

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 7/24 (2006.01)



# [12] 发明专利说明书

专利号 ZL 200380104150.7

[45] 授权公告日 2009年7月1日

[11] 授权公告号 CN 100507830C

[22] 申请日 2003.10.23

[21] 申请号 200380104150.7

[30] 优先权

[32] 2002.10.25 [33] JP [31] 310618/2002

[86] 国际申请 PCT/JP2003/013559 2003.10.23

[87] 国际公布 WO2004/038582 日 2004.5.6

[85] 进入国家阶段日期 2005.5.26

[73] 专利权人 特博数据实验室公司

地址 日本神奈川

[72] 发明人 古庄晋二

[56] 参考文献

JP2002-169689A 2002.6.14

US6032207A 2000.2.29

US5544312A 1996.8.6

审查员 徐 春

[74] 专利代理机构 中国国际贸易促进委员会专利  
商标事务所

代理人 付建军

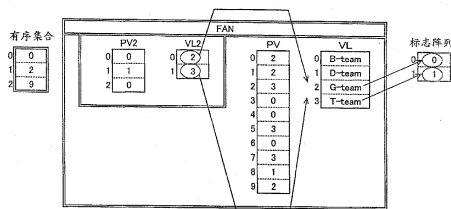
权利要求书 5 页 说明书 27 页 附图 45 页

[54] 发明名称

数据处理方法和数据处理程序

[57] 摘要

可以缩短处理阵列的子集所需的时间。表格格式数据被分成信息块，每一个信息块都包括值列表 (VL) 和指针阵列 (PV)，在所述值列表中，项目值以对应于每一个项目的项目值的项目值号码的顺序存储，在所述指针阵列中，指针值以唯一记录号的顺序存储，每一个指针值都指出项目值号码。一部分记录被选择作为子集，并且响应该选择，创建包括子集的记录号的有序集合阵列。有序集合阵列中的每一个记录号都指出指针阵列中的其中一个指针值的位置。在对应于其中排列了记录号的位置的项目值号码阵列中排列指针阵列中的指针值。参考项目值号码排列中的值，生成第二指针阵列 (PV2)，其中，排列了位置元素，第二值列表 (VL2)，其中，排列了值元素。



1. 一种用于从被表达为记录阵列的表格格式数据中提取子集作为处理对象的数据处理方法，每一个记录都包括项目和属于该项目的项目值，包括下列步骤：

a) 通过将表格格式数据划分为信息块来构造表格格式数据，每一个信息块都包括第一值列表和第一指针阵列，在所述第一值列表中，项目值以对应于属于特定项目的项目值的项目值号码的顺序存储，在所述第一指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储；

b) 响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列；

c) 将在由有序集合阵列的每一个记录号指出的位置上的第一指针阵列中的指针值对应于其中在有序集合阵列中排列了记录号的位置排列到项目值号码阵列中；以及

d) 通过引用项目值号码阵列，创建存储项目值号码阵列中的值元素的第二值列表，和存储指出对应于记录号的第二值列表中的元素的位置元素的第二指针阵列，并且其中

通过由记录号指出的位置上的第二指针阵列中的元素和由第二指针阵列中的元素指出的位置上的第二值列表中的元素，根据有序集合阵列的记录号指定第一值列表中的值。

2. 根据权利要求 1 所述的数据处理方法，其特征在于，创建第二指针阵列和第二值列表的步骤包括下列步骤：

将项目值号码阵列中的元素分类，在重复值被排除的状态下创建阵列，并使创建阵列成为第二值列表；以及

通过转换项目值号码阵列中的元素以反映元素的分类和重复值的排除，来创建第二指针阵列。

3. 根据权利要求 1 或 2 所述的数据处理方法，其特征在于，进一步包括将有序集合阵列的大小与所述第一值列表的大小进行比较的步骤

骤，

其中，作为比较的结果，在所述有序集合阵列的大小按预先确定的比率小于第一值列表的大小的情况下，执行步骤 c) 和 d)。

4. 一种用于从被表达为记录阵列的表格格式数据中提取子集作为处理对象的数据处理方法，每一个记录都包括项目和属于该项目的值，包括下列步骤：

a) 通过将表格格式数据划分为信息块来构造表格格式数据，每一个信息块都包括第一值列表和第一指针阵列，在所述第一值列表中，项目值以对应于属于特定项目的项目值的项目值号码的顺序存储，在所述第一指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储；

b) 响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列；

c) 将在由有序集合阵列的每一个记录号指出的位置上的第一指针阵列中的指针值对应于其中在有序集合阵列排列了记录号的位置排列到项目值号码阵列中；以及

e) 通过由记录号指出的位置上的项目值号码阵列中的元素，根据有序集合阵列的记录号指定第一值列表中的值。

5. 根据权利要求4所述的数据处理方法，其特征在于，进一步包括将有序集合阵列的大小与所述值列表的大小进行比较的步骤，

其中，作为比较的结果，在所述有序集合阵列的大小按预先确定的比率小于第一值列表的大小的情况下，执行步骤c) 和e)。

6. 使用根据权利要求1或2所述的数据处理方法创建的第二指针阵列和/或第二值列表的检索方法，其特征在于，所述方法包括下列步骤：

在第二值列表中的元素中指定指出项目值的元素，作为检索对象；

将指出标志为on的值放入这样一个位置，所述位置对应于与第二值列表具有同样大小的标志阵列中的指定元素的位置；

指定由有序集合阵列中的记录号指出的第二指针阵列的元素；  
引用由第二指针阵列的元素指出的位置处的标志阵列中标志的状态；以及

在所述标志的状态为 on 的情况下，连续地将记录号排列到新提供的用于输出的有序集合阵列中。

7. 使用根据权利要求1或2所述的数据处理方法创建的第二指针阵列和/或第二值列表的聚合方法，其特征在于，所述方法包括下列步骤：

创建分类号阵列，在该阵列中，对应于第二值列表的元素，排列指出值的类别的分类号；

指定对应于指定的分类号的有序集合阵列中的记录号；以及  
使用由指定的记录号指出的第二值列表中的元素执行聚合。

8. 使用根据权利要求 1或2所述的数据处理方法创建的第二指针阵列和/或第二值列表的分类方法，其特征在于，所述方法包括下列步骤：

计算存在号码，作为由第二值列表的每一个值的第二指针阵列指出的元素的数量；

基于存在号码，创建对应于第二值列表的值并且指出其中有序集合阵列中的记录号将被排列的头部位置的累积号码阵列；以及

引用累积号码阵列，并将有序集合阵列的记录号排列到用于输出的阵列，以便反映所述第二值列表中的值的分类顺序。

9. 通过使用涉及由根据权利要求1或2所述的数据处理方法创建的多个表格格式数据中的每一个数据的信息块中的第二指针阵列和/或第二值列表联接表格格式数据的方法，其特征在于所述方法包括下列步骤：

查找要在多个表格格式数据中的每一个数据中共享的项目；

使涉及项目的信息块的第二值列表中的值相等；以及

响应使值相等，随着项目值的排列中的变化，更新每一个信息块中的第二指针阵列中的元素。

10. 一种用于从被表达为记录阵列的表格格式数据中提取子集作为处理对象的计算机系统，每一个记录都包括项目和属于该项目的值，所述计算机系统包括：

a) 用于通过将表格格式数据划分为信息块来构造表格格式数据的装置，每一个信息块都包括第一值列表和第一指针阵列，在所述第一值列表中，项目值以对应于属于特定项目的值的项目值号码的顺序存储，在所述第一指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储；

b) 用于响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列的装置；

c) 用于将在由有序集合阵列的每一个记录号指出的位置上的第一指针阵列中的指针值对应于其中在有序集合阵列中排列了记录号的位置排列到项目值号码阵列中的装置；以及

d) 用于通过引用项目值号码阵列中的值，创建存储位置元素的第二指针阵列和存储值元素的第二值列表的装置，以及其中

通过由记录号指出的位置上的第二指针阵列中的元素和由第二指针阵列中的元素指出的位置上的第二值列表中的元素，根据有序集合阵列的记录号指定第一值列表中的值。

11. 根据权利要求 10 所述的计算机系统，其特征在于

用于创建第二指针阵列和第二值列表的装置进一步包括：

将项目值号码阵列中的元素分类，在重复值被排除的状态下创建阵列，并使创建阵列成为第二值列表的装置；以及

用于通过转换项目值号码阵列中的元素以反映元素分类和重复值的排除，创建第二指针阵列的装置。

12. 一种用于从被表达为记录阵列的表格格式数据中提取子集作为处理对象的计算机系统，每一个记录都包括项目和属于该项目的值，所述计算机系统包括：

a) 用于通过将表格格式数据划分为信息块来构造表格格式数据的装置，每一个信息块都包括第一值列表和第一指针阵列，在所述第

一值列表中，项目值以对应于属于特定项目的项目值的项目值号码的顺序存储，在所述第一指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储；

b) 用于响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列的装置；

c) 用于将在由有序集合阵列的每一个记录号指出的位置上的指针阵列中的指针值对应于其中有序集合阵列中排列了记录号的位置排列到项目值号码阵列中的装置；以及

e) 用于通过由记录号指示位置上的项目值号码阵列中的元素，根据有序集合阵列的记录号指定第一值列表中的值的装置。

## 数据处理方法和数据处理程序

### 技术领域

本发明涉及用于有效地处理表格格式数据的子集并涉及程序。

### 背景技术

对诸如数据库之类的大量数据进行检索、聚合或分类，或联接或更新这些数据需要花费很多时间。为了解决这一问题，本发明人提出了一种用于以非常高的速度对表格格式数据进行检索、聚合和分类的方法以及用于联接或更新表格格式数据或执行事务处理的方法（例如，参见国际出版物 No. WO00/10103，JP-A-2000-339390，以及 JP-A-2001-043290）。

这些文档中说明的一系列方法的创新之处在于，没有使用索引，可以执行统一的处理，甚至对于子集也可以执行处理，而不会降低效率。

由本发明人始创的表格格式数据具有如下所述的数据结构。图 1(a) 是显示了表格格式数据的示例的视图。在此示例中，使存储记录号的阵列 100 中的每一个记录号对应于项目“member name”、“area”和“fan”的值（项目值）（参见附图标记 101）。由本发明人始创的数据结构以如图 2(a) 到 2(c) 所示的格式持有如图 1(a) 所示的表格格式数据。

例如，如图 2(a) 所示，对于项目“member name”，提供了值列表（在某些情况下，简称为“VL”）211，其中，按照片假名字音表的顺序（按英语字母顺序）对项目值进行排序，以及值列表的指针阵列（在某些情况下，简称为“PV”）212，其中，存储了对应于相应的记录号的号码以及由记录号指出的值列表 211 中的指示元素位置。如上所述，对于某些项目，包括指针阵列和值列表的阵列组，除了这些外，在某些情况下包括一个或多个伴随的阵列被称为信息块（参见附

图标记 201)。从图 2(a) 中可以理解, 指针阵列包含其数量等于表格格式数据的记录的数量的元素。还是对于项目“area”, 信息块 202 包括 PV 和 VL, 同样, 对于项目“fan”, 信息块 203 包括 PV 和 VL。

注意记录号“0”, 由于信息块 201 中的 PV 的值为“5”, 因此, 指定了其中的存储位置为“5”的 VL 中的项目值“Green”。同样, 由于信息块 202 和 203 的 PV 的值分别为“5”和“2”, 因此, 指定了对应的存储位置号码中的项目值“Tokyo”和“G-Team”。可以理解, 这些都对应于图 1(a) 中的记录号“0”中包含的项目值。

使用如上所述的数据结构执行检索、聚合、分类和联接, 以便实现非常高速的处理。

然而, 在使用上面的数据结构对检索等等进行处理的过程中, 已经认识到, 可能会产生随后的问题。

甚至在使少量的子集成为所有记录之中的处理对象的情况下, 从子集中检索所希望的记录, 根据由本申请人始创的方法, 提供与 VL 具有一样大小的阵列, 排列指出它是否包含在要检索的值的范围内的标志。同样, 对于检索的通过/失败判断, 需要与值列表具有一样大小的阵列。

也是在分类或聚合时, 需要与值列表具有一样大小的一个或多个排列。此外, 也是在联接过程中, 必须提供与值列表具有一样大小的各种排列。

相应地, 甚至在当从非常多的记录中使非常小的数量的子集成为的处理对象时执行检索、聚合、分类和联接的过程的情况下, 当在信息块中提供大的值列表时, 需要进行处理所需的阵列的大的存储区域。此外, 需要在阵列中排列值的过程, 以及扫描值的过程等等。如此, 一直有这样的问題, 与使所有记录成为处理对象的情况相比, 虽然处理对象比较小, 但是, 不能相应地缩短处理时间。

本发明的目的是提供一种方法: 该方法通过使阵列与子集的大小或根据此受限的值列表的子集的大小一致, 实现对应于子集的大小的



处理时间的缩短。

### 发明内容

本发明的目的通过一种用于从被表达为记录阵列的表格格式数据中提取子集作为处理对象的数据处理方法来实现，每一个记录都包括项目和属于该项目的项目值，该数据处理方法包括下列步骤：

a) 通过将表格格式数据划分为信息块来构造表格格式数据，每一个信息块都包括值列表和指针阵列，在所述值列表中，项目值以对应于属于特定项目的项目值的项目值号码的顺序存储，在所述指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储，

b) 响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列，

c) 将在由有序集合的每一个记录号指出的位置上的指针阵列中的指针值对应于其中排列了记录号的位置排列到项目值号码阵列中，以及

d) 通过引用项目值号码阵列中的值，创建存储位置元素的第二指针阵列和存储值元素的第二值列表，以及其中

通过由记录号指出的位置上的第二指针阵列中的元素和由第二指针阵列中的元素指出的位置上的第二值列表中的元素，根据有序集合阵列的记录号指定值列表中的值。

根据本发明，创建指出构成子集的与记录号相关的值列表的项目值的第二值列表，通过引用第二值列表，可以执行数据处理（稍后描述的检索、聚合、分类，以及联接处理）。相应地，由于可以使对其进行引用的数据量比较小，因此，可以缩短处理时间。此外，由于还可以使在数据处理时创建的阵列的大小与第二值列表或第二指针阵列相同，因此，也可以降低所需要的存储区。

在优选实施例中，创建第二指针阵列和第二值列表的步骤包括下列步骤：将项目值号码阵列中的元素分类，在重复值被排除的状态下创建阵列，并使其成为第二值列表，通过转换项目值号码阵列中的元素以反映分类和重复值的排除，创建第二指针阵列。

此外，在另一个优选实施例中，一种用于从被表达为记录阵列的表格格式数据中提取子集作为处理对象的数据处理方法，每一个记录都包括项目和属于该项目的项目值，包括下列步骤：

a) 通过将表格格式数据划分为信息块来构造表格格式数据，每一个信息块都包括值列表和指针阵列，在所述值列表中，项目值以对应于属于特定项目的项目值的项目值号码的顺序存储，在所述指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储，

b) 响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列，

c) 将在由有序集合的每一个记录号指出的位置上的指针阵列中的指针值对应于其中排列了记录号的位置排列到项目值号码阵列中，以及

e) 使用项目值号码阵列作为第二值列表并使用原始值列表作为第二值列表，

其中，通过由记录号指出的位置上的第二指针阵列中的元素和由第二指针阵列中的元素指出的位置上的第二值列表中的元素，根据有序集合阵列的记录号指定值列表中的值。

在此实施例中，由于第二值列表为原始值列表，因此，其量不能被缩小。然而，与有序集合阵列具有一样大小的项目值号码阵列被用作第二指针阵列。相应地，与使用原始指针阵列和原始值列表的情况相比，可以使对其进行引用的数据量比较小，因此，可以缩短处理时间。

优选实施例进一步包括将有序集合阵列的大小与值列表的大小进行比较的步骤，作为比较的结果，在子集的大小按预先确定的比率小于值列表的大小的情况下，执行步骤 c) 和 d) 或步骤 c) 和 e)。

这里，在甚至在考虑到第二指针阵列和第二值列表的创建的处理时间的情况下识别缩短处理时间的情况下，创建第二指针阵列和第二值列表，并执行使用这些东西的处理。例如，预先确定的速率通过(子集的大小，即，记录的数量)/(值列表的大小，即，记录的数量)= $1/n$

( $n$  是大于 1 的数字)。数字“ $n$ ”可以通过由操作员输入的所希望的数字确定，或者可以预先确定为预先确定的数字。

此外，本发明的目的通过使用上面的数据处理方法创建的第二指针阵列和/或第二值列表的检索方法来实现，该方法的特征在于，包括下列步骤：在第二值列表中的元素中指定指出项目值的元素，作为检索对象，将指出一个标志在一个位置为  $on$  的值组成阵列，所述位置对应于与第二值列表具有同样大小的标志阵列中的指定元素的位置，指定由有序集合中的记录号指出的第二指针阵列的元素，引用由第二指针阵列的元素指出的位置处的标志阵列中的标志的状态，以及在标志的状态为  $on$  的情况下，连续地将记录号排列到新提供的用于输出的有序集合阵列中。

在另一个实施例中，使用上面的数据处理方法创建的第二指针阵列和/或第二值列表的聚合方法包括下列步骤：创建分类号阵列，在该阵列中，对应于第二值列表的元素，排列指出值的类别的分类号，指定对应于指定的分类号的有序集合阵列中的记录号，以及使用由指定的记录号指出的预先确定的值列表项目值执行聚合。

此外，在另一个实施例中，使用上面的数据处理方法创建的第二指针阵列和/或第二值列表的分类方法，包括下列步骤：作为第二值列表的每一个值的第二指针阵列指出的元素的数量，计算存在号码，基于存在号码，创建对应于第二值列表的值并且指出其中有序集合阵列中的记录号将被排列的头部位置的累计号码阵列，以及，引用累积阵列，并将有序集合阵列的记录号排列到用于输出的阵列，以便反映值列表中的项目值的分类顺序。

此外，在另一个实施例中，通过使用涉及由上面的数据处理方法创建的多个表格格式数据中的每一个数据的信息块中的第二指针阵列和/或第二值列表，联接表格格式数据的方法，包括下列步骤：查找要在多个表格格式数据中的每一个数据中共享的项目，使涉及项目的信息块的第二值列表中的项目值相等，以及，响应使项目值相等，随着项目值的配置中的变化，更新每一个信息块中的第二指针阵列中的元

素。

此外，本发明的目的通过一种用于操作计算机以从被表达为记录阵列的表格格式数据中提取子集作为处理对象的数据处理程序来实现，每一个记录都包括项目和属于该项目的项目值，使计算机执行下列步骤：

a) 通过将表格格式数据划分为信息块来构造表格格式数据，每一个信息块都包括值列表和指针阵列，在所述值列表中，项目值以对应于属于特定项目的项目值的项目值号码的顺序存储，在所述指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储，

b) 响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列，

c) 将在由有序集合的每一个记录号指出的位置上的指针阵列中的指针值对应于其中排列了记录号的位置排列到项目值号码阵列中，以及

d) 通过引用项目值号码阵列中的值，创建存储位置元素的第二指针阵列和存储值元素的第二值列表，以及其中

操作计算机，以便通过由记录号指出的位置上的第二指针阵列中的元素和由第二指针阵列中的元素指出的位置上的第二值列表中的元素，根据有序集合阵列的记录号指定值列表中的值。

在优选实施例中，在创建第二指针阵列和第二值列表的步骤中，使计算机执行下列步骤：将项目值号码阵列中的元素分类，在重复值被排除的状态下创建阵列，并使其成为第二值列表；以及通过转换项目值号码阵列中的元素以反映分类和重复值的排除，创建第二指针阵列。此外，在另一个优选实施例中，一种用于操作计算机以从被表达为记录阵列的表格格式数据中提取子集作为处理对象的数据处理程序，每一个记录都包括项目和属于该项目的项目值，使计算机执行下列步骤：

a) 通过将表格格式数据划分为信息块来构造表格格式数据，每一个信息块都包括值列表和指针阵列，在所述值列表中，项目值以对

应于属于特定项目的项目值的项目值号码的顺序存储，在所述指针阵列中，指出项目值号码的指针值以唯一记录号的顺序存储，

b) 响应对作为记录的一部分的子集的选择，创建包含子集的记录号的有序集合阵列，

c) 将在由有序集合的每一个记录号指出的位置上的指针阵列中的指针值对应于其中排列了记录号的位置排列到项目值号码阵列中，以及

e) 使用项目值号码阵列作为第二值列表并使用原始值列表作为第二值列表，

以及使计算机被操作，以便通过由记录号指出的位置上的第二指针阵列中的元素和由第二指针阵列中的元素指出的位置上的第二值列表中的元素，根据有序集合阵列的记录号指定值列表中的值。

优选情况下，使计算机进一步执行将有序集合阵列的大小与值列表的大小进行比较的步骤，作为比较的结果，在子集的大小按预先确定的比率小于值列表的大小的情况下，使计算机执行步骤 c) 和 d) 或步骤 c) 和 e)。

顺便说一句，在本说明书中，虽然各种阵列中包含的数值被称为“元素”或“值”，但是，它们只是表示方式的不同，而没有本质的区别。例如，“值列表”中包含的东西被表达为“项目值”，显然，这也是值列表中的“元素”。

#### 附图说明

通过参考实施例以及附图，本发明的目的以及其他目标将变得显而易见。这里，

图 1 是显示表格格式数据的示例和其子集的示例的视图。

图 2 是显示基于图 1 的表格格式数据的信息块的示例的视图。

图 3 是显示执行实现本发明的实施例的过程的程序的计算机系统的硬件结构的视图。

图 4 是显示由实施例中的计算机系统执行的过程的概述的流程图。

- 图 5 是显示实施例的初步处理的流程图。
- 图 6 是说明实施例的项目值号码的创建的视图。
- 图 7 是显示实施例的编译过程的示例的流程图。
- 图 8 是说明实施例的编译过程的视图。
- 图 9 是显示实施例的检索过程的流程图。
- 图 10 是说明实施例的检索过程的视图。
- 图 11 是说明实施例的检索过程的视图。
- 图 12 是说明实施例的检索过程的视图。
- 图 13 是显示实施例的分类过程的流程图。
- 图 14 是显示实施例的分类过程的流程图。
- 图 15 是显示实施例的分类过程的视图。
- 图 16 是显示实施例的分类过程的视图。
- 图 17 是显示实施例的分类过程的视图。
- 图 18 是显示实施例的分类过程的视图。
- 图 19 是显示实施例的分类过程的视图。
- 图 20 是说明实施例的联接的概述的视图。
- 图 21 是显示实施例中的联接表（视图）的示例的视图。
- 图 22 是显示实施例中的要联接的表格格式数据的信息块的视图。
- 图。
- 图 23 是说明实施例的联接过程的视图。
- 图 24 是说明实施例的联接过程的概述的流程图。
- 图 25 是显示实施例的值列表共享过程的流程图。
- 图 26 是说明实施例的联接过程的视图。
- 图 27 是说明实施例的联接过程的视图。
- 图 28 是说明实施例的联接过程的视图。
- 图 29 是说明实施例的联接过程的视图。
- 图 30 是说明实施例的联接过程的视图。
- 图 31 是说明实施例的联接过程的视图。
- 图 32 是说明实施例的联接过程的视图。

- 图 33 是说明实施例的联接过程的视图。
- 图 34 是说明实施例的联接过程的视图。
- 图 35 是说明实施例的联接过程的视图。
- 图 36 是说明实施例的联接过程的视图。
- 图 37 是显示实施例的联接过程中的记录号计算过程的流程图。
- 图 38 是说明实施例的记录计算过程的视图。
- 图 39 是说明实施例的记录计算过程的视图。
- 图 40 是说明实施例的记录计算过程的视图。
- 图 41 是说明实施例的记录计算过程的视图。
- 图 42 是说明实施例的记录计算过程的视图。
- 图 43 是说明实施例的记录计算过程的视图。
- 图 44 是说明实施例的记录计算过程的视图。
- 图 45 是显示由实施例的联接过程最后获得的表(视图)的视图。

#### 具体实施方式

下面将参考附图描述本发明的实施例。图 3 是显示计算机系统的硬件结构的方框图，该计算机系统用于实现根据本发明的实施例的作为表格格式数据中的处理对象为子集创建转换阵列的过程，以及检索、聚合、分类和联接的过程。在此实施例中，在检索、聚合、分类并联接的过程之前，参考如图 2(a) 到 2(c) 所示的信息块，创建只包括值列表的记录和元素的阵列，该阵列成为处理对象。此过程被称为“初步处理”。接下来，使由初步处理创建的阵列成为工作区，执行检索、聚合等等各种过程。

如图 3 所示，此计算机系统 10 与普通计算机系统具有相同结构，并包括 CPU 12，用于通过程序来控制整个系统和单个结构部分，RAM（随机存取存储器）14，用于存储工作数据等等，ROM（只读存储器）16，用于存储程序等等，诸如硬盘之类的固定存储介质 18，CD-ROM 驱动器 20，用于访问 CD-ROM 19，在 CD-ROM 驱动器 20 和连接到外部网络(未显示)的外部终端之间提供的接口 (I/F) 22，包括键盘和鼠标的输入设备 24，以及 CRT 显示设备 26。CPU 12、

RAM 14、ROM 16、外部存储介质 18、I/F 22、输入设备 24，以及显示设备 26 通过总线 28 相互连接。

根据实施例的检索、聚合、分类、联接等等的预处理程序以及程序包含在 CD-ROM 19 中，并可以由 CD-ROM 驱动器 20 读取，或者可以预先存储在 ROM 16 中。此外，曾经从 CD-ROM 19 中读出的内容可以存储在外部存储介质 18 的预先确定的区域。或者，程序可以通过网络（未显示）、外部终端和 I/F 22 从外部提供。

