



(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 I789546 B

(45)公告日：中華民國 112 (2023) 年 01 月 11 日

(21)申請案號：108132448 (22)申請日：中華民國 108 (2019) 年 09 月 09 日

(51)Int. Cl. : **G06F12/02 (2006.01)** **G06N20/00 (2019.01)**

(30)優先權：2018/11/20 美國 62/770,151
2019/06/03 美國 16/430,408

(71)申請人：南韓商三星電子股份有限公司(南韓) SAMSUNG ELECTRONICS CO., LTD. (KR)
南韓

(72)發明人：卡佳瑞 瑞姆達斯 P KACHARE, RAMDAS P. (IN)；沙爾瑪 瑪納利 SHARMA,
MANALI (IN)

(74)代理人：林孟閱；盧佩君；陳怡如

(56)參考文獻：

TW	201839642A	US	2010/0318718A1
US	2016/0232440A1	US	2018/0074723A1
US	2018/0314985A1		

審查人員：陳泰龍

申請專利範圍項數：20 項 圖式數：4 共 46 頁

(54)名稱

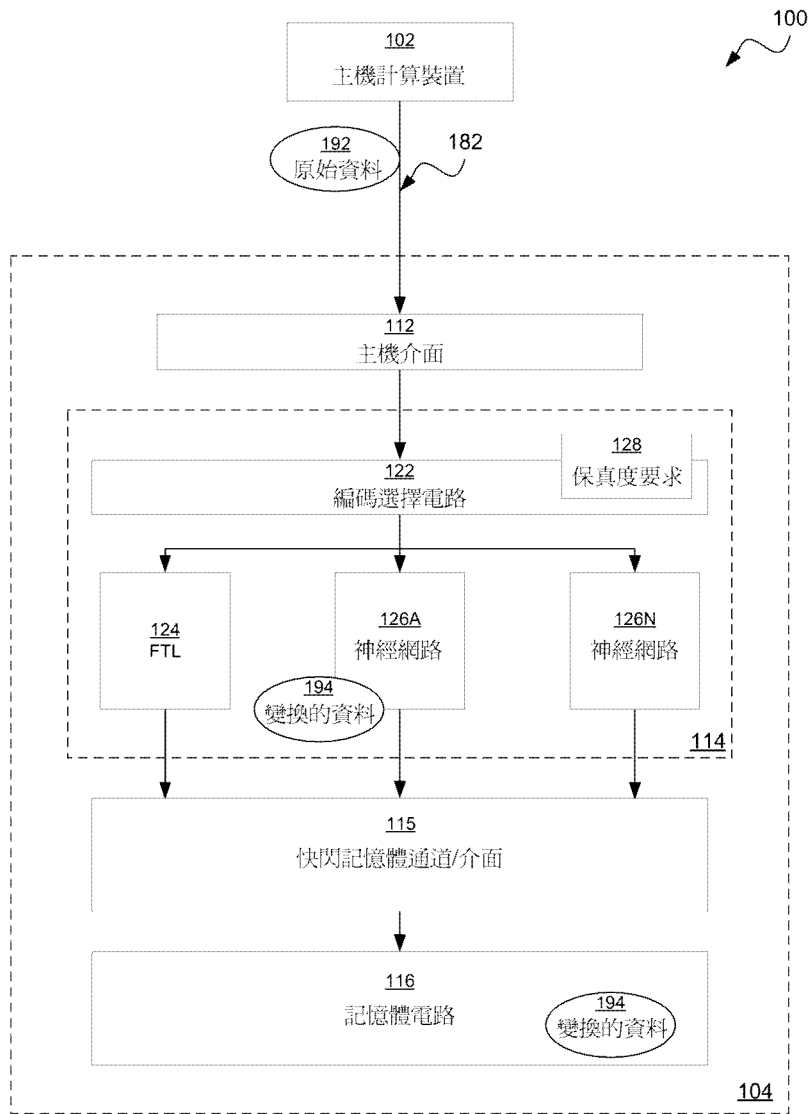
儲存設備、儲存系統及儲存方法

(57)摘要

根據一個一般方面，一種儲存設備可包含配置成接收記憶體存取請求的主機介面電路，其中記憶體存取請求與資料集相關聯。設備可包含配置成創建並儲存變換的資料集的非揮發性記憶體儲存電路。設備可包含包括至少一個機器學習電路的轉換電路。轉換電路可配置成：回應於寫入記憶體存取，將資料集的原始版本轉變為變換的資料集，以及回應於讀取記憶體存取，將變換的資料集轉變為重構的資料集。也提供一種儲存系統及儲存方法。

According to one general aspect, a storage apparatus may include a host interface circuit configured to receive a memory access request, wherein the memory access request is associated with a data set. The apparatus may include a non-volatile memory storage circuit configured to create and store a transformed data set. The apparatus may include a translation circuit comprising at least one machine learning circuit. The translation circuit may be configured to: in response to a write memory access, convert an original version of the data set to the transformed data set, and in response to a read memory access, convert the transformed data set to a reconstructed data set. A storage system and a storage method are also provided.

指定代表圖：



符號簡單說明：

- 100:系統
- 102:主機計算裝置
- 104:儲存裝置
- 112:主機介面電路
- 114:轉換電路
- 115:快閃記憶體通道或介面電路
- 116:儲存/記憶體電路
- 122:編碼/解碼選擇電路
- 124:快閃記憶體轉換層
- 126A、126N:神經網路
- 128:保真度要求
- 182:寫入記憶體存取
- 192:原始資料集
- 194:變換的資料集

【圖 1A】

**公告本**

I789546

【發明摘要】**【中文發明名稱】** 儲存設備、儲存系統及儲存方法**【英文發明名稱】** STORAGE APPARATUS, STORAGE SYSTEM

AND STORAGE METHOD

【中文】 根據一個一般方面，一種儲存設備可包含配置成接收記憶體存取請求的主機介面電路，其中記憶體存取請求與資料集相關聯。設備可包含配置成創建並儲存變換的資料集的非揮發性記憶體儲存電路。設備可包含包括至少一個機器學習電路的轉換電路。轉換電路可配置成：回應於寫入記憶體存取，將資料集的原始版本轉變為變換的資料集，以及回應於讀取記憶體存取，將變換的資料集轉變為重構的資料集。也提供一種儲存系統及儲存方法。

【英文】 According to one general aspect, a storage apparatus may include a host interface circuit configured to receive a memory access request, wherein the memory access request is associated with a data set. The apparatus may include a non-volatile memory storage circuit configured to create and store a transformed data set. The apparatus may include a translation circuit comprising at least one machine learning circuit. The translation circuit may be configured to: in response to a write memory access, convert an original version of the data set to the transformed data set, and in response to a read

第1頁，共2頁(發明摘要)

memory access, convert the transformed data set to a reconstructed data set. A storage system and a storage method are also provided.

【指定代表圖】圖1A。

【代表圖之符號簡單說明】

- 100：系統
- 102：主機計算裝置
- 104：儲存裝置
- 112：主機介面電路
- 114：轉換電路
- 115：快閃記憶體通道或介面電路
- 116：儲存/記憶體電路
- 122：編碼/解碼選擇電路
- 124：快閃記憶體轉換層
- 126A、126N：神經網路
- 128：保真度要求
- 182：寫入記憶體存取
- 192：原始資料集
- 194：變換的資料集

【特徵化學式】

無

【發明說明書】

【中文發明名稱】 儲存設備、儲存系統及儲存方法

【英文發明名稱】 STORAGE APPARATUS, STORAGE SYSTEM

AND STORAGE METHOD

【技術領域】

【0001】 本說明書涉及資料的儲存，且更具體地說，涉及基於神經網路的永久資料儲存。

【先前技術】

【0002】 在現代資訊技術（Information Technology；IT）基礎設施中，各種裝置和工藝生成非常大量的資料。這些資料生成器的一些實例為智慧型機器、自動汽車、社交網路以及物聯網（Internet-of-things；IOT）裝置。正在開發新的人工智慧（artificial intelligence；AI）和機器學習（machine learning；ML）技術，以有效地分析收集到的資料且使用收集到的資料來實現應用的甚至更高的效率和生產率。此類大量的資料儲存在由固態裝置（solid-state device；SSD）、硬碟驅動器（hard disk drive；HDD）、儲存節點以及儲存互連件組成的高性能、可信賴的儲存系統中。

【0003】 人工神經網路（neural network；NN）或連接系統是由構成動物大腦的生物神經網路模糊地啟發的計算系統。神經網路是用於許多不同機器學習技術一起工作並處理複雜資料登錄的框

架。此類系統通常在沒有用任何特定任務規則來程式化的情況下通過考慮實例來“學習”執行任務。舉例來說，在圖像識別中，他們可能通過分析已經手動標記為“貓”或“沒有貓”的實例圖像及使用所述結果識別其它圖像中的貓來學習識別含有貓的圖像。這是他們在無任何關於貓的先驗知識（例如其具有皮毛、尾巴、鬍鬚以及貓類的臉）的情況下進行的。相反，他們從處理的學習材料中自動生成識別特性。

【0004】 NN 是基於稱為人工神經元的連接單元或節點的集合，所述人工神經元鬆散地類比生物大腦中的神經元。每個連接（如生物大腦中的突觸）可將訊號從一個人工神經元傳輸到另一人工神經元。接收訊號的人工神經元可處理訊號，且接著向與其連接的額外人工神經元發送訊號。在常見的 NN 實施方案中，人工神經元之間的連接處的訊號是實數，且每個人工神經元的輸出通過其輸入的總和的一些非線性函數來計算。人工神經元之間的連接稱為‘邊緣’。人工神經元和邊緣通常具有隨著學習進行而調整的權重。權重增加或減小連接處的訊號強度。人工神經元可具有閾值，使得僅在集合訊號超過所述閾值時才發送訊號。通常，人工神經元聚集成層。不同的層可對其輸入執行不同種類的變換。訊號可能在多次穿過層之後從第一層（輸入層）傳播到最後一層（輸出層）。

【發明內容】

【0005】 根據一個一般方面，一種儲存設備可包含配置成接收記憶體存取請求的主機介面電路，其中記憶體存取請求與資料集相關聯。設備可包含配置成儲存變換的資料集的非揮發性記憶體儲存電路。設備可包含包括至少一個機器學習電路的轉換電路。轉換電路可配置成：回應於寫入記憶體存取，將資料集的原始版本轉變為變換的資料集，以及回應於讀取記憶體存取，將變換的資料集轉變為重構的資料集。

