202321994

(19)中華民國智慧財產局

(12)發明說明書公開本 (11)公開編號：TW 202321994 A

(43)公開日： 中華民國 112 (2023) 年 06 月 01 日

(21)申請案號：111108074 (22)申請日： 中華民國 111 (2022) 年 03 月 04 日

(51)Int. Cl.： *G06N3/08 (2006.01)* *G06F17/16 (2006.01)*
G11C11/54 (2006.01)

(30)優先權：2021/11/24 美國 63/282,696
2021/11/24 美國 63/282,698

(71)申請人：旺宏電子股份有限公司 (中華民國) MACRONIX INTERNATIONAL CO., LTD. (TW)
新竹縣科學工業園區力行路 16 號

(72)發明人：王韋程 WANG, WEI-CHEN (TW) ；王侑邦 WANG, YU-PANG (TW) ；張原豪 CHANG, YUAN-HAO (TW) ；郭大維 KUO, TEI-WEI (TW)

(74)代理人：祁明輝；林素華

申請實體審查：有 申請專利範圍項數：10 項 圖式數：20 共 37 頁

(54)名稱

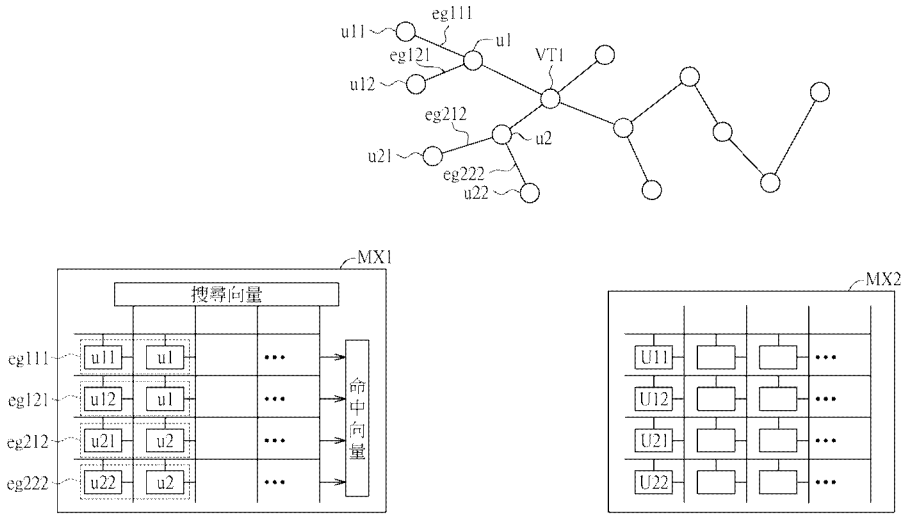
採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置

(57)摘要

一種採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置。三態內容尋址記憶體之圖神經網路的訓練方法包括以下步驟。從一資料集抽取資料。根據資料集之資料，對圖神經網路進行訓練。對圖神經網路進行訓練之步驟包括一特徵擷取階段、一聚合階段及一更新階段。在聚合階段中，一 TCAM 交叉桿矩陣儲存對應於一頂點之數個邊線並輸出一命中向量以選擇部分之邊線，一乘積和交叉桿矩陣儲存這些邊線之數個特徵，以根據命中向量進行一乘積和運算。

A Ternary Content Addressable Memory (TCAM)-based training method for graph neural network and a memory device using the same are provided. The TCAM-based training method for the Graph Neural Network includes the following steps. Data are sampled from a dataset. The Graph Neural Network is trained according to the data from the dataset. The step of training the Graph Neural Network includes a feature extraction phase, an aggregation phase and an update phase. In the aggregation phase, one TCAM crossbar matrix stores a plurality of edges corresponding to one vertex and outputs a hit vector for selecting some of the edges, and a Multiply Accumulate (MAC) crossbar matrix stores a plurality of features in the edges for performing a multiply accumulate operation according to the hit vector.

指定代表圖：



第 6 圖

符號簡單說明：

eg111,eg121,eg212,eg222:邊線

MX1:TCAM 交叉桿矩陣

MX2:MAC 交叉桿矩陣

u1,u2,u11,u12,u21,u22:節點

U11,U12,U21,U22:特徵

VT1:頂點

202321994
專利案號：111108074

【發明摘要】

【中文發明名稱】採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置

【英文發明名稱】TERNARY CONTENT ADDRESSABLE MEMORY (TCAM)-BASED TRAINING METHOD FOR GRAPH NEURAL NETWORK AND MEMORY DEVICE USING THE SAME

【中文】

一種採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置。三態內容尋址記憶體之圖神經網路的訓練方法包括以下步驟。從一資料集抽取資料。根據資料集之資料，對圖神經網路進行訓練。對圖神經網路進行訓練之步驟包括一特徵擷取階段、一聚合階段及一更新階段。在聚合階段中，一 TCAM 交叉桿矩陣儲存對應於一頂點之數個邊線並輸出一命中向量以選擇部分之邊線，一乘積和交叉桿矩陣儲存這些邊線之數個特徵，以根據命中向量進行一乘積和運算。

【英文】

A Ternary Content Addressable Memory (TCAM)-based training method for graph neural network and a memory device using the same are provided. The TCAM-based training method for the Graph Neural Network includes the

following steps. Data are sampled from a dataset. The Graph Neural Network is trained according to the data from the dataset. The step of training the Graph Neural Network includes a feature extraction phase, an aggregation phase and an update phase. In the aggregation phase, one TCAM crossbar matrix stores a plurality of edges corresponding to one vertex and outputs a hit vector for selecting some of the edges, and a Multiply Accumulate (MAC) crossbar matrix stores a plurality of features in the edges for performing a multiply accumulate operation according to the hit vector.

【指定代表圖】第6圖。

【代表圖之符號簡單說明】

eg111, eg121, eg212, eg222: 邊線

MX1: TCAM 交叉桿矩陣

MX2: MAC 交叉桿矩陣

u1, u2, u11, u12, u21, u22: 節點

U11, U12, U21, U22: 特徵

VT1: 頂點

【特徵化學式】

無

【發明說明書】

【中文發明名稱】採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置

【英文發明名稱】TERNARY CONTENT ADDRESSABLE MEMORY (TCAM)-BASED TRAINING METHOD FOR GRAPH NEURAL NETWORK AND MEMORY DEVICE USING THE SAME

【技術領域】

【0001】本揭露是有關於一種神經網路的訓練方法及應用其之記憶體裝置，且特別是有關於一種採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置。

【先前技術】

【0002】隨著人工智慧技術的發展，記憶體內運算技術（Computing in Memory）已應用於單晶片系統（system-on-chip, SoC）。記憶體內運算技術可以加速人工智慧演算法的訓練與辨識。因此，記憶體內運算技術已成為一個重要研發方向。

【0003】然而，透過記憶體進行訓練時，大量的資料遷移可能會降低運算速度。研究人員正致力於改善記憶體內運算技術之訓練效率。

【發明內容】

【0004】本揭露係有關於一種採用三態內容尋址記憶體之圖神經網路的訓練方法及應用其之記憶體裝置，其在抽樣的步驟中運用了自適性資料重覆使用策略，並在聚合階段運用了TCAM資料處理策略與動態固定點格式。於是，資料遷移量能夠大幅降低，且能夠維持準確度。記憶體內運算技術（特別是圖神經網路）的訓練效率能夠有效改善。

【0005】根據本揭露之一方面，提出一種採用三態內容尋址記憶體（Ternary Content Addressable Memory, TCAM）之圖神經網路（Graph Neural Network）的訓練方法。三態內容尋址記憶體之圖神經網路的訓練方法包括以下步驟。從一資料集抽取資料。根據資料集之資料，對圖神經網路進行訓練。對圖神經網路進行訓練之步驟包括一特徵擷取階段（feature extraction phase）、一聚合階段（aggregation phase）及一更新階段（update phase）。在聚合階段中，一TCAM交叉桿矩陣（crossbar matrix）儲存對應於一頂點（vertex）之數個邊線並輸出一命中向量以選擇部分之邊線，一乘積和（Multiply Accumulate, MAC）交叉桿矩陣儲存這些邊線之數個特徵，以根據命中向量進行一乘積和運算。

【0006】根據本揭露之另一方面，提出一種記憶體裝置。記憶體裝置包括一控制器及一記憶體陣列。記憶體陣列連接於控制

器。在記憶體陣列中，一三態內容尋址記憶體（Ternary Content Addressable Memory, TCAM）交叉桿矩陣（crossbar matrix）儲存對應於一頂點（vertex）之數個邊線並輸出一命中向量以選擇部分之這些邊線，一乘積和（Multiply Accumulate, MAC）交叉桿矩陣儲存這些邊線之數個特徵，以根據命中向量進行一乘積和運算。

【0007】為了對本揭露之上述及其他方面有更佳的瞭解，下文特舉實施例，並配合所附圖式詳細說明如下：

【圖式簡單說明】

【0008】

第1圖繪示應用圖神經網路之一圖形的一個例子。

第2圖繪示根據一實施例之採用TCAM之圖神經網路的訓練方法的流程圖。

第3圖繪示執行步驟S110之一個例子。

第4圖說明特徵擷取階段、聚合階段與更新階段。

第5圖繪示一交叉桿矩陣（crossbar matrix）。

第6圖繪示一TCAM交叉桿矩陣與一乘積和交叉桿矩陣。

第7～10圖說明TCAM交叉桿矩陣與MAC交叉桿矩陣。

第11～13圖說明在數個批次之TCAM交叉桿矩陣及MAC交叉桿矩陣。

第14圖說明TCAM資料處理策略之管線運算架構。

第15圖說明動態固定點格式。

第16圖說明拔靴法。

第17圖說明圖形分割法。

第18圖說明非均勻拔靴法。

第19圖繪示根據一實施例之自適性資料重覆使用策略的流程圖。

第20圖繪示採用上述訓練方法的記憶體裝置。

【實施方式】

【0009】在本揭露之實施例中，提供一種採用三態內容尋址記憶體（Ternary Content Addressable Memory, TCAM）之圖神經網路（Graph Neural Network）的訓練方法。請參照第1圖，其繪示應用圖神經網路之一圖形GP的一個例子。圖形GP包括數個頂點（vertex）VTi及數個節點（node）Nj。頂點VTi與節點Nj可以是某一個人、某一組織、某一部門。頂點VTi與節點Nj之間的邊線（edge）儲存其特徵（features）。圖神經網路可以用來識別或分析兩個頂點VTi之間的關係。

【0010】採用TCAM之圖神經網路的訓練方法可以改善記憶體內運算技術的訓練效率。請參照第2圖，其繪示根據一實施例之採用TCAM之圖神經網路的訓練方法的流程圖。在步驟S110中，從一資料集900抽取資料。請參照第3圖，其繪示執行步驟S110

之一個例子。在第3圖中，係以數個批次BCq來執行訓練步驟的數次迭代。

【0011】在步驟S120中，根據資料集900之資料，對圖神經網路進行訓練。步驟S120包括一特徵擷取階段（feature extraction phase）P1、一聚合階段（aggregation phase）P2及一更新階段（update phase）P3。請參照第4圖，其說明特徵擷取階段P1、聚合階段P2與更新階段P3。在特徵擷取階段P1中，擷取邊線與節點上之特徵。在聚合階段P2中，則是進行乘積和（Multiply Accumulate）等數種運算。在更新階段P3，對權重進行更新。聚合階段P2係為密集輸入/輸出的工作，而容易形成大量的資料遷移，故訓練效能的瓶頸通常發生在聚合階段P2。

