

(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
2018年3月1日 (01.03.2018)



(10) 国际公布号
WO 2018/036364 A1

- (51) 国际专利分类号:
G06F 12/1027 (2016.01)
- (21) 国际申请号: PCT/CN2017/095845
- (22) 国际申请日: 2017年8月3日 (03.08.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610738487.2 2016年8月26日 (26.08.2016) CN
- (71) 申请人: 上海寒武纪信息科技有限公司 (SHANGHAI CAMBRICON INFORMATION TECHNOLOGIES LTD.) [CN/CN]; 中国上海市浦东新区祖冲之路2290号1号楼1004, Shanghai 201203 (CN)。
- (72) 发明人: 陈天石 (CHEN, Tianshi); 中国上海市浦东新区祖冲之路2290号1号楼1004, Shanghai 201203 (CN)。 郭崎 (GUO, Qi); 中国上海市浦东新区祖冲之路2290号1号楼1004, Shanghai 201203 (CN)。 陈云霁 (CHEN, Yunji); 中国上海市浦东新区祖冲之路2290号1号楼1004, Shanghai 201203 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) Title: TLB DEVICE SUPPORTING MULTIPLE DATA FLOWS AND UPDATE METHOD FOR TLB MODULE

(54) 发明名称: 支持多数据流的TLB装置和TLB模块的更新方法

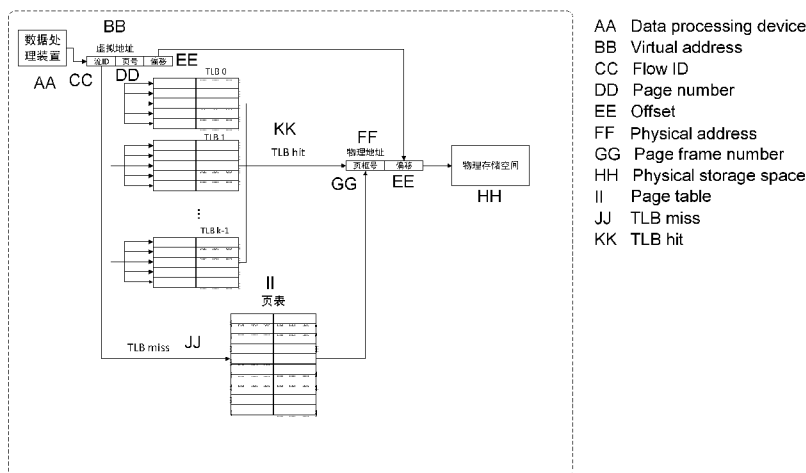


图 1

(57) Abstract: A TLB device supporting multiple data flows and an update method for a TLB module. The device comprises: a control unit corresponding to k data flows of a streaming application to be processed. k TLB modules are set, each of the TLB modules has two parts, namely a page and a page frame in one-to-one correspondence, and conversion from a logical address to a physical address is completed by means of a mapping relationship between the page and the page frame, wherein k is a natural number. According to the device and method, with regard to inherent features of streaming application data flows, the occurrence of the TLB miss situation can be greatly reduced during the conversion from a logical address to a physical address, so as to reduce the number of times a memory

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, 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))。

is accessed, thus greatly improving the performance of data accessing.

(57) 摘要: 一种支持多数据流的TLB装置和TLB模块的更新方法, 该装置包括: 控制单元, 对待处理的流式应用的k个数据流, 设置k个TLB模块, 每一个TLB模块具有一一对应的页与页框两部分, 通过页与页框的映射关系完成逻辑地址向物理地址的转换; 其中k为自然数。该装置和方法针对于流式应用数据流本身的特点, 可以极大减少逻辑地址与物理地址转换过程中TLB miss情况的发生, 从而减少访问内存的次数, 极大提高数据访问的性能。

支持多数据流的 TLB 装置和 TLB 模块的更新方法

技术领域

本公开涉及人工智能领域，更具体地涉及一种支持多数据流的 TLB 装置和 TLB 模块的更新方法。

5 背景技术

内存管理单元 (Memory Management Unit, MMU) 的出现，通过逻辑地址向物理地址的映射，使得程序的数据、堆栈的总的大小可以超过物理存储器的大小，MMU 的页表，即逻辑地址与物理地址的转换表，存储在内存中。由于逻辑地址到物理地址的转换需要多次访问内存，大大降低数据访问的性能，于是出现了传输后备缓存 (Translation Lookaside Buffer, TLB) 模块。TLB 模块存储页表中的一部分页项，当数据处理装置发出一个逻辑地址时，MMU 首先访问 TLB 模块，如果 TLB 模块中含有能转换这个逻辑地址的页，即 TLB 命中 (TLB hit)，直接利用此页进行地址转换，否则称为 TLB 失败 (TLB miss)，MMU 访问内存中的页表找到相关地址转换的页进行地址转换，同时，将这个页更新到 TLB 模块里。TLB 模块的出现减少访问内存的频率，极大提高数据访问的性能，特别是在某些数据需要经常访问的情况下。

对于类似于人工神经网络的流式应用，其特点是存在多个逻辑地址连续的数据流，并且对每个数据只访问一次。对于传统的 TLB 模块，由于流式应用数据通常只访问一次，因此经常会出现 TLB miss 的情况，每次 TLB miss 都要从内存中寻找相关地址转换的页并更新 TLB 模块，因此，对于流式应用，传统的 TLB 模块并不能很好的提高数据访问的性能。

发明内容

25 有鉴于此，本公开的一个目的在于提供一种支持多数据流的 TLB 装置，本公开的另一个目的在于提供一种 TLB 模块的更新方法，以解决传

统 TLB 模块在流式应用上的不足。

为了实现上述目的，作为本公开的一个方面，本公开提供了一种支持多数据流的 TLB 的装置，包括：

5 控制单元，对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数。

作为本公开的另一个方面，本公开还提供了一种 TLB 模块的更新方法，包括以下步骤：

10 对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数；

当发生 TLB miss 的情况时，控制单元访问内存中的页表找到相关地址转换的页进行地址转换，同时，将所述页更新到对应 TLB 模块里，替换掉所述 TLB 页中一个页项。