【0006】 根據另一一般方面，一種儲存系統可包含配置成執行具有儲存裝置的記憶體存取的主機計算裝置，其中記憶體存取請求與資料集相關聯。系統可包含儲存裝置。儲存裝置可包含配置成儲存變換的資料集的記憶體儲存電路。儲存裝置可包含包括至少一個機器學習電路的轉換電路。轉換電路可配置成響應於記憶體存取來執行以下中的一個：將資料集的原始版本轉變為變換的資料集，將變換的資料集轉變為重構的資料集或返回變換的資料集。

【0007】 根據另一一般方面，一種儲存方法可包含接收原始資料集，以儲存在記憶體電路中。所述方法可包含經由機器學習技術將原始資料集轉變為變換的資料集，其中變換的資料集小於原始資料集。所述方法可包含儲存變換的資料集而不是原始資料集。

【0008】 以下隨附圖式及描述中闡述一或多個實施方案的細節。其它特徵將從描述和圖式以及從權利要求書顯而易見。

【0009】 一種用於資料儲存的系統和/或方法，且更具體地說，涉及基於神經網路的永久資料儲存，基本上如至少一個圖中所示和/

或結合至少一個圖描述，如權利要求書中更完全地闡述。

【圖式簡單說明】

【0010】

圖 1A 是根據所揭露主題的系統的示例性實施例的框圖。

圖 1B 是根據所揭露主題的系統的示例性實施例的框圖。

圖 1C 是根據所揭露主題的系統的示例性實施例的框圖。

圖 2 是根據所揭露主題的系統的示例性實施例的框圖。

圖 3 是根據所揭露主題的系統的示例性實施例的框圖。

圖 4 是可包含根據所揭露主題的原理所形成的裝置的資訊處理系統的示意性框圖。

【0011】 各圖中的相同附圖符號指示相同元件。

【實施方式】

【0012】 將在下文中參考隨附圖式更全面地描述各種示例性實施例，隨附圖式中繪示了一些示例性實施例。然而，本發明所揭露主題可以用許多不同形式實施，並且不應被解釋為限於本文所闡述的示例性實施例。確切地說，提供這些示例性實施例以使得本揭露將為透徹及全面的，且將向所屬領域的技術人員充分地傳達本發明所揭露主題的範圍。在圖式中，為了清楚起見，可能會放大層及區域的大小及相對大小。

【0013】 將理解，當元件或層稱為在另一元件或層“上”、“連接到”另一元件或層或“耦接到”另一元件或層時，所述元件或層可直接在另一元件或層上、直接連接到另一元件或層或耦接到另一元件或層，或可能存在中間元件或層。相比之下，當元件稱為“直接在另一元件或層上”、“直接連接到”或“直接耦接到”另一元件或層時，不存在中間元件或層。相同標號始終指代相同元件。如本文中所使用，術語“和/或”包含相關聯的所列項中的一或多個的任何以及所有組合。

【0014】 將理解，雖然本文中可使用術語第一、第二、第三等來描述各種元件、元件、區域、層和/或區段，但是這些元件、元件、區域、層和/或區段不應受到這些術語的限制。這些術語僅用於區分一個元件、元件、區域、層或區段與另一區域、層或區段。因此，在不脫離本發明所揭露主題的教示的情況下，下文所論述的第一元件、元件、區域、層或區段可稱為第二元件、元件、區域、層或區段。

【0015】 可在本文中使用例如“在...下面（beneath）”、“在...下方（below）”、“下部（lower）”、“在...上方（above）”、“上部（upper）”及類似的空間相對術語，以便於描述如圖式中所說明的一個元件或特徵與另一（些）元件或特徵的關係。將理解，空間相對術語既定涵蓋裝置在使用或操作中除圖式中描繪的取向外的不同取向。舉例來說，如果圖中的裝置倒過來，那麼描述為“在”其它元件或特徵“下方”或“在”其它元件或特徵

“下面”的元件的取向則將變成“在”其它元件或特徵“上方”。因此，示例性術語“在...下方”可涵蓋上方及下方兩個取向。裝置可以以其它方式取向（旋轉 90 度或處於其它取向），且本文中所用的空間相對描述符可以相應地進行解釋。

【0016】 本文中所使用的術語僅出於描述特定示例性實施例的目的，且並不意圖限制本發明所揭露的主題。如本文中所使用，除非上下文另外明確指示，否則單數形式“一(a、an)”以及“所述”意圖還包含複數形式。應更理解，術語“包括(comprises 和/或 comprising)”在用於本說明書中時指定所陳述的特徵、整數、步驟、操作、元件及/或元件的存在，但並不排除一或多個其它特徵、整數、步驟、操作、元件、元件及/或其群組的存在或添加。

【0017】 本文中參考橫截面圖解來描述示例性實施例，所述橫截面圖解為理想化示例性實施例（和中間結構）的示意性圖解。如此，將預期到，作為例如製造技術和/或公差的結果而與圖解的形狀的差異。因此，示例性實施例不應解釋為限於本文中所說明的區域的特定形狀，而是應包含例如由製造引起的形狀偏差。舉例來說，圖解說明為矩形的植入區域通常將具有圓形或彎曲特徵和/或植入物濃度在其邊緣上的梯度，而不是從植入區域到非植入區域的二元變化。同樣地，通過植入形成的埋入區域可在埋入區域與發生植入所在的表面之間的區域中產生一些植入。因此，圖式中所說明的區域本質上是示意性的，並且其形狀並不意圖說明裝置的區域的實際形狀且並不意圖限制本發明所揭露主題的範圍。

【0018】 除非另外定義，否則本文中所有術語（包含技術和科學術語）具有本揭露主題所屬領域的普通技術人員通常所理解的相同含義。應更理解，例如常用詞典中所定義的術語等術語應解釋為具有與其相關技術的上下文中的含義一致的含義，且將不在理想化或過度正式意義上進行解釋，除非明確地如此界定。

【0019】 下文將參考隨附圖式詳細解釋示例性實施例。

【0020】 圖 1A 是根據所揭露主題的系統 100 的示例性實施例的框圖。在各種實施例中，可使用系統 100 來儲存大量資料，如上文所描述。在所示出的實施例中，示出寫入記憶體存取。

【0021】 在所示出的實施例中，系統 100 可包含主機計算裝置 102。在一些實施例中，主機計算裝置 102 可包含各種計算裝置，例如筆記型電腦、臺式電腦、工作站、個人數位助理、智慧手機、平板電腦以及其它適合的電腦或其虛擬機器或虛擬計算裝置。在各種實施例中，用戶（未示出）可使用主機計算裝置 102。在一些實施例中，主機計算裝置 102 可為多個裝置，每個裝置從儲存裝置 104 讀取資料且將資料寫入儲存裝置 104。

【0022】 在所示出的實施例中，系統 100 可包含儲存裝置 104。在各種實施例中，儲存裝置 104 可配置成儲存資料。此外，在各種實施例中，系統 100 可包含多個儲存裝置 104，即使僅一個在圖中示出。在各種實施例中，儲存裝置 104 可包含固態裝置（SSD）、鐵磁性硬碟驅動器（HDD）或任何其它揮發性或非揮發性的儲存技術。

【0023】 傳統地，儲存裝置配置成精確地儲存寫入到其中的資料。當有人將資料寫入到儲存裝置時，他期望稍後提取完全相同的資料。在所示出的實施例中，儲存裝置 104 可不配置成如此操作。相反，儲存裝置 104 可配置成儲存及稍後產生原始資料的表示。在各種實施例中，此表示可為原始資料或與原始資料相等，或可不與原始資料相等。在此實施例中，系統 100 可配置成提供無損儲存和有損儲存中的兩者或任一個。

【0024】 在所示出的實施例中，主機計算裝置 102 可發出寫入記憶體存取 182。寫入記憶體存取 182 可包含待由儲存裝置 104 儲存的原始資料集 192。

【0025】 在所示出的實施例中，儲存裝置 104 可包含主機介面電路 112。在此實施例中，主機介面電路 112 可配置成接收記憶體存取請求（例如請求 182）。在一些實施例中，主機介面電路 112 或層（HIL）可處理儲存 116 的邏輯視圖。在此實施例中，主機計算裝置 102 可使用邏輯位址範圍來存取儲存裝置 104 的儲存容量。在各種實施例中，主機介面電路 112 可配置成在由主機計算裝置 102 使用的邏輯協定與由儲存裝置 104 使用的內部實際物理協定或機制之間進行轉換。在各種實施例中，快閃記憶體通道或介面電路 115 可執行邏輯位址到物理位址轉換。在一些實施例中，儲存裝置 104 可能不使用傳統的物理定址方案，如下所述。

【0026】 在所示出的實施例中，儲存裝置 104 可包含轉換電路 114。在各種實施例中，轉換電路 114 可配置成將原始資料集 192

轉變成變換的資料集 194。在各種實施例中，變換的資料集 194 可小於原始資料集 192。在此實施例中，儲存裝置 104 可配置成儲存原始資料集 192 的減少表示，而不是原始資料自身。

【0027】 在各種實施例中，在需要時，減少表示與神經網路(NN)的狀態一起可用於再創建原始資料或重構原始資料。就原始資料的大小和/或原始資料的值而言，資料的再創建或重構可或可不與原始資料精確匹配。在此實施例中，可減少儲存資料所需的空間的量。如上文所描述，在各種實施例中，可使用無損或有損減少方案來創建變換的資料集 194。當重構準確度（也稱為保真度）小於 100%時，永久儲存據稱為有損的。當重構準確度為 100%時，永久儲存是無損的。

【0028】 使用的實際永久儲存媒體（例如 NAND 快閃記憶體）的量是減少表示的大小和儲存使用的神經網路的狀態所需的儲存媒體的量。神經網路所需的儲存的量取決於在每個網路中的層的數目和節點的數目。舉例來說，對於每一層中具有 128 個節點的雙層完全連接的神經網路，此神經網路儲存需要的參數數目將為 128×128 ，且每一層有一個偏壓參數。在此實例中，為簡單起見，我們假定我們不儲存偏壓參數。假定每個參數需要 64 位元，那麼此雙層神經網路的整個儲存將會是 $(128 \times 128 \times 8) / 1024 = 128 \text{ KB}$ 。假定在儲存裝置中使用兩個網路，那麼儲存網路的狀態所需的 NAND 快閃記憶體容量將為 256 KB。此儲存將在儲存在此儲存裝置中的所有使用者資料上攤銷。也就是說，此儲存量將是固定的