【0012】為了增進訓練效率，在從資料集900抽取資料之步驟S110採用了自適性資料重覆使用策略（adaptive data reusing policy自適性資料重覆使用策略），並且於聚合階段P2採用了TCAM資料處理策略（data processing strategy）及動態固定點格式（dynamic fixed-point formatting）。以下先說明TCAM資料處理策略與動態固定點格式，接著再說明自適性資料重覆使用策略。

【0013】應用於聚合階段P2之TCAM資料處理策略包括一頂點內平行運算架構（intra-vertex parallelism architecture）及一頂點間平行運算架構（inter-vertex parallelism architecture）。請參照第5圖，其繪示一交叉桿矩陣（crossbar

matrix）MX。在此實施例中，數個特徵x11、x12、x13、x21、x22、x23、x31、x32、x33可以儲存於交叉桿矩陣MX中。交叉桿矩陣MX例如是電阻式隨機存取記憶體（Resistive random-access memory, ReRAM）。交叉桿矩陣MX包括數條字元線WL1、WL2、WL3、數條位元線BT1、BT2、BT3及數個記憶胞。記憶胞儲存這些特徵x11、x12、x13、x21、x22、x23、x31、x32、x33，而不是儲存權重。在聚合階段P2中，數個係數a1、a2、a3輸入至字元線WL1、WL2、WL3，數個乘積和結果v1、v2、v3可以從位元線BL1、BT2、BT3獲得。0或1可以用來選擇任何節點X1、X2、X3。如第4圖所示，[1, 0, 1]係為選擇節點X1、X3之命中向量HV。

【0014】請參照第6圖，其繪示一TCAM交叉桿矩陣MX1與一乘積和交叉桿矩陣MX2。在聚合階段P2中，TCAM交叉桿矩陣MX1儲存對應於頂點VT1之數個邊線eg111、eg121、eg212、eg222、...並輸出選擇部分邊線eg111、eg121、eg212、eg222、...之命中向量HV。邊線eg111包括來源節點u11與目的節點u1。邊線eg121包括來源節點u12與目的節點u1。邊線eg212包括來源節點u21與目的節點u2。邊線eg222包括來源節點u22與目的節點u2。

【0015】在頂點內平行運算架構下，MAC交叉桿矩陣MX2儲存邊線eg111、eg121、eg212、eg222、...之特徵U11、U12、

U21、U22、...，以根據命中向量HV執行乘積和運算。以下透過圖示說明數個例子。

【0016】請參照第7～10圖，其說明TCAM交叉桿矩陣MX1與MAC交叉桿矩陣MX2。如第7圖所示，一搜尋向量SV1輸入至TCAM交叉桿矩陣MX1。搜尋向量SV1之內容為目的節點u1。邊線eg111之目的節點u1匹配於搜尋向量SV1，故輸出1。邊線eg121之目的節點u1匹配於搜尋向量SV1，故輸出1。邊線eg212之目的節點u2並不匹配於搜尋向量SV1，故輸出0。邊線eg222之目的節點u2並不匹配於搜尋向量SV1，故輸出0。因此，內容為[1, 1, 0, 0]之命中向量HV1輸入至MAC交叉桿矩陣MX2。

【0017】命中向量HV1輸入至MAC交叉桿矩陣MX2，以選擇特徵U11、U12。如第7圖所示，獲得了一乘積和結果U1(1)（乘積和結果U1(1)=特徵U11+特徵U12）。

【0018】如第8圖所示，一搜尋向量SV2輸入至TCAM交叉桿矩陣MX1。搜尋向量SV2的內容為目的節點u2。邊線eg111之目的節點u1並不匹配於搜尋向量SV2，故輸出0。邊線eg121之目的節點u1並不匹配於搜尋向量SV2，故輸出0。邊線eg212之目的節點u2匹配於搜尋向量SV2，故輸出1。邊線eg222之目的節點u2匹配於搜尋向量SV2，故輸出1。因此，內容為[0, 0, 1, 1]之命中向量HV2輸入至MAC交叉桿矩陣MX2。

【0019】命中向量HV2輸入至MAC交叉桿矩陣MX22，以選擇特徵U21、U22。如第8圖所示，獲得了乘積和結果U2(1)（乘積和結果U2(1)=特徵U21+特徵U22）。

【0020】如第9圖所示，一TCAM交叉桿矩陣MX21可以更儲存頂點VT1、...、階層L0、L1、...與邊線eg11、eg21。邊線eg111、eg121、eg212、eg222對應於頂點VT1與階層L0。邊線eg11、eg21對應於頂點VT1與階層L1。一搜尋向量SV3輸入至TCAM交叉桿矩陣MX21。搜尋向量SV3之內容係為頂點VT1及階層L0。頂點VT1、階層L0與對應之邊線eg111、eg212匹配於搜尋向量SV3，故輸出1。頂點VT1、階層L0與對應之邊線eg121、eg222匹配於搜尋向量SV3，故輸出1。頂點VT1、階層L1與對應之邊線eg11並未匹配於搜尋向量SV3，故輸出0。頂點VT1、階層L1與對應之邊線eg21並未匹配於搜尋向量SV3，故輸出0。因此，內容為[1, 1, 0, 0]之命中向量HV3輸出至MAC交叉桿矩陣MX22。

【0021】命中向量HV3輸入至MAC交叉桿矩陣MX22，以選擇特徵U11、U21，並選擇特徵U12、U22。如第9圖所示，獲得了乘積和結果U1(1)、U2(1)。

【0022】如第10圖所示，MAC交叉桿矩陣MX22更儲存對應於邊線eg11、eg21之乘積和結果U1(1)、U2(1)。一搜尋向量SV4輸入至TCAM交叉桿矩陣MX21。搜尋向量SV4之內容係為頂點VT1及階層L1。頂點VT1、階層L0與對應之邊線eg111、eg212並未匹配於搜尋向量SV4，故輸出0。頂點VT1、階層L0與對應之

邊線eg121、eg222並未匹配於搜尋向量SV4，故輸出0。頂點VT1、階層L1與對應之邊線eg11匹配於搜尋向量SV4，故輸出1。頂點VT1、階層L1與對應之邊線eg21匹配於搜尋向量SV4，故輸出1。因此，內容為[0, 0, 1, 1]之命中向量HV4輸出至MAC交叉桿矩陣MX22。

【0023】命中向量HV4輸入至MAC交叉桿矩陣MX22，以選擇乘積和結果U1(1)、U2(1)。如第10圖所示，獲得了乘積和結果。

【0024】在一實施例中，在頂點間平行運算架構下，TCAM交叉桿矩陣MX21可以更儲存對應於另一頂點之邊線。搜尋向量可以用來選擇頂點。

【0025】如上所述，在頂點間平行運算架構之下，儲存單元（bank）/矩陣層級平行架構可以應用於不同頂點之聚合運算。在頂點內平行運算架構之下，交叉桿矩陣之寬度可以有效率地利用，以分散聚合運算。

【0026】請參照第11～13圖，其說明在數個批次B1、B2、...、Bk之TCAM交叉桿矩陣MX311、MX312、...及MAC交叉桿矩陣MX321、MX322、...。如第11圖所示，數個TCAM交叉桿矩陣MX311、MX312、...與數個MAC交叉桿矩陣MX321、MX322、...設置數個記憶區塊（bank）中。對於批次B1，記憶體區域A3111用以儲存頂點VT31之邊線，記憶體區域A3211用以儲存頂點VT31之特徵。記憶體區域A3121用以儲存頂點VT32之邊線，記憶體區域A3221用以儲存頂點VT32之特徵。

【0027】如第12圖所示，對於批次B2，記憶體區域A3112用以儲存頂點VT33之邊線，記憶體區域A3212用以儲存頂點VT33之特徵。記憶體區域A3122用以儲存頂點VT34之邊線，記憶體區域A3222用以儲存頂點VT34之特徵。

【0028】如第13圖所示，對於批次Bk，記憶體區域A3111用以儲存頂點VT35之邊線，記憶體區域A3211用以儲存頂點VT35之特徵。記憶體區域A3121用以儲存頂點VT36之邊線，記憶體區域A3221用以儲存頂點VT36之特徵。也就是說，相同的記憶體區域可以被重複使用於不同的頂點，使得記憶體能夠有效利用。

【0029】在某些情況下，MAC交叉桿矩陣的寬度可能不夠儲存一個節點或一個頂點的特徵。為了避免速度下降，在此可以運用管線運算架構。請參照第14圖，其說明TCAM資料處理策略之管線運算架構。如第14圖所示，特徵U11被分為兩個分段pt21、pt22且儲存於兩列中。邊線eg111儲存於TCAM交叉桿矩陣MX41的兩列中。分段pt21、pt22的排列是獨立的。在時間點T1，對分段pt21執行聚合階段P2；在時間點T2，對分段pt21可以開始執行更新階段P3。在時間點T2，對分段pt22執行聚合階段P2；在時間點T3，對分段pt22可以開始執行更新階段P3。

【0030】此外，在聚合階段P2中，更可以進一步採用動態固定點格式。儲存於交叉桿矩陣之權重或特徵可能具有浮點格式。在本技術中，權重或特徵可以利用動態固定點格式儲存於交叉桿

矩陣中。請參照第15圖，其說明動態固定點格式。如表一所示，權重可以表示為浮點格式。

權重	浮點格式	尾數	指數
0.2165	1.10111011x2^-3	10111011	2^-3
0.214	1.10110110x2^-3	10110110	2^-3
0.202	1.10011101x2^-3	10011101	2^-3
0.0096	1.00111010x2^-7	00111010	2^-7
0.472	1.11100011x2^-2	11100011	2^-2

表一

【0031】指數的範圍從2^-0到2^-7。在此實施例中，指數可以分為兩個群組G0、G1。群組G0係為2^-0到2^-3，群組G1係為2^-4到2^-7。如第15圖所示，當指數落於群組G0，則以「0」儲存；當指數落於群組G1，則以「1」儲存。為了精確表示出「2^-0」，尾數位移0個位元。為了精確表示出「2^-1」，尾數位移1個位元。為了精確表示出「2^-2」，尾數位移2個位元。為了精確表示出「2^-3」，尾數位移3個位元。為了精確表示出「2^-4」，尾數位移0個位元。為了精確表示出「2^-5」，尾數位移1個位元。為了精確表示出「2^-6」，尾數位移2個位元。為了精確表示出「2^-7」，尾數位移3個位元。舉例來說，權重wt1係為「0.2165」，其尾數係為「10111011」，末位元的「0」表示群組G0，且尾數「10111011」被位移3個位元，以精確表示出「2^-3」。權重wt2係為「0.472」，

其尾數係為「11100011」，末位元的「0」表示群組G0，且尾數「11100011」被位移2個位元，以精確表示出「2^-2」。

【0032】根據動態固定點格式，7種指數被分類至僅有兩個群組G0、G1，故運算週期數可以從7降到2，大幅增加的運算速度。

【0033】此外，以下更進一步說明步驟S110之自資料集900抽樣資料之自適性資料重覆使用策略。自適性資料重覆使用策略包括拔靴法（Bootstrapping）、圖形分割法（Graph Partitioning）及非均勻拔靴法（non-uniform Bootstrapping）。

