



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I852888 B

(45)公告日：中華民國 113 (2024) 年 08 月 11 日

(21)申請案號：113105811

(22)申請日：中華民國 113 (2024) 年 02 月 19 日

(51)Int. Cl. : G06F3/06 (2006.01)

(30)優先權：2023/11/14 美國

63/548,546

(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：林柏榕 LIN, BO-RONG (TW)；胡瀚文 HU, HAN-WEN (TW)；李永駿 LI, YUNG-CHUN (TW)；王淮慕 WANG, HUAI-MU (TW)

(74)代理人：祁明輝；林素華

(56)參考文獻：

TW 201826115A

TW 201833823A

TW 202143066A

審查人員：黃鴻鈞

申請專利範圍項數：11 項 圖式數：15 共 49 頁

(54)名稱

用於記憶體內運算之累加器、記憶體裝置及其操作方法

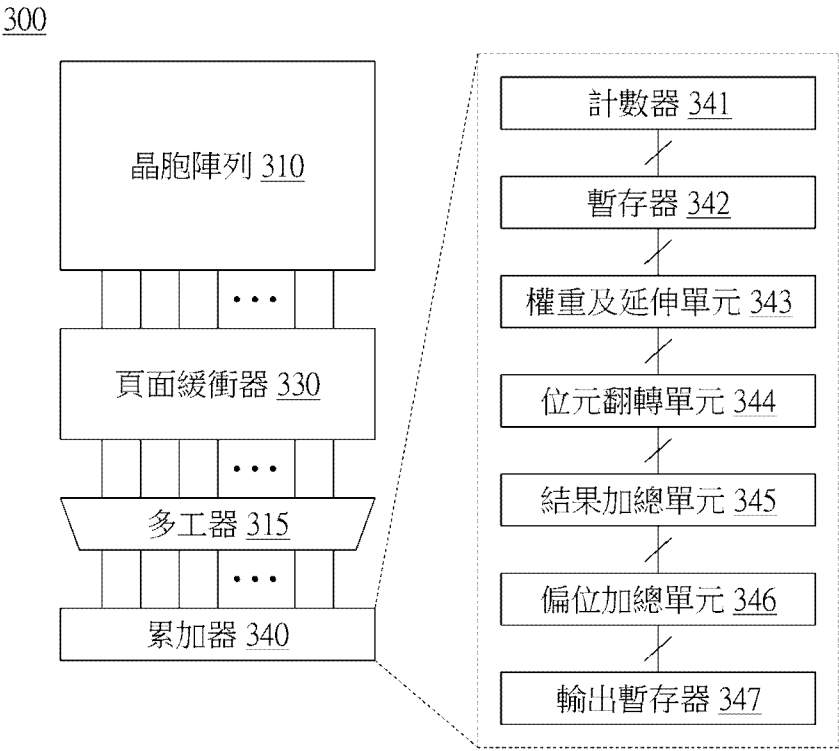
(57)摘要

本案提供的累加器包括：計數器用於將運算產生的部分結果進行計數，以產生計數結果；權重及延伸單元用於將計數結果中各值的位元延伸，產生位元延伸結果，且根據權重及延伸資料表，移動位元延伸結果之對應位元，以產生權重結果；位元翻轉單元，用於根據位元翻轉資料表將權重結果中對應位元進行翻轉，以產生翻轉結果；結果加總單元，用於加總翻轉結果以產生加總結果；以及偏位加總單元，用於將加總結果加上偏位值以產生累加結果。

An accumulator provided by the present disclosure comprises: a counter for counting partial results generated by computing, to generate counting results; a weighting and extension unit for extending bits of each values of the counting results to generate bit extension results, and for moving corresponding bits of the bit extension results to generate weighting results based on a weight & extension table; a bit-flip unit for flipping the corresponding bits of the weighting results to generate flipping results based on a bit-flip table; a result summing unit for summing the flipping results to generate summing results; and a bias add unit for adding a bias value to the summing results to obtain final results.

指定代表圖：

- 符號簡單說明：
- 300:記憶體裝置
 - 310:晶胞陣列
 - 315:多工器
 - 330:頁面緩衝器
 - 340:累加器
 - 341:計數器
 - 342:暫存器
 - 343:權重及延伸單元
 - 344:位元翻轉單元
 - 345:結果加總單元
 - 346:偏位加總單元
 - 347:輸出暫存器



第 3 圖



I852888

【發明摘要】

【中文發明名稱】 用於記憶體內運算之累加器、記憶體裝置及其操作方法

【英文發明名稱】 ACCUMULATOR AND MEMORY DEVICE
FOR IN-MEMORY COMPUTING AND OPERATION
METHOD THEREOF

【中文】

本案提供的累加器包括：計數器用於將運算產生的部分結果進行計數，以產生計數結果；權重及延伸單元用於將計數結果中各值的位元延伸，產生位元延伸結果，且根據權重及延伸資料表，移動位元延伸結果之對應位元，以產生權重結果；位元翻轉單元，用於根據位元翻轉資料表將權重結果中對應位元進行翻轉，以產生翻轉結果；結果加總單元，用於加總翻轉結果以產生加總結果；以及偏位加總單元，用於將加總結果加上偏位值以產生累加結果。

【英文】

An accumulator provided by the present disclosure comprises: a counter for counting partial results generated by computing, to generate counting results; a weighting and extension unit for extending bits of each values of the counting results to generate bit extension results, and for moving corresponding bits of the bit extension results to generate weighting results

based on a weight & extension table; a bit-flip unit for flipping the corresponding bits of the weighting results to generate flipping results based on a bit-flip table; a result summing unit for summing the flipping results to generate summing results; and a bias add unit for adding a bias value to the summing results to obtain final results.

【指定代表圖】第 3 圖。

【代表圖之符號簡單說明】

300:記憶體裝置

310:晶胞陣列

315:多工器

330:頁面緩衝器

340:累加器

341:計數器

342:暫存器

343:權重及延伸單元

344:位元翻轉單元

345:結果加總單元

346:偏位加總單元

347:輸出暫存器

【發明說明書】

【中文發明名稱】 用於記憶體內運算之累加器、記憶體裝置及其操作方法

【英文發明名稱】 ACCUMULATOR AND MEMORY DEVICE
FOR IN-MEMORY COMPUTING AND OPERATION
METHOD THEREOF

【技術領域】

【0001】 本發明係有關於一種用於記憶體內運算(In-Memory-Computing, IMC)的累加器、記憶體裝置及其操作方法。

【先前技術】

【0002】 人工智慧(AI)已在許多領域中成為高度有效解決方案。AI的關鍵操作在於對大量的輸入資料(如輸入特徵圖(input feature maps))與權重值進行乘積累加運算(multiply-and-accumulate, MAC)，而其中也包含了向量對向量乘積(vector-vector multiplication, VVM)操作。然而，以目前的AI架構而言，例如具有外部記憶體的CPU/GPU的馮紐曼架構(Von Neumann architecture)，容易遇到輸出入瓶頸(I/O bottleneck)與低效率的MAC運算流程。然而，隨著所需運算的資料量增加，大量的運算資源消耗在資料的輸入/輸出(I/O)，使I/O瓶頸變得更加嚴重，且效率將更低。

【發明內容】

【0003】本揭露之第一方面提供一種用於記憶體內運算(in-memory computing)的累加器，包括：計數器(counter)，用於將記憶體內運算產生的多個部分結果的多個位元中具有第一位元值的對應對個位元進行計數，以產生計數結果；權重及延伸(weighting and extension)單元，耦接至計數器，用於將計數結果中的各位元使用多個第二位元值延伸至一預設位元數，以產生位元延伸結果，及用於根據權重及延伸資料表，將位元延伸結果中具有第一位元值的各位元，移動至各預設位元數中的對應第W位元，以產生權重結果；位元翻轉(bit-flip)單元，耦接至權重及延伸單元，用於根據位元翻轉資料表將權重結果中的對應各位元值中的各位元進行翻轉，以產生翻轉結果，且位元翻轉資料表具有P個第一位元值；一結果加總(result summing)單元，耦接至位元翻轉單元，用於將翻轉結果中的各值進行加總，以產生加總結果；以及偏位加總(bias add)單元，耦接至結果加總單元，用於將加總結果加上偏位值以產生累加結果，偏位值為P。P及W為大於等於0的正整數。

【0004】本揭露之第二方面提供一種用於記憶體內運算的記憶體裝置，包括：多個記憶體平面，多個記憶體平面之各者包括：晶胞陣列，用於儲存多個權重；頁面緩衝器，耦接至晶胞陣列，頁面緩衝器用於讀取位於該晶胞陣列中的多個權重，並根據多個輸入運算產生多個部分結果，；多工器，耦接至頁面緩衝器，用

於選擇位於頁面緩衝器中多個部分結果中的對應部分結果；以及如本揭露之第一方面提供之累加器，耦接至多工器。

【0005】本揭露之第三方面提供一種用於記憶體內運算的記憶體裝置，包括：多個記憶體平面，多個記憶體平面之各者包括：晶胞陣列，用於儲存多個權重；頁面緩衝器，耦接至晶胞陣列，頁面緩衝器用於讀取位於該晶胞陣列中的多個權重，並根據多個輸入運算產生多個部分結果，；多工器，耦接至頁面緩衝器，用於選擇位於頁面緩衝器中多個部分結果中的對應部分結果；；以及計數器，耦接至多工器，計數器用於將多個部分結果的多個位元中具有第一位元值的對應多個位元進行計數，以產生計數結果；以及累加器，耦接至多個記憶體平面，累加器包括：暫存器，用於暫存於各記憶體平面中的多個計數結果；權重及延伸單元，耦接至暫存器，權重及延伸單元用於將各計數結果中的各位元使用多個第二位元值延伸至預設位元數，產生位元延伸結果，及用於根據權重及延伸資料表，將位元延伸結果中具有第一位元值的各位元，移動至預設位元數中的對應第W個位元，以產生權重結果；位元翻轉單元，耦接至權重及延伸單元，用於根據位元翻轉資料表將權重結果中的對應各位元值中的各位元進行翻轉，以產生翻轉結果，且位元翻轉資料表具有P個第一位元值；結果加總單元，耦接至位元翻轉單元，用於將翻轉結果中的各值進行加總，以產生加總結果；以及偏位加總單元，耦接至結果加總單元，用

於將加總結果加上偏位值以產生累加結果，偏位值為P。P及W為大於等於0的正整數。

【0006】根據本揭露之第一方面之多實施例，計數器至偏位加總單元間的一或多個匯流排的各者的頻寬對應於資料的權重及輸入的N位元、M維度及進位位元，N及M為大於等於1之整數。

【0007】根據本揭露之第三方面之多實施例，計數器至偏位加總單元間的一或多個匯流排的各者的頻寬對應於各記憶體平面的數量O、輸入的N位元、M維度及進位位元(carry bit)，M、N及O為大於等於1之整數。

【0008】根據本揭露之第一方面或第三方面之多實施例，結果加總單元具有一或多個暫存器，一或多個暫存器的數量與對翻轉結果中的各值進行加總以產生加總結果所需的循環的數量成反比。

【0009】根據本揭露之第一方面或第三方面之多實施例，累加器更包括另至少之一移位加總單元，以對應被進行錯誤位元容忍資料(error-bit-tolerance)編碼的資料。

【0010】本揭露之第四方面提供一種用於累加器之操作方法，包括：將記憶體內運算產生的多個部分結果的多個位元中具有第一位元值的對應多個位元進行計數，以產生計數結果；將計數結果中的各位元使用多個第二位元值延伸預設位元數，以產生位元延伸結果；根據權重及延伸資料表，將位元延伸結果中具有第一位元值的各位元，移動至預設位元數中的對應第W個位元，

以產生權重結果；根據位元翻轉資料表將權重結果中的對應各位元值中的各位元進行翻轉，以產生翻轉結果，且位元翻轉資料表具有P個第一位元值；將翻轉結果中的各值進行加總，以產生加總結果；以及將加總結果加上偏位值以產生累加結果，偏位值為P。P及W為大於等於0的正整數。

