(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2017 年 11 月 2 日 (02.11.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/185392 A1

(51) 国际专利分类号:
G06F 17/16 (2006.01)

(21) 国际申请号: PCT/CN2016/081107

(22) 国际申请日: 2016 年 5 月 5 日 (05.05.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610266989.X 2016年4月26日 (26.04.2016) CN

(71) 申请人: 北京中科寒武纪科技有限公司 (CAMBRICON TECHNOLOGIES CO., LTD.) [CN/CN]; 中国北京市海淀区科学院南路6号科研综合楼644室, Beijing 100190 (CN)。

(72) 发明人: 陶劲桦(TAO, Jinhua); 中国北京市海淀区科学院南路6号科研综合楼644室, Beijing 100190 (CN)。 支天(ZHI, Tian); 中国北京市海淀区科学院南路6号科研综合楼644室, Beijing 100190 (CN)。 刘少礼(LIU, Shaoli); 中国北京市海淀区科学院南路6号科研综合楼644室, Beijing 100190 (CN)。 陈天石(CHEN, Tianshi); 中国北京市海淀区科学院南路6号科研综合楼644室, Beijing 100190 (CN)。 陈云霁(CHEN, Yunji); 中国北京市海淀区科学院南路6号科研综合楼644室, Beijing 100190 (CN)。

(74) 代理人: 中科专利商标代理有限责任公司 (CHINA SCIENCE PATENT & TRADEMARK AGENT LTD.); 中国北京市海淀区西三环北路87号4-1105室, Beijing 100089 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA,

(54) **Title:** DEVICE AND METHOD FOR PERFORMING FOUR FUNDAMENTAL OPERATIONS OF ARITHMETIC OF VECTORS

(54) **发明名称**: 一种用于执行向量四则运算的装置和方法

图 1

AA Register unit
BB Unit for four fundamental operations of arithmetic of vectors
CC Storage unit

(57) **Abstract:** A device and method for performing four fundamental operations of arithmetic of vectors, used for cooperating with a corresponding instruction set to perform four fundamental operations of arithmetic of vectors. The device comprises a storage unit, a register unit and a unit for four fundamental operations of arithmetic of vectors; there are vectors storing in the storage unit; there are addresses of the vectors storing in the register unit; the unit for four fundamental operations of arithmetic of vectors obtains vector addresses from the register unit according to matched instructions, then obtains the corresponding vectors from the storage unit according to the vector addresses, and performs the four fundamental operations of arithmetic of vectors according to the obtained vectors to obtain operation results. By temporarily storing the vector data involved in the operation in a cache, data with different width the can be supported during the process of four fundamental operations of arithmetic of vectors and the operation performance of an application which comprises a great deal of four fundamental operations of arithmetic is improved.

(57) **摘要**: 一种执行向量四则运算的装置及方法，用于配合一套相应的指令集，执行向量四则运算，装置包括存储单元、寄存器单元和向量四则运算单元，存储单元中存储有向量，寄存器单元中存储有向量存储的地址，向量四则运算单元根据配套指令在寄存器单元中获取向量地址，然后，根据该向量地址在存储单元中获取相应的向量，接着，根据获取的向量进行向量四则运算，得到运算结果。通过将参与计算的向量数据暂存在高速暂存存储器上，使得向量四则运算过程中可以更加灵活有效地支持不同宽度的数据，提升包含大量向量四则运算应用的执行性能。

WO 2017/185392 A1

[见续页]

MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

一种用于执行向量四则运算的装置和方法

技术领域

本发明涉及一种向量四则运算装置及方法，用于根据向量四则运算指令高效灵活地执行向量四则运算，能够很好地解决当前计算机领域越来越多的算法包含大量向量四则运算的问题。

背景技术

在已有的计算机领域应用中，与向量运算相关的应用十分普遍。以目前的热门应用领域人工智能中的主流算法机器学习算法为例，几乎所有已有的经典算法中都含有大量的向量四则运算。向量四则运算是指对向量的对应分量进行加减乘除这四种运算。具体来说，对于两个向量 a = [a1, a2,…, an]和 b = [b1, b2,…, bn]，向量加法定义为：a+b=[a1+b1, a2+b2,…, an+bn]，向量减法定义为：a-b=[a1-b1, a2-b2,…, an-bn]，向量乘法定义为：[a1*b1, a2*b2,…, an*bn]向量除法定义为：[a1/b1, a2/b2,…, an/bn]。

在现有技术中，一种进行向量四则运算的已知方案是使用通用处理器，该方法通过通用寄存器堆和通用功能部件来执行通用指令，从而执行向量四则运算。然而，该方法的缺点之一是单个通用处理器多用于标量计算，在进行向量四则运算时运算性能较低。而使用多个通用处理器并行执行时，通用处理器之间的相互通讯又有可能成为性能瓶颈。在另一种现有技术中，使用图形处理器（GPU）来进行向量计算，其中，通过使用通用寄存器堆和通用流处理单元执行通用 SIMD 指令来进行向量四则运算。然而，上述方案中，GPU 片上缓存太小，在进行大规模向量四则运算时需要不断进行片外数据搬运，片外带宽成为了主要性能瓶颈。在另一种现有技术中，使用专门定制的向量四则运算装置来进行向量计算，其中，使用定制的寄存器堆和定制的处理单元进行向量四则运