15 作为本公开的再一个方面，本公开还提供了一种 TLB 页的更新方法，包括以下步骤：

对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数；

20 每次发生 TLB miss 时，控制单元访问内存中的页表找到相关地址转换的页进行地址转换，同时，将所述页及所述页之后连续的页项替换 TLB 模块中原有的所有页项。

作为本公开的还一个方面，本公开还提供了一种 TLB 页的更新方法，包括以下步骤：

25 对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数，且所述 k 个 TLB 模块中 TLB 存放的页项为连续页项，即所存页项表示的逻辑地址连续，设当前 TLB 存放页项为 $\langle P_1, F_1 \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$ ，其中 $\langle P, F \rangle$ 表示页号与页

框号的映射关系，P 表示页号，F 表示页框号，页号 $P_1, P_2, P_3, \dots, P_n$ 所对应的逻辑地址范围连续；

在每次发生 TLB hit 后，完成地址转换的页被替换，即 $\langle P_1, F_1 \rangle$ 被替换，替换的页为当前 TLB 存放的连续页项的下一个连续页项，即 $\langle P_{n+1}, F_{n+1} \rangle$ ，
5 替换后 TLB 存放的页项为 $\langle P_{n+1}, F_{n+1} \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$ ，仍为连续页项。

基于上述技术方案可知，本公开的装置和方法具有如下有益效果：由于采用不同的 TLB 模块对应不同的数据流，使得每个 TLB 模块中可以存放逻辑地址连续的页项，有利于后续的更新操作；针对于流式应用的特性，采用逻辑地址连续的 TLB 模块，如此的 TLB 模块更新方法可以极大的减小 TLB miss 的发生，也就是可以减少 MMU 搜索内存中的
10 页表，极大的提高数据访问的性能。

附图说明

图 1 为根据本公开一实施例的装置的总体流程的示例框图；

图 2 为根据本公开一实施例的内存管理单元 (MMU) 的页表结构的
15 示例框图；

图 3 为根据本公开一实施例的逻辑地址的表示方法；

图 4 至图 6 为根据本公开一实施例的 TLB 模块的更新方法的示例框图。

具体实施方式

20 为使本公开的目的、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，对本公开作进一步的详细说明。

在本说明书中，下述用于描述本公开原理的各种实施例只是说明，不应该以任何方式解释为限制发明的范围。参照附图的下述描述用于帮助全面理解由权利要求及其等同物限定的本公开的示例性实施例。下述
25 描述包括多种具体细节来帮助理解，但这些细节应认为仅仅是示例性的。因此，本领域普通技术人员应认识到，在不背离本公开的范围和精神的情况下，可以对本文中描述的实施例进行多种改变和修改。此外，为了

清楚和简洁起见,省略了公知功能和结构的描述。此外,贯穿整个附图,相同附图标记用于相似功能和操作。

本公开公开了一种支持多数据流的 TLB 装置,包括:

控制单元,对应待处理的流式应用的 k 个数据流,设置 k 个 TLB 模块,每一个 TLB 模块具有一一对应的页与页框两部分,通过页与页框的映射关系完成逻辑地址向物理地址的转换;其中 k 为自然数。

其中,在数据处理装置传给所述控制单元的逻辑地址中,除了页号和偏移两个字段之外,还包括一个流 ID 字段,该流 ID 字段占用的比特数为 $\lceil \log_2 k \rceil$, k 表示 TLB 的个数, $\lceil \cdot \rceil$ 表示取向上取整数操作。

10 在一些实施例中,对于人工神经网络,该控制单元设置 4 个 TLB 模块,分别对应权值、输入、输出、部分和等四个数据流。

本公开还公开了几种 TLB 模块的更新方法,其中:

对应待处理的流式应用的 k 个数据流,设置 k 个 TLB 模块,每一个 TLB 模块具有一一对应的页与页框两部分,通过页与页框的映射关系完成逻辑地址向物理地址的转换;其中 k 为自然数;

方法一还包括以下步骤:

当发生 TLB miss 的情况时,控制单元访问内存中的页表找到相关地址转换的页进行地址转换,同时,将所述页更新到对应 TLB 模块里,替换掉所述 TLB 页中一个页项。

20 在一些实施例中,替换页项的步骤采用随机替换算法或 LRU 算法。

方法二还包括以下步骤:

每次发生 TLB miss 时,控制单元访问内存中的页表找到相关地址转换的页进行地址转换,同时,将所述页及所述页之后连续的页项替换 TLB 模块中原有的所有页项。

25 方法三还包括以下步骤:

将 k 个 TLB 模块中 TLB 存放的页项设为连续页项,即所存页项表示的逻辑地址连续,设当前 TLB 存放页项为 $\langle P_1, F_1 \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$, 其中 $\langle P, F \rangle$ 表示页号与页框号的映射关系, P 表示页号, F 表示页框号, 页号 $P_1, P_2, P_3, \dots, P_n$ 所对应的逻辑地址范围连续;

在每次发生 TLB hit 后,完成地址转换的页被替换,即 $\langle P_1, F_1 \rangle$ 被替换,替换的页为当前 TLB 存放的连续页项的下一个连续页项,即 $\langle P_{n+1}, F_{n+1} \rangle$,替换后 TLB 存放的页项为 $\langle P_{n+1}, F_{n+1} \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$, 仍为连续页项。

下面通过具体实施例对本公开的技术方案进行进一步阐述说明。

5 具体实施例一

根据本公开一实施例的支持多数据流的 TLB 的装置,可用于人工神经网络的相关应用,对于人工神经网络,一般存在权值、输入、输出和部分和等四个数据流,根据实际应用需要也可以设置其他不同的数据流。

图 1 为根据本公开一实施例的装置的总体流程的示例框图。对于人工神经网络,存在权值、输入、输出和部分和四种数据流,因此,示例框图中 TLB 模块可以设置为 4 个,每个 TLB 模块对应一个数据流。当数据处理装置传入一个逻辑地址时,根据逻辑地址中的流 ID 部分,选择对应 TLB 查找是否有该逻辑地址对应的页项,若有对应页项,则为 TLB hit,完成地址转换,若需要更新该 TLB,则执行更新操作,若没有对应页项,则为 TLB miss,从内存的页表总查找对应页项,同时对该 TLB 执行更新操作。