【0011】根據本揭露之第四方面之多實施例，對應被進行錯誤位元容忍資料編碼的資料，及根據該資料被運算產生的該些部分結果進行解碼及加總，以產生加總結果。

【0012】前述方面之多個變化選擇性地包含硬體、方法或程序、或電腦軟體於電腦可存取媒體。

【0013】為了對本發明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】

【0014】

第1圖顯示示例記憶體裝置的示意圖。

第2圖顯示示例記憶體裝置的累加器的架構的示意圖。

第3圖顯示根據本揭露一實施例之記憶體裝置的累加器的架構的示意圖。

第4圖顯示根據具有本揭露一實施例之累加器的記憶體裝置的架構的示意圖。

第5圖顯示根據本揭露一實施例之權重及延伸的操作示意圖。

第6圖顯示根據本揭露一實施例之位元翻轉的操作示意圖。

第7圖顯示根據本揭露一實施例之結果加總的操作示意圖。

第8圖顯示根據本揭露一實施例之記憶體裝置的累加器的示例匯流排頻寬設計的示意圖。

第9圖顯示根據本揭露一實施例之記憶體裝置的累加器的另一示例匯流排頻寬的示意圖。

第10圖顯示根據本揭露一實施例之記憶體裝置的累加器的結果加總單元的示例暫存器的示意圖。

第11圖顯示根據本揭露另一實施例之記憶體裝置的累加器的架構的示意圖。

第12圖顯示根據本揭露又另一實施例之記憶體裝置的累加器的架構的示意圖。

第13圖顯示根據本揭露另一實施例之記憶體裝置的架構的示意圖。

第14圖顯示根據本揭露又另一實施例之記憶體裝置的架構的示意圖。

第15圖顯示根據本揭露一實施例之累加器之操作方法的流程圖。

各圖式中相同的參考編號與名稱表示相同的元件。這也應理解圖式中所示的各種範例性實施例僅是說明性的表示，並不一定按比例繪製。

【實施方式】

【0015】本說明書的技術用語係參照本技術領域之習慣用語，如本說明書對部分用語有加以說明或定義，該部分用語之解釋係以本說明書之說明或定義為準。本揭露之各個實施例分別具有一或多個技術特徵。在可能實施的前提下，本技術領域具有通常知識者可選擇性地實施任一實施例中部分或全部的技術特徵，或者選擇性地將這些實施例中部分或全部的技術特徵加以組合。

【0016】第 1 圖顯示示例記憶體裝置 100 的示意圖。記憶體裝置 100 可用於記憶體內運算(in-memory computing, IMC)。記憶體裝置 100 具有晶胞陣列 110、頁面緩衝器 130 及累加器 140。記憶體裝置 100 可使用頁面緩衝器 130 及累加器 140 完成 IMC 的向量對向量乘積(vector-vector multiplication, VVM)操作。在執行 VVM 操作時，會將晶胞陣列 110 中的頁面 120 的所選頁面 121 中的資料讀取至頁面緩衝器 130，這些資料的資料形式可為向量，且包括輸入 101 及權重 102。在此示例記憶體裝置 100 中，執行 VVM 操作的累加運算時，輸入 101 需要逐步的與權重 102 進行相乘，當資料為具有多組多個維度的向量時，需要多個循環才能將資料處理完畢(例如資料形式為 64 組 4 位元 512 維 (64sets of 4-bit 512-dimension)，容量為 16KB)，也就是需要較多的時間及功耗。

【0017】第 2 圖顯示示例記憶體裝置 200 的累加器 240 的架構的示意圖。記憶體裝置 200 與第 1 圖的記憶體裝置 100 類似，同樣可用於記憶體內運算(in-memory computing, IMC)，並執

行 VVM 操作。記憶體裝置 200 類似於第 1 圖的記憶體裝置 100，具有晶胞陣列 210、頁面緩衝器 230 及累加器 240。與第 1 圖的記憶體裝置 100 不同在於，記憶體裝置 200 的累加器 240 中設置有多個移位加總 241(shift and add)，用於分別地處理輸入 201 及權重 202 的乘積。由於具有多個移位加總 241，累加 240 在執行 VVM 操作時，可同時處理這些輸入資料，因而可以減少運算所需的時間及功耗。然而，為了同時處理大量資料(例如資料形式為 64 組 4 位元 512 維，容量為 16KB)，隨之增加的移位加總 241 會增加大量成本及增加累加器 240 的體積。

【0018】為解決上述記憶體內運算中針對大量資料部分乘積使用大量累加器的成本及體積問題，本揭露提出的記憶體裝置及記憶體裝置的累加器使用脈動輸入(systolic input)及配合計數器(counter-co-work)的操作架構，並使用位元翻轉資料表(bit-flip table)來更廣泛的支援有符號數資料的整數運算。本揭露的多個實施例的記憶體裝置及記憶體裝置的累加器的技術將於下列詳細說明。

【0019】此處所揭露的技術提供用於在半導體裝置中利用配合計數器(counter-co-work)支持執行 VVM、乘積累加運算(MAC)、高維度運算(hyperdimensional computing, HDC)或只加總(SUM-only)等操作。針對所述之目的，半導體裝置於此係以說明成一記憶體裝置作為範例。需注意的是，此些技術可應用於執行 VVM 或 MAC 運算的累加的任何形式之電路、裝置、或系統。

舉例來說，除了記憶體裝置以外，此些技術可亦應用於任何其他儲存裝置或邏輯裝置例如神經網路加速器。

【0020】此些技術可應用於數種形式的非揮發性記憶體裝置，例如是 NOR 快閃記憶體、NAND 快閃記憶體、可抹除可程式唯讀記憶(erasable programmable read-only memory, EPROM)、鐵電式隨機存取記憶體(ferroelectric random access memory, FeRAM)、磁性隨機存取記憶體(magnetoresistive random-access memory, PCRAM)及其他等。此些技術可應用於數種記憶體形式，例如是單層記憶胞(single-level cell, SLC)裝置、多層記憶胞(multi-level cell, MLC)裝置如二層記憶胞裝置(2-level cell devices)、三層記憶胞(triple-level cell, TLC)裝置、四層記憶胞(quad-level cell, QLC)裝置、或五層記憶胞(penta-level cell, PLC)裝置。此些技術可額外地或替代地應用於各種形式的裝置及系統，例如是安全數位(secure digital, SD)卡、嵌入型多媒體卡(embedded multimedia cards, eMMC)、或固態硬碟(solid-state drives, SSDs)、嵌入型系統，或是網路計算機裝置，如網路路由器或網路處理器、快取控制器以及位址轉換暫存器、查找表、資料庫引擎、資料壓縮硬體、人工神經網路、入侵防禦系統、訂製電腦以及其它。

【0021】第3圖顯示根據本揭露一實施例之記憶體裝置300的累加器340的架構的示意圖。記憶體裝置300與第1圖的記憶體裝置100及第2圖的記憶體裝置200類似，同樣可用於記憶體內運算

(in-memory computing, IMC)，並執行VVM或MAC操作。記憶體裝置300，具有晶胞陣列310、頁面緩衝器330、多工器315及累加器340。與第1圖的記憶體裝置100及第2圖的記憶體裝置200不同在於，記憶體裝置300的累加器340具有計數器341、暫存器342、權重及延伸單元343、位元翻轉單元344、結果加總單元345、偏位加總單元346及輸出暫存器347。在執行VVM或MAC操作時，記憶體裝置300的累加器340接收作為部分結果的資料(圖中未示，請參照第4圖所說明之內容)，並利用具有計數器341，以致能於脈動輸入(systolic input)及配合計數器(counter-co-work)的操作。累加器340的權重及延伸單元343、位元翻轉單元344、結果加總單元345及偏位加總單元346透過位元翻轉資料表(bit-flip table) (圖中未示，請參照第6圖所說明之內容)，致能於支援有符號數資料的整數運算。根據本揭露的多個實施例的半導體裝置的累加器中的各元件的具體功能將於下列第4圖至第7圖中詳細說明。

【0022】第4圖顯示具有根據本揭露一實施例之累加器440的記憶體裝置400的架構的示意圖。於多個實施例中，在進行VVM操作前，資料(如向量資料)可先進行預處理(pre-processing)。在預處理之前，輸入的資料形式可為L組N位元M維之向量。作為示例，進行預處理之後的資料可將L組N位元M維的權重進行編碼，得到編碼後權重402，其中L、M及N為大於等於1的正整數。於此示例中，將以64組4位元512維(L=64, N=4, M=512)的向量

(總容量為16KB)為例以方便說明。權重402可被資料預處理，將原本64組4位元512維的資料形式重新依據權重編碼為各L組中有N種不同對應權重值(如圖中權重 W_3 、 W_2 、 W_1 及 W_0)，且各對應權重值具有M位元，也就是64組有4種權重的而各權重具有512位元的資料形式，如圖中所示。而為對應於各權重具有512位元的資料形式，輸入401可經由N循環(對應於N種權重)的M位元形式(對應權重的M位元)，與權重進行運算，每個循環段應產生多個部分結果，如圖中所示。根據於本揭露之資料預處理，能將資料先進行編碼處理及計算(如相乘)，以致能於使用前述之脈動輸入(systolic input)功能，將儲存於晶胞陣列中(如第3圖中的晶胞陣列310)中的權重透過頁面緩衝器讀取，並與輸入進行運算產生多個部分結果，再由多工器(如第3圖中的頁面緩衝器330及多工器315)選擇對應的部分結果，可進一步被使用於本揭露前述之配合計數器(counter-co-work)操作架構(可配合累加器440之計數器441、權重及延伸單元443、位元翻轉單元444、結果加總單元445、偏位加總單元446及輸出暫存器447)。其他元件的具體功能將於下列圖式中詳細說明。

【0023】 第5圖顯示根據本揭露一實施例之權重及延伸500的操作示意圖。如前述，用於多個運算(例如VVM或MAC操作)的權重可儲存於晶胞陣列中(如第3圖中的晶胞陣列310)，並透過頁面緩衝器(如第3圖中的頁面緩衝器330)將讀取出的權重根據輸入進行運算，並產生多個部分結果，多工器(如第3圖中的多工器315)

選擇對應的部分結果分批輸出至累加器(例如第4圖中的累加器440)進行處理。為方便說明，於此示例中，用於計算的輸入及權重以4位元資料格式的輸入501： $A_3A_2A_1A_0$ 及權重502： $B_3B_2B_1B_0$ 為例，且以輸入501的值為-3(1101)及權重502的值為-1(1111)為例，如圖中所示。

【0024】輸入501及權重502進行乘積時，權重502可儲存於晶胞陣列中(如第3圖中的晶胞陣列310)，並透過頁面緩衝器讀取權重502並根據輸入501進行運算。多工器(如第3圖中的頁面緩衝器330及多工器315)可選擇對應的部分結果輸出至累加器中的計數器(如第4圖中的累加器440及計數器441)，也就是對應於第5圖中的部分結果510(例如為 A_3 與 $B_3B_2B_1B_0$ 的乘積部分結果)。多個部分結果510進行由計數器(例如第4圖的計數器441)進行計數，以得到計數結果520，如圖中所示。其中，計數為將具有第一位元值的位元進行計數。接下來為方便說明，將以第一位元值為1，而第二位元值為0為例，但本揭露不以此作為限制。

【0025】接著權重及延伸(weighting and extension)單元(例如第4圖的權重及延伸單元443)，將計數結果520中的各位元使用多個第二位元值(0)延伸至預設位元數，產生位元延伸結果530。預設位元數關聯於輸入資料及權重的位元、資料維度及進位位元(carry bit)。於一些示例中，預設位元數可根據於下列等式(1)計算：

$$2^{\text{預設位元數}} = (2^{(\text{輸入位元}(Ni) - 1)} * 2^{(\text{權重位元}(Nw) - 1)}) * ((2^m > \text{資料維度})$$

$(M)) * 2^l$ (進位位元(*carry bit*))

等式(1)

