



(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 201421240 A

(43)公開日：中華民國 103 (2014) 年 06 月 01 日

(21)申請案號：103105076

(22)申請日：中華民國 96 (2007) 年 05 月 30 日

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

G06F13/00 (2006.01)

(30)優先權：2006/06/08 美國

11/450,023

(71)申請人：畢特微網路工作公司 (美國) BITMICRO NETWORKS, INC. (US)
美國

(72)發明人：布魯斯 瑞 H BRUCE, REY H. (US)；麥特羅 挪門 帕斯 MATEO, NOEME PAZ
(PH)；尼特 瑞齊 塞服拉 NITE, RICKY SEVILLA (PH)

(74)代理人：洪武雄；陳昭誠

申請實體審查：有 申請專利範圍項數：18 項 圖式數：19 共 62 頁

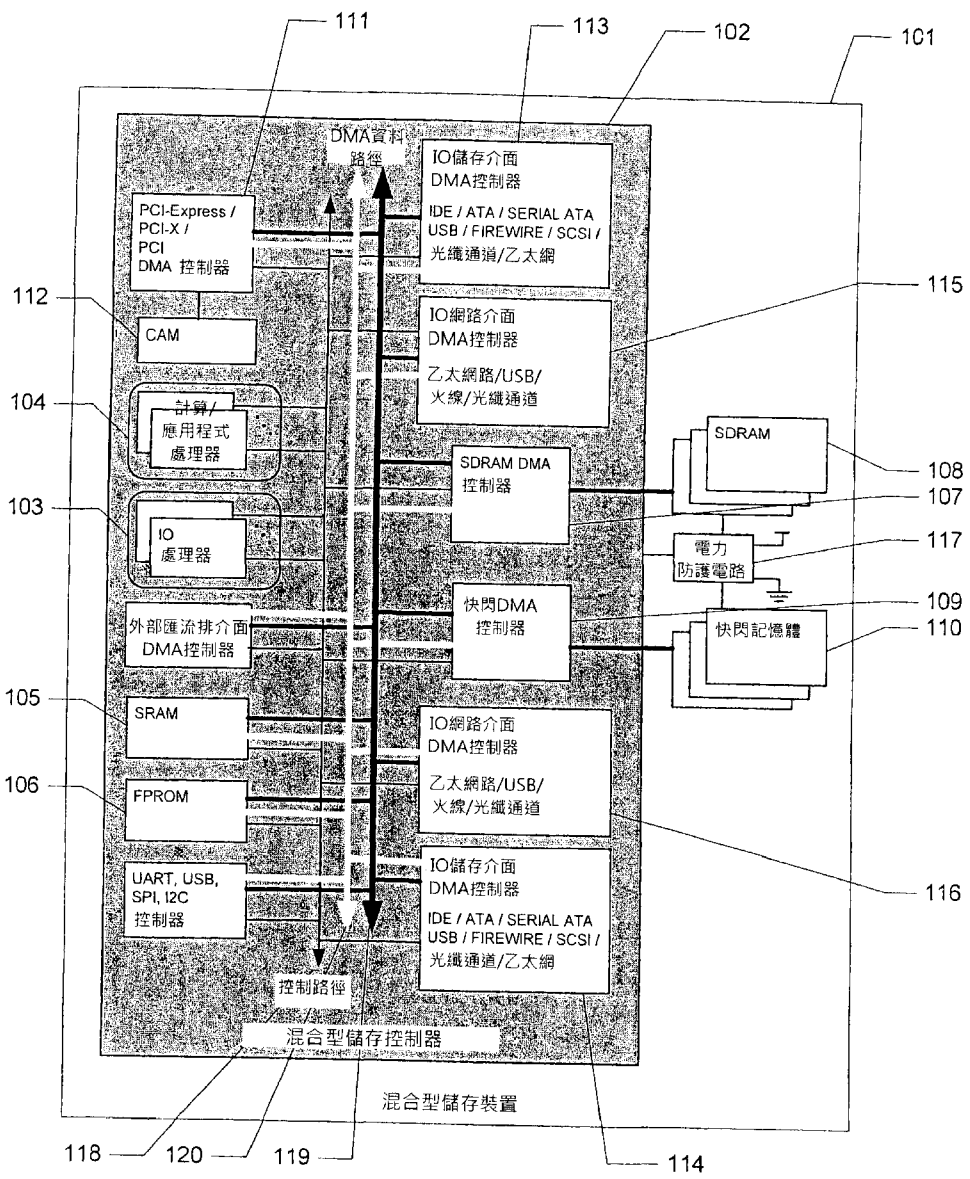
(54)名稱

資料儲存系統、資料結構及資料儲存方法

DATA STORAGE SYSTEM, DATA STRUCTURE AND DATA STORAGE METHOD

(57)摘要

本發明係揭露一種混合型儲存系統，包括機械式磁碟機機制(mechanical disk drive means)、快閃記憶體機制、SDRAM 記憶體機制以及 SRAM 記憶體機制。設計出 IO 處理器機制和 DMA 控制器機制，用以消除主機干預(intervention)。多層快取系統以及用於將邏輯位址映射到實體位址的新穎資料結構造成可組構及可縮放尺寸之高效能電腦資料儲存解決方案。



第1圖

- 101：混合型儲存裝置
- 102：混合型儲存控制器
- 103：IO 處理器
- 104：計算/應用程式處理器
- 105：SRAM
- 106：場可程式化ROM
- 107：SDRAM 控制器
- 108：SDRAM
- 109：快閃 DMA 控制器
- 110：快閃裝置、快閃陣列、快閃
- 111：PCI-Express/PCI-X/PCI DMA 控制器
- 112：內容可定址記憶體
- 113：IO 儲存介面 DMA 控制器
- 114：IO 儲存介面 DMA 控制器
- 115：IO 網路介面 DMA 控制器
- 116：IO 網路介面 DMA 控制器
- 117：電力防護電路
- 118：控制路徑
- 119：DMA 路徑
- 120：DMA 路徑



(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 201421240 A

(43)公開日：中華民國 103 (2014) 年 06 月 01 日

(21)申請案號：103105076 (22)申請日：中華民國 96 (2007) 年 05 月 30 日
(51)Int. Cl. : G06F12/02 (2006.01) G06F13/00 (2006.01)
(30)優先權：2006/06/08 美國 11/450,023
(71)申請人：畢特微網路工作公司 (美國) BITMICRO NETWORKS, INC. (US)
美國
(72)發明人：布魯斯 瑞 H BRUCE, REY H. (US)；麥特羅 挪門 帕斯 MATEO, NOEME PAZ
(PH)；尼特 瑞齊 塞服拉 NITE, RICKY SEVILLA (PH)
(74)代理人：洪武雄；陳昭誠
申請實體審查：有 申請專利範圍項數：18 項 圖式數：19 共 62 頁

(54)名稱

資料儲存系統、資料結構及資料儲存方法

DATA STORAGE SYSTEM, DATA STRUCTURE AND DATA STORAGE METHOD

(57)摘要

本發明係揭露一種混合型儲存系統，包括機械式磁碟機機制(mechanical disk drive means)、快閃記憶體機制、SDRAM 記憶體機制以及 SRAM 記憶體機制。設計出 IO 處理器機制和 DMA 控制器機制，用以消除主機干預(intervention)。多層快取系統以及用於將邏輯位址映射到實體位址的新穎資料結構造成可組構及可縮放尺寸之高效能電腦資料儲存解決方案。

發明摘要

※ 申請案號：

※ 申請日：

IPC 分類：

E06F 13/00 (2006.01)

【發明名稱】(中文/英文)

資料儲存系統、資料結構及資料儲存方法

DATA STORAGE SYSTEM, DATA STRUCTURE AND DATA
STORAGE METHOD

【中文】

本發明係揭露一種混合型儲存系統，包括機械式磁碟機機制 (mechanical disk drive means)、快閃記憶體機制、SDRAM 記憶體機制以及 SRAM 記憶體機制。設計出 IO 處理器機制和 DMA 控制器機制，用以消除主機干預(intervention)。多層快取系統以及用於將邏輯位址映射到實體位址的新穎資料結構造成可組構及可縮放尺寸之高效能電腦資料儲存解決方案。

【英文】

A hybrid storage system comprising mechanical disk drive means, flash memory means, SDRAM memory means, and SRAM memory means is described. IO processor means and DMA controller means are devised to eliminate host intervention. Multi-tiered caching system and novel data structure for mapping logical address to physical address results in a configurable and scalable high performance computer data storage solution.

【代表圖】

【本案指定代表圖】：第（ 1 ）圖。

【本代表圖之符號簡單說明】：

101	混合型儲存裝置	102	混合型儲存控制器
103	IO 處理器	104	計算/應用程式處理器
105	SRAM	106	場可程式化 ROM
107	SDRAM 控制器	108	SDRAM
109	快閃 DMA 控制器		
110	快閃裝置、快閃陣列、快閃		
111	PCI-Express/PCI-X/PCI DMA 控制器		
112	內容可定址記憶體		
113、114	IO 儲存介面 DMA 控制器		
115、116	IO 網路介面 DMA 控制器		
117	電力防護電路	118	控制路徑
119、120	DMA 路徑		

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

本案無化學式

發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】(中文/英文)

資料儲存系統、資料結構及資料儲存方法

DATA STORAGE SYSTEM, DATA STRUCTURE AND DATA
STORAGE METHOD

【技術領域】

【0001】本發明有關於一種應用於電腦系統之資料儲存系統，該資料儲存系統包括揮發性(例如 SRAM、SDRAM)和非揮發性(例如快閃記憶體、機械式硬碟)的儲存組件。

【先前技術】

【0002】在傳統的電腦系統中，硬碟機(hard disk drive; HDD)是作為一種外部記憶體裝置，其中，磁碟是作為一種儲存媒介。HDD 可以作為是一種大容量檔案設備。然而，與包括了半導體記憶體(例如 DRAM)的主記憶體相比，HDD 的存取速度較慢。用於 HDD 的快取系統已為人所熟知是用於加快 HDD 之存取速度的一種機制。動態隨機存取記憶體(DRAM)以及快閃記憶體已被用來實作針對 HDD 的快取系統。然而，由邏輯位址到適用於存取快閃記憶體以及 HDD 之實體位址格式的轉換消耗了主機大量的資源並影響效能。因此有混合型儲存系統的需要，其中，經由消除主機干預而改進了效能。

【0003】企業級(enterprise-level)的儲存系統典型地會使用硬碟機(HDD)的陣列來作為大量儲存單元，或是組構成獨立磁碟備援陣列(RAID)系統。資料使用者或是客戶使用標準的基於區塊的

IO 介面或是在網路上使用標準的基於檔案的存取協定來存取資料。HDD 陣列資料是由運行儲存管理應用程式的專用主機來加以管理。由於在主機以及用戶端系統中使用了數個介面控制器，企業級儲存系統可由整合對資料的基於區塊的存取以及基於檔案或是隨機的存取的控制器架構中獲利。在前述針對 HDD 實作多層快取系統之混合型儲存系統中，用於不同介面之資料傳遞控制器的整合將延伸 HDD 資料快取的優點到 HDD 陣列系統。

【發明內容】

【0004】本發明係揭露一種混合型儲存裝置，包括用於資料儲存之機械式磁碟機機制、快閃記憶體機制、SDRAM 記憶體機制以及 SRAM 記憶體機制。設計出 IO 處理器機制以及多個 DMA 控制器機制以消除主機干預。多層快取系統以及用於將邏輯位址映射(mapping)到實體位址的新穎資料結構造成了可組構及可縮放尺寸的高效能電腦資料儲存解決方案。

【0005】LBA_Flash_HDD 表具有用於將邏輯位址映射到快閃位址的第一部分、以及用於將邏輯位址映射到磁碟機位址的第二部分。永久的副本(copy)是儲存在快閃記憶體中，最常用的部份是儲存在 SRAM 內，而剩餘部份則儲存在 SDRAM 中，最常用的部份和剩餘部份共同形成了工作副本(working copy)。備份副本是儲存在機械式磁碟機中。LBA_SDRAM 表係用來將邏輯位址映射到 SRAM 和 SDRAM 位址。這是儲存在 SDRAM 中並在 SRAM 中快取。數個 DMA 控制器是提供用以在多層儲存裝置之間移動資料。IO 處理器藉由以鏈結列表格式準備 DMA 指令來實作用於 DMA 的一致方法(uniform method)。

【0006】在本發明的實施例中，IO 處理器、DMA 控制器以及所有必須的控制功能均整合在晶片上系統(SOC)裝置中。提供了有八種組構的儲存解決方案。在本發明的第一個範例實施例中，儲存系統 SOC 裝置是組構成爲從屬裝置(slave device)，並經由可隨機存取的系統匯流排以及諸如是 PCI/PCI-X/PCI Express 之 DMA 來和主機系統介接(interface)，並同時透過諸如是 ATA 或 SCSI 之標準 IO 儲存介面來和機械式磁碟機介接。

【0007】在本發明的第二個範例實施例中，該儲存系統 SOC 裝置是組構成主機系統，其係經由可隨機存取的系統匯流排以及諸如是 PCI/PCI-X/PCI Express 之 DMA 來和外部儲存裝置介接。

【0008】在本發明的第三個範例實施例中，該儲存系統 SOC 裝置是組構成獨立的主機系統，其係透過諸如是 SCSI 之標準 IO 儲存介面來和機械式磁碟機介接，並且亦透過諸如是乙太網路之標準 IO 網路介面來和網路介接。

【0009】在本發明的第四個範例實施例中，該儲存系統 SOC 裝置是組構成從屬裝置，且係透過諸如是光纖通道的內部標準 IO 來和主機系統介接，且透過諸如是通用序列匯流排(USB)之第二標準 IO 介面來和機械式磁碟機介接。

【0010】在本發明的第五個範例實施例中，該儲存系統 SOC 裝置是組構成主機系統，其係透過諸如是光纖通道的外部標準 IO 儲存介面來和外部儲存裝置介接，且亦透過諸如是乙太網路之標準 IO 網路介面來和網路介接。

【0011】在本發明的第六個範例實施例中，該儲存系統 SOC 裝置是組構成從屬裝置，且係透過諸如是光纖通道的外部標準 IO

介面來和主機系統介接，並透過諸如是 USB 之第二標準 IO 介面來和機械式磁碟機介接。

【0012】在本發明的第七個範例實施例中，該儲存系統 SOC 裝置是組構成為主機系統，其係透過諸如是光纖通道的外部標準 IO 儲存介面來和外部儲存裝置介接，且亦透過諸如是乙太網路之標準 IO 網路介面來和網路介接。

【0013】在本發明的第八個範例實施例中，該儲存系統 SOC 裝置是組構成從屬裝置，且在單一的包封體(enclosure)中整合到機械式磁碟機中，並透過諸如是 SCSI 之標準 IO 介面來和主機系統介接，且透過低階直接連結(low level direct connection)來和磁性磁碟控制器介接。

【圖式簡單說明】

【0014】藉由參考圖示於附圖之實施例，可更特定敘述如發明內容中所概要之本發明，而能詳細了解並達到本發明之上述特徵、優點及目的。

【0015】然而，應注意的是，所附圖式僅例示本發明典型的實施例，並且不應視為是其範疇的限制，因本發明可加入其它等效的實施例。

第 1 圖為顯示根據本發明實施例之組成混合型儲存裝置的組件的圖；

第 2 圖為顯示本發明實施例之範例組構的圖，其中，該儲存系統 SOC 裝置是組構成從屬裝置，且係透過可以隨機存取的系統匯流排以及諸如是 PCI/PCI-X/PCI Express 之 DMA 來和主機系統介接，且亦透過諸如是 ATA 或是 SCSI 之標準 IO 儲存介面來和機械

式磁碟機介接；

第 3 圖為顯示本發明實施例之第二個範例組構的圖，其中，該儲存系統 SOC 裝置是組構成主機系統，其係透過可隨機存取的系統匯流排以及諸如是 PCI/PCI-X/PCI Express 之 DMA 來和外部儲存裝置介接；

第 4 圖為顯示本發明實施例之第三個範例組構的圖，其中，該儲存系統 SOC 是組構成獨立的主機系統，其係透過諸如是 SCSI 之標準 IO 儲存介面來和機械式磁碟機介接，且亦透過諸如是乙太網路之標準 IO 網路介面來和網路介接；

第 5 圖為顯示本發明實施例之第四個範例組構的圖，其中，該儲存系統 SOC 裝置是組構成從屬裝置，且係透過諸如是光纖通道的內部標準 IO 來和主機系統介接，且透過諸如是 USB 之第二標準 IO 介面來和機械式磁碟機介接；

第 6 圖為顯示本發明實施例之第五個範例組構的圖，其中，該儲存系統 SOC 裝置是組構成主機裝置，其係透過諸如是光纖通道的外部標準 IO 儲存介面來和外部儲存裝置介接，且透過諸如是乙太網路之標準 IO 網路介面來和網路介接；

第 7 圖為顯示本發明實施例之第六個範例組構的圖，其中，該儲存系統 SOC 裝置是組構成從屬裝置，且係透過諸如是光纖通道的外部標準 IO 介面來和主機系統介接，並透過諸如是 USB 之第二標準 IO 介面來和機械式磁碟機介接；

第 8 圖為顯示本發明實施例之第七個範例組構的圖，其中，該儲存系統 SOC 裝置是組構成主機系統，其係透過諸如是光纖通道的內部標準 IO 儲存介面來和外部儲存裝置介接，且亦透過諸如

是乙太網路之標準 IO 網路介面來和網路介接；

第 9 圖為顯示本發明實施例之第八個範例組構的圖，其中，該儲存系統 SOC 裝置是組構成從屬裝置，且在單一的包封體中整合到機械式磁碟機中，並透過諸如是 SCSI 之標準 IO 介面來和主機系統介接，且透過低階直接連結來和磁性磁碟控制器介接；