此外，在实施例中，为了实现高速度的检索、聚合、分类和联接的预处理，以及处理，必须创建图 2(a) 到 2(c) 所示的数据格式的信息块。信息块创建程序也可以包含在 CD-ROM 19 中，存储在 ROM 16 中，或存储在外部存储介质 18 中。或者，不用说，这些程序也可以通过网络（未显示）从外部提供。此外，在实施例中，由信息块创建程序创建的数据（信息块）存储在 RAM 14 中，或存储在外部存储介质 18 的预先确定的区域中。

接下来，将描述由计算机系统 10 执行的处理的概述。如图 4 所示，在实施例中，根据检索、聚合或分类的过程（其中，所有记录或其组成部分是处理对象），限制性较强地选择记录（步骤 401）。此外，使具有所选择的记录的集合成为处理对象，当开始检索、聚合、分类或联接的过程时（步骤 402），在明确的条件下，执行初步处理（步骤 404）。此初步处理在某些情况下被称为“范围”的设置过程，在执行了初步处理之后要执行的检索等等的过程被称为“使用范围”的检索等等的过程（步骤 405）。

顺便说一句，具有所选择的记录的集合被称为“有序集合”。在有序集合的阵列中，所选择的记录号以指定的顺序来排列（例如，升序）。在初步处理中，创建第二指针阵列 (PV2) 和第二值列表 (VL2)，它们用于检索等等的后处理。稍后将描述此创建过程。第二指针阵列包含等于有序集合的元素数量的元素，并指出对应于有序集合的阵列中的元素的第二值列表中的元素。第二值列表包含值列表 (VL) 的存储位置号码。第二值列表的元素数量不大于有序集合的阵列中的元素数量。



即，只有在对应于有序集合的阵列中的元素（所选择的记录）的所有项目值彼此不同的情况下，这些元素的数量彼此一致。相应地，有序集合的阵列中的元素数量成为第二值列表的元素数量可以采取的最大值。

然后，在此实施例中，在原始值列表 (VL) 的元素数量 (大小) 充分大于有序集合的阵列中的元素数量 (大小) 的情况下 (参见步骤 403)，执行使用此实施例中的第二指针阵列和第二 VL 的过程，即，使用范围的检索等等的过程 (步骤 405)。例如，使有序集合的大小不大于 VL 的大小的  $1/n$  ( $n$ : 不小于 1 的集合值)。此值“ $n$ ”可以根据记录数量、处理器的处理能力等等来进行设置。

由于下列原因而执行如上所述的判断。在使用范围的过程中，需要初步处理 (步骤 404)，其执行需要花费一些时间。如此，只有在有序集合的大小 (实际上，第二值列表的大小) 充分小于 VL 的大小的情况下，可以想象，甚至在初步处理的执行时间增大的情况下，与正常的处理 (步骤 406) 的执行相比，可以缩短整个处理时间，即，只有在可以缩短处理时间的情况下，与正常的处理相比，通过使用初步处理和范围执行检索等等的过程，应该执行步骤 404 和步骤 405 的处理。

相应地，陈述“充分小”意味着，其小的程度致使，即使执行初步处理，并且绝对数的大小没有定义，也可以缩短处理时间。

接下来，将描述实施例的初步处理。图 5 是显示此实施例的初步处理的流程图。如图 5 所示，在初始化指定有序集合的阵列中的位置 (存储位置号码) 的参数之后 (步骤 501)，指定由参数指出的有序集合的阵列中的元素 (步骤 502)。最初，在图 6 中，指定有序集合的阵列 601 中的存储位置号码“0”中的元素“0”。

接下来，通过引用由有序集合的阵列中的元素指出的指针阵列中的元素，指定由元素指出的位置上的项目值的号码 (项目值号码) (步骤 503)。在初始状态的示例中，指定由有序集合的阵列中的元素“0”指出的指针阵列中的元素“2”。由于此元素表示指出值列表中的相关元

素（项目值）的位置的号码，因此，这被称为“项目值号码”。

此项目值号码位于由单独提供的项目值号码阵列中的参数指出的位置（步骤 504）。在初始状态的示例中，指针阵列 212 中的元素（项目值号码）“2”存储在由项目值号码阵列 611 中的参数“0”指出的位置。顺便说一句，项目值号码阵列包含其数量等于有序集合的阵列的数量的元素。

通过重复如上所述的过程（参见步骤 505 和 506），可以创建包含有序集合中的每一个元素（所选择的记录号）的对应的项目值号码的项目值号码阵列（参见图 6 中的附图标记 611）。

当创建项目值号码阵列的过程完成时，执行编译过程（步骤 507）。在此实施例中，在记录的子集是处理对象的情况下，只对与子集相关的值列表中的项目值进行参考。相应地，由于作为处理对象的范围是有限的，因此，在本说明书中，它也被称为“范围”。

为了实现该范围，需要值列表的子集（部分集合中的元素要引用该子集）以及记录的子集中的元素（记录号）的转换阵列和值列表的子集。前者是第二值列表 (VL2)，而后者是第二指针阵列 (PV2)。创建第二值列表和第二指针阵列的过程被称为编译过程。

图 7 是显示编译过程的示例的流程图。在此示例的编译过程中，首先，创建项目值号码的副本（步骤 701）。在涉及如图 6 所示的信息块 201 和有序集合的阵列 601 作为记录的子集的示例中，如图 8(a) 所示，创建包括与项目值号码阵列（原始阵列）相同元素的副本阵列。

接下来，在对副本阵列进行分类之后（参见步骤 702 和图 8(b)），排除经过分类的副本阵列中的重复值，并浓缩阵列（参见步骤 703 和图 8(c)）。如此，其中元素被分类并且重复值被排除的阵列成为第二值列表 (VL2)。然后，指定项目值号码阵列（原始阵列）中的每一个元素在第二值列表中的位置，并将指出指定的位置的值作为原始阵列的存储位置号码的元素存储到新准备的阵列中（参见步骤 704 和图 8(d)）中。其中在步骤 704 中存储元素的阵列成为第二指

针阵列 (PV2)。

可以想象，在项目值号码阵列中，值的组件和位置的组件是混合的。然后，在步骤 701 到 703 的过程中，只提取值的组件，并创建只包括被提取的值的组件的第二值列表 (VL2)。此外，在随后的步骤 704 中，提取位置的组件，并创建只包括位置的组件的第二指针阵列 (PV2)。

使用编译过程如此创建的第二值列表和第二指针阵列作为信息块的一部分，用于稍后描述的检索、聚合、分类、联接等等的过程。

接下来，将描述使用第二值列表和第二指针阵列（使用范围）的检索过程。图 9 是显示根据此实施例的使用范围的检索过程的流程图。在此过程中，首先，确保与第二值列表 (VL2) 具有一样大小的标志阵列区域（步骤 901）。在此标志阵列中，“1”与位置（存储位置）位于相同的位置，在该位置中，存储值列表的元素（项目值）之中将成为检索对象的元素，而“0”位于另一个位置。相应地，在下一个步骤中，“1”或“0”的值位于标志阵列中。

图 10 显示了一个示例，在该示例中，对于所选择的记录（有序集合）的项目“fan”，检索具有项目值“T-team”的记录。这里，指定了第二值列表 (VL2) 的元素，判断由该元素指出的值列表 (VL) 中的元素（项目值）是否为检索对象。当项目值是检索对象时，在标志阵列中，“1”位于对应于第二值列表 (VL2) 的元素的存储位置的位置。在图 10 的示例中，由于项目值“T-team”是检索对象，值“1”位于标志阵列中的指出第二值列表 (VL2) 中的值列表 (VL) 的项目值“T-team”的元素的存储位置（存储位置号码“1”）。

如此，当值位于标志阵列中时，在初始化指出有序集合的阵列中的位置（存储位置号码）的参数之后（步骤 903），指定由参数指出的有序集合的阵列中的元素（步骤 904）。在图 11 的示例中，最初，在有序集合的阵列中，指定位于存储位置号码“0”的元素“0”。接下来，指定由元素指出的第二指针阵列 (PV2) 中的元素，然后，指定由 PV2 中的元素指出的标志阵列中的值（步骤 905）。在初始状态的示例中，

指定，位于由第二指针阵列 (PV2) 元素“0”指出的位置的标志阵列的值为“0”。

接下来，判断指定的标志值是否为“1”（步骤 906）。当标志值为“1”（在步骤 906 中为“是”），有序集合的阵列的元素位于提供的用于输出的阵列中（步骤 907）。在图 11 的示例中，在有序集合的阵列中，对于存储位置号码“1”的元素“2”，通过 PV2 和标志配置，标志阵列中的标志值成为“1”。相应地，有序集合的阵列的元素“2”存储在用于输出的阵列中。

对于有序集合的存储位置号码的所有参数，执行如上所述的过程（参见步骤 908 和 909）。在用于输出的阵列中获得的元素分别成为通过检索提取的记录号。

如上所述，根据此实施例，指出值列表中的元素的信息（值列表的存储位置号码）（使该信息对应于作为记录的子集的有序集合中的元素）存储在第二值列表 (VL2)。在检索过程中，创建与此大小具有相同大小的标志阵列，参考标志阵列中的值，并就它是否为检索对象的记录作出判断。由于标志阵列的大小等于第二值列表的大小，与涉及所有记录的值列表的大小相比，它足够小。相应地，创建标志阵列以及对值进行参考所需的时间可以缩短，由于此，可以足够地缩短检索过程的处理时间。

接下来，将描述使用范围的聚合过程。聚合过程用于计算具有一个或多个特殊条款值的记录数量。在此实施例中，提供了与第二值列表具有同样大小的存在号码阵列，与第二值列表 (VL2) 中的项目值相关的记录数量存储在存在号码阵列，提取存在号码阵列中的元素，以便可以立刻实现聚合。

图 12 是显示了涉及项目“fan”的信息块中的第二指针阵列 (PV2)、第二值列表 (VL2) 和存在号码阵列的视图。在 PV2 中，存储位置号码“0”和“2”中的两个元素与 VL2 的存储位置号码“0”一致。相应地，在存在号码阵列中，位于存储位置号码“0”的元素成为“2”。另一方面，在 PV2 中，位于存储位置号码“1”的一个（单个）元素与

VL2 的存储位置号码“1”一致。相应地，在存在号码阵列中，位于存储位置号码“1”的元素成为“1”。顺便说一句，在接下来描述的分类过程中将详细描述了创建存在号码阵列的过程。

根据此实施例，也是聚合过程中，可以使要使用的存在号码阵列的大小与第二值列表一样大小，该列表充分小于原始值列表。相应地，可以缩短创建存在号码阵列所需的时间，由于此，可以缩短整个处理时间。

接下来，将描述使用范围分类过程的流程图。图 13 和 14 是显示使用范围分类过程的流程图。如图 13 所示，首先，创建初始存在号码阵列，该阵列与第二值列表 (VL2) 具有一样大小，其中，排列了初始值“0”(步骤 1301)。接下来，在初始化指定第二指针阵列 (PV2) 的位置 (存储位置号码) 的参数之后 (步骤 1302)，在有序集合的阵列中，指定了由参数指出的位置上的元素，接下来，在存在号码阵列中，增大由元素指出的位置的值 (步骤 1303)。重复步骤 1303，直到最后的参数 (参见步骤 1304 和 1305)。到在步骤 1305 中的判断为“是”的阶段，完成存在号码阵列。

图 15 和 16 是显示了稍后描述的涉及项目“fan”和累计号码阵列的存在号码阵列的创建过程的示例的视图。如图 15(a) 所示，最初，指定位于参数“0”的位置处的 PV2 的元素“0”，在存在号码阵列中，将位于由 PV2 的元素指出的位置处的值从“0”增大到“1”。接下来，如图 15(b) 所示，指定位于参数“1”的位置处的 PV2 的元素“1”，在存在号码阵列中，将位于由 PV2 的元素指出的位置处的值从“0”增大到“1”。然后，如图 16(a) 所示，指定位于参数“2”的位置处的 PV2 的元素“0”，在存在号码阵列中，将位于由 PV2 的元素指出的位置处的值从“1”增大到“2”。

当完成存在号码阵列时，基于存在号码阵列，创建存储累计号码 (其中累积了元素) 的累计号码阵列 (步骤 1306)。累计号码阵列指出对于在某些存储位置号码之前的存储位置号码总共存在多少元素。在图 16(b) 的示例中，在累计号码阵列 1612 中，位于存储位置号码

“0”的元素成为“0”，由于没有较小的存储位置号码。位于存储位置号码“1”的元素等于小于此号码的存储位置号码“0”的（存在号码阵列 1611 中的元素）的存在号码，并成为“2”。此外，在此累计号码阵列中，作为另一个元素（位于存储位置号码“2”的元素），可以提供小于存储位置号码“2”的存储位置号码的存在号码的总和“3”（参见附图标记 1612 的括号）。

当如此创建累计号码阵列时，通过使用累计号码阵列，在新确保的用于输出的阵列中排列记录号。如图 14 所示，首先，在初始化指出有序集合的阵列中的位置（存储位置号码）的参数之后（步骤 1401），指定由参数指出的有序集合的阵列中的元素，接下来，在第二指针阵列（PV2）中，指定了由指定的元素指出的位置处的元素（步骤 1402）。

接下来，在累计号码阵列中，指定由第二指针阵列（PV2）中的指定元素指出的位置（步骤 1403）。然后，在单独提供的用于输出的阵列中在步骤 1403 中指定的累计号码指出的位置，排列由参数指出的有序集合的阵列中的元素（步骤 1404）。此后，在累计号码阵列中，增大在步骤 1403 中指定的累计号码（步骤 1405）。

在步骤 1403 中指定的累计号码指出作为指定累计号码的来源的有序集合的阵列中的元素，即，记录号存储在用于输出的阵列中的位置。相应地，在排列了某一记录号之后，接下来指定同一个累计号码，以便在步骤 1405 中将记录号存储在用于输出的阵列中，增大累计号码。此外，由于存储有序集合的阵列中的元素，用于输出的阵列与有序集合的阵列具有一样的大小。

重复如上所述的过程，直到最后的参数（参见步骤 1406 和 1407）。在如此获得的用于输出的阵列中，对于作为键的项目，指出分类的记录号。

图 17 到 19 是显示分类过程中的阵列的状态的视图，此时，使项目“fan”成为键。在图 17 中，最初，指定由参数“0”指出的有序集合阵列的元素（记录号）“0”。相应地，通过 PV2 指定由 PV2 中的相关元素指出的位置上的累计号码阵列的值（累计号码）。相应地，

有序集合的阵列的元素“0”存储在用于输出的阵列中的由累计号码指出的存储位置号码“0”的位置，然后，累计号码从“0”增大到“1”。

也是在图 18 中，同样，由参数“1”指出的有序集合的阵列的值“2”存储在用于输出的阵列中的通过 PV2 获得的累计号码阵列中的对应于累计号码“2”的位置，然后，累计号码阵列中的累计号码从“2”增大到“3”。此外，也是在图 19 中，同样，由参数“2”指出的有序集合的阵列的值“9”存储在用于输出的阵列中的通过 PV2 获得的累计号码阵列中的对应于累计号码“1”的位置，然后，累计号码阵列中的累计号码从“1”增大到“2”。

如上所述，根据此实施例，也是在分类的过程中，可以使用于该过程的累计号码阵列与第二值列表具有一样的大小，由于此，可以缩短处理时间。

接下来，将描述使用范围的联接过程。联接过程是指，从多个表格中，预先确定的项目被用作键，并提供多个表格之间的共同性，以联接它们。例如，考虑这样的情况：从图 20(a) 所示的棒球爱好者俱乐部成员的源表（原始表）2001，提取项目“区域”的值为“Tokyo”的记录，并准备表格（棒球爱好者俱乐部成员表格），如图 20(b) 所示，从棒球比赛时间表的原始表 2011 中，提取项目“棒球场”的值为“东京”的记录，并准备表格（比赛时间表）2012。在联接过程中，棒球爱好者俱乐部成员表格 2011 中的项目“区域”和比赛时间表 2012 的项目“棒球场”被用作键，创建其中键项目被联接的表，如图 21 所示。这里，反映（保留）了分类顺序的表格 2101 被称为主端表格，而另一个表格 2111 被称为从端表格。

在联接过程中，在主端和从端，通过使用项目作为键，提取具有预先确定的值的记录，并创建表格。相应地，当提取记录时，在原始值列表的元素的数目（大小）充分大于有序集合的阵列中的元素的数目（大小）的情况下，执行使用范围的联接过程。

根据此实施例，比赛时间表原始表具有如图 22 所示的结构。即，对于项目“棒球队”、“比赛日期”和“棒球场”中的每一个项目，提供包

括值列表的指针阵列 (PV) 和值列表 (VL) 的信息块。此外, 如上所述, “棒球场”的值为“Tokyo”的记录具有记录号“0”、“4”、“9”和“10”。相应地, 其中上面的记录号按照升序排序的阵列成为有序集合的阵列。

图 23 是用于说明涉及项目“棒球队”的信息块的初步处理(参见图 5)的过程的视图。通过图 5 中的步骤 501 到 505 的处理, 创建项目值号码阵列, 然后, 通过编译过程(参见图 5 和图 7 的步骤 507), 创建第二值列表 (VL2) 和第二指针阵列 (PV2)。

图 24 是显示使用第二值列表和第二指针阵列的联接过程(即, 使用范围的联接过程)的流程图。在联接过程中, 首先, 作为键共享项目的值列表(步骤 2401)。图 25 是比较详细地显示值列表共享过程的流程图。在值列表共享过程中, 首先, 将主端值列表中的值与从端值列表中的值进行比较, 并判断哪个值比较小(例如, 在字母表中, 哪一个位于前面)(步骤 2501)。

图 26 是显示了在编译之后构成主端表格的项目“fan”的信息块和构成从端表格的项目“棒球队”的信息块的初始状态的视图。如图 26 所示, 对于每一个信息块, 提供共享之后的值列表 (NewVL) 的指示值的新的第二值列表(新的 VL2), 在第二值列表中, 排列指出元素的位置作为处理对象(即, 将提供值的位置)的指针 (PUT 指针)。由此指针指出的位置被称为“PUT 位置”。同样, 也准备指出在共享之后值位于值列表中的位置的指针 (PUT 指针)。

最初, 将主端第二值列表 (VL2) 的头元素“2”指出的主端值列表 (VL) 的值“G-team”与由从端第二值列表 (VL2) 的头元素“2”指出的从端值列表 (VL) 的值“D-team”进行比较, 并判断是否以字母顺序建立了“D-team < G-team”(参见图 27)。

接下来, 从比较的结果被判断为小的值(在图 27 中, “D-team”)存储在共享之后的值列表 (newVL) 中的“PUT 指针”指出的位置 (PUT 位置)。接下来, 在新的第二值列表(在图 27 中, 从端“新的 VL2”)(在步骤 2502 中将值存储在共享之后的值列表中), 共享之后的值列表 (newVL) 的“PUT 位置”(在图 27 中, “0”)存储



在由“PUT 指针”指出的位置 (PUT 位置) (步骤 2503)。由于此, 其中存储“PUT 位置”的新的第二值列表 (新的 VL2) 指出共享之后的值列表 (newVL) 的对应的元素。

接下来, 其中存储“PUT 位置”的第二值列表 (在图 27 中, 从端 VL2) 和共享之后的值列表 (newVL) 的“PUT 位置”增大 (步骤 2504)。

重复如上所述的过程, 直到主端第二值列表 (VL2) 的末尾和从端第二值列表 (VL2) 的末尾 (参见步骤 2505 和步骤 2506)。在图 27 的示例之后, 如图 28 所示, 提取主端值列表 (VL2) 的头元素“2”指出的值列表“VL”的元素“G-team”, 另一方面, 提取从端第二值列表 (VL2) 的第二元素“3”指出的值列表“VL”的元素“G-team”。当将两者彼此进行比较时, 由于“G-team = G-team”, “G-team”存储在共享之后的值列表 (newVL) 的“PUT 位置”, 共享之后的值列表 (newVL) 的“PUT 位置”存储在主端和从端的新的第二值列表 (新的 VL2) 的“PUT 位置”。此后, 共享之后的每一个值列表 (新的 VL2、newVL) 中的每一个值列表的“PUT 位置”增大。

此后, 如图 29 所示, 提取主端第二值列表 (VL2) 的第二元素“3”指出的值列表“VL”的元素“T-team”, 另一方面, 提取从端第二值列表 (VL2) 的第三元素“5”指出的值列表“VL”的元素“T-team”。这里, 由于“T-team = T-team”, “T-team”存储在共享之后的值列表 (newVL) 的“PUT 位置”, 共享之后的值列表 (newVL) 的“PUT 位置”存储在主端和从端的新的第二值列表 (新的 VL2) 的“PUT 位置”。此后, 每一个阵列 (新的 VL2、newVL) 中的每一个阵列的“PUT 位置”增大。如此, 主端和从端新的第二值列表 (新的 VL2) 和共享之后的值列表 (NewVL) 完成。

顺便说一句, 在实践中, 新的第二值列表 (新的 VL2) 区域不是单独创建的, 但是, 第二值列表 VL2 可以在处理的过程中被覆盖。此外, 为便于说明, 确保共享之后的值列表 (newVL) 的区域, 并排列值, 然而, 这不是不可缺少, 如果只获取共享之后的值列表的大小

并存储在存储器等等中，就足够了。

当结束值列表的共享时，连续地创建从端存在号码阵列、主端累计号码阵列，以及从端累计号码阵列（步骤 2403）。下面将参考图 31 到 34 比较详细地描述此处理。首先，指定由从从端第二指针阵列 (PV2) 提取的元素指出的第二值列表（新的 VL2）中的值。接下来，由存在号码阵列中的第二值列表 (newVL2) 的值指出的位置上的值增大。

在图 31(a) 中，指定“PV2”的头元素“2”指出的“新的 VL2”的值“2”，接下来，在存在号码阵列中，存储位置号码“2”中的值从“0”增大到“1”。同样，对于“PV2”的第二元素“1”，在存在号码阵列中，存储位置号码“1”中的值从“0”增大到“1”（参见图 31(b)）。对于“PV2”的第三元素“0”，在存在号码阵列中，存储位置号码“0”中的值从“0”增大到“1”（参见图 32(a)）。对于“PV2”的第四元素“1”，在存在号码阵列中，存储位置号码“1”中的值从“1”增大到“2”（参见图 32(b)）。如此，可以获得从端存在号码阵列的最终形式（参见附图标记 3211）。

接下来，通过使用从端存在号码阵列，创建主端存在号码阵列，以及累计号码阵列（其中，累积存在号码阵列中的元素（存在号码））。主端存在号码阵列的元素（存在号码）指出由主端第二指针阵列指定的值列表中的多少值存在于从属阵列中。

如图 33 所示，由主端第二指针阵列 (PV2) 的头元素“0”指定的新的第二值列表（新的 VL2）的值为“1”。通过使用此，当对从端存在号码阵列进行参考时，由值“1”指出的位置处的元素变成“2”。相应地，“2”作为存在号码阵列的头元素来排列。此外，由主端第二指针阵列 (PV2) 的第二元素“1”指定的新的第二值列表（新的 VL2）的值为“2”。当通过使用此对从端存在号码阵列进行参考时，由值“2”指出的位置处的元素变成“1”。相应地，“1”作为存在号码阵列的第二元素来排列。同样，由主端第二指针阵列 (PV2) 的第三元素“0”指定的新的第二值列表（新的 VL2）的值为“1”。当通过使用此对从端存在号码阵列进行参考时，由值“1”指出的位置处的元素变成“2”。相应地，“2”

作为存在号码阵列的第三元素来排列。

接下来，基于主端存在号码阵列，创建累计号码阵列。主端累计号码阵列在某些情况下被称为“投影阵列 (1)”。在第四位置中排列的值“5”指出联接的表格中的所有记录的数量。此外，如图 34 所示，基于从端存在号码阵列，创建累计号码阵列。从端累计号码阵列在某些情况下被称为“投影阵列 (3)”。此外，主端有序集合在某些情况下被称为“投影阵列 (2)”。

如上所述，在创建各种存在号码阵列和累计号码阵列之后，对其中排列了从端记录号的有序集合（步骤 2403）。将参考图 35 和 36 比较详细地描述有序集合的分类。从从端有序集合（即，具有从端记录号的集合）提取元素（即，记录号），随后，指定由元素（在图 35(a) 中，“0”）指出的第二指针阵列 (PV2) 的元素（在图 35(a)，“2”）。此外，指定由 PV2 的元素指出的新的第二值列表（新的 VL2）的值（在图 35(a) 中，“2”）。接下来，在累计号码阵列中，提取由新的 VL2 的值指出的位置处的值（累计号码）（在图 35(a)，“3”），在由累计号码指出的位置处的分类之后，在有序集合的阵列中提供了有序集合的元素（记录号）。此后，累计号码增大（例如，在图 35(a) 中，从“3”到“4”）。

如图 35(b) 所示，也对于有序集合的阵列的第二元素“4”，同样，在累计号码阵列中通过第二指针阵列 (PV2) 和新的第二值列表(新的 VL2) 指定第二（存储位置号码“1”）元素（累计号码）“1”。相应地，在分类之后的有序集合的阵列中，在由累计号码“1”指出的位置排列原始有序集合的阵列的元素“4”。此后，累计号码从“1”增大到“2”。