其中輸入位元(Ni)、權重位元(Nw)及資料維度(M)皆為大於等於1之整數。需注意的是其中($2^m > \text{資料維度}(M)$)代表 2^m 需要大於資料維度，而m同樣為大於等於1的整數。舉例來說，當資料維度(M)=1， $2^{m=1}=2$ 大於資料維度(M)(1)，因此m為1；而當資料維度(M)=2時， $2^{m=1}=2$ ，並無大於資料維度(M)(為2)，因此需以 $2^{m=2}$ 表示，使 2^2 大於資料維度(M)(為2)，其他m值的計算方法以此類推。於此示例中，由於輸入501及權重502的位元皆為4($A_3A_2A_1A_0$ 及 $B_3B_2B_1B_0$ 對應於輸入位元Ni及權重位元Nw)，而資料維度為1(M=1)，因此計數結果520中的各位元被延伸至8(2^8)位元數。例如於圖中的延伸結果530中所框示的部分， A_3B_0 從原本的1(第一位元值)使用了7個0(第二位元值)延伸至8位元數。而由於本揭露的數值在進算後可進位，因此包含了進位位元531，如圖中所示。

【0026】接著，參照於第5圖中所示的權重及延伸資料表550，其中包括了 $A_3A_2A_1A_0$ 及 $B_3B_2B_1B_0$ 所對應的權重位移資料，例如於表中W00~W03分別對應至0~3，代表了需要將延伸結果530中的各位元中具有第一位元值(於此示例中為1)的位元所需移動到的位置(預設位元數中的第W個)。舉例來說，於權重及延伸資料表550中， A_3B_3 的值為 $2^{6(W33)}$ ，也就是W33為6，因此， A_3B_3 中第一位元值(於此示例中為1)原本在第0個位置(右邊數第1位)，將根據權重及延伸資料表550中的資料(A_3B_3 的值為 $2^{6(W33)}$)，移動至

第W33個位置(W33=6，也就是右邊數來第7位)。以此類推，位元延伸結果540中具有第一位元值(於此示例中為1)的對應位元(例如除了 A_1 列之外的各位元)，同樣根據權重及延伸資料表550，移動第一位元值(1)的對應位元至預設位元數中的對應第W(W00~W33)個位元，以產生權重結果540，如第5圖中所示。

【0027】第6圖顯示根據本揭露一實施例之位元翻轉600的操作示意圖。為了簡化於本揭露之累加器中之乘積運算所需使用之元件，於多個實施例中，位元翻轉(bit-flip)單元(例如第4圖的位元翻轉單元444)具有位元翻轉資料表。位元翻轉資料表對應於下列等式：

$$\begin{aligned}
 A \times B &= \left(-2^{n-1}A_{n-1} + \sum_{i=0}^{n-2} 2^i A_i \right) \times \left(-2^{n-1}B_{n-1} + \sum_{i=0}^{n-2} 2^i B_i \right) \\
 &= \sum_{i=0}^{n-2} \sum_{j=0}^{n-2} A_i B_j 2^{i+j} + 2^{2n-2} A_{n-1} B_{n-1} - 2^{n-1} A_{n-1} \sum_{i=0}^{n-2} 2^i B_i \\
 &\quad - 2^{n-1} B_{n-1} \sum_{i=0}^{n-2} 2^i A_i \\
 &= \left(-2^2 A_3 + \sum_{i=0}^2 2^i A_i \right) \times \left(-2^2 B_3 + \sum_{i=0}^2 2^i B_i \right) = \left(\sum_{i=0}^2 \sum_{j=0}^2 A_i B_j 2^{i+j} \right) \\
 &\quad + (2^6 A_3 B_3) + \left(-2^3 A_3 \sum_{i=0}^2 2^i B_i \right) + \left(-2^3 B_3 \sum_{i=0}^2 2^i A_i \right) \\
 &= A_0 B_0 + 2^1 A_0 B_1 + 2^2 A_0 B_2 + 2^1 A_1 B_0 + 2^2 A_1 B_1 + 2^3 A_1 B_2 \\
 &\quad + 2^2 A_2 B_0 + 2^3 A_2 B_1 + 2^4 A_2 B_2 + 2^6 A_3 B_3 - 2^3 A_3 B_0 - 2^4 A_3 B_1 \\
 &\quad - 2^5 A_3 B_2 - 2^3 A_0 B_3 - 2^4 A_1 B_3 - 2^5 A_2 B_3 \\
 &\quad \text{等式(2)}
 \end{aligned}$$

【0028】經由等式(2)，可獲得位元翻轉資料表中需要進行位元翻轉的位元位置(如第6圖的位元翻轉資料表605中所標示)。於此示例中，由位元翻轉資料表605可以得知需要進行位元翻轉的位元為 $A_3(B_2B_1B_0)$ 及 $B_3(A_2A_1A_0)$ ，其中以第一位元值(於此示例中為1)。位元翻轉資料表中，上述第一位元值的數量為P，也就是具有P個第一位元值。於此示例中，位元翻轉資料表605中第一位元值(於此示例中為1)的數量(P)為6，6可由(0110)表示，位元翻轉資料表中第一位元值的數量於後續其他操作中，可用作為偏位值P(可參照於第7圖之說明)。

【0029】如於第6圖中所示，權重結果640(與此實施例中與第5圖的權重結果540相同)的對應位元 $A_3(B_2B_1B_0)$ 及 $B_3(A_2A_1A_0)$ 被根據位元翻轉資料表605的對應欄位進行翻轉，將第一位元值翻轉成第二位元值(例如將1翻轉為0)及將第二位元值翻轉成第一位元值(例如將0翻轉為1)，以得到翻轉結果660。

【0030】第7圖顯示根據本揭露一實施例之結果加總700的操作示意圖。結果加總700的操作可藉由結果加總(results summing)單元(例如第4圖中的結果加總單元445)來實施，將翻轉結果760(與此實施例中與第6圖的翻轉結果660相同)中的各值先進行分組加總770，再進行加總以產生加總結果780(例如為1111101，如圖中所示)。

【0031】由於根據本揭露之多個實施例中有進行如前述之位元翻轉操作，因此本揭露之累加器具有偏位加總(bias add)單元

(例如第4圖中的偏位加總單元446)，用於將加總結果加上偏位值以產生累加結果。

【0032】如於第7圖中所示，加總結果780(例如為1111101)加上前述之偏位值P(位元翻轉資料表605中之第一位元值之數樣)，而P為00000110(代表6)，兩者相加後之累加結果790如圖中所示為00000011，也就是最後累加的結果為3。而如前述，於此示例中，輸入的值為-3(1101)(例如輸入501)及權重的值為-1(1111)(例如權重502)，因此累加結果為3等同於輸入值-3與權重值-1的乘積。

【0033】因此，根據於上述多個實施例，本揭露所提出之記憶體裝置及記憶體裝置的累加器的架構，以脈動輸入(systolic input)及配合計數器(counter-co-work)的操作架構，可達成VVM或MAC的部分結果的累加運算，不需要配置大量的移位加總元件，有效降低IMC運算的時間、功耗及成本。

【0034】除了前述根據本揭露的記憶體裝置及記憶體裝置的累加器的架構外，本揭露進一步可根據所處理的VVM資料形式來調整元件間的匯流排頻寬，以增加運算的效率及彈性。下列將配合於第8圖及第9圖進行詳細說明。

【0035】第8圖顯示根據本揭露一實施例之記憶體裝置800的累加器840的示例匯流排頻寬設計的示意圖。於此示例中，如圖中所示，累加器840中之權重及延伸單元843至輸出暫存器847間之一或多個匯流排頻寬(如圖中虛線框所標示)，可根據輸入及權重

的位元(Ni及Nw)、資料維度及進位位元(carry bits)來進行設計，也就是根據下列等式：

$$2^{\text{匯流排頻寬}} = (2^{(\text{輸入位元}(Ni) - 1)} * 2^{(\text{權重位元}(Nw) - 1)}) * \text{資料維度}(M) * 2^1 \text{ (進位位元)}$$

等式(3)

【0036】 例如，在資料形式為4位元512維(等於2⁹)時，匯流排的頻寬根據上列等式(3)為 (2⁽⁴⁻¹⁾ * 2⁽⁴⁻¹⁾) * 2⁹ * 2¹ = 2¹⁶，被設計為16，如第8圖中所示。

【0037】 第9圖顯示根據本揭露一實施例之記憶體裝置900的累加器940的另一示例匯流排頻寬的示意圖。類似於第8圖的示例，與此示例中，在資料形式為8位元512維(等於2⁹)時，匯流排的頻寬根據上列等式(3)為 (2⁽⁸⁻¹⁾ * 2⁽⁸⁻¹⁾) * 2⁹ * 2¹ = 2²⁴，被設計為24，如圖中所示。因此，本揭露的多個實施例可根據所處理的VVM資料形式來調整元件間的匯流排頻寬的設計，以增加運算的效率及彈性。

【0038】 第10圖顯示根據本揭露一實施例之記憶體裝置1000的累加器1040的結果加總單元1045的示例暫存器1042的示意圖。於多個實施例中，本揭露進一步可根據需求來設計結果加總單元1045的示例暫存器1042的數量。以資料形式為64組4位元512維向量，頁面緩衝器容量為16KB為例，若暫存器的數量設計為64個(策略1)，可暫存64組的部分值(partial value)，一次可處理16KB，因此需要較少的循環；若暫存器的數量設計為16個(策

略2)，可暫存16組的部分值，一次可處理4KB，因此需要較多的循環；若暫存器的數量設計為1個(策略3)，可暫存1組的部分值，一次可處理256B，因此需要更多的循環。然而較多的暫存器意味於需要更多的成本但較少的循環(運算時間)，而較少的暫存器意味有較低的成本但需要較多的循環(運算時間)。根據於反覆的實驗，於上述條件下所得到的運算速度於策略1約為1306us、策略2約為1306us及策略3約為1351us。

【0039】策略1、策略2及策略3所使用的參數及計算方式分別如下：

策略1：

Assume the latency parameters =

Req_VVM = 512

$t_{\text{Read}} = 70\mu\text{s}$, one-time $t_{\text{Input\_import}} = 30.72\mu\text{s}$,

Global $t_{\text{Logic-AND then to CDL}} = 5\mu\text{s}$, $t_{\text{Internal-transfer}} = 5\mu\text{s}$,

$t_{\text{Accumulation-base}} = 30.72\mu\text{s}$

$t_{\text{Accumulation}} = (t_{\text{Accumulation-base}} / (\text{Exe\_VMM\_PB})) * N$ (N = 64)

$\text{Exe\_VMM\_PB} = \text{Page buffer size} / \text{Required memory cells}$

$\text{Request\_Read} = \text{Req\_VVM} / \text{Exe\_VMM\_PB}$

Computing time =

$\text{Input\_bit} * (t_{\text{Input\_import}}) +$

$(\text{Request\_Read} * (((\text{Exe\_VMM\_PB} / N) *$

第 18 頁，共 27 頁(發明說明書)

$$\begin{aligned}
 & (\text{Input\_bit} * (t_{\text{Logic-function then to CDL}} + t_{\text{Accumulation}}))) + \\
 & t_{\text{Internal-transfer}})) \\
 & = (4 * 30.72) + (8 * (((64/64) * (4 * (5 + 30.72)))) + 5)) \sim = \\
 & 1306\mu\text{s}
 \end{aligned}$$

策略2：

Assume the latency parameters =

$$\text{Req\_VVM} = 512$$

$$t_{\text{Read}} = 70\mu\text{s}, \text{ one-time } t_{\text{Input\_import}} = 30.72\mu\text{s},$$

$$\text{Local } t_{\text{Logic-AND then to CDL}} = 1.25\mu\text{s}, t_{\text{Internal-transfer}} = 5\mu\text{s},$$

$$t_{\text{Accumulation-base}} = 30.72\mu\text{s}$$

$$t_{\text{Accumulation}} = (t_{\text{Accumulation-base}} / (\text{Exe\_VMM\_PB})) * N \quad (N = 16)$$

$$\text{Exe\_VMM\_PB} = \text{Page buffer size} / \text{Required memory cells}$$

$$\text{Request\_Read} = \text{Req\_VVM} / \text{Exe\_VMM\_PB}$$

Computing time =

$$\begin{aligned}
 & \text{Input\_bit} * (t_{\text{Input\_import}}) + \\
 & (\text{Request\_Read} * (((\text{Exe\_VMM\_PB} / N) * \\
 & (\text{Input\_bit} * (t_{\text{Logic-function then to CDL}} + t_{\text{Accumulation}}))) + \\
 & t_{\text{Internal-transfer}})) \\
 & = (4 * 30.72) + (8 * (((64/16) * (4 * (1.25 + 7.68)))) + 5)) \\
 & \sim = 1306\mu\text{s}
 \end{aligned}$$