第 10a 圖為顯示在根據本發明實施例之非揮發性及揮發性儲存組件內之資料結構的圖；

第 10b 圖為顯示根據本發明實施例之用於 LBA-Flash-HDD 映射表的範例資料結構的圖；

第 10c 圖為顯示根據本發明實施例之用於 LBA-SDRAM 映射表的範例資料結構的圖；

第 11 圖為顯示根據本發明實施例之啓動初始化過程流程圖；

第 12 圖為顯示根據本發明實施例之來自主機電腦系統的區塊讀取命令的流程圖；

第 13 圖為顯示根據本發明實施例之來自主機電腦系統的區塊寫入命令的流程圖；

第 14 圖為顯示根據本發明實施例之對混合型儲存裝置的隨機位元組-定址讀取存取的流程圖；

第 15 圖為顯示根據本發明實施例之對混合型儲存裝置的隨機位元組-定址寫入存取的流程圖；

第 16 圖為顯示根據本發明實施例之 DMA 傳送的流程圖；

第 17 圖為顯示根據本發明實施例之回應讀取要求而更新映射表的流程圖；

第 18 圖為顯示根據本發明實施例之回應寫入要求而更新映

射表的流程圖；以及

第 19 圖為顯示根據本發明實施例之回應寫入要求或是啓動來作為背景程序而在資料清理操作期間更新映射表的流程圖。

【實施方式】

【0016】第 1 圖為顯示根據本發明實施例之組成混合型儲存裝置 101 的組件的圖。

【0017】混合型儲存裝置 101 包括數個在下列所列用以增加儲存容量和減少存取時間的儲存裝置，依序為-嵌入式 SRAM 105、SDRAM 裝置陣列 108、快閃裝置陣列 110 以及硬碟機裝置陣列(未顯示)。在該儲存系統內是實施三階快取：快閃陣列快取硬碟機內的資料、SDRAM 陣列快取快閃陣列內的資料、SRAM 快取 SDRAM 陣列內的資料。該主要的非揮發性儲存組件包括一個或多個硬碟(未顯示)。

【0018】混合型儲存控制器 102 是一種管理該儲存系統的晶片，其含有多個嵌入式 DMA 控制器：

【0019】PCI-Express/PCI-X/PCI DMA 控制器 111 藉由透過諸如是 PCI-Express、PCI-X、PCI 介面之系統匯流排而連結的任何裝置來處理對儲存資料之位元組或是字組定址(word addressable)存取。內容可定址記憶體(content addressable memory; CAM)112 儲存了一種由 PCI-Express/PCI-X/PCI 介面 DMA 控制器所使用的查詢表(look-up table)，以查詢與位元組或字元位址相關聯的區塊位址。

【0020】IO 儲存 DMA 控制器 113 以及 IO 儲存 DMA 控制器 114 透過諸如是整合式驅動電子/進階技術附加(IDE/ATA)、序列 ATA(Serial ATA)、USB、SCSI 等標準區塊存取 IO 介面來處理

DMA。該 IO DMA 控制器可透過 IO 介面來和主機電腦系統相連結，其亦可用以控制硬碟的陣列。

【0021】IO 網路介面 DMA 控制器 115 以及 IO 網路介面 DMA 控制器 116 處理對諸如是 ETHERNET、USB、FIREWIRE、FIBER CHANNEL 之網路介面的 DMA。

【0022】快閃 DMA 控制器 109 處理對快閃陣列 110 的 DMA。

【0023】SDRAM 控制器 107 處理對 SDRAM 陣列 108 的 DMA。

【0024】混合型儲存控制器 102 含有 IO 處理器 103，IO 處理器 103 可使用一個或多個嵌入式處理器來加以實施。該 IO 處理器處理主機命令(讀取/寫入)的程序，並進行用以管理不同的儲存媒介之演算法。實作快取演算以及諸如是翻譯表(translation table)之控制結構的維護對使用或是連結到諸如是主機電腦系統的儲存系統之外界實體而言是顯而易見的。

【0025】一個或多個額外的嵌入式處理器 114 可作用為在諸如是 Windows、Linux 等上運行之傳統 O/S 的電腦/應用處理器。SDRAM 105 可在 IO 處理器以及應用處理器之間分享。提供了數個 DMA 路徑 119、120 以避免資料瓶頸(data bottleneck)。不同的儲存媒介能以不同的方式來和 DMA 路徑相連結，以達到最佳化的流量分佈(traffic distribution)。例如，該 DMA 路徑 120 中的至少一個可作用為是在 SRAM 105 和快閃陣列 110 之間的專屬路徑。在讀取快閃陣列內的資料時，若是在其它路徑的流量很大，則此 DMA 路徑 119 可作用為將資料由快閃陣列 110 傳送到在 SRAM 105 內的暫存緩衝器。此外，個別的控制路徑 118 是提供給嵌入式處理器，以存取不同 DMA 控制器的登錄介面(register interface)，以

減少在高速 DMA 路徑 119、120 上的控制負荷(overhead)。可使用一種場可程式化(Field-Programmable)ROM 106 來儲存用於 IO 處理器的啓動碼(Boot Code)。

【0026】在正常操作期間，資料以及控制資訊是分佈在儲存組件之中，如第 10a 圖所示，且在此敘述中將會在稍後詳細說明。電力防護電路(power guard)117 確保了在快閃 110、SDRAM 108 內以及在混合型儲存控制器 102 內所有組件的資料在失去電力時均受到保護。在揮發性 SDRAM 以及 SRAM 內的資料將被清除至快閃中。因此，快閃保留所有的快取資料以及控制資訊。電力防護電路保護著在混合型儲存控制器 101 內所有的組件，包括該嵌入式處理器，以及所有的嵌入式 FIFO、內部 RAM。在失去電力時，在這些組件內持續處理所保留的資料。來自連結於該混合型儲存裝置之外部實體的暫時性資料將不會被接受。

【0027】IO 處理器 103 對 IO 儲存介面 DMA 控制器 113 和 114、IO 網路介面 DMA 控制器 115 和 116、PCI-Express/PCI-X/PCI DMA 控制器 111 和快閃 DMA 控制器 109 發送指令，以在 SRAM 105 或是 SDRAM 108 以及其個別的介面之間傳遞資料。因為該儲存系統被管理著，以致於由該 DMA 控制器傳遞到其個別介面的資料總是在 SRAM 105 或 SDRAM 108 內進行快取，SRAM 105 或 SDRAM 108 與快閃 110 以及硬碟機(未顯示)相比之下提供較快速的存取。IO 處理器 103 在其指令中含有諸如是下列的資訊-資料傳遞的方向、來源以及目的位址、預備傳送之資料的大小、以及所有其它介面-特定的控制資訊。這些指令是儲存到 SRAM 105 或是 SDRAM 108。每一指令含有到下一個指令的鏈結(link)。因此，在

IO 處理器 103 在透過控制匯流排而將初始指令發送到任何 DMA 控制器後，該 DMA 控制器可自動地由 SRAM 105 或是 SDRAM 108 提取(fetch)下一個指令。該 DMA 控制器隨後則通知 IO 處理器 103 資料已傳送完畢。執行 DMA 傳送的程序流程是在第 16 圖中所顯示，且在稍後於此說明內將會予以詳細說明。

【0028】第 2 圖為顯示本發明實施例之範例組構的圖。在此組構中，混合型儲存控制器 201 是組構成從屬裝置，並透過可隨機存取的系統匯流排以及諸如是 PCI/PCI-X/PCI Express 之 DMA 來和主機系統 202 介接。該混合型儲存裝置透過諸如是序列 ATA 之標準 IO 儲存介面來控制硬碟機 204 的陣列。

【0029】第 3 圖顯示的是本發明實施例的第二個範例組構的圖。其顯示的是混合型儲存控制器 301 亦可組構成爲一種主機系統，其係透過諸如是 PCI/PCI-X/PCI Express 的系統匯流排 303 來控制從屬裝置 302。在圖式中，從屬裝置 302 事實上是爲硬碟陣列控制器。由於混合型儲存裝置本身可組構成作用爲具有至系統匯流排之介面的硬碟陣列控制器，因此兩個混合型儲存裝置可使用其 PCI/PCI-X/PCI Express DMA 控制器來互連，其中一個爲另一個的從屬裝置。

【0030】第 4 圖為顯示本發明實施例的第三個範例組構的圖。其顯示著混合型儲存控制器 401 是組構成獨立的主機系統，其係透過諸如是 SCSI 之標準 IO 儲存介面來和硬碟機陣列介接，而且同時也透過諸如是乙太網路之標準 IO 網路介面來和網路介接。IO 儲存介面 DMA 控制器 402 處理對硬碟陣列 404 的 DMA。IO 儲存介面 DMA 控制器 403 處理對硬碟陣列 405 的 DMA。IO 網

路介面 DMA 控制器 406 處理對網路 408 的連結。IO 網路介面 DMA 控制器 407 處理對網路 409 的連結。嵌入式 IO 處理器 410 協調該等 DMA 控制器的運作。由於該混合型儲存裝置作用成獨立的系統，因此其它軟體的應用程式亦可在額外的嵌入式計算/應用處理器 411 上運行。

【0031】第 5 圖為顯示本發明實施例的第四個範例組構的圖。混合型儲存控制器 501 是組構成從屬裝置，且係透過諸如是光纖通道之標準資料區塊存取 IO 匯流排(Block-access IO bus)503 來和主機系統 502 介接，且透過諸如是 IDE 之第二標準 IO 介面來和硬碟機介接。在此情形下之該混合型儲存裝置使用嵌入於混合型儲存控制器 501 的內部 IO 儲存介面 DMA 控制器 504 和 505 來處理與主機系統 502 和硬碟機兩者的介接。

【0032】第 6 圖為顯示本發明實施例的第五個範例組構的圖。混合型儲存控制器 601 是組構成主機系統，其係透過外部標準 IO 介面控制器 603 來和外部儲存裝置 602 介接，且外部標準 IO 介面控制器 603 係透過諸如是光纖通道 604 來和外部儲存裝置 602 連結。該外部儲存裝置亦可是另一個混合型儲存裝置。該混合型儲存控制器也透過諸如是乙太網路 605 之標準 IO 網路介面來和網路介接。該混合型儲存控制器使用 PCI-Express/PCI-X/PCI DMA 控制器 606 來組構並且控制該外部 IO 介面控制器的操作，並且透過該外部 IO 介面控制器來將 IO 命令、資料以及狀態資訊傳送至該外部儲存裝置及從該外部儲存裝置傳送出來。該 PCI-Express/PCI-X/PCI DMA 控制器具有主要(master)以及從屬介面，以致於不論是混合型儲存控制器或是外部 IO 控制器均可啟動

DMA 交易(DMA transaction)。要將 IO 命令傳送到外部儲存裝置(其中該外部 IO 控制器是作為 DMA 從屬(DMA slave))，該混合型儲存控制器使用混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 主要介面(DMA master interface)605 來將 IO 命令寫入至該外部 IO 控制器。該外部 IO 控制器在 IO 匯流排建立了命令階段(command phase)來傳送 IO 命令。要將 IO 命令傳送到外部儲存裝置，其中該外部 IO 控制器是作為 DMA 主要(DMA master)，該混合型儲存控制器將該 IO 命令緩衝器在 SRAM 或 SDRAM 內之位置指示給該外部 IO 控制器。該外部 IO 控制器使用混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 從屬介面由該混合型儲存控制器讀取 IO 命令，並在 IO 匯流排上建立命令階段，以將 IO 命令傳送到外部儲存裝置。為了將資料傳送到外部儲存裝置或是由外部儲存裝置傳送資料，其中該外部 IO 控制器是作為 DMA 從屬，該混合型儲存控制器於該 IO 資料階段係建立在該 IO 匯流排時是使用該混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 主要介面讀取從該外部 IO 控制器來之資料或將資料寫入至該外部 IO 控制器。將資料傳送到外部儲存裝置或是由外部儲存裝置傳送出來，其中該外部 IO 控制器為 DMA 主要，該混合型儲存控制器將在 SRAM 或是 SDRAM 內之資料快取緩衝器的位置指示給外部 IO 控制器，以致於當該 IO 資料階段係建立在該 IO 匯流排上時，該外部 IO 控制器可使用混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 從屬介面來將資料寫入至該混合型儲存控制器或是讀取從該混合型儲存控制器來之資料。該資料是寫入至或讀取自 SRAM 或是 SDRAM 內的資料快取。為了能由該外部儲存裝置接收 IO 狀態資訊，真中該外

部 IO 控制器是作為 DMA 主要，該外部 IO 控制器於狀態階段在該 IO 匯流排上完成時，會中斷該混合型儲存控制器，以致於該混合型儲存控制器可以透過該混合型儲存控制器 PCI/Express/PCI-X/PCI DMA 主要介面從外部 IO 控制器讀取收到的 IO 狀態資訊，並且將其傳遞到位在 SRAM 707 或是 SDRAM 708 內的 IO 狀態緩衝器。為了從外部儲存裝置收到該 IO 狀態資訊，其中該外部 IO 控制器是作為 DMA 主要，該混合型儲存控制器初期會將該 IO 狀態緩衝器在 SRAM 或是 SDRAM 內之位置指示給該外部 IO 控制器，所以當狀態階段在該 IO 匯流排上完成時，該外部 IO 控制器可將所收到的 IO 狀態資訊透過該混合型儲存控制器 PCI/Express/PCI-X/PCI DMA 從屬介面寫入至位在混合型儲存控制器內的 IO 狀態緩衝器。

【0033】第 7 圖為顯示本發明實施例的第七個範例組構的圖。混合型儲存控制器 701 是組構成從屬裝置，並且透過使用諸如是光纖通道之標準 IO 介面的外部 IO 介面控制器 703 來和主機系統 702 介接，以及透過諸如是序列連接(serial attached)SCSI 以及序列 ATA 704、705 之內部標準 IO 介面來和機械式磁碟機介接。該混合型儲存控制器使用 PCI-Express/PCI-X/PCI DMA 控制器 706 來組構以及控制外部 IO 介面控制器的操作，並透過該外部 IO 介面控制器將 IO 命令、資料和狀態資訊傳送至主機系統以及從主機系統傳送出來。該 PCI-Express/PCI-X/PCI DMA 控制器具有主要以及從屬介面，以致於不論是混合型儲存控制器或是外部 IO 控制器均可啟動 DMA 交易。為了從該主機傳送 IO 命令，其中該外部 IO 控制器是作為 DMA 從屬，該外部 IO 控制器於命令階段在 IO 匯流

排 709 上完成時，中斷該混合型儲存控制器，以便該混合型儲存控制器可透過該混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 主要介面而從外部 IO 控制器來讀取收到的 IO 命令，並將其傳送到位在 SRAM 707 或是 SDRAM 708 內的 IO 命令緩衝器。爲了從主機傳送 IO 命令，其中該外部 IO 控制器是作爲 DMA 主要，該混合型儲存控制器初期將 IO 命令緩衝器位在 SRAM 或 SDRAM 內的位置指示給外部 IO 控制器，以致於當命令階段在該 IO 匯流排上完成時，該外部 IO 控制器可將所收到的 IO 命令透過該混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 從屬介面寫入至位在混合型儲存控制器內的 IO 命令緩衝器。爲了從主機傳送資料或傳送資料到主機處，其中該外部 IO 控制器是作爲 DMA 從屬，該混合型儲存控制器於 IO 資料階段係建立在該 IO 匯流排上時，使用該混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 主要介面將資料寫入外部 IO 控制器或是從外部 IO 控制器讀取資料。爲了由主機傳送資料或傳送資料到主機處，其中該外部 IO 控制器是作爲 DMA 主要，該混合型儲存控制器將資料快取緩衝器位在 SRAM 或 SDRAM 內的位置指示給該外部 IO 控制器，以致於當該 IO 資料階段已建立在該 IO 匯流排上時，該外部 IO 控制器可以使用該混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 從屬介面來將資料寫入該混合型儲存控制器或是從該混合型儲存控制器讀取資料。資料是被寫入位在 SRAM 或 SDRAM 內的資料快取或是從位在 SRAM 或 SDRAM 內的資料快取讀取。爲了將 IO 狀態資訊傳送到主機，其中當該外部 IO 控制器是作爲 DMA 從屬，該混合型儲存控制器使用混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 主要介面來將 IO 狀態資

訊寫入到該外部 IO 控制器。該外部 IO 控制器在該 IO 匯流排上建立 IO 狀態階段，並且將 IO 狀態資訊傳送到主機。爲了將 IO 狀態資訊傳送到主機，其中該外部 IO 控制器是作爲 DMA 主要，混合型儲存控制器將 IO 狀態資訊位在 SRAM 或 SDRAM 內的位置指示給外部 IO 控制器，以致於當該 IO 狀態階段已建立在該 IO 匯流排上時，該外部 IO 控制器可以使用該混合型儲存控制器 PCI-Express/PCI-X/PCI DMA 從屬介面從該混合型儲存控制器處讀取該 IO 狀態資訊。

【0034】第 8 圖爲顯示本發明實施例的第七個範例組構的圖，其中，混合型儲存控制器 801 是組構成主機系統，其係透過諸如是序列連接介面 SCSI 803 之內部標準 IO 儲存介面來和外部儲存裝置 802 介接，且亦透過諸如是乙太網路 804 之標準 IO 網路介面來和網路介接。該外部儲存裝置亦可以是另一個混合型儲存裝置。

【0035】第 9 圖爲顯示本發明實施例的第八個範例組構的圖，其中混合型儲存控制器裝置 901 是組構成從屬裝置，且整合到包含在單一磁碟機包封體內的混合型硬碟中，且透過諸如是序列 ATA 的標準 IO 介面來和主機系統介接，並透過低階直接連結和磁碟控制器介接。

【0036】第 10a 圖爲顯示根據本發明實施例之儲存系統的非揮發性以及揮發性記憶體組件內的資料結構的圖。第 10a 圖顯示不同的儲存媒介以及每一個媒介是如何用於儲存以及快取資料、碼以及其它控制資料結構。資料 1001 是永久儲存在硬碟機 1002 內之資料的區塊。資料 1003 是永久儲存在硬碟機 1004 內之資料

