



(12) 发明专利申请

(10) 申请公布号 CN 118519613 A

(43) 申请公布日 2024. 08. 20

(21) 申请号 202410985037.8

(22) 申请日 2024.07.23

(71) 申请人 珠海皓泽科技有限公司

地址 519085 广东省珠海市高新区唐家湾
镇鼎兴路129号9栋9层904-2、904-3房
间

(72) 发明人 谢成兴 王军 杨帆

(74) 专利代理机构 北京中强智尚知识产权代理
有限公司 11448

专利代理师 高荣艳

(51) Int. Cl.

G06F 7/544 (2006.01)

G06F 17/16 (2006.01)

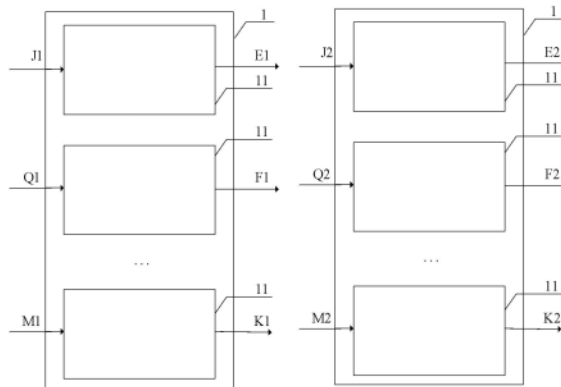
权利要求书2页 说明书10页 附图5页

(54) 发明名称

乘累加器运算集群及数据处理方法

(57) 摘要

本申请公开一种乘累加器运算集群及数据处理方法。乘累加器运算集群包括多个乘累加器单元,每个乘累加器单元用于完成至少一组待处理数据的乘累加运算,在待处理数据的计算维度小于乘累加器单元的总维度时,通过乘累加器单元中的至少一个乘累加器子单元完成每组待处理数据的乘累加运算,并输出待处理数据的乘累加运算结果;和/或,在待处理数据包括多个顺序排列的数据块时,通过乘累加器运算集群中的多个乘累加器分集群分别完成一个数据块的乘累加运算,并将数据块的乘累加运算结果输出到顺序排列的下一个乘累加器分集群中,直至完成所有数据块的乘累加运算,输出待处理数据的乘累加运算结果。该乘累加器运算集群提高了乘累加器运算单元的利用率。



1. 一种乘累加器运算集群, 其特征在于, 所述乘累加器运算集群包括多个乘累加器单元, 每个所述乘累加器单元用于完成至少一组待处理数据的乘累加运算, 每组所述待处理数据包括两个多维度数据, 每个所述乘累加器单元均包括多个乘累加器子单元, 每个乘累加器子单元用于完成一次乘累加运算, 并输出乘累加运算结果, 其中,

在所述待处理数据中的两个所述多维度数据的计算维度小于所述乘累加器单元的总维度时, 通过所述乘累加器单元中的至少一个乘累加器子单元完成每组所述待处理数据的乘累加运算, 并输出所述待处理数据的乘累加运算结果; 和/或,

在所述待处理数据包括多个顺序排列的数据块时, 通过所述乘累加器运算集群中的多个乘累加器分集群分别完成一个数据块的乘累加运算, 并将所述数据块的乘累加运算结果输出到顺序排列的下一个所述乘累加器分集群中, 直至完成所有所述数据块的乘累加运算, 输出所述待处理数据的乘累加运算结果。

2. 根据权利要求1所述的乘累加器运算集群, 其特征在于, 同一个乘累加器单元中的乘累加器子单元的维度相同, 每个所述乘累加器单元中的所有乘累加器子单元的维度的总和与所述乘累加器单元的总维度相同, 且多个不同的乘累加器单元中的乘累加器子单元的维度彼此相同或不同。

3. 根据权利要求2所述的乘累加器运算集群, 其特征在于, 当一组待处理数据的计算维度小于或等于至少一个乘累加器单元中乘累加器子单元的维度时, 根据所述待处理数据的计算维度和所述乘累加器单元中乘累加器子单元的维度, 确定运算所述待处理数据的目标乘累加器单元, 通过所述目标乘累加器单元中的乘累加器子单元对所述待处理数据进行乘累加运算。

4. 根据权利要求2所述的乘累加器运算集群, 其特征在于, 当一组待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度且小于每个所述乘累加器单元的总维度时, 通过所述乘累加器单元中的多个乘累加器子单元完成所述待处理数据的乘累加运算。

5. 根据权利要求4所述的乘累加器运算集群, 其特征在于, 在所述待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度且小于每个所述乘累加器单元的总维度时, 根据所述待处理数据的计算维度和每个所述乘累加器单元中乘累加器子单元的维度, 确定运算所述待处理数据的目标乘累加器以及所述目标乘累加器中乘累加器子单元的目标数量, 通过所述目标乘累加器中所述目标数量的乘累加器子单元对所述待处理数据进行乘累积运算。

6. 根据权利要求1所述的乘累加器运算集群, 其特征在于, 所述乘累加器运算集群中的至少一个乘累加器子单元作为一个乘累加器分集群。

7. 根据权利要求6所述的乘累加器运算集群, 其特征在于, 当所述待处理数据中所有数据块的计算维度均小于或等于任一乘累加器单元中乘累加器子单元的维度时, 将所述乘累加器单元中的每个乘累加器子单元作为一个乘累加器分集群, 通过所述乘累加器单元中的多个乘累加器子单元分别完成一个数据块的乘累加运算, 并将所述数据块的乘累加运算结果输出到顺序排列的下一个所述乘累加器子单元, 直至完成所有所述数据块的乘累加运算, 输出所述待处理数据的乘累加运算结果。

8. 根据权利要求1-7中任一项所述的乘累加器运算集群, 其特征在于, 所述乘累加器单

元包括多个数据输入模块、多个乘运算模块、多个多级加运算模块和多个数据输出模块,其中,一个乘运算模块、一个多级加运算模块和多个数据输出模块构成一个乘累加器子单元,每个所述乘运算模块包括多个乘法器,每个所述多级加运算模块包括多个加法器;

每个所述数据输入模块的输入端分别与数据提供模块电连接,每个所述数据输入模块的输出端分别与一个所述乘运算模块的输入端电连接,每个所述乘运算模块的输出端与一个多级加运算模块的输入端电连接,每个所述多级加运算模块的输出端分别与多个所述数据输出模块电连接。

9. 一种数据处理方法,应用于上述权利要求1-8中任一项所述的乘累加器运算集群,其特征在于,所述数据处理方法,包括:

获取待处理数据,根据所述待处理数据的计算维度和每个乘累加器单元中的乘累加器子单元的维度,确定所述待处理数据对应的目标乘累加器单元;