同样，如图 36(a) 所示，对于有序集合的阵列的第三元素“9”，在累计号码阵列中通过第二指针阵列 (PV2) 和新的第二值列表(新的 VL2) 指定第一（存储位置号码“0”）元素（累计号码）“0”。相应地，在分类之后的有序集合的阵列中，在由累计号码“0”指出的位置排列原始有序集合的阵列的元素“9”，然后，累计号码从“0”增大到“1”。此外，如图 36(b) 所示，对于有序集合的阵列的第四元素“10”，在累计号码

阵列中通过第二指针阵列 (PV2) 和新的第二值列表 (新的 VL2) 指定第二 (存储位置号码“1”) 元素 (累计号码) “2”。相应地, 在分类之后的有序集合的阵列中, 在由累计号码“2”指出的位置排列原始有序集合的阵列的元素“10”, 然后, 累计号码从“2”增大“3”。如此, 完成从端记录号的分类。

接下来, 计算主端表格的记录号和对应于联接的表格的记录号的从端表格。联接的表格的记录号的总数 (记录数目) 与在主端累计号码阵列 (投影阵列 (1)) 结尾处排列的值一致。相应地, 在参考图 31 到 36 中说明的示例中, 记录的数量变成“5”。

图 37 和 38 是显示从联接的表格的记录号计算对应的主端和从端记录号的过程 (记录号计算过程) 的流程图。此外, 图 39 是显示通过对于图 33 到 36 所示的示例执行图 37 和 38 所示的过程获得的主端记录号阵列和从端记录号阵列的视图。

下面, 使用在图 33 到 36 中获得的投影阵列并使用获取图 39 所示的记录号的阵列的情况作为示例, 详细描述记录号计算过程。

首先, 提取联接的表格的记录号 (步骤 3701)。此记录号在记录号计算过程中被称为“值 A”。最初, 如图 40 所示, 从联接的表格的记录号阵列中提取头元素“0”。

接下来, 在主端累计号码阵列 (投影阵列 (1)) 中, 指定在步骤 3701 中提取的不大于元素 (记录号) 的最大值以及其存储位置 (步骤 3702)。此最大值在记录号计算过程中被称为“值 B”。在图 40 的示例中, 在投影阵列 (1) 中, 由于不大于记录号“0”的最大值为“0”, 则“值 B”为“0”, 指定其存储位置为“0”。

此后, 在主端有序集合阵列 (投影阵列 (2)) 中, 指定存储位置的元素 (主端记录号), 在主端记录号阵列中的对应于联接表格中的存储位置的位置排列此元素 (记录号) (步骤 3703)。在图 40 的示例中, 在投影阵列 (2) 中, 提取存储位置“0”处的值“0”, 即, 头值, 并且这将作为如图 39 所示的主端记录号阵列的头元素来存储 (参见附图标记 3901)。如此, 计算对应于联接表格的记录号的主端记录号。

接下来，取“值 A”和“值 B”之间的差（步骤 3704）。在记录号计算过程中，此差值被称为“值 C”。在图 40 的示例中，“值 C=0”（参见 4010）。

此外，在主端第二值列表 (PV2) 中，指定在步骤 3702 中指定的存储位置处的元素，此外，指定由 VL2 的元素指出的位置处的新的第二值列表（新的 VL2）中的值（步骤 3706）。在图 40 的示例中，在 PV2 中，指定存储位置“0”处的元素“0”，接下来，在新的 VL2 中，提取存储位置“0”处的值“1”。

此外，在从端累计号码阵列（投影阵列 (3)）中，指定由新的第二值列表（新的 VL2）指出的位置处的元素（步骤 3801）。此元素在记录号计算过程中被称为“值 D”。在图 40 的示例中，在投影阵列 (3) 中，指定新的 VL2 的值“1”指出的位置处的元素（累计号码）“1”作为值 D。

此后，将“值 C”和“值 D”相加（步骤 3802）。在记录号计算过程中，相加的值被称为“值 E”。在图 40 的示例中，“ $E = 0 + 1 = 1$ ”。

接下来，在从端有序集合阵列（投影阵列 (4)）中，指定由“值 E”指出的位置处的元素（从端记录号），在从端记录号阵列中对应于联接表格中的存储位置处排列此元素（记录号）（步骤 3803）。在图 40 的示例中，提取投影阵列 (4) 的对应于“1”的位置处的值“4”，并且这将成为图 39 所示的从端记录号阵列的头元素来存储（参见步骤 3911）。

对联接表格的每一个记录号连续地执行步骤 3701 到步骤 3803 的过程（参见步骤 3804 和 3805），以便计算对应于联接表格的相应的记录号的主端和从端记录号，并在记录号阵列的对应的位置排列相应的记录号。

图 41 到 44 是分别说明计算从联接表格的记录号阵列中的第二（存储位置号码“1”）到第五（存储位置号码“4”）记录号的主端和从端记录号的过程的视图。

在图 41 中，联接表格的记录号阵列的元素（记录号）为“1”，

而在投影阵列 (1) 中, 指定不大于“1”的最大值“0”以及其存储位置“0” (步骤 3701 和 3702)。在投影阵列 (2) 中, 由存储位置“0”指出的位置处的元素“0”变成主端记录号, 并存储在主端记录号阵列中的对应的位置 (参见步骤 3703 和 3902)。此外, 计算“值  $C = 1 - 0 = 1$ ” (步骤 3704)。

另一方面, 指定主端 PV2 中的存储位置“0”处的元素“0”, 以及由 PV2 的元素“0”指出的位置处的新的 VL2 中的值“1”。此外, 在投影阵列 (3) 中, 指定由新的 VL2 的值“1”指出的元素 3 (累计号码) “1” (步骤 3705、3706 和 3801)。接下来, 计算“值  $E = 1 + 1 = 2$ ” (步骤 3802), 并在投影阵列 (2) 中, 提取存储位置“2”中的元素 (记录号) “10”, 并存储在从端记录号阵列中的对应的位置 (参见步骤 3803 和附图标记 3912)。

在图 42 中, 联接表格的记录号阵列的元素 (记录号) 为“2”, 而在投影阵列 (1) 中, 指定不大于“2”的最大值“2”以及其存储位置“1” (步骤 3701 和 3702)。在投影阵列 (2) 中, 由存储位置“1”指出的位置处的元素“2”变成主端记录号, 并存储在主端记录号阵列中的对应的位置 (参见步骤 3703 和附图标记 3903)。此外, 计算“值  $C = 2 - 2 = 0$ ” (步骤 3724)。

另一方面, 指定主端 PV2 中的存储位置“1”中的元素“1”, PV2 的元素“1”指出的位置上的新的 VL2 中的值“2”。此外, 在投影阵列 (3) 中, 指定由新的 VL2 的值“2”指出的元素 (累计号码) (步骤 3705、3706 和 3801)。接下来, 计算“值  $E = 0 + 3 = 3$ ” (步骤 3802), 并在投影阵列 (3) 中, 提取存储位置“3”中的元素 (记录号) “0”, 并存储在从端记录号阵列中的对应的位置 (参见步骤 3803 和附图标记 3913)。

同样, 在图 43 中, 联接表格的记录号阵列的元素 (记录号) 为“3”, 并在投影阵列 (1) 中, 指定不大于“3”的最大值“3”以及其存储位置“2” (步骤 3701 和 3702)。在投影阵列 (2) 中, 由存储位置“2”指出的位置处的元素“9”变成主端记录号, 并存储在主端记录号阵列中

的对应的位置(参见步骤 3703 和附图标记 3904)。此外,计算“值  $C = 3 - 3 = 0$ ”(步骤 3704)。

另一方面,指定主端 PV2 中的存储位置“2”中的元素“0”,PV2 的元素“0”指出的位置上的新的 VL2 中的值“1”,然后,在投影阵列 (3) 中,指定由新的 VL2 的值“1”指出的元素(累计号码)“1”(步骤 3705、3706 和 3801)。接下来,计算“值  $E = 0 + 1 = 1$ ”(步骤 3802),并在投影阵列 (2) 中,提取存储位置“1”中的元素(记录号)“4”,并存储在从端记录号阵列中的对应的位置(参见步骤 3803 和附图标记 3914)。

此外,在图 44 中,联接表格的记录号阵列的元素(记录号)为“4”,并在投影阵列 (1) 中,指定不大于“4”的最大值“4”以及其存储位置“2”(步骤 3701 和 3702)。在投影阵列 (2) 中,由存储位置“2”指出的位置处的元素“9”变成主端记录号,并存储在主端记录号阵列中的对应的位置(参见步骤 3703 和附图标记 3905)。此外,计算“值  $C = 4 - 3 = 1$ ”(步骤 3704)。

另一方面,指定主端 PV2 中的存储位置“2”处的元素“0”,以及由 PV2 的元素“0”指出的位置处的新的 VL2 中的值“1”,然后,在投影阵列 (3) 中,指定由新的 VL2 的值“1”指出的元素(累计号码)“1”(步骤 3705、3706 和 3801)。接下来,计算“值  $E = 1 + 1 = 2$ ”(步骤 3802),并在投影阵列 (2) 中,提取存储位置“2”中的元素(记录号)“10”,并存储在从端记录号阵列中的对应的位置(参见步骤 3803 和附图标记 3915)。

图 45 是显示主端表格的记录号阵列和作为处理的结果获得的从端表格的记录号阵列的视图,以及联接的视图等等。根据主端表格的记录号阵列 3900(其记录号),从主动原始表格(棒球爱好者俱乐部成员原始表 4501)提取记录,以便获得其中保留了原始表的顺序的表格(棒球爱好者俱乐部成员表格 4521)。另一方面,根据从端表格的记录号阵列 3910(其记录号),从从端原始表(比赛时间表原始表 4511)提取记录,以便可以获得表格(比赛时间表 4531)(使该表格

对应于主端记录)。

如上所述,根据此实施例,也是在联接的过程中,可以使每一个投影阵列与第二值列表相同。此外,共享值列表(共享新的值列表)至多只限于通过将主端第二值列表和从端第二值列表相加所获得的大小。相应地,当指定投影阵列中的元素或指定了第二值列表的值时,可以缩短处理时间。

本发明不仅限于上述实施例,在权利要求书中列举的本发明的范围内可以作出各种修改,毫无疑问,它们也包括在本发明的范围内。

例如,在实施例中,作为聚合过程,虽然描述了创建存在号码阵列,以及提取存在号码阵列中的元素的情况,但是,本发明不仅限于此。一般而言,聚合基本上被分成“0 或大于 0 的多个维”和“0 或 1 的度量”。例如,当考虑“性别 x 每一个重量的重量的平均值”,维度是“二维”(“性别”和“重量”),度量是“重量”。在度量为“0”的情况下,元素的数量变成通过聚合获得的值(聚合值)。在此情况下,存在号码阵列中的元素变成聚合值。此外,还可以将总和、最大值、最小值、平均值等等作为度量。

作为维,可以有各种类别。例如,在年龄中,可以使 0 到 9 岁小于 10 岁,并可以使其成为一个类别。这里,准备与值列表具有相同元素数量的分类号阵列,在分类号阵列中,将“0”作为对应于值列表的值(年龄)“0”到“9”的位置处的元素排列,“1”作为另一个元素排列。当然,基本上,可以将一个类别指定到可以被区别的每一个值。即,在年龄中,在分类号阵列中,例如,值“0”、“1”、“2”...可以按顺序排列。在聚合中,获取具有与某一分类号相关的记录号的集合,对于从记录号指定的值,执行预先确定的操作。由于此,可以获取诸如总和、最大值和最小值之类的各种聚合值。也是在此实施例中,提供与第二值列表具有同样大小的分类号阵列,当指定了对应于类别的值时,可以获取各种聚合值。此外,在实施例中,创建了第二指针阵列和第二值列表,以便压缩用于检索、聚合、分类、联接等等的阵列的量,由于此,所需的存储区缩小,并且处理时间缩短。然而,对此没有限制,



可以使用在创建了第二指针阵列并且第二值列表的中间阶段创建的项目值号码阵列。在此情况下，代替第二指针阵列 (PV2)，可以使用项目值号码阵列，代替第二值列表 (VL2)，可以使用原始值列表 (VL)。即，在上述实施例中，例如，为了指定值列表中的项目值，按照有序集合阵列中的元素、第二指针阵列中的元素、第二值列表中的元素，以及值列表中的元素（项目值）的顺序指定元素。另一方面，在修改的示例中，可以按照有序集合阵列中的元素、项目值号码阵列中的元素，以及值列表中的元素（项目值）的顺序来指定元素。

此外，也是在检索、聚合、分类、联接等等的过程中，作为其中可以对值进行参考的阵列，代替第二指针阵列，可以使用项目值号码阵列，代替第二值列表，可以使用原始值列表。

此外，在上述实施例中，将预先确定的程序读取到通用计算机系统 10 中，并执行程序，以便实现多个表格格式数据的联接和涉及联接的表格格式数据的过程。然而，本发明不仅限于此，毫无疑问，可以采用一种结构，以便专用于数据库处理的单板计算机连接到诸如个人计算机之类的通用计算机系统，并且单板计算机可以执行该过程。相应地，在本说明书中，装置不一定表示物理装置，而包括通过软件和/或硬件实现相应的装置的功能的情况。此外，一个装置的功能可以通过两个或更多物理装置来实现，或者两个或更多装置的功能可以通过一个物理装置来实现。

根据本发明，可以提供这样的方法，其中，使阵列与子集的大小或根据此受限的值列表的子集的大小一致，实现对应于子集的大小的处理时间的缩短。

#### 工业实用性

本发明可以用于管理大量的数据的系统，例如，数据库或数据仓库。具体来说，它可以用于大型的科学计算，诸如接受和排序管理之类的基本业务管理或股票交易，以及办公室管理。

图1

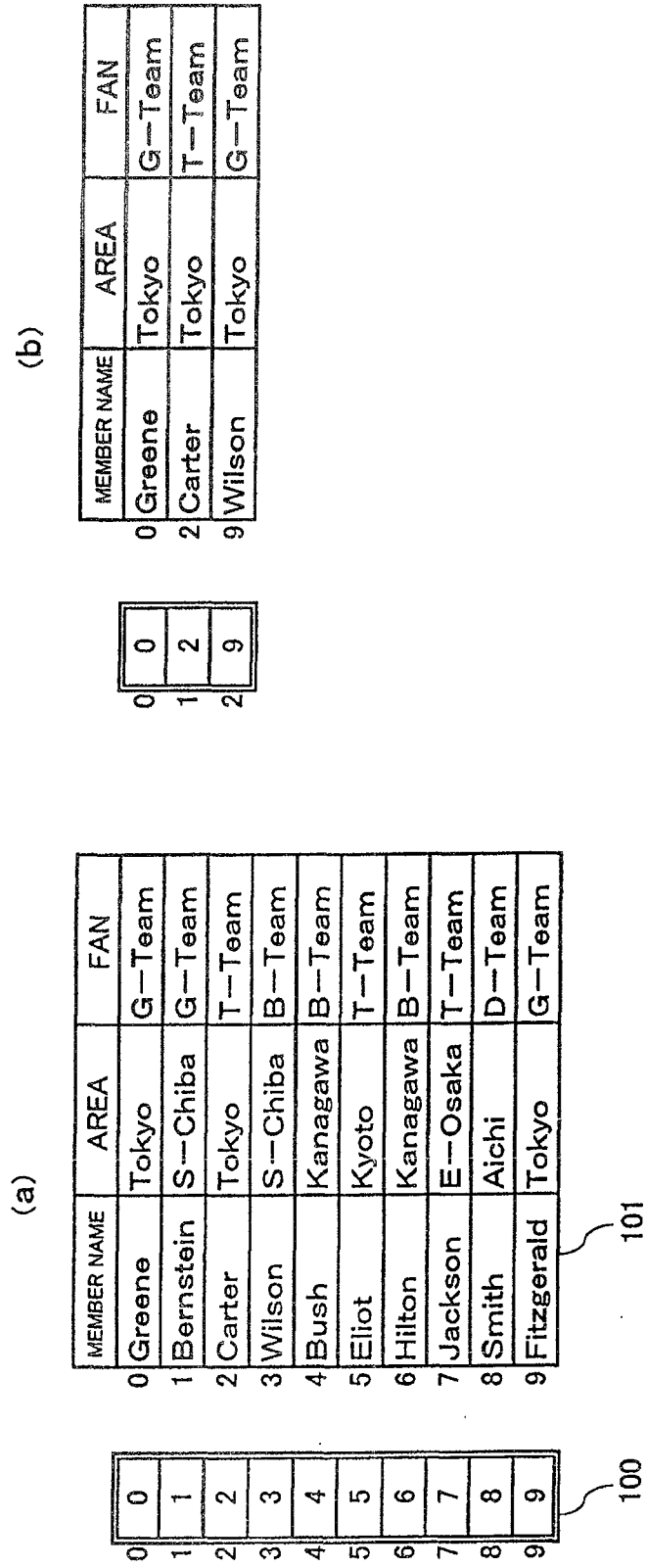
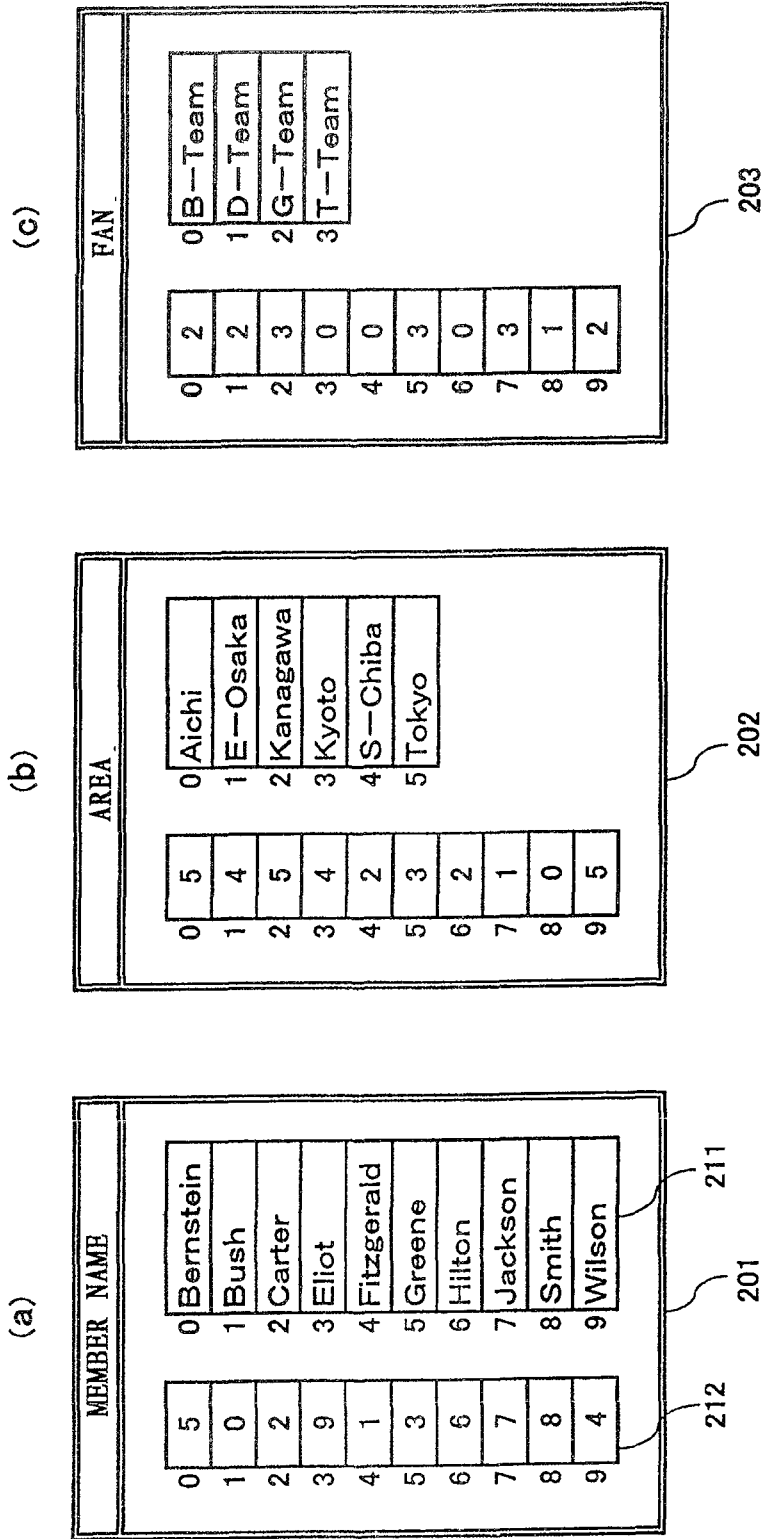