的區塊。快閃亦提供資料永久的儲存。在快閃 1007 內的資料 1005 是在硬碟機 1002 內之資料 1001 的快取部份。在快閃 1007 內的資料 1006 是在硬碟機 1004 內之資料 1003 的快取部份。與快閃 1007 以及硬碟機 1002 和 1004 相比，SDRAM 1009 提供資料更快速的存取儲存。在 SDRAM 1009 內的資料 1008 是在快閃 1007 內之資料 1005 和 1006 的快取部份。這些快取部份是最近藉由主機系統(未顯示)從儲存系統 1000 讀取或寫入到儲存系統 1000 的。SRAM 1026 是最快速存取的儲存裝置，其可用以儲存資料。在圖式中，在 SRAM 1026 內的資料 1024 也是在快閃 1007 內之資料 1005 的快取部份。在 SRAM 1026 內的資料 1025 也是在快閃 1007 內之資料 1006 的快取部份。在 SRAM 1026 內快取的資料可以和在 SDRAM 1009 內快取的資料相同的方式來處理。快取的資料是假定為最近或是最常從主機存取的資料。然而，對於非必要夠資格被快取的隨機一次讀取存取而言，也可使用 SRAM 作為此種讀取資料的暫時儲存。一旦資料傳送到主機後，緩衝器便立即地清空(freed)。因為儲存系統是被管理著，使得由 DMA 控制器(未顯示)傳送到主機系統/從主機系統傳送，以及傳送到硬碟或快閃/從硬碟或快閃傳送的資料總是在 SDRAM 或 SRAM 內被快取。

【0037】碼 1012 代表著可在嵌入式處理器上運行的低階軟體。此碼實作用於管理儲存系統的演算法。“碼：O/S, Apps”，Apps 1013 代表著 OS 核心(kernel)以及應用程式碼。視需要地，可在諸如是 Windows，Linux 等傳統的 O/S 下使用另一個嵌入式處理器來執行應用程式。碼 1012 和 1013 是永遠地儲存在快閃 1007 內。由於這些是重要的資訊，備份副本(back-up copy)1016、1017 是儲

存在硬碟中。FPROM 1015 是另一個小容量、非揮發性的儲存媒介，其可用以儲存在啓動(power-up)時所載入的小量的碼 1014。然而，啓動時所載入之初期碼亦可從快閃 1007 所載入。剩餘的碼 1012、1013 則從快閃 1007 到 SRAM 1026 而被分頁(get paged)。IO 處理器從 SRAM 1026 執行碼，除非其是在第一階處理器內部快取(未顯示)中所快取者。因此，該 SRAM 1026 作為 IO 處理器的第二階快取。

【0038】 LBA-Flash-HDD 表 1010 是控制結構，其係在維持資料邏輯區塊位址(LBA)到其在快閃內和在硬碟內之實體位置的映射。快閃媒體快取該硬碟內之資料。除了實體位置外，也有相關於在快閃內之資料的狀態之資訊(是否修改、是否是暫時性的、是否準備重新映射至其它快閃位置、是否準備清回(flushed back)至硬碟)。LBA-Flash-HDD 表 1010 是由 IO 處理器來維護。更多有關於維護系統邏輯區塊位址到快閃實體區塊位址以及硬碟區塊位址的映射之詳細資料可在申請號為 11/450,005 的美國專利申請案中 找到，該美國專利申請案係在 2006 年 6 月 08 日提出申請，且標題為 “Optimized Placement Policy for Solid State Storage Divices” (在下文中稱為專利申請案)在此併入作為參考。最常存取的部份 1027 是緩衝在 SRAM 1026 內，SRAM 1026 可由 IO 處理器最快存取。較不常存取的部份 1011 是緩衝在 SDRAM 1009 內。在電源切斷(power-down)時，合併這些表 1027、1011，且更新的部份則是清回至它們可永久儲存的快閃 1007 內的表 1010。由於這些是重要的資訊，因此備份副本 1016 儲存在硬碟機內。

【0039】 LBA-SDRAM 表 1018、1019 延伸 LBA-Flash-HDD 表

以亦包含了資料邏輯區塊位址到其在 SDRAM 內的位置的映射，因為這些資料區塊是在 SDRAM 內快取的。除了 SDRAM 位置外，該表尚具有額外的資訊，其係關於快取資料區塊的狀態(是否修改、是否是暫時性的、是否準備清回硬碟或快閃)。LBA-SDRAM 表 1018、1019 是由 IO 處理器來維護。最常存取的部分 1019 是儲存在可由 IO 處理器最快存取之 SRAM 1026 內。較不常存取的部分 1018 是儲存在 SDRAM 1009 內。由於 SDRAM 1009 和 SRAM 1026 是揮發性儲存體，LBA-SDRAM 表 1018、1019 初期為空的，而在當資料區塊從儲存系統讀取或是寫入儲存系統時，才慢慢地建立起來。

【0040】暫存緩衝器(scratch buffer)1020 是集合性用語，指的是暫時的儲存區，其針對 IO 處理器在執行時期而緩衝資訊(例如佇列 IO 命令用於處理的那些緩衝器)，或由 OS 以及應用程式所使用的暫存記憶體(scratchpad memory)。SRAM 1026 以及 SDRAM 1009 兩者均可用以儲存此種資訊。LBA-SDRAM 表 1018、1019 是僅在執行時期才產生的控制資訊，並且是在 SRAM 1026 以及 SDRAM 1009 內使用暫存緩衝器之執行時期資訊的特殊狀況。

【0041】DMA 指令 1021 是由 IO 處理器所產生之執行時期控制資訊的另一組特殊狀況。產生該等 DMA 指令是提供給 DMA 控制器來使用。為了回應來自主機系統的讀取/寫入要求，IO 處理器針對 IO DMA 控制器或是 PCI-Express/PCI-X/PCI DMA 控制器產生 DMA 指令 1021，並將其儲存在 SDRAM 1009 或 SRAM 1026 內。當 IO 處理器傳送資料區塊到 SDRAM/從 SDRAM 傳送資料區塊到快閃時，IO 處理器產生 DMA 指令給快閃 DMA 控制器。當 IO 處

理器傳送資料到 SDRAM/從 SDRAM 傳送資料到硬碟機時，IO 處理器對和該硬碟機連結的 IO DMA 控制器產生 DMA 指令。DMA 指令包含了到下一指令的鏈結，因此 DMA 控制器可自動地提取儲存在 SDRAM 或 SRAM 內的 DMA 指令。這些 DMA 指令包含了在 SDRAM 內之位置供 DMA 控制器來提取/儲存資料。

【0042】ByteAdr-LBA 表 1022 指的是由 PCI-Express/PCI-X/PCI 介面 DMA 控制器所使用之位元組位址查詢表，用以查閱與位元組或字元位址相關聯之區塊位址。CAM 1023 則是爲此目的而使用的。

【0043】第 10b 圖係顯示根據本發明實施例之用於 LBA-Flash-HDD 映射表之範例資料結構的圖。在機械式硬碟以及在快閃內之快取位置之資料的真實位置是由嵌入式 IO 處理器在沒有主機干預下而獨立決定的。例如，若是主機使用邏輯區塊位址或 LBA 來參考資料，此種 LBA 會由 IO 處理器翻譯成實體位置。IO 處理器將在硬碟以及快閃內的實體位置最佳化，使得通常或是最近存取的資料會儲存在快閃內，而使其可以最快的方式來存取。此種最佳化的範例係藉由主機來將一致存取的一組 LBA 分配到在快閃陣列內的不同裝置中，使得該 LBA 組的部份可同時地存取。主機的存取可加以追蹤，而且利用存取行爲來最佳化存取效能。LBA_Flash_HDD 表指的是由嵌入式處理器所維護的資料結構，以便將由該主機存取的資料關聯到在快閃以及硬碟內的實體位置，並且對那些最近或是最常存取的資料在快閃陣列內分配其位置。將此類資料放置在快閃內是有利的，因爲在快閃與 SDRAM 之間傳送資料是較在硬碟與 SDRAM 之間傳送資料來的快。在該表內

的每個登錄(entry)將由該主機所定址的一組 LBA 關聯到關於其在快閃以及硬碟機內的位置的資訊。

【0044】在表內所含的資訊再分成快閃重新映射表以及 HDD 重新映射表。快閃重新映射表包含了快取資料區塊在快閃陣列內之實體位置(實體區塊位址或 PBA)的資訊。此一特別的資訊是由 IO 處理器來使用，以建立 DMA 指令，該 DMA 指令是由快閃 DMA 控制器來加以解譯(interpreted)，以控制快閃裝置或是快閃裝置群。除了目前的實體位置外，該表亦包含了資料的快取狀態的資訊。此資訊指示著在快閃內快取的資料和永久儲存在硬碟機內之對應者(counterpart)是如何的不同。此種資訊包含：是否那組資料尚未儲存在硬碟機內、其是否全部或是部份修改、是否其目前是暫時的，且尚不應准許存取…等。最後，快閃重新映射表亦包含了其它有關於實體快閃區塊之使用的控制資訊。此種資訊決定了資料是否符合移到其它快閃區塊，以延長快閃記憶體的使用年限或是作為最佳化的一部份，以改善由主機對資料的存取。

【0045】HDD 重新映射表包含了位置資訊(實體區塊位址或 PBA)以及其它諸如是 HDD 使用統計之控制資訊。位置資訊是由 IO 處理器來加以使用，以建立 DMA 指令，該 DMA 指令係由 IO DMA 控制器加以解譯，以准許其在硬碟機中唯一地定址資料。諸如 SCSI 或 ATA 之 IO 介面通常使用 LBA 或 CHS 的定址方案(scheme)，以將在硬碟機內的資料定址。使用統計是關於定址磁碟區或位置的使用頻率以及方式之額外的資訊。此資訊可由 IO 處理器使用在演算法中，其係將資料分配到磁碟最佳化，並改善主機對資料的存取。對於顯示在第 10b 圖中用以維持 LBA-Flash-HDD

映射表之最佳化方法的更多細節可在專利申請案中找到。

【0046】第 10c 圖為顯示根據本發明實施例用於 LBA-SDRAM 映射表的範例資料結構的圖。若是 SRAM 以和 SDRAM 相同的方式也用作為快取時，則 LBA-SDRAM 表亦適用於在 SRAM 內的資料。SDRAM/SRAM 提供了對主機最快的存取，因此所有由主機寫入的資料均會先緩衝在 SDRAM/SRAM 內。同時，由主機讀取的資料是先由快閃或硬碟機讀取到 SDRAM/SRAM。SDRAM/SRAM 表指的是由嵌入式處理器所維護的資料結構，以便將主機資料存取關聯到其在 SDRAM/SRAM 快取的位置。在表中的每個登錄將一組由主機所定址的 LBA 關聯到關於其在 SDRAM/SRAM 內的暫時位置的資訊，以及其在永久儲存器(快閃或是硬碟機)內的原始或是最終位置。一般而言，LBA_SDRAM 表包含了資訊，諸如是：在 SDRAM/SRAM 內之快取資料區塊的位置、諸如是資料區塊之快取狀態的控制資訊等。位置資訊可讓 SDRAM DMA 控制器或是 SRAM 控制器來實質地控制 SDRAM 裝置以及存取資料。快取狀態指示在 SDRAM 內的快取資料與在快閃內快取的或在硬碟機內儲存的版本相比是如何的不同。快取狀態包括了資訊，諸如是：是否該組資料尚未組構在永久儲存、是否其是完全地或是部份地修改、是否其目前是暫時的，且尚不應予存取等。若是資料得以永久地組構在快閃或是硬碟機內，在 LBA_Flash_HDD 表內就有對應的登錄。對於顯示在第 10c 圖中用以維持 LBA-SDRAM 映射表之最佳化方法的更多細節可在專利申請案中找到。

【0047】第 11 圖係顯示可應用在如第 1、2 和 3 圖所示之本發明實施例之啟動初始化流程。此程序是關於碼的移動以及控制

結構和資料的移動和更新。載入到嵌入式處理器內部快取記憶體
的初始碼是永久地儲存在諸如在第 10a 圖內之 1015 的 FEPROM
中。由 IO 處理器所執行之碼的初始部份將 FEPROM 內剩餘的碼傳
送到 SRAM，該 SRAM 是對 IO 處理器提供最快速的存取並對 IO
處理器而言作為是一種二階(level-2)快取之記憶體。由於 FEPROM
是為小容量裝置，用於 IO 處理器之剩餘的碼以及諸如是 OS 核心
之其它的碼、或是視需要地由另一嵌入式處理器來執行的應用程
式是儲存在快閃內。這些碼的部份是被分頁(paged)至 SRAM 中用
於執行。由 FEPROM 載入到 SRAM 的初始碼包含用於指示快閃 DMA
控制器來將下一組準備執行的常式(routine)予以分頁的常式。在將
碼由 FEPROM 分頁到 SRAM 的常式後，啟動的初始化亦必須將
SDRAM 分割成用於快取資料的區域以及用於儲存控制結構的區
域。快閃 DMA 控制器被指示提取控制結構的初始組，例如，將
LBA_Flash_HDD 表的位置資訊維持在快閃內的控制結構。下一步
則是初始化 LBA_SDRAM 表以指示 SDRAM 快取區域是空的。在
對 SDRAM 初始化之後，系統已準備好開始正常的操作。在正常
操作期間，IO 處理器服務來自主機的讀取/寫入要求，以及管理不
同的儲存媒介(SRAM、SDRAM、快閃以及 HDD)。其它的處理器
可執行其它的應用程式。資料在系統內於不同的裝置間傳送時，
儲存媒介的管理必須更新至不同的控制結構，並且將此類結構定
期地儲存在永久儲存媒介中(快閃以硬碟)。

【0048】第 12 圖為顯示主要可應用於如第 5、7 和 9 圖所示
之本發明實施例的來自主機電腦系統之區塊讀取命令的流程圖。
圖式的左側顯示著主機系統對混合型儲存裝置執行區塊讀取操作

的活動。圖式的右側顯示著在混合型儲存裝置內，從主機系統處收到區塊讀取命令時的活動。區塊 IO 傳送協定(ATA 或 SCSI)讓儲存裝置佇列所收到的命令，並在延長的時間內以所要求的資料區塊作回應。

【0049】混合型儲存裝置藉由使用在 SRAM/SDRAM 以及硬碟之間的快閃中間快取(flash intermediate cache)來減少回應時間。因為到 DMA 控制器的指令可加以鏈結，IO 處理器可在背景中建立數個此類指令，並將它們鏈結。使用指令鏈結，DMA 控制器可自動地由記憶體提取下一個指令，並在沒有來自 IO 處理器之額外的干預下執行所指派的傳送。一旦從非揮發性儲存(快閃以及硬碟)到揮發性儲存(SRAM/SDRAM)所傳送的資料量達到預定的臨界值，IO 介面 DMA 控制器則啟動以開始將資料區塊傳送到主機系統。區塊讀取命令的狀態資訊在混合型儲存裝置傳送了所要求的資料區塊後才予以送出。

【0050】第 13 圖為顯示主要可應用於如第 5、7 和 9 圖所示之本發明實施例的來自主機電腦系統之區塊寫入命令的流程圖。圖式的左側顯示著主機系統對混合型儲存裝置執行區塊寫入操作的活動。圖式的右側顯示著在混合型儲存裝置內，從主機系統處收到區塊寫入命令時的活動。區塊 IO 傳送協定(ATA 或 SCSI)讓儲存裝置佇列所收到的命令，並指示何時其準備好從主機系統接收資料區塊。

【0051】混合型儲存裝置藉由使用在 SRAM/SDRAM 以及硬碟之間的快閃中間快取來減少回應時間。因為到 DMA 控制器的指令可加以鏈結，所以 IO 處理器可在背景中建立數個此類指令，

並將它們鏈結。使用指令鏈結，DMA 控制器可自動地由記憶體提取下一個指令，並在沒有來自 IO 處理器之額外的干預下執行所指派之傳送。如有需要，資料必須從揮發性記憶體(SRAM/SDRAM)清回至快閃。在此狀況下，則啟動快閃 DMA 控制器以將資料由 SRAM/SDRAM 處傳送到快閃。如有需要，資料必須從揮發性記憶體(SRAM/SDRAM)清回至硬碟。在此狀況下，則啟動連接至硬碟的 IO 介面 DMA 控制器以將資料由 SRAM/SDRAM 處傳送到硬碟。當在揮發性儲存(SRAM/SDRAM)緩衝器的可用空間達到第二預定的臨界定值時，則啟動 IO 介面 DMA 控制器以持續地接收來自主機系統的資料區塊。用於區塊寫入命令的狀態資訊是在混合型儲存系統可寫入所有資料後才加以傳送。

【0052】第 14 圖為顯示透過可應用於顯示在第 2 和 3 圖中本發明實施例之系統匯流排而收到之隨機存取位元組讀取要求的流程圖。圖式的左側顯示著連接至系統匯流排之要求裝置的活動。圖式的右側顯示著在混合型儲存裝置內，於收到隨機存取位元組讀取要求時的活動。

【0053】給定了要求的讀取資料的位址後，顯示在第 1 圖的 PCI-Express/PCI-X/PCI DMA 控制器 111 可查詢在 CAM 112 內的位址。若是 CAM 回覆了位址有效的匹配，則亦由 CAM 所回覆的登錄索引係對應於含有所要求之讀取資料之 SRAM 或 SDRAM 內資料區塊的索引。SRAM 105 以及 SDRAM 108 亦顯示在第 1 圖。PCI-Express/PCI-X/PCI DMA 可將資料區塊索引翻譯成 SRAM 或 SDRAM 位址，並繼續讀取資料。然而，若是 CAM 不回覆有效的匹配時，則表示目前在 SRAM 或是 SDRAM 內快取的資料區塊沒