算。然而，目前已有的专用向量四则运算装置受限于寄存器堆，不能够灵活地支持不同长度的向量四则运算。

发明内容

（一）要解决的技术问题

本发明的目的在于，提供一种向量四则运算装置及方法，解决现有技术中存在的受限于片间通讯、片上缓存不够、支持的向量长度不够灵活等问题。

（二）技术方案

本发明提供一种向量四则运算装置，用于根据向量四则运算指令执行向量四则运算，包括：

存储单元，用于存储向量；

寄存器单元，用于存储向量地址，其中，向量地址为向量在存储单元中存储的地址；

向量四则运算单元，用于获取向量四则运算指令，根据向量四则运算指令在寄存器单元中获取向量地址，然后，根据该向量地址在存储单元中获取相应的向量，接着，根据获取的向量进行向量四则运算，得到向量四则运算结果。

（三）有益效果

本发明提供的向量四则运算装置及方法，将参与计算的向量数据暂存在高速暂存存储器（Scratchpad Memory）上。在仅发送同一条指令的情况下，向量四则运算单元中可以更加灵活有效地支持不同宽度的数据，并可以解决数据存储中的相关性问题，从而提升了包含大量向量计算任务的执行性能，本发明采用的指令具有精简的格式，使得指令集使用方便、支持的向量长度灵活。

本发明可以应用于以下（包括但不限于）场景中：数据处理、机器人、电脑、打印机、扫描仪、电话、平板电脑、智能终端、手机、行车记录仪、导航仪、传感器、摄像头、云端服务器、相机、摄像机、投影

仪、手表、耳机、移动存储、可穿戴设备等各类电子产品；飞机、轮船、车辆等各类交通工具；电视、空调、微波炉、冰箱、电饭煲、加湿器、洗衣机、电灯、燃气灶、油烟机等各类家用电器；以及包括核磁共振仪、B 超、心电图仪等各类医疗设备。

附图说明

图 1 是本发明提供的向量四则运算装置的结构示意图。

图 2 是本发明提供的指令集的格式示意图。

图 3 是本发明实施例提供的向量四则运算装置的结构示意图。

图 4 是本发明实施例提供的向量四则运算装置执行向量四则指令的流程图。

图 5 为本发明实施例提供的向量四则运算单元的结构示意图。

具体实施方式

本发明提供一种向量四则运算装置及配套指令集，装置包括存储单元、寄存器单元和向量四则运算单元，存储单元中存储有向量，寄存器单元中存储有向量存储的地址向量四则运算单元根据向量四则运算指令在寄存器单元中获取向量地址，然后，根据该向量地址在存储单元中获取相应的向量，接着，根据获取的向量进行向量四则运算，得到向量四则运算结果。本发明将参与计算的向量数据暂存在高速暂存存储器上，使得向量四则运算过程中可以更加灵活有效地支持不同宽度的数据，提升包含大量向量计算任务的执行性能。

图 1 是本发明提供的向量四则运算装置的结构示意图，如图 1 所示，向量四则运算装置包括：

存储单元，用于存储向量，在一种实施方式中，该存储单元可以是高速暂存存储器，能够支持不同大小的向量数据；本发明将必要的计算数据暂存在高速暂存存储器（Scratchpad Memory）上，使本运算装置在

进行向量四则运算过程中可以更加灵活有效地支持不同宽度的数据。存储单元可以通过各种不同存储器件（SRAM、eDRAM、DRAM、忆阻器、3D-DRAM 或非易失存储等）实现。

寄存器单元，用于存储向量地址，其中，向量地址为向量在存储单元中存储的地址；在一种实施方式中，寄存器单元可以是标量寄存器堆，提供运算过程中所需的标量寄存器，标量寄存器不只存放向量地址，还存放有标量数据。当涉及到向量与标量的运算时，向量四则运算单元不仅要从寄存器单元中获取向量地址，还要从寄存器单元中获取相应的标量。另外，寄存器单元的数量一般为多个，以组成寄存器堆，用于存储多个向量地址及标量。

向量四则运算单元，用于获取向量四则运算指令，根据向量四则运算指令在寄存器单元中获取向量地址，然后，根据该向量地址在存储单元中获取相应的向量，接着，根据获取的向量进行向量四则运算，得到向量四则运算结果，并将向量四则运算结果存储于存储单元中。向量四则运算单元包含包括向量四则加法部件、向量四则减法部件、向量四则乘法部件和向量四则除法部件，并且，向量四则运算单元为多流水级结构，其中，加法部件和减法部件处于第一流水级，乘法部件和除法部件处于第二流水级。这些单元处于不同的流水级，当连续串行的多条向量四则运算指令的先后次序与相应单元所在流水级顺序一致时，可以更加高效地实现这一连串向量四则运算指令所要求的操作。