图 2



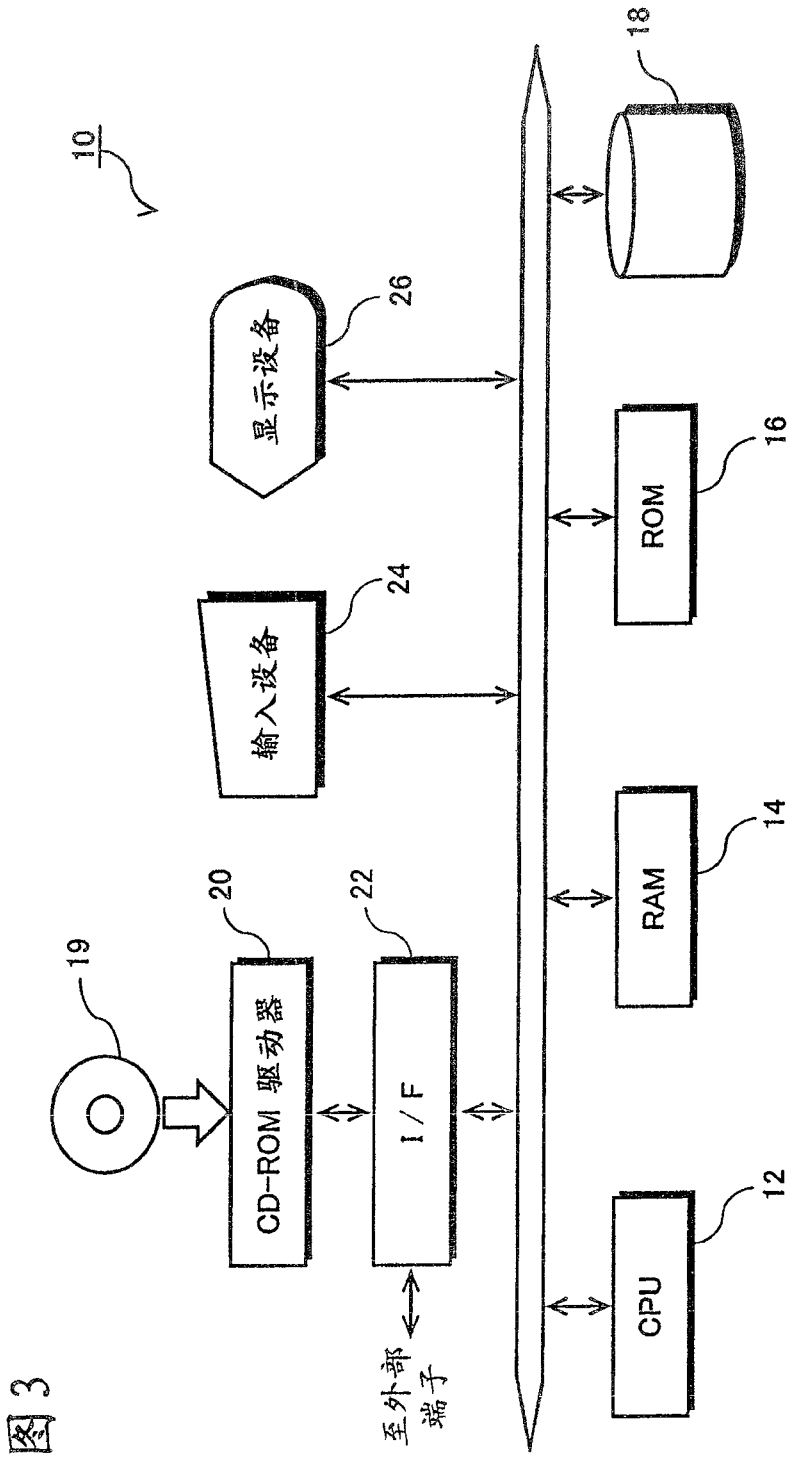


图3

图4

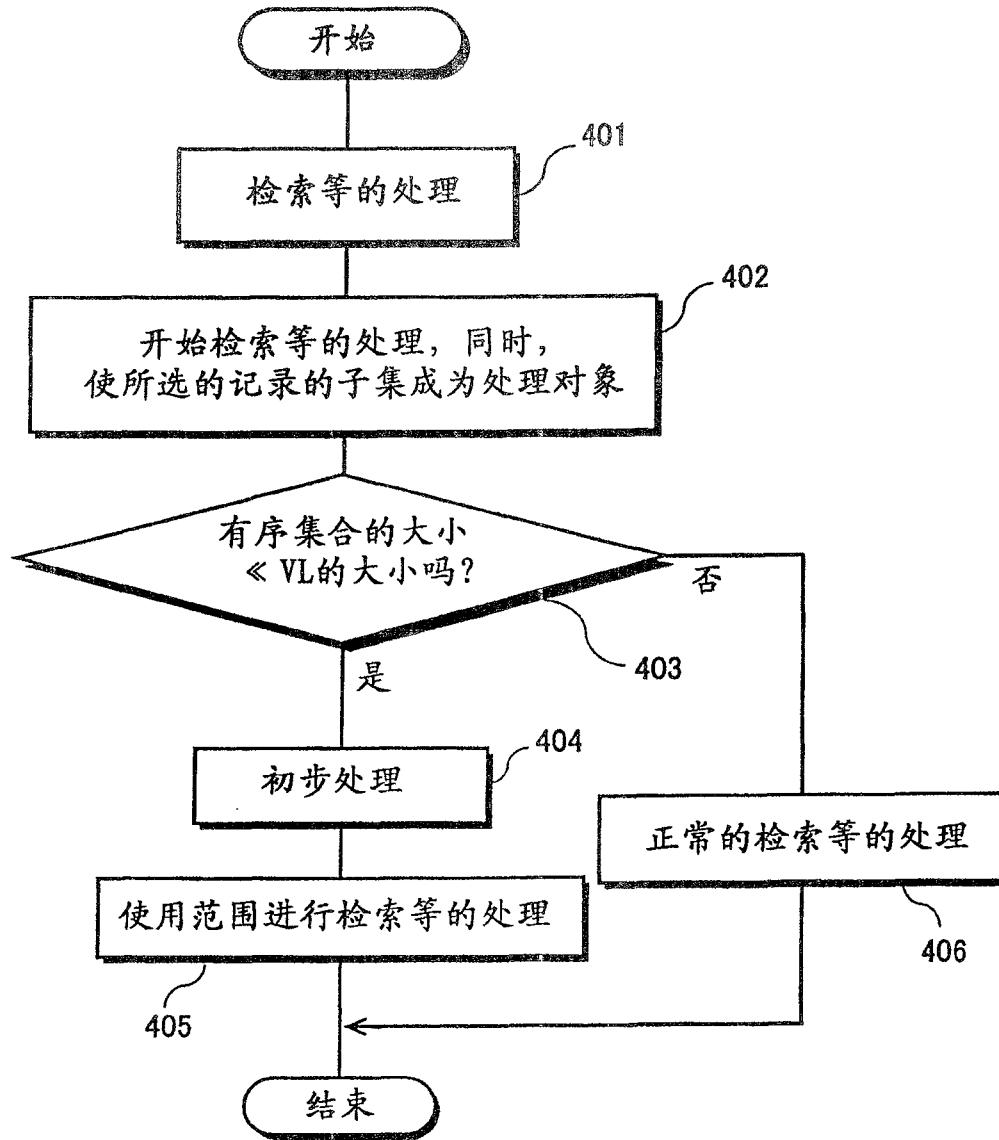


图 5

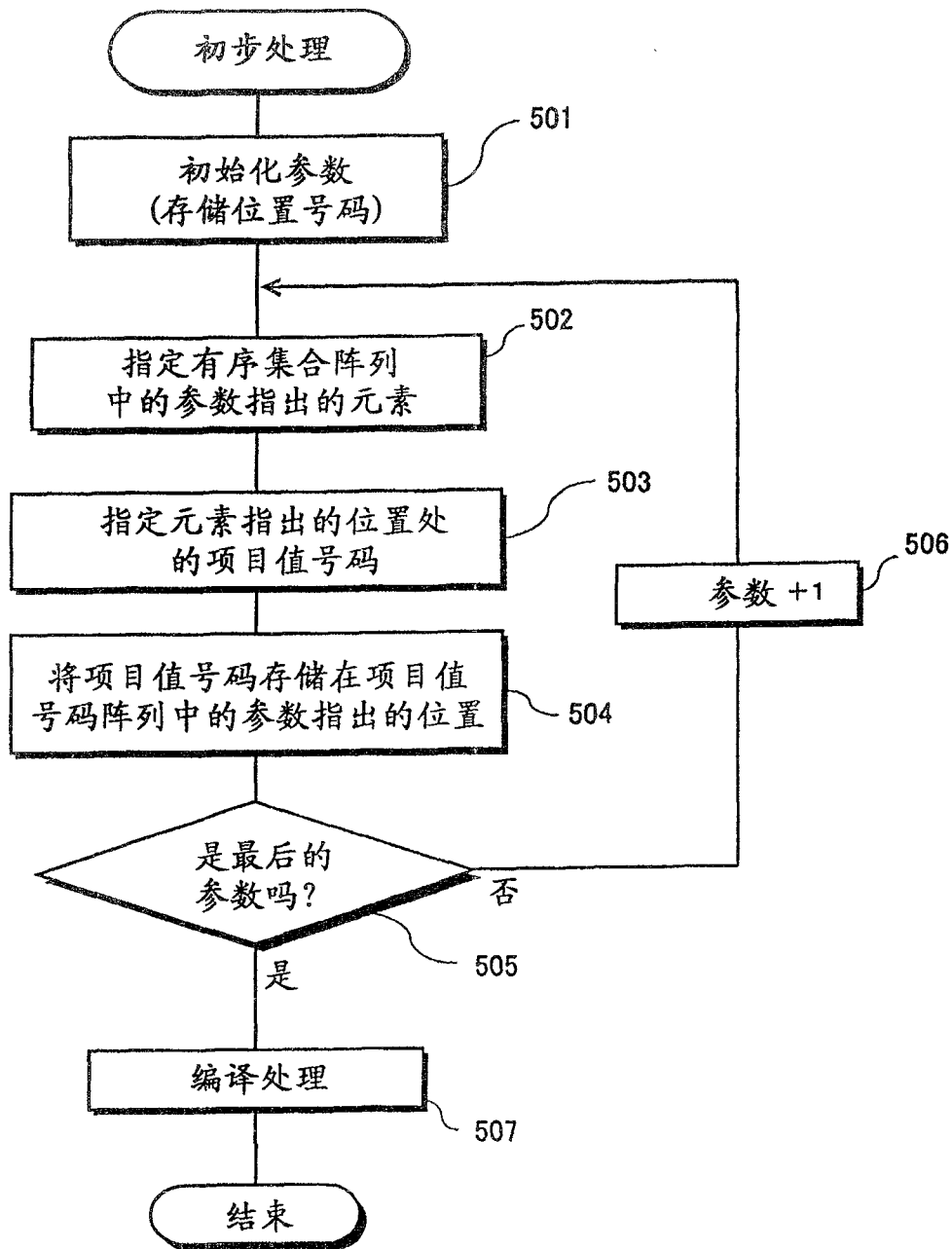


图6

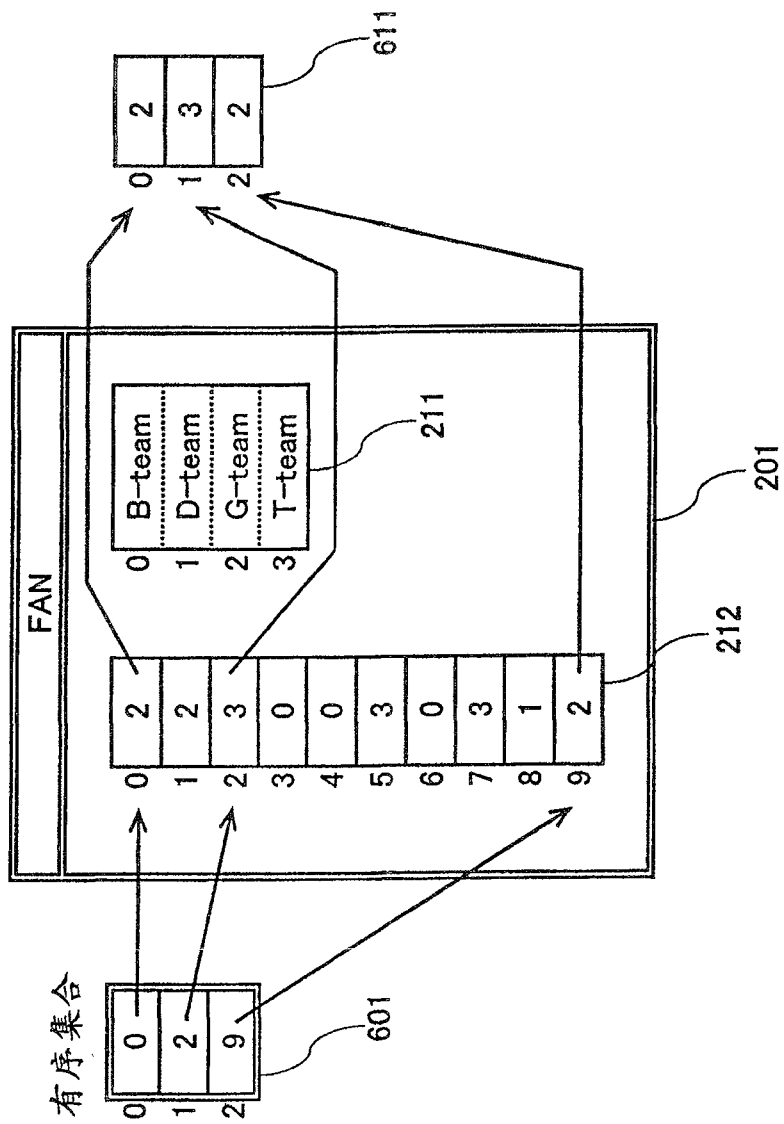


图7

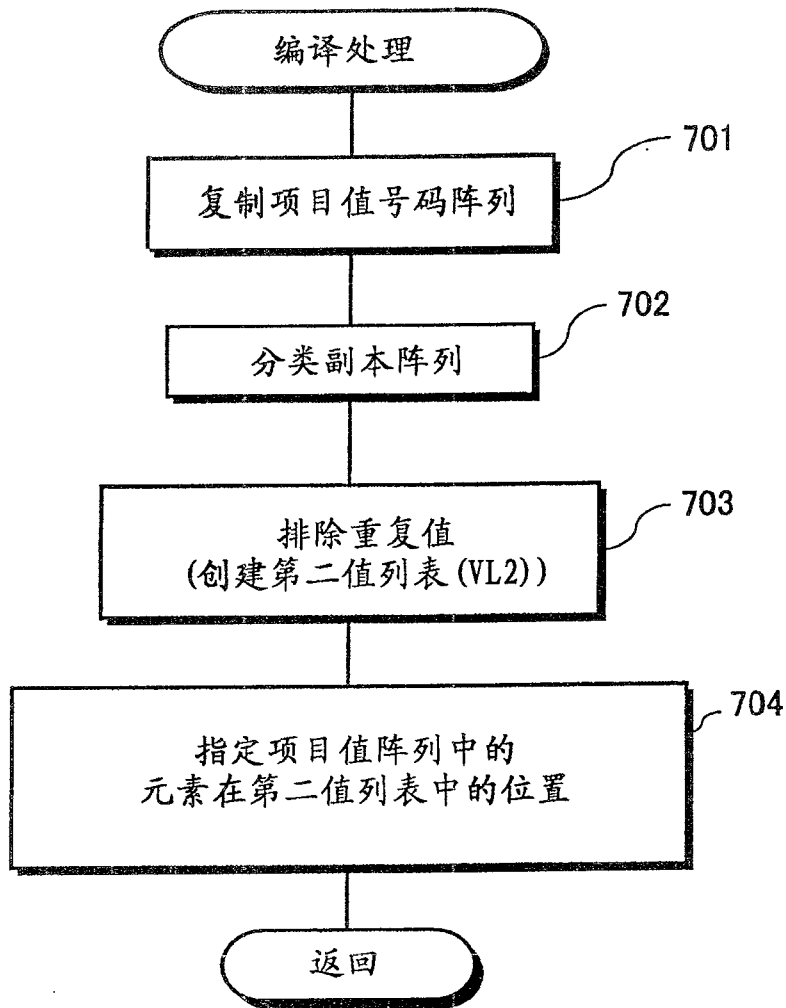




图 8

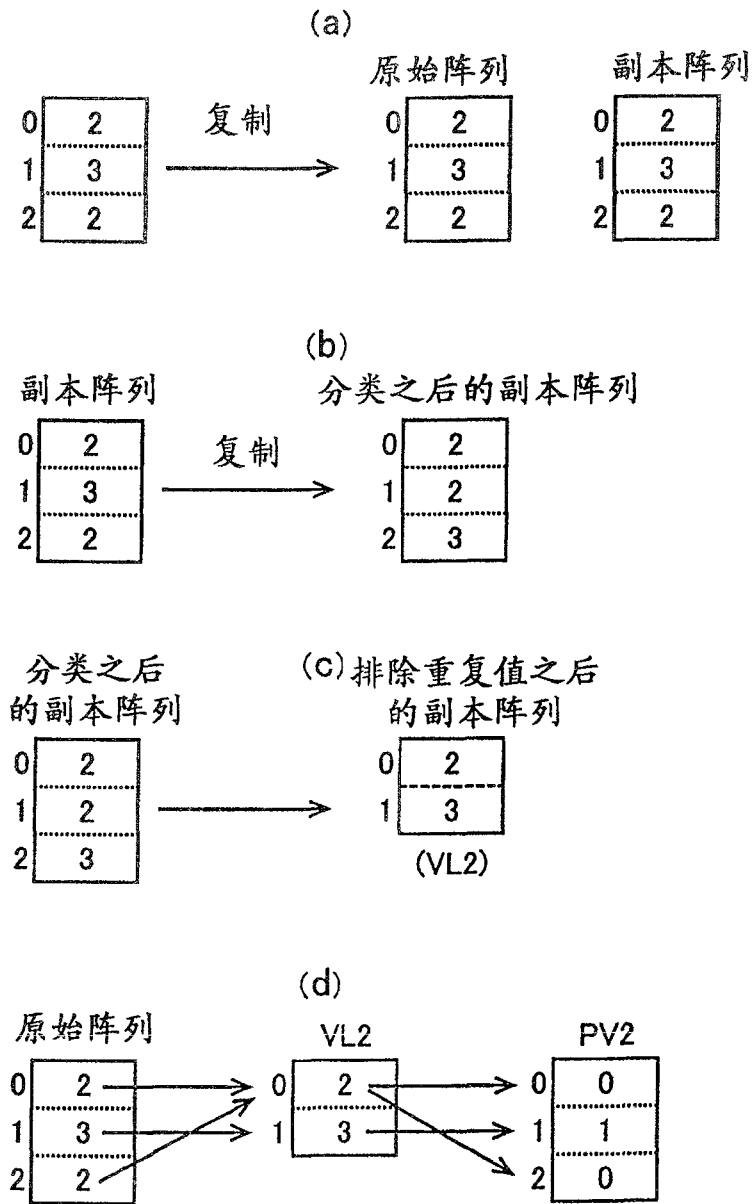


图9

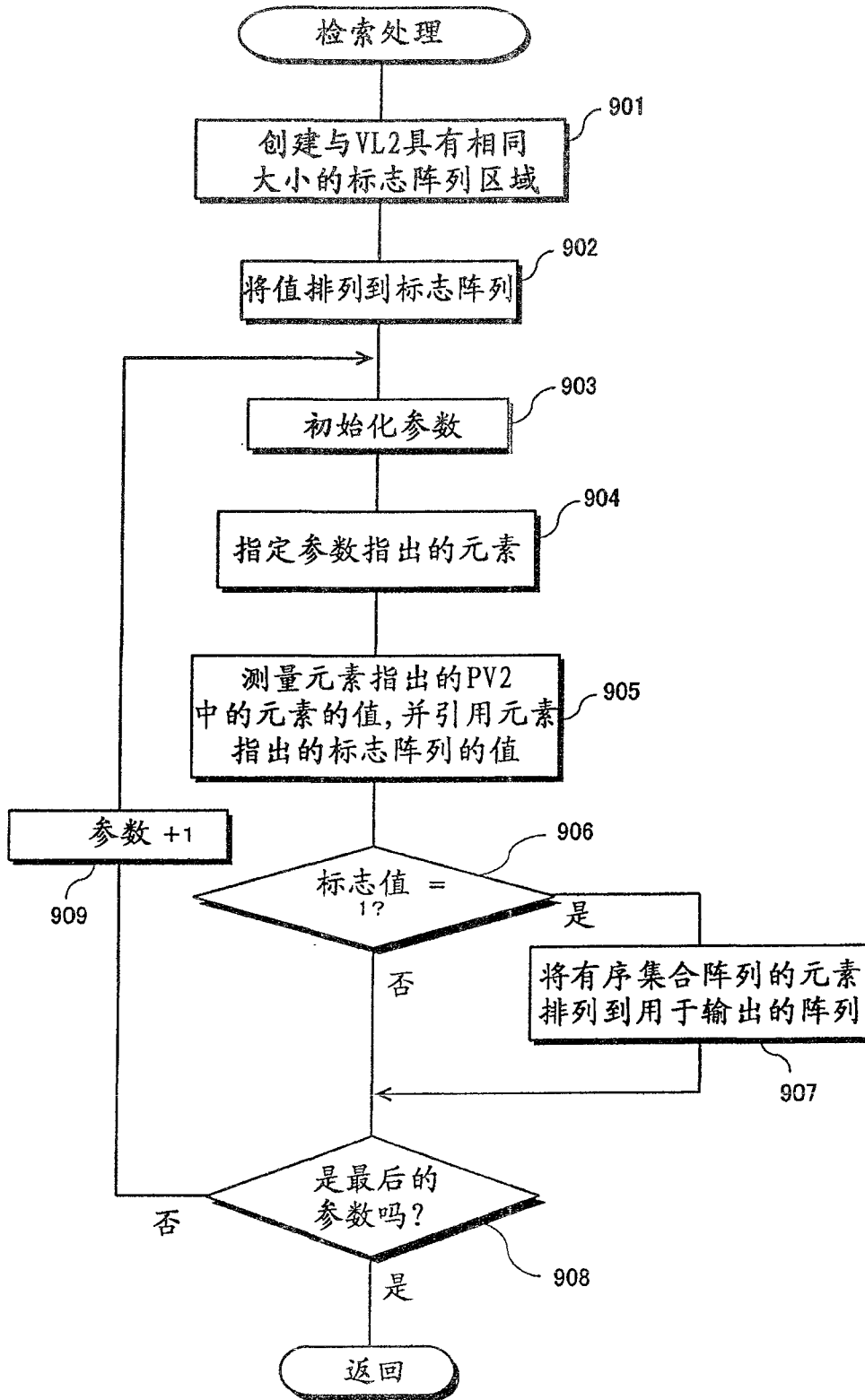


图10

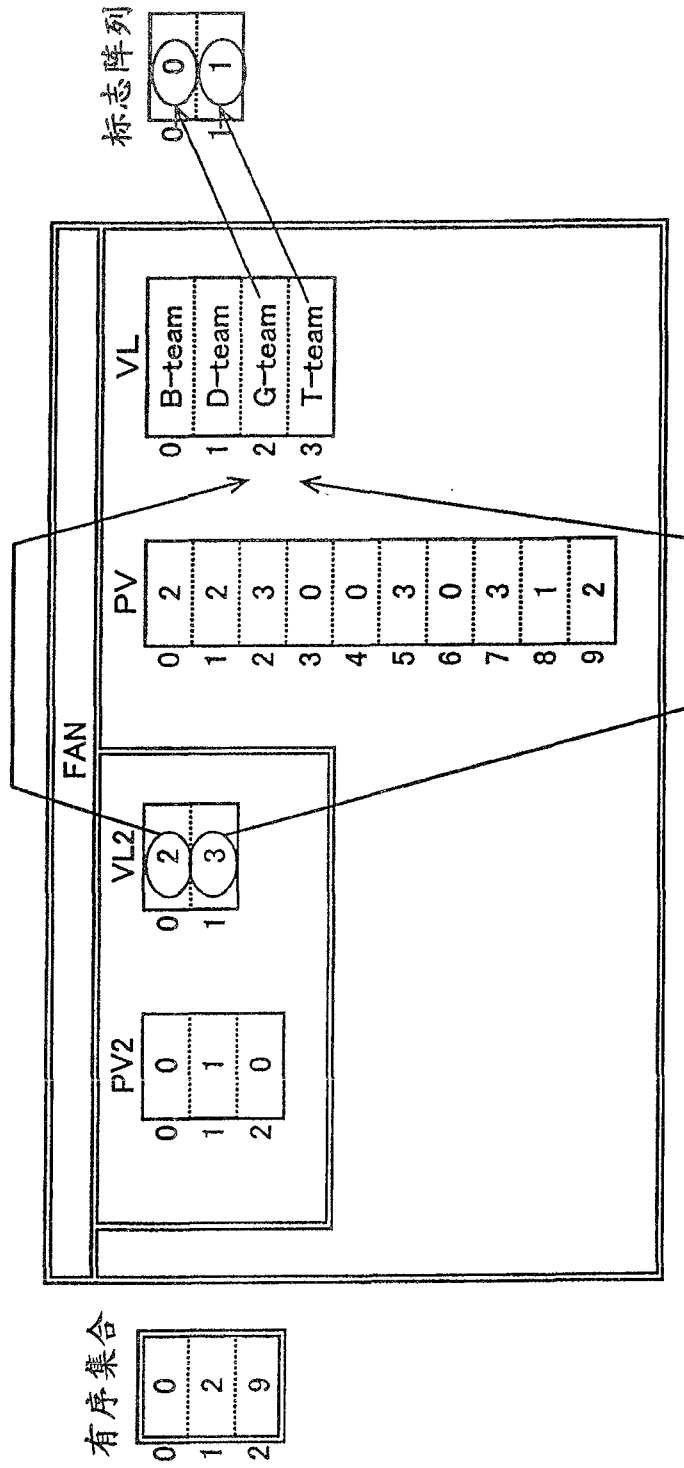


图11

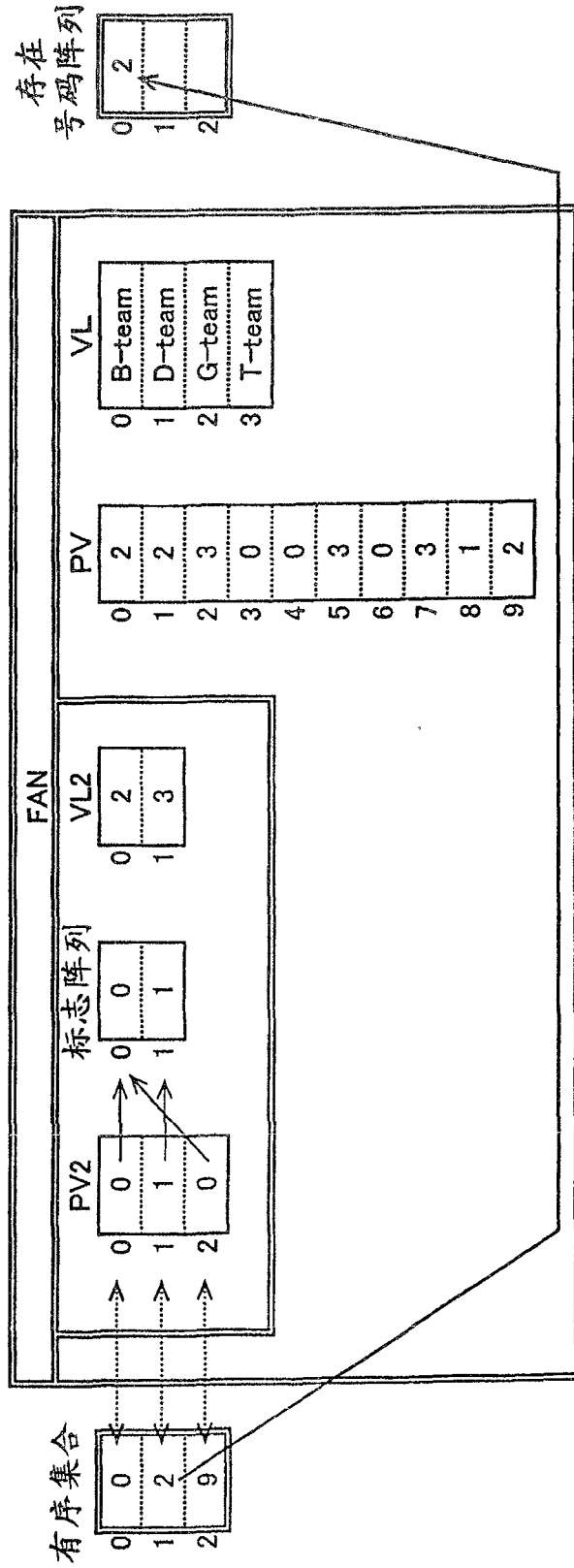


图12

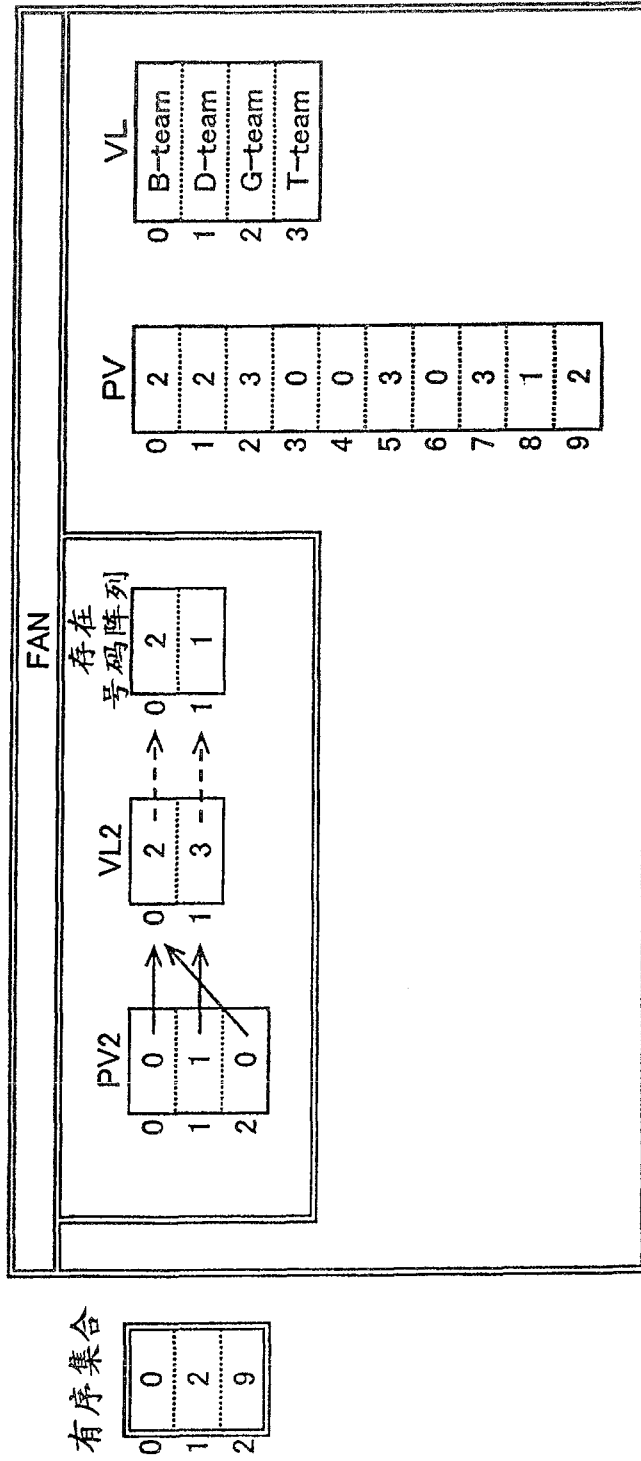


图 13

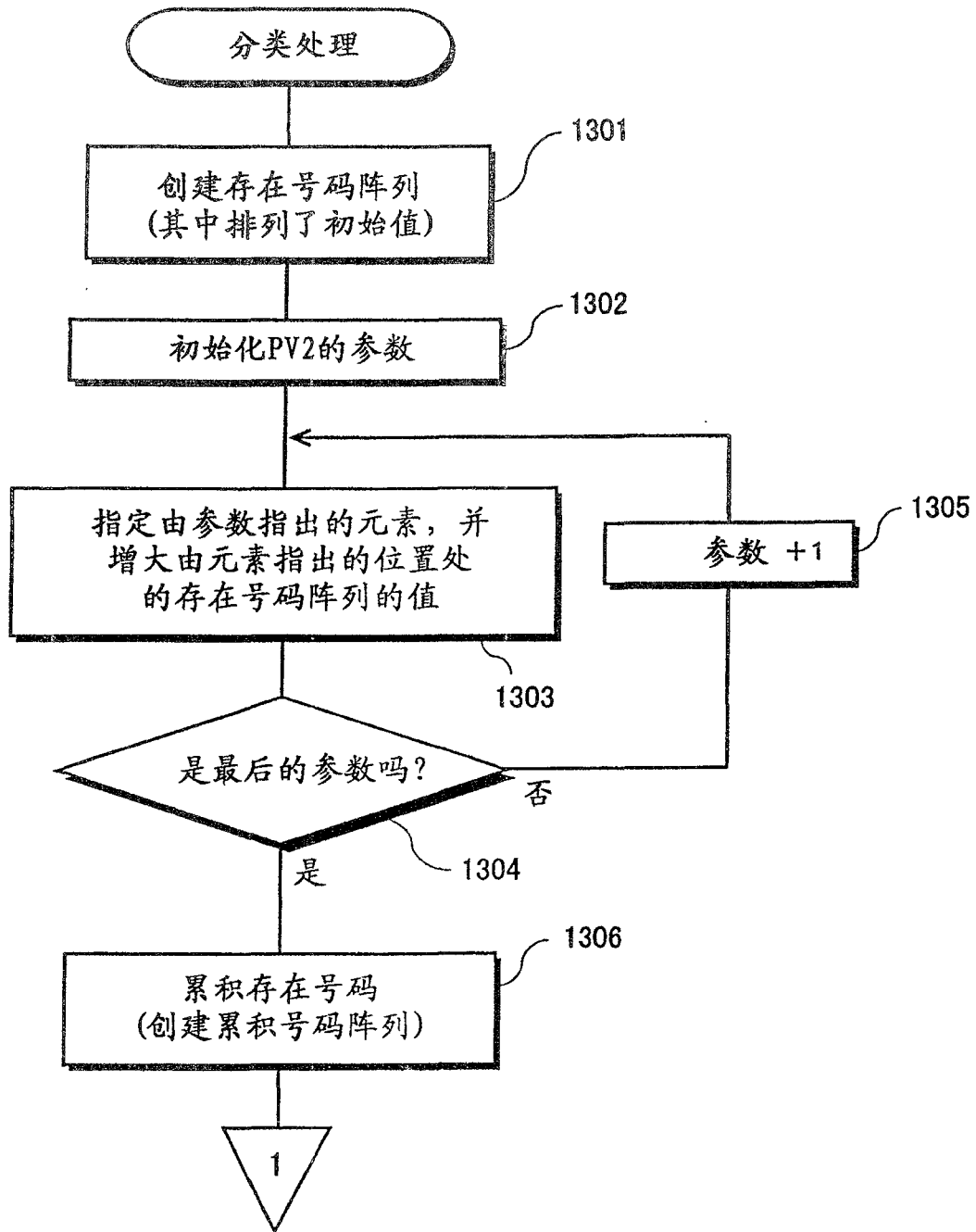


图 14

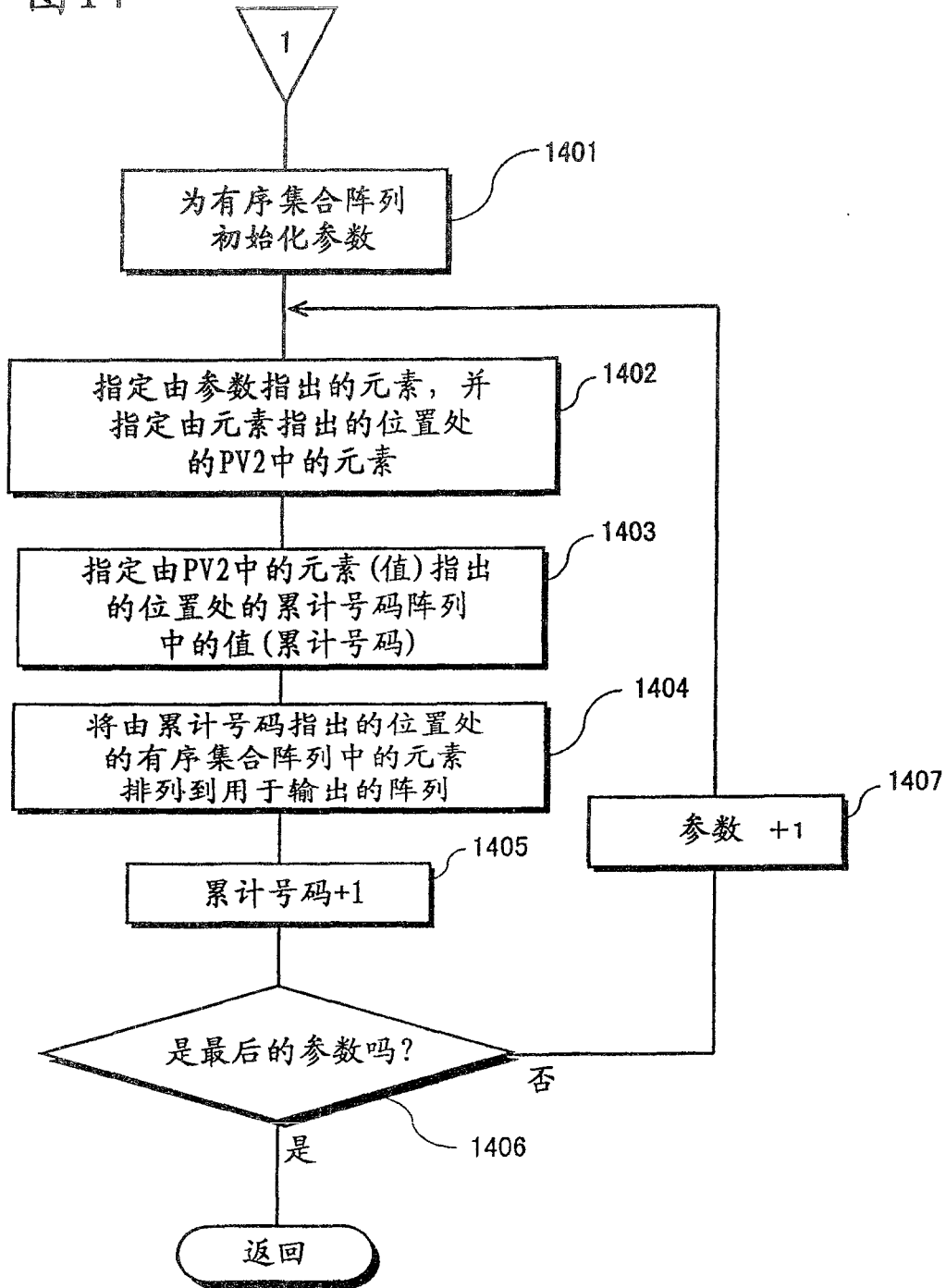


图15

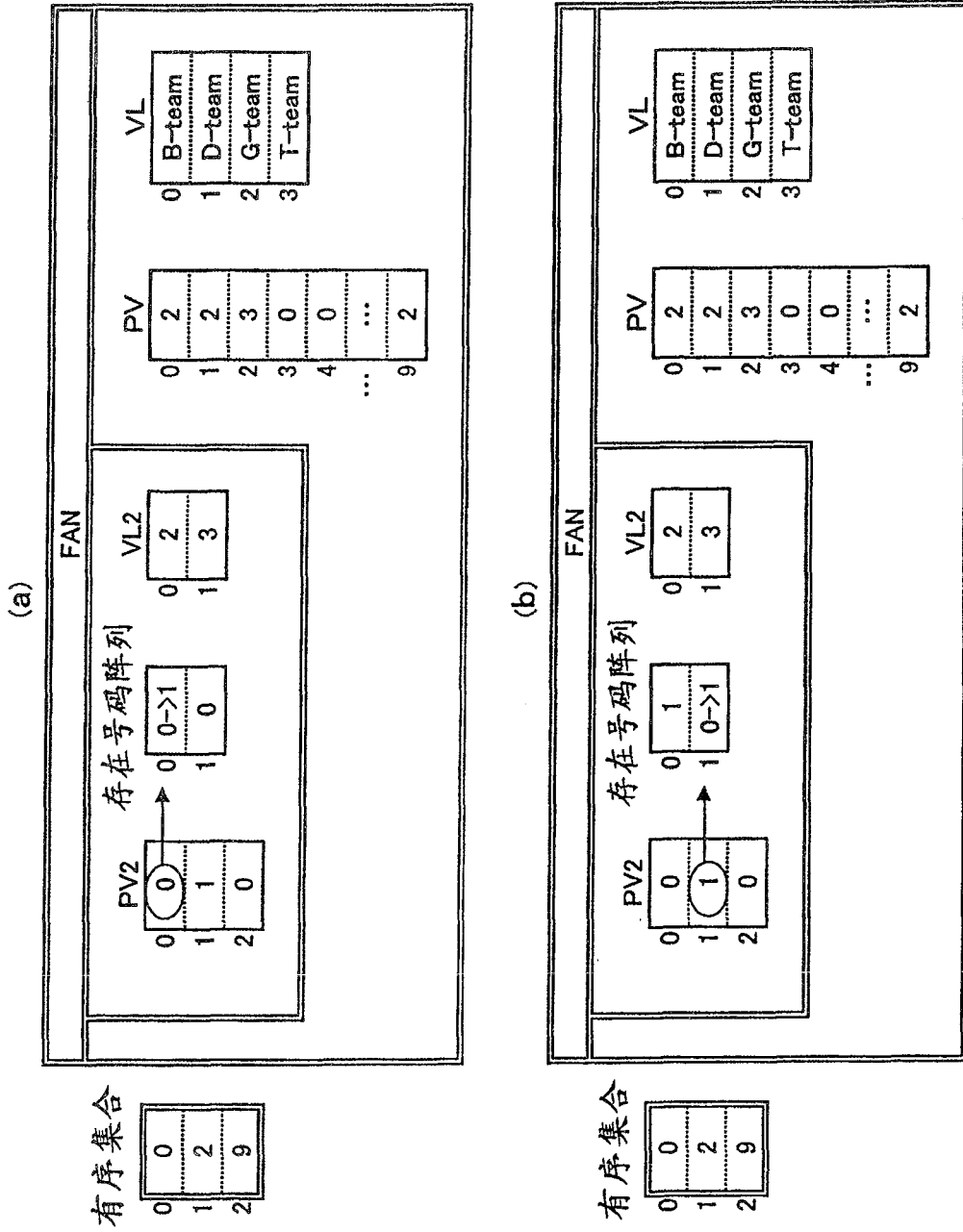




图16

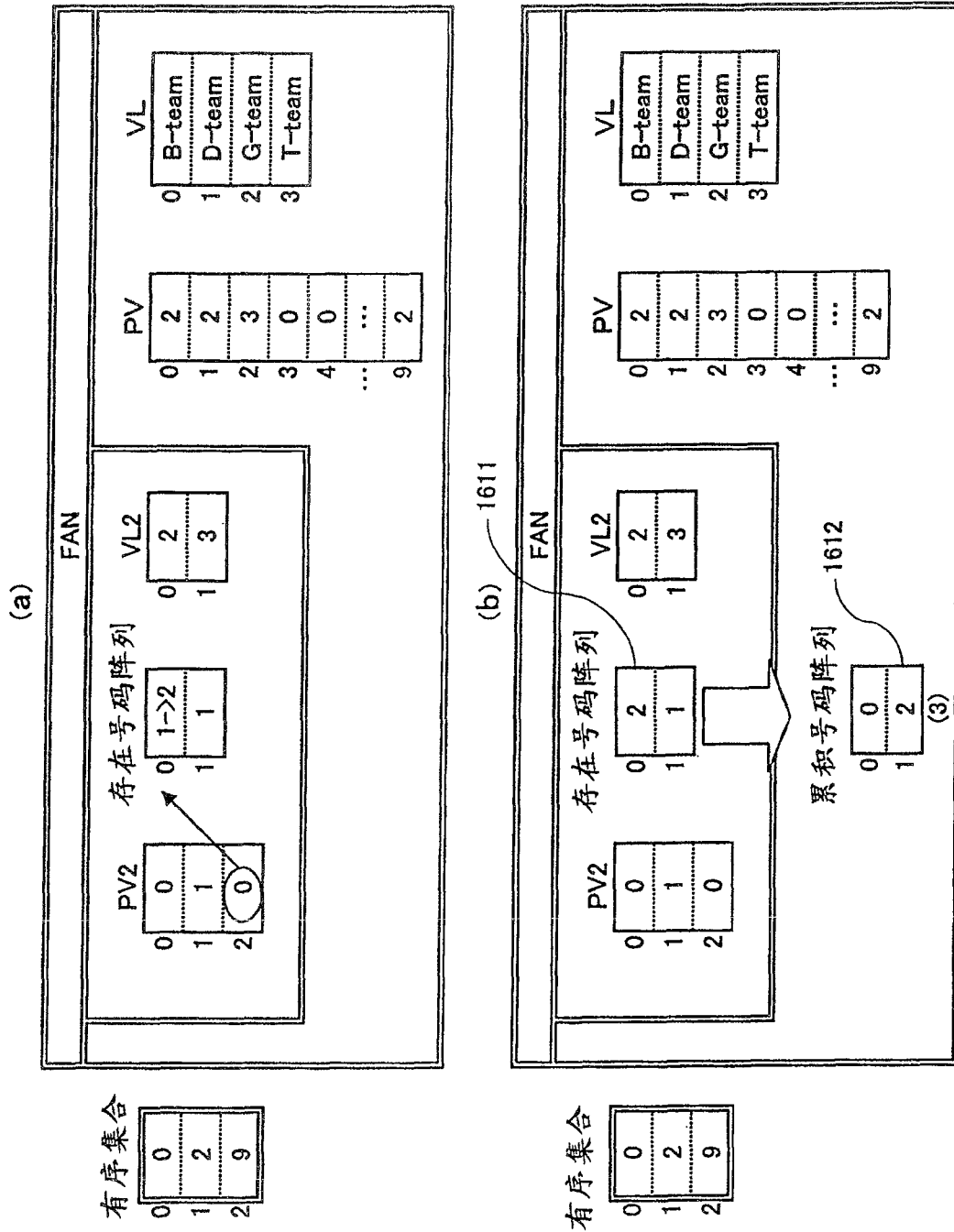
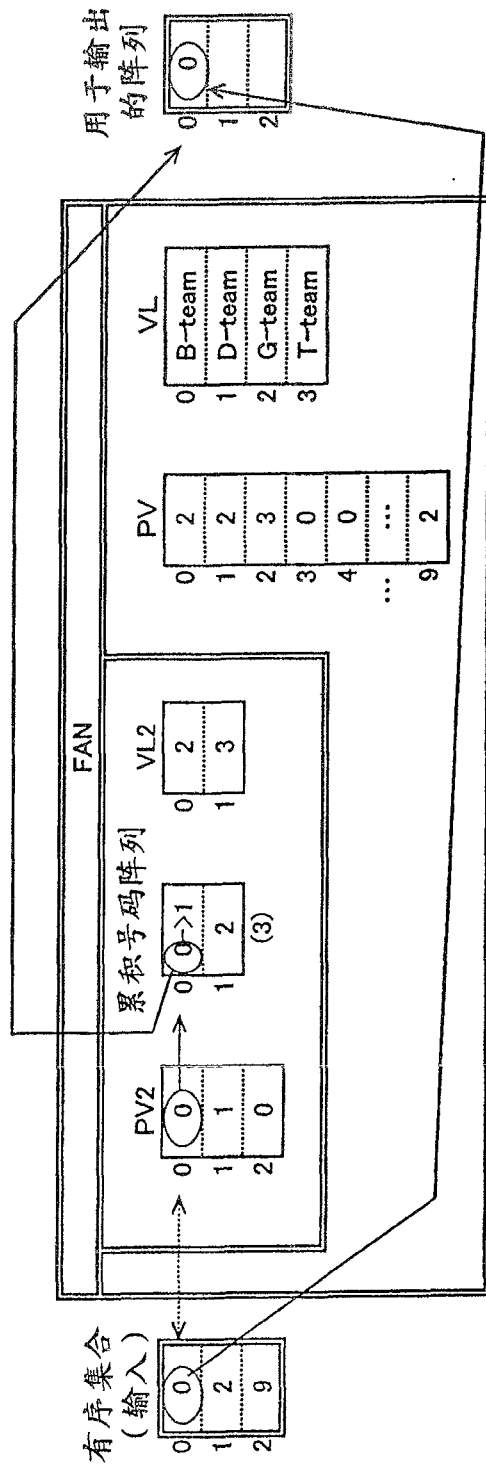


图17



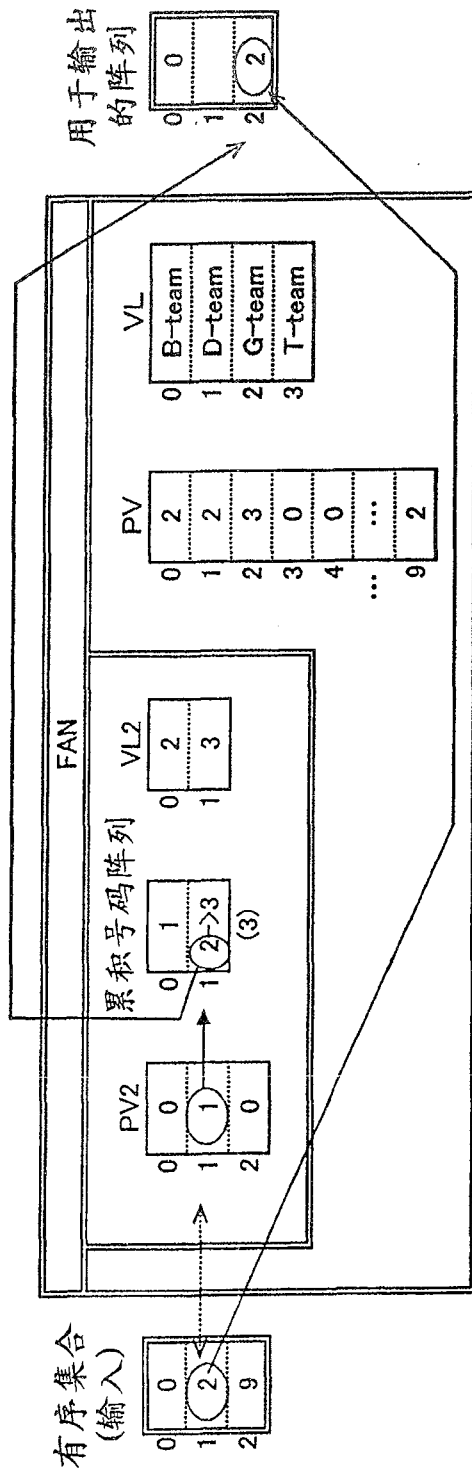


图18

图19

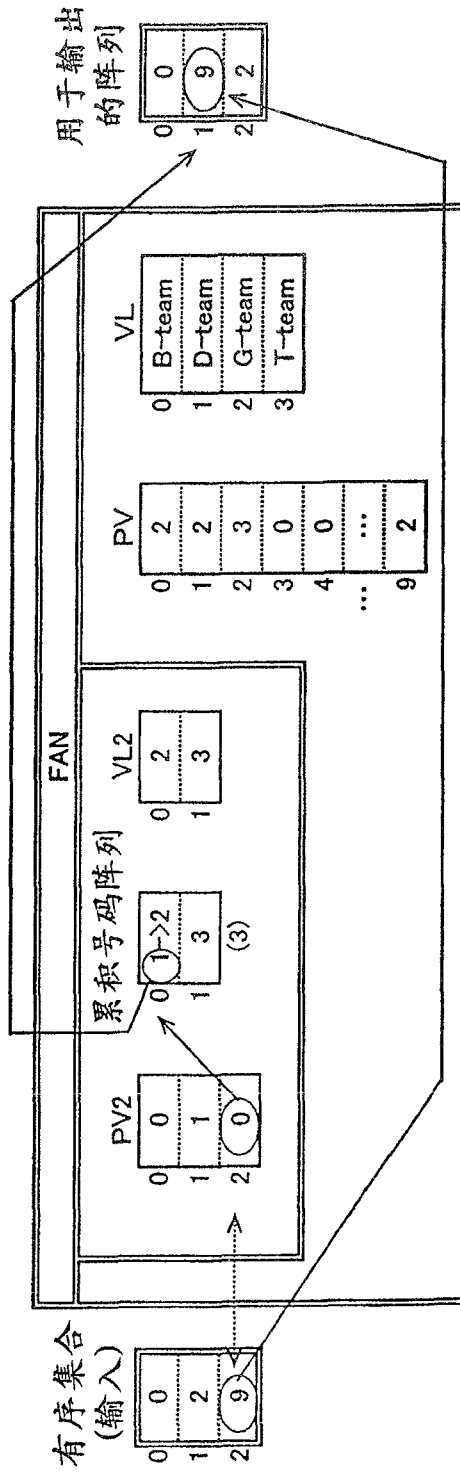


图20

棒球爱好者俱乐部成员原始表 (a)

	MEMBER NAME	AREA	FAN
0	Greene	Tokyo	G-team
1	Bernstein	S-Chiba	G-team
2	Carter	Tokyo	T-team
3	Wilson	S-Chiba	B-team
4	Bush	Kanagawa	B-team
5	Eliot	Kyoto	T-team