有含有所要求的讀取資料。在這種狀況下，PCI-Express/PCI-X/PCI DMA 控制器會通知顯示在第 1 圖的 IO 處理器 103 並給定位址。其後，IO 處理器使用映射表以及在第 17 圖所顯示的程序來尋找資料，並將資料從快閃或是硬碟傳送到在 SRAM 或 SDRAM 內閒置的資料區塊位置。當傳送完成時，IO 處理器將所要求的讀取資料位址寫入 CAM 登錄，該 CAM 登錄之索引對應於現在含有所要求讀取資料之 SRAM 或 SDRAM 內的資料區塊索引。PCI-Express/PCI-X/PCI DMA 控制器在偵測到有效的 CAM 匹配時會進行讀取所要求的資料。

【0054】第 15 圖為顯示透過可應用於顯示在第 2 和 3 圖中所示之本發明實施例的系統匯流排所收到之隨機存取位元組寫入要求的流程圖。圖式的左側顯示著連結到系統匯流排之要求裝置的活動。圖式的右側顯示著在收到隨機存取位元組寫入要求時，在混合型儲存裝置內的活動。

【0055】給定了寫入要求的位址後，顯示在第 1 圖之 PCI-Express/PCI-X/PCI DMA 控制器 111 可在 CAM 112 內查詢位址。若是 CAM 112 對該位址回覆了有效的匹配時，亦由 CAM 所回覆之登錄的索引則對應到在資料可寫入之 SRAM 或 SDRAM 內的資料區塊的索引。SRAM 105 以及 SDRAM 108 亦是顯示在第 1 圖中。PCI-Express/PCI-X/PCI DMA 可將資料區塊索引翻譯成 SRAM 或 SDRAM 位址，並接受寫入資料，然後再寫入到 SRAM 或 SDRAM。然而，若是 CAM 並未回覆有效的匹配時，其意謂著目前在 SRAM 或 SDRAM 內快取的資料區塊沒有含有可和資料一同寫入的資料位置。在這種狀況下，PCI-Express/PCI-X/PCI DMA 控

制器應通知顯示在第 1 圖內的 IO 處理器 103，並給定寫入位址。IO 處理器然後進行顯示在第 18 圖中的程序，以在資料可寫入之 SRAM 或 SDRAM 內獲得資料區塊位置。IO 處理器將把寫入位址寫入到 CAM 登錄中，該 CAM 登錄之索引係對應到在 SRAM 或 SDRAM 內的資料區塊索引，而該 SRAM 或 SDRAM 目前則含有所要求的寫入資料。PCI-Express/PCI-X/PCI DMA 控制器在偵測到有效的 CAM 匹配時，可進行接受寫入的資料，並將其寫入到 SRAM 或 SDRAM 中。

【0056】第 16 圖為顯示根據本發明實施例之 DMA 傳送的流程圖。圖式的左側顯示著 IO 處理器在決定其需要去指示 DMA 控制器以執行 DMA 傳送時的活動。圖式的右側顯示的是 DMA 控制器在由 IO 處理器所啟動以執行資料傳送時的活動。因為 DMA 指令可加以鏈結，IO 處理器可在背景中建立數個此類指令，並將它們加以鏈結。使用指令鏈結，DMA 控制器可以自動地由記憶體提取下一個指令，並在沒有來自 IO 處理器的額外干預下執行所指派的傳送。此流程的執行是作為回應來自主機系統之讀取/寫入要求的部份，或者作為用於不同儲存媒介之管理功能或用於有關 DMA 傳送的任何其他目的的部份。

【0057】第 17、18 和 19 圖為顯示可應用在多層儲存系統之基本快取演算法(algorithm)的範例圖式。資料以及控制資訊兩者均在系統內快取。作為是提供給 IO 處理器最快存取速度的媒介，SRAM 對於快取由 IO 處理器所使用之控制結構而言是理想的對象，諸如亦儲存在 SDRAM 內之不同的映射表。SRAM 亦可作為用於儲存由處理器所執行之碼的二階快取。然而，其亦可作為一種

資料快取，以補充 SDRAM。在第 17、18 和 19 圖中，雖祇提到 SDRAM，但亦可使用 SRAM 來快取資料。資料快取方案係由 IO 處理器碼來實行，並可程式化，以對系統的應用程式予以最佳化。尤其，顯示在第 17、18 和 19 圖中的快取方案顯示了使用 SDRAM 作為一階資料快取以及使用快閃作為二階資料快取。第 17 圖是一種高階流程圖，其顯示回應來自主機系統的讀取要求，如何使用以及更新映射表。圖式顯示了選項，其中對於儲存在快閃內之讀取要求的資料尚未準備快取，相反地，其是暫時地儲存在 SRAM 1701 中。第 18 圖是一種高階流程圖，其顯示回應來自主機系統的寫入要求，如何使用以及更新映射表。第 19 圖顯示了將資料清回到 HDD 以及快取在快閃內資料的程序，上述兩者在當快取滿載 (Cache Full) 發生時，均可作為在背景程序或是在寫入期間來加以啟動。為了在寫入時最小化快取滿載，在 L1 表內登錄的最小計量 (count) 維持著其被替換的合適性，使得寫入要求可立即地被接受。若是對應的資料已被清回至 HDD 的快閃或是資料區塊已清空時，在 L1 內的登錄可被替換。在所顯示的方案中，來自連結的主機系統之所有的資料存取要求是由作為一階(L1)快取的 SDRAM 來讀取以及寫入。快閃作為是二階(L2)快取，以儲存永久存在 HDD 內資料的某部份的副本。用以決定在 L2 內快取哪部份的演算法可能為不同。基本的標準可以是儲存那些最近存取的部份。

【0058】前述本發明的實施例是作為例示以及解釋。其並非意圖限制本發明於所描述的特定形式。尤其，應考慮到在此所描述之本發明的功能性實作可實作於硬體、軟體、韌體及/或其它可用的功能性組件或是建立區塊，同時該資料儲存系統亦可包括和

網路連結的裝置，並且該網路可以是有線、無線或是有線和無線的合併。其它的變化以及實施例在按照上述的教示後當成爲可能，因此本發明的範疇並不受限於詳細說明，而是由下列的申請專利範圍來限制。

【符號說明】

【0059】

101	混合型儲存裝置		
102、201、301、401、501、601、701、801	混合型儲存控制器		
103、410	IO 處理器	104、411	計算/應用程式處理器
105、707	SRAM	106	場可程式化 ROM
107	SDRAM 控制器	108	SDRAM
109	快閃 DMA 控制器		
110	快閃裝置、快閃陣列、快閃		
111	PCI-Express/PCI-X/PCI DMA 控制器		
112、1023	內容可定址記憶體		
113、114、402、403、504、505	IO 儲存介面 DMA 控制器		
115、116、406、407	IO 網路介面 DMA 控制器		
117	電力防護電路	118	控制路徑
119、120	DMA 路徑	202、502、702	主機系統
203	系統匯流排	204	硬碟機
302	從屬裝置	303	系統匯流排
404、405	硬碟陣列	408、409	網路
503、709	IO 匯流排	602、802	外部儲存裝置
603、703	IO 介面控制器	604	光纖通道

605	乙太網路、混合型儲存控制器 PCI-Express/PCI-X/ PCI DMA 主要介面		
606、706	PCI-Express/PCI-X/PCI DMA 控制器		
708	SDRAM	704、705	序列 ATA
803	序列連接介面 SCSI		
804	乙太網路	901	混合型儲存控制器裝置
1000	儲存系統		
1001、1003、1005、1006、1008、1024、1025	資料		
1002、1004	硬碟機	1007	快閃
1009	SDRAM	1010、1011、1027	LBA-Flash-HDD 表
1012、1013、1014	碼	1015	FEPROM
1016、1017	備份副本	1018、1019	LBA-SDRAM 表
1020	暫存緩衝器	1021	DMA 指令
1022	ByteAdr-LBA 表	1026	SRAM

申請專利範圍

1. 一種資料儲存系統，用於在沒有來自外部主機的干預下使用至少一個邏輯位址來儲存以及擷取電腦資料，該資料儲存系統包括：

IO 處理器，用於控制該資料儲存系統的資料輸入或輸出；

磁碟機，用於儲存第一區塊的資料，該第一區塊的資料包括從該外部主機傳送來的資料以及用於該 IO 處理器的控制資料，該磁碟機係相關於磁碟機位址；

快閃記憶體機制，包括至少一個快閃記憶體，該快閃記憶體係用於儲存第二區塊的資料，該第二區塊的資料包括從該外部主機傳送來的資料、儲存在該磁碟機內之資料的快取部份、以及用於該 IO 處理器的控制資料，該快閃記憶體係相關於快閃位址；以及

LBA_Flash_HDD 表，包括第一部份和第二部份，該第一部份係用於將邏輯位址映射至該快閃位址，而該第二部份係用於將該邏輯位址映射至該磁碟機位址，該 LBA_Flash_HDD 表係用於在沒有從該外部主機來之干預下，致能該 IO 處理器在該快閃記憶體、該磁碟機、或兩者上實施記憶操作；

其中，該 FAST 記憶體機制包括：SDRAM 記憶體機制及 SRAM 記憶體機制，該 SDRAM 記憶體機制係用於儲存一個或多個區塊的資料，該資料包括從該外部主機傳送來的該資料、儲存在該快閃記憶體機制內的資料的該快取部份、以及用於該 IO 處理器的該控制資料，而該 SRAM 記憶體機制則包括一個或多個 SRAM，該等 SRAM 係用於儲存該一個或多個區塊的資

料，該資料包括用於該 IO 處理器機制的該控制資料；其中，該 LBA_FAST table 係儲存在該 SDRAM 記憶體機制中，並在該 SRAM 記憶體機制中被快取；其中，該 LBA_Flash_HDD 表係儲存在該快閃記憶體機制中，並在該 FAST 記憶體機制中被快取，其中，第一快取部份係儲存在該 SRAM 記憶體機制中，而包含該 LBA_Flash_HDD 表之剩餘部份的第二快取部份則係儲存在該 SDRAM 記憶體機制中；以及其中，該 LBA_Flash_HDD 表的副本係儲存在該磁碟機中。

2. 如申請專利範圍第 1 項的資料儲存系統，其中，該 LBA_Flash_HDD 表是在沒有該外部主機的干預下，根據該外部主機之資料存取行為模式而不時的予以更新。
3. 如申請專利範圍第 1 項的資料儲存系統，復包括：

FAST 記憶體機制，用於儲存一個或多個區塊的資料，該資料包含從該外部主機傳送來的資料、儲存在該快閃記憶體機制內之資料的快取部份、以及用於該 IO 處理器機制的控制資料；以及

LBA_FAST 表，用於將該邏輯位址映射至該一個或多個區塊的資料，並儲存該一個或多個區塊的資料的控制資訊，

其中，該 IO 處理器係設置以在沒有從該外部主機來之干預下，藉由使用該 LBA_FAST 表，以存取儲存在該快閃記憶體機制中的資料的該快取部份。

4. 如申請專利範圍第 3 項的資料儲存系統，其中，該 LBA_FAST 表以及該 LBA_Flash_HDD 表是在沒有從該外部主機來之干預下，根據該外部主機之資料存取行為模式而不時的予以更新。

5. 如申請專利範圍第 1 項的資料儲存系統，復包括：

SDRAM DMA 控制器，用於將資料傳送到該 SDRAM 記憶體機制或是從該 SDRAM 記憶體機制傳送資料出來，以回應一個或多個 DMA 指令；以及

快閃 DMA 控制器，用於將資料傳送到該快閃記憶體機制及從該快閃記憶體機制傳送資料出來，以回應至少一個或多個 DMA 指令；

其中，該 IO 處理器用於藉由使用該 LBA_FAST 表以及該 LBA_Flash_HDD 表，以準備一個或多個 DMA 指令；以及

其中，儲存於該 FAST 記憶體機制中用於該 IO 處理器的該控制資料復包括該一個或多個 DMA 指令。

6. 如申請專利範圍第 1 項的資料儲存系統，復包括：

主機 DMA 控制器，用於將資料傳送到該外部主機或是從該外部主機將資料傳送出來，以回應接收該一個或多個 DMA 指令；

CAM 記憶體，用於儲存位元組位址查詢表，其中，該主機 DMA 控制器機制將資料傳送到該外部主機以及從該外部主機傳送資料出來，而不需該 IO 處理器在該位元組位址查詢表含有用於由該外部主機所要求之該資料的有效登錄之情況下準備該 DMA 指令；以及

IO DMA 控制器，用於回應該一個或多個 DMA 指令而將資料傳送到該磁碟機以及從該磁碟機傳送資料出來；

其中，該外部主機是耦接至該主機 DMA 控制器；以及

其中，該磁碟機是透過 IO 介面而耦接至該 IO DMA 控制

器，該 IO 介面是選自由 IDE/ATA、序列 ATA、USB、FIREWIRE、SCSI、FIBER CHANNEL、以及 ETHERNET 所組成的群組。

7. 如申請專利範圍第 1 項的資料儲存系統，復包括：

第一 IO DMA 控制器，用於回應一個或多個 DMA 指令而將資料傳送到該外部主機以及從該外部主機傳送資料出來；以及

第二 IO DMA 控制器，用於回應一個或多個 DMA 指令而將資料傳送到該磁碟機以及從該磁碟機傳送資料出來；

其中，該外部主機是耦接至該第一 IO DMA 控制器；以及

其中，該磁碟機是耦接至該第二 IO DMA 控制器。

8. 如申請專利範圍第 1 項的資料儲存系統，復包括：

IO DMA 控制器，用於回應一個或多個 DMA 指令而將資料傳送到該外部主機及從該外部主機將資料傳送出來；以及

外部匯流排介面 DMA 控制器，用於回應一個或多個 DMA 指令而將資料傳送到該磁碟機及從該磁碟機將資料傳送出來；

其中，該外部主機是耦接至該 IO DMA 控制器；以及

其中，該磁碟機是耦接至該外部匯流排介面 DMA 控制器。

9. 一種資料結構，用於儲存資料儲存系統的映射資訊，該資料儲存系統包括磁碟機、快閃記憶體、SDRAM 記憶體、及 SRAM 記憶體，該資料結構包括：

LBA_Flash_HDD 表，包括一個或多個邏輯位址、一個或多個對應的快閃位址、以及一個或多個對應的磁碟機位址；以及

LBA_SDRAM 表，包括一個或多個邏輯位址、一個或多個對應的 SRAM 位址、以及一個或多個對應的 SDRAM 位址；

其中，該 LBA_Flash_HDD 表之工作副本的第一部份是儲存在該 SRAM 記憶體內，該工作副本的第二部份是儲存在該 SDRAM 記憶體內；以及

其中，該 LBA_SDRAM 表是儲存在該 SDRAM 記憶體內，而該 LBA_SDRAM 表的快取部份是儲存在該 SRAM 記憶體機制內。

10. 一種資料儲存方法，用於在沒有來自外部主機或是用戶端系統的干預下使用一個或多個邏輯位址來儲存以及擷取電腦資料，該方法包括：

控制資料儲存系統的資料輸入或輸出；

儲存一個或多個區塊的資料，該資料包括從該主機傳送來的資料以及用於 IO 處理器的控制資料；

儲存複數個資料區塊，該資料區塊包括從該主機傳送來的資料、儲存在磁碟機內之資料的快取部份、以及用於該 IO 處理器的控制資料；

儲存至少一個邏輯位址，該邏輯位址係映射到非揮發性記憶體陣列的至少一個實體位址及該磁碟機的至少一個實體位址；以及

在沒有從該主機來之干預下致能至該非揮發性記憶體機制之記憶存取。

11. 如申請專利範圍第 10 項的資料儲存方法，復包括：在沒有該主機干預下，根據該主機之資料存取行為模式而更新第一表。
12. 如申請專利範圍第 10 項的資料儲存方法，復包括：

儲存一個或多個區塊的資料，該資料包含從該主機傳送來

的資料、儲存在該非揮發性記憶體陣列內之資料的快取部份、以及用於該 IO 處理器機制的控制資料；

在第二表內查詢該一個或多個邏輯位址中之至少一者，一旦找到後，將該邏輯位址映射到至少一個 FAST 記憶體位址，以在沒有從該主機來之干預下，提供該邏輯位址與資料之實體位址間之相關，並致能至該資料之存取。

13. 如申請專利範圍第 12 項的資料儲存方法，復包括：在沒有從該主機來之干預下，根據該主機之資料存取行為模式，更新該第一表以及該第二表。

14. 如申請專利範圍第 13 項的資料儲存方法，復包括：

儲存一個或多個區塊的資料，該資料包括由該主機傳送來的資料、儲存在該非揮發性記憶體陣列內之資料的快取部份、用於該 IO 處理器的控制資料、以及該第二表；

儲存一個或多個區塊的資料，該資料包括用於該 IO 處理器及用於快取該第二表的控制資料；以及

其中，在運作期間，將該一表儲存在該非揮發性記憶體陣列內，並在該 FAST 記憶體內快取該第一表，將第一快取部份儲存在該 SRAM 記憶體內，並將含有該第一表之剩餘部份的第二快取部份儲存在該 SDRAM 記憶體內，以及將該第一表的副本儲存在該磁碟機內。

15. 如申請專利範圍第 14 項的資料儲存方法，復包括：

將資料傳送到該 SDRAM 記憶體或是從該 SDRAM 記憶體傳送資料出來，以回應至少一個 DMA 指令；

將資料傳送到該非揮發性記憶體陣列及從該非揮發性記

憶體陣列傳送資料出來，以回應至少一個 DMA 指令；

準備至少一個 DMA 指令，並使用該第二表以及該第一表，以用於映射該一個或多個邏輯位址；以及

將該控制資料在該 FAST 記憶體內，該控制資料復包含至少一個 DMA 指令。

16. 如申請專利範圍第 14 項的資料儲存方法，復包括：

將資料傳送到該主機或是從該主機將資料傳送出來，以回應至少一個 DMA 指令，以及將該主機介接於該資料儲存系統；

儲存位元組位址查詢表，其中，資料係傳送到該主機以及從該主機傳送資料出來，而不需該 IO 處理器在該位元組位址查詢表含有用於由該主機所要求之該資料的有效登錄之情況下準備該 DMA 指令；以及