且獨立於實際使用者資料。更假定，每 4 KB 原始資料可減少到 8 位元組用於儲存，這將表示所需儲存容量中減少的 512 倍。換句話說，比方說所提出的儲存裝置具有 1 MB (1024 KB) NAND 快閃記憶體容量。在減去如上文所描述兩個神經網路所需的空間 (256 KB) 之後，剩餘容量 (1024 KB-256 KB) =768 KB 可用于永久儲存使用者資料。假定 4 KB 資料的 8 B 減少表示，768KB 空間可用於儲存每個資料集大小為 4 KB 的 (768 KB /8 B) =96 K 資料集。換句話說，具有 1 MB 實際 NAND 快閃記憶體媒體的儲存裝置可儲存 (96 K X 4 KB) =384 MB 使用者資料。

【0029】 在各種實施例中，轉換電路 114 可包含一或多個機器學習 (ML) 電路。這些 ML 電路可配置成減小原始資料集 192 的大小，且將原始資料集 192 轉變 (convert) 或轉換 (translate) 到變換的資料集 194。在一實施例中，ML 電路可實施神經網路 (NN)，儘管貝葉斯 (Bayesian) 網路、基因技術、主成分分析和/或支援向量機是預期的。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0030】 在所示出的實施例中，轉換電路 114 可包含不僅一個 ML 電路，而是多個 ML 電路 (例如神經網路 126A 和神經網路 126N)。在各種實施例中，經由編碼選擇電路 122，轉換電路 114 可配置成基於一或多個因素選擇 ML 電路中的一個來執行變換。在此實施例中，使用多個 ML 電路可允許轉換電路 114 (基於任何所期望的準則；例如重構保真度) 動態地選擇最佳適合於變換原

始資料集 192 的目的的 ML 電路。編碼選擇電路 122 可選擇 FTL 124 以 100% 保真度按原樣儲存資料，而不需執行原始資料的任何變換。

【0031】 在所示出的實施例中，可設置保真度要求 128，所述保真度要求 128 規定原始資料集 192 與從變換的資料集 194 中重構的資料集之間的變換損失的程度。在此實施例中，轉換電路 114 可基於資料類型或其它因素選擇產生所需保真度（或更佳）的給定 ML 電路。可由主機計算裝置 102 針對每個資料寫入操作提供保真度要求 128。在另一實施例中，保真度要求 128 可配置成由主機計算裝置 102 提供為各種儲存參數的固定值，例如名字空間識別字、主機識別字、LBA 位址範圍、非揮發性記憶體（NonVolatile Memory；NVM）集識別字、NVM 快速（NVM Express；NVMe）提交佇列識別字、流識別字、乙太網媒體存取控制（Media Access Control；MAC）識別字、傳輸控制協議（Transmission Control Protocol；TCP）/網際網路協定（Internet Protocol；IP）位址和其它傳輸/網路參數以及一般參數，如日期和時間等。

【0032】 在另一實施例中，可基於一或多個其它因素來選擇使用哪個 ML 電路或技術。在此實施例中，這些因素可包含資料或檔案類型、源主機（例如主機計算裝置 102）、原始資料集 192 的大小、記憶體電路 116 中的可用空間的量等。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0033】 在各種實施例中，這些要求 128 可包含在記憶體存取

182 中。在另一實施例中，這些要求 128 可經由單獨命令設置，由主機介面電路 112 或其它介面電路（未示出）接收。在又一實施例中，保真度要求 128 可由儲存裝置自身基於各種儲存參數來決定，例如名字空間識別字、主機識別字、LBA 位址範圍、NVM 集識別字、NVMe 提交佇列識別字、流識別字、乙太網 MAC 識別字、TCP/IP 位址和其它傳輸/網路參數以及一般參數，如日期和時間等。

【0034】 在各種實施例中，如下文參考圖 3 所描述，可聚集轉換電路 114 以動態地更改 ML 電路或技術的操作從而調整輸出特性（例如保真度）。在此實施例中，特定 ML 電路的輸出特性可為可調節的。

【0035】 在一些實施例中，轉換電路 114 可包含無損 ML 電路。在此實施例中，變換的資料集 194 可小於或等於原始資料集 192 的大小。

【0036】 在所示出的實施例中，轉換電路 114 可包含基於傳統的非神經網路的直接轉換電路 124，所述轉換電路創建與原始資料集 192 相同或相等的變換的資料集 194 的版本。在所示出的實施例中，轉換電路 114 可包含快閃記憶體轉換層（Flash Translation Layer；FTL）124 或使用快閃記憶體轉換層（FTL）124。在此實施例中，FTL 電路 124 可配置成執行邏輯位址到物理位址轉換且將原始資料集 192（如變換的資料集 194）直接寫入記憶體電路 116 中。在此實施例中，轉換電路 114 可使用 FTL 124 和 ML 電路當

中的一或多個選項。

【0037】 在所示出的實施例中，儲存裝置 104 可包含記憶體電路 116。在此實施例中，記憶體電路 116 可配置成儲存變換的資料集 194。在此實施例中，一旦創建變換的資料集 194，則所述變換的資料集就可儲存在記憶體電路 116 中。原始資料集 192 可全部或部分丟棄。舉例來說，如果記憶體存取 182 使用邏輯塊定址(logical block addressing；LBA)，那麼可丟棄全部原始資料集 192，且儲存裝置 104 可依靠從變換的資料集 194 中重構資料集 196 的能力。在另一實施例中，如果記憶體存取 182 使用鍵值(key-value；KV)對，那麼原始資料集 192 的值部分可丟棄，但鍵部分可保留。在此實施例中，鍵部分(通常比值部分小得多)可保留作為變換的資料集 194(目前用作值部分)的索引。

【0038】 在各種實施例中，儘管有損失，但變換的資料集的使用可為足夠的。舉例來說，在保真度和回應時間都不是非常重要的應用程式中，例如感測器資料、IoT 資料、視頻以及圖像等。另一實例可能需要保真度不重要但回應時間重要的的應用程式，例如需要來自系統的快速回應的網路應用程式。在另一實施例中，資料儲存的調整能力可為合意的。例如保真度重要但回應時間不重要的應用程式，例如定期監視動物園中動物行為的應用程式。或在另一實例中，保真度和回應時間都非常重要的應用程式，例如儲備區段。本文所述的儲存裝置稱為深固態裝置(深 SSD)，所述儲存裝置利用轉換電路 114 將原始資料集 192 變換為減少表示資

料集（例如變換的資料集 194），且從資料集（例如變換的資料集 194 中）重構資料集（例如重構的資料集 196）。在此上下文中，深 SSD 可包含永久資料儲存裝置的類型，所述資料儲存裝置使用神經網路或類似技術實施一或多個“功能”，所述神經網路或類似技術將原始資料變換為減小大小表示，且在需要時，從減小大小表示中重構原始資料或原始資料的近似。“功能”可基本上使用輸入值，稱為深鍵，所述輸入值從根本上為維持到裝置內部的原始資料的減少表示，以產生符合給定程度的原始資料的輸出。深 SSD 可通過“按原樣”儲存資料像常規 SSD 一樣運作。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0039】 圖 1B 是根據所揭露主題的系統 100 的示例性實施例的框圖。在各種實施例中，可使用系統 100 儲存大量資料，如上文所描述。在所示出的實施例中，示出讀取記憶體存取。

【0040】 在所示出的實施例中，系統 100 可同樣包含主機計算裝置 102 和儲存裝置 104。在此實施例中，儲存裝置 104 可包含主機介面電路 112、轉換電路 114 以及記憶體電路 116。

【0041】 在所示出的實施例中，轉換電路 114 可包含一或多個 ML 電路，所述 ML 電路配置成解碼或再創建已知為重構的資料集 196 的原始資料集（或其複寫或相似物）。在包含多個 ML 電路（例如 NN 126A 和 NN 126N）和/或 FTL 124 的各種實施例中，轉換電路 114 可包含編碼（或解碼）選擇電路 122。

【0042】 在所示出的實施例中，主機計算裝置 102 可發出讀取記

憶體存取或請求 183。讀取請求 183 可與原始資料集（圖 1A 中所示）相關聯。在此實施例中，讀取請求 183 可包含用於指示正在請求哪個資料集的邏輯位址或（鍵值對的）鍵。

【0043】 在所示出的實施例中，儲存裝置 104 可尋求滿足讀取請求 183，但其可能不具有請求的原始資料集。相反，儲存裝置 104 可具有變換的資料集 194。在此實施例中，轉換電路 114 可經由用於在第一位置中創建變換的資料集 194 的 ML 電路（或相關聯的解碼 ML 電路）運行變換的資料集 194。在此實施例中，ML 電路可從變換的資料集 194 中創建重構的資料集 196。此重構的資料集 196 可為儲存裝置 104 在再創建原始資料集的最佳或所選嘗試。

【0044】 在一個實施例中，如果變換的資料集 194 為無損編碼（或經由 FTL 124 的直接複製），那麼重構的資料集 196 可與原始資料集相同。在另一實施例中，如果使用有損編碼建構變換的資料集 194，那麼重構的資料集 196 可不與原始資料集相同或相等。在此實施例中，系統 100 可設置其中重構的資料集 196 僅需部分精確（例如 90%精確、80%精確等）的保真度要求 128。

【0045】 在此實施例中，一旦創建重構的資料集 196，那麼重構的資料集 196 就可發送回主機計算裝置 102（如讀取回應 184）。在各種實施例中，主機介面電路 112 可配置成傳輸此讀取響應 184。在各種實施例中，如果重構的資料集 196 是有損的，那麼重構的資料集 196 可小於原始資料集，且因此消耗更少頻寬或其它系統資源以便滿足讀取請求 183。在另一實施例中，重構的資料仍

可與原始資料的精確大小匹配，但重構的內容或值可能不精確匹配。

【0046】 圖 1C 是根據所揭露主題的系統 100 的示例性實施例的框圖。在各種實施例中，可使用系統 100 來儲存大量資料，如上文所描述。在所示出的實施例中，示出讀取記憶體存取。

【0047】 在所示出的實施例中，系統 100 可同樣包含主機計算裝置 102 和儲存裝置 104。在此實施例中，儲存裝置 104 可包含主機介面電路 112、轉換電路 114 以及記憶體電路 116。

【0048】 在所示出的實施例中，轉換電路 114 可包含一或多個 ML 電路，所述 ML 電路配置成解碼或再創建已知為重構的資料集 196 的原始資料集（或其複寫或相似物）。在其中包含多個 ML 電路（例如 NN 126A 和 NN 126N）和/或 FTL 124 的各種實施例中，轉換電路 114 可包含編碼（或解碼）選擇電路 122。

【0049】 在所示出的實施例中，主機計算裝置 102 可發出讀取記憶體存取或請求 185。讀取請求 185 可與原始資料集（圖 1A 中所示）相關聯。在此實施例中，讀取請求 185 可包含用於指示正在請求哪個資料集的邏輯位址或（鍵值對的）鍵。