6	Hilton	Kanagawa	B-team
7	Jackson	E-Osaka	T-team
8	Smith	Aichi	D-team
9	Fitzgerald	Tokyo	G-team

2001

棒球爱好者俱乐部成员表

	MEMBER NAME	AREA	FAN
0	Greene	Tokyo	G-team
2	Carter	Tokyo	T-team
9	Fitzgerald	Tokyo	G-team

2002

比赛时间表原始表 (b)

	BASEBALL TEAM	GAME DAY	BASEBALL FIELD
0	T-team	MAY 10	Tokyo
1	S-team	MAY 10	U-Nagoya
2	C-team	MAY 10	V-Hiroshima
3	D-team	MAY 10	U-Nagoya
4	G-team	MAY 10	Tokyo
5	B-team	MAY 10	V-Hiroshima
6	T-team	MAY 11	Yokohama
7	S-team	MAY 11	V-Hiroshima
8	C-team	MAY 11	V-Hiroshima
9	D-team	MAY 11	Tokyo
10	G-team	MAY 11	Tokyo
11	B-team	MAY 11	Yokohama

2011

比赛时间表

	BASEBALL TEAM	GAME DAY	BASEBALL FIELD
0	T-team	MAY 10	Tokyo
4	G-team	MAY 10	Tokyo
9	D-team	MAY 11	Tokyo
10	G-team	MAY 11	Tokyo

2012

图 21

主端			从端		
MEMBER NAME	AREA	FAN	BASEBALL TEAM	GAME DAY	BASEBALL FIELD
Greene	Tokyo	G-team	G-team	MAY 10	Tokyo
Greene	Tokyo	G-team	G-team	MAY 11	Tokyo
Carter	Tokyo	T-team	T-team	MAY 10	Tokyo
Fitzgerald	Tokyo	G-team	G-team	MAY 10	Tokyo
Fitzgerald	Tokyo	G-team	G-team	MAY 11	Tokyo