回應至少一個 DMA 指令而將資料傳送到該磁碟機以及從該磁碟機傳送資料出來。

17. 如申請專利範圍第 14 項的資料儲存方法，復包括：

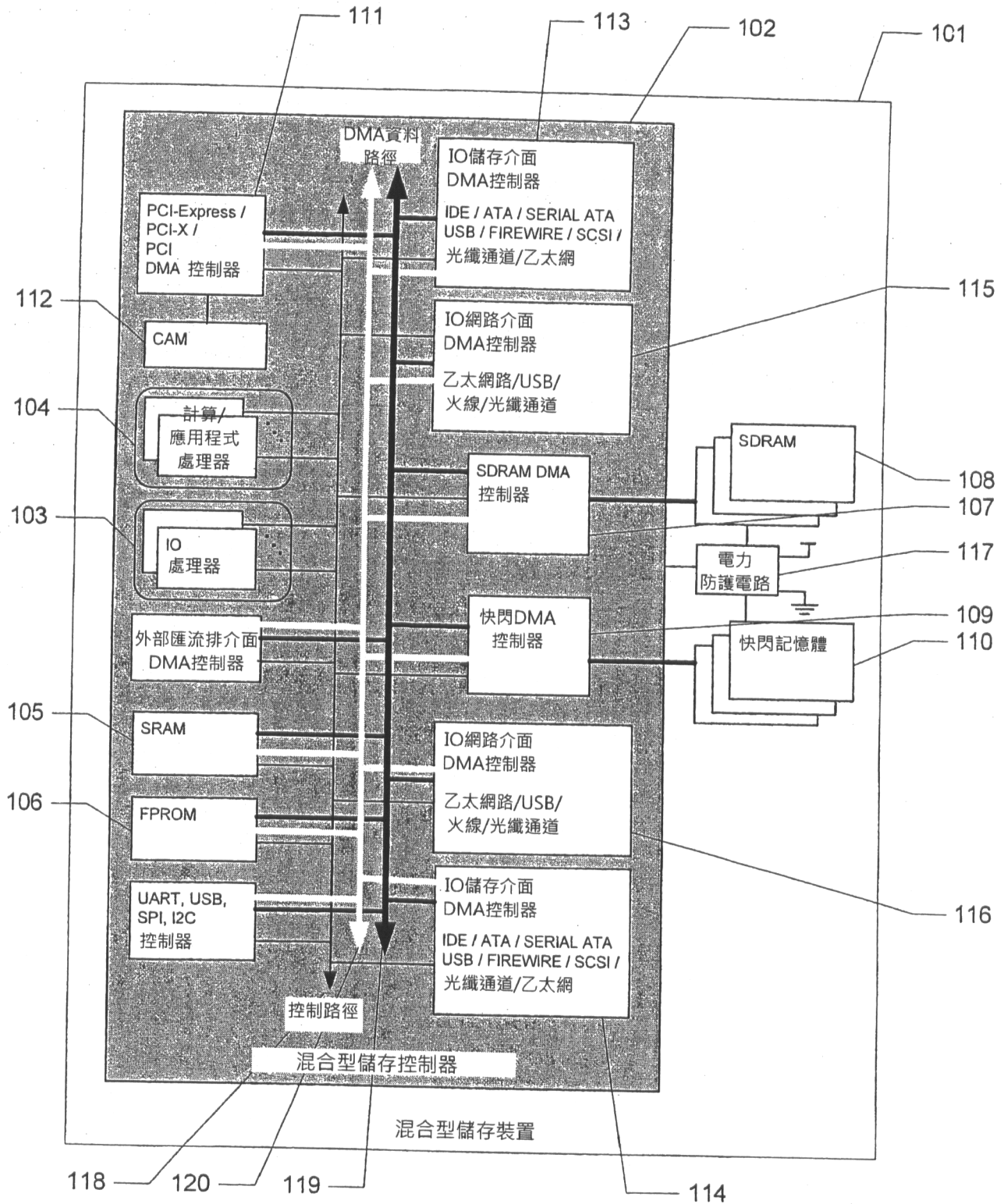
回應至少一個 DMA 指令而將資料傳送到該主機以及從該主機傳送資料出來；以及

回應至少一個 DMA 指令而將資料傳送到該磁碟機以及從該磁碟機傳送資料出來。

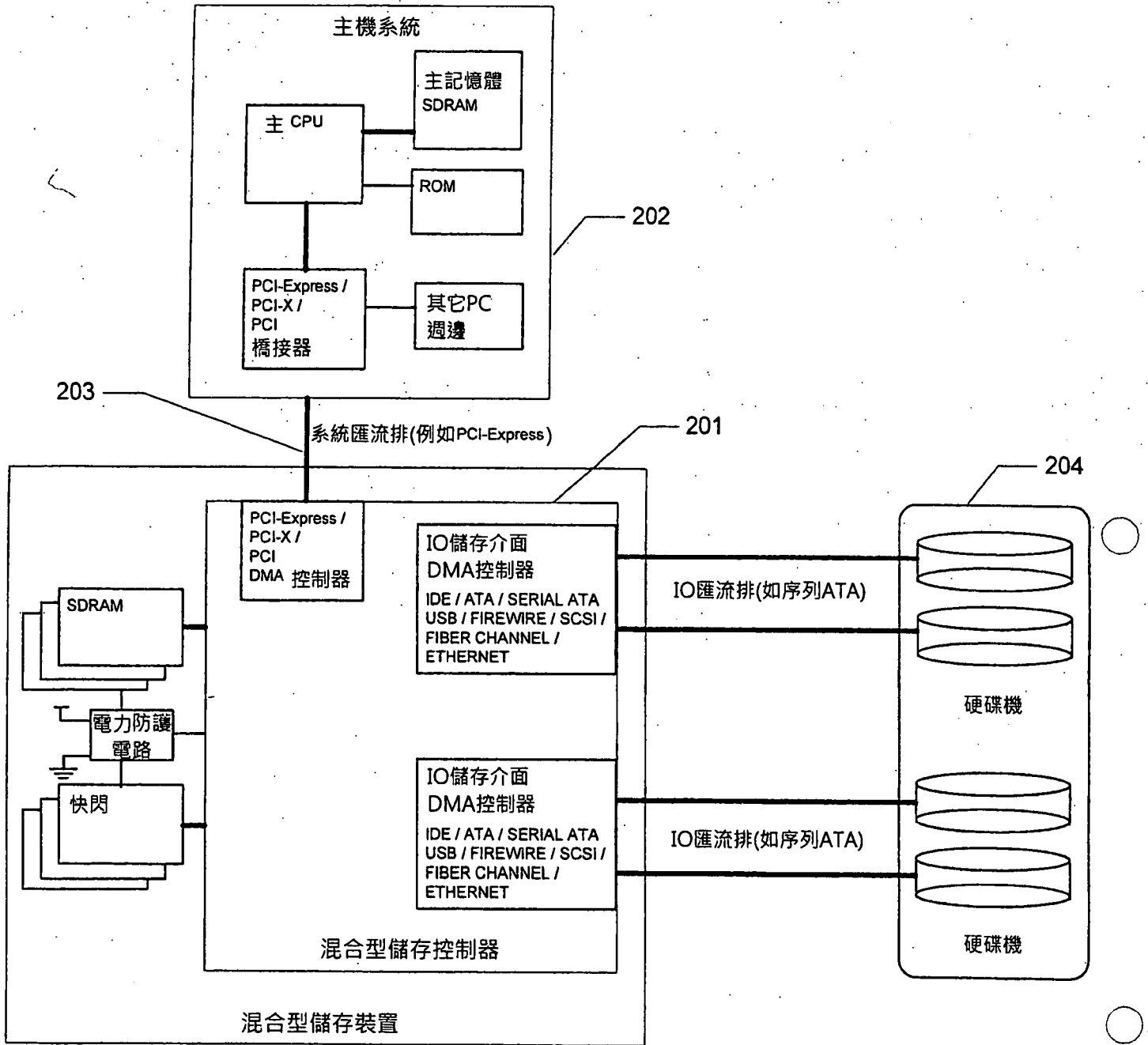
18. 如申請專利範圍第 14 項的資料儲存方法，復包括：

回應至少一個 DMA 指令而將資料傳送到該主機及從該主機將資料傳送出來；以及

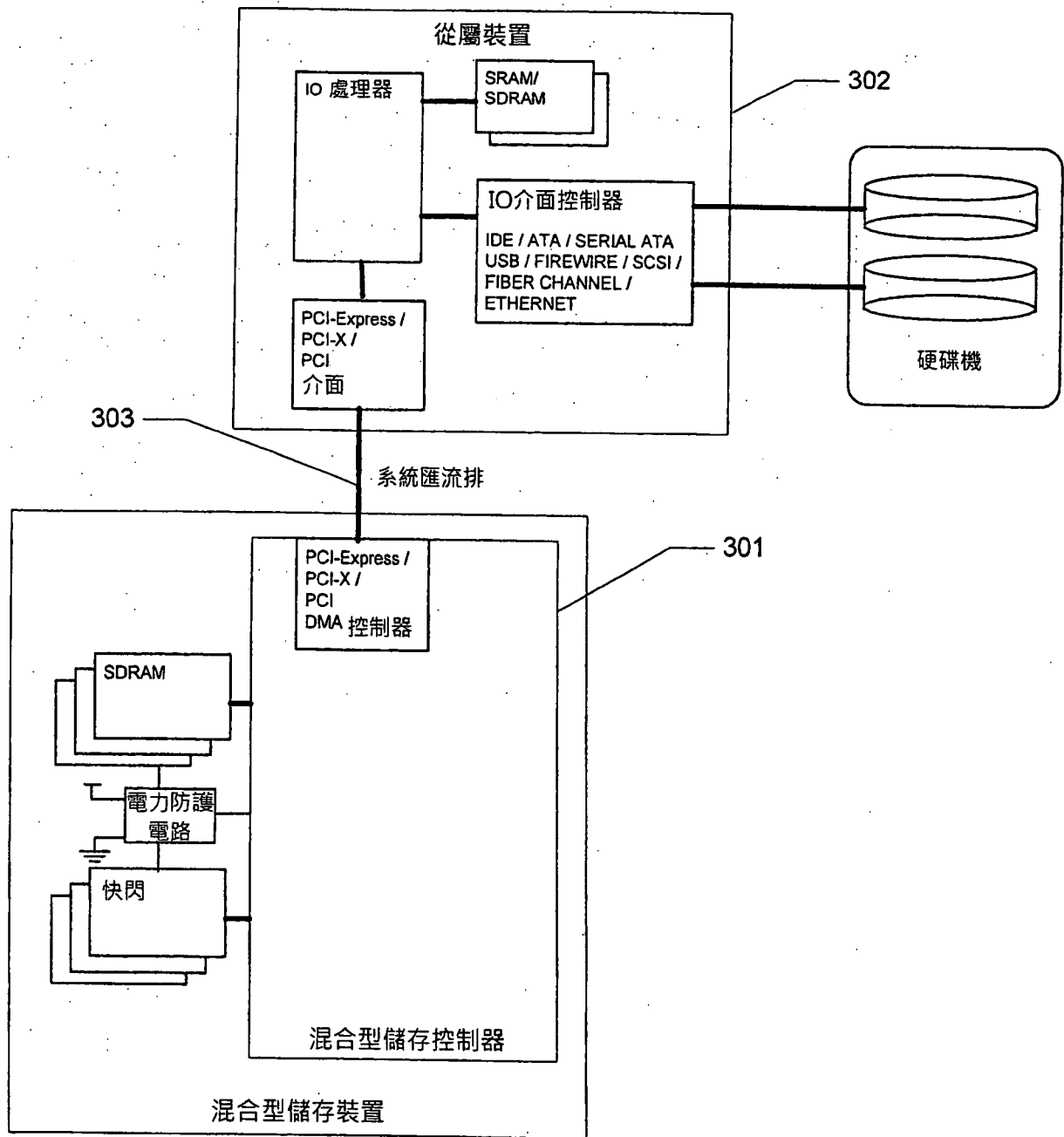
回應至少一個 DMA 指令而將資料傳送到該磁碟機及從該磁碟機將資料傳送出來。