图 2 为根据本公开一实施例的内存管理单元 (MMU) 的页表结构 (TLB 模块结构) 的示例框图。通过分页存储管理将逻辑地址划分成一系列同等大小的部分,称为页 (page),并依次对每个页做编号,示例框图中页面的大小为 4kB,也可以是其他大小,根据具体的需求而定,每个页用一个编号代表 4kB 的页面大小,并非该页需占用 4kB 大小空间,例如页号 0 代表 0-4kB 大小的逻辑地址范围。对于物理内存也划分成同样大小的连续的部分,称为页框 (frame),并对每个页框做编号。页表存储的为页号与页框号的一个映射关系,例如,逻辑地址 8640 对应的范围为 [8k,16k],则对应的页号为 2,按照示例框图中的映射关系,映射为页框号为 6 所对应的物理地址范围,再根据偏移地址确定最终该逻辑地址转换后的物理地址。图 1 中的每个 TLB 存储的页项为页表中的部分页项,TLB 存储页项的数目由 TLB 存储空间大小决定。

图 3 为根据本公开一实施例的逻辑地址的表示方法,本公开的逻辑

地址的表示方法相对于传统的逻辑地址表示方法，添加了一些比特位表示流 ID，用于数据流与对应 TLB 模块的映射，示例图中只画出需要关心的部分。对于神经网络存在 4 个数据流，而 $\lceil \log_2 4 \rceil = 2$ ，因此流 ID 只需要 2bit 就可以区分四个数据流。由于图 2 中设定的页面大小为 4kB，
5 $\lceil \log_2 4k \rceil = 12$ ，因此 32 位的逻辑地址中，有 20 位对应为页号，12 位对应为偏置地址，根据页号与页框号的对应关系，可以获得对应物理地址的前 20 位，对应物理地址的后 12 位与逻辑地址后 12 位一致。

图 4-6 为根据本公开一实施例的 TLB 页的更新方法的示例框图。示例框图只示出一个 TLB 的一种更新方法，对装置任何 TLB 都可以使用
10 任何其中一种或混用多种更新方法。

图 4 为根据本公开一实施例的 TLB 页的第一更新方法。第一更新方法 1 与传统的 TLB 页的更新方法一样，当发生 TLB miss 的情况时，MMU 访问内存中的页表找到相关地址转换的页进行地址转换，同时，将这个页更新到 TLB 里，替换掉 TLB 中一个页项。此处替换页项可以采用不
15 用的替换算法，例如，随机替换算法、LRU 算法等。随机替换算法根据数据处理装置生成的随机数决定要替换的 TLB 页项；LRU 算法选择替换的页项根据访问时间离当前时间的长短决定，上一次访问时间距当前最长的页项作为被替换的页项。

图 5 为根据本公开一实施例的 TLB 页的第二更新方法。第二更新方法
20 利用流式应用的特点，即数据流逻辑地址连续且每个数据只访问一次。其更新方法为，每次发生 TLB miss 时，MMU 访问内存中的页表找到相关地址转换的页进行地址转换，同时，将这个页及该页之后连续的页项替换 TLB 原有的所有页项。其中，图中 P、P1、P2、P3... 为页号连续的页项。

图 6 为根据本公开一实施例的 TLB 页的第三更新方法。第三更新方法
25 同样利用了流式应用的特点。其更新前提是 TLB 存放的页项为连续页项，即所存页项表示的逻辑地址连续，设当前 TLB 存放页项为 $\langle P_1, F_1 \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$ ，其中 $\langle P, F \rangle$ 表示页号与页框号的映射关系，P 表示页号，F 表示页框号，页号 $P_1, P_2, P_3, \dots, P_n$ 所对应的逻辑地址范围连

续。在每次发生 TLB hit 后,完成地址转换的页被替换,即 $\langle P_1, F_1 \rangle$ 被替换,替换的页为当前 TLB 存放的连续页项的下一个连续页项,即 $\langle P_{n+1}, F_{n+1} \rangle$,替换后 TLB 存放的页项为 $\langle P_{n+1}, F_{n+1} \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$, 仍为连续页项,该更新方式类似于滑动窗口的滑动机制。

5

具体实施例二

根据本公开一实施例的支持多数据流的 TLB 的装置,可用于机器学习的相关流式应用,例如自然语音处理、手写识别、人脸识别等相关应用,设本实施例中存在 k 个数据流, k 的大小根据不同应用的需求而定。

10 图 1 为根据本公开一实施例的装置的总体流程的示例框图。对于存在 k 个数据流的流式应用,因此,示例框图中 TLB 模块可以设置为 k 个,每个 TLB 模块对应一个数据流。当数据处理装置传入一个逻辑地址时,根据逻辑地址中的流 ID 部分,选择对应 TLB 查找是否有该逻辑地址对应的页项,若有对应页项,则为 TLB hit,完成地址转换,若需要更新该 TLB,则执行更新操作,若没有对应页项,则为 TLB miss,从内存的页表总查找对应页项,同时对该 TLB 执行更新操作。

图 2 为根据本公开一实施例的内存管理单元 (MMU) 的页表结构的示例框图。通过分页存储管理将逻辑地址划分成一系列同等大小的部分,称为页 (page),并依次对每个页做编号,示例框图中页面的大小为 20 4kB,也可以是其他大小,根据具体的需求而定,每个页用一个编号代表 4kB 的页面大小,并非该页需占用 4kB 大小空间,例如页号 0 代表 0-4kB 大小的逻辑地址范围。对于物理内存也划分成同样大小的连续的部分,称为页框 (frame),并对每个页框做编号。页表存储的为页号与页框号的一个映射关系,例如,逻辑地址 8640 对应的范围为 [8k,16k], 25 则对应的页号为 2,按照示例框图中的映射关系,映射为页框号为 6 所对应的物理地址范围,再根据偏移地址确定最终该逻辑地址转换后的物理地址。图 1 中的每个 TLB 存储的页项为页表中的部分页项,TLB 存储页项的数目由 TLB 存储空间大小决定。