2101

2111

图 22

BASEBALL TEAM		GAME DAY		BASEBALL FIELD	
PV	VL	PV	VL	PV	VL
0 5	0 B-team	0 0	0 MAY 10	0 0	0 Tokyo
1 4	1 C-team	1 0	1 MAY 11	1 1	1 U-Nagoya
2 1	2 D-team	2 0		2 2	2 V-Hiroshima
3 2	3 G-team	3 0		3 1	3 Yokohama
4 3	4 S-team	4 0		4 0	
5 0	5 T-team	5 0		5 2	
6 5		6 1		6 3	
7 4		7 1		7 2	
8 1		8 1		8 2	
9 2		9 1		9 0	
10 3		10 1		10 0	
11 0		11 1		11 3	

图23

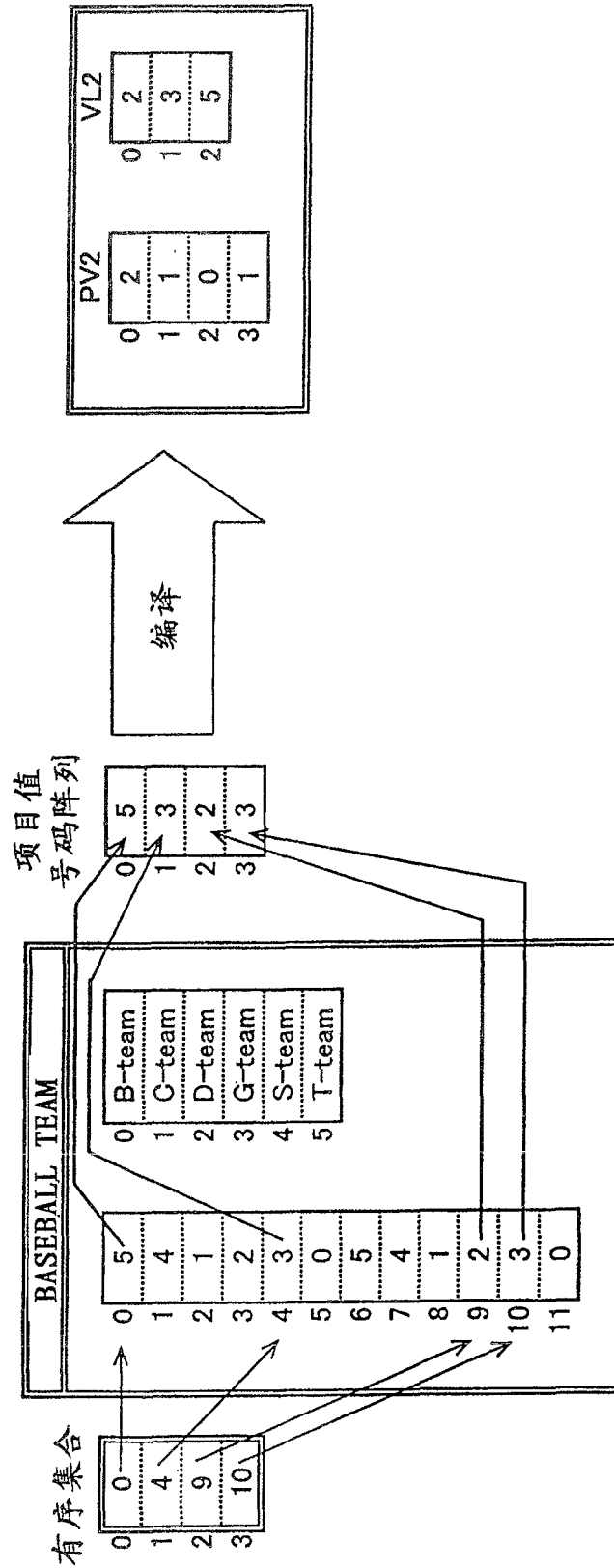




图 24

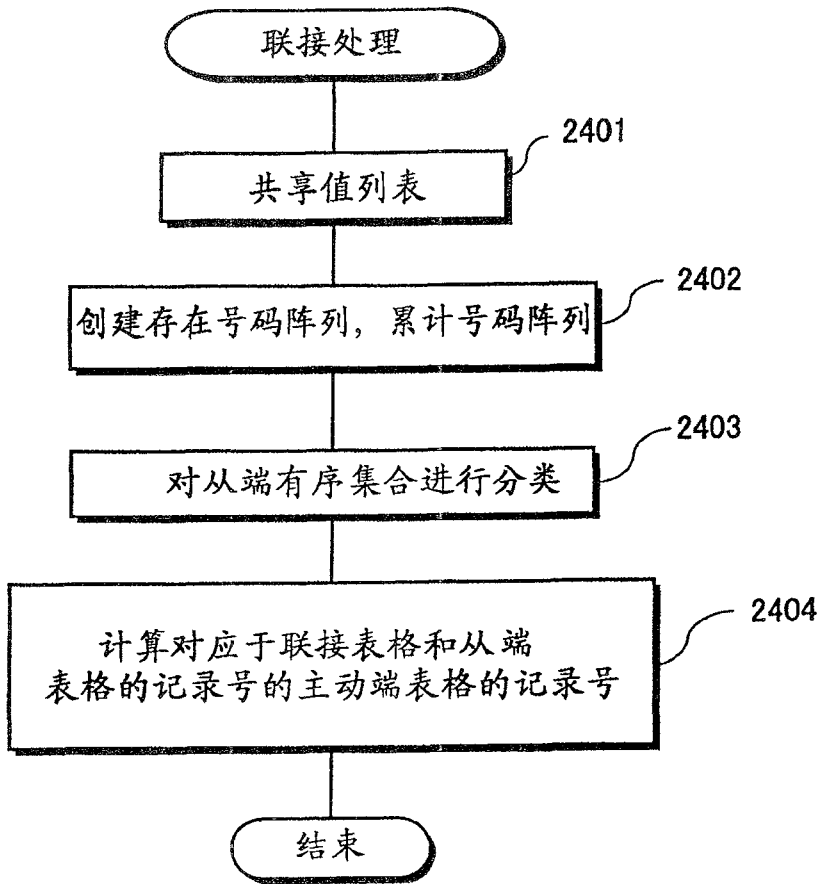


图 25

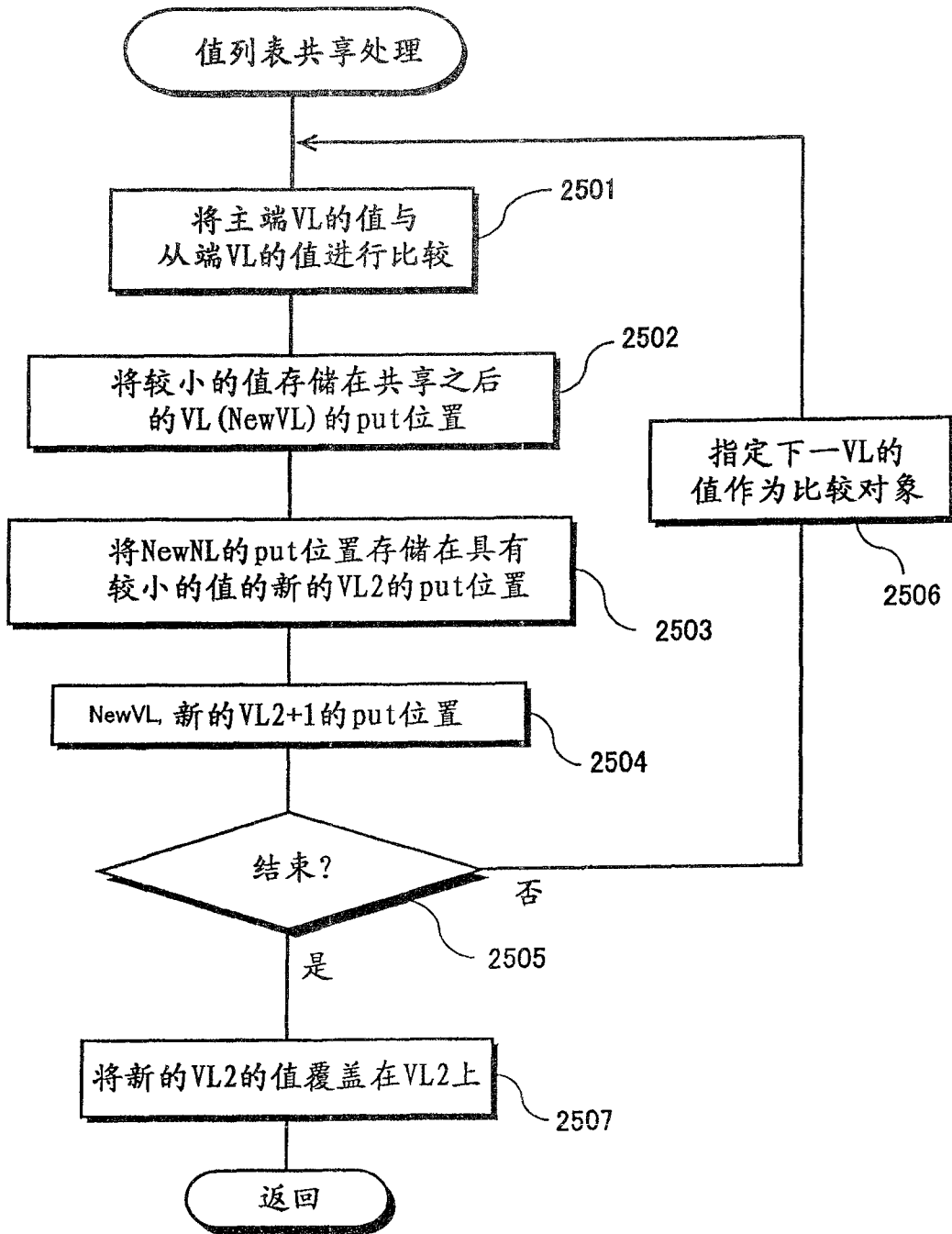


图 26

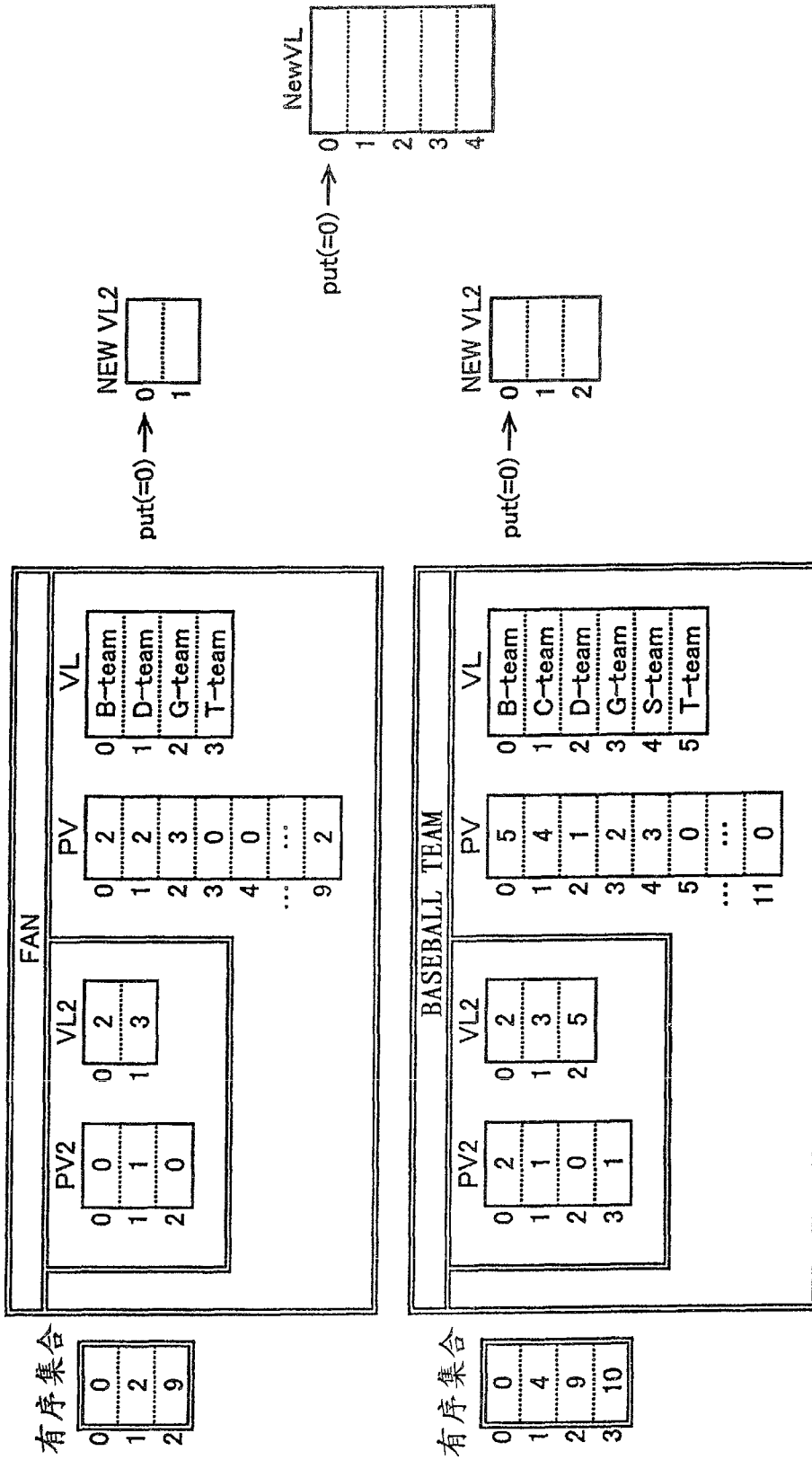


图 27

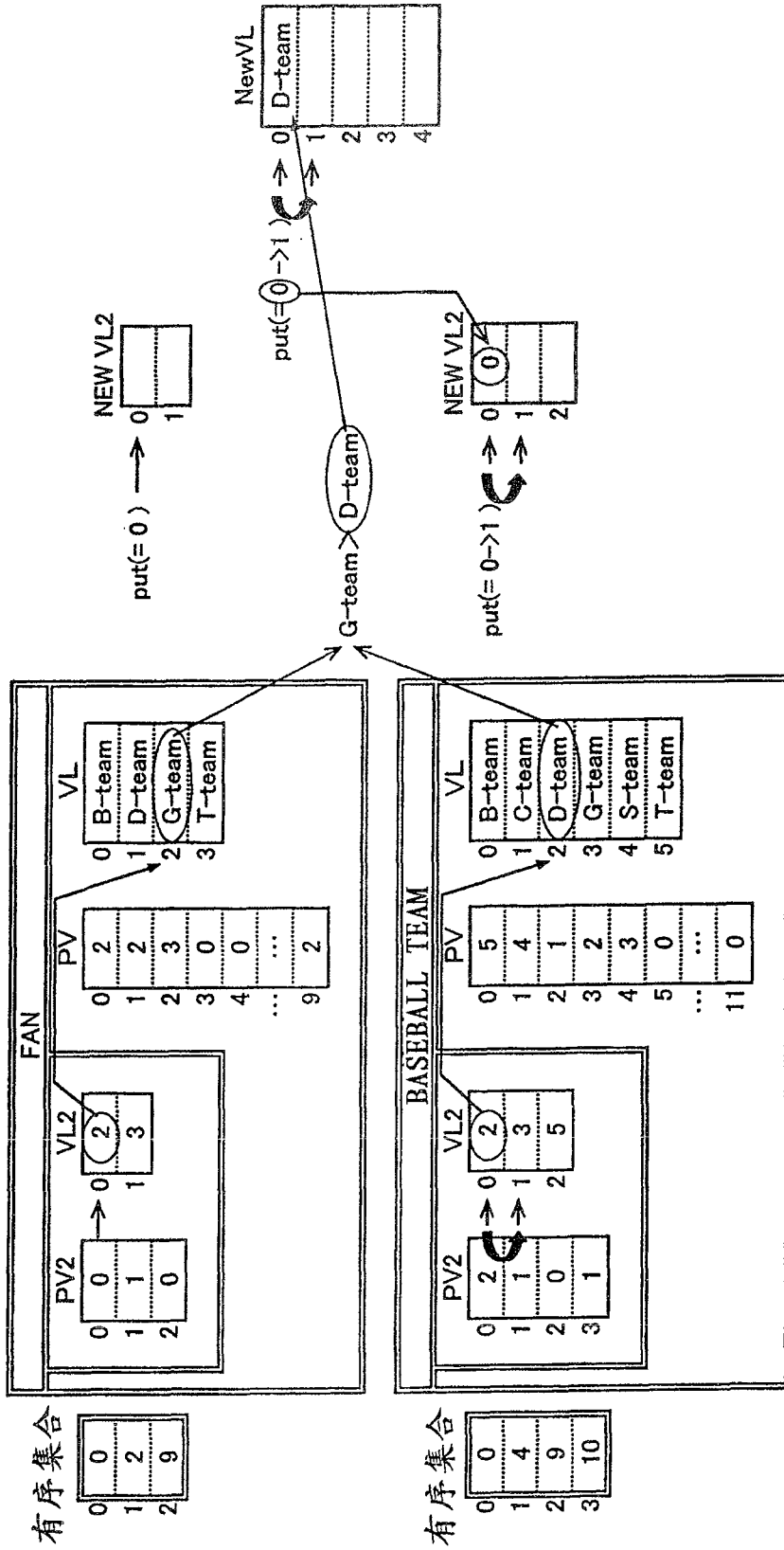


图 28

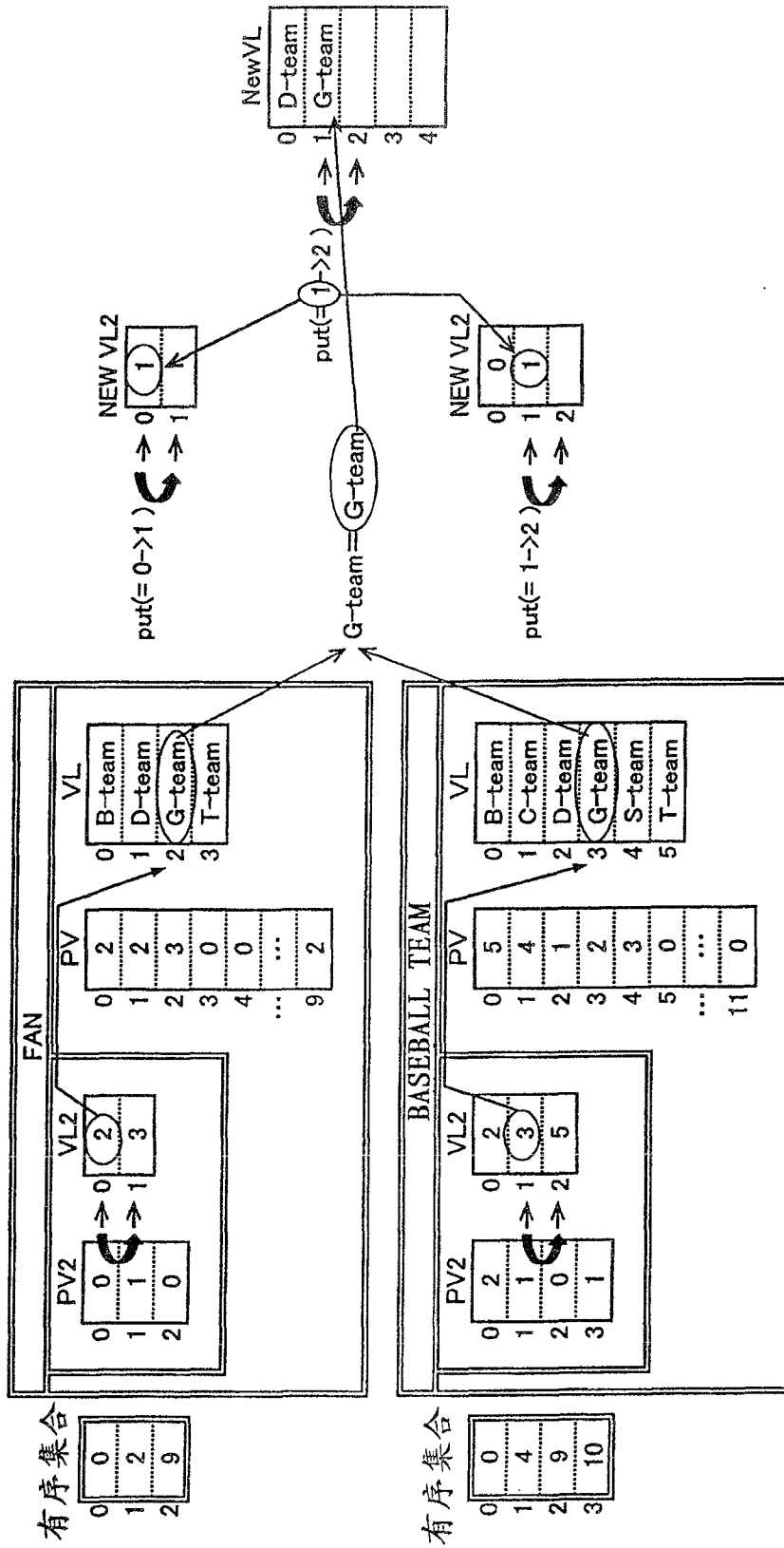


图 29

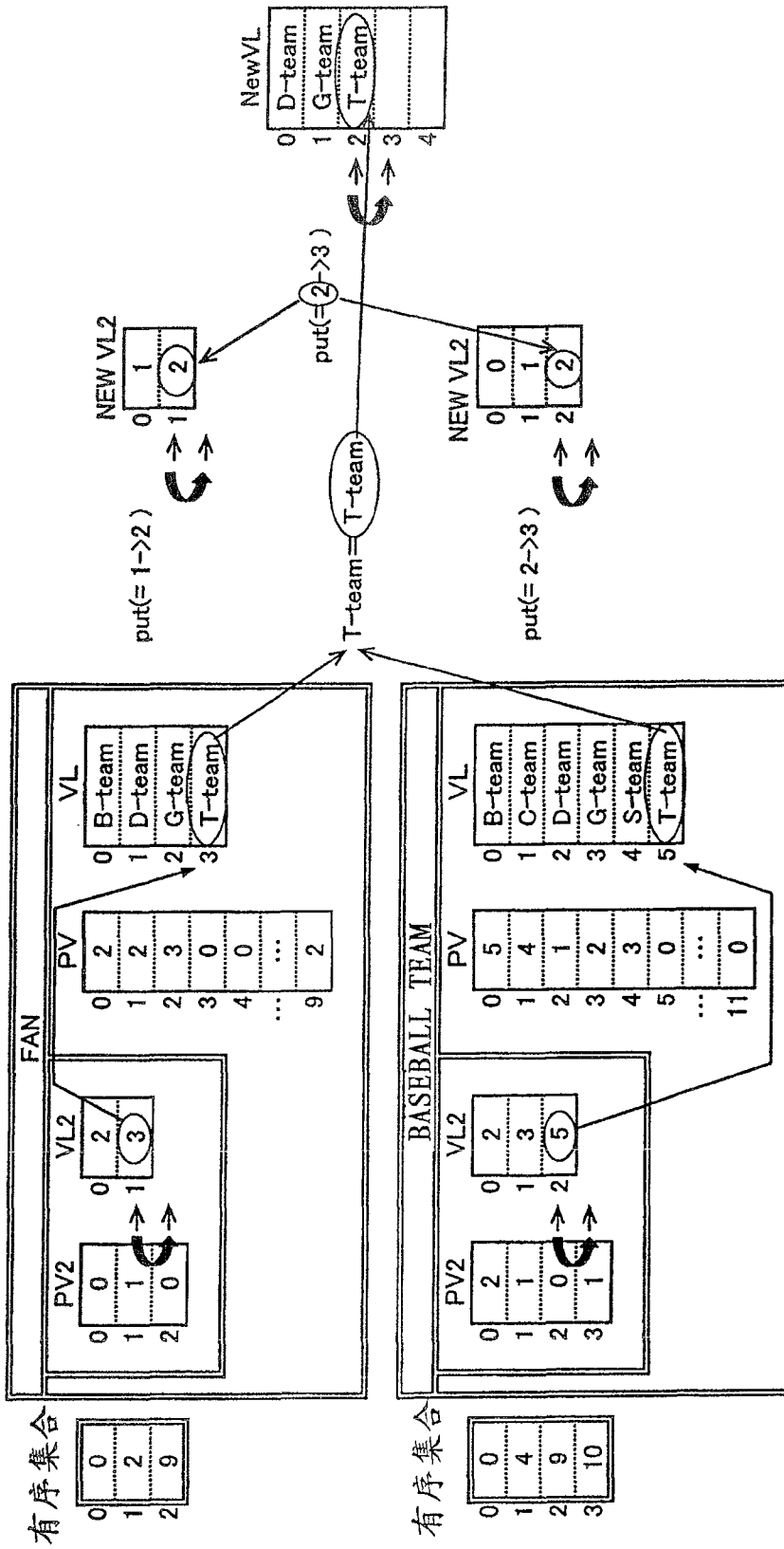


图 30

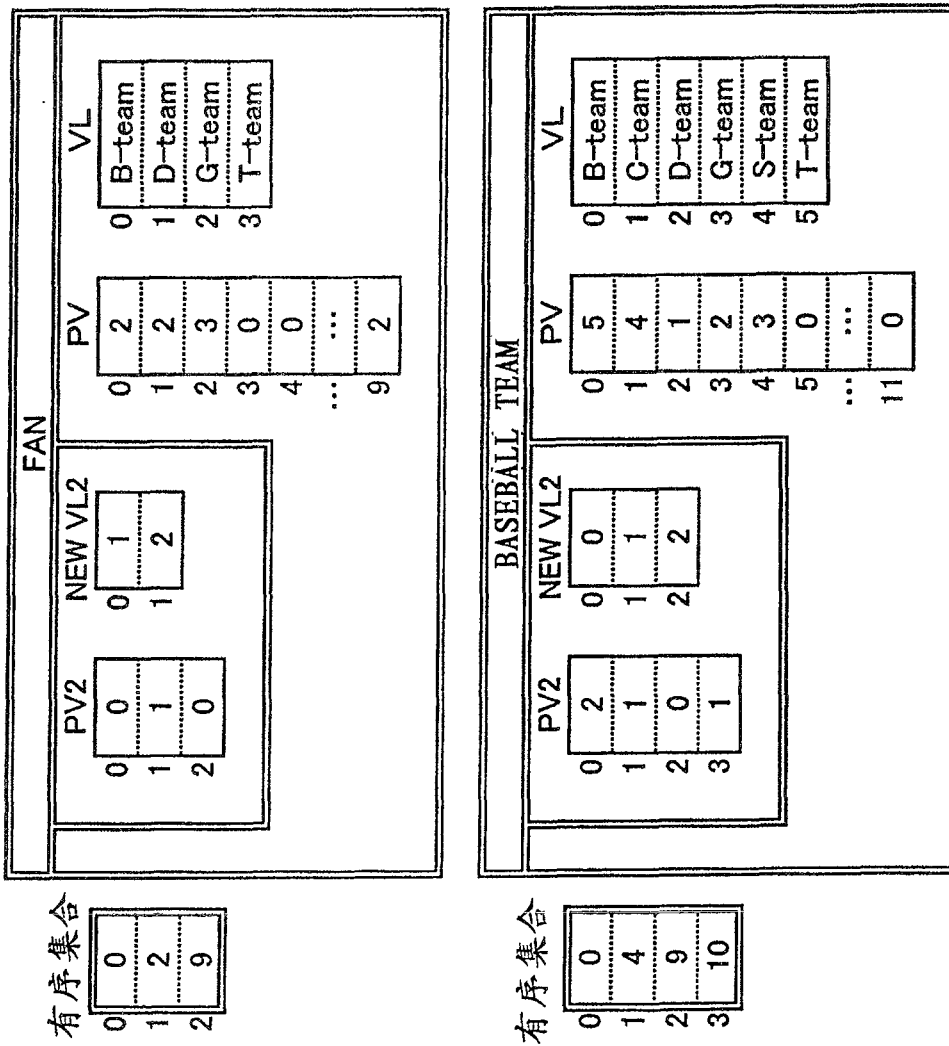


图 31

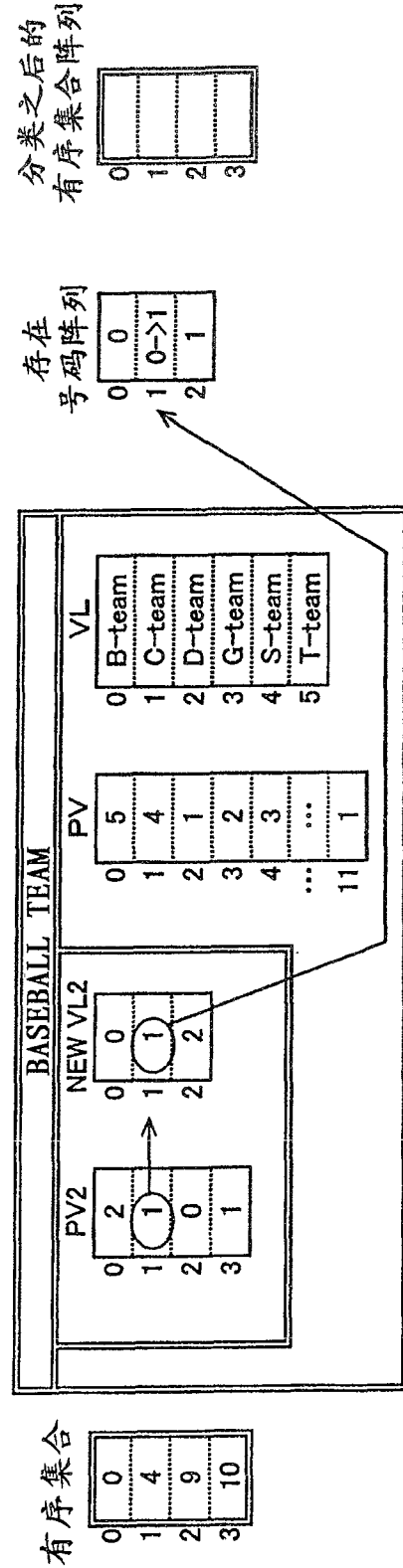
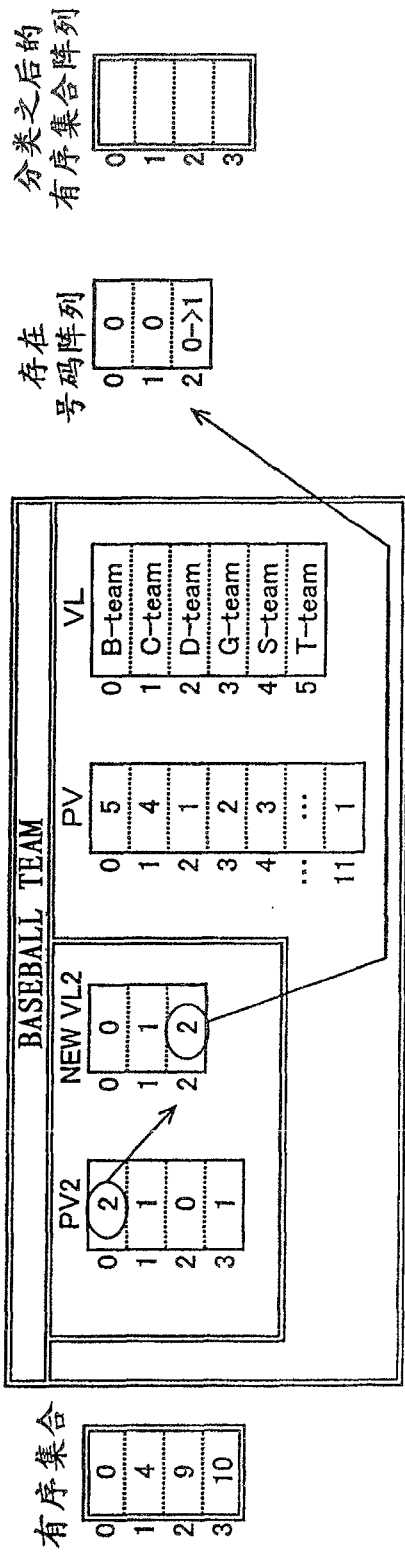
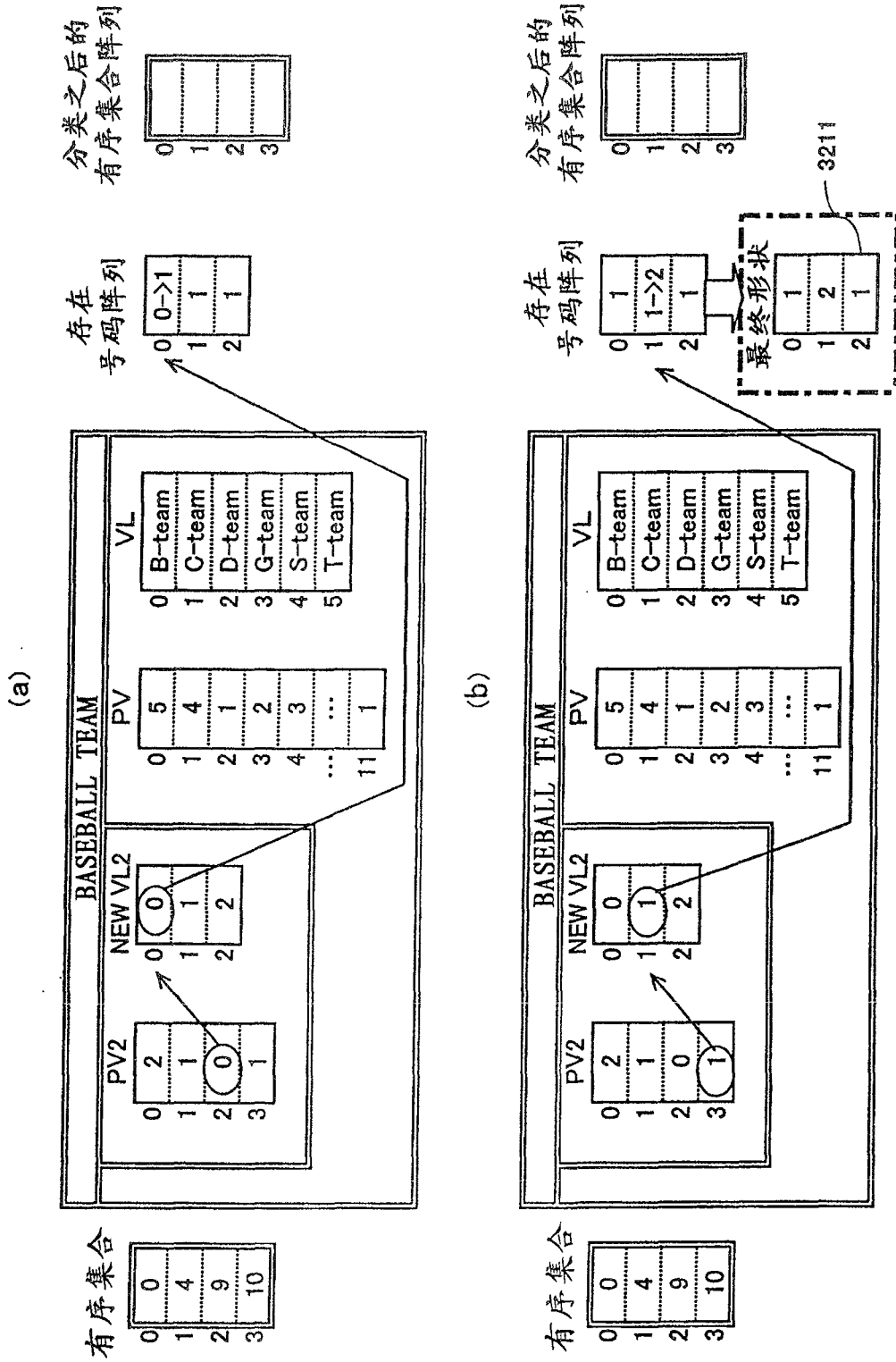