策略3：

Assume the latency parameters =

$$\text{Req\_VVM} = 512$$

$$t_{\text{Read}} = 70\mu\text{s}, \text{ one-time } t_{\text{Input\_import}} = 30.72\mu\text{s},$$

$$\text{Local } t_{\text{Logic-AND then to CDL}} = 0.1\mu\text{s}, t_{\text{Internal-transfer}} = 5\mu\text{s},$$

$$t_{\text{Accumulation-base}} = 30.72\mu\text{s}$$

$$t_{\text{Accumulation}} = (t_{\text{Accumulation-base}} / (\text{Exe\_VMM\_PB})) * N \quad (N = 1)$$

$$\text{Exe\_VMM\_PB} = \text{Page buffer size} / \text{Required memory cells}$$

$$\text{Request\_Read} = \text{Req\_VVM} / \text{Exe\_VMM\_PB}$$

Computing time =

$$\text{Input\_bit} * (t_{\text{Input\_import}}) +$$

$$(\text{Request\_Read} * (((\text{Exe\_VMM\_PB} / N) * ($$

$$(\text{Input\_bit} * (t_{\text{Logic-function then to CDL}} + t_{\text{Accumulation}})))) +$$

$$t_{\text{Internal-transfer}}))$$

$$= (4 * 30.72) + (8 * (((64 / 1) * (4 * (0.1 + 0.48)))) + 5)) \sim =$$

$$1351\mu\text{s}$$

【0040】 綜合上述，可以理解的是本揭露所提供之暫存器的數量與所需的循環(對翻轉結果中的各值進行加總以產生加總結果所需循環)的數量成反比。

【0041】除了根據於本揭露多個實施例中結果加總單元中的暫存器的數量設計可根據需求調整，於本揭露多個實施例中累加器的輸出暫存器(例如第10圖中的輸出暫存器1047)也可根據於需求調整數量。例如，較少輸出暫存器具有較低成本，但需要記憶體裝置的控制器(如狀態機控制器，未於圖中示出)在運算還沒結束時，隨時將VVM輸出結果取出；較多輸出暫存器具有較高成本，但記憶體控制器可在運算完成後再將VVM輸出結果一併取出。

【0042】第 11 圖顯示根據本揭露另一實施例之記憶體裝置 1100 的累加器 1140 的架構的示意圖。前述的資料預處理(圖中未示)可將資料進行錯誤位元容忍(error-bit-tolerance)資料編碼後進行運算，以得到對應的部分結果。錯誤位元容忍資料編碼乃是使用輸入資料複製(duplication)及權重資料平坦化(flattening)技術。如圖中所示，累加器 1140 進一步可包括結果加總單元與移位加總單元(如移位加總單元 1160)，以對應處理被進行錯誤位元容忍資料編碼所運算出的部分結果。比較於本實施例與如第 4 圖所說明之實施例，可以得知本揭露之所述之累加器架構可大致相同，可經調整(如增加移位加總單元)後對應於不同的資料預處理(編碼)方式。圖中累加器 1140 中之其他元件之對應操作方式及功能類似於上述說明，於此不再贅述(如第 1 圖至第 10 圖之說明)。

【0043】第 12 圖顯示根據本揭露又另一實施例之記憶體裝置 1200 的累加器 1240 的架構的示意圖。除了前述配合計數器

(counter-co-work)的操作架構，本揭露之累加器亦可應用於不同的操作架構，例如配合乒乓緩衝器(ping-pong buffer-co-work)的操作架構。如圖中所示，比較於如第 4 圖所說明之實施例，取代於計數器 441，累加器 1240 進一步包括多個頁面緩衝器 1231 及頁面緩衝器 1232，及多個權重及延伸單元 1242 及權重及延伸單元 1243)，以對應 Page ADC(頁面類比數位轉換器)設計的架構。比較於本實施例與如第 4 圖所說明之實施例，可以得知本揭露之所述之累加器架構可大致相同，可調整(如增加多個頁面緩衝器及多個權重及延伸單元取代於計數器)後實施不同的累加運算架構。圖中累加器 1240 中之其他元件之對應操作方式及功能類似於上述說明，於此不再贅述(如第 1 圖至第 10 圖之說明)。

【0044】 第 13 圖顯示根據本揭露另一實施例之記憶體裝置 1300 的架構的示意圖。如圖所示，記憶體裝置 1300，包括記憶體平面 1311、記憶體平面 1312、記憶體平面 1313 及記憶體平面 1314，各記憶體平面(例如記憶體平面 1311)包括晶胞陣列 1310、頁面緩衝器 1330、多工器 1315 及累加器 1340。於此實施例中，由於各記憶體平面皆具有根據本揭露之多個實施例的累加器之架構，因此可在各記憶體平面直接執行 VVM 操作。圖中累加器 1340 中之其他元件之對應操作方式及功能類似於上述說明(如第 1 圖至第 12 圖之說明)，於此不再贅述。

【0045】 第 14 圖顯示根據本揭露又另一實施例之記憶體裝置的架構的示意圖。如圖所示，記憶體裝置 1400，包括記憶體平

面 1411、記憶體平面 1412、記憶體平面 1413 及記憶體平面 1414，各記憶體平面包括晶胞陣列 1410、頁面緩衝器 1430、多工器 1415 及計數器 1441。而記憶體平面 1411、記憶體平面 1412、記憶體平面 1413 及記憶體平面 1414 耦接至累加器 1440，且各記憶體平面可對應於不同的權重。在此配置中，具有不同權重的各記憶體平面，在執行 VVM 時，可先由各記憶體平面的計數器(如計數器 1441)產生計數結果(如第 5 圖之計數結果 520)，而各平面之計數結果可先存放於累加器 1440 中的暫存器 1442，再經由其他元件進行累加運算，以得到這些記憶體平面的整體累加結果(如第 7 圖之累加結果 790)。圖中累加器 1440 中之其他元件之對應操作方式及功能類似於上述說明(如第 1 圖至第 10 圖之說明)，於此不再贅述。

【0046】第 15 圖顯示根據本揭露一實施例之累加器之操作方法 1500 的流程圖。於步驟 S1501 中，可藉由計數器(例如第 4 圖之計數器 441)，將運算產生的部分結果的多個位元中具有第一位元值(例如為 1)的對應多個位元進行計數，以產生計數結果(例如第 5 圖之計數結果 520)。接著，於步驟 S1502 中，藉由權重及延伸單元(例如第 4 圖之權重及延伸單元 443)將計數結果中的各位元使用多個第二位元值(例如為 0)延伸至預設位元數，產生位元延伸結果(例如第 5 圖之位元延伸結果 530)。於步驟 S1503 中，根據權重及延伸資料表(例如第 5 圖之權重及延伸資料表 550)，將位元延伸結果中具有第一位元值(例如為 1)的各位元，移動至預

設位元數中的對應第 W 個位元，以產生權重結果(例如第 5 圖之權重結果 540)。

【0047】於步驟 1504 中，可藉由位元翻轉單元(例如第 4 圖之位元翻轉單元 444)，根據位元翻轉資料表(例如第 6 圖之位元翻轉資料表 605)將權重結果中的對應各位元值中的各位元進行翻轉，以產生翻轉結果(例如第 6 圖之翻轉結果 660)，且位元翻轉資料表具有 P 個第一位元值(例如為 1)。接著，於步驟 1505 中，可藉由結果加總單元(例如第 4 圖之結果加總單元 445)，將翻轉結果中的各值進行加總，以產生加總結果(例如第 7 圖之加總結果 780)。最後於步驟 S1506，可藉由偏位及加總單元(例如第 4 圖之偏位及加總單元 446)，將加總結果加上偏位值 P 以產生累加結果(例如第 7 圖之加總結果 790)。

【0048】本揭露的上述多個實施例可應用於需要運行 VVM 操作的 IMC 架構設計中，並且可運用於 AI 技術，例如但不受限於，完全連接層(fully-connection layer)、卷積層(convolution layer)、多層感知器(multiple layer Perceptron)、支援向量機器(support vector machine)等 AI 技術之中。

【0049】本本揭露的上述多個實施例不只可應用於應用取向的記憶體內運算(application-driven in-memory computing)，也可應用於記憶體內搜尋(in-memory search)應用及高維度運算(hyperdimensional computing, HDC)等。

【0050】雖然本文可說明許多細節，但此些細節不應解釋為

對本發明的所主張或可主張的範疇的限制，而是解釋為針對特定實施例的特徵的說明。在分開之數個實施例中說明於本文中的某些特徵可亦在單一實施例中組合實施。相反地，說明於單一實施例中的各種特徵可亦分開地或以任何適合的子組合的方式在多個實施例中實施。再者，儘管數個特徵於上文可說明為以某些組合來作用且甚至最初按此來主張，但來自所主張的組合的一或多個特徵在一些情況下可從所述的組合排除(*excised*)，且所主張的組合可針對次組合或次組合的變化。類似地，雖然數個操作在圖式中以特定次序來描繪，但不應將理解為此些操作須以所繪示的特定次序或以順序次序執行，或須執行所有所說明的操作以達成合乎期望的結果。

【0051】綜上所述，雖然本發明已以實施例揭露如上，然其並非用以限定本發明。本發明所屬技術領域中具有通常知識者，在不脫離本發明之精神和範圍內，當可作各種之更動與潤飾。因此，本發明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】

【0052】

100,200,300,800,900,1000,1100,1200,1300,1400:記憶體裝置

101,201,401,501:輸入

102,202,402,502:權重

110,210,310,810,910,1010,1110,1210,1310,1410: 晶胞陣
列

120,220: 頁面

121: 所選頁面

130,230,330,830,930,1030,1130,1230,1231,1232,1330,
1430: 頁面緩衝器

140,240,340,400,840,940,1040,1140,1240,1340,1440: 累
加器

241: 移位加總

341,441,841,941,1041,1141,1341,1441: 計數器

342,842,942,1042,1142,1161,1248,1342,1442,1449: 暫
存器

343,443,843,943,1043,1143,1242,1243,1245,1343,1443
: 權重及延伸單元

344,444,844,944,1044,1144,1244,1344,1444: 位元翻轉單
元

345,445,845,945,1045,1145,1245,1345,1445: 結果加總單
元

346,446,846,946,1046,1146,1246,1346,1446: 偏位加總單
元

347,447,847,947,1047,1147,1247,1347,1447: 輸出暫存器

315,415,815,915,1015,1115,1215,1315,1415: 多工器

500: 權重及延伸

550: 權重及延伸資料表

510: 部分結果

520: 計數結果

530: 延伸結果

531: 進位位元

540,640: 權重結果

600: 位元翻轉

605,805,905,1005,1105,1205,1305,1405: 位元翻轉資料表

660,760: 翻轉結果

700: 結果加總

770: 分組加總

780: 加總結果

790: 累加結果

1160: 移位加總單元

1311,1312,1313,1314,1411,1412,1413,1414: 記憶體平面

1500: 操作方法

S1501,S1502,S1503,S1504,S1505,S1506: 步驟

P: 偏位值

【發明申請專利範圍】

【請求項1】一種用於記憶體內運算(in-memory computing)的累加器，包括：

一計數器(counter)，用於將記憶體內運算產生的複數個部分結果的複數個位元中具有一第一位元值的對應該些位元進行計數，以產生一計數結果；

一權重及延伸(weighting and extension)單元，耦接至該計數器，用於將該計數結果中的各位元使用複數個第二位元值延伸至預設位元數，以產生一位元延伸結果，及用於根據一權重及延伸資料表，將該位元延伸結果中具有該第一位元值的各該些位元，移動至預設位元數中的對應第W個位元，以產生一權重結果；