根据本发明的一种实施方式，向量四则运算装置还包括：指令缓存单元，用于存储待执行的向量四则运算指令。指令在执行过程中，同时也被缓存在指令缓存单元中，当一条指令执行完之后，如果该指令同时也是指令缓存单元中未被提交指令中最早的一条指令，该指令将被提交，一旦提交，该条指令进行的操作对装置状态的改变将无法撤销。在一种实施方式中，指令缓存单元可以是重排序缓存。

根据本发明的一种实施方式，向量四则运算装置还包括：指令处理单元，用于从指令缓存单元获取向量四则运算指令，并对该向量四则运算指令进行处理后，提供给所述向量四则运算单元。其中，指令处理单

元包括：

取指模块，用于从指令缓存单元中获取向量四则运算指令；

译码模块，用于对获取的向量四则运算指令进行译码；

指令队列，用于对译码后的向量四则运算指令进行顺序存储，考虑到不同指令在包含的寄存器上有可能存在依赖关系，用于缓存译码后的指令，当依赖关系被满足之后发送指令。

根据本发明的一种实施方式，向量四则运算装置还包括：依赖关系处理单元，用于在向量四则运算单元获取向量四则运算指令前，判断该向量四则运算指令与前一向量四则运算指令是否访问相同的向量，若是，将该向量四则运算指令存储在一存储队列中，待前一向量四则运算指令执行完毕后，将存储队列中的该向量四则运算指令提供给所述向量四则运算单元；否则，直接将该向量四则运算指令提供给所述向量四则运算单元。具体地，向量四则运算指令访问高速暂存存储器时，前后指令可能会访问同一块存储空间，为了保证指令执行结果的正确性，当前指令如果被检测到与之前的指令的数据存在依赖关系，该指令必须在存储队列内等待至依赖关系被消除。

根据本发明的一种实施方式，向量四则运算装置还包括：输入输出单元，用于将向量存储于存储单元，或者，从存储单元中获取向量四则运算结果。其中，输入输出单元可直接存储单元，负责从内存中读取向量数据或写入向量数据。

本发明还提供一种向量四则运算方法，用于根据向量四则运算指令执行向量四则运算，方法包括：

S1，存储向量；

S2，存储向量地址，向量地址指示了向量在步骤 S1 中所存储的位置；

S3，获取向量四则运算指令，根据向量四则运算指令获取向量地址，然后，根据该向量地址获取存储的向量，接着，根据获取的向量进行向量四则运算，得到向量四则运算结果。

根据本发明的一种实施方式，在步骤 S3 之前还包括：

存储向量四则运算指令；

获取存储的向量四则运算指令；

对获取的向量四则运算指令进行译码；

对译码后的向量四则运算指令进行顺序存储。

根据本发明的一种实施方式，在步骤 S3 之前还包括：

判断该向量四则运算指令与前一向量四则运算指令是否访问相同的向量，若是，将该向量四则运算指令存储在一存储队列中，待前一向量四则运算指令执行完毕后，再执行步骤 S3；否则，直接执行步骤 S3。

根据本发明的一种实施方式，方法还包括，存储所述向量四则运算结果。

根据本发明的一种实施方式，步骤 S1 包括，将向量存储至一高速暂存存储器中。

根据本发明的一种实施方式，向量四则运算指令包括一操作码和至少一操作域，其中，所述操作码用于指示该向量运算指令的功能，操作域用于指示该向量运算指令的数据信息。

根据本发明的一种实施方式，向量四则运算包括向量加法运算、向量减法运算、向量乘法运算和向量除法运算。

根据本发明的一种实施方式，向量运算单元为多流水级结构，包括第一流水级和第二流水级，其中，在第一流水级执行向量加法运算和向量减法运算，在第二流水级执行向量乘法运算和向量除法运算。

本发明的指令集采用 Load/Store 结构，向量四则运算单元不会对内存中的数据进行操作。本指令集采用精简指令集架构，指令集只提供最基本的向量四则运算操作，复杂的向量四则运算都由这些简单指令通过组合进行模拟，使得可以在高时钟频率下单周期执行指令。另外，本指令集同时采用定长指令，使得本发明提出的向量四则运算装置在上一条指令的译码阶段对下一条指令进行取指。

在本装置执行向量四则运算的过程中，装置取出指令进行译码，然后送至指令队列存储，根据译码结果，获取指令中的各个参数，这些参数可以是直接写在指令的操作域中，也可以是根据指令操作域中的寄存

器编号从指定的寄存器中读取。这种使用寄存器存储参数的好处是无需改变指令本身，只要用指令改变寄存器中的值，就可以实现大部分的循环，因此大大节省了在解决某些实际问题时所需要的指令条数。在全部操作数之后，依赖关系处理单元会判断指令实际需要使用的数据与之前指令中是否存在依赖关系，这决定了这条指令是否可以被立即发送至运算单元中执行。一旦发现与之前的数据之间存在依赖关系，则该条指令必须等到它依赖的指令执行完毕之后才可以送至运算单元执行。在定制的运算单元中，该条指令将快速执行完毕，并将结果，即生成的向量四则运算结果写回至指令提供的地址，该条指令执行完毕。