【0034】請參照第16圖，其說明拔靴法。各個批次BC1、BC2、BC3、BC4用以進行一次迭代。批次BC1包括節點N1、N2、N5之資料；批次BC2包括節點N1、N3、N6之資料；批次BC3包括節點N5、N3、N6之資料；批次BC4包括節點N4、N3、N2之資料。節點N1之資料重複使用於批次BC1與批次BC2。節點N3之資料重複使用於批次BC3與批次BC4。

【0035】根據拔靴法，部分資料重複使用於兩個批次中，故資料遷移的次數可以有效降低，且訓練效率能夠改善。

【0036】請參照第17圖，其說明圖形分割法。在一圖形中，圖形尺寸（即所有節點之數量）係為n，且批次尺寸（即一批次中節點的數量）係為b。重複使用率係為b/n。如果重複使用率太低，拔靴法將無法獲得大幅的改善，故圖形需要進一步分割以提升重複使用率。如第17圖所示，圖形內的節點被隨機地分割入3個分區。重複使用率將會提升至3倍。節點N11～N145之資料被排入

批次BC11～BC13。節點N12與節點N14之資料重複使用於批次BC11與批次BC12。節點N13與節點N14之資料重複使用於批次BC12與批次BC13。

【0037】節點N21～N25之資料被排入批次BC21～BC23。節點N23與節點N25之資料重複使用於批次BC21與批次BC22。節點N21之資料重複使用於批次BC22與批次BC23。

【0038】根據圖形分割法，重複使用率能夠提升，且即使圖形過大時，拔靴法仍然具有有效的改善能力。

【0039】請參照第18圖，其說明非均勻拔靴法。在拔靴法中，節點的資料被重複使用，故某些節點可能會被抽樣過多次而影響到準確度。如第18圖所示，節點之抽樣機率設置為不均勻。在數次迭代後，節點N8之抽樣次數超出界線，故節點N8之抽樣機率被降低至0.826%（低於其他節點之抽樣機率）。

【0040】根據非均勻拔靴法，任何節點不會過度抽樣且能夠維持準確度。

【0041】上述自適性資料重覆使用策略之拔靴法、圖形分割法與非均勻拔靴法可以透過以下流程圖來執行。請參照第19圖，其繪示根據一實施例之自適性資料重覆使用策略的流程圖。在步驟S111中，判斷重複使用率是否低於一預定值。若重複使用率低於預定值，則進入步驟S112；若重複使用率不低於預定值，則進入步驟S113。

【0042】在步驟S112中，執行圖形分割法。

【0043】在步驟S113中，判斷是否有任一節點之抽樣次數超出界線。若有節點之抽樣次數超出界線，則進入步驟S114；若所有節點之抽樣次數均未超出界線，則進入步驟S115。

【0044】在步驟S114中，執行非均勻拔靴法。

【0045】在步驟S115中，執行（均勻）拔靴法。

【0046】再者，請參照第20圖，其繪示採用上述訓練方法的記憶體裝置1000。記憶體裝置1000包括一控制器100及一記憶體陣列200。記憶體陣列200連接於控制器100。記憶體陣列200包括至少一TCAM交叉桿矩陣MXm1及至少一MAC交叉桿矩陣MXm2。TCAM交叉桿矩陣MXm1儲存對應於一頂點之邊線egij。TCAM交叉桿矩陣MXm1接收搜尋向量SVt後，輸出命中向量HVt，以選擇部分邊線egij。MAC交叉桿矩陣MXm2儲存邊線egij之特徵，以根據命中向量HVt執行乘積和運算。

【0047】根據上述實施例，在採用TCAM三態內容尋址記憶體之圖神經網路的訓練方法中，在抽樣的步驟S110中運用了自適性資料重覆使用策略，並在聚合階段P2運用了TCAM資料處理策略與動態固定點格式。於是，資料遷移量能夠大幅降低，且能夠維持準確度。記憶體內運算技術（特別是圖神經網路）的訓練效率能夠有效改善。

【0048】綜上所述，雖然本揭露已以實施例揭露如上，然其並非用以限定本揭露。本揭露所屬技術領域中具有通常知識者，

在不脫離本揭露之精神和範圍內，當可作各種之更動與潤飾。因此，本揭露之保護範圍當視後附之申請專利範圍所界定者為準。

【符號說明】

【0049】

900: 資料集

1000: 記憶體裝置

100: 控制器

200: 記憶體陣列

a1, a2, a3: 係數

A3111, A3211, A3121, A3221, A3112, A3212, A3122, A3222: 記憶體區域

B1, B2, Bk, BC1, BC2, BC3, BC4, BC11, BC12, BC13, BC21, BC22, BC23, BCq: 批次

BL1, BL2, BL3: 位元線

egij, eg111, eg121, eg212, eg222, eg11, eg21: 邊線

G0, G1: 群組

GP: 圖形

HV, HV1, HV2, HVt: 命中向量

L0, L1: 階層

MX: 交叉桿矩陣

MX1, MX21, MX311, MX312, MX41, MXm1: TCAM 交叉桿矩陣

MX2, MX22, MX321, MX322, MX42, MXm2: MAC 交叉桿矩陣

Nj, N1, N2, N3, N4, N5, N6, N8, N11, N12, N13, N14, N21, N22, N23, N24, N25, N31, N32, N33, N34: 節點

P1: 特徵擷取階段

P2: 聚合階段

P3: 更新階段

pt21, pt22: 分段

S110, S111, S112, S113, S114, S115, S120: 步驟

SV1, SV2, SV3, SV4, SVt: 搜尋向量

T1, T2, T3: 時間點

u1, u2, u11, u12, u21, u22: 節點

U11, U12, U21, U22: 特徵

U1(1), U2(1), v1, v2, v3: 乘積和結果

VTi, VT1, VT31, VT32, VT33, VT34, VT35, VT36: 頂點

WL1, WL2, WL3: 字元線

wt1, wt2: 權重

X1, X2, X3: 節點

x11, x12, x13, x21, x22, x23, x31, x32, x33: 特徵

【發明申請專利範圍】

【請求項1】一種採用三態內容尋址記憶體(Ternary Content Addressable Memory, TCAM）之圖神經網路（Graph Neural Network）的訓練方法，包括：

從一資料集抽取資料；以及

根據該資料集之該資料，對該圖神經網路進行訓練，其中對該圖神經網路進行訓練之步驟包括：

一特徵擷取階段（feature extraction phase）；

一聚合階段（aggregation phase）；及

一更新階段（update phase）；

在該聚合階段中，一 TCAM 交叉桿矩陣（crossbar matrix）儲存對應於一頂點（vertex）之複數個邊線並輸出一命中向量以選擇部分之該些邊線，一乘積和（Multiply Accumulate, MAC）交叉桿矩陣儲存該些邊線之複數個特徵，以根據該命中向量進行一乘積和運算。

【請求項 2】如請求項 1 所述採用三態內容尋址記憶體之圖神經網路的訓練方法，其中該 TCAM 交叉桿矩陣儲存各該邊線之一來源節點與一目的節點。

【請求項 3】如請求項 2 所述採用三態內容尋址記憶體之圖神經網路的訓練方法，其中該交叉桿矩陣更儲存對應於另一頂點之複數個邊線。

【請求項 4】如請求項 1 所述採用三態內容尋址記憶體之圖神經網路的訓練方法，其中該些特徵或複數個權重之每一個具有一尾數（mantissa）及一指數（exponent），各該指數被歸類至二群組之其中之一，並且各該尾數根據各該指數偏移。

【請求項 5】如請求項 1 所述採用三態內容尋址記憶體之圖神經網路的訓練方法，其中在從該資料集抽取資料之步驟中，至少一節點之資料重複使用於兩個批次（batch）。

【請求項 6】一種記憶體裝置，包括：

一控制器；以及

一記憶體陣列，連接於該控制器，其中在該記憶體陣列中，一三態內容尋址記憶體（Ternary Content Addressable Memory, TCAM）交叉桿矩陣（crossbar matrix）儲存對應於一頂點（vertex）之複數個邊線並輸出一命中向量以選擇部分之該些邊線，一乘積和（Multiply Accumulate, MAC）交叉桿矩陣儲存該些邊線之複數個特徵，以根據該命中向量進行一乘積和運算。

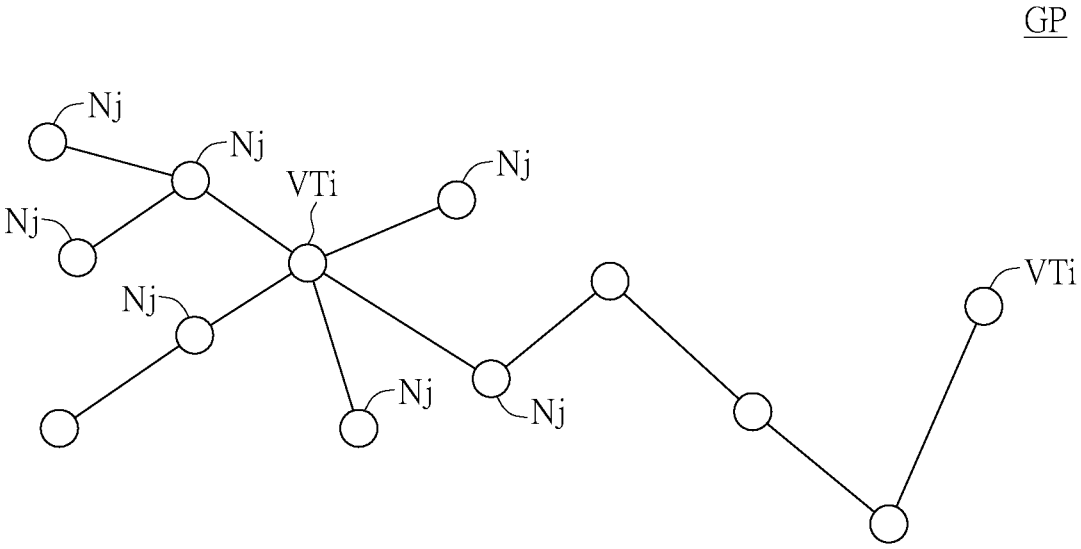
【請求項 7】如請求項 6 所述之記憶體裝置，其中該 TCAM 交叉桿矩陣儲存各該邊線之一來源節點與一目的節點。

【請求項 8】如請求項 7 所述之記憶體裝置， 其中該交叉桿矩陣更儲存對應於另一頂點之複數個邊線。

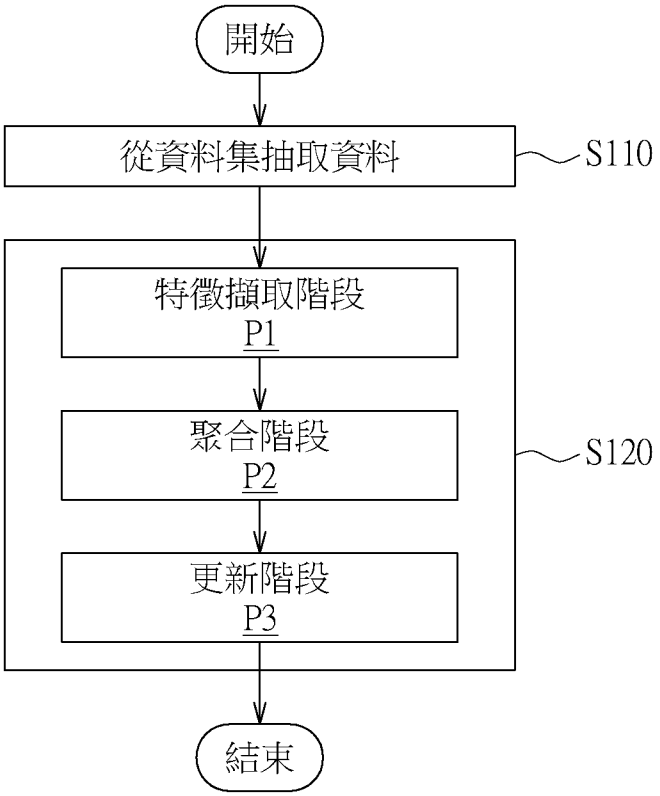
【請求項 9】如請求項 6 所述之記憶體裝置，其中該些特徵或複數個權重之每一個具有一尾數（mantissa）及一指數（exponent），各該指數被歸類至二群組之其中之一，並且各該尾數根據各該指數偏移。