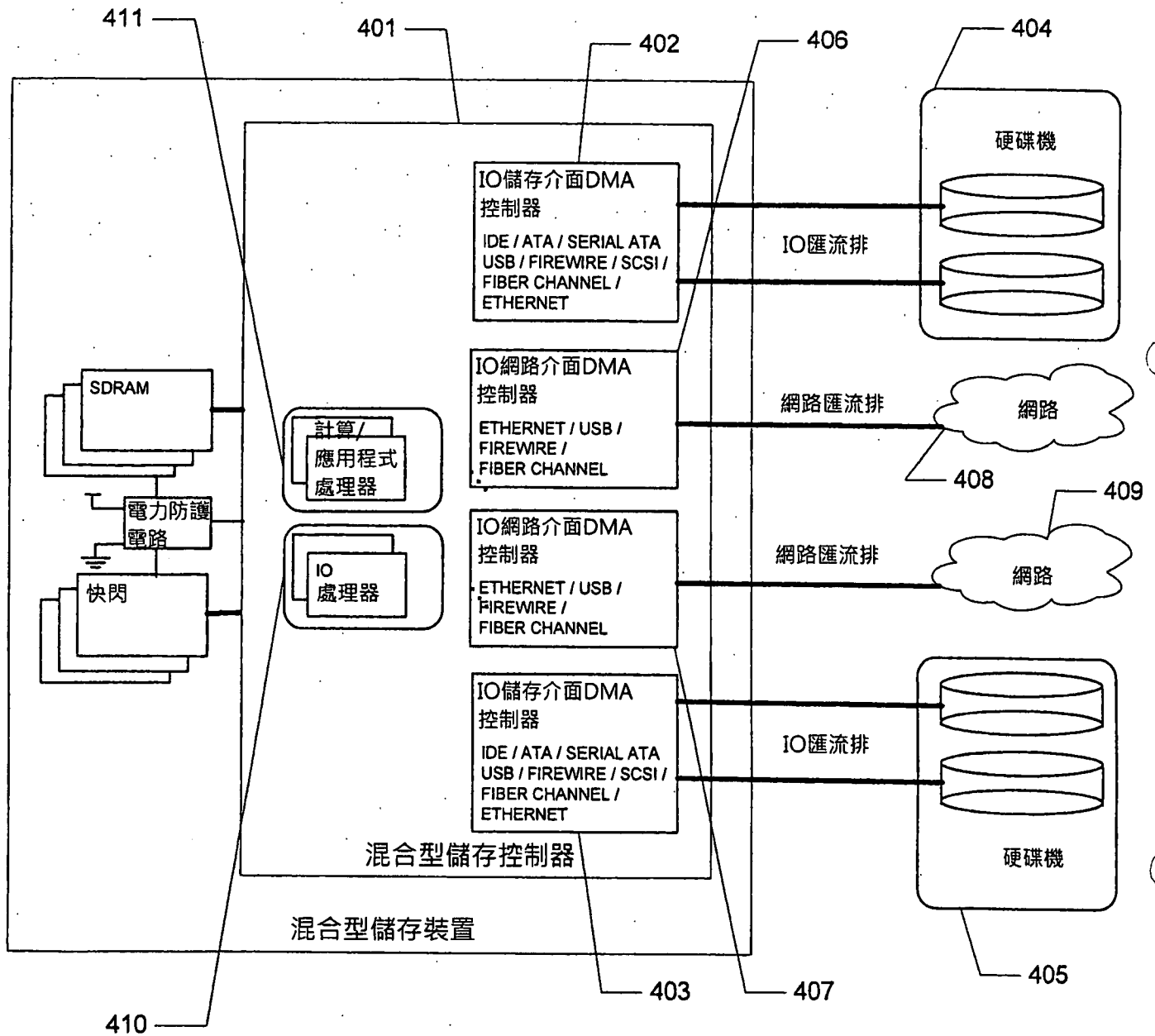
第1圖



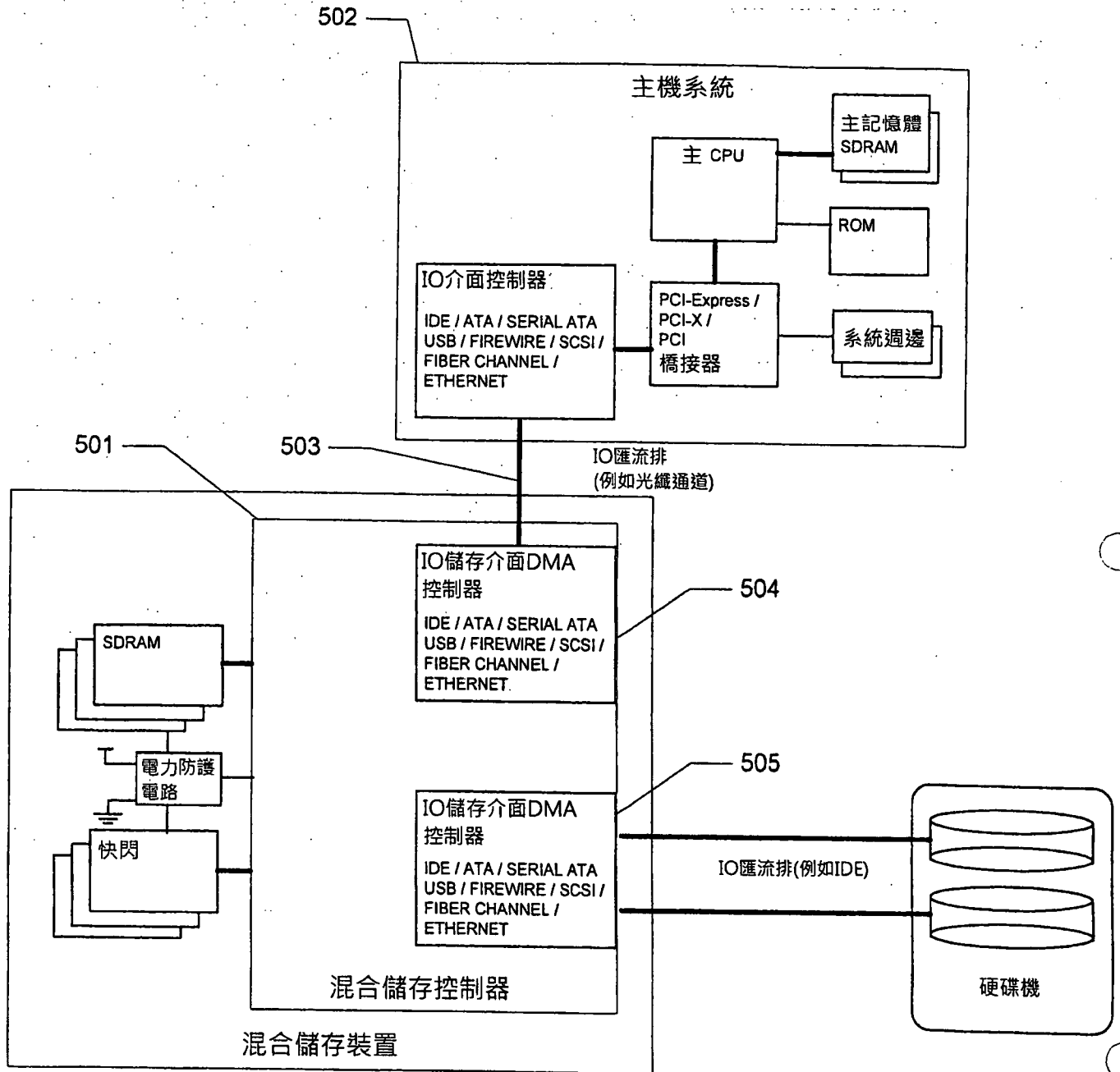
第2圖



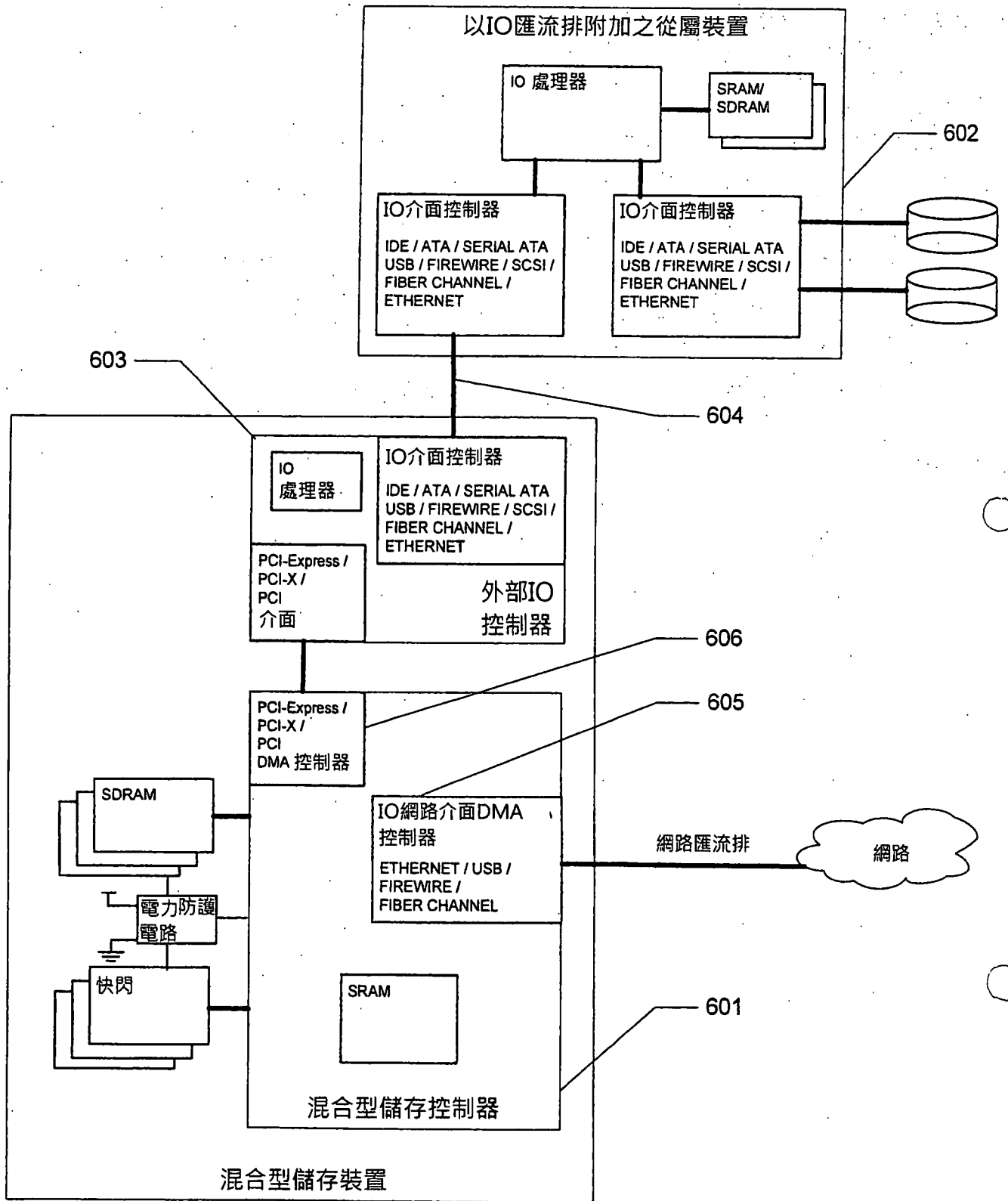
第3圖



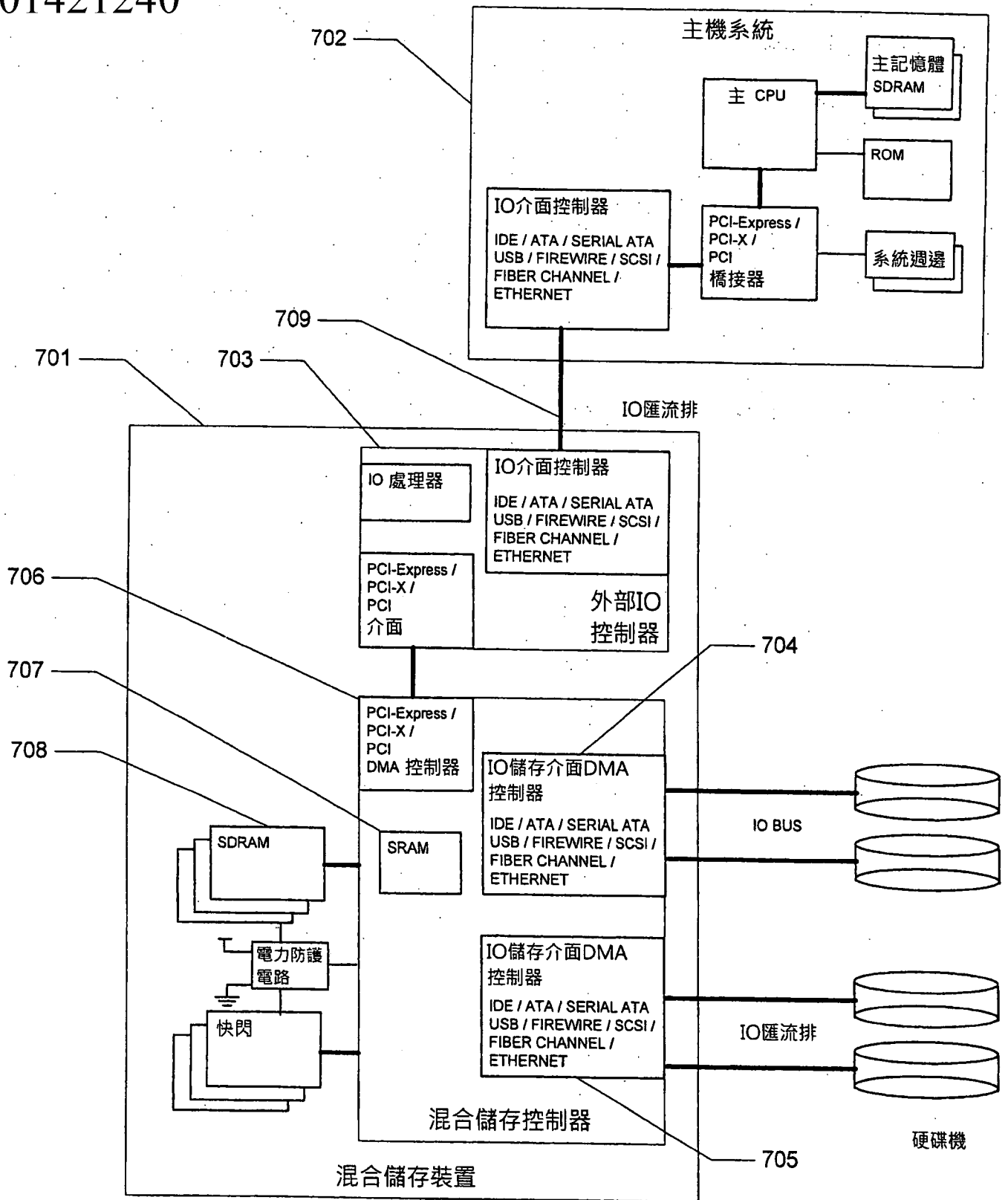
第4圖



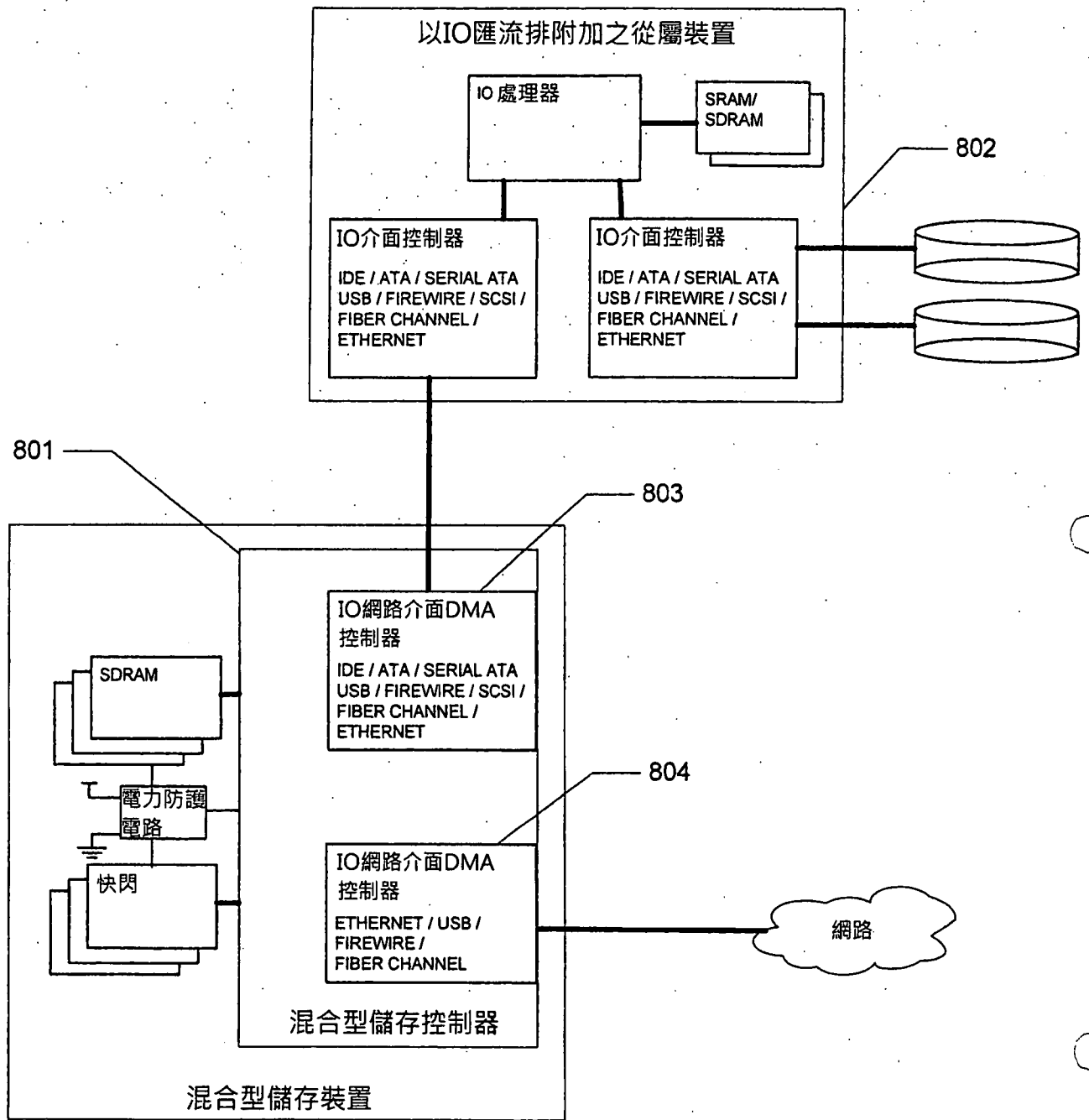
第5圖



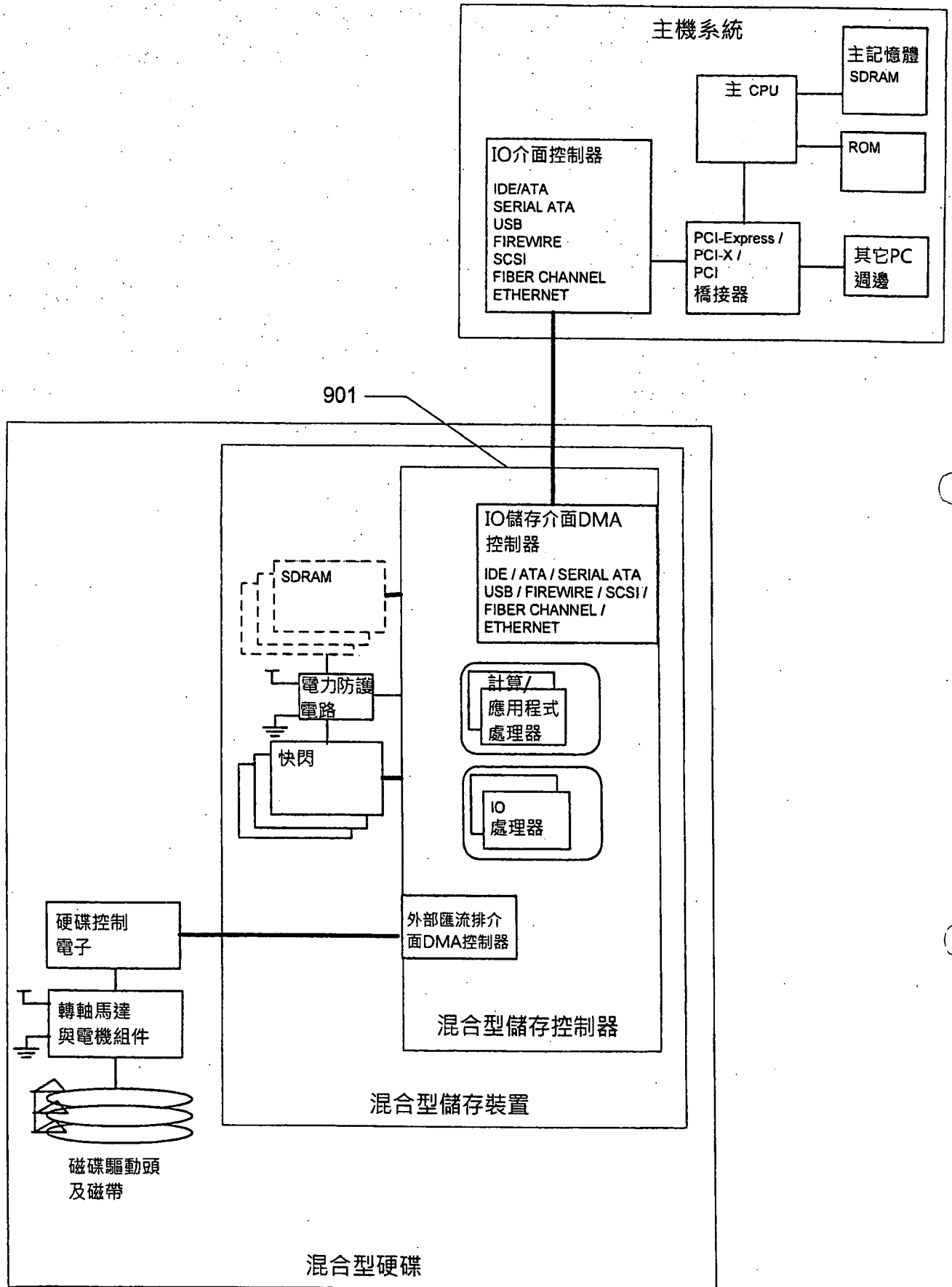
第6圖



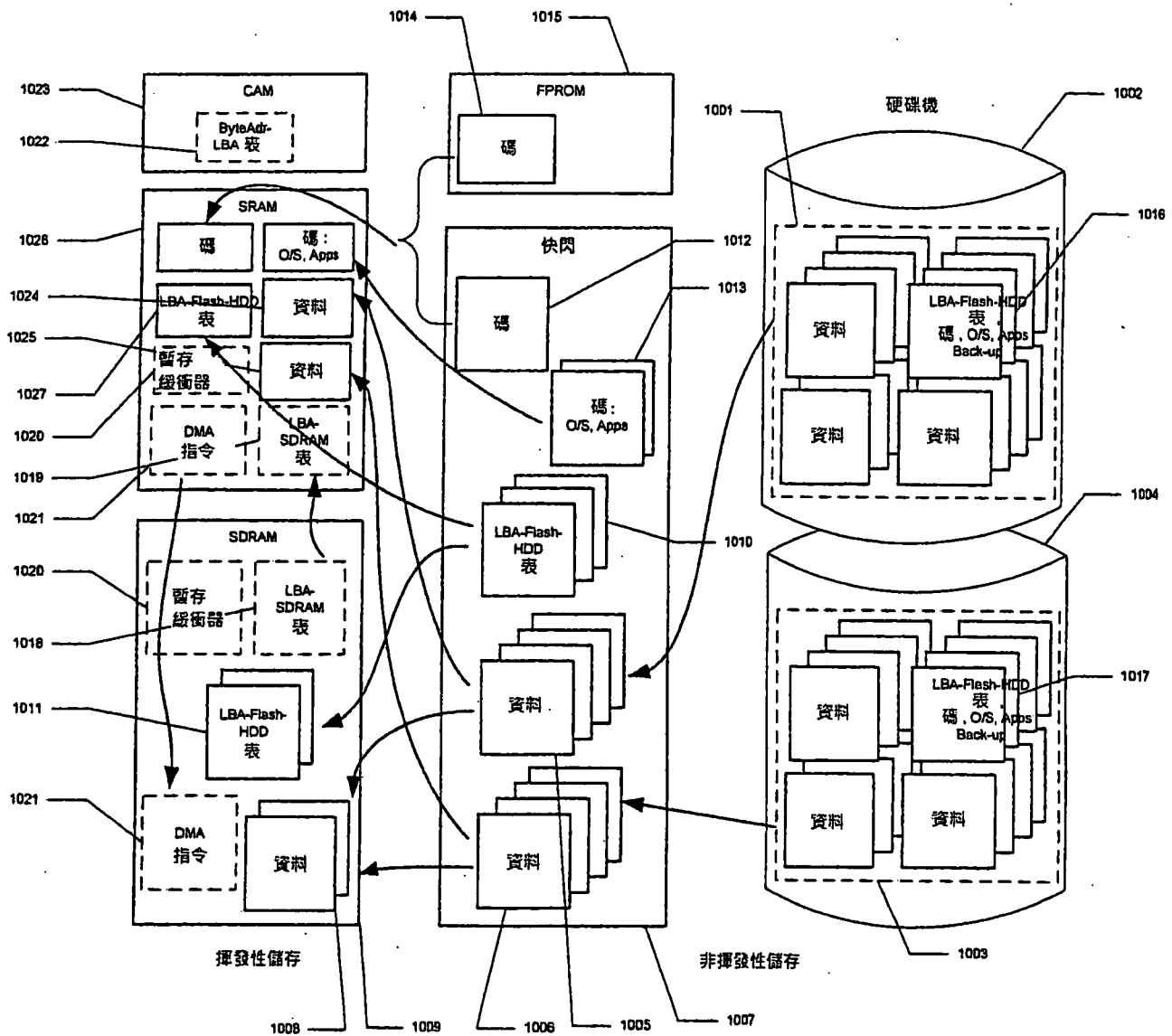
第7圖



第8圖



第9圖



第 10a 圖