图 3 为根据本公开一实施例的逻辑地址的表示方法,本公开的逻辑

地址的表示方法相对于传统的逻辑地址表示方法，添加了一些比特位表示流 ID，用于数据流与对应 TLB 模块的映射，示例图中只画出需要关心的部分。对于神经网络存在 k 个数据流，因此流 ID 只需要 $\lceil \log_2 k \rceil^{bit}$ 就可以区分 k 个数据流。由于图 2 中设定的页面大小为 4kB，
 5 $\lceil \log_2 4k \rceil = 12$ ，因此 32 位的逻辑地址中，有 20 位对应为页号，12 位对应为偏置地址，根据页号与页框号的对应关系，可以获得对应物理地址的前 20 位，对应物理地址的后 12 位与逻辑地址后 12 位一致。

图 4-6 为根据本公开一实施例的 TLB 页的更新方法的示例框图。示例框图只示出一个 TLB 的一种更新方法，对装置任何 TLB 都可以使用
 10 任何其中一种或混用多种更新方法。

图 4 为根据本公开一实施例的 TLB 页的第一更新方法。第一更新方法与传统的 TLB 页的更新方法一样，当发生 TLB miss 的情况是，MMU 访问内存中的页表找到相关地址转换的页进行地址转换，同时，将这个页更新到 TLB 里，替换掉 TLB 中一个页项。此处替换页项可以采用不
 15 用的替换算法，例如，随机替换算法、LRU 算法等。随机替换算法根据数据处理装置生成的随机数决定要替换的 TLB 页项；LRU 算法选择替换的页项根据访问时间离当前时间的长短决定，上一次访问时间距当前最长的页项作为被替换的页项。

图 5 为根据本公开一实施例的 TLB 页的第二更新方法。第二更新方法利用流式应用的特点，即数据流逻辑地址连续且每个数据只访问一次。
 20 其更新方法为，每次发生 TLB miss 时，MMU 访问内存中的页表找到相关地址转换的页进行地址转换，同时，将这个页及该页之后连续的页项替换 TLB 原有的所有页项。其中，图中 P、P1、P2、P3... 为页号连续的页项。

图 6 为根据本公开一实施例的 TLB 页的第三更新方法。第三更新方法同样利用了流式应用的特点。其更新前提是 TLB 存放的页项为连续页
 25 项，即所存页项表示的逻辑地址连续，设当前 TLB 存放页项为 $\langle P_1, F_1 \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$ ，其中 $\langle P, F \rangle$ 表示页号与页框号的映射关系，P 表示页号，F 表示页框号，页号 $P_1, P_2, P_3, \dots, P_n$ 所对应的逻辑地址范围连

续。在每次发生 TLB hit 后,完成地址转换的页被替换,即 $\langle P_1, F_1 \rangle$ 被替换,替换的页为当前 TLB 存放的连续页项的下一个连续页项,即 $\langle P_{n+1}, F_{n+1} \rangle$,替换后 TLB 存放的页项为 $\langle P_{n+1}, F_{n+1} \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$, 仍为连续页项,该更新方式类似于滑动窗口的滑动机制。

5 对于类似于人工神经网络的流式应用,由于其具有多个逻辑地址连续的数据流的特点,通过采用多个 TLB 模块,使得每个 TLB 可以存放逻辑地址连续的页项,对于页项逻辑连续的 TLB,采用本公开所述的 TLB 页的更新方法,可以极大减小 TLB miss 的发生,即可以极大减少从页表中查找的次数,极大提高数据访问的性能。

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

权利要求

1、一种支持多数据流的 TLB 装置，其特征在于，包括：

控制单元，对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数。

2、如权利要求 1 所述的 TLB 装置，其特征在于，数据处理装置传给所述控制单元的逻辑地址中，除了页号和偏移两个字段之外，还包括一个流 ID 字段，其中，所述流 ID 字段占用的比特数为 $\lceil \log_2 k \rceil$ ， k 表示 TLB 的个数， $\lceil \cdot \rceil$ 表示取向上取整数操作。

3、如权利要求 1 所述的 TLB 装置，其特征在于，对于人工神经网络，所述控制单元设置 4 个 TLB 模块，分别对应权值、输入、输出、部分和四个数据流。

4、一种 TLB 模块的更新方法，其特征在于，包括以下步骤：

对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数；

当发生 TLB miss 的情况时，控制单元访问内存中的页表找到相关地址转换的页进行地址转换，同时，将所述页更新到对应 TLB 模块里，替换掉所述 TLB 页中一个页项。

5、权利要求 4 所述的更新方法，其特征在于，其中替换页项的步骤采用随机替换算法或 LRU 算法。

6、一种 TLB 页的更新方法，其特征在于，包括以下步骤：

对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数；

每次发生 TLB miss 时，控制单元访问内存中的页表找到相关地址转换的页进行地址转换，同时，将所述页及所述页之后连续的页项替换 TLB 模块中原有的所有页项。

7、一种 TLB 页的更新方法，其特征在于，包括以下步骤：

5 对应待处理的流式应用的 k 个数据流，设置 k 个 TLB 模块，每一个 TLB 模块具有一一对应的页与页框两部分，通过页与页框的映射关系完成逻辑地址向物理地址的转换；其中 k 为自然数，且所述 k 个 TLB 模块中 TLB 存放的页项为连续页项，即所存页项表示的逻辑地址连续，设当前 TLB 存放页项为 $\langle P_1, F_1 \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$ ，其中 $\langle P, F \rangle$ 表示页号与页框号的映射关系， P 表示页号， F 表示页框号，页号 $P_1, P_2, P_3, \dots, P_n$ 所对应的逻辑地址范围连续；

10 在每次发生 TLB hit 后，完成地址转换的页被替换，即 $\langle P_1, F_1 \rangle$ 被替换，替换的页为当前 TLB 存放的连续页项的下一个连续页项，即 $\langle P_{n+1}, F_{n+1} \rangle$ ，替换后 TLB 存放的页项为 $\langle P_{n+1}, F_{n+1} \rangle, \langle P_2, F_2 \rangle, \langle P_3, F_3 \rangle, \dots, \langle P_n, F_n \rangle$ ，仍为连续页项。