图 2 是本发明提供的指令集的格式示意图，如图 2 所示，向量四则运算指令包括 1 个操作码和多个操作域，其中，操作码用于指示该向量四则运算指令的功能，功能如加、减、乘、除等。向量四则运算单元通过识别该操作码可进行向量四则运算，操作域用于指示该向量四则运算指令的数据信息，其中，数据信息可以是立即数或寄存器编号，例如，要获取一个向量时，根据寄存器编号可以在相应的寄存器中获取向量起始地址和向量长度，再根据向量起始地址和向量长度在存储单元中获取相应地址存放的向量。

指令集包含有不同功能的向量四则运算指令：

向量加法指令（VA）。根据该指令，装置从高速暂存存储器的指定地址处分别取出两块指定大小的向量数据，在向量运算单元中进行加法运算，并将结果写回至高速暂存存储器的指定地址；

向量加标量指令（VAS）。根据该指令，装置从高速暂存存储器的指定地址取出指定大小的向量数据，从标量寄存器堆的指定地址取出标量数据，在标量运算单元中将向量的每一个元素加上该标量值，并将结果写回至高速暂存存储器的指定地址；

向量减法指令（VS）。根据该指令，装置从高速暂存存储器的指定地址处分别取出两块指定大小的向量数据，在向量运算单元中进行减法运算，并将结果写回至高速暂存存储器的指定地址；

标量减向量指令（SSV）。根据该指令，装置从标量寄存器堆的指定

地址取出标量数据，从高速暂存存储器的指定地址取出向量数据，在向量计算单元中用该标量减去向量中的相应元素，并将结果写回高速暂存存储器的指定地址；

向量乘法指令（VMV）。根据该指令，装置从高速暂存存储器的指定地址分别取出指定大小的向量数据，在向量计算单元中将两向量数据对位相乘，并将结果写回高速暂存存储器的指定地址；

向量乘标量指令（VMS）。根据该指令，装置从高速暂存存储器的指定地址取出指定大小的向量数据，从标量寄存器堆的指定地址取出指定大小的标量数据，在向量寄存单元中进行向量乘标量运算，并将结果写回高速暂存存储器的指定地址；

向量除法指令（VD）。根据该指令，装置从高速暂存存储器的指定地址取出分别取出指定大小的向量数据，在向量运算单元中将两向量对位相除，并将结果写回至高速暂存存储器的指定地址；

标量除向量指令（SDV）。根据该指令，装置从标量寄存器堆的指定位置取出标量数据，从高速暂存存储器的指定位置取出指定大小的向量数据，在向量计算单元中用标量分别除以向量中的相应元素，并将结果写回至高速暂存存储器的指定位置；

向量检索指令（VR）。根据该指令，装置从高速暂存存储器的指定地址取出指定大小的向量数据，在向量计算单元中根据指定位置取出向量中的相应元素作为输出，并将结果写回至标量寄存器堆的指定地址；

向量加载指令（VLOAD）。根据该指令，装置从指定外部源地址载入指定大小的向量数据至高速暂存存储器的指定地址；

向量存储指令（VS）。根据该指令，装置将高速暂存存储器的指定地址的指定大小的向量数据存至外部目的地址处；

向量搬运指令（VMOVE）。根据该指令，装置将高速暂存存储器的指定地址的指定大小的向量数据存至高速暂存存储器的另一指定地址处。

为使本发明的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本发明进一步详细说明。

图 3 是本发明实施例提供的向量四则运算装置的结构示意图，如图 3 所示，装置包括取指模块、译码模块、指令队列、标量寄存器堆、依赖关系处理单元、存储队列、重排序缓存、向量四则运算单元、高速暂存器、IO 内存存取模块；

取指模块，该模块负责从指令序列中取出下一条将要执行的指令，并将该指令传给译码模块；

译码模块，该模块负责对指令进行译码，并将译码后指令传给指令队列；

指令队列，考虑到不同指令在包含的标量寄存器上有可能存在依赖关系，用于缓存译码后的指令，当依赖关系被满足之后发射指令；

标量寄存器堆，提供装置在运算过程中所需的标量寄存器；

依赖关系处理单元，该模块处理处理指令与前一条指令可能存在的存储依赖关系。向量四则运算指令会访问高速暂存存储器，前后指令可能会访问同一块存储空间。为了保证指令执行结果的正确性，当前指令如果被检测到与之前的指令的数据存在依赖关系，该指令必须在存储队列内等待至依赖关系被消除。

存储队列，该模块是一个有序队列，与之前指令在数据上有依赖关系的指令被存储在该队列内直至存储关系被消除；

重排序缓存，指令在执行过程中，同时也被缓存在给该模块中，当一条指令执行完之后，如果该指令同时也是重排序缓存中未被提交指令中最早的一条指令，该指令将被提交。一旦提交，该条指令进行的操作对装置状态的改变将无法撤销；