【0050】 然而，在所示出的實施例中，讀取請求 185 可能不要求或期望返回重構的資料集。相反，讀取請求 185 可要求變換的資料集 194 自身。

【0051】 在此實施例中，儲存裝置 104 可包含繞過轉換電路 114 或至少 ML 電路的路徑 170。在各種實施例中，路徑 170 可利用

FTL 124，而不是圖 1C 中所示的。應理解，上述僅僅是並未限制所揭露主題的一個說明性實例。

【0052】 在此實施例中，儲存裝置 104 可確定正在請求哪個變換的資料集 194，且直接（經由讀取回應 186）將變換的資料集 194 提供到主機計算裝置 102。在此實施例中，儲存裝置 104 與主機計算裝置 102 之間傳輸的資料量可顯著減少。

【0053】 在一個實施例中，主機計算裝置 102 可包含與編碼變換的資料集 194 的 ML 電路類似的 ML 電路。在此實施例中，主機計算裝置 102 可能夠本地創建重構的資料集（到主機 102）。在此實施例中，儲存裝置 104 和主機裝置 102 可能夠轉移 ML 電路權重或設置，使得 ML 電路可保持同步。

【0054】 圖 2 是根據所揭露主題的系統 200 的示例性實施例的框圖。在所示出的實施例中，系統 200 可包含圖 1A 的儲存裝置。

【0055】 在所示出的實施例中，系統 200 可包含主機介面電路 112，如上文所描述。在此實施例中，主機介面電路 112 可配置成將資料傳輸到外部計算裝置（未示出）及從所述外部計算裝置（未示出）接收資料。

【0056】 在所示出的實施例中，系統 200 可包含轉換電路 114，如上文所描述。在此實施例中，轉換電路 114 可包含一或多個機器學習（ML）電路。

【0057】 在所示出的實施例中，每個 ML 電路可包含編碼電路（例如編碼電路 222A、編碼電路 222B 以及編碼電路 222C）。編碼電

路可配置成將原始資料集變換為變換的資料集，如上文所描述。

【0058】 在所示出的實施例中，每個 ML 電路可包含解碼電路(例如解碼電路 224A、解碼電路 224B 以及解碼電路 224C)。解碼電路可配置成將變換的資料集變換為重構的資料集，如上文所描述。

【0059】 在各種實施例中，轉換電路 114 可包含單個編碼電路和多個解碼電路。在此實施例中，轉換電路 114 可選擇哪個解碼電路產生最小差互斥或最高保真度。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0060】 在所示出的實施例中，每個 ML 電路可包含差分或 Δ 電路(例如差分電路 226A、差分電路 226B 以及差分電路 226C)。在各種實施例中，轉換電路 114 可包含一個差分或 Δ 電路，或多個 ML 電路可共用差分或 Δ 電路。在此實施例中，差分電路可配置成將原始資料集與重構的資料集進行比較。在此實施例中，差分電路可指示或計算提供給相應 ML 電路的保真度的水準或量。如上文所描述，當確定使用哪個 ML 電路來編碼原始資料集時，轉換電路 114 可使用此保真度水準。在另一實施例中，觀測到的 Δ 可用于更新在編碼器和/或解碼器電路中使用的神經網路的權重和其它此類參數，以實現所期望的重構保真度。也就是說，在資料寫入操作期間，觀測到的重構 Δ 可用於訓練神經網路，使得稍後可重構輸入資料到所期望的準確度水準。在一些實施例中，神經網路為預先訓練的，且僅改變網路的層數以實現所期望的重構保真度。

【0061】 在各種實施例中，系統 200 可包含寫入快取記憶體 217。寫入快取記憶體 217 可配置成提供臨時記憶體。在各種實施例中，此臨時記憶體可用於原始資料集和/或 ML 電路的編碼/解碼程式期間的任何中間值。

【0062】 在各種實施例中，系統 200 可包含記憶體電路 216，如上文所描述。在各種實施例中，記憶體電路 216 可儲存變換的資料集。記憶體電路 216 還可儲存所期望的保真度水準，所述保真度水準可在資料的重構期間使用。

【0063】 在所示出的實施例中，記憶體電路 216 可被組織成資料結構，例如查詢表、索引陣列或一系列欄位。在所示出的實施例中，記憶體電路 216 可包含索引欄位 252，所述索引欄位 252 允許系統 200 找到變換的資料集。在各種實施例中，索引欄位 252 可包含邏輯塊位址（LBA）或鍵值（KV）對的鍵。

【0064】 記憶體電路 216 可包含資料欄位 254，所述資料欄位 254 儲存變換的資料集或變換的資料集的指針。在此實施例中，因為原始資料集的相同實例可產生相同的變換的資料集，因此系統 200 可執行重復資料刪除，且因此節省額外空間。

【0065】 在一個實施例中，記憶體電路 216 可在資料欄位 254 中儲存變換的資料集 194。我們稱儲存在資料欄位 254 中的變換的資料集 194 為深鍵。在此實施例中，變換的資料集 194（即，深鍵）可或可不單獨地儲存在第二記憶體電路 218 中。包含在第一記憶體電路 216 中的深鍵永久地儲存到第二記憶體電路 218 中。

【0066】 記憶體電路 216 可包含編碼欄位 256，所述編碼欄位儲存且識別哪個 ML 電路編碼及/或將用於對變換的資料集進行解碼。在此實施例中，轉換電路 114 可能夠快速確定在決定過程中應使用哪個 ML 電路。

【0067】 在所示出的實施例中，系統 200 還可包含第二記憶體電路 218 和快閃記憶體通道或介面電路 215。在各種實施例中，第一記憶體電路 216 和第二記憶體電路 218 可包含在同一記憶體電路中，但在單獨分區或部分內。

【0068】 在各種實施例中，第二記憶體電路 218 可配置成儲存各種 ML 電路的狀態或權重值（例如 NN 狀態 228A、NN 狀態 228B 以及 NN 狀態 228C）。在此實施例中，轉換電路 114 可在 ML 電路或 ML 電路的配置之間快速地且動態地切換。在各種實施例中，轉換電路 114 可包含裝置處理器或控制器，所述裝置處理器或控制器配置成通過載入不同 ML 狀態或設置在 ML 電路之間切換。

【0069】 如上文所描述，系統 200 可包含以下方面或優勢。在所示出的實施例中，系統 200 可使用神經網路作為儲存媒體而不是在記憶體電路上“按原樣”儲存資料。系統 200 與傳統的儲存裝置相比可顯著地增加儲存容量。在各種實施例中，某些操作（例如檢索）可由神經網路和/或機器學習電路執行，而不是對全部資料執行操作，從而使資料操作更快。舉例來說，機器學習電路可將資料從第一格式變換為第二格式。反過來，機器學習電路可對第二格式的資料執行一或多個檢索，而不將資料變換為第一格式。

【0070】 在各種實施例中，系統 200 可自動對所儲存資料執行重復資料刪除。在此實施例中，重復資料刪除可位於塊層級或資料組塊的一些細微性，從而產生更多的資料的重復資料刪除。

【0071】 在一些實施例中，主機可指定用於儲存資料的保真度的不同程度。如果保真度是 100%，那麼資料（無論是無損地還是直接地儲存）可“按原樣”返回到主機。如果保真度小於 100%，那麼返回到主機的資料將為原始資料的近似，所述近似可為或可不為原始資料大小。也就是說，在一些實施例中，返回的資料可具有與原始資料完全相同的大小，但返回的資料值可能不與原始資料值精確地匹配。

【0072】 在此實施例中，系統 200 可補充機器學習應用程式。在各種實施例中，機器學習應用程式可能不需要重構的資料；而是可使用減小的特徵表示。舉例來說，可直接使用變換的資料集來加速 ML 應用程式。

【0073】 圖 3 是根據所揭露主題的系統 300 的示例性實施例的框圖。在各種實施例中，系統 300 可包含機器學習（ML）電路或軟體實施方案。在一個具體實施例中，系統 300 可包含神經網路。

【0074】 在所示出的實施例中，系統 300 可包含編碼器神經網路 302 和解碼器神經網路 304。在各種實施例中，神經網路 302 和神經網路 304 可各自包含多個層（例如層 312A、層 312B、層 312C、層 312D 以及層 314A、層 314B、層 314C 以及層 314D）。每一層 312A-312D/層 314A-314D 可包含處理層的輸入以產生輸出的各種

權重和神經元的數目。神經網路中的每一層可具有不同的類型且可包含不同的神經網路節點、操作以及層，例如卷積、兩層之間的密集連接、遞迴神經網路單元、長短期記憶單元、閘控遞迴單元、生成對抗網路、S 型 (sigmoid)、柔性最大傳遞函數 (softmax) 等。應理解，這些僅是並未限制所揭露主題的說明性實例。接著將輸出用作另一層的輸入，直到達到神經網路的最終輸出層 (例如層 312D 或層 314D)。

【0075】 在所示出的實施例中，原始資料集 192 可輸入到編碼神經網路 302 以產生變換的資料集 194。同樣地，變換的資料集 194 可輸入到解碼神經網路 304 以產生重構的資料集 196。

【0076】 在此實施例中，系統 300 可包含差分或 Δ 電路 306。 Δ 電路 306 可將原始資料集 192 與重構的資料集 196 進行比較，且確定神經網路 302 和神經網路 304 重現原始資料集 192 的精確程度。在各種實施例中，此準確度可與保真度要求相比，如上文所描述。

【0077】 在各種實施例中，系統 300 可配置成通過跳過或繞過神經網路 302 和神經網路 304 內的各層來動態地調整系統 300 的準確度。舉例來說，在一個實施例中，如果保真度比所需的高，那麼系統 300 可繞過層 312B。如果保真度仍比所需的高，那麼系統 300 可繞過層 312B 和層 312C。同樣地，可跳過解碼神經網路 304 中的層 (例如層 314C)。相反，如果獲得的保真度比所需的水準低，那麼先前繞過的層可再添加到計算以增加保真度水準。

【0078】 在各種實施例中，可在編碼神經網路 302 和解碼神經網路 304 內同時添加/繞過層。在另一實施例中，可獨立地添加/繞過層。

【0079】 在各種實施例中，系統 300 可包含各種設置或受到各種設置限制。舉例來說，系統 300 可受到可跳過或繞過的層的最大數目的限制。在另一實施例中，系統 300 可受到最大次數或實現所期望的保真度要求（或其它要求，例如計算時間等）可能進行的嘗試的限制。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0080】 在各種實施例中，系統 300 可使此保真度計算作為資料儲存操作的部分。在此實施例中，系統 300 確定使用什麼神經網路設置來實現所期望的保真度或其它資源要求。在此實施例中，神經網路的設置可與用於編碼資料的神經網路一起儲存。