将所述待处理数据输入至所述目标乘累加器中处于空闲状态的乘累加器子单元中,通过所述目标乘累加器对所述待处理数据进行乘累加运算,得到所述待处理数据的乘累加运算结果。

10. 根据权利要求9所述的数据处理方法,其特征在于,所述获取待处理数据,根据所述待处理数据的计算维度和每个乘累加器单元中的乘累加器子单元的子维度,确定所述待处理数据对应的目标乘累加器单元,包括:

将所述乘累加器单元中乘累加器子单元的维度大于或等于所述待处理数据的计算维度的乘累加器单元作为候选乘累加器单元;

将处于空闲工作状态且乘累加器子单元的维度最大的候选乘累加器单元作为所述待处理数据对应的目标乘累加器单元。

乘累加器运算集群及数据处理方法

技术领域

[0001] 本申请属于数字电路技术领域,具体涉及一种乘累加器运算集群及数据处理方法。

背景技术

[0002] 乘累加器(Multiply and Accumulate,简称 MAC)用于完成向量相乘、矩阵相乘和向量矩阵互乘等乘累加运算,是协处理器、数字信号处理器、中央处理器(Central Processing Unit,简称 CPU)等处理器中极其重要的运算子系统,尤其是人工智能(AI)加速器中的最基本的运算单元。

[0003] AI加速器中的MAC一般可实现X组M维向量的点积,例如两组M维向量(A向量和B向量)的乘加运算,先完成 $A_1 \cdot B_1, A_2 \cdot B_2 \dots A_X \cdot B_X$ 运算,然后再将乘积相加求和,其中每个向量(A或B)的处理维数以M表示。每个数据可能是不同位宽的整数或浮点数,以输入为8bits整数为例,则MAC的每个时钟周期的算力为M乘以X乘以2(2理解为一个乘法和一个加法的两次运算操作)的8bits整数运算。但是实际的应用过程中,由于一些原因会导致MAC运算单元利用率低。例如,当A向量或B向量的处理维数小于乘累加器的维度M时(乘累加器的维度代表包含的乘法器的数量),进行处理数据时会有一些乘法器处于空闲状态,只有当这次处理结束后,空闲状态的乘法器才可以进行下一次的使用,造成了乘累加器中的乘法器使用率低的情况。例如,乘累加器的输入数据没有及时准备好或者乘累加器的输出数据没有及时输出,造成线路阻塞,乘累加器不能进入下一次的使用,也会导致乘累加器运算单元利用率低。

[0004] 目前没有好的电路结构能解决乘累加器运算单元利用率低的问题。

发明内容

[0005] 有鉴于此,本申请提供了一种乘累加器运算集群及数据处理方法,主要目的在于解决现有的乘累加器运算单元利用率低的问题。

[0006] 为解决上述问题,本申请提供一种乘累加器运算集群,所述乘累加器运算集群包括多个乘累加器单元,每个所述乘累加器单元用于完成至少一组待处理数据的乘累加运算,每组所述待处理数据包括两个多维度数据,每个所述乘累加器单元均包括多个乘累加器子单元,每个乘累加器子单元用于完成一次乘累加运算,并输出乘累加运算结果,其中,

在所述待处理数据中的两个所述多维度数据的计算维度小于所述乘累加器单元的总维度时,通过所述乘累加器单元中的至少一个乘累加器子单元完成每组所述待处理数据的乘累加运算,并输出所述待处理数据的乘累加运算结果;和/或,在所述待处理数据包括多个顺序排列的数据块时,通过所述乘累加器运算集群中的多个乘累加器分集群分别完成一个数据块的乘累加运算,并将所述数据块的乘累加运算结果输出到顺序排列的下一个所述乘累加器分集群中,直至完成所有所述数据块的乘累加运算,输出所述待处理数据的乘累加运算结果。

[0007] 在本申请的一个实施例中,可选地,同一个乘累加器单元中的乘累加器子单元的维度相同,每个所述乘累加器单元中的所有乘累加器子单元的维度的总和与所述乘累加器单元的总维度相同,且多个不同的乘累加器单元中的乘累加器子单元的维度彼此相同或不同。

[0008] 在本申请的一个实施例中,可选地,当一组待处理数据的计算维度小于或等于至少一个乘累加器单元中乘累加器子单元的维度时,根据待处理数据的计算维度和所述乘累加器单元中乘累加器子单元的维度,确定运算所述待处理数据的目标乘累加器单元,通过所述目标乘累加器单元中的乘累加器子单元对所述待处理数据进行乘累加运算。

[0009] 在本申请的一个实施例中,可选地,当一组待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度且小于每个所述乘累加器单元的总维度时,通过所述乘累加器单元中的多个乘累加器子单元完成所述待处理数据的乘累加运算。

[0010] 在本申请的一个实施例中,可选地,在所述待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度且小于每个所述乘累加器单元的总维度时,根据待处理数据的计算维度和每个乘累加器单元中乘累加器子单元的维度,确定运算所述待处理数据的目标乘累加器以及所述目标乘累加器中乘累加器子单元的目标数量,通过所述目标乘累加器中目标数量的乘累加器子单元对所述待处理数据进行乘累积运算。

[0011] 在本申请的一个实施例中,可选地,所述乘累加器运算集群中的至少一个乘累加器子单元作为一个乘累加器分集群。

[0012] 在本申请的一个实施例中,可选地,当所述待处理数据中所有数据块的计算维度均小于或等于任一乘累加器单元中乘累加器子单元的维度时,将所述乘累加器单元中的每个乘累加器子单元作为一个乘累加器分集群,通过所述乘累加器单元中的多个乘累加器子单元分别完成一个数据块的乘累加运算,并将所述数据块的乘累加运算结果输出到顺序排列的下一个所述乘累加器子单元,直至完成所有所述数据块的乘累加运算,输出所述待处理数据的乘累加运算结果。

[0013] 在本申请的一个实施例中,可选地,所述乘累加器单元包括多个数据输入模块、多个乘运算模块、多个多级加运算模块和多个数据输出模块,其中,一个乘运算模块、一个多级加运算模块和多个数据输出模块构成一个乘累加器子单元,每个所述乘运算模块包括多个乘法器,每个所述多级加运算模块包括多个加法器;

每个所述数据输入模块的输入端分别与数据提供模块电连接,每个所述数据输入模块的输出端分别与一个所述乘运算模块的输入端电连接,每个所述乘运算模块的输出端与一个多级加运算模块的输入端电连接,每个所述多级加运算模块的输出端分别与多个数据输出模块电连接。

[0014] 本申请还提供一种数据处理方法,应用于上述所述的乘累加器运算集群,所述数据处理方法,包括:

获取待处理数据,根据所述待处理数据的计算维度和每个乘累加器单元中的乘累加器子单元的维度,确定所述待处理数据对应的目标乘累加器单元;

将所述待处理数据输入至所述目标乘累加器中处于空闲状态的乘累加器子单元中,通过所述目标乘累加器对所述待处理数据进行乘累加运算,得到所述待处理数据的乘累加运算结果。

[0015] 在本申请的一个实施例中,可选地,所述获取待处理数据,根据所述待处理数据的计算维度和每个乘累加器单元中的乘累加器子单元的子维度,确定所述待处理数据对应的目标乘累加器单元,包括:

将所述乘累加器单元中乘累加器子单元的维度大于或等于所述待处理数据的计算维度的乘累加器单元作为候选乘累加器单元;将处于空闲工作状态且乘累加器子单元的维度最大的候选乘累加器单元作为所述待处理数据对应的目标乘累加器单元。

[0016] 本申请中的有益效果:本申请提供一种乘累加器运算集群及数据处理方法,乘累加器运算集群包括多个乘累加器单元,每个乘累加器单元均包括多个乘累加器子单元,当待处理数据中的两个多维度数据的计算维度小于乘累加器单元的总维度时,将多个待处理数据各自输入一个乘累加器子单元,在同一个时钟周期内,每个乘累加器单元中的多个乘累加器子单元能够同时对各自输入的待处理数据进行乘累加运算,提高乘累加器运算单元的利用率,还增大了乘累加器的输出带宽;当待处理数据包括多个顺序排列的数据块时,一个乘累加器分集群完成一个数据块的一次乘累加运算,并将该数据块的乘累加运算结果输出到顺序排列的下一个乘累加器分集群中,实现待处理数据的流水线工作方式的处理,在同一个时钟周期内,多个乘累加器分集群同时对各自输入的数据进行处理,且乘累加器的输入数据及时准备好,输出端及时输出运算结果,避免存在堵塞线路的问题,提高了乘累加器运算单元的利用率。

[0017] 上述说明仅是本申请技术方案的概述,为了能够更清楚了解本申请的技术手段,而可依照说明书的内容予以实施,并且为了让本申请的上述和其它目的、特征和优点能够更明显易懂,以下特举本申请的具体实施方式。

附图说明

[0018] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本申请的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

图1为一种传统的乘累加器单元的结构示意图;

图2为本申请示例性实施例的一种乘累加器运算集群的结构框图;

图3为本申请示例性实施例的另一种乘累加器运算集群的结构框图;

图4为一种传统的乘累加器单元的运算示意图;

图5为本申请示例性实施例的一种乘累加器单元的运算示意图;

图6为一种流水线处理方式的待处理数据使用传统的乘累加器单元的示意图;

图7为本申请示例性实施例的一种流水线处理方式的待处理数据使用乘累加器单元的示意图;

图8为本申请示例性实施例的又一种乘累加器运算集群的结构框图;

图9为本申请示例性实施例的一种乘累加器运算集群的结构连接示意图;

图10为本申请示例性实施例的一种乘累加器子单元的结构连接示意图;

图11为本申请示例性实施例的一种数据处理方法的流程图;

其中,

图中的标号如下:1-乘累加器单元;11-乘累加器子单元;2-乘累加器分集群;12-

数据输入模块;111-乘运算模块;112-多级加运算模块;113-数据输出模块。

具体实施方式

[0019] 下文中将参考附图并结合实施例来详细说明本申请。需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。

[0020] 为更进一步阐述本申请为达成预定发明目的所采取的技术手段及功效,以下结合附图及较佳实施例,对依据本申请申请的具体实施方式、结构、特征及其功效,详细说明如后。在下述说明中,不同的“一实施例”或“实施例”指的不一定是同一实施例。此外,一或多个实施例中的特定特征、结构、或特点可由任何合适形式组合。

[0021] 在现有技术中,通常使用的乘累加器单元只有一个数据输出端,如图1所示,图中D是数据输出端。如果某AI处理器中有8bits整数的MAC,其中 $M=32$, $X=16$,M是乘法器的数量。在做通常的卷积运算时,例如做 3乘以3的depthwise (深度方向) 卷积运算,进行3乘以3卷积运算时仅使用9个乘法器。通常架构下的MAC只有一个数据输出端,进行运算时用到的乘法器参与计算,用不到的乘法器处于空闲等待状态,如果计算这个3乘以3的卷积,使用的乘法器个数是 $X'=9$,乘累加器单元中的乘法器是32个,会有很多乘法器处于空闲状态,只有等这次卷积计算结束,计算结果输出之后,整个MAC再等待下一次的运算,3乘以3的卷积运算时MAC的运算利用率仅为 $9/32$,运算利用率比较低。

[0022] 本申请提出了一种乘累加器运算集群,如图2所示,乘累加器运算集群包括多个乘累加器单元1,每个乘累加器单元1用于完成至少一组待处理数据的乘累加运算,每组待处理数据包括两个多维度数据,每个乘累加器单元1均包括多个乘累加器子单元11,每个乘累加器子单元11用于完成一次乘累加运算,并输出乘累加运算结果,其中,在待处理数据中的两个多维度数据的计算维度小于乘累加器单元1的总维度时,通过乘累加器单元1中的至少一个乘累加器子单元11完成每组待处理数据的乘累加运算,并输出待处理数据的乘累加运算结果;和/或,如图3所示,在待处理数据包括多个顺序排列的数据块时,通过乘累加器运算集群中的多个乘累加器分集群2分别完成一个数据块的乘累加运算,并将数据块的乘累加运算结果输出到顺序排列的下一个乘累加器分集群2中,直至完成所有数据块的乘累加运算,输出待处理数据的乘累加运算结果。

[0023] 具体的,乘累加器运算集群包括多个乘累加器单元,每个乘累加器单元包括至少一个乘累加器子单元,每个乘累加器子单元能够对待处理数据进行独立的处理,在同一个时钟周期内,当多组待处理数据分别送入同一个乘累加器单元中的多个乘累加器子单元时,多个乘累加器子单元分别对各自输入的待处理数据进行乘累加运算,各自输出运算结果,实现在同一时钟周期内,同一个乘累加器单元对多个待处理数据进行运算处理,提高了乘累加器单元的运算利用率。

[0024] 待处理数据的计算维度小于乘累加器单元的维度时,待处理数据使用传统的乘累加器时只能有效利用到很少的MAC算力,如图4,当有乘累加器1/4维度的待处理数据输入时,传统乘累加器单元中只有1/4的乘法器使用,其余乘法器空闲,运算利用率很低。当乘累加器单元被划分为多个乘累加器子单元时,例如如图5所示,乘累加器单元被划分为4个乘累加器子单元时,乘累加器单元的输入总维度不变,每个乘累加器子单元的输入维度为总维度的1/4,在一个时钟周期内,有4组1/4维度的待处理数据各自输入一个乘累加器子单