图 32



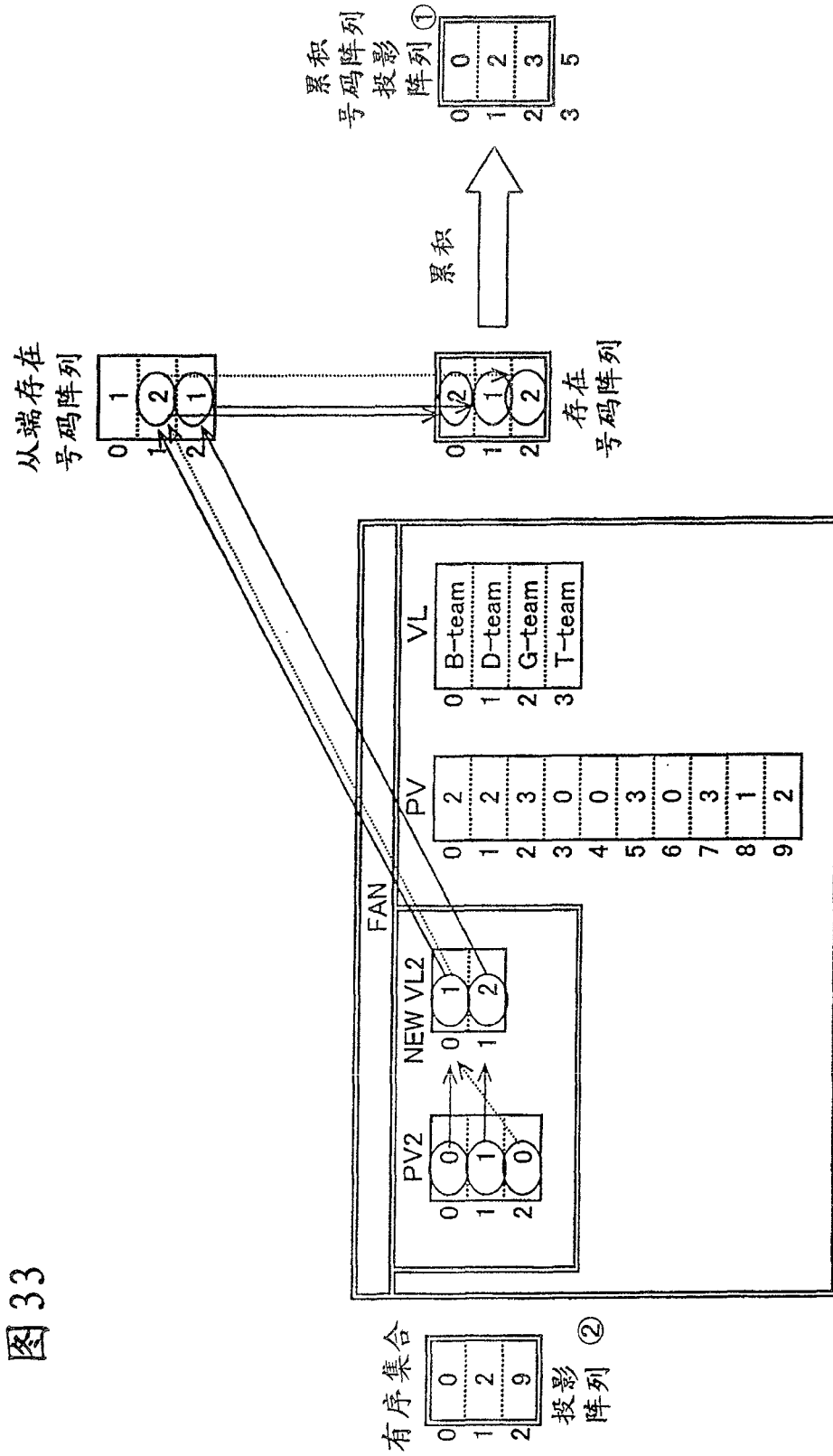


图 33

图 34

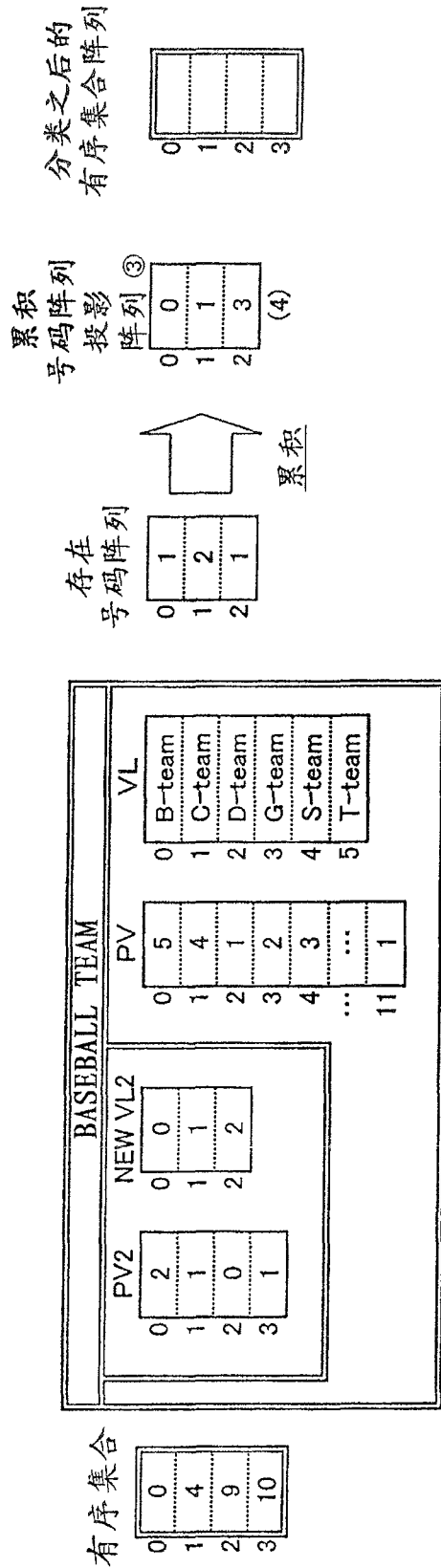


图 35

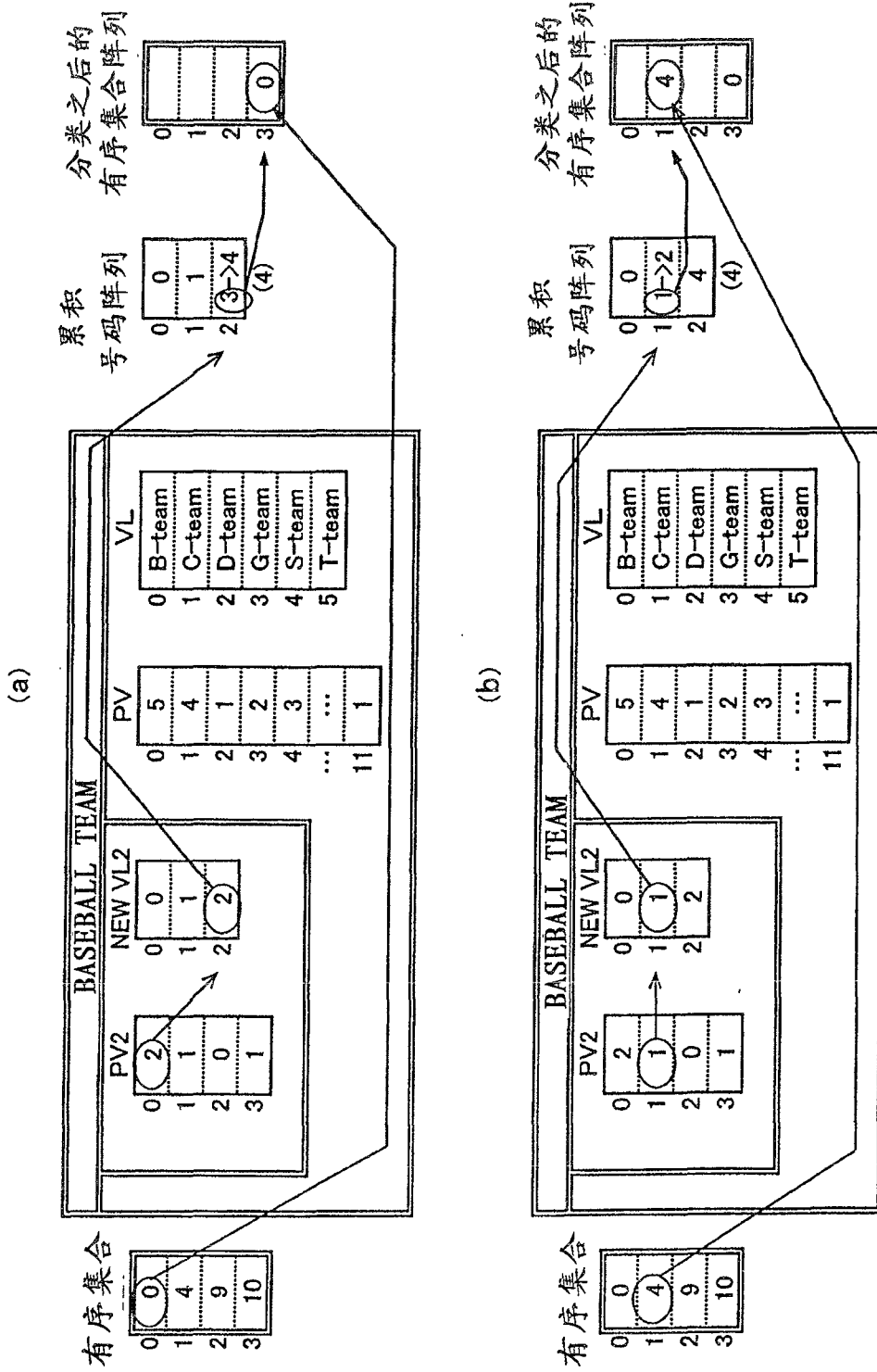


图 36

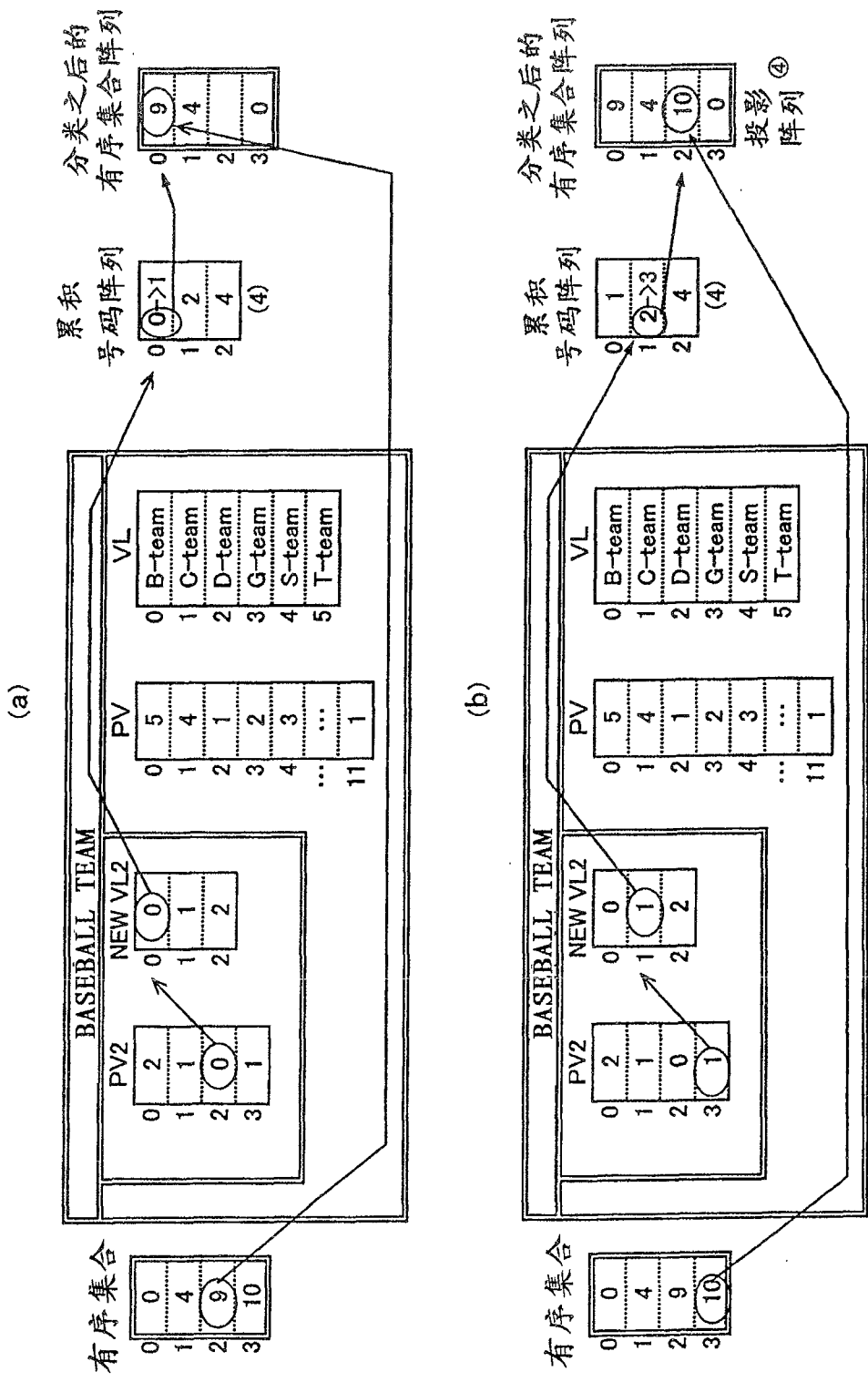


图 37

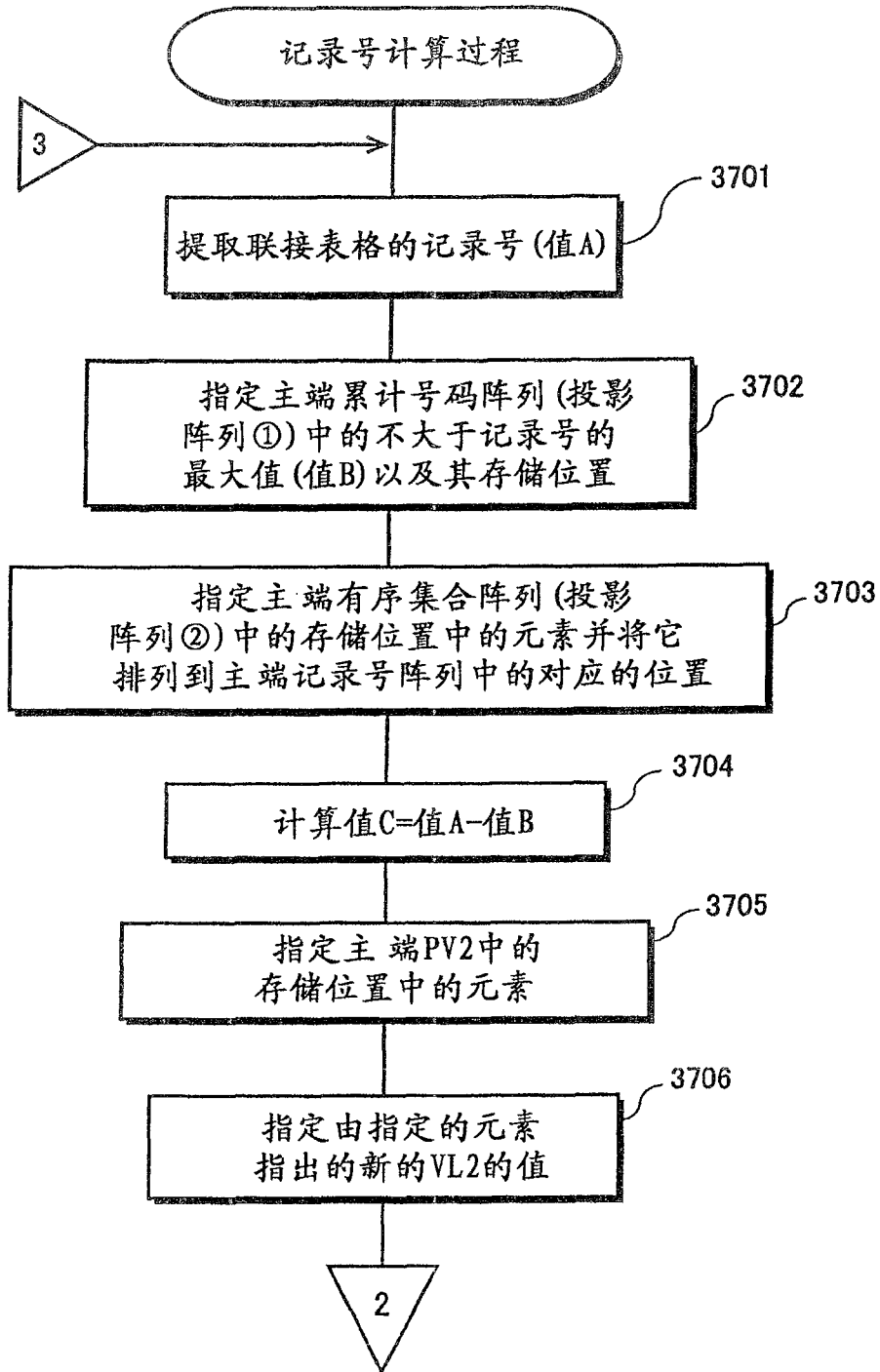


图 38

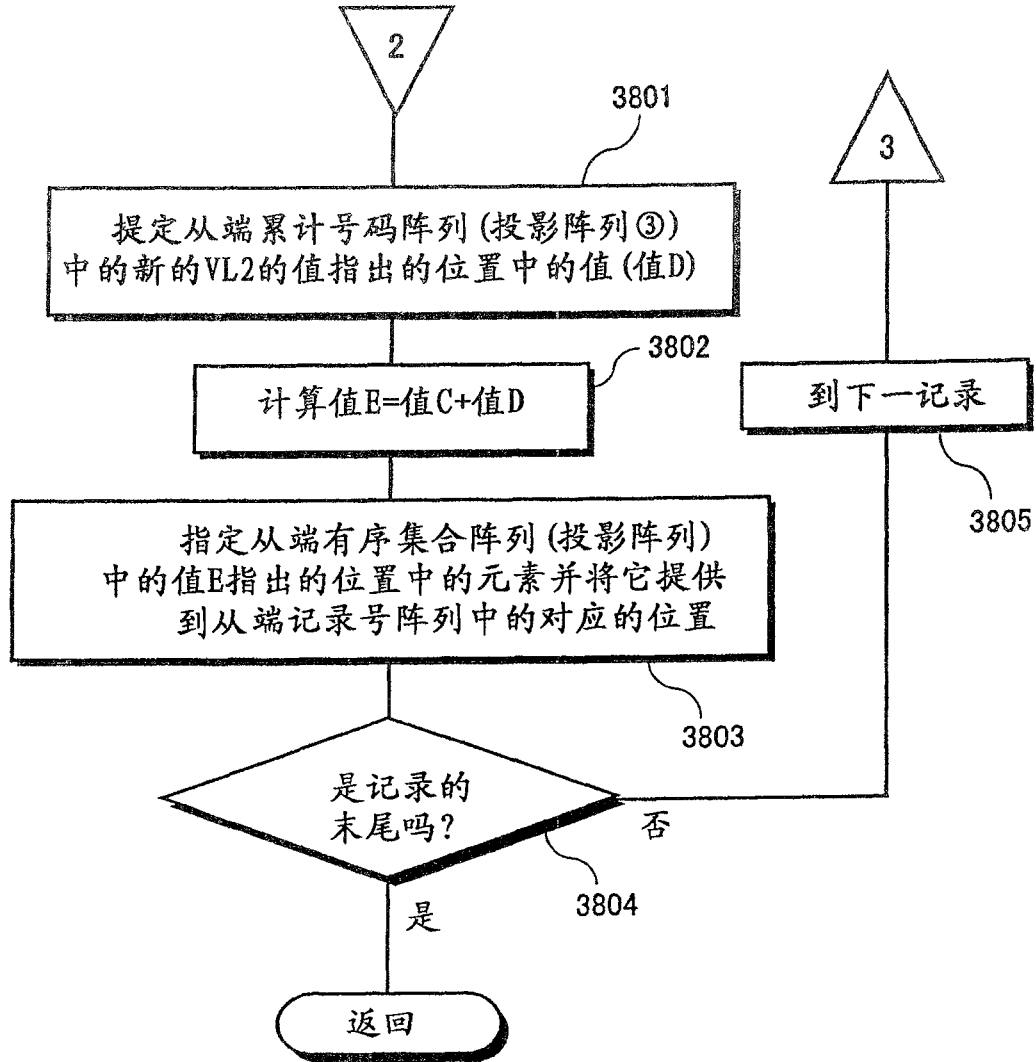