【0081】 在各種實施例中，如果系統 300 不能夠在設定的嘗試次數內或根本不可能實現所期望的保真度，那麼系統 300 可使用實現盡可能接近保真度要求的保真度水準的設置。在另一實施例中，如果系統 300 不能夠實現所期望的保真度，那麼系統 300 可“按原樣”或以其原始狀態儲存原始資料集 192（例如經由 FTL）。在另一實施例中，如果系統 300 不能夠實現所期望的精確保真度，那麼系統 300 可以高於所期望的保真度水準但小於 100% 保真度水準的保真度水準儲存原始資料。在又一實施例中，如果系統 300 不能夠實現所期望的保真度（或其它要求），那麼系統 300

可切換到多個 ML 電路中的另一個且再次嘗試。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0082】 圖 4 是資訊處理系統 400 的示意性框圖，所述資訊處理系統可包含根據所揭露主題的原理形成的半導體裝置。

【0083】 參考圖 4，資訊處理系統 400 可包含根據所揭露主題的原理建構的裝置中的一或多個。在另一實施例中，資訊處理系統 400 可使用或執行根據所揭露主題的原理的一或多種技術。

【0084】 在各種實施例中，資訊處理系統 400 可包含計算裝置，例如筆記型電腦、臺式電腦、工作站、伺服器、刀片伺服器、個人數位助理、智慧手機、平板電腦以及其它適合的電腦或其虛擬機器或虛擬計算裝置。在各種實施例中，資訊處理系統 400 可由使用者（未示出）使用。

【0085】 根據所揭露主題的資訊處理系統 400 可更包含中央處理單元（central processing unit；CPU）、邏輯或處理器 410。在一些實施例中，處理器 410 可包含一或多個功能單元塊（functional unit block；FUB）或組合邏輯塊（combinational logic block；CLB）415。在此實施例中，組合邏輯塊可包含各種布林邏輯操作（Boolean logic operation）（例如反及（NAND）、反或（NOR）、反相（NOT）、互斥或（XOR）、穩定邏輯裝置（例如正反器、栓鎖）、其它邏輯裝置或其組合。這些組合邏輯操作可以簡單或複雜方式配置成處理輸入訊號以實現所需結果。應理解，在描述同步組合邏輯操作的幾個說明性實例時，所揭露主題不受如此限制且可包含非同步

作業或其混合。在一個實施例中，組合邏輯操作可包括多個互補金屬氧化物半導體（complementary metal oxide semiconductor；CMOS）電晶體。在各種實施例中，這些 CMOS 電晶體可佈置到執行邏輯操作的閘中；但應理解，可使用其它技術且所述其它技術在所揭露主題的範圍內。

【0086】 根據所揭露主題的資訊處理系統 400 可更包含揮發性記憶體 420（例如隨機存取記憶體（Random Access Memory；RAM））。根據所揭露主題的資訊處理系統 400 可更包含非揮發性記憶體 430（例如硬碟驅動器、光學記憶體、NAND 記憶體或快閃記憶體）。在一些實施例中，揮發性記憶體 420、非揮發性記憶體 430 或其組合或部分可稱為“儲存媒體”。在各種實施例中，揮發性記憶體 420 和/或非揮發性記憶體 430 可配置成以半永久或基本上永久的形式儲存資料。

【0087】 在各種實施例中，資訊處理系統 400 可包含一或多個網路介面 440，所述介面配置成允許資訊處理系統 400 為通訊網路的部分且經由通訊網路通訊。Wi-Fi 協定的實例可包含但不限於電氣和電子工程師學會（Institute of Electrical and Electronics Engineer；IEEE）802.11g, IEEE 802.11n。蜂窩協定的實例可包含但不限於：IEEE 802.16m（也稱為無線都會區域網路（Metropolitan Area Network；MAN）改進）、長期演進（Long Term Evolution；LTE）改進、全球移動通訊系統（Global System for Mobile Communications；GSM）的增強資料速率演進（Enhanced Data rates

for GSM Evolution ; EDGE)、演進高速封包存取 (Evolved High-Speed Packet Access ; HSPA+)。有線協定的實例可包含但不限於 IEEE 802.3 (也稱為乙太網路 (Ethernet))、光纖通道 (Fibre Channel)、電力線通訊 (Power Line communication) (例如 HomePlug 、 IEEE 1901)。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0088】 根據所揭露主題的資訊處理系統 400 可更包含使用者介面單元 450 (例如顯示卡、觸覺介面、人機介面裝置)。在各種實施例中，此使用者介面單元 450 可配置成從使用者接收輸入和/或將輸出提供給用戶。其它種類的裝置同樣可用於提供與用戶的互動；例如，向用戶提供的回饋可以是任何形式的感覺回饋 (例如視覺回饋、聽覺回饋或觸覺回饋)；且來自用戶的輸入可以任何形式接收，包含聲音、語音或觸覺輸入。

【0089】 在各種實施例中，資訊處理系統 400 可包含一或多個其它裝置或硬體元件 460 (例如顯示器或監視器、鍵盤、滑鼠、相機、指紋讀取器、視頻處理器)。應理解，上述僅是並未限制所揭露主題的一些說明性實例。

【0090】 根據所揭露主題的資訊處理系統 400 可更包含一或多個系統匯流排 405。在此實施例中，系統匯流排 405 可配置成以通訊方式耦接處理器 410、揮發性記憶體 420、非揮發性記憶體 430、網路介面 440、使用者介面單元 450 以及一或多個硬體元件 460。由處理器 410 處理的資料或從非揮發性記憶體 430 外部輸入的資

料可儲存在非揮發性記憶體 430 或揮發性記憶體 420 中。

【0091】 在各種實施例中，資訊處理系統 400 可包含或執行一或多個軟體元件 470。在一些實施例中，軟體元件 470 可包含作業系統（operating system；OS）和/或應用程式。在一些實施例中，OS 可配置成向應用程式提供一或多個服務且管理或充當資訊處理系統 400 的應用程式與各種硬體元件（例如處理器 410、網路介面 440）之間的中繼裝置。在此實施例中，資訊處理系統 400 可包含一或多個原生應用程式，所述程式可本地（例如在非揮發性記憶體 430 內）安裝且配置成由處理器 410 直接執行且與 OS 直接交互。在此實施例中，原生應用程式可包含預編譯機器可執行碼。在一些實施例中，原生應用程式可包含配置成將原始程式碼或目標代碼轉譯成隨後由處理器 410 執行的可執行碼的腳本解譯器（例如 C shell（csh）、蘋果腳本（AppleScript）、熱鍵腳本（AutoHotkey））或虛擬執行機（virtual execution machine；VM）（例如 Java 虛擬機器（Java Virtual Machine）、微軟公共語言運行時（Microsoft Common Language Runtime））。

【0092】 上文所描述的半導體裝置可使用各種封裝技術來包封。舉例來說，根據所揭露主題的原理構建的半導體裝置可使用以下技術中的任一種來包封：層疊封裝（package on package；POP）技術、球柵陣列（ball grid array；BGA）技術、晶片大小封裝（chip scale package；CSP）技術、塑膠引線晶片載體（plastic leaded chip carrier；PLCC）技術、塑膠雙列直插式封裝（plastic dual in-line

package ; PDIP) 技術、華夫包裝式裸片 (die in waffle pack) 技術、晶片式裸片 (die in wafer form) 技術、板上晶片 (chip on board ; COB) 技術、陶瓷雙列直插封裝 (ceramic dual in-line package ; CERDIP) 技術、塑膠公制四方扁平封裝 (plastic metric quad flat package ; PMQFP) 技術、塑膠四方扁平封裝 (plastic quad flat package ; PQFP) 技術、小外形積體電路封裝 (small outline integrated circuit package ; SOIC) 技術、緊縮小外形封裝 (shrink small outline package ; SSOP) 技術、薄型小外形封裝 (thin small outline package ; TSOP) 技術、薄型四方扁平封裝 (thin quad flat package ; TQFP) 技術、系統級封裝 (system in package ; SIP) 技術、多晶片封裝 (multi-chip package ; MCP) 技術、晶片級構造封裝 (wafer-level fabricated package ; WFP) 技術、晶片級處理堆疊封裝 (wafer-level processed stack package ; WSP) 技術或如所屬領域的技術人員將已知的其它技術。

【0093】 方法步驟可由執行電腦程式的一或多個可程式化處理器來執行，從而通過對輸入資料進行操作並產生輸出來執行功能。方法步驟還可由可實施為專用邏輯電路（例如現場可程式化閘陣列 (field programmable gate array ; FPGA) 或專用積體電路 (application-specific integrated circuit ; ASIC)) 的設備來執行。

【0094】 在各種實施例中，電腦可讀媒體可包含在執行時使得裝置執行方法步驟的至少一部分的指令。在一些實施例中，電腦可讀媒體可包含在磁性媒體、光學媒體、其它媒體或其組合（例如

CD-ROM、硬碟驅動器、唯讀記憶體、快閃記憶體驅動器) 中。

在此實施例中，電腦可讀媒體可以是有形且非暫時性實施的製品。

【0095】 在已參考示例性實施例描述所揭露主題的原理時，所屬領域的技術人員將顯而易見的是可在不脫離這些揭露概念的精神和範圍的情況下對其作出各種改變以及修改。因此，應理解，上述實施例並非限制性的，而僅是說明性的。因此，所揭露概念的範圍將通過所附權利要求書和其等效物所最廣泛容許的解釋來確定，且不應受前文描述的約束或限制。因此，應理解，所附權利要求書旨在涵蓋如屬於實施例的範圍內的所有此類修改和改變。

【符號說明】

【0096】

100、200、300：系統

102：主機計算裝置

104：儲存裝置

112：主機介面電路

114：轉換電路

115、215：快閃記憶體通道或介面電路

116：儲存/記憶體電路

122：編碼/解碼選擇電路

124：快閃記憶體轉換層

126A、126N：神經網路

- 128：保真度要求
- 170：路徑
- 182：寫入記憶體存取
- 183、185：讀取記憶體存取或請求
- 184、186：讀取回應
- 192：原始資料集
- 194：變換的資料集
- 196：資料集/重構的資料集
- 216：記憶體電路
- 217：寫入快取記憶體
- 218：第二記憶體電路
- 222A、222B、222C：編碼電路
- 224A、224B、224C：解碼電路
- 226A、226B、226C：差分電路
- 228A、228B、228C：NN 狀態
- 252：索引欄位
- 254：資料欄位
- 256：編碼欄位
- 302：編碼神經網路
- 304：解碼神經網路
- 306：差分或 Δ 電路
- 312A、312B、312C、312D、314A、314B、314C、314D：層