元,每个乘累加器子单元各自进行数据处理,乘累加器单元中每个乘法器都得到了使用,利用到了所有的运算能力,因此提高了运算利用率。

[0025] 例如,一个乘累加器单元包括4个乘累加器子单元,每个乘累加器子单元包括的乘法器数量都是9,也就是说乘累加器子单元的维度是9,这个乘累加器单元的总维度是36。如果有两组待输入数据,两组待处理数据的计算维度分别是8和9,在同一个时钟周期内,通过该乘累加器单元中的两个乘累加器子单元对这两组待处理数据分别进行乘累加运算,各自得到输出结果,此时乘累加器单元的运算利用率是 $8+9/36$,如果这两个待处理数据通过传统的乘累加器单元进行运算处理,需要两个时钟周期,每个时钟周期处理一组待处理数据,两个时钟周期内的乘累加器单元的运算利用率分别是 $8/36$ 和 $9/36$,而通过本申请的乘累加器运算集群在同一个时钟周期即可完成,且乘累加器单元的运算利用率高。

[0026] 当乘累加器运算集群的输入数据没有及时准备好或者乘累加器运算集群的输出数据没有及时输出,造成线路阻塞,导致乘累加器运算集群不能进入下一次的使用,也会导致乘累加器运算单元利用率低。

[0027] 当待处理数据是按流水作业处理方式进行处理的数据时,如果使用传统的乘累加器,待处理数据通过乘累加器进行第一次乘累加运算后,运算结果再输入这个乘累加器,进行第二次乘累加运算,运算结果再输入这个乘累加器单元,直至待处理数据的运算结束为止,乘累加器的运算利用率比较低。

[0028] 因此针对按流水作业处理方式进行处理的数据,将乘累加器运算集群分成多个乘累加器分集群,多个乘累加器分集群按顺序排列,待处理数据包括多个顺次排列的数据块,每个时钟周期有一个数据块输入至第一排列次序的累加器分集群,在第一个时钟周期内,第一排列次序的乘累加器分集群对输入的数据进行处理,处理结果被传输至第二排列次序的乘累加器分集群的数据输入端,在第二个时钟周期内,第一排列次序的乘累加器分集群有新的数据块进入,第一排列次序的乘累加器分集群和第二排列次序的乘累加器分集群各自对输入的数据进行处理,第二排列次序的乘累加器分集群的处理结果被传输至第三排列次序的乘累加器分集群的数据输入端,第一排列次序的乘累加器分集群的处理结果被传输至第二排列次序的乘累加器分集群的数据输入端,在第三个时钟周期,新的数据块被送入第一排列次序的乘累加器分集群的输入端,重复上述过程,将前一排列次序的乘累加器分集群的输出数据输入至相邻的后一排列次序的乘累加器分集群的数据输入端,使得在同一个时钟周期内,多个乘累加器分集群同时对各自输入的数据进行处理,且使得每个乘累加器分集群的数据输入端及时准备好数据,数据输出端及时将数据取走,提高了乘累加器的运算利用率。

[0029] 例如,神经网络模型的处理过程经常被定义为按先后顺序的若干个layer网络层的运算过程。通常情况下若干个layer网络层的计算过程是,layer1网络层调用整个MAC来进行运算,其输入输出带宽皆为W,MAC运算结果是layer2网络层的输入数据,layer1网络层全部运算完毕后,layer2网络层继续调用整个MAC来进行运算,直到完成整个计算为止,如图6所示,可以认为是AI加速器系统中对MAC的分时复用,也就是说每个layer网络层使用整个MAC,MAC的运算利用率比较低,此时的带宽可认为是每个layer网络层的带宽。

[0030] 本申请提出的乘累加器运算集群包括多个乘累加器分集群,例如乘累加器运算集群包括4个乘累加器分集群P1_MAC、P2_MAC、P3_MAC、P4_MAC。多个layer网络层之间采用流

水作业的处理方式,将layer1网络层使用乘累加器分集群P1_MAC进行运算处理,layer2网络层使用乘累加器分集群P2_MAC进行运算处理,layer3网络层使用乘累加器分集群P3_MAC进行运算处理,layer4网络层使用乘累加器分集群P4_MAC进行运算处理。在第一个时钟周期,待处理数据输入至乘累加器分集群P1_MAC的输入端,乘累加器分集群P1_MAC运算处理结束之后,将乘累加器分集群P1_MAC的运算处理结果作为乘累加器分集群P2_MAC的输入数据,在第二个时钟周期,新的待处理数据输入至乘累加器分集群P1_MAC的输入端,乘累加器分集群P1_MAC和乘累加器分集群P2_MAC同时对各自输入的数据进行运算处理,运算结束之后,将乘累加器分集群P1_MAC的运算处理结果作为乘累加器分集群P2_MAC的输入数据,乘累加器分集群P2_MAC的运算处理结果作为乘累加器分集群P3_MAC的输入数据,类似地,将前一排列次序的乘累加器分集群的处理结果作为下一排列次序的乘累加器分集群的输入数据,如图7所示,P1_MAC有待处理数据依次输入,相应的P2_MAC~P4_MAC也依次有待处理数据依次输入,在同一个时钟周期内,P1_MAC~P4_MAC各自对输入的数据进行运算处理,整个乘累加器分集群的总的算力得到提高,输出带宽变为每个layer网络层的带宽之和,可以认为输出带宽变为原来的4倍,满足了增大带宽的目的,但是每个乘累加器分集群的输出带宽不变。

[0031] 乘累加器运算集群包括多个乘累加器单元,每个乘累加器单元包括至少一个乘累加器子单元,每个乘累加器子单元能够对待处理数据进行独立的处理,实现在同一时钟周期内,提高了乘累加器单元的运算利用率,但是乘累加器单元的输出带宽增加很多,增加的输出带宽可能会造成后续的处理模块不能及时进行处理,如果乘累加器单元输出的数据不能及时被取走,该乘累加器单元不能进行下一次的运算,造成线路堵塞,影响乘累加器单元的运算效率,因此,将乘累加器子单元与乘累加器分集群结合,至少一个乘累加器子单元构成乘累加器分集群,如图8所示,一个乘累加器子单元作为一个乘累加器分集群,按排列顺次,将前一乘累加器子单元的输出数据送入下一乘累加器子单元的数据输入端,使得乘累加器子单元的输出数据及时处理,不造成线路堵塞,提高了乘累加运算集群的运算利用率。

[0032] 与现有技术相比,本申请提供的一种乘累加器运算集群,乘累加器运算集群包括多个乘累加器单元,每个乘累加器单元均包括多个乘累加器子单元,当待处理数据中的两个多维度数据的计算维度小于乘累加器单元的总维度时,在同一个时钟周期内,每个乘累加器单元中的多个乘累加器子单元能够同时对各自输入的待处理数据进行乘累加运算,提高乘累加器运算单元的利用率,还增大了乘累加器的输出带宽;当待处理数据包括多个顺序排列的数据块时,一个乘累加器分集群完成一个数据块的一次乘累加运算,并将该数据块的乘累加运算结果输出到顺序排列的下一个乘累加器分集群中,实现待处理数据的流水线工作方式的处理,在同一个时钟周期内,多个乘累加器分集群同时对各自输入的数据进行处理,且乘累加器的输入数据及时准备好,输出端及时输出运算结果,避免存在堵塞线路的问题,提高了乘累加器运算单元的利用率。