向量四则运算单元，该模块负责装置的所有向量四则运算，向量四则运算指令被送往该运算单元执行；

高速暂存器，该模块是向量数据专用的暂存存储装置，能够支持不同大小的向量数据；

IO 内存存取模块，该模块用于直接访问高速暂存存储器，负责从高速暂存存储器中读取数据或写入数据。

图 4 是本发明实施例提供的向量四则运算装置执行任一向量加法指

令（VA）的流程图，如图 4 所示，执行向量加法指令（VA）的过程包括：

S1，取指模块取出该条向量加法指令（VA），并将该指令送往译码模块。

S2，译码模块对指令译码，并将向量加法指令（VA）送往指令队列。

S3，在指令队列中，该向量加法指令（VA）需要从标量寄存器堆中获取指令中四个操作域所对应的标量寄存器里的数据，包括向量 vin0 的起始地址、向量 vin0 的长度、向量 vin1 的起始地址、向量 vin1 的长度。

S4，在取得需要的标量数据后，该指令被送往依赖关系处理单元。依赖关系处理单元分析该指令与前面的尚未执行结束的指令在数据上是否存在依赖关系。该条指令需要在存储队列中等待至其与前面的未执行结束的指令在数据上不再存在依赖关系为止。

S5：依赖关系不存在后，该条向量加法指令（VA）被送往向量四则运算单元。向量四则运算单元根据所需数据的地址和长度从数据暂存器中取出需要的向量，然后在向量四则运算单元中完成向量加法运算。

S6，运算完成后，将结果写回至高速暂存存储器的指定地址，同时提交重排序缓存中的该向量四则指令。

图 5 为本发明实施例提供的向量四则运算单元的结构示意图，如图 5 所示，向量四则运算单元内包含向量四则运算单元等。并且，向量四则运算单元为多流水级结构，

其中，向量加法部件和向量减法部件处于流水级 1，向量四则乘法部件和向量四则除法部件处于流水级 2。这些单元处于不同的流水级，当连续串行的多条向量四则运算指令的先后次序与相应单元所在流水级顺序一致时，可以更加高效地实现这一连串向量四则运算指令所要求的操作。

以上所述的具体实施例，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施例而已，并不用于限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权利要求

1、一种向量四则运算装置，用于根据向量四则运算指令执行向量四则运算，包括：

存储单元，用于存储向量；

寄存器单元，用于存储向量地址，其中，向量地址为向量在存储单元中存储的地址；

向量四则运算单元，用于获取向量四则运算指令，根据向量四则运算指令在寄存器单元中获取向量地址，然后，根据该向量地址在存储单元中获取相应的向量，接着，根据获取的向量进行向量四则运算，得到向量四则运算结果。

2、根据权利要求 1 所述的向量四则运算装置，其特征在于，还包括：指令缓存单元，用于存储待执行的向量四则运算指令。

3、根据权利要求 2 所述的向量四则运算装置，其特征在于，还包括：指令处理单元，用于从指令缓存单元获取向量四则运算指令，并对该向量四则运算指令进行处理后，提供给所述向量四则运算单元。

4、根据权利要求 3 所述的向量四则运算装置，其特征在于，所述指令处理单元包括：

取指模块，用于从指令缓存单元中获取向量四则运算指令；

译码模块，用于对获取的向量四则运算指令进行译码；

指令队列，用于对译码后的向量四则运算指令进行顺序存储。

5、根据权利要求 1 所述的向量四则运算装置，其特征在于，还包括：

依赖关系处理单元，用于在所述向量四则运算单元获取向量四则运算指令前，用于在向量四则运算单元获取向量四则运算指令前，判断该向量四则运算指令与前一向量四则运算指令是否访问相同的向量，若是，将该向量四则运算指令存储在一存储队列中，待前一向量四则运算指令执行完毕后，将存储队列中的该向量四则运算指令提供给所述向量四则运算单元；否则，直接将该向量四则运算指令提供给所述向量四则

运算单元。

6、根据权利要求 1 所述的向量四则运算装置，其特征在于，所述存储单元还用于存储所述向量四则运算结果。

7、根据权利要求 6 所述的向量四则运算装置，其特征在于，还包括：

输入输出单元，用于将向量存储于所述存储单元，或者，从所述存储单元中获取向量四则运算结果。

8、根据权利要求 1 所述的向量四则运算装置，其特征在于，所述存储单元为高速暂存存储器。

9、根据权利要求 1 所述的向量运算装置，其特征在于，所述向量四则运算指令包括一操作码和至少一操作域，其中，所述操作码用于指示该向量运算指令的功能，操作域用于指示该向量运算指令的数据信息。

10、根据权利要求 9 所述的向量运算装置，其特征在于，所述数据信息为寄存器单元编号，所述向量四则运算单元根据该寄存器单元编号访问对应的寄存器单元，并获取向量地址。

11、根据权利要求 1 所述的向量四则运算装置，其特征在于，所述向量四则运算单元包括向量加法部件、向量减法部件、向量乘法部件和向量除法部件。

12、根据权利要求 11 所述的向量运算装置，其特征在于，所述向量运算单元为多流水级结构，包括第一流水级和第二流水级，其中，向量加法部件和向量减法部件处于第一流水级，向量乘法部件和向量除法部件处于第二流水级。