- 400：資訊處理系統
- 405：系統匯流排
- 410：處理器
- 415：組合邏輯塊
- 420：揮發性記憶體
- 430：非揮發性記憶體
- 440：網路介面
- 450：使用者介面單元
- 460：硬體元件
- 470：軟體元件

【發明申請專利範圍】

【第1項】 一種儲存設備，包括：

主機介面電路，配置成接收記憶體存取請求，其中所述記憶體存取請求與資料集相關聯；

至少一個非揮發性記憶體儲存電路，配置成儲存變換的資料集；以及

轉換電路，包括機器學習電路，且配置成：

基於寫入記憶體存取，將所述資料集的原始版本轉變為所述變換的資料集，以及

基於讀取記憶體存取，將所述變換的資料集轉變為重構的資料集，所述重構的資料集包括不同於所述資料集的所述資料集的近似；

其中所述至少一個非揮發性記憶體儲存電路配置成儲存所述機器學習電路的永久狀態，所述機器學習電路用以將所述資料集的所述原始版本轉變為所述變換的資料集；以及

其中所述機器學習電路的所述永久狀態包括將所述變換的資料集轉變為所述重構的資料集的配置。

【第2項】 如申請專利範圍第1項所述的儲存設備，其中所述機器學習電路包括第一機器學習電路，並且其中所述轉換電路包括第二機器學習電路，以及

其中所述轉換電路配置成至少部分地基於主機所期望的保真度的量以及可由所述機器學習電路提供的保真度的量來選擇所述

第一機器學習電路或所述第二機器學習電路中的一個以轉變所述資料集。

【第3項】 如申請專利範圍第2項所述的儲存設備，其中所述主機所期望的所述保真度是用於儲存參數的集合的固定值，其中所述儲存參數的集合包括下列的一或多個：

名字空間識別字、主機識別字、邏輯塊位址範圍、非揮發性記憶體集識別字、非揮發性記憶體快速提交佇列識別字、流識別字、乙太網媒體存取控制識別字、網路位址、傳輸參數、日期或時間。

【第4項】 如申請專利範圍第2項所述的儲存設備，

其中至少部分地基於與所述記憶體存取請求相關聯的資料類型來調整所期望的所述保真度；以及

其中至少部分地基於與所述記憶體存取請求相關聯的軟體應用來調整所期望的所述保真度。

【第5項】 如申請專利範圍第1項所述的儲存設備，其中所述變換的資料集具有小於或等於所述資料集的所述原始版本的大小；以及

其中所述重構的資料集不同於所述資料集的所述原始版本。

【第6項】 如申請專利範圍第1項所述的儲存設備，其中所述轉換電路配置成對所述資料集執行重復資料刪除，且其中在塊層級處執行資料的重復資料刪除。

【第7項】如申請專利範圍第1項所述的儲存設備，其中，基於讀取記憶體存取，主機介面配置成返回所述變換的資料集，其中所述變換的資料集小於所述資料集的所述原始版本或與所述資料集的所述原始版本大小相同，且所述變換的資料集是所述資料集的近似。

【第8項】如申請專利範圍第1項所述的儲存設備，其中所述至少一個非揮發性記憶體儲存電路包括：

使包含在記憶體存取中的定址值與所述變換的資料集相關聯的第一欄位，以及指示所述機器學習電路用於創建所述變換的資料集的第二欄位。

【第9項】如申請專利範圍第1項所述的儲存設備，其中所述轉換電路包含快閃記憶體轉換層電路，所述快閃記憶體轉換層電路配置成創建與所述資料集的所述原始版本相等的所述變換的資料集的版本，且創建與所述資料集的所述原始版本相等的重構的資料集的版本；以及

其中所述轉換電路配置成基於保真度要求，在使用所述機器學習電路以處理所述變換的資料集。

【第10項】如申請專利範圍第1項所述的儲存設備，其中所述機器學習電路包含神經網路；以及

其中所述轉換電路配置成至少部分地基於保真度要求調整在所述神經網路中的層的數目。

【第11項】如申請專利範圍第1項所述的儲存設備，其中所述機器學習電路包括第一機器學習電路，並且其中所述轉換電路包括第二機器學習電路，並且所述轉換電路配置成：

使用觀測到的重構 Δ 來訓練所述第一機器學習電路，且使用所述觀測到的重構 Δ 來確定基於使用所述第一機器學習電路的保真度；以及

基於所述保真度選擇所述第一機器學習電路。

【第12項】如申請專利範圍第1項所述的儲存設備，其中所述機器學習電路包含編碼器神經網路以及二或多個解碼器神經網路，或多個編碼器神經網路及多個解碼器神經網路。

【第13項】如申請專利範圍第1項所述的儲存設備，其中所述轉換電路配置成至少部分地根據保真度目標來確定執行所述資料集的所述原始版本到所述變換的資料集的有損轉換；以及

其中所述轉換電路配置成至少部分地基於保真度目標來調整損失的量。

【第14項】一種儲存系統，包括：

主機計算裝置，配置成執行使用儲存裝置的寫入記憶體存取及讀取記憶體存取，其中所述寫入記憶體存取及所述讀取記憶體存取與資料集相關聯；

所述儲存裝置包括：

至少一個記憶體儲存電路，配置成儲存變換的資料集；以及

轉換電路，包括機器學習電路，且配置成：

基於所述寫入記憶體存取，將所述資料集的原始版本轉變為所述變換的資料集，以及

基於所述讀取記憶體存取，將所述變換的資料集轉變為重構的資料集，所述重構的資料集包括不同於所述資料集的所述資料集的近似；

其中所述至少一個記憶體儲存電路配置成儲存所述機器學習電路的永久狀態，所述機器學習電路用以將所述資料集的所述原始版本轉變為所述變換的資料集；以及

其中所述機器學習電路的所述永久狀態包括將所述變換的資料集轉變為所述重構的資料集的配置。

【第15項】 如申請專利範圍第14項所述的儲存系統，

其中所述機器學習電路包括第一機器學習電路，所述轉換電路包括第二機器學習電路，且其中所述轉換電路配置成至少部分地基於由所選擇的機器學習電路提供的保真度的量來選擇所述第一機器學習電路及所述第二機器學習電路中的一個以轉變所述資料集；

其中所述變換的資料集具有小於所述資料集的所述原始版本的大小。

【第16項】 如申請專利範圍第14項所述的儲存系統，其中所述機器學習電路包括第一機器學習電路，其中所述轉換電路包括第二機器學習電路，並且其中，基於讀取記憶體存取，所述儲存裝置

配置成將所述變換的資料集返回到所述主機計算裝置，其中所述變換的資料集在大小上比所述資料集的所述原始版本更小；

其中所述轉換電路包含直接儲存電路，所述直接儲存電路配置成創建與所述資料集的所述原始版本相等的所述變換的資料集的版本，且創建與所述資料集的所述原始版本相等的重構的資料集的版本；以及