【請求項 10】如請求項 6 所述之記憶體裝置，其中該控制器於兩個批次（batch）重複使用至少一節點之資料。

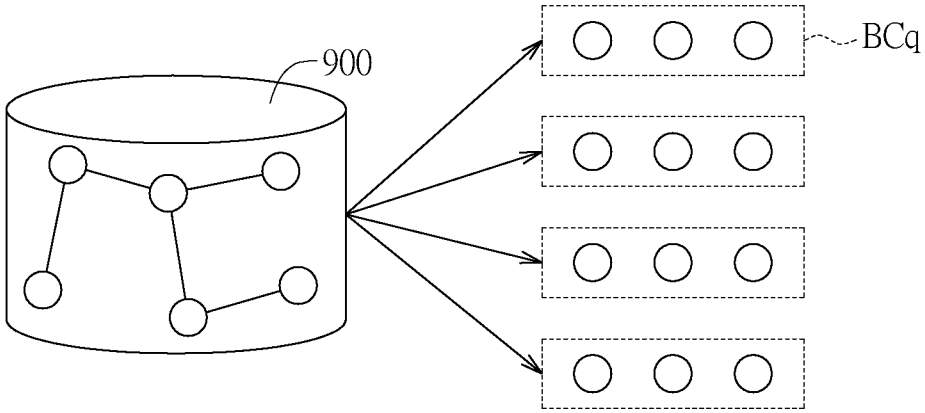
【發明圖式】


GP
Nj
Nj
VTi
Nj
Nj
Nj
VTi
Nj
Nj

第 1 圖


開始
從資料集抽取資料
S110
特徵擷取階段
P1
聚合階段
P2
S120
更新階段
P3
結束

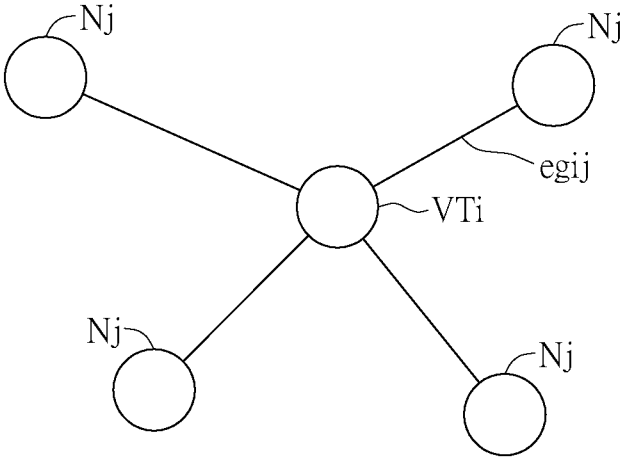
第 2 圖



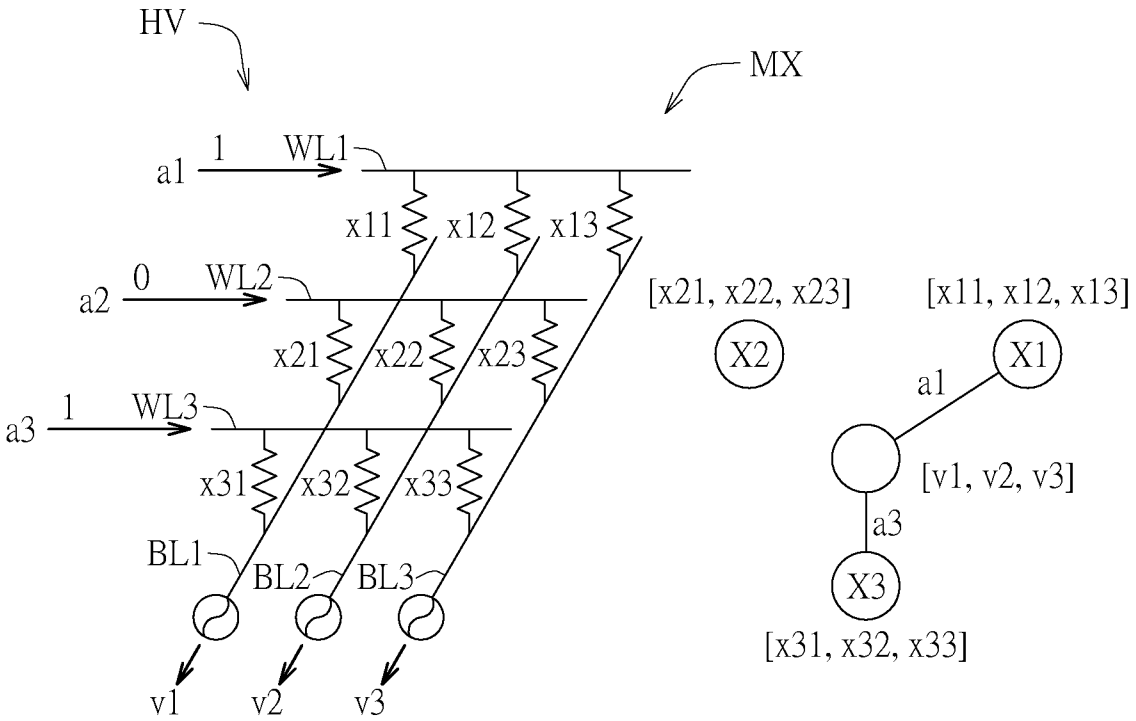
第 3 圖

$$h_v^{k+1} = W_{com}^{k} \left(\Sigma_{u \in N(v)} \alpha_{u,v}^{k} h_u^{k} \right)$$