[0033] 在一个实施例中,同一个乘累加器单元中的乘累加器子单元的维度相同,每个乘累加器单元中的所有乘累加器子单元的维度的总和与乘累加器单元的总维度相同,且多个不同的乘累加器单元中的乘累加器子单元的维度彼此相同或不同。

[0034] 具体的,乘累加器单元包含的乘法器的数量就是乘累加器单元的维度,同一乘累加器单元包含的乘累加器子单元中的乘法器数量相同,每个乘累加器中所有乘累加子单元

的乘法器的总数量等于每个乘累加器单元的总维度。乘累加器运算集群中有多个乘累加器单元,多个乘累加器单元中必须存在乘法器数量不同的一些乘累加器单元,相同乘法器数量的乘累加单元的数量可以为多个。例如乘累加器运算集群包括多个乘累加器单元,其中,两个乘累加器单元中的乘累加器子单元包括的乘法器数量都是9,一个乘累加器单元中的乘累加器子单元包括的乘法器数量都是12,一个乘累加器单元中的乘累加器子单元包括的乘法器数量都是16等。

[0035] 在一个实施例中,当一组待处理数据的计算维度小于或等于至少一个乘累加器单元中乘累加器子单元的维度时,根据待处理数据的计算维度和乘累加器单元中乘累加器子单元的维度,确定运算待处理数据的目标乘累加器单元,通过目标乘累加器单元中的乘累加器子单元对待处理数据进行乘累加运算。

[0036] 具体的,当一组待处理数据的计算维度小于或等于至少一个乘累加器单元中乘累加器子单元的维度时,将该组处理数据的计算维度与这些乘累加器单元中乘累加器子单元的维度做比较,选择计算维度与乘累加器子单元的维度的差值最小的乘累加器单元作为目标乘累加器单元,通过目标乘累加器单元中处于空闲状态的乘累加器子单元进行乘加运算。例如,多个乘累加器单元中的乘累加子单元的维度分为是6、9、12和16,待处理数据需要使用的乘法器数量是8个时,将乘累加器子单元的维度大于8的乘累加器单元作为候选乘累加器单元,即乘累加器子单元的维度是9、12和16的乘累加器单元作为候选乘累加器单元,获取计算维度与每个候选乘累加器中乘累加器子单元的维度的差值,差值最小的是乘累加器子单元的维度是9的候选乘累加器单元,将这个候选乘累加器单元作为目标乘累加器单元。

[0037] 在一个实施例中,当一组待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度且小于每个乘累加器单元的总维度时,通过乘累加器单元中的多个乘累加器子单元完成待处理数据的乘累加运算。

[0038] 具体的,当一组待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度时,说明不能使用乘累加器单元中的一个乘累加器子单元进行运算处理,需要使用乘累加器单元中的多个乘累加器子单元合并处理该待处理数据,即通过乘累加器单元中的多个乘累加器子单元共同处理该待处理数据。

[0039] 在一个实施例中,在待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度且小于每个乘累加器单元的总维度时,根据待处理数据的计算维度和每个乘累加器单元中乘累加器子单元的维度,确定运算待处理数据的目标乘累加器以及目标乘累加器中乘累加器子单元的目标数量,通过目标乘累加器中目标数量的乘累加器子单元对待处理数据进行乘累积运算。

[0040] 具体的,当一组待处理数据的计算维度大于所有乘累加器单元中乘累加器子单元的维度时,需要使用乘累加器单元中的多个乘累加器子单元沟通处理该待处理数据。将该待处理数据的计算维度与任一乘累加器单元中乘累加器子单元的维度的商加一作为目标数量,通过这个乘累加器单元中目标数量个乘累加器子单元对该待处理数据进行运算处理。

[0041] 在一个实施例中,乘累加器运算集群中的至少一个乘累加器子单元作为一个乘累加器分集群。

[0042] 具体的,当同一个乘累加器单元中的多个乘累加器子单元各自独立进行运算处理时,提高了乘累加器单元的运算利用率,但同时乘累加器单元的输出带宽变为原来的多倍,输出带宽的增加会导致处理线路阻塞。因此,将至少一个乘累加器子单元作为一个乘累加器分集群,通过乘累加器分集群对待处理数据进行流水作业方式处理,也就是将前一排列次序的乘累加器分集群输出的数据输入至后一排列次序的乘累加器分集群的数据输入端,这样虽然乘累加器单元的输出带宽增大,但是乘累加器子单元的输出数据及时被处理,不会产生堵塞现象,提高了乘累加器的运算利用率。

[0043] 在一个实施例中,当待处理数据中所有数据块的计算维度均小于或等于任一乘累加器单元中乘累加器子单元的维度时,将乘累加器单元中的每个乘累加器子单元作为一个乘累加器分集群,通过乘累加器单元中的多个乘累加器子单元分别完成一个数据块的乘累加运算,并将数据块的乘累加运算结果输出到顺序排列的下一个乘累加器子单元,直至完成所有数据块的乘累加运算,输出待处理数据的乘累加运算结果。

[0044] 具体的,当待处理数据中所有数据块的计算维度均小于或等于任一乘累加器单元中乘累加器子单元的维度时,且这些数据块的第一次运算结果需要进行再一次的乘累加运算时,即按流水作业处理方式处理时,将乘累加器单元中的每个乘累加器子单元作为一个乘累加器分集群,通过乘累加器分集群对待处理数据进行流水作业方式处理,也就是将前一排列次序的乘累加器子单元输出的数据输入至后一排列次序的乘累加器子单元,这样虽然乘累加器单元的输出带宽增大,但是乘累加器子单元的输出带宽不变,不会产生堵塞现象,提高了乘累加器的运算利用率。

[0045] 在一个实施例中,如图9所示,乘累加器单元包括多个数据输入模块12、多个乘运算模块111、多个多级加运算模块112和多个数据输出模块113,其中,一个乘运算模块111、一个多级加运算模块112和多个数据输出模块113构成一个乘累加器子单元11,每个乘运算模块包括多个乘法器,每个多级加运算模块包括多个加法器;每个数据输入模块12的输入端分别与数据提供模块电连接,每个数据输入模块12的输出端分别与一个乘运算模块111的输入端电连接,每个乘运算模块111的输出端与一个多级加运算模块112的输入端电连接,每个多级加运算模块112的输出端分别与多个数据输出模块113电连接。

[0046] 具体的,乘累加器单元包括多个数据输入模块、多个乘运算模块、多个多级加运算模块和多个数据输出模块,一个乘运算模块、一个多级加运算模块和多个数据输出模块构成一个乘累加器子单元。针对每个乘累加器子单元,数据提供模块将待处理数据输入至数据输入模块的输入端,数据输入模块的输出端将数据传输至乘运算模块的输入端,乘运算模块对输入的数据进行乘运算,然后将乘运算结果输出至多级加运算模块的输入端,多级加运算模块对输入的数据进行累加运算,得到多个累加运算结果,将每个累加运算结果输出至一个数据输出模块。