一位元翻轉(bit-flip)單元，耦接至該權重及延伸單元，用於根據一位元翻轉資料表將該權重結果中的對應各位元值中的各位元進行翻轉，以產生一翻轉結果，且該位元翻轉資料表具有P個該第一位元值；

一結果加總單元(result summing)單元，耦接至該位元翻轉單元，用於將該翻轉結果中的各值進行加總，以產生一加總結果；以及

一偏位加總(bias add)單元，耦接至該結果加總單元，用於將該加總結果加上一偏位值以產生一累加結果，該偏位值為P，

其中， P及W為大於等於0的正整數。

【請求項2】如請求項1所述之累加器，其中，該計數器至該偏位加總單元間的一或多個匯流排的各者的一頻寬對應於資料的一權重及一輸入的N位元、M維度及一進位位元，N及M為大於等於1之整數。

【請求項3】如請求項1所述之累加器，其中，該結果加總單元具有一或多個暫存器，該一或多個暫存器的數量與對該翻轉結果中的各該值進行加總以產生該加總結果所需的循環的數量成反比。

【請求項4】如請求項1所述之累加器，其中，更包括另至少之一移位加總單元，以對應被進行錯誤位元容忍資料(error-bit-tolerance)編碼的資料。

【請求項5】一種用於記憶體內運算(in-memory computing)記憶體裝置，包括：

複數個記憶體平面，該些記憶體平面之各者包括：

一晶胞陣列，用於儲存複數個權重；

一頁面緩衝器，耦接至該晶胞陣列，該頁面緩衝器用於讀取位於該晶胞陣列中的該些權重，並根據複數個輸入運算產生複數個部分結果；

一多工器，耦接至該頁面緩衝器，用於選擇輸出位於該頁面緩衝器中的該些部分結果；以及

如請求項1至4中任一項所述之累加器，耦接至該多工器。

【請求項6】一種用於記憶體內運算的記憶體裝置，包括：

複數個記憶體平面，該些記憶體平面之各者包括：

一晶胞陣列，用於儲存複數個權重；

一頁面緩衝器，耦接至該晶胞陣列，該頁面緩衝器用於讀取位於該晶胞陣列中的該些權重，並根據複數個輸入運算產生複數個部分結果；

一多工器，耦接至該頁面緩衝器，用於選擇輸出位於該頁面緩衝器中的該些部分結果；；以及

一計數器，耦接至該多工器，該計數器用於將被選擇的該些部分結果的複數個位元中具有一第一位元值的對應該些位元進行計數，以產生一計數結果；以及

一累加器，耦接至該些記憶體平面，該累加器包括：

一暫存器，用於暫存於各該些記憶體平面中的該些計數結果；

一權重及延伸單元，耦接至該暫存器，該權重及延伸單元用於將各該些計數結果中的各位元使用複數個第二位元值延伸一預設位元數，以產生一位元延伸結果，及用於根據一權重及延伸資料表，將該位元延伸結果中具有該第一位元值的各該些位元，移動至該預設位元數中的對應第W個位元，以產生一權重結果；

一位元翻轉單元，耦接至該權重及延伸單元，用於根據一位元翻轉資料表將該權重結果中的對應各位元值中的各位元進行翻轉，以產生一翻轉結果，且該位元翻轉資料表具有P個該第一位元值；

一結果加總單元，耦接至該位元翻轉單元，用於將該翻轉結果中的各值進行加總，以產生一加總結果；以及

一偏位加總單元，耦接至該結果加總單元，用於將該加總結果加上一偏位值以產生一累加結果，該偏位值為P，

其中，P及W為大於等於0的正整數。

【請求項7】如請求項6所述之記憶體裝置，其中，該累加器中複數個單元間的一或多個匯流排的各者的一頻寬對應於該些記憶體平面的數量O、輸入的N位元、M維度及一進位位元(carry bit)，M、N及O為大於等於1之整數。

【請求項8】如請求項6所述之記憶體裝置，其中，該結果加總單元具有一或多個暫存器，該一或多個暫存器的數量與對該翻轉結果中的各該值進行加總以產生該加總結果所需的循環的數量成反比。

【請求項9】如請求項6所述之記憶體裝置，其中，該累加器更包括另至少之一移位加總單元，以對應被進行錯誤位元容忍資料編碼的資料。

【請求項10】一種用於累加器之操作方法，包括：

將記憶體內運算產生的複數個部分結果的複數個位元中具有一第一位元值的對應該些位元進行計數，以產生一計數結果；

將該計數結果中的各位元使用複數個第二位元值延伸至一預設位元數，以產生一位元延伸結果；

根據一權重及延伸資料表，將該位元延伸結果中具有該第一位元值的各該些位元，移動至該預設位元數中的對應第 W 個位元，以產生一權重結果；

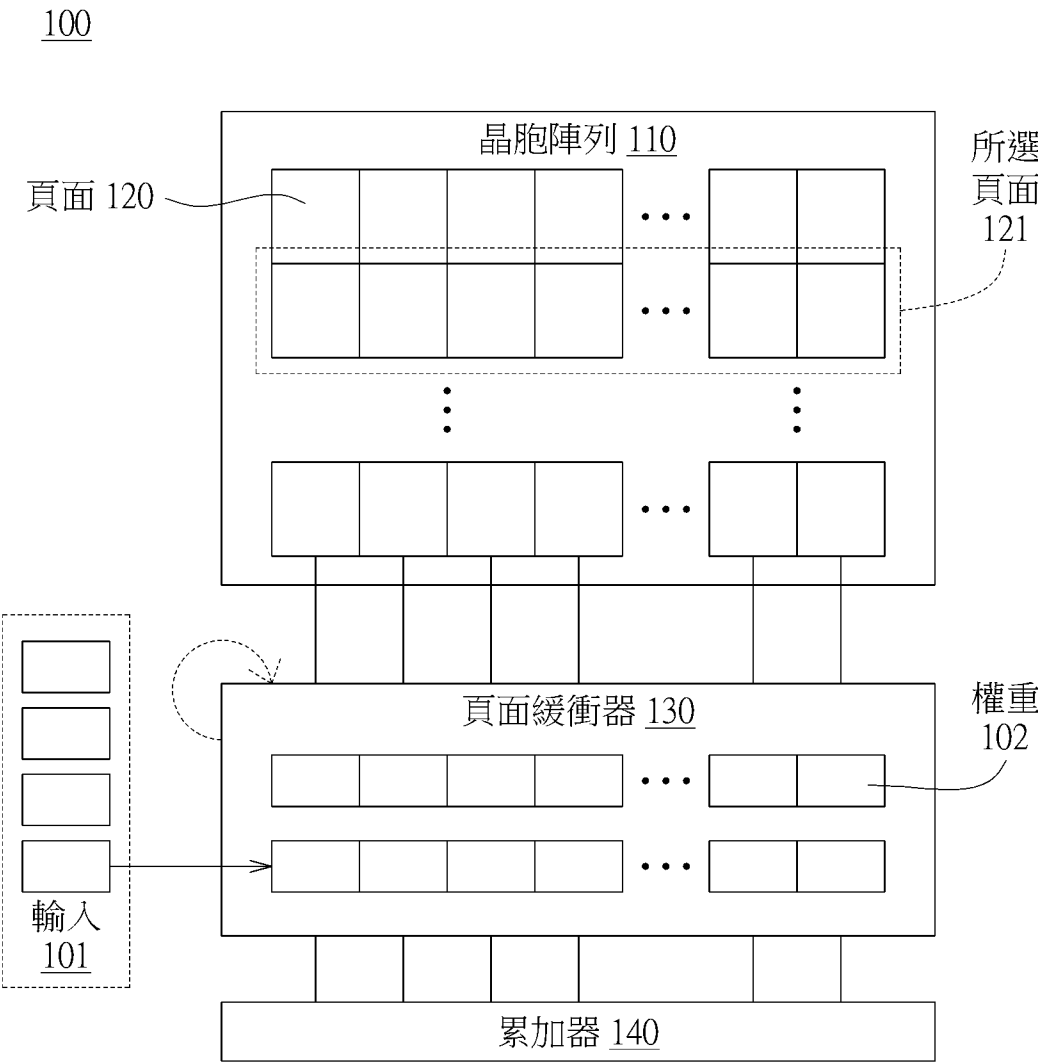
根據一位元翻轉資料表將該權重結果中的對應各位元值中的各位元進行翻轉，以產生一翻轉結果，且該位元翻轉資料表具有 P 個該第一位元值；