LBA-FLASH-HDD 表

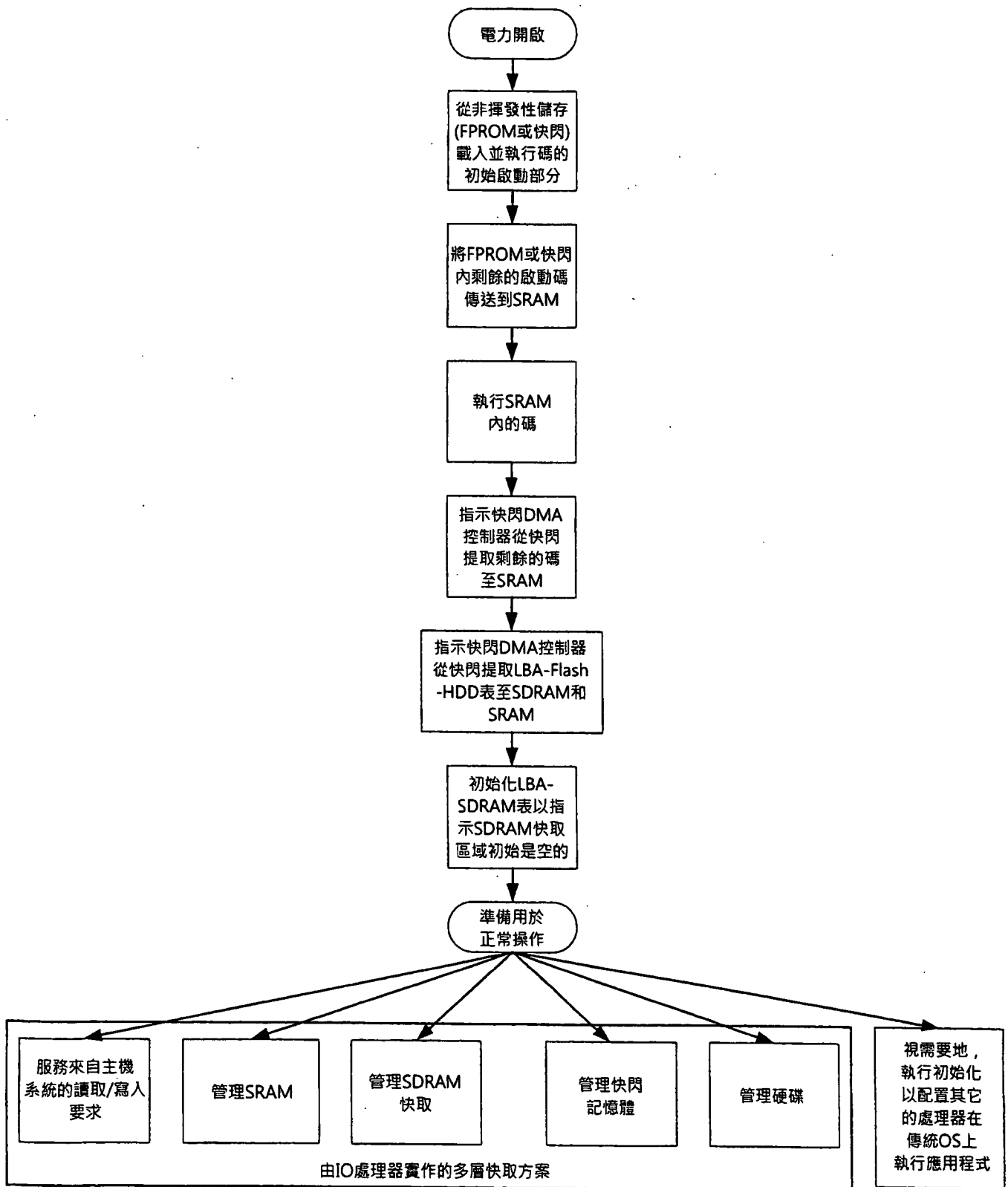
LBA Set Idx	快閃重新映射表			硬碟重新映射表	
	PBA	控制資訊		HDD PBA or LBA	控制資訊
		快取狀態	其它快閃 控制資訊		
0x0			用於主LBA組0的登錄		
0x1			用於主LBA組1的登錄		
0x2			用於主LBA組2的登錄		
⋮			⋮		

第 10b 圖

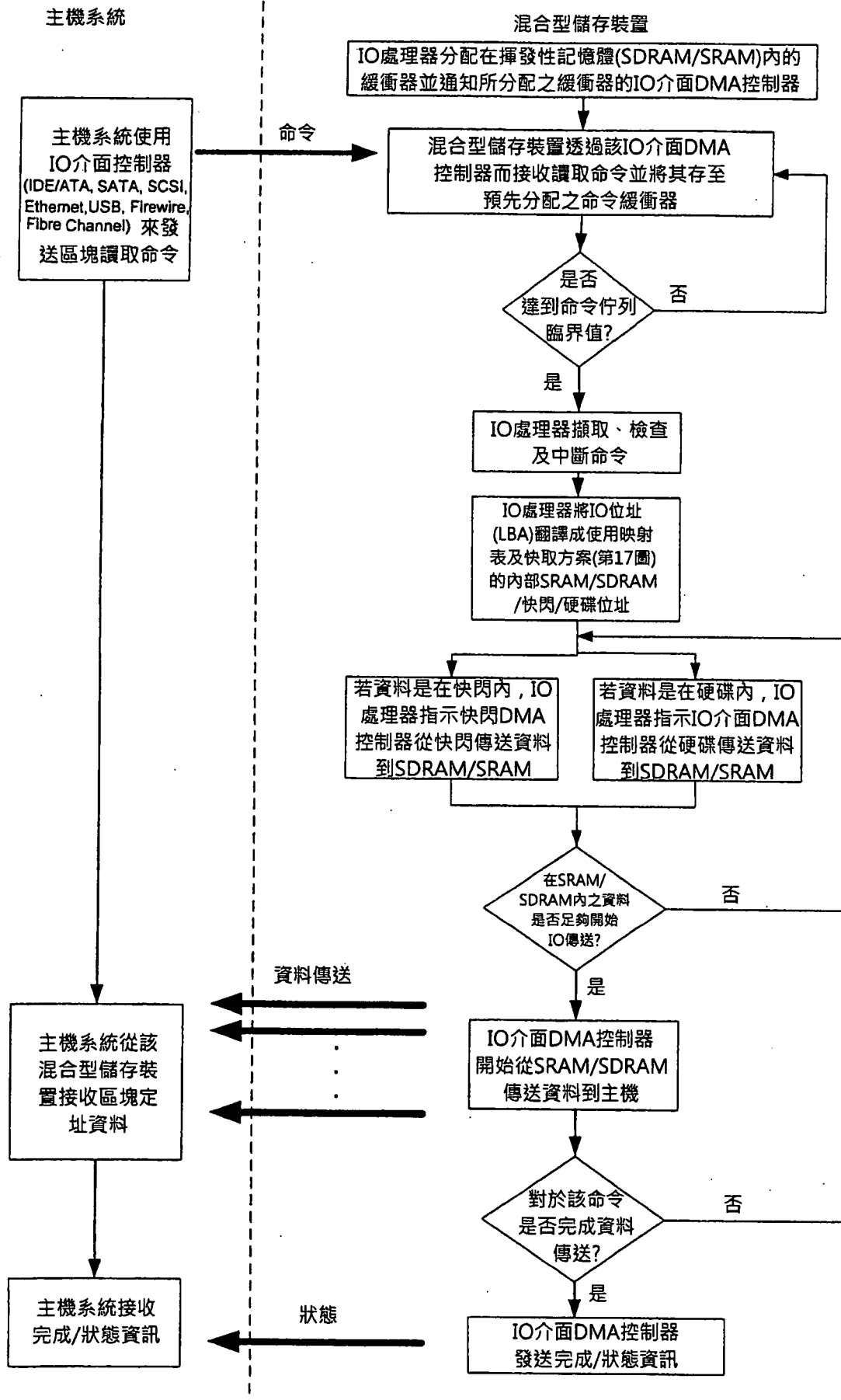
LBA-SDRAM 表

LBA Set Idx	PBA	控制資訊	
		快取狀態	其它 控制資訊
0x0		用於主LBA組0的登錄	
0x1		用於主LBA組1的登錄	
0x2		用於主LBA組2的登錄	
⋮		⋮	