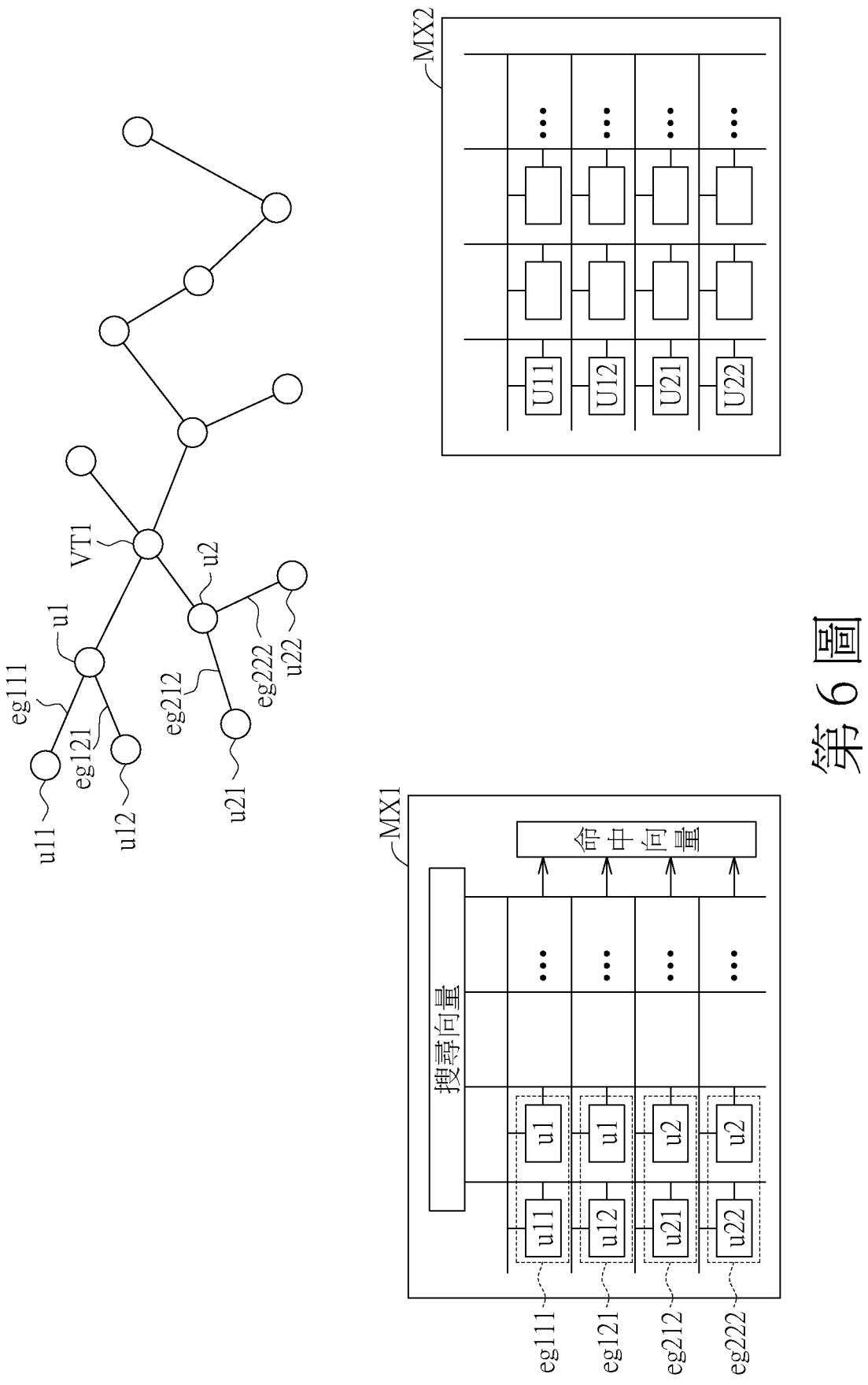
P3 P2 P1



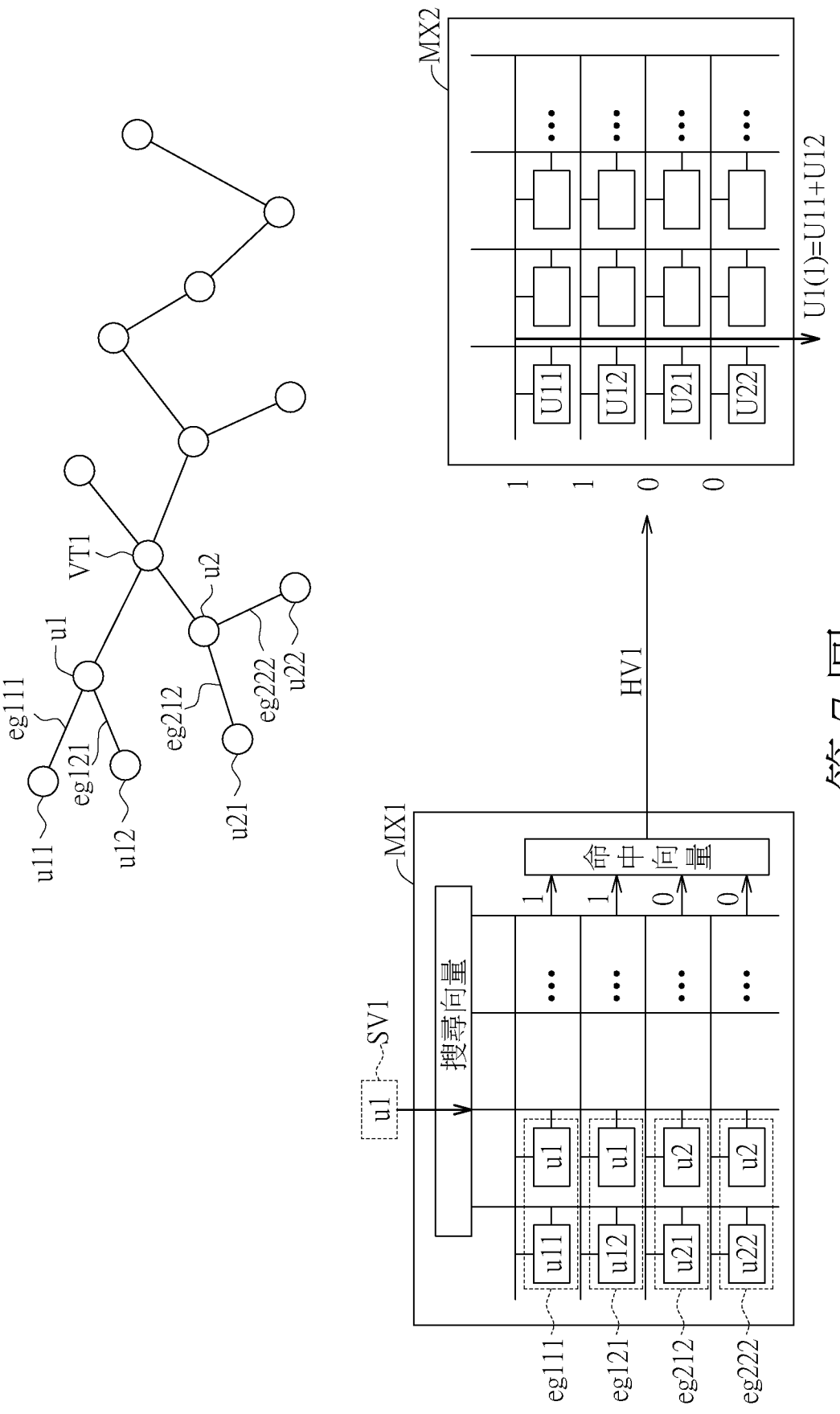
第 4 圖


HV
MX
a1
1
WL1
x11
x12
x13
a2
0
WL2
x21
x22
x23
a3
1
WL3
x31
x32
x33
BL1
BL2
BL3
v1
v2
v3
[x21, x22, x23]
X2
[x11, x12, x13]
X1
a1
[v1, v2, v3]
a3
X3
[x31, x32, x33]

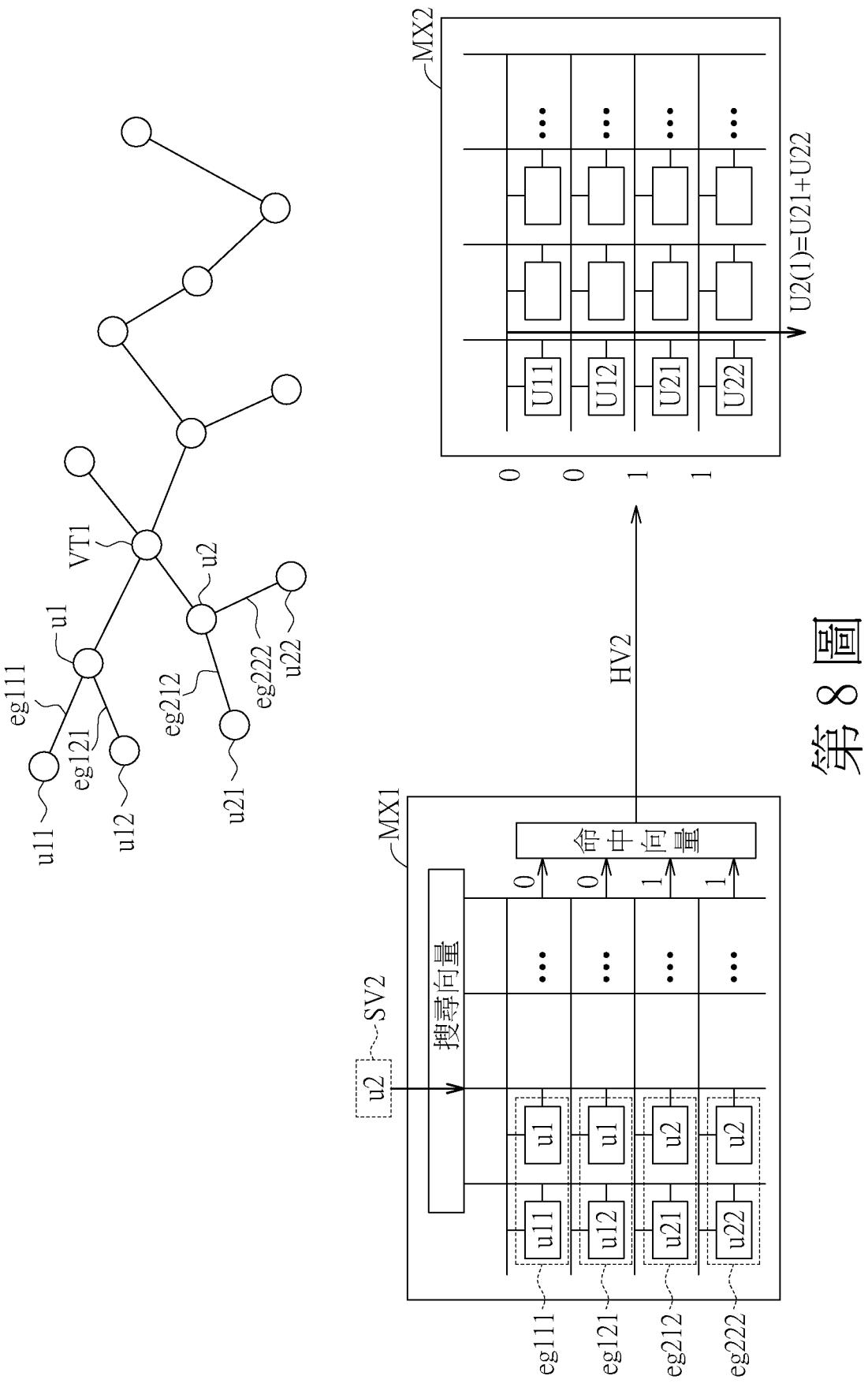
第 5 圖


MX2
U11
U12
U21
U22
VT1
u1
u2
eg111
eg121
eg212
eg222
u11
u12
u21
u22
MX1
搜尋向量
命中向量
u1
u1
u2
u2
u11
u12
u21
u22
eg111
eg121
eg212
eg222

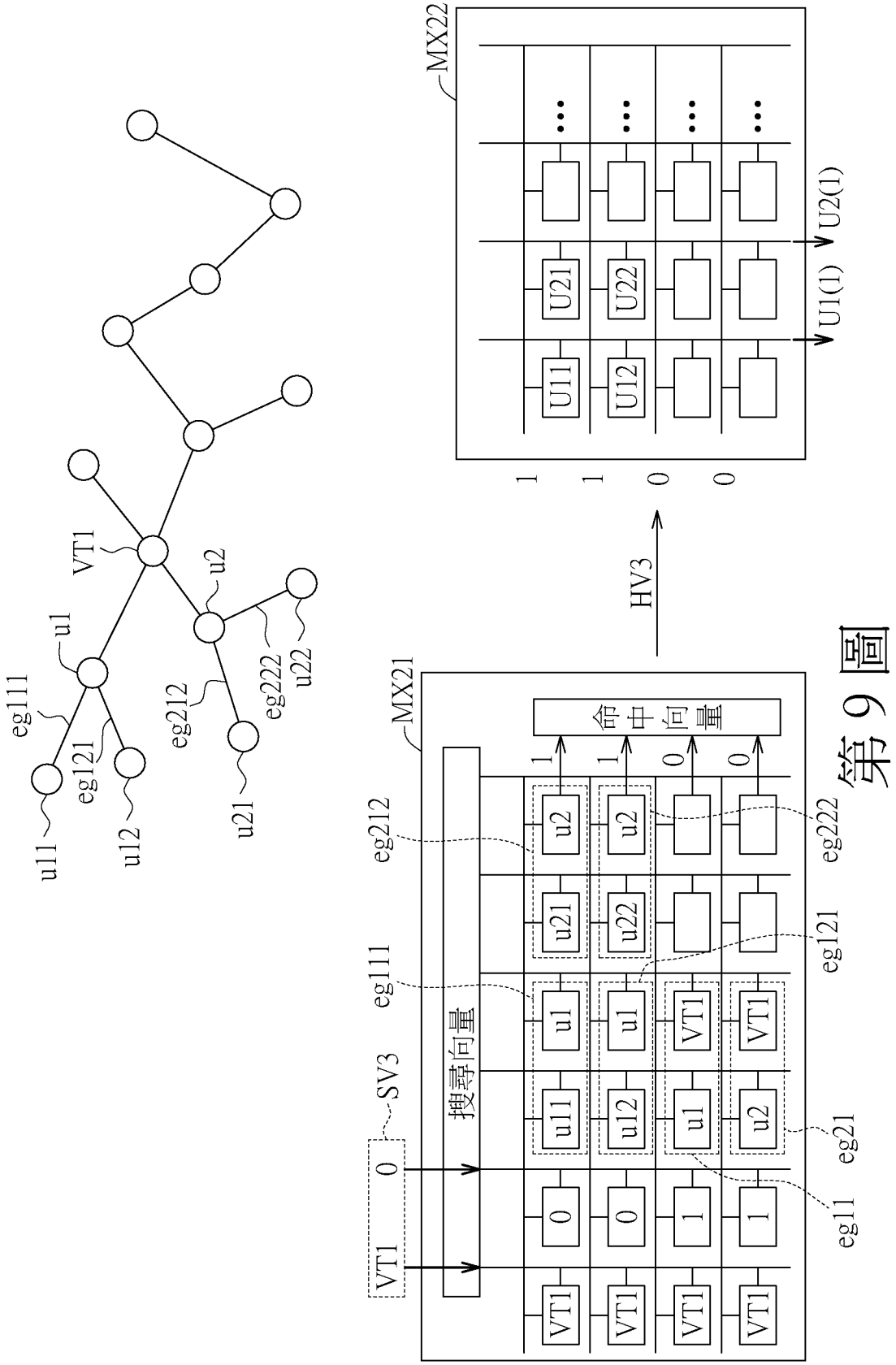
第 6 圖


MX2
U11
U12
U21
U22
U1(1)=U11+U12
1
1
0
0
HV1
MX1
命中向量
搜尋向量
SV1
u1
u11
u12
u21
u22
VT1
u2
eg111
eg121
eg212
eg222