將該翻轉結果中的各值進行加總，以產生一加總結果；以及將該加總結果加上一偏位值以產生一累加結果，該偏位值為 P ，

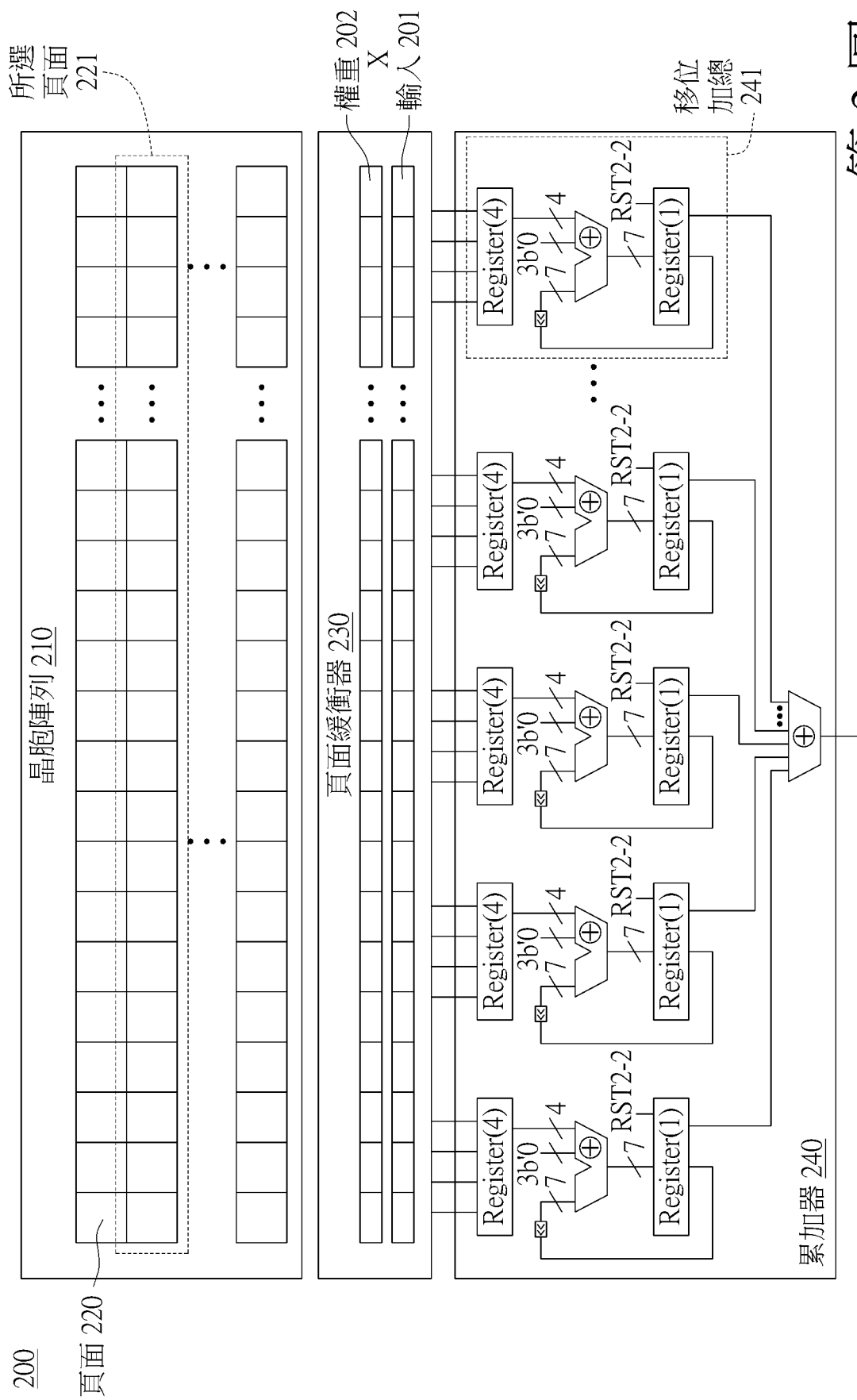
其中， P 及 W 為大於等於0的正整數。

【請求項11】如請求項10所述之操作方法，其中，對應被進行錯誤位元容忍資料編碼的資料，及根據該資料被運算產生的該些部分結果進行解碼及加總，以產生該加總結果。

【發明圖式】

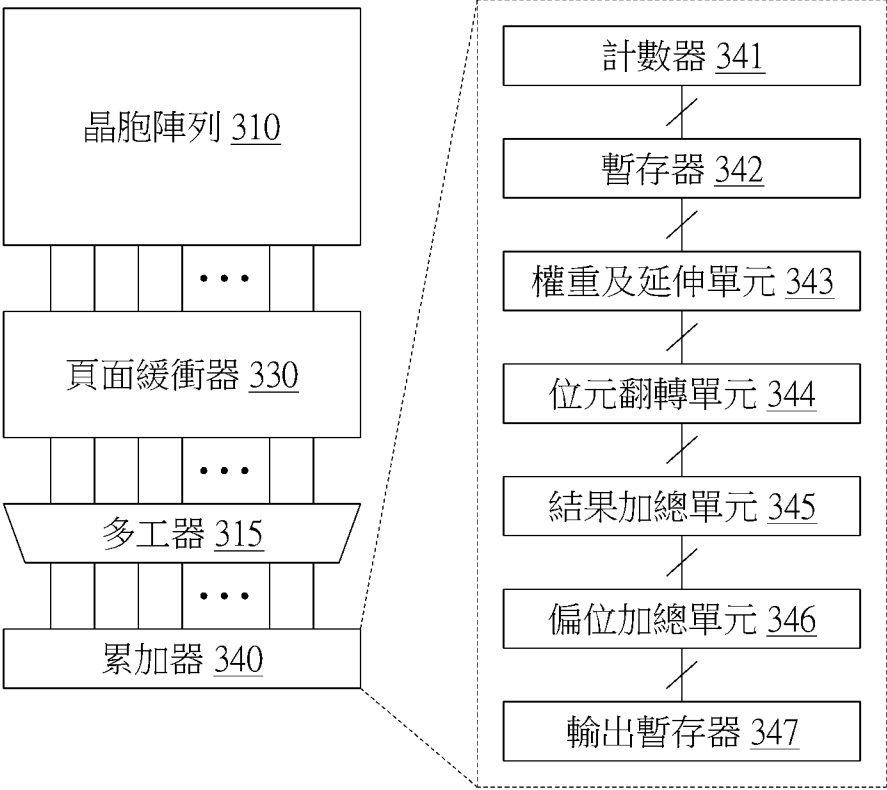


第 1 圖

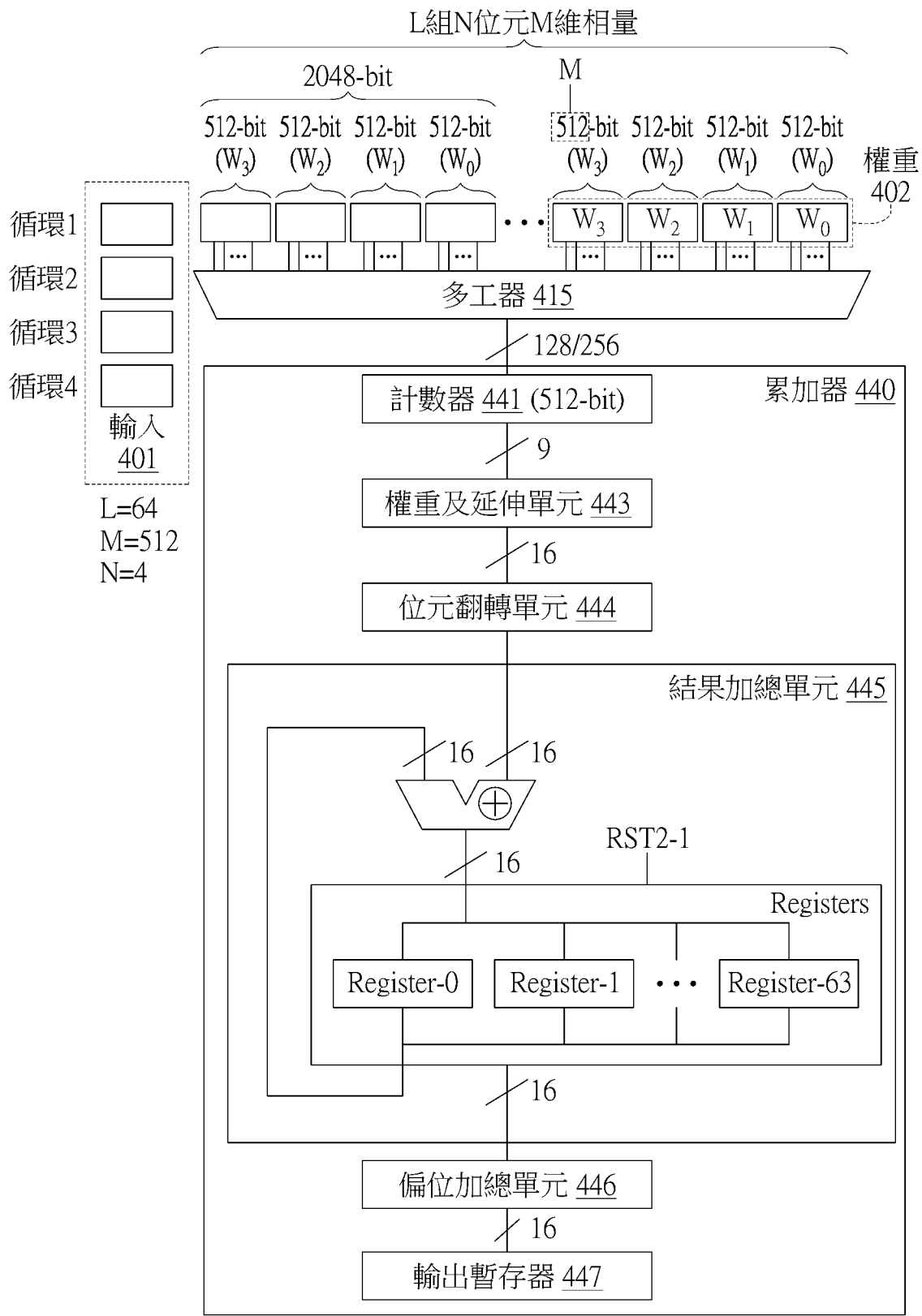


第2圖

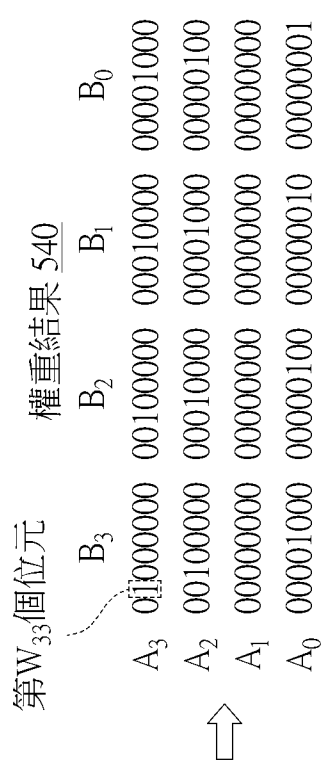
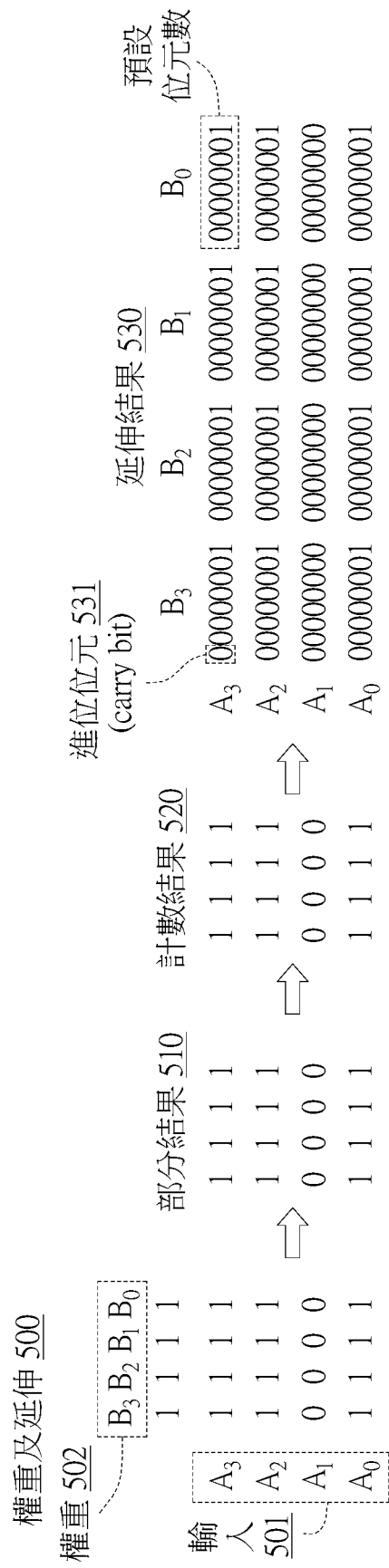
300



第 3 圖



第 4 圖



權重及延伸資料表 550				
	B ₃	B ₂	B ₁	B ₀
A ₃	2 ^{6(W33)}	2 ^{5(W32)}	2 ^{4(W31)}	2 ^{3(W30)}
A ₂	2 ^{5(W23)}	2 ^{4(W22)}	2 ^{3(W21)}	2 ^{2(W20)}
A ₁	2 ^{4(W13)}	2 ^{3(W12)}	2 ^{2(W11)}	2 ^{1(W10)}
A ₀	2 ^{3(W03)}	2 ^{2(W02)}	2 ^{1(W01)}	2 ^{0(W00)}

第 5 圖

位元翻轉 600

權重結果 640				
	B ₃	B ₂	B ₁	B ₀
A ₃	01000000	00100000	00010000	00001000
A ₂	00100000	00010000	00001000	00000100
A ₁	00000000	00000000	00000000	00000000
A ₀	00001000	00000100	00000010	00000001