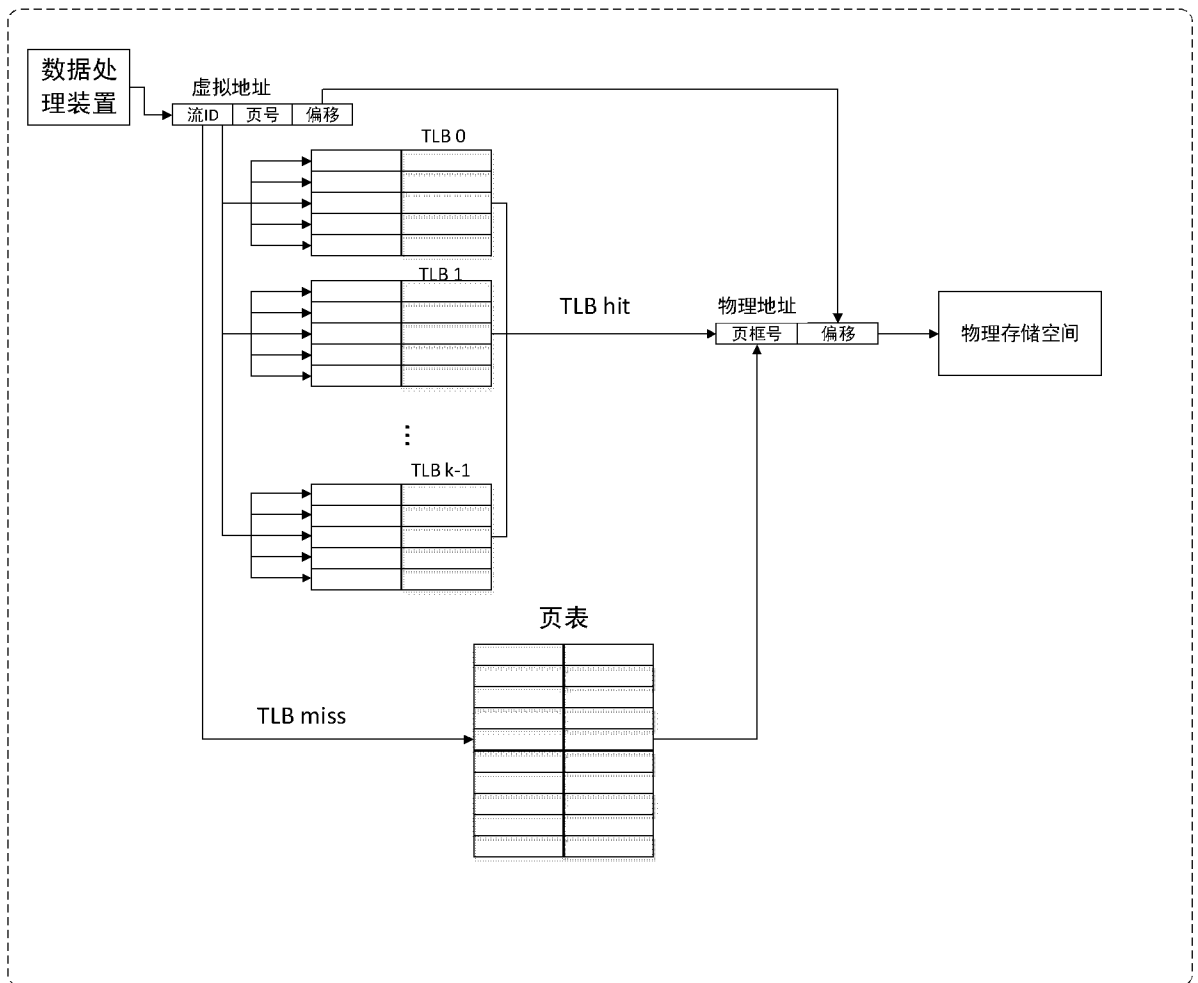


图 1

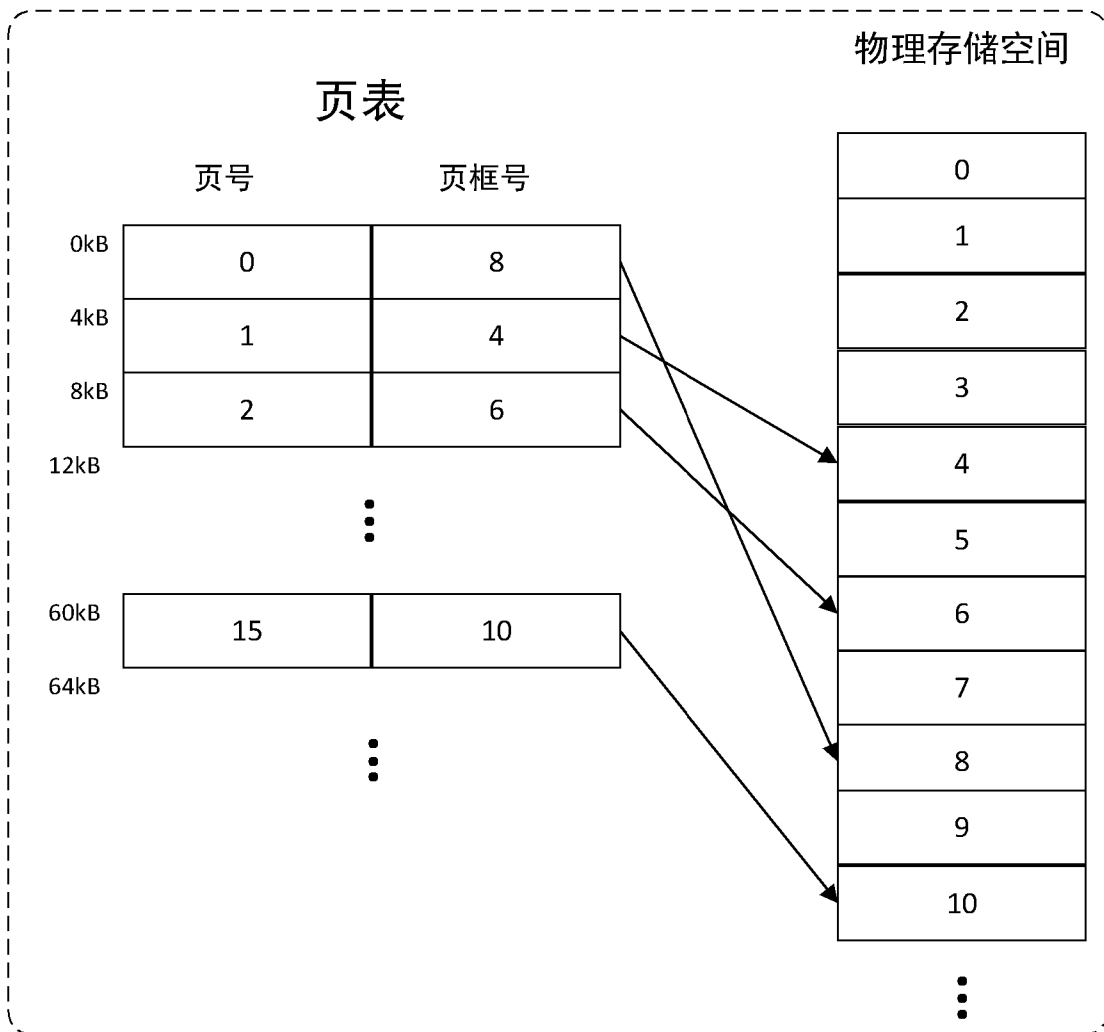


图 2

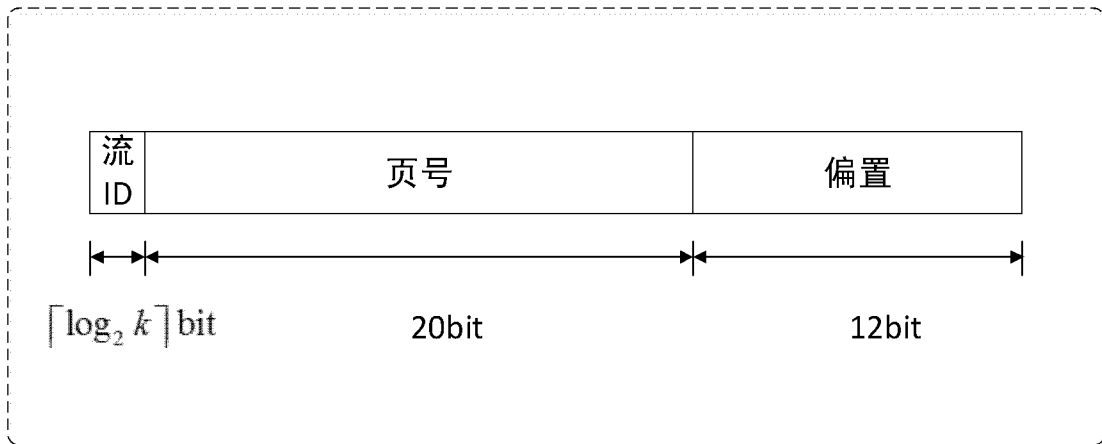


图 3

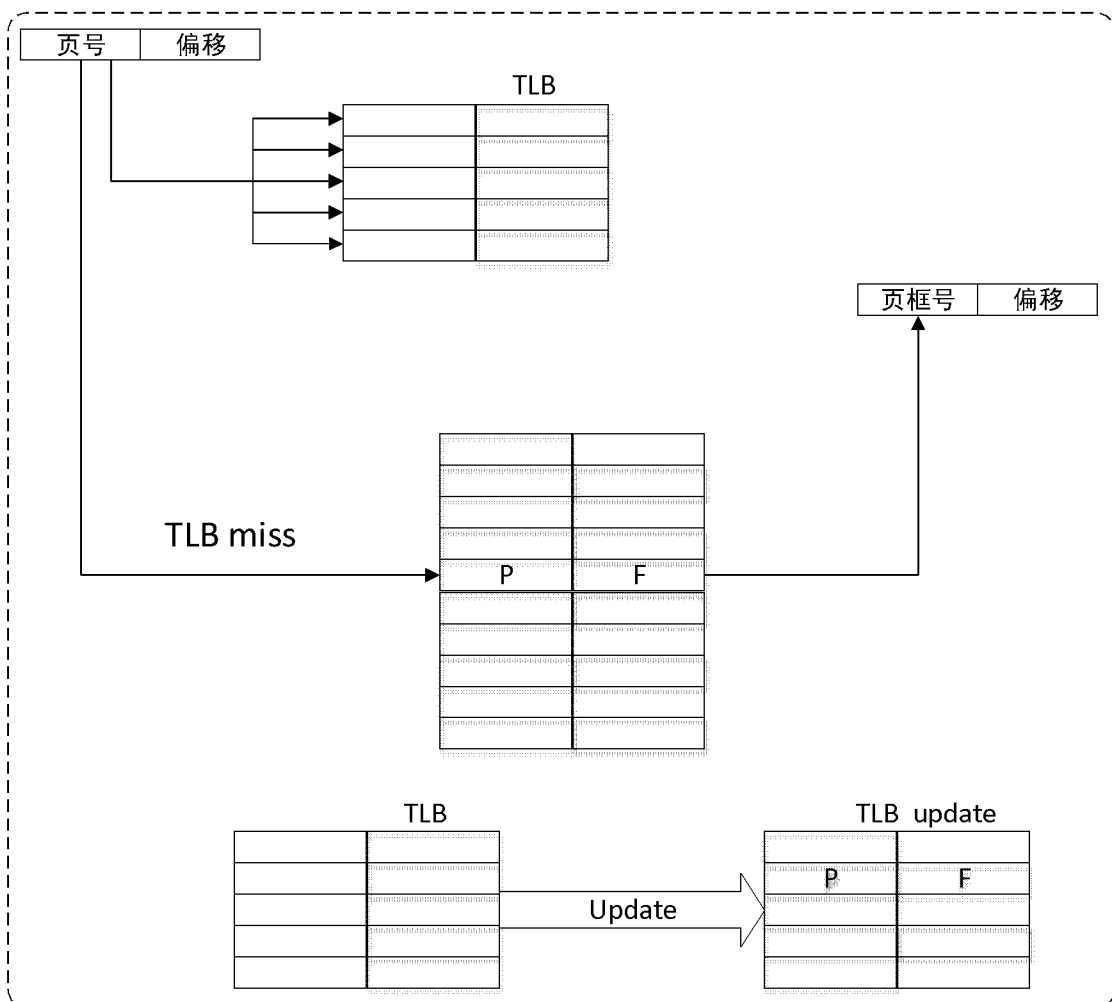


图 4

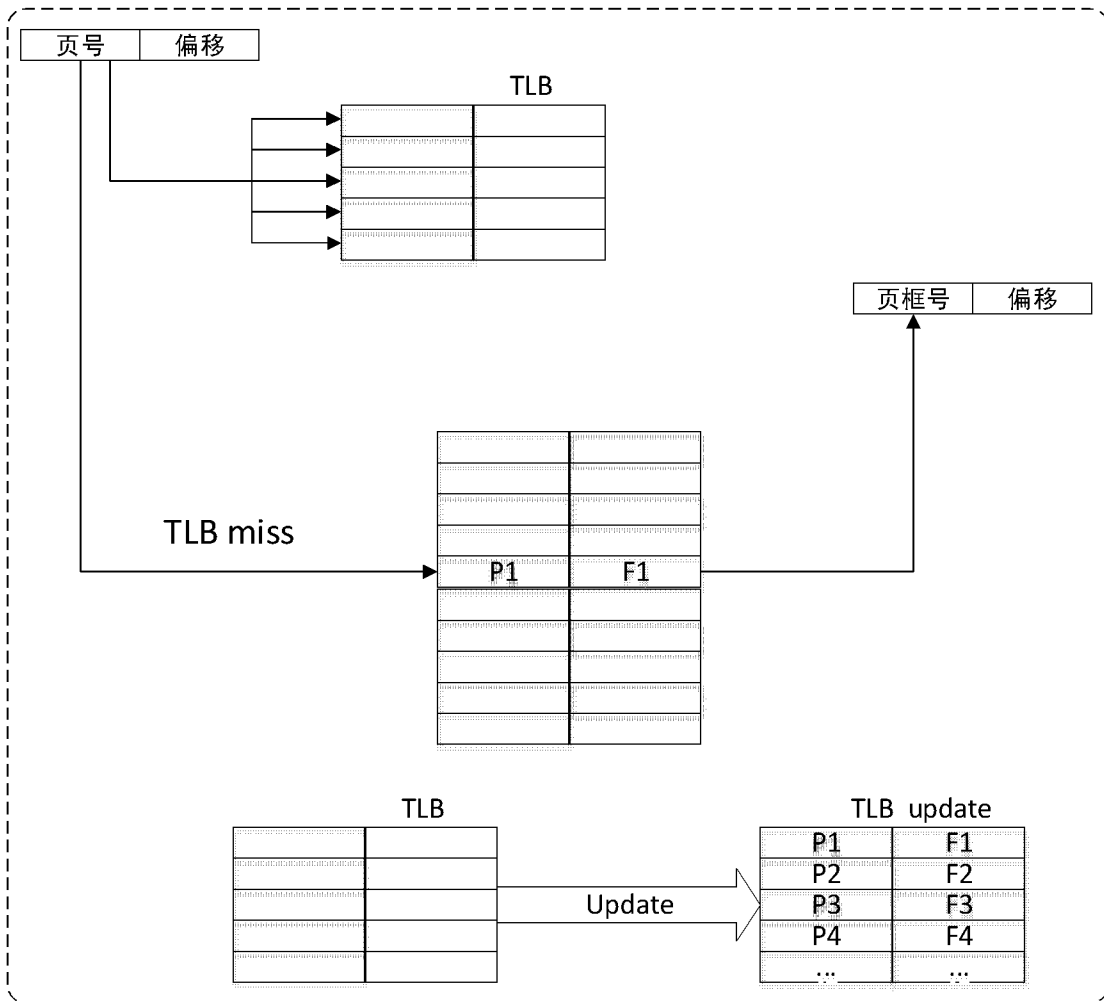


图 5

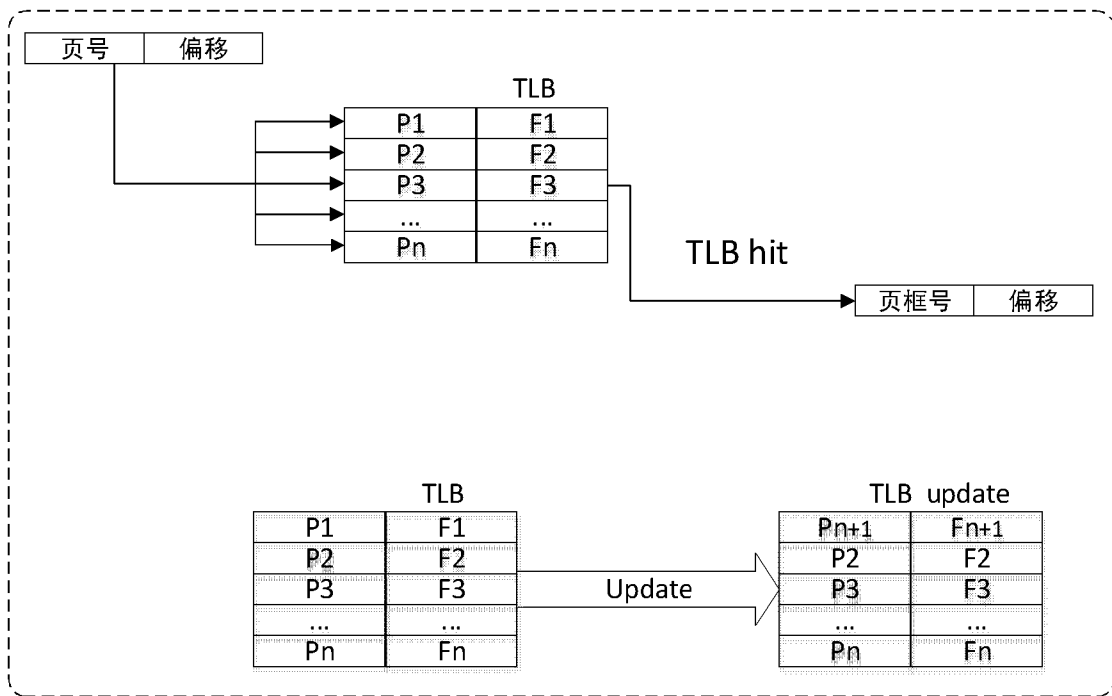


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/095845

A. CLASSIFICATION OF SUBJECT MATTER <p style="text-align: center;">G06F 12/1027 (2016.01) i</p> <p>According to International Patent Classification (IPC) or to both national classification and IPC</p>		
B. FIELDS SEARCHED <p>Minimum documentation searched (classification system followed by classification symbols)</p> <p style="text-align: center;">G06F</p> <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNPAT, WPI, EPODOC, GOOGLE: 缓冲, 页框, 映射, 逻辑, 物理, 转换, 命中, TLB, page, frame, PFT, mapped, logic, physics, switch, conversion, transform, hit, miss</p>		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104298616 A (HUAWEI TECHNOLOGIES CO., LTD. et al.), 21 January 2015 (21.01.2015), claims, and description, paragraphs [0018] and [0064]-[0065]	1-7
A	CN 104346284 A (HUAWEI TECHNOLOGIES CO., LTD. et al.), 11 February 2015 (11.02.2015), entire document	1-7
A	CN 103455443 A (HUAWEI TECHNOLOGIES CO., LTD.), 18 December 2013 (18.12.2013), entire document	1-7
A	CN 102662860 A (C*CORE TECHNOLOGY (TIANJIN) CO., LTD.), 12 September 2012 (12.09.2012), entire document	1-7