[0047] 如图10所示,多个乘法器组成乘运算模块,多个加法器组成多级加运算模块,多个数据输出端C1、C2、C3、C4和D,输入数据是A1...A16,输入数据是B1...A16。乘运算模块、多级加运算模块和多个数据输出端作为一个乘累加器子单元。

[0048] 本申请还提供一种数据处理方法,应用于上述的乘累加器运算集群,数据处理方法,如图11所示,包括:

112:获取待处理数据,根据待处理数据的计算维度和每个乘累加器单元中的乘累

加器子单元的维度,确定待处理数据对应的目标乘累加器单元;

114:将待处理数据输入至目标乘累加器中处于空闲状态的乘累加器子单元中,通过目标乘累加器对待处理数据进行乘累加运算,得到待处理数据的乘累加运算结果。

[0049] 具体的,根据待处理数据的计算维度和每个乘累加器单元中的乘累加器子单元的维度,确定待处理数据对应的目标乘累加器单元,通过目标乘累加器中处于空闲状态的乘累加器子单元对待处理数据进行乘累加运算,得到待处理数据的乘累加运算结果。

[0050] 例如,当待处理数据的计算维度是8,多个乘累加器单元中乘累加器子单元的维度分别是6,9,12,16,选择乘累加器单元中每个乘累加器子单元的维度是9的乘累加器单元作为目标乘累加器单元,通过目标乘累加器单元中处于空闲状态的乘累加器子单元对待处理数据进行乘累加运算。

[0051] 当有多组待处理数据都需要进行处理时,分别确定每组待处理数据对应的目标乘累加器单元,将该组待处理数据输入至其对应的目标乘累加器单元,实现多个目标乘累加器单元对各自输入的数据进行同时处理,提高了乘累加器运算集群的数据处理效率。

[0052] 在一个实施例中,获取待处理数据,根据待处理数据的计算维度和每个乘累加器单元中的乘累加器子单元的子维度,确定待处理数据对应的目标乘累加器单元,包括:

将乘累加器单元中乘累加器子单元的维度大于或等于待处理数据的计算维度的乘累加器单元作为候选乘累加器单元;将处于空闲工作状态且乘累加器子单元的维度最大的候选乘累加器单元作为待处理数据对应的目标乘累加器单元。

[0053] 具体的,比较待处理数据的计算维度和每个乘累加器单元中乘累加器子单元的维度,将乘累加器子单元的维度大于或等于待处理数据的计算维度的乘累加器单元作为候选乘累加器单元,候选乘累加器单元有多个,从多个候选乘累加器单元中选出处于空闲状态且乘累加器子单元的维度最大的候选乘累加器单元作为目标乘累加器单元,使用目标乘累加器单元对待处理数据进行乘累加运算。

[0054] 与现有技术相比,本申请提供一种数据处理方法,根据待处理数据的计算维度和每个乘累加器单元中乘累加器子单元包括的维度,确定待处理数据对应的目标乘累加器单元,将待处理数据输入至目标乘累加器单元中的乘累加器子单元的数据输入端,由于目标乘累加器单元中的多个乘累加器子单元能够独立工作,多个乘累加器子单元能够同时进行数据处理,提高了乘累加器模块的运算利用率和总输出带宽。

[0055] 应理解的是,可以对此处申请的实施例做出各种修改。因此,上述说明书不应该视为限制,而仅是作为实施例的范例。本领域的技术人员将想到在本申请的范围和精神内的其他修改。

[0056] 包含在说明书中并构成说明书的一部分的附图示出了本申请的实施例,并且与上面给出的对本申请的大致描述以及下面给出的对实施例的详细描述一起用于解释本申请的原理。

[0057] 通过下面参照附图对给定为非限制性实例的实施例的优选形式的描述,本申请的这些和其它特性将会变得显而易见。

[0058] 还应当理解,尽管已经参照一些具体实例对本申请进行了描述,但本领域技术人员能够确定地实现本申请的很多其它等效形式。

[0059] 当结合附图时,鉴于以下详细说明,本申请的上述和其他方面、特征和优势将变得

更为显而易见。

[0060] 此后参照附图描述本申请的具体实施例；然而，应当理解，所申请的实施例仅仅是本申请的实例，其可采用多种方式实施。熟知和/或重复的功能和结构并未详细描述以避免不必要或多余的细节使得本申请模糊不清。因此，本文所申请的具体的结构性和功能性细节并非意在限定，而是仅仅作为权利要求的基础和代表性基础用于教导本领域技术人员以实质上任意合适的详细结构多样地使用本申请。

[0061] 本说明书可使用词组“在一种实施例中”、“在另一个实施例中”、“在又一实施例中”或“在其他实施例中”，其均可指代根据本申请的相同或不同实施例中的一个或多个。

[0062] 以上实施例仅为本申请的示例性实施例，不用于限制本申请，本申请的保护范围由权利要求书限定。本领域技术人员可以在本申请的实质和保护范围内，对本申请做出各种修改或等同替换，这种修改或等同替换也应视为落在本申请的保护范围内。