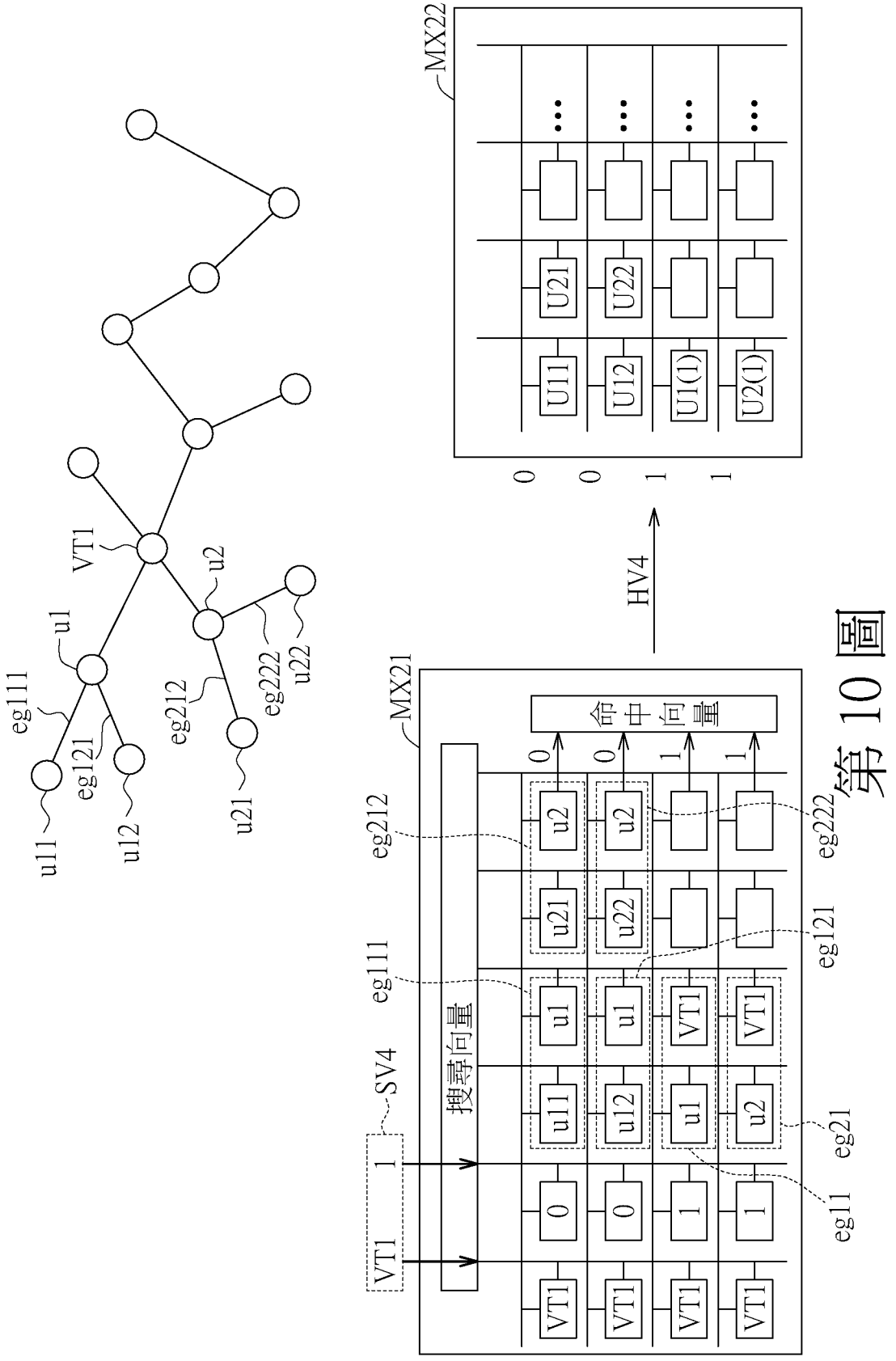
第 7 圖


MX2
U11
U12
U21
U22
U2(1)=U21+U22
0
0
1
1
HV2
MX1
命中向量
搜尋向量
SV2
u2
u1
u11
u12
u21
u22
VT1
eg111
eg121
eg212
eg222

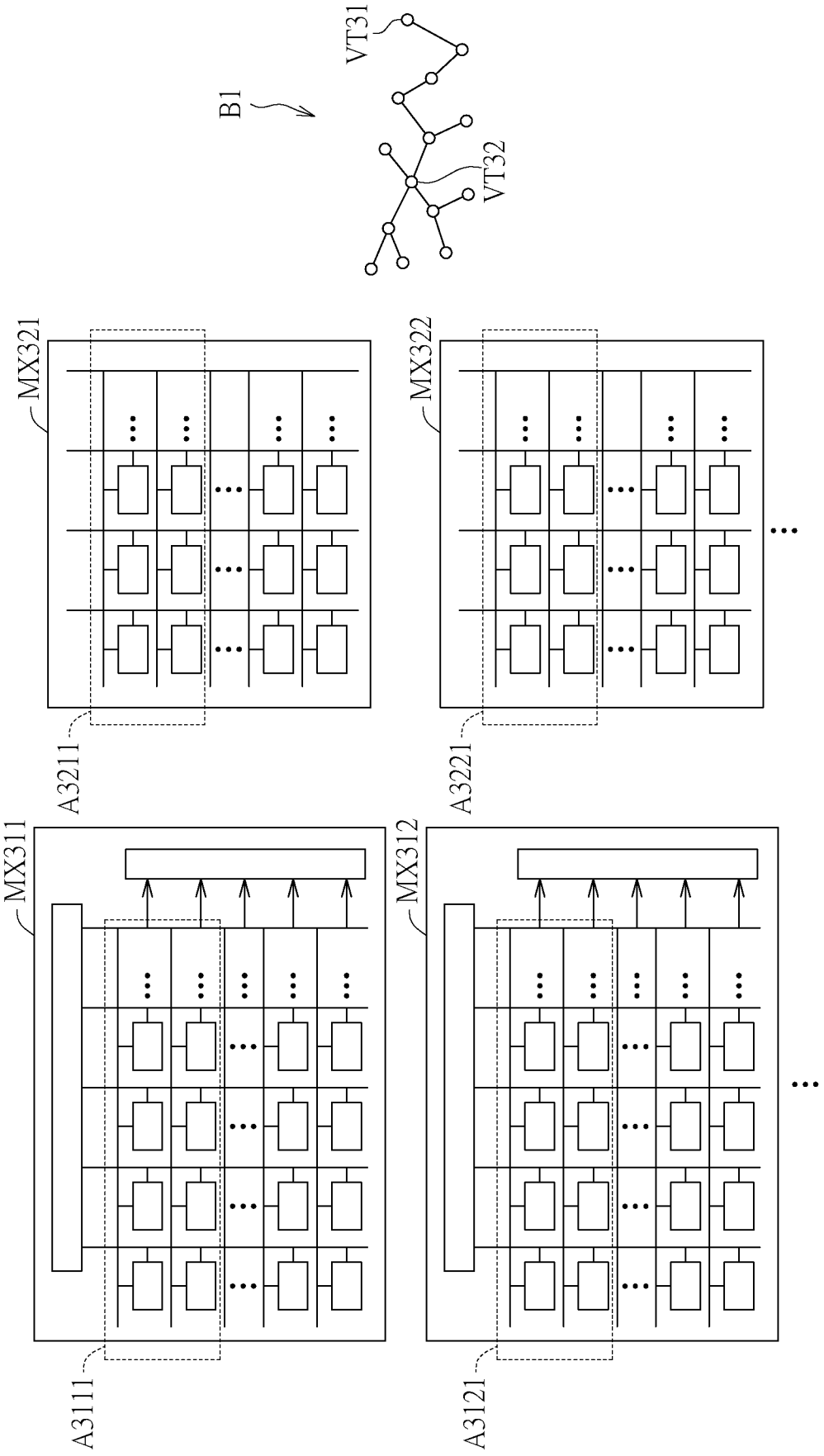
第 8 圖


VT1
u1
u2
eg111
eg121
eg212
eg222
u11
u12
u21
u22
MX22
U11
U12
U21
U22
U1(1)
U2(1)
1
1
0
0
HV3
MX21
命中向量
搜尋向量
SV3
VT1
eg111
eg212
eg121
eg222
eg11
eg21

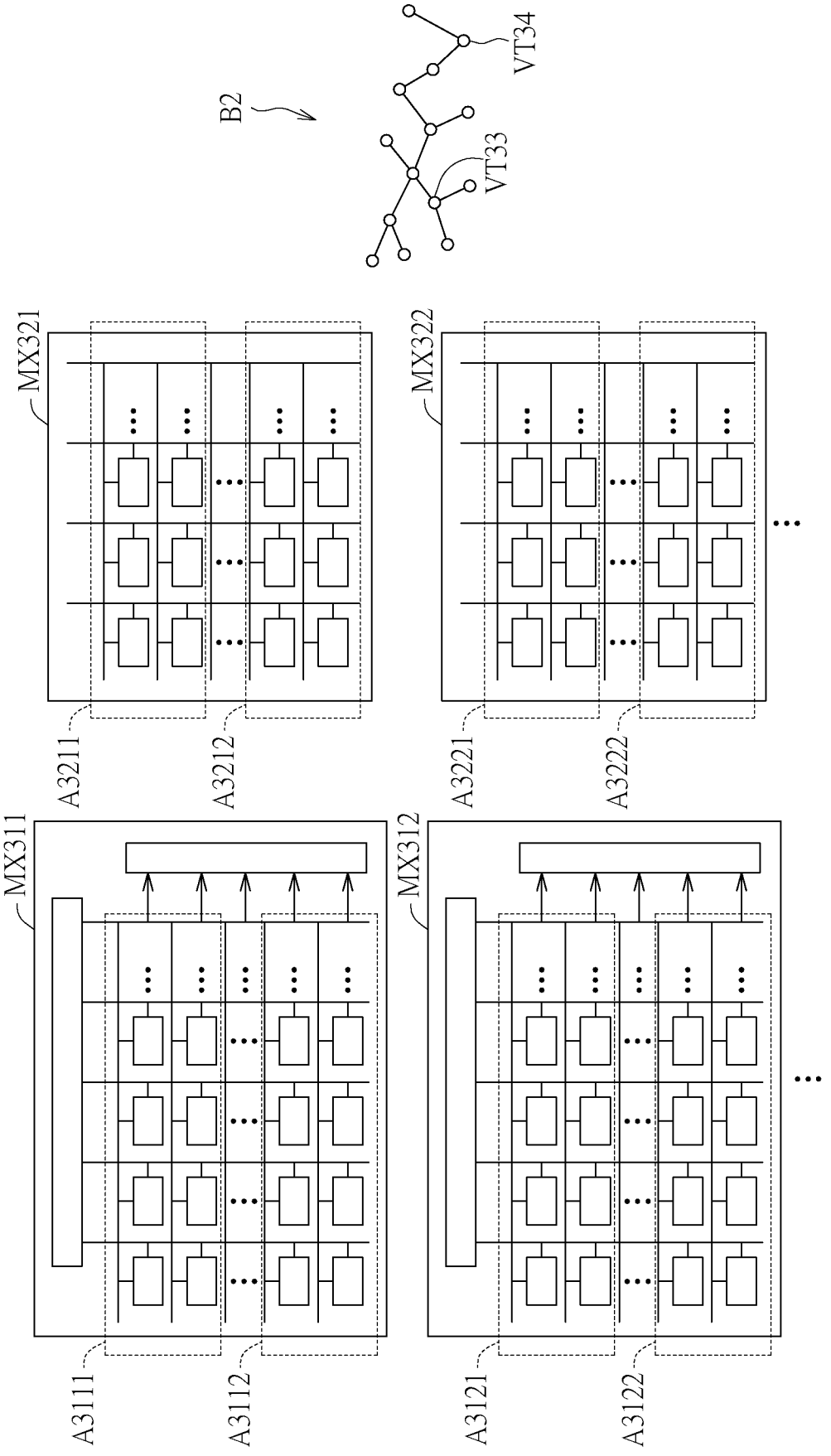
第 9 圖


VT1
u1
u2
eg111
eg121
eg212
eg222
u11
u12
u21
u22
MX22
U11
U12
U1(1)
U2(1)
U21
U22
0
0
1
1
HV4
MX21
SV4
搜尋向量
命中向量
VT1
eg111
eg212
eg121
eg222
eg11
eg21