其中所述轉換電路配置成基於保真度要求，在所述第一機器學習電路及所述第二機器學習電路中的一個之間進行選擇以處理所述變換的資料集。

【第17項】 如申請專利範圍第14項所述的儲存系統，其中所述至少一個記憶體儲存電路包括：

使包含在記憶體存取中的定址值與所述變換的資料集相關聯的欄位，以及指示所述機器學習電路用於創建所述變換的資料集的欄位。

【第18項】 如申請專利範圍第14項所述的儲存系統，其中所述機器學習電路包含一或多個神經網路；以及

其中所述轉換電路配置成至少部分地基於保真度要求，調整在所述神經網路中的層的數目。

【第19項】 一種儲存方法，包括：

接收原始資料集以儲存在至少一個記憶體電路中；

經由機器學習技術將所述原始資料集轉變成變換的資料集，其中所述變換的資料集小於所述原始資料集，且其中自所述變換

的資料集創建的重構的資料集包括不同於資料集的所述資料集的近似；以及

儲存所述變換的資料集在所述至少一個記憶體電路中；以及

儲存所述機器學習技術的永久狀態在所述至少一個記憶體電路中，所述機器學習技術用以將所述原始資料集轉變為所述變換的資料集；

其中所述機器學習技術的所述永久狀態包括自所述變換的資料集創建所述重構的資料集的配置。