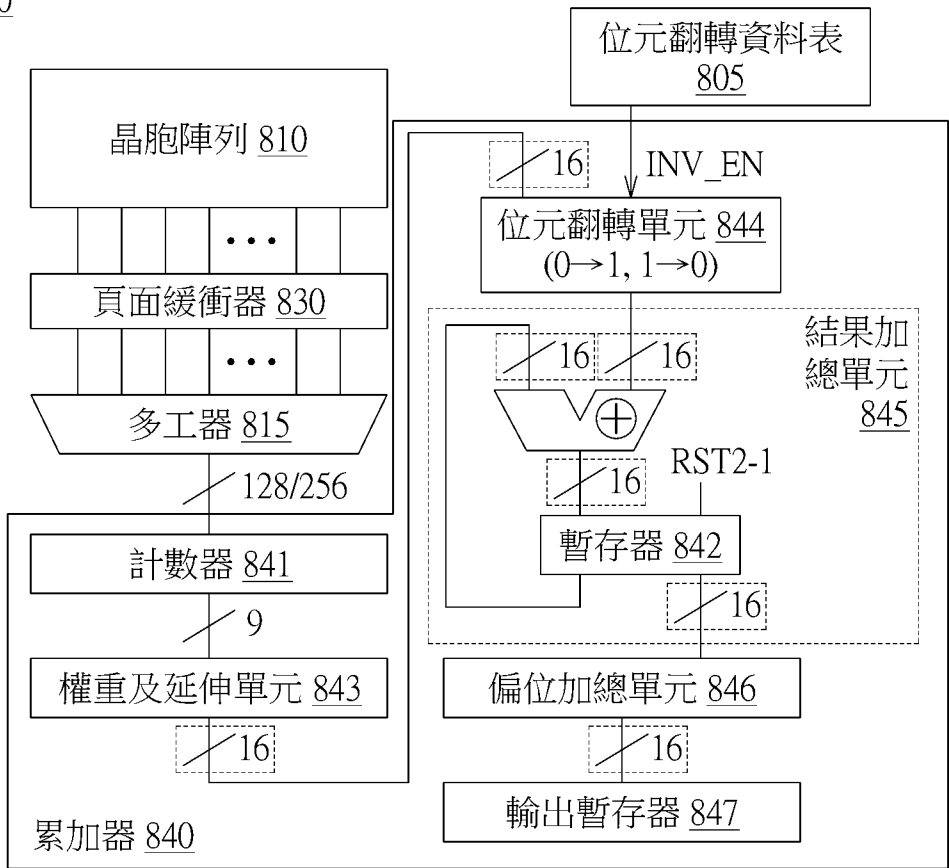
翻轉結果 660				
	B ₃	B ₂	B ₁	B ₀
A ₃	01000000	11011111	11101111	11110111
A ₂	11011111	00010000	00001000	00000100
A ₁	11111111	00000000	00000000	00000000
A ₀	11110111	00000100	00000010	00000001

位元翻轉資料表 605				
	B ₃	B ₂	B ₁	B ₀
A ₃	0	1	1	1
A ₂	1	0	0	0
A ₁	1	0	0	0
A ₀	1	0	0	0

第一位元值(1)之對應位元為要翻轉之位元
根據於此表偏位值P為6(6個第一位元值(1))，6由(0110)表示

第6圖

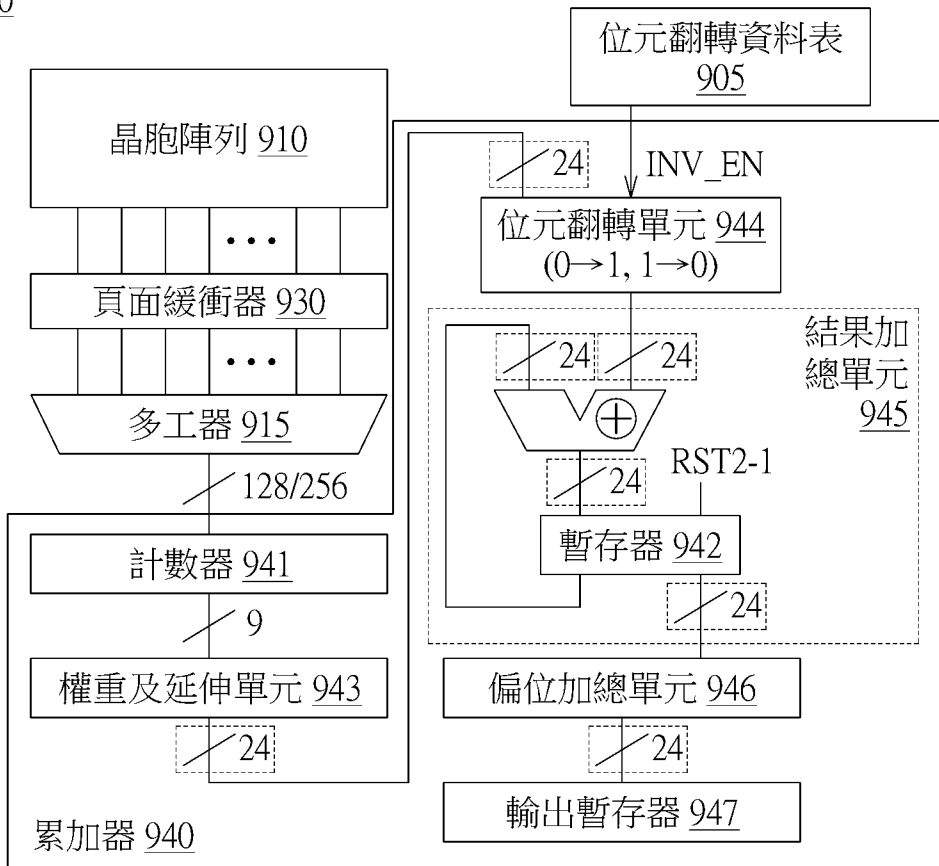
800



用於4位元512維度向量的累加器設計
匯流排頻寬= $(2^{(\text{輸入位元}(N_i) - 1)} * 2^{(\text{權重位元}(N_w) - 1)}) * \text{資料維度}(M) * 2^1$ (進位位元)= $(2^{(4-1)} * 2^{(4-1)}) * 2^9 * 2^1 = 2^{16}$

第 8 圖

900

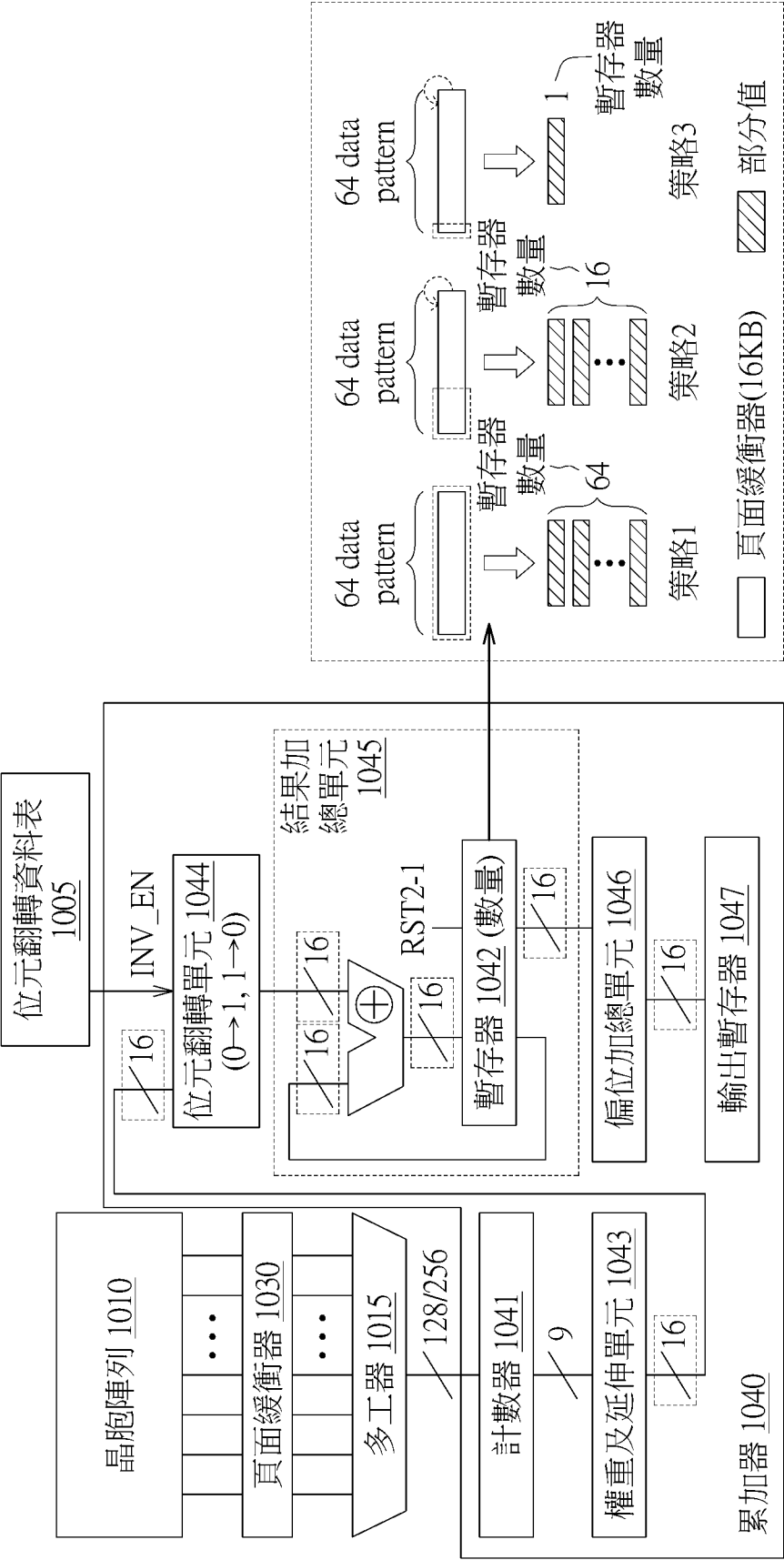


用於8位元512維度向量的累加器設計

$$\text{匯流排頻寬} = (2^{(\text{輸入位元(Ni)} - 1)} * 2^{(\text{權重位元(Nw)} - 1)}) * \text{資料維度(M)} * 2^1 (\text{進位位元}) = (2^{(8-1)} * 2^{(8-1)}) * 2^9 * 2^1 = 2^{24}$$