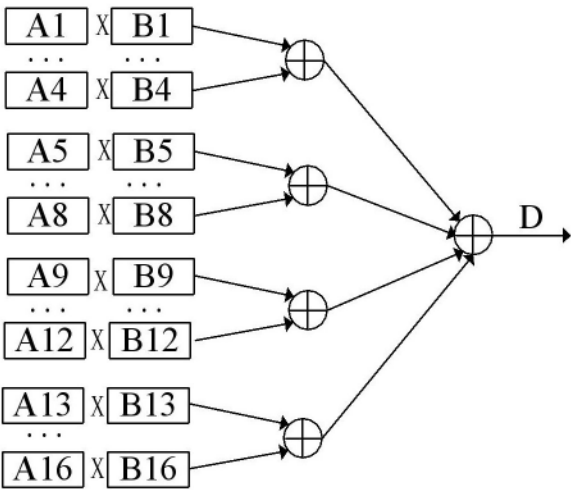


图1

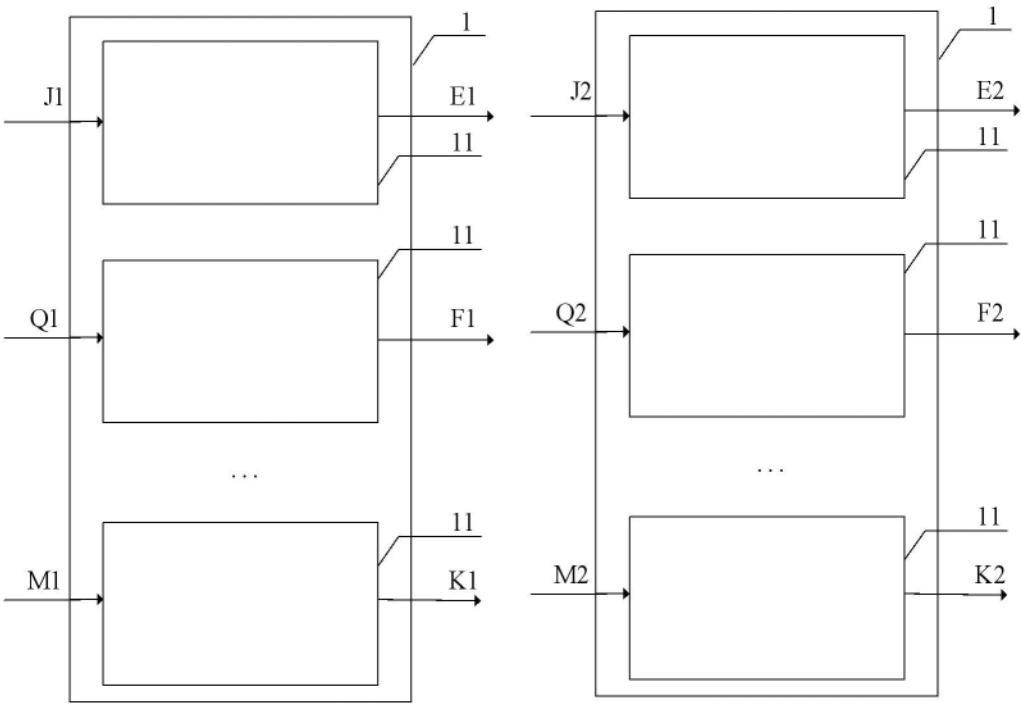


图2

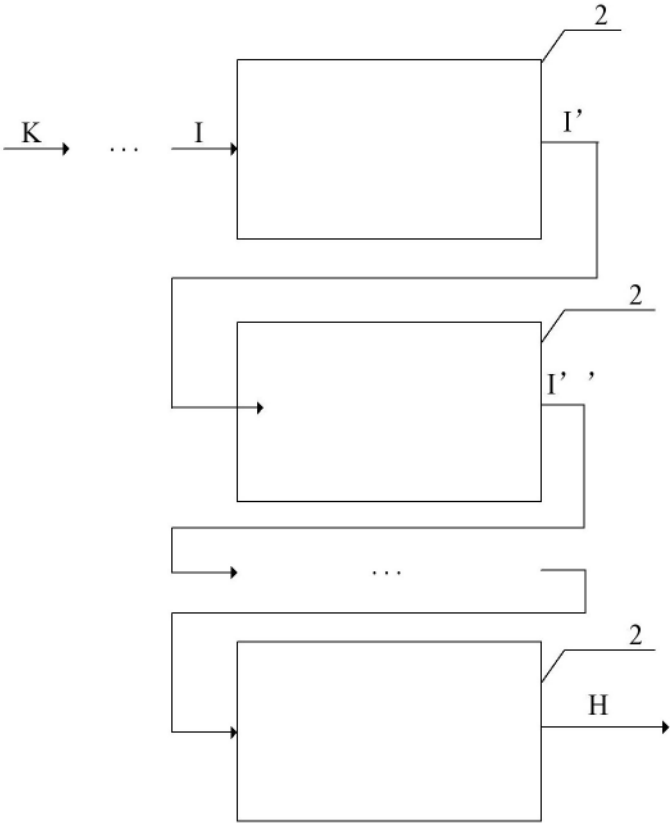


图3

只有1/4的运算单元参与计算，其余空闲

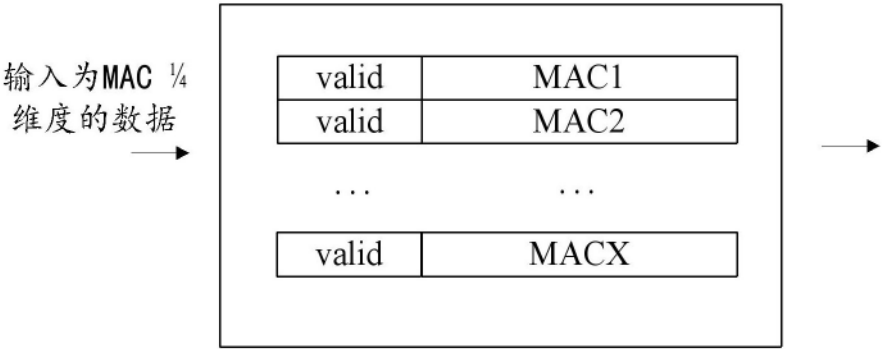


图4

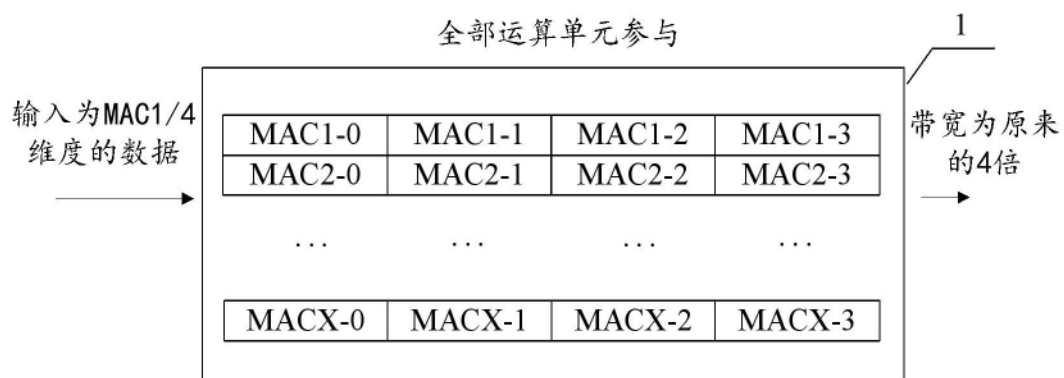


图5