13、一种向量四则运算方法，用于根据向量四则运算指令执行向量四则运算，方法包括：

S1，存储向量；

S2，存储向量地址；

S3，获取向量四则运算指令，根据向量四则运算指令获取向量地址，然后，根据该向量地址获取存储的向量，接着，根据获取的向量进行向

量四则运算，得到向量四则运算结果。

14、根据权利要求 13 所述的向量四则运算方法，其特征在于，在步骤 S3 之前还包括：

存储向量四则运算指令；

获取存储的向量四则运算指令；

对获取的向量四则运算指令进行译码；

对译码后的向量四则运算指令进行顺序存储。

15、根据权利要求 13 所述的向量四则运算方法，其特征在于，在步骤 S3 之前还包括：

判断该向量四则运算指令与前一向量四则运算指令是否访问相同的向量，若是，将该向量四则运算指令存储在一存储队列中，待前一向量四则运算指令执行完毕后，再执行步骤 S3；否则，直接执行步骤 S3。

16、根据权利要求 13 所述的向量四则运算方法，其特征在于，还包括，存储所述向量四则运算结果。

17，根据权利要求 13 所述的向量四则运算方法，其特征在于，所述步骤 S1 包括，将向量存储至一高速暂存存储器中。

18、根据权利要求 13 所述的向量四则运算方法，其特征在于，所述向量四则运算指令包括一操作码和至少一操作域，其中，所述操作码用于指示该向量运算指令的功能，操作域用于指示该向量运算指令的数据信息。

19、根据权利要求 13 所述的向量四则运算方法，其特征在于，所述向量四则运算包括向量加法运算、向量减法运算、向量乘法运算和向量除法运算。

20、根据权利要求 19 所述的向量四则运算方法，其特征在于，所述向量运算单元为多流水级结构，包括第一流水级和第二流水级，其中，在第一流水级执行向量加法运算和向量减法运算，在第二流水级执行向量乘法运算和向量除法运算。

图 1

操作码	寄存器 0	寄存器 1	寄存器 2	寄存器 3	寄存器 4
VA	向量 0 起始地址	向量 0 长度	向量 1 起始地址	向量 1 长度	输出向量地址
VAS	向量 0 起始地址	向量 0 长度	输出向量地址	输入标量	
VS	向量 0 起始地址	向量 0 长度	向量 1 起始地址	向量 1 长度	输出向量地址
SSVS	向量 0 起始地址	向量 0 长度	输出向量地址	输入标量	
VMV	向量 0 起始地址	向量 0 长度	向量 1 起始地址	向量 1 长度	输出向量地址
VMS	向量 0 起始地址	向量 0 长度	输出向量地址	输入标量	
VDV	向量 0 起始地址	向量 0 长度	向量 1 起始地址	向量 1 长度	输出向量地址
SDV	向量 0 起始地址	向量 0 长度	输出向量地址	输入标量	
VR	向量输入地址	位置参数	标量输出	--	--
VLOAD	输入地址	矩阵大小	输出地址	--	--
VS	向量 0 起始地址	向量 0 长度	向量 1 起始地址	向量 1 长度	输出向量地址
VMOVE	输入地址	矩阵大小	输出地址		

图 2

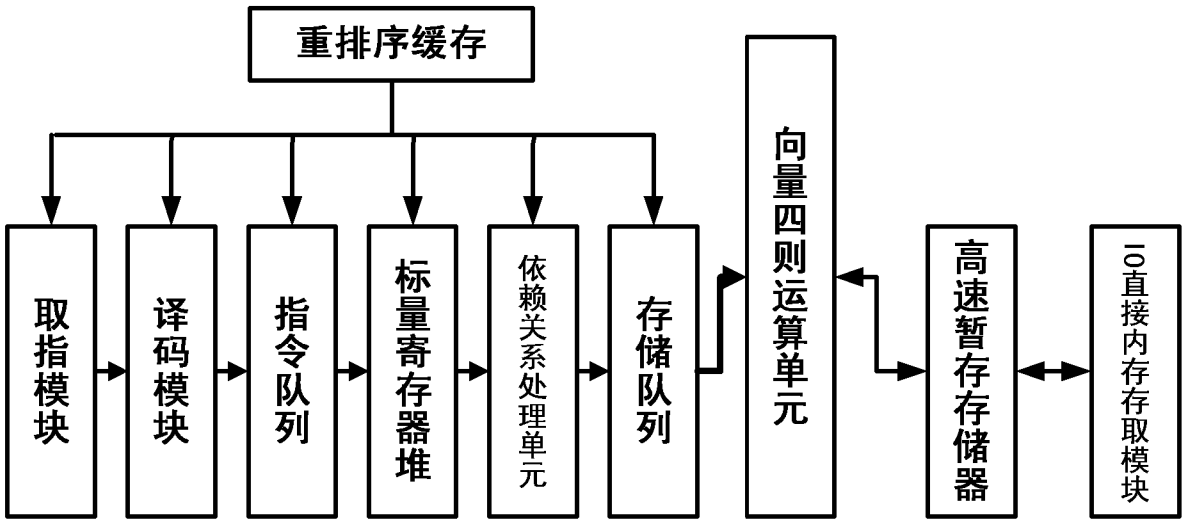
重排序缓存
取指模块
译码模块
指令队列
标量寄存器堆
依赖关系处理单元
存储队列
向量四则运算单元
高速暂存存储器
I/O直接内存存取模块