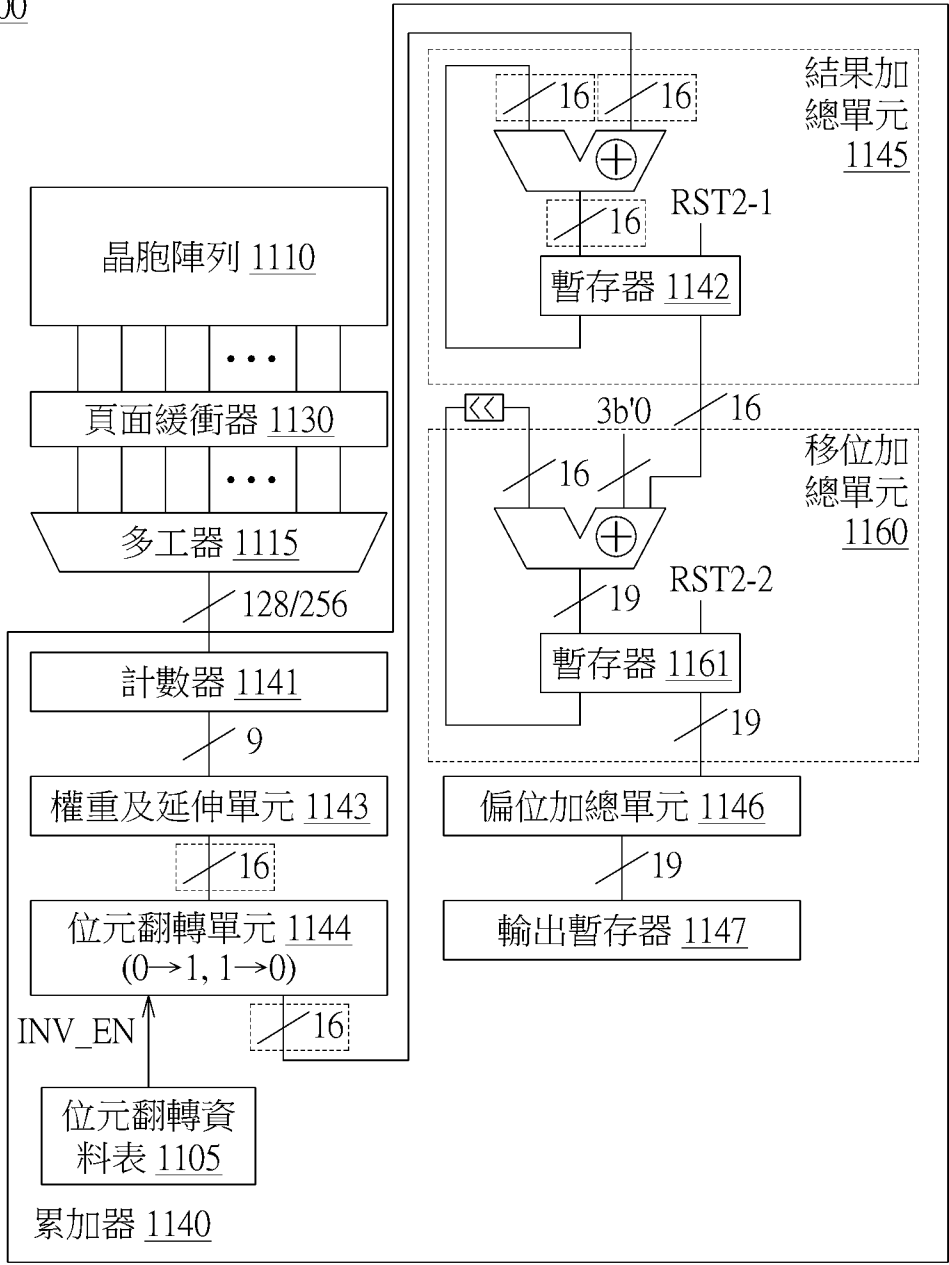
第 9 圖

1000



第 10 圖

1100



第 11 圖

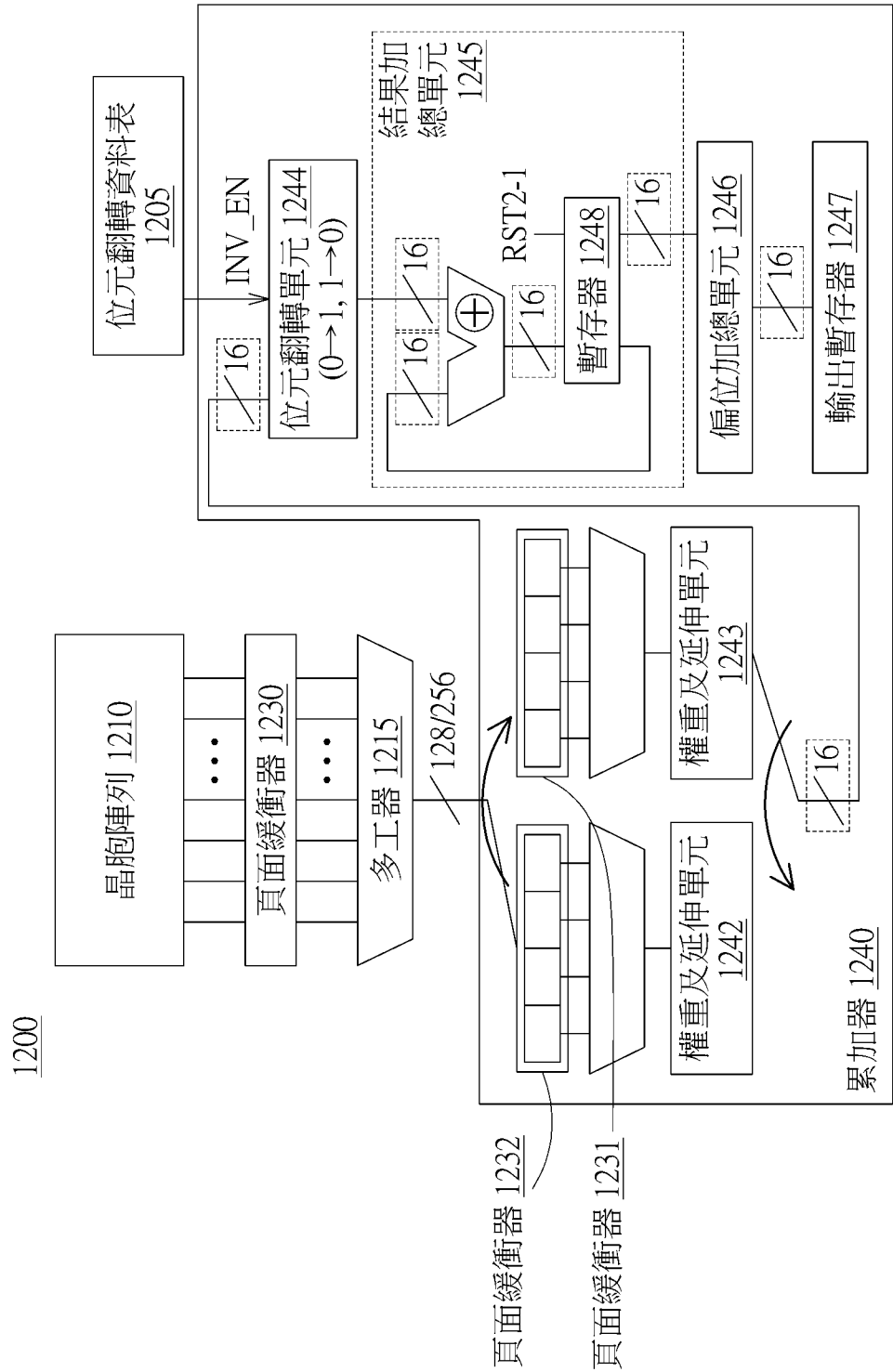
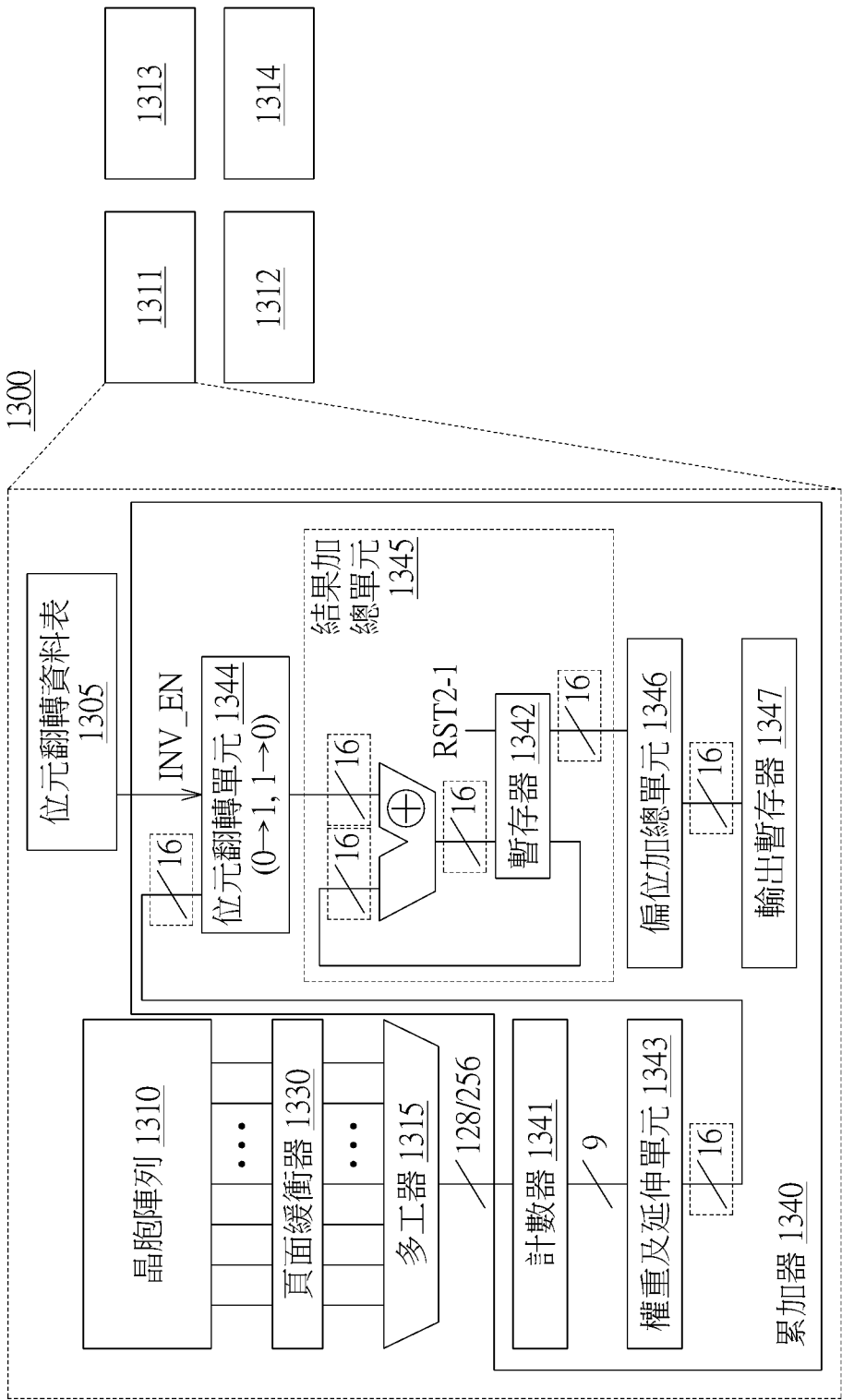


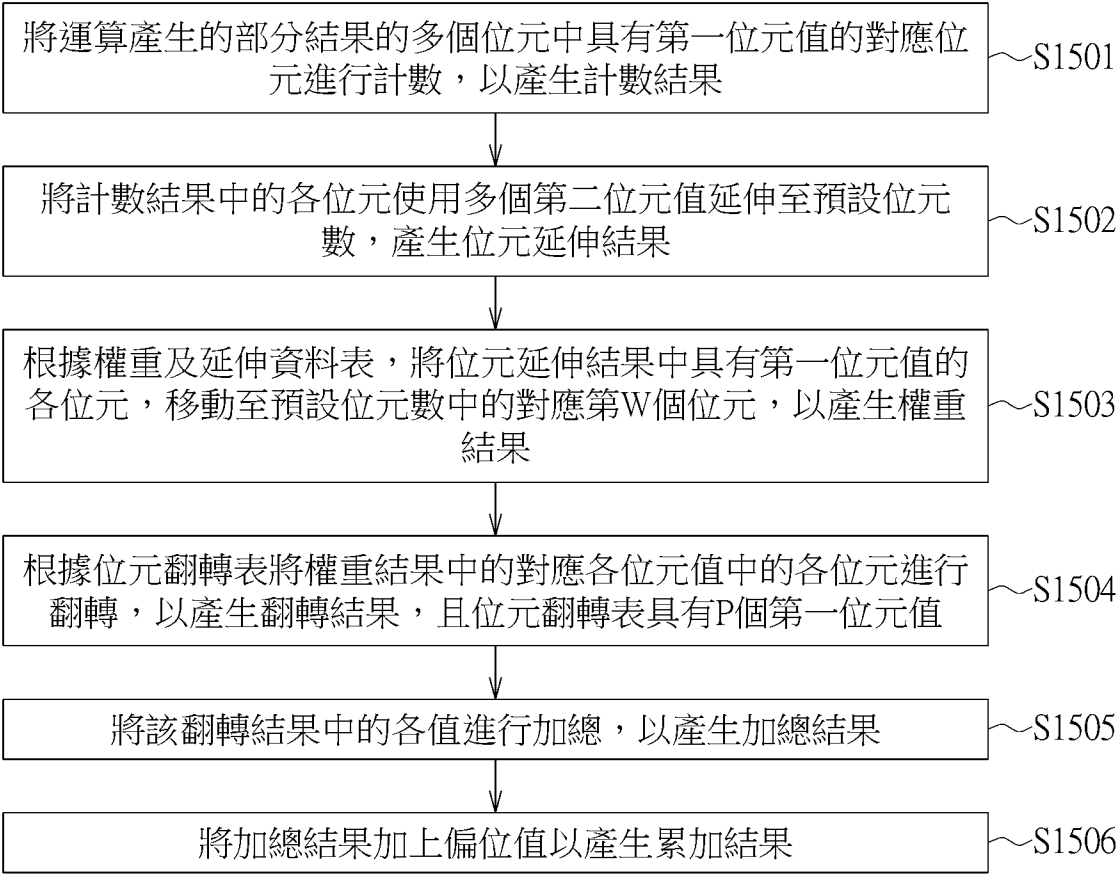
圖 12



第13圖



1500



第 15 圖