A	US 2004073766 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION), 15 April 2004 (15.04.2004), entire document	1-7
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents:	“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	
“A” document defining the general state of the art which is not considered to be of particular relevance	“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	
“E” earlier application or patent but published on or after the international filing date	“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	
“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)	“&” document member of the same patent family	
“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		
Date of the actual completion of the international search <p style="text-align: center;">29 August 2017</p>	Date of mailing of the international search report <p style="text-align: center;">27 August 2017</p>	
Name and mailing address of the ISA 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 <p style="text-align: center;">HU, Beibei</p> Telephone No. (86-10) 82246931	

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/095845

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104298616 A	21 January 2015	None	
CN 104346284 A	11 February 2015	None	
CN 103455443 A	18 December 2013	None	
CN 102662860 A	12 September 2012	None	
US 2004073766 A1	15 April 2004	CN 1492332 A	28 April 2004

国际检索报告

国际申请号

PCT/CN2017/095845

<p>A. 主题的分类 G06F 12/1027(2016.01) i</p> <p>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</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号) G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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)) CNPAT, WPI, EPODOC, GOOGLE: 缓冲, 页框, 映射, 逻辑, 物理, 转换, 命中, TLB, page, frame, PFT, mapped, logic, physics, switch, conversion, transform, hit, miss</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104298616 A (华为技术有限公司 等) 2015年 1月 21日 (2015 - 01 - 21) 权利要求书、说明书第[0018]、[0064]-[0065]段</td> <td>1-7</td> </tr> <tr> <td>A</td> <td>CN 104346284 A (华为技术有限公司 等) 2015年 2月 11日 (2015 - 02 - 11) 全文</td> <td>1-7</td> </tr> <tr> <td>A</td> <td>CN 103455443 A (华为技术有限公司) 2013年 12月 18日 (2013 - 12 - 18) 全文</td> <td>1-7</td> </tr> <tr> <td>A</td> <td>CN 102662860 A (天津国芯科技有限公司) 2012年 9月 12日 (2012 - 09 - 12) 全文</td> <td>1-7</td> </tr> <tr> <td>A</td> <td>US 2004073766 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2004年 4月 15日 (2004 - 04 - 15) 全文</td> <td>1-7</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 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</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 104298616 A (华为技术有限公司 等) 2015年 1月 21日 (2015 - 01 - 21) 权利要求书、说明书第[0018]、[0064]-[0065]段	1-7	A	CN 104346284 A (华为技术有限公司 