图 3

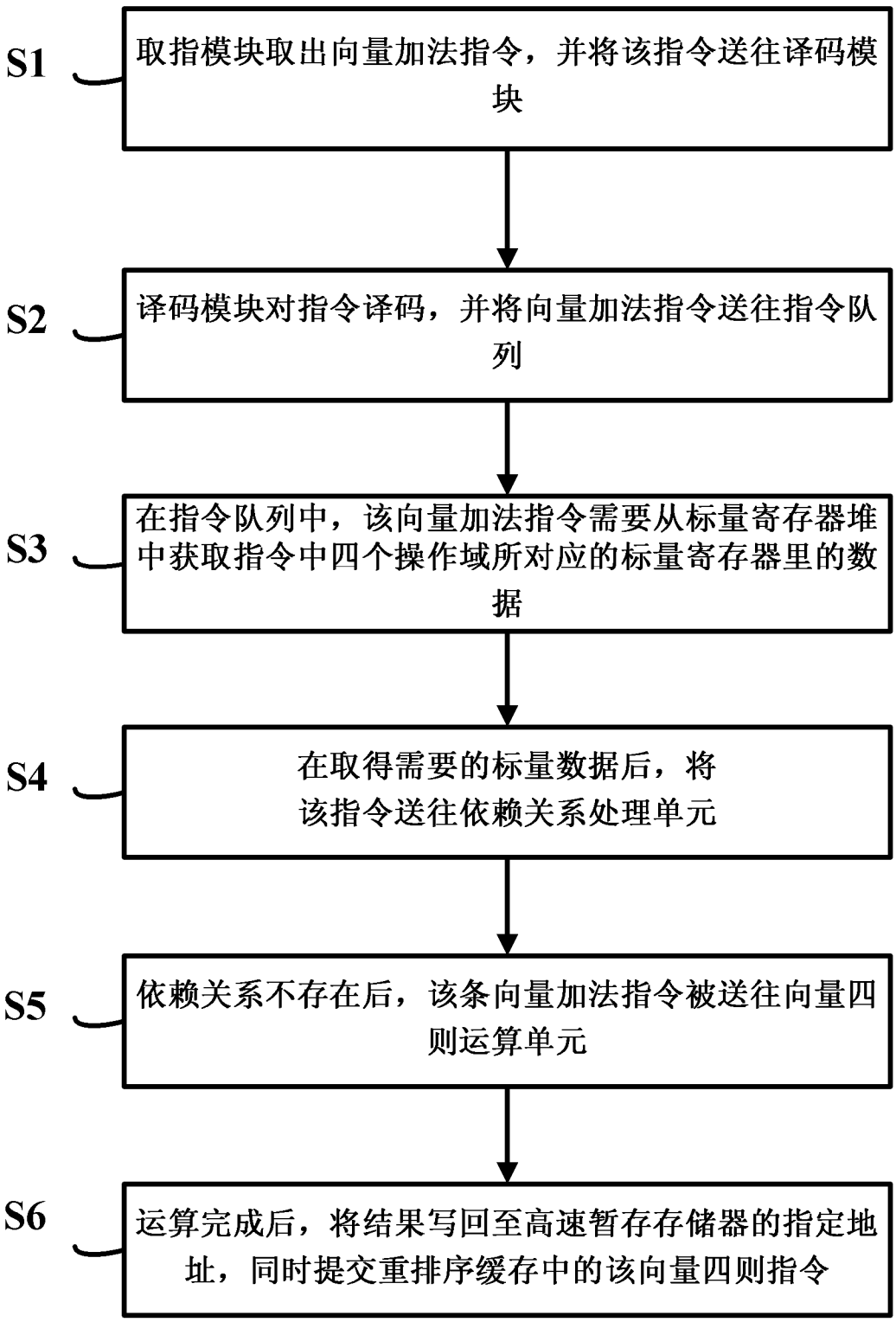
S1
取指模块取出向量加法指令，并将该指令送往译码模块
S2
译码模块对指令译码，并将向量加法指令送往指令队列
S3
在指令队列中，该向量加法指令需要从标量寄存器堆中获取指令中四个操作域所对应的标量寄存器里的数据
S4
在取得需要的标量数据后，将该指令送往依赖关系处理单元
S5
依赖关系不存在后，该条向量加法指令被送往向量四则运算单元
S6
运算完成后，将结果写回至高速暂存存储器的指定地址，同时提交重排序缓存中的该向量四则指令

