匹配的任务段从所述存储器中读出并发送至所述处理核。

6、如权利要求 1 所述的任务分配方法，其中，所述方法还包括：

响应于不需要给所述处理核分配任务段，所述微处理器执行自身程序；

所述处理核执行前一次接收到的任务段；

响应于满足第二预设条件，所述处理核发送所述同步请求信号。

- 7、如权利要求 6 所述的任务分配方法，其中，所述第二预设条件为：所述处理核执行到所述前一次接收到的任务段中的同步请求信号发送指令、或所述处理核执行所述前一次接收到的任务段的时间达到预设的时间阈值。

8、如权利要求 1-7 任一项所述的任务分配方法，其中所述芯片包括至少两个执行所述任务的处理核，所述微处理器分别针对所述至少两个处理核指示分配所述任务段。

- 9、如权利要求 8 所述的任务分配方法，其中所述同步请求信号包括所述至少两个执行所述任务的处理核所生成的所述同步请求信号。

10、如权利要求 8 或 9 任一项所述的任务分配方法，所述同步请求信号用于触发生成所述同步信号，包括：

响应于每个所述执行所述任务的处理核都发送了所述同步请求信号，触发生成所述同步信号。

- 11、如权利要求 1-10 中任一项所述的任务分配方法，其特征在于，所述微处理器包括计数器，在接收到所述同步信号之后，还包括：

启动所述计数器以记录所述微处理器接收到的所述同步信号的次数。

12、一种芯片，其特征在于，包括：

处理核，所述处理核用于执行任务；

- 20 微处理器，用于响应于接收到同步信号，根据所述同步信号判断是否需要给所述处理核分配任务段，其中所述任务段为所述任务的一部分；响应于需要给所述处理核分配任务段，指示分配任务段至所述处理核；

存储控制器，用于基于所述微处理器的指示，从外部存储空间取出与所述指示对应的任务段，并将所述任务段发送至所述处理核。

- 25 13、如权利要求 12 所述的芯片，其特征在于，所述处理核还用于：

接收所述任务段；响应于满足第一预设条件，发送同步请求信号，其中所述同步请求信号用于触发生成所述同步信号。

- 30 14、如权利要求 12 或 13 中任一项所述的芯片，其中：所述芯片包括至少两个执行所述任务的处理核，所述微处理器用于分别针对所述至少两个处理核指示分配所述任务段。

15、一种电子设备，包括：存储器，用于存储计算机可读指令；以及一个或多个处理器，用于运行所述计算机可读指令，使得所述处理器运行时实现前述权利要求 1-11 所述的任务分配方法。

5 16、一种非暂态计算机可读存储介质，其特征在于，该非暂态计算机可读存储介质存储计算机指令，该计算机指令用于使计算机执行前述权利要求 1-11 所述的任务分配方法。

17、一种计算机程序产品，其特征在于：包括计算机指令，当所述计算机指令被计算设备执行时，所述计算设备可以执行前述权利要求 1-11 所述的任务分配方法。

18、一种计算装置，其特征在于，包括一个或多个前述权利要求 12-14 所述的芯片。