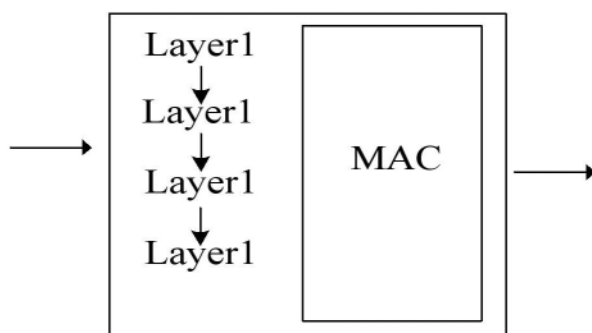


图6

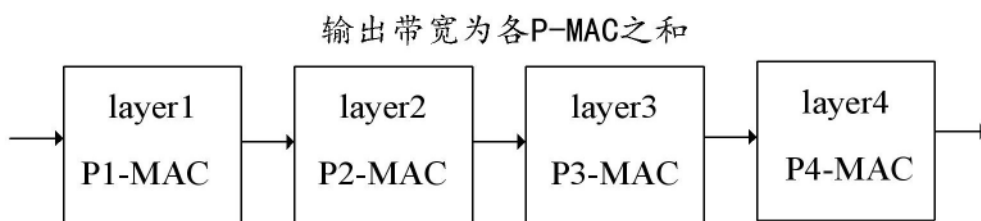


图7

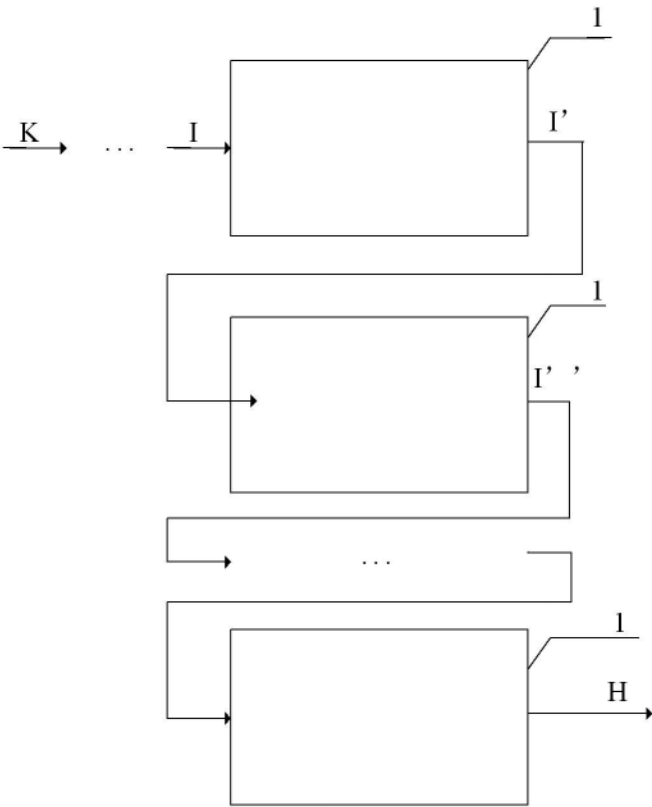


图8

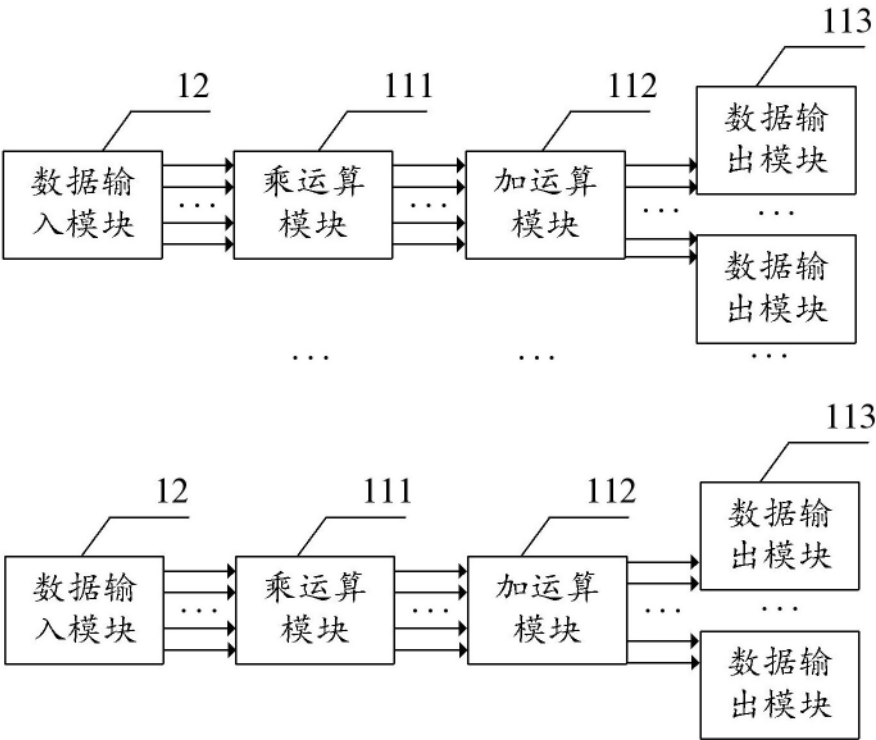


图9

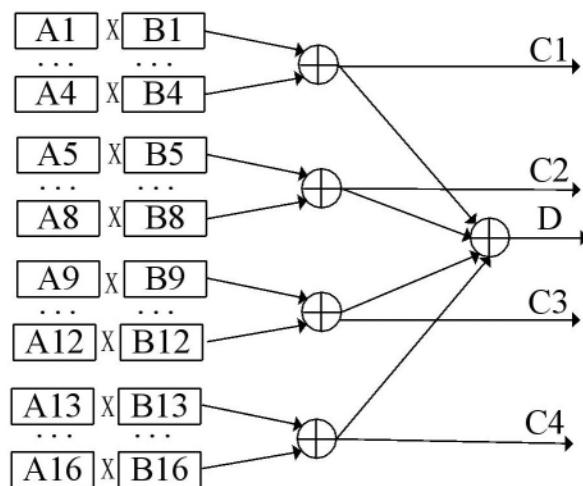


图10

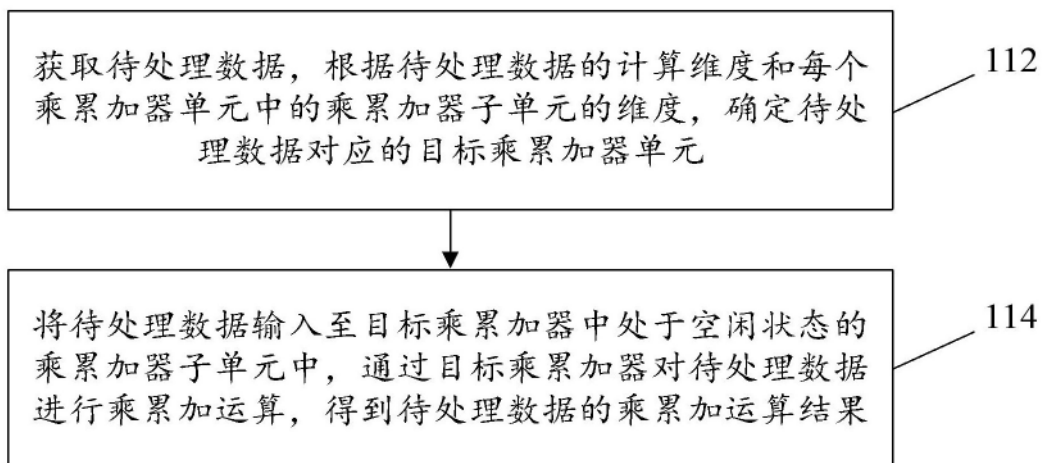


图11