图 4

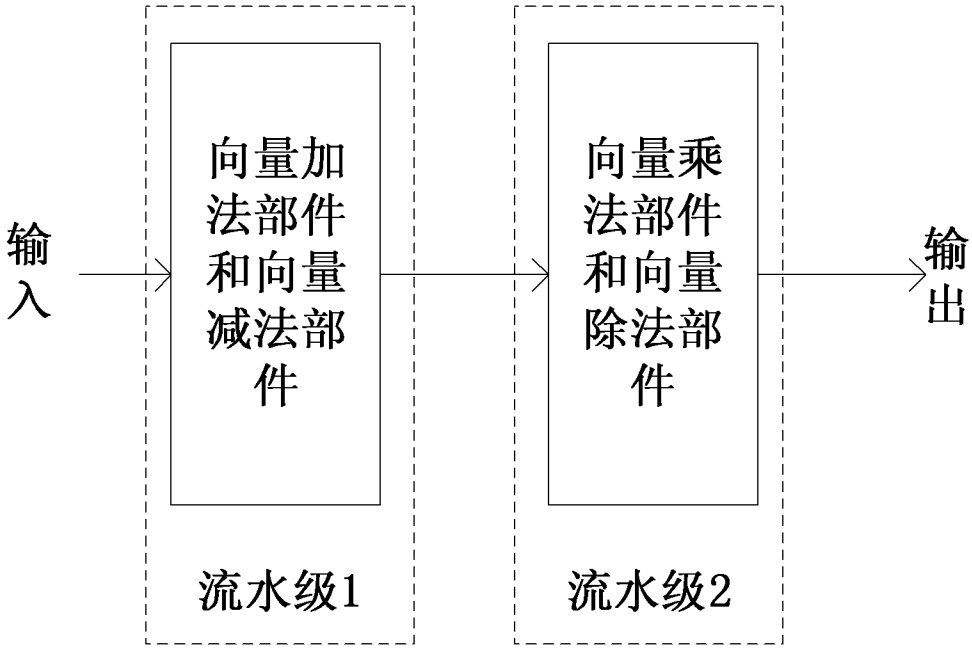

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/081107

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/16 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNKI, CNPAT, IEEE, GOOGLE: store, compute, vector, instruction, operation, add, subtration, multiplication, division, register, arithmetic, buffer, queue

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103699360 A (BEIJING ZHONGKE JINGSHANG TECHNOLOGY CO., LTD.), 02 April 2014 (02.04.2014), description, paragraphs [0073]-[0093]	1-20
A	CN 102629238 A (NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY), 08 August 2012 (08.08.2012), the whole document	1-20
A	US 2003037221 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION), 20 February 2003 (20.02.2003), the whole document	1-20
A	CN 104407997 A (NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY), 11 March 2015 (11.03.2015), the whole document	1-20
A	CN 1349159 A (NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY), 15 May 2002 (15.05.2002), the whole document	1-20
A	CN 101847093 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES), 29 September 2010 (29.09.2010), the whole document	1-20

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 January 2017 (09.01.2017)	**25 January 2017 (25.01.2017)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **MA, Chunli** Telephone No.: (86-10) **62413712**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/081107

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103699360 A	02 April 2014	None	
CN 102629238 A	08 August 2012	None	
US 2003037221 A1	20 February 2003	None	
CN 104407997 A	11 March 2015	None	
CN 1349159 A	15 May 2002	None	
CN 101847093 A	29 September 2010	None	

国际检索报告

国际申请号
PCT/CN2016/081107

A. 主题的分类

G06F 17/16(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

WPI, EPODOC, CNKI, CNPAT, IEEE, GOOGLE:向量，指令，存储，寄存器，运算，计算，加，减，乘，除，四则运算，缓存，队列，vector，instruction，operation，add，subtration，multiplication，division，register，arithmetic，buffer，queue，

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 103699360 A (北京中科晶上科技有限公司) 2014年 4月 2日 (2014-04-02) 说明书第[0073]-[0093]段	1-20
A	CN 102629238 A (中国人民解放军国防科学技术大学) 2012年 8月 8日 (2012-08-08) 全文	1-20
A	US 2003037221 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2003年 2月 20日 (2003-02-20) 全文	1-20
A	CN 104407997 A (中国人民解放军国防科学技术大学) 2015年 3月 11日 (2015-03-11) 全文	1-20
A	CN 1349159 A (中国人民解放军国防科学技术大学) 2002年 5月 15日 (2002-05-15) 全文	1-20
A	CN 101847093 A (中国科学院自动化研究所) 2010年 9月 29日 (2010-09-29) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2017年 1月 9日	2017年 1月 25日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	马春黎
传真号 (86-10)62019451	电话号码 (86-10)62413712

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/081107

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 103699360 A	2014年 4月 2日	无	
CN 102629238 A	2012年 8月 8日	无	
US 2003037221 A1	2003年 2月 20日	无	
CN 104407997 A	2015年 3月 11日	无	
CN 1349159 A	2002年 5月 15日	无	
CN 101847093 A	2010年 9月 29日	无	