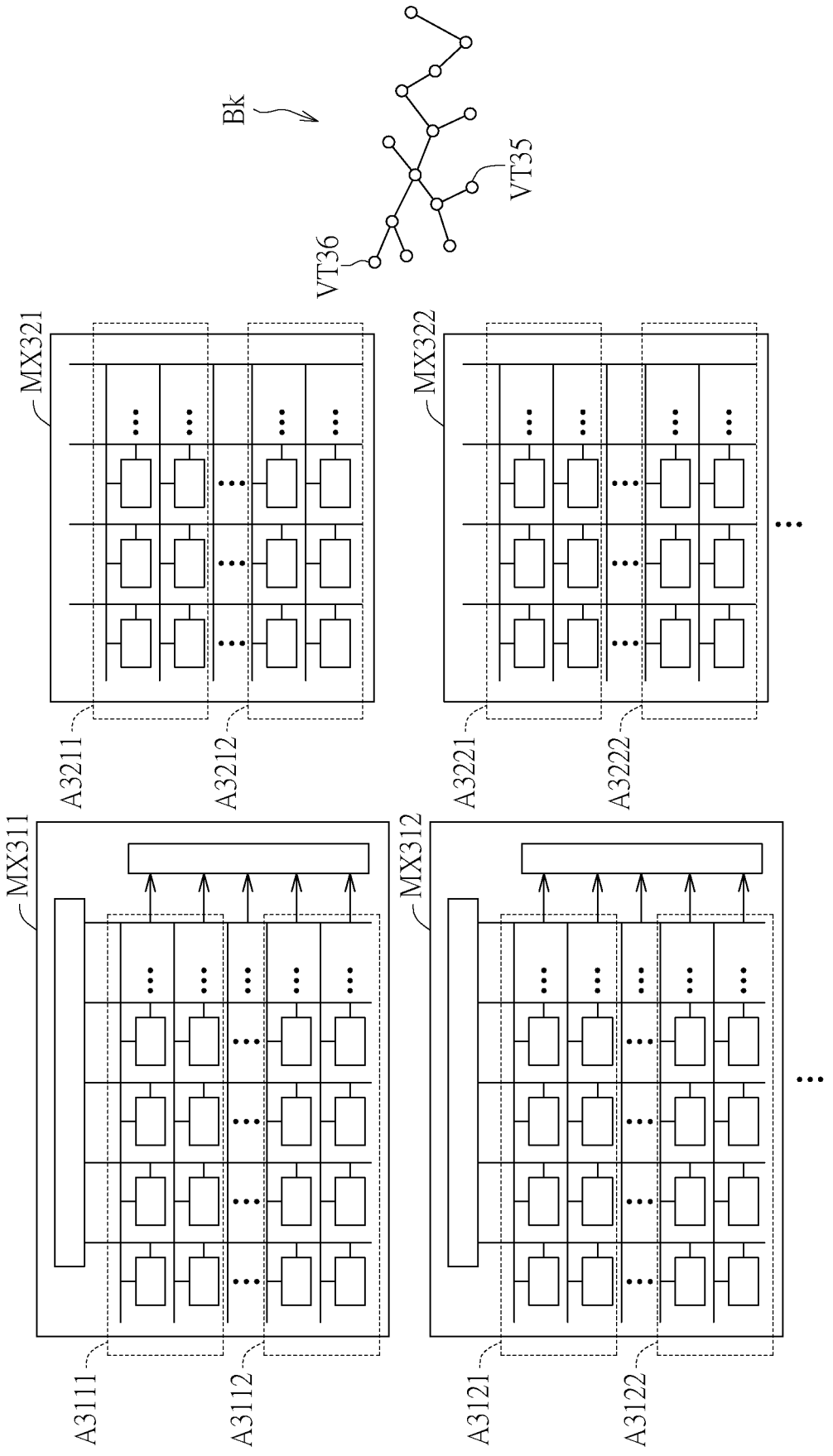
第 10 圖


B1
VT31
VT32
MX321
MX322
MX311
MX312
A3211
A3221
A3111
A3121

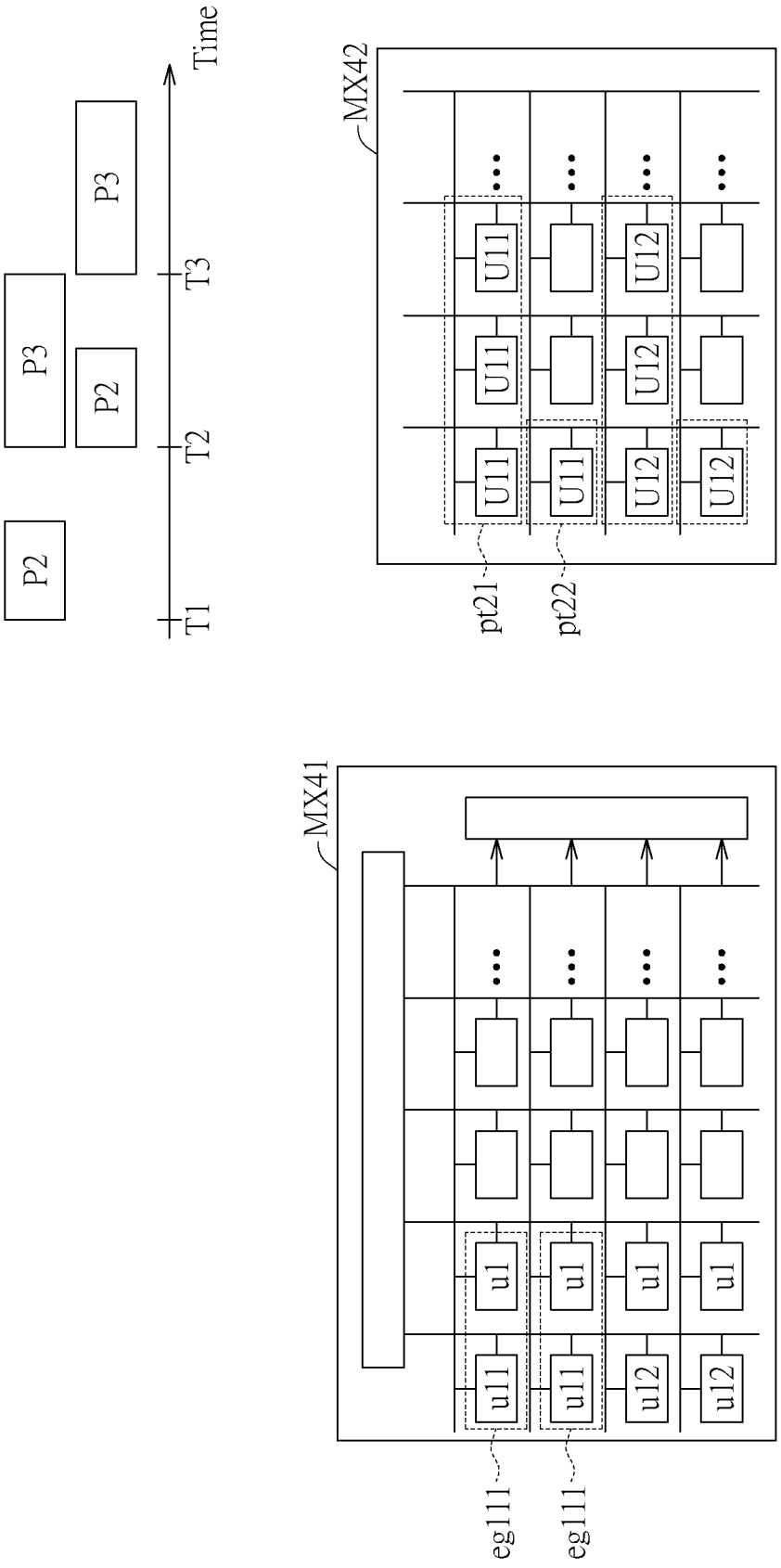
第 11 圖


B2
VT34
VT33
MX321
MX322
MX311
MX312
A3211
A3212
A3221
A3222
A3111
A3112
A3121
A3122

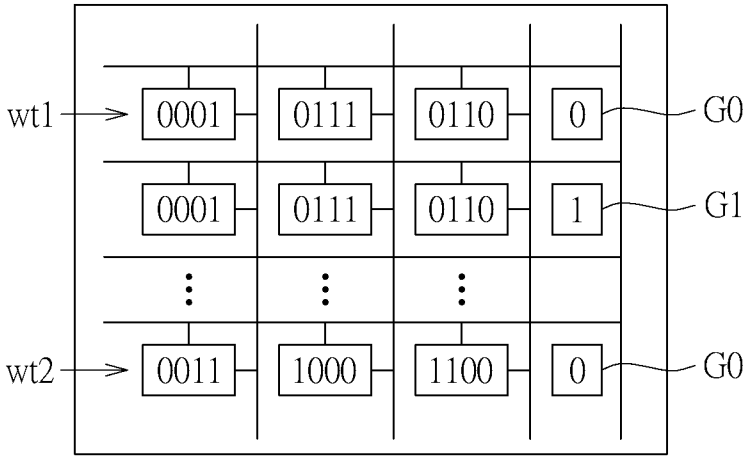
第 12 圖


Bk
VT35
VT36
MX321
MX322
MX311
MX312
A3211
A3212
A3221
A3222
A3111
A3112
A3121
A3122

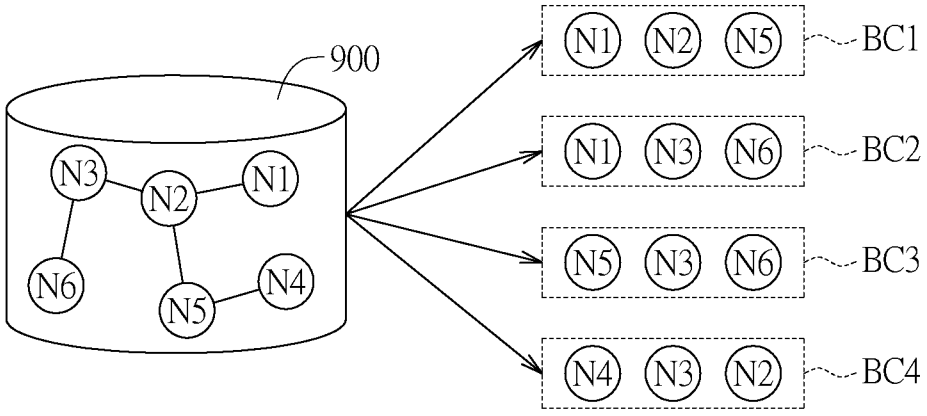
第 13 圖


Time
T1
T2
T3
P2
P3
P2
P3
MX42
U11
U12
pt21
pt22
MX41
u1
u11
u12
eg111
eg111

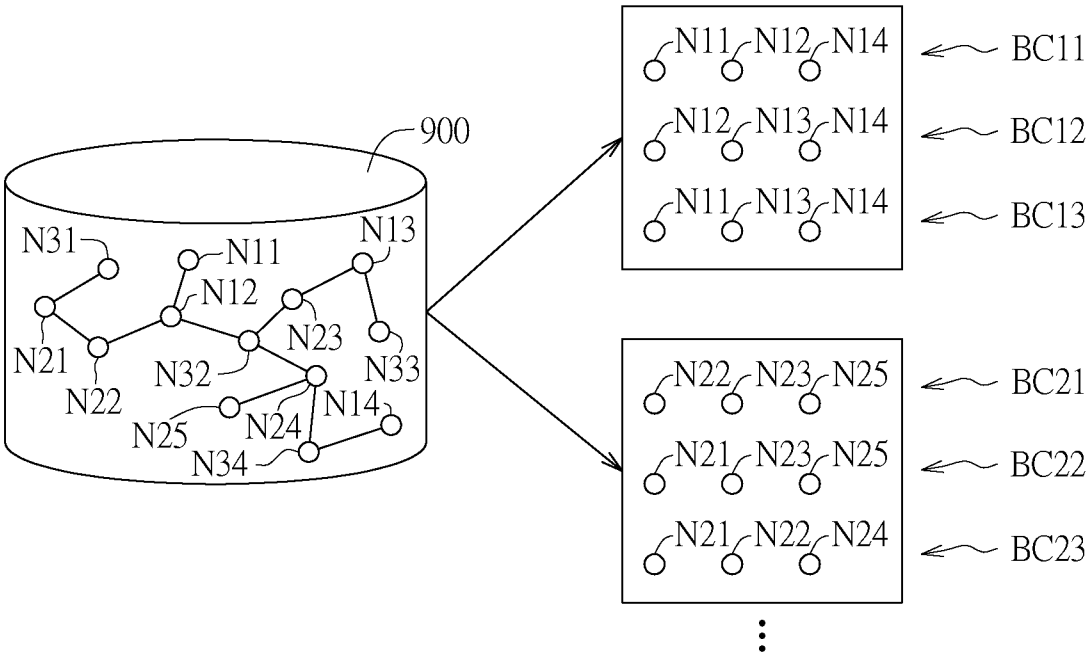
第 14 圖


wt1
0001
0111
0110
0
G0
0001
0111
0110
1
G1
wt2