等) 2015年 2月 11日 (2015 - 02 - 11) 全文	1-7	A	CN 103455443 A (华为技术有限公司) 2013年 12月 18日 (2013 - 12 - 18) 全文	1-7	A	CN 102662860 A (天津国芯科技有限公司) 2012年 9月 12日 (2012 - 09 - 12) 全文	1-7	A	US 2004073766 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2004年 4月 15日 (2004 - 04 - 15) 全文	1-7
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
X	CN 104298616 A (华为技术有限公司 等) 2015年 1月 21日 (2015 - 01 - 21) 权利要求书、说明书第[0018]、[0064]-[0065]段	1-7																		
A	CN 104346284 A (华为技术有限公司 等) 2015年 2月 11日 (2015 - 02 - 11) 全文	1-7																		
A	CN 103455443 A (华为技术有限公司) 2013年 12月 18日 (2013 - 12 - 18) 全文	1-7																		
A	CN 102662860 A (天津国芯科技有限公司) 2012年 9月 12日 (2012 - 09 - 12) 全文	1-7																		
A	US 2004073766 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2004年 4月 15日 (2004 - 04 - 15) 全文	1-7																		
国际检索实际完成的日期	国际检索报告邮寄日期																			
2017年 8月 29日	2017年 9月 27日																			
ISA/CN的名称和邮寄地址	授权官员																			
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	胡贝贝																			
传真号 (86-10)62019451	电话号码 (86-10)82246931																			

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/095845

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	104298616	A	2015年 1月 21日	无			
CN	104346284	A	2015年 2月 11日	无			
CN	103455443	A	2013年 12月 18日	无			
CN	102662860	A	2012年 9月 12日	无			
US	2004073766	A1	2004年 4月 15日	CN	1492332	A	2004年 4月 28日