图 39

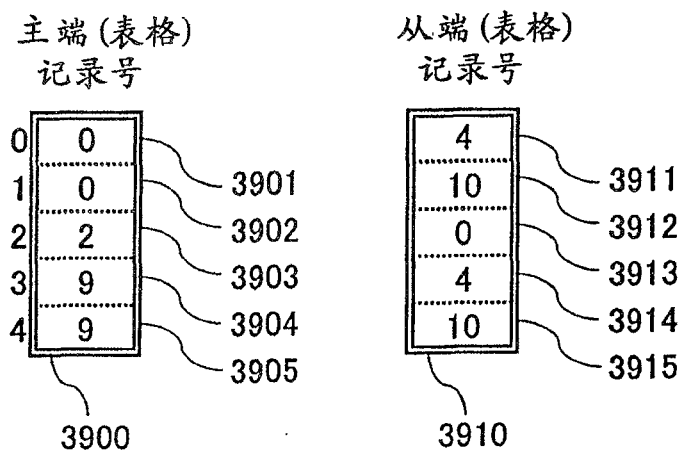




图 40

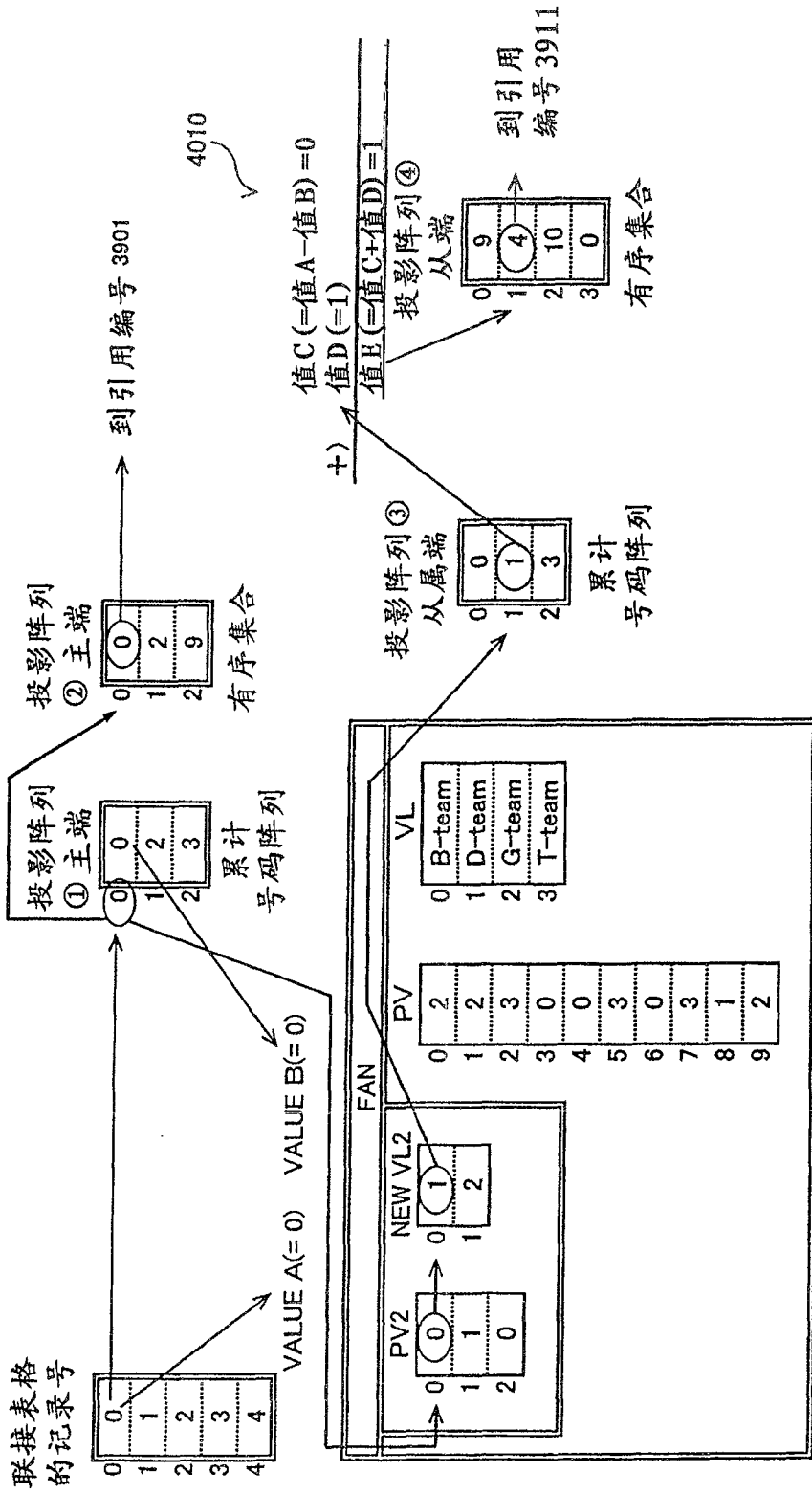


图41

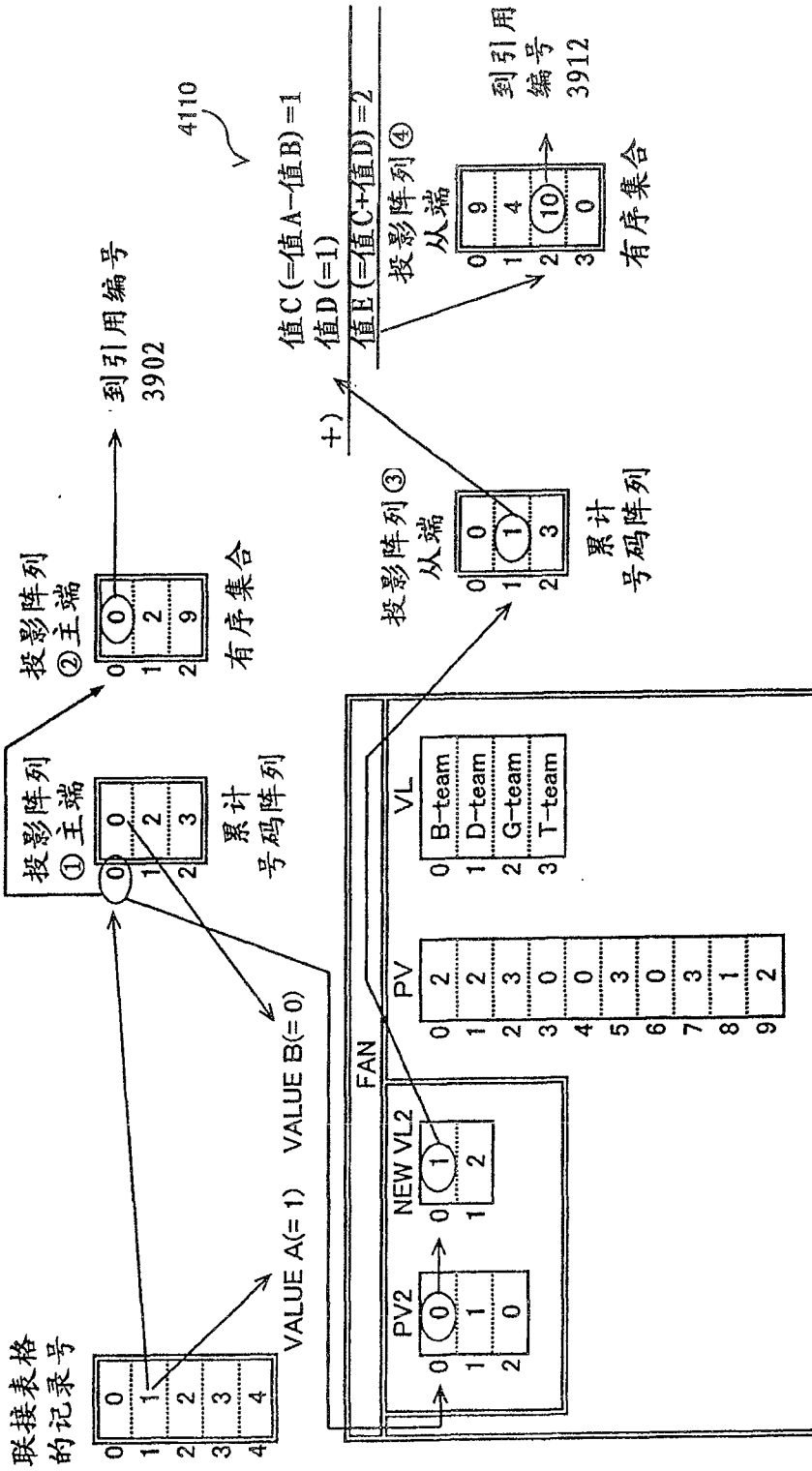


图42

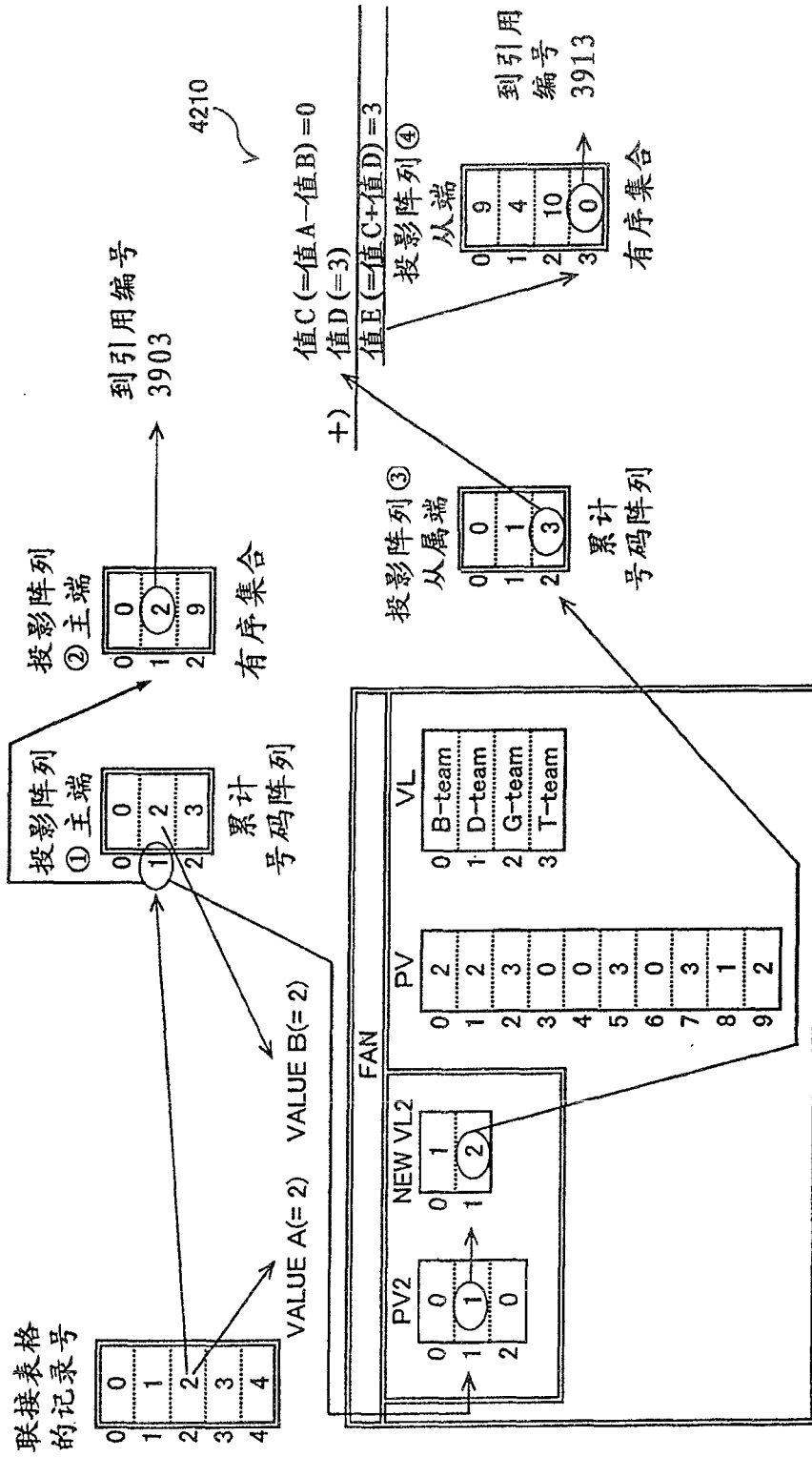


图 43

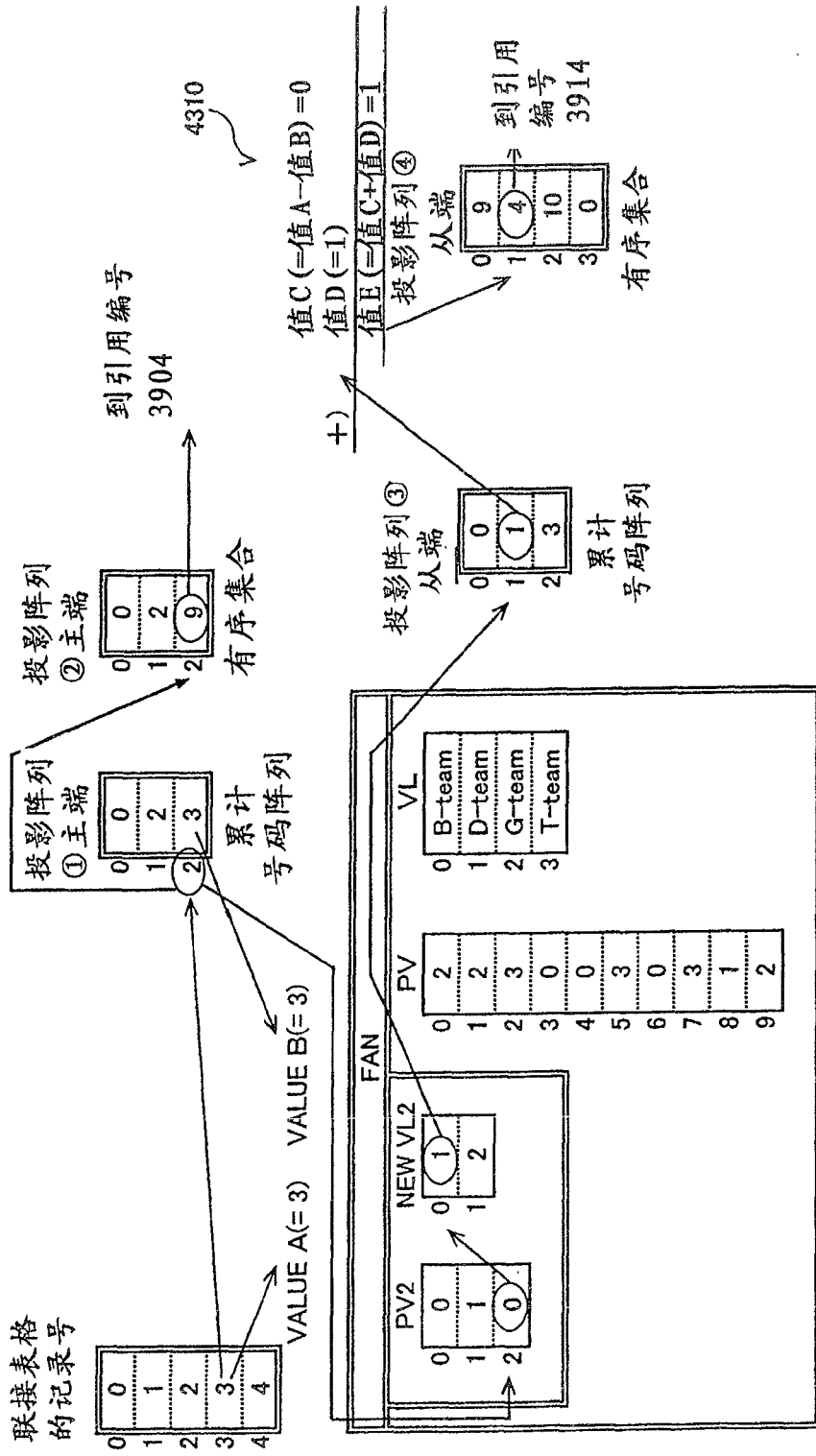


图 44

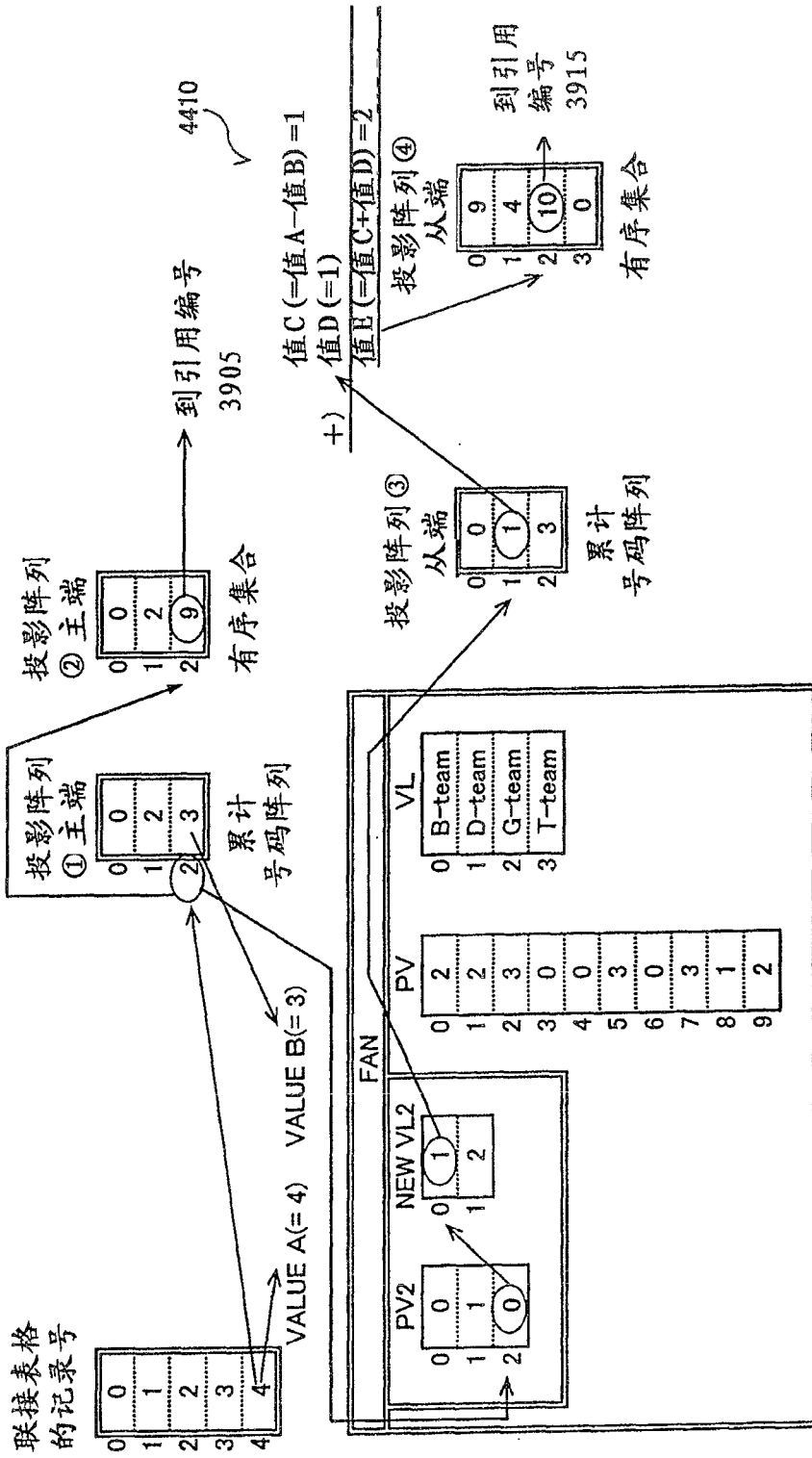


图 45