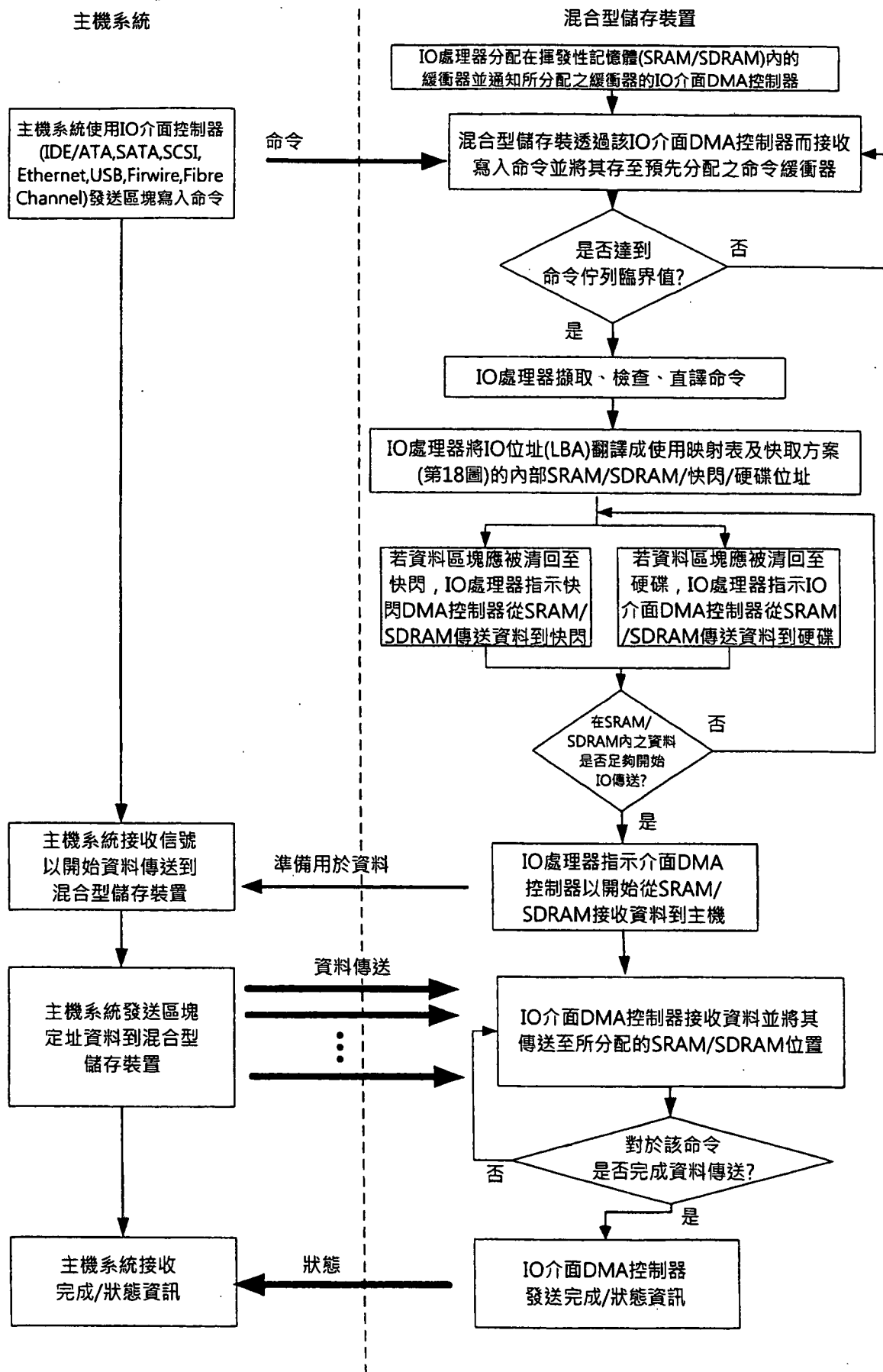
第 10c 圖



第11圖



第12圖

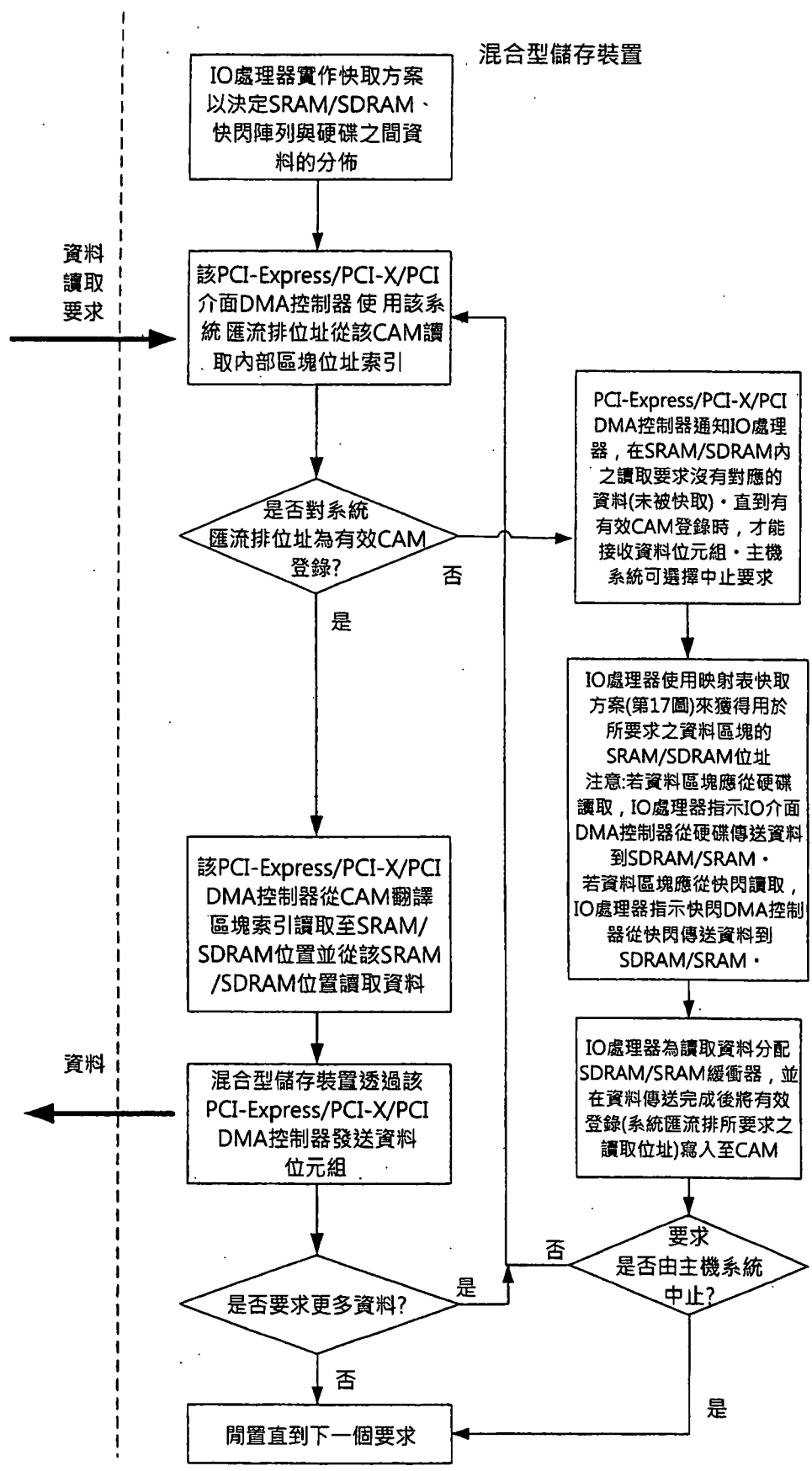


第13圖

主機系統

混合型儲存裝置

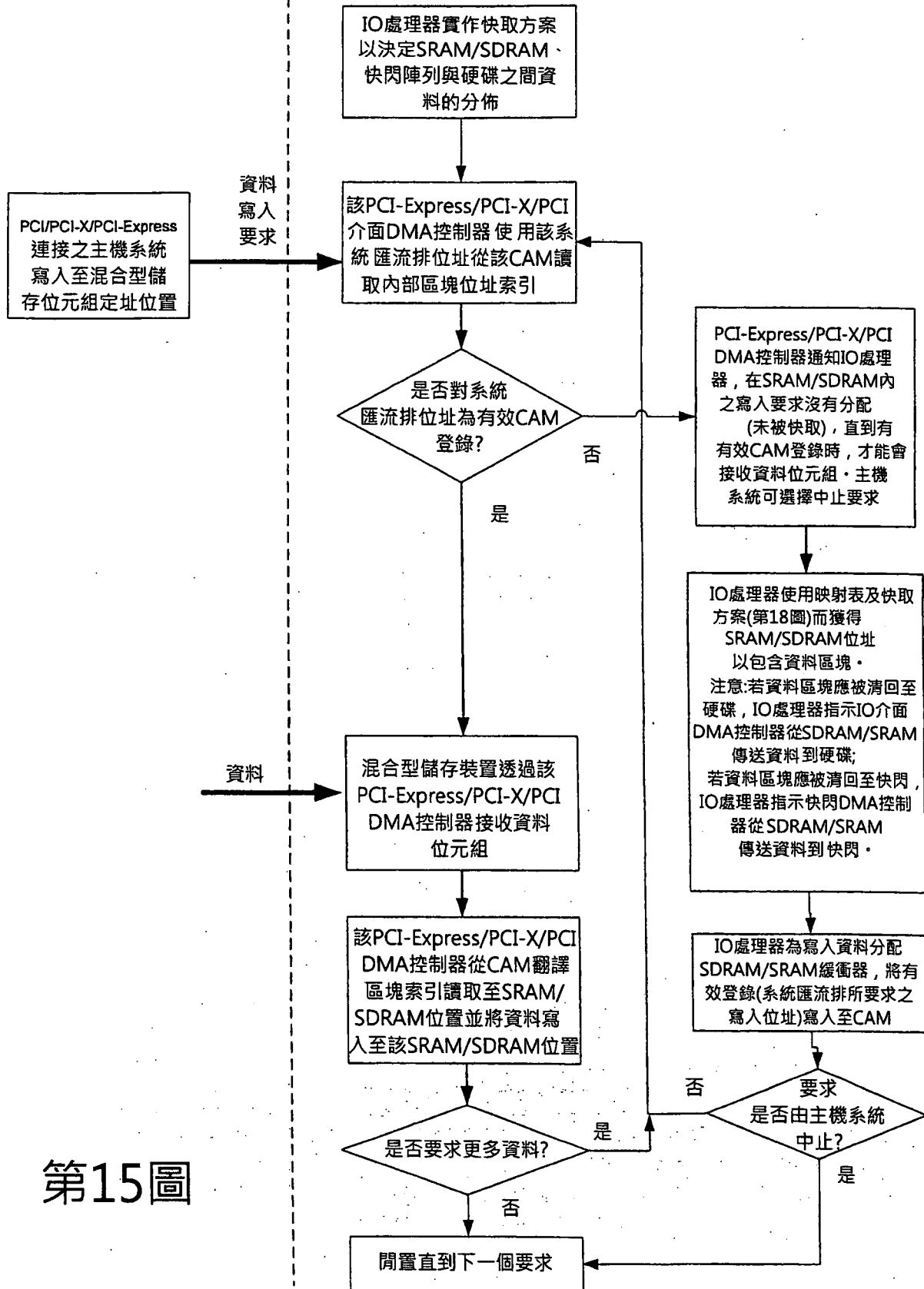
PCI/PCI-X/PCI-Express
連接之主機系統從
混合型儲存位元組
定址位置讀取



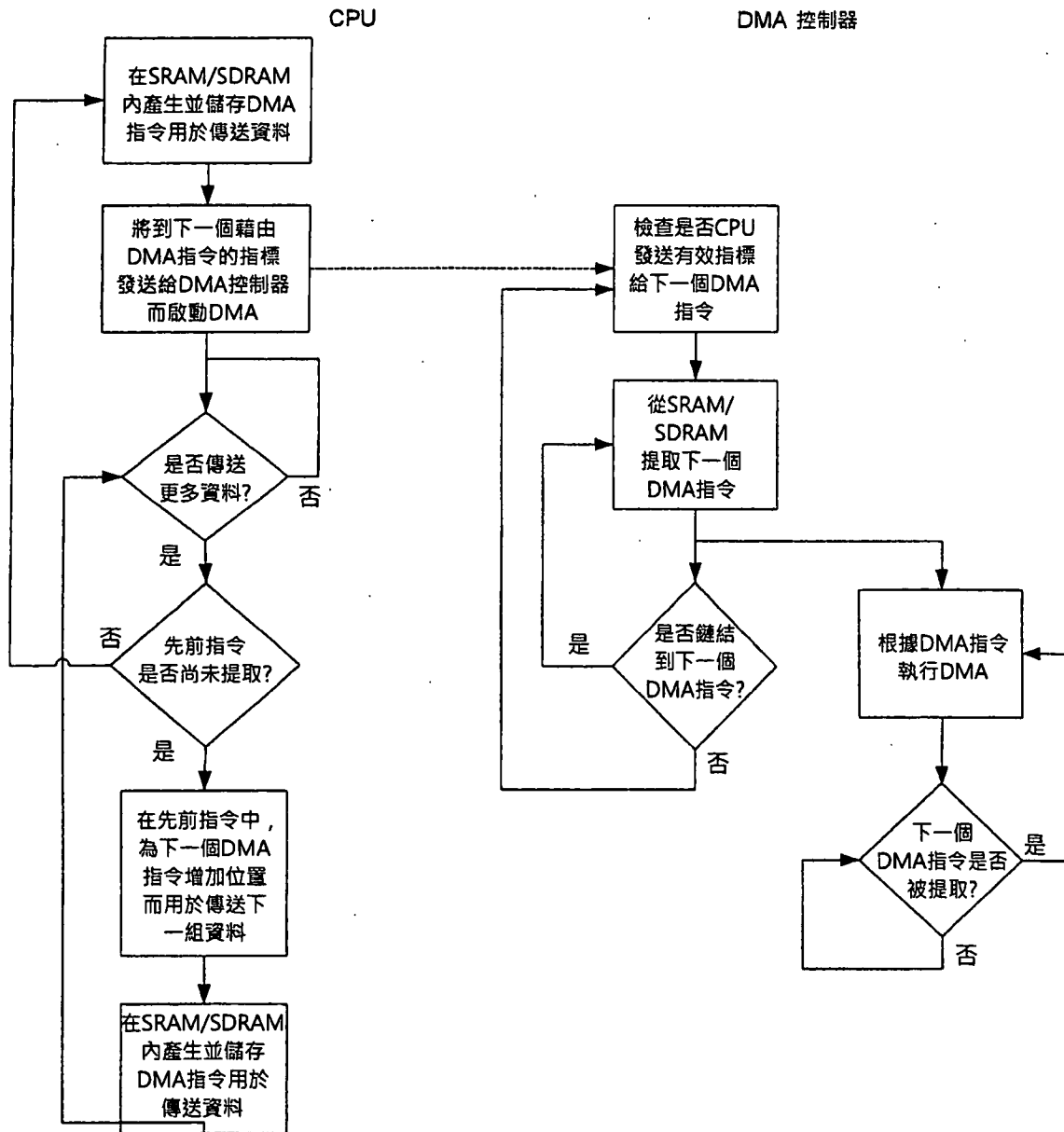
第14圖

主機系統

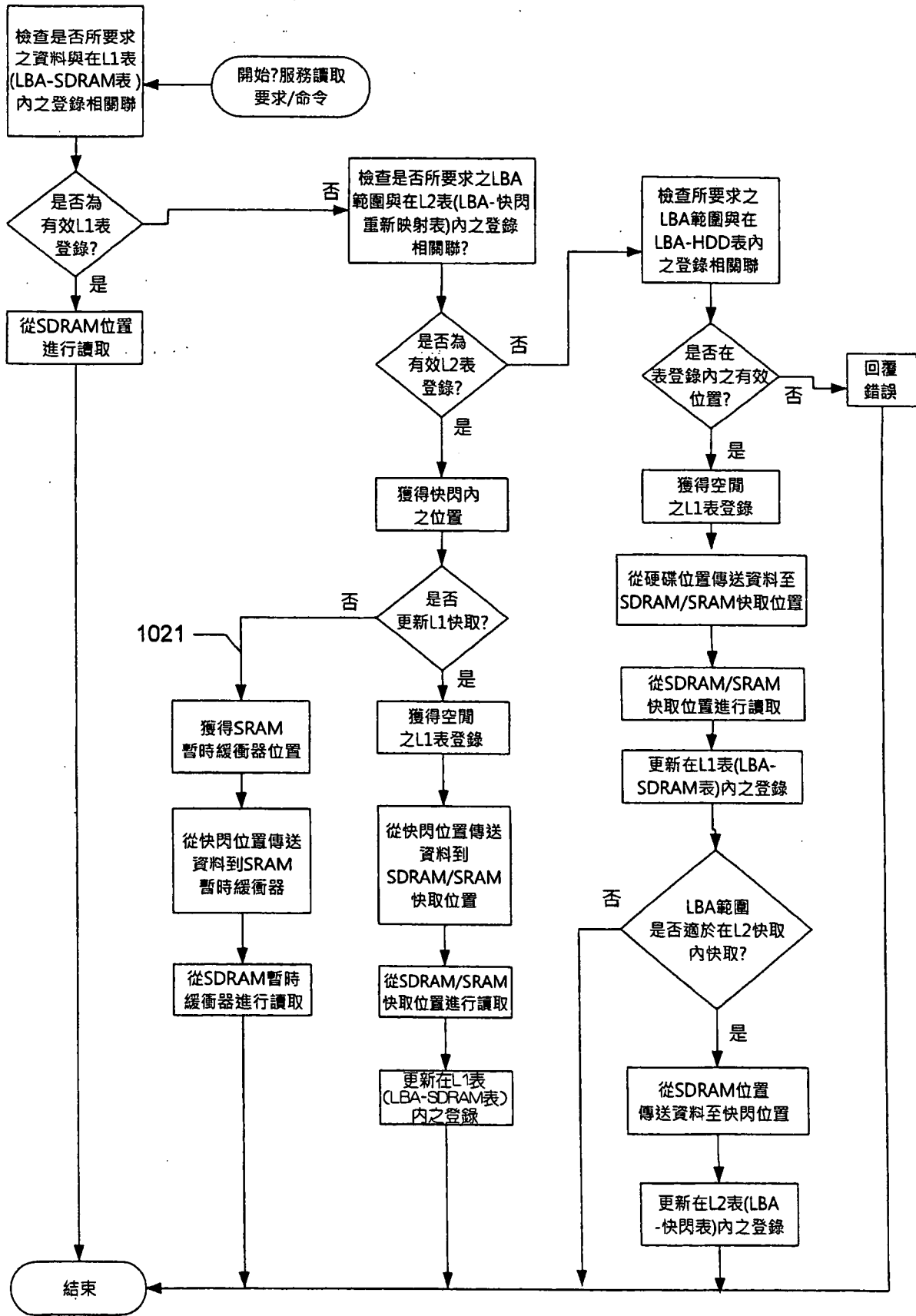
混合型儲存裝置



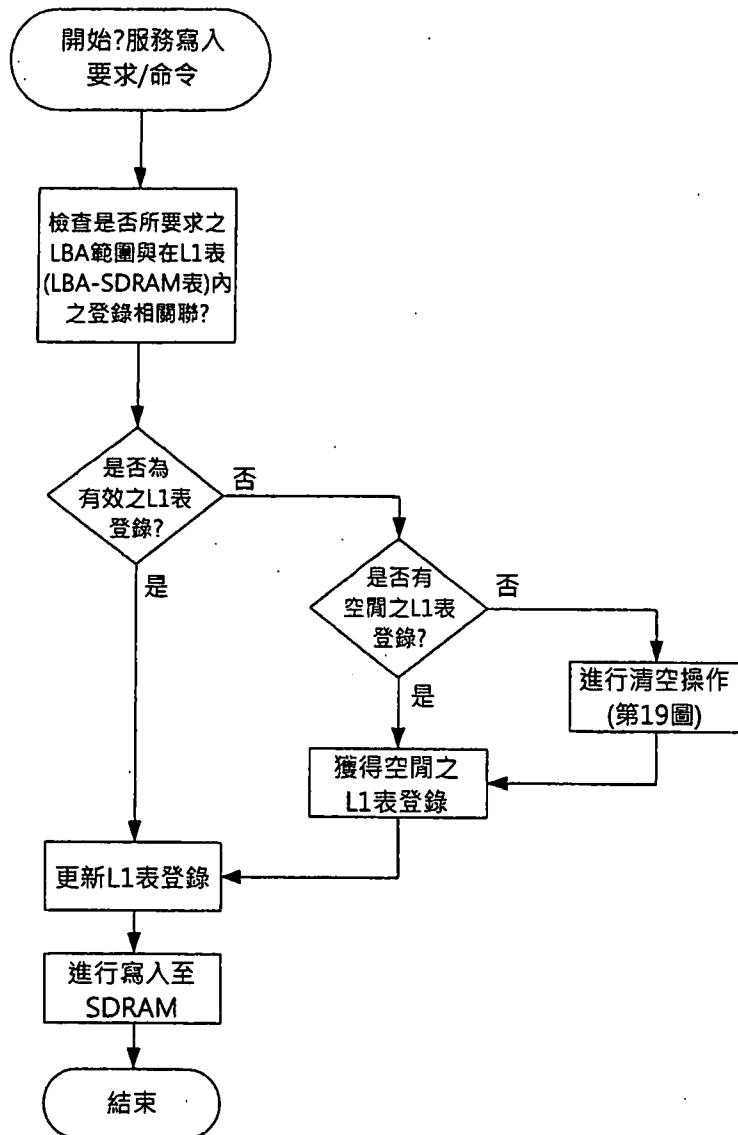
第15圖



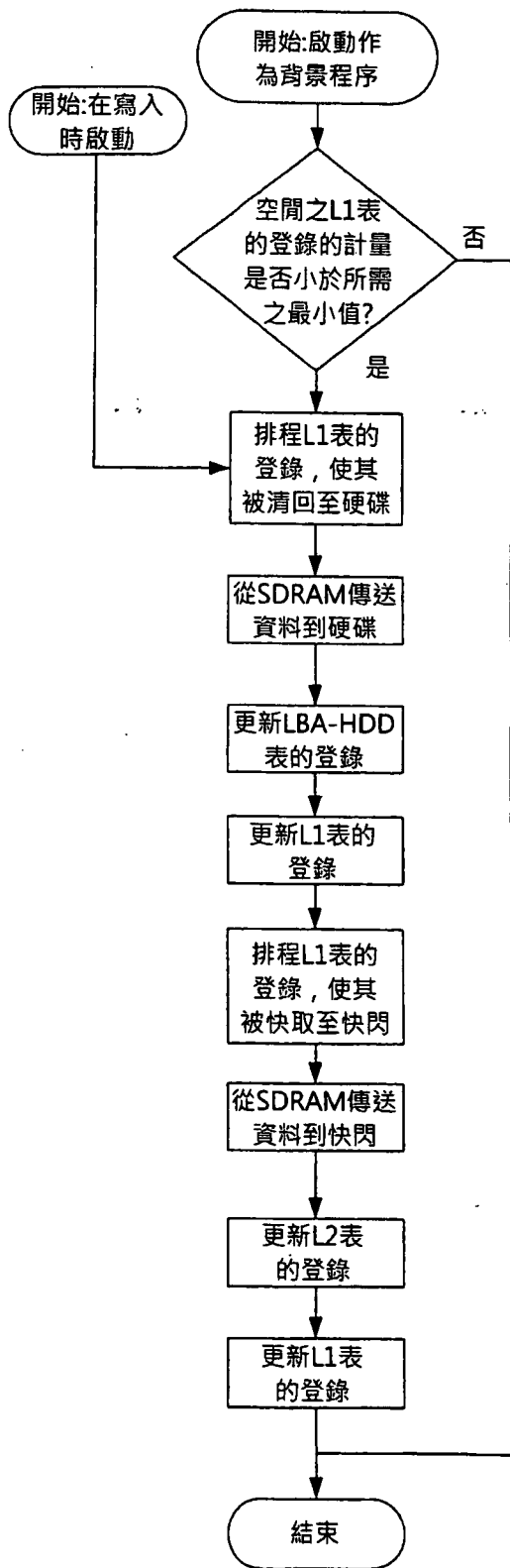
第16圖



第17圖



第18圖



第19圖