【第20項】 如申請專利範圍第19項所述的儲存方法，更包括

自主機裝置接收請求以自所述至少一個記憶體電路中檢索所述原始資料集；

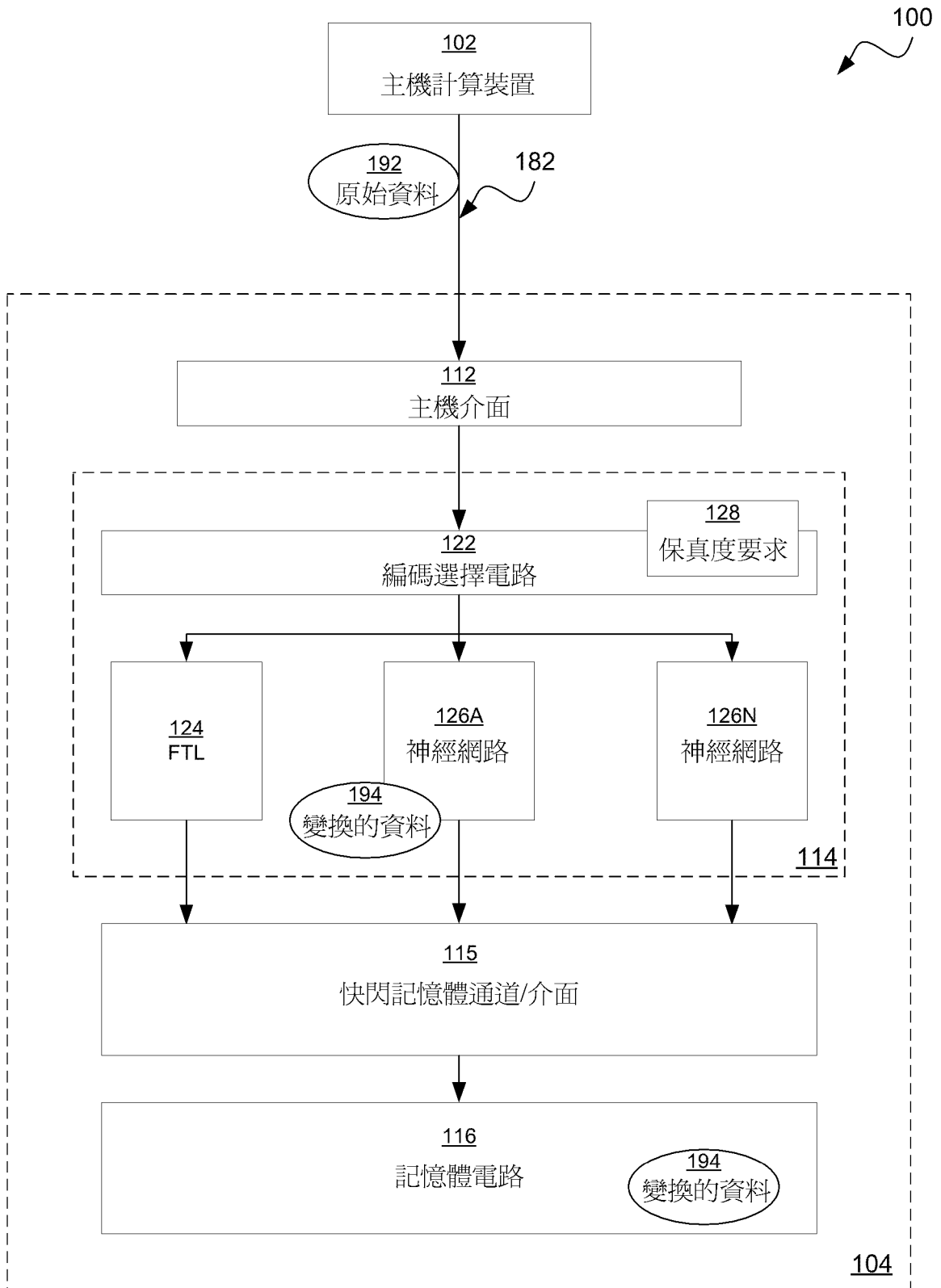
自所述至少一個記憶體電路中檢索所述變換的資料集；

自所述至少一個記憶體電路中檢索所述機器學習技術的所述永久狀態；

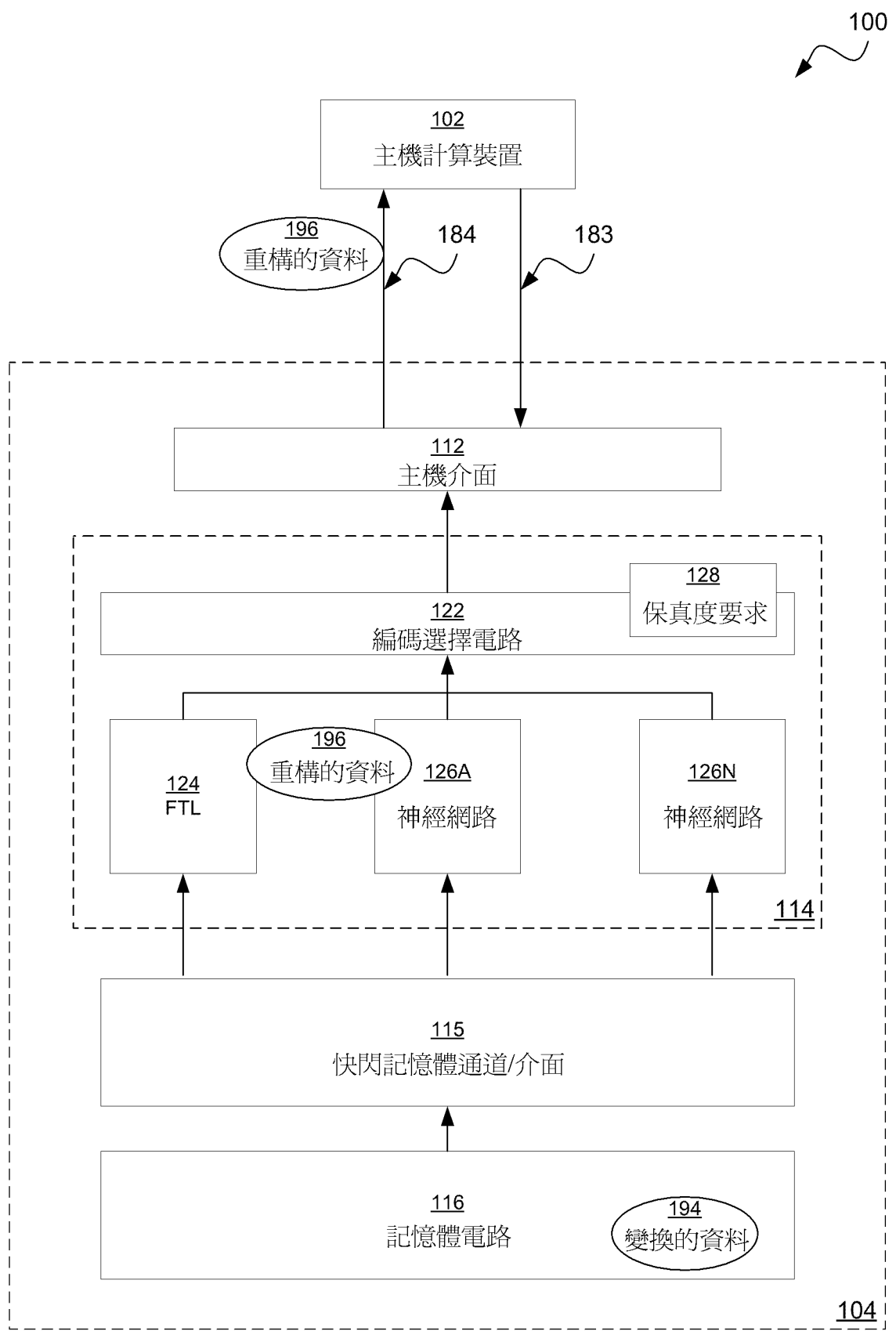
通過所述機器學習技術將所述變換的資料集轉變為所述重構的資料集；以及

將所述重構的資料集返回至所述主機裝置。

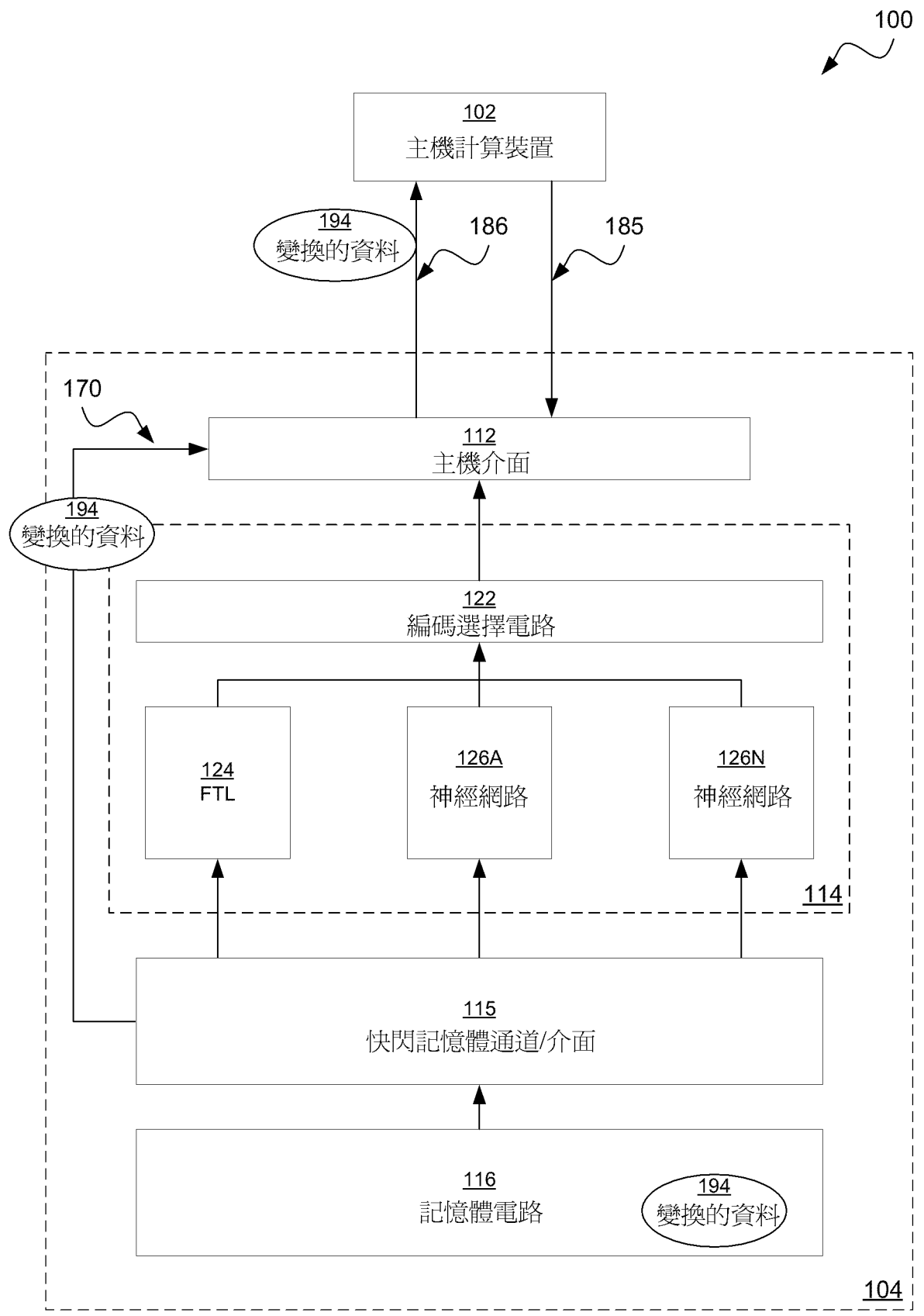
【發明圖式】



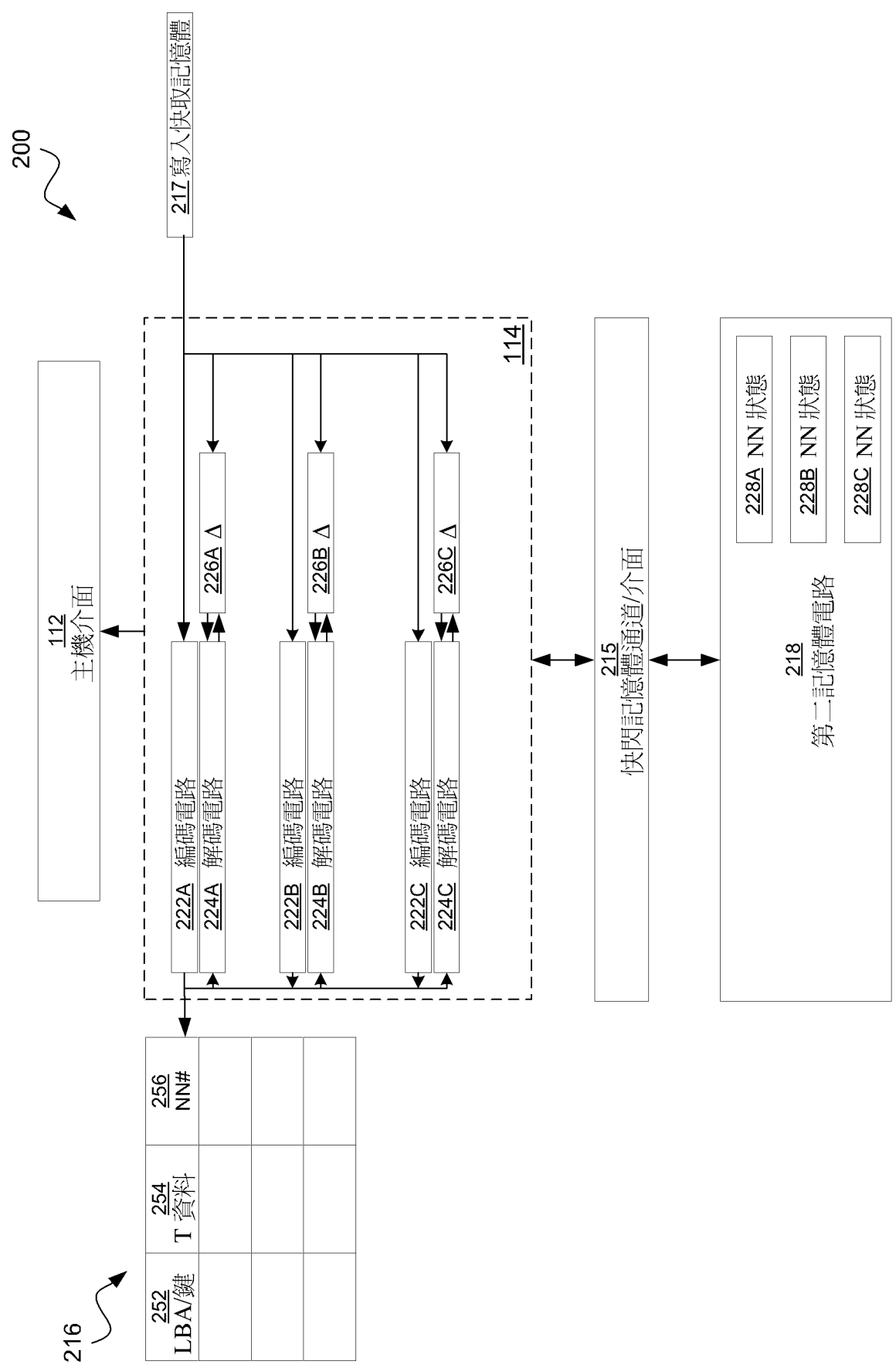
【圖 1A】



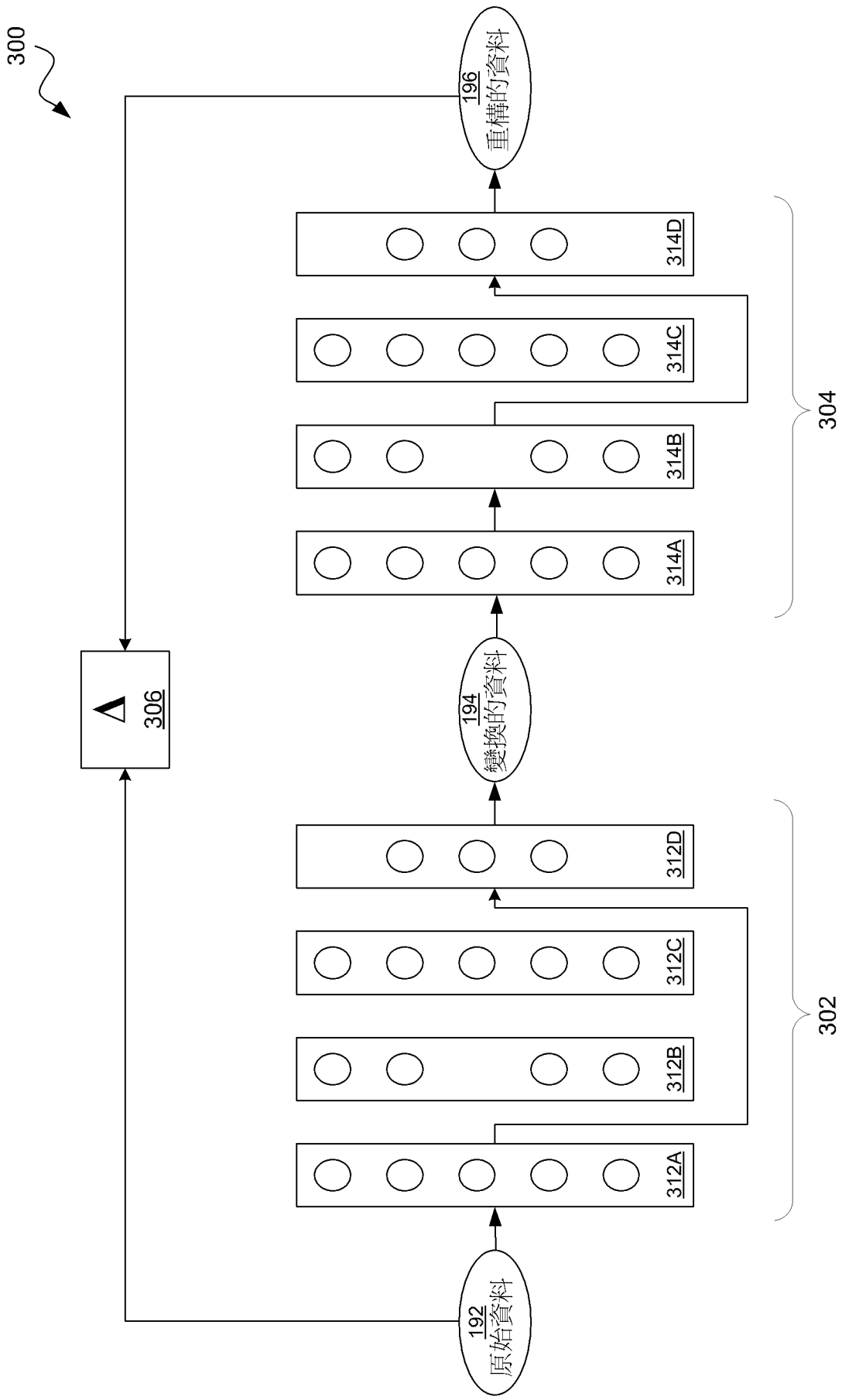
【圖 1B】



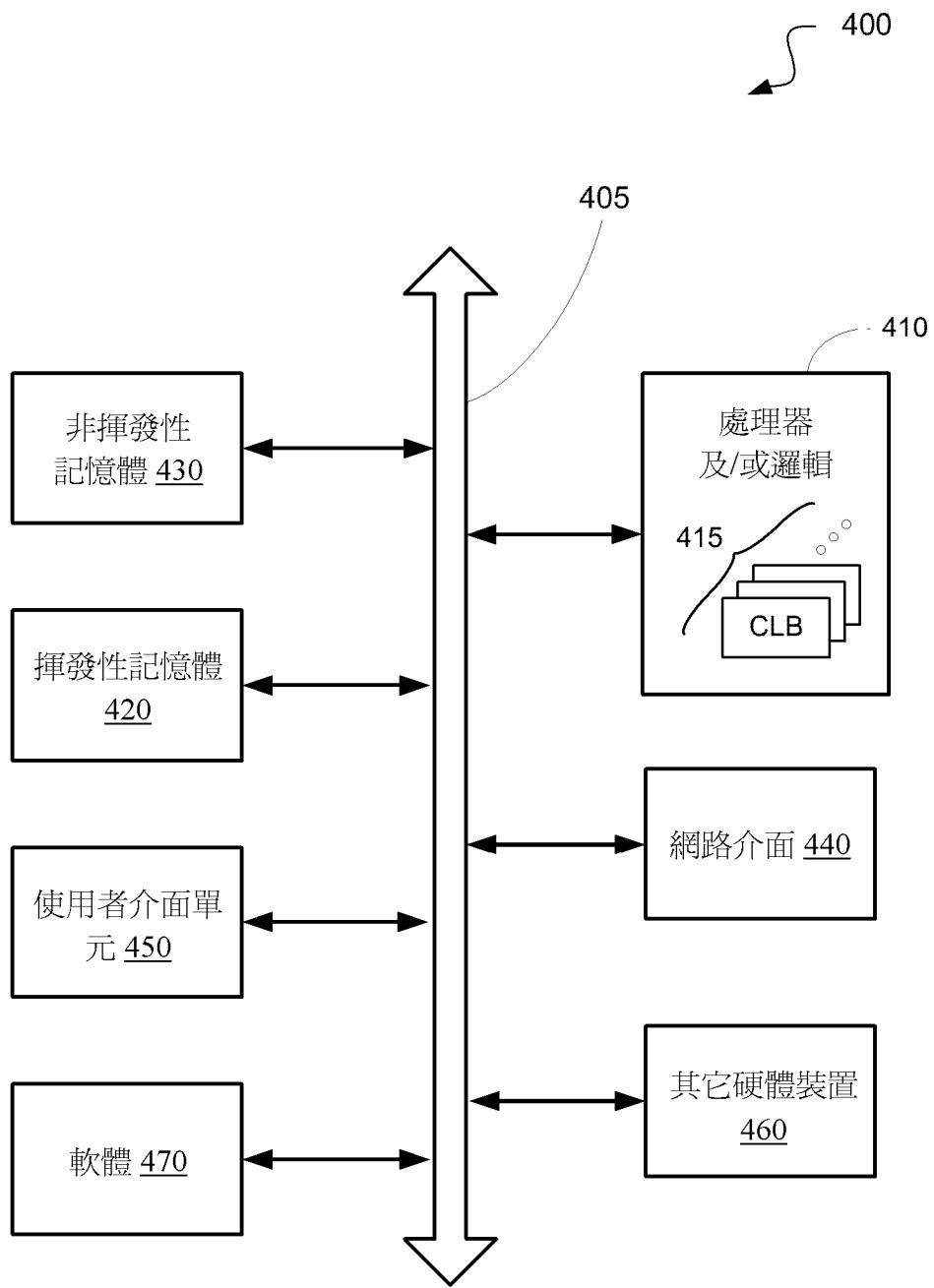
【圖 1C】



【圖 2】



【圖 3】



【圖 4】