0011
1000
1100
0
G0

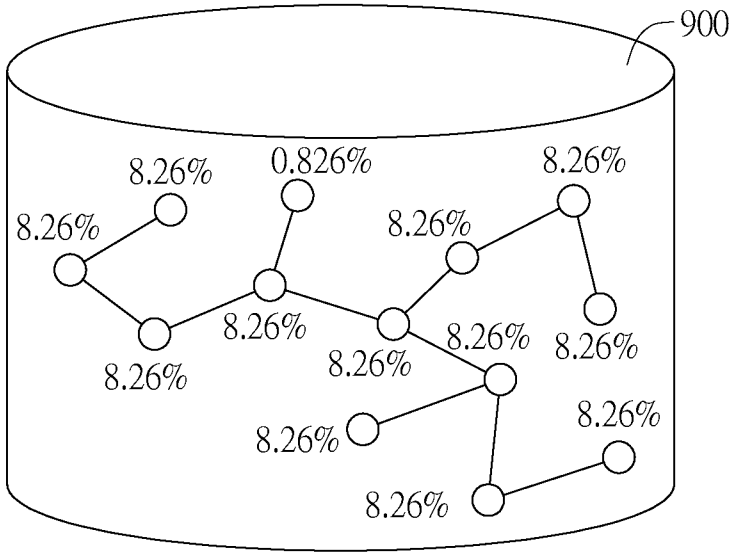
第 15 圖


900
N3
N2
N1
N6
N5
N4
N1
N2
N5
BC1
N1
N3
N6
BC2
N5
N3
N6
BC3
N4
N3
N2
BC4

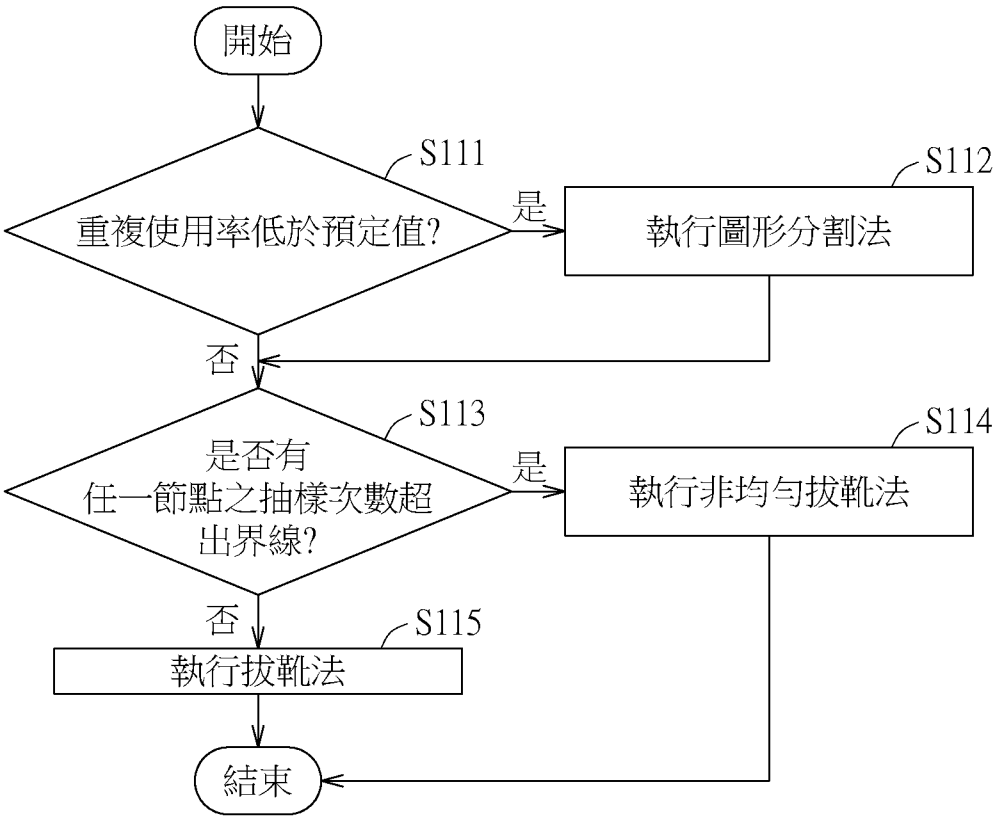
第 16 圖


900
N31
N11
N13
N12
N23
N21
N32
N33
N22
N14
N25
N24
N34
N11 N12 N14
BC11
N12 N13 N14
BC12
N11 N13 N14
BC13
N22 N23 N25
BC21
N21 N23 N25
BC22
N21 N22 N24
BC23

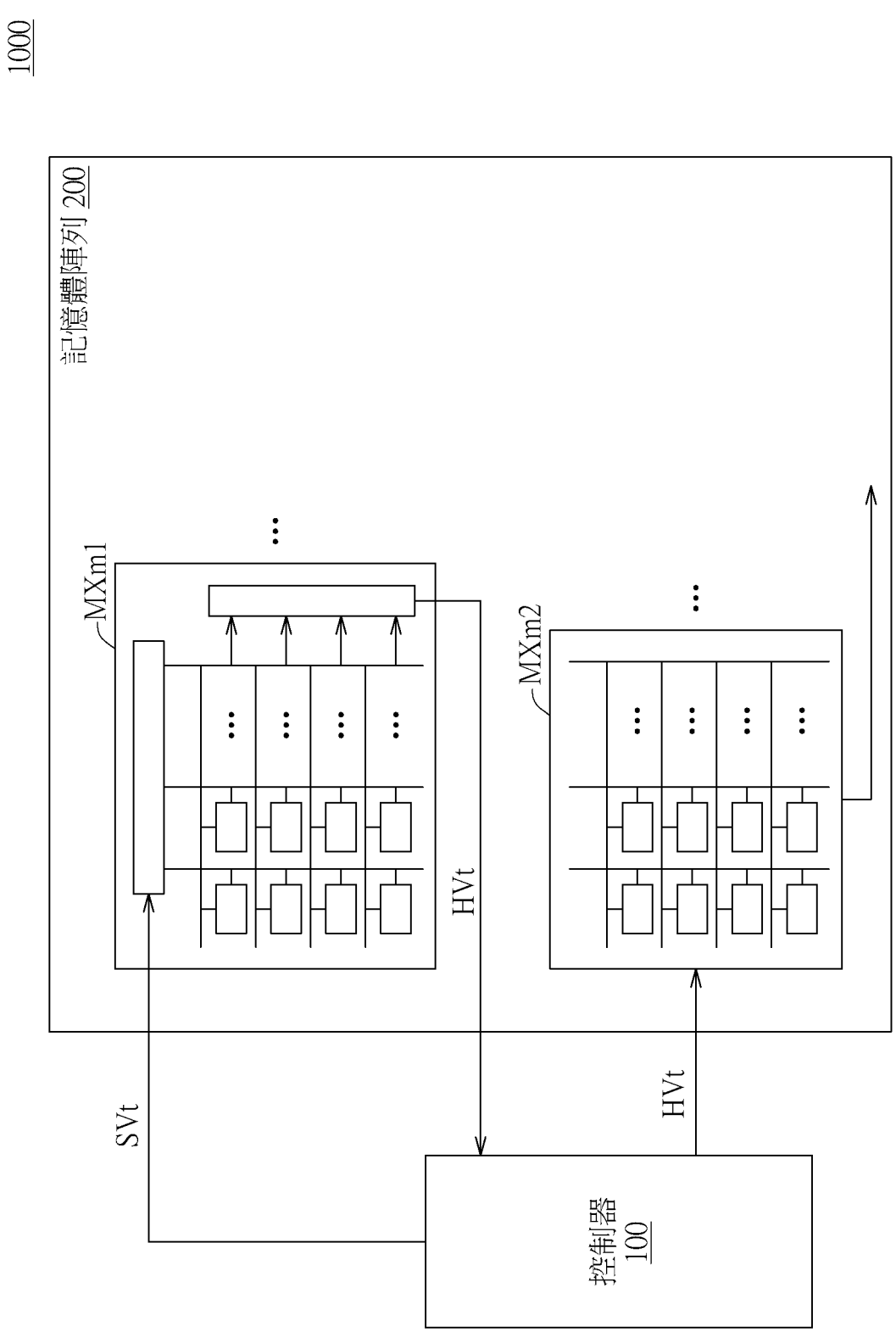
第 17 圖


900
8.26%
0.826%
8.26%
8.26%
8.26%
8.26%
8.26%
8.26%
8.26%
8.26%
8.26%
8.26%
8.26%

第 18 圖


開始
S111
重複使用率低於預定值?
是
S112
執行圖形分割法
否
S113
是否有
任一節點之抽樣次數超
出界線?
是
S114
執行非均勻拔靴法
否
S115
執行拔靴法
結束

第 19 圖



第 20 圖