



(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 I738737 B

(45)公告日：中華民國 110 (2021) 年 09 月 11 日

(21)申請案號：106107679

(22)申請日：中華民國 106 (2017) 年 03 月 09 日

(51)Int. Cl. : G06F12/08 (2016.01)

G06F9/44 (2018.01)

G06F9/22 (2006.01)

(30)優先權：2016/04/22 美國

62/326,537

2016/10/24 美國

15/333,010

(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：古素拉 馬諾基 K GUTHULA, MANOJ K (IN)；葛拉普迪 文卡塔 巴努 普拉
卡什 GOLLAPUDI, VENKATA BHANU PRAKASH (IN)

(74)代理人：林孟閔；盧珮君；陳怡如

(56)參考文獻：

US 7979645B2

US 9092426B1

US 9256474B2

US 2015/0095610A1

審查人員：易昶霈

申請專利範圍項數：10 項 圖式數：3 共 19 頁

(54)名稱

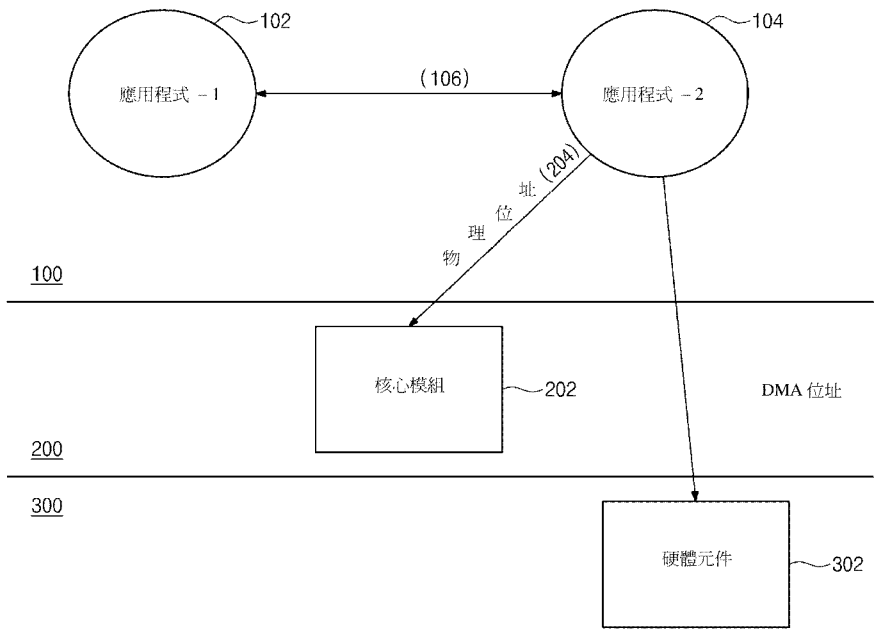
使用伴隨記憶體預配置的緩衝器映射方案的電腦實施方法及控制資料傳遞的裝置

(57)摘要

本發明概念是有關於使應用程式、作業系統、及硬體彼此通訊的由電腦實施的方法。所述方法使所述應用程式將應用程式層階虛擬位址轉換成物理位址並將所述物理位址傳送至作業系統。所述作業系統接著利用所述物理位址來確定作業系統層階虛擬位址並完成資料傳遞。

The inventive concept pertains to a computer-implemented method by which an application, an operating system, and hardware communicate with one another. The method entails the application converting an application-level virtual address to a physical address and communicating the physical address to the operating system. The operating system then uses the physical address to determine OS-level virtual address and complete the data transfer.

指定代表圖：



符號簡單說明：

100 . . . 應用程式/
應用程式使用者空間

102 . . . 應用程式-1

104 . . . 應用程式-2

106 . . . 偏移量

200 . . . 作業系統/
核心

202 . . . 核心模組

204 . . . 物理位址

300 . . . 硬體

302 . . . 硬體元件

【圖3】



申請日: 106年3月9日

I738737

【發明摘要】

IPC分類: G06F 12/08 (2016.01)
G06F 9/44 (2018.01)
G06F 9/22 (2006.01)

【中文發明名稱】 使用伴隨記憶體預配置的緩衝器映射方案的電腦實施方法及控制資料傳遞的裝置

【英文發明名稱】 COMPUTER-IMPLEMENTED METHOD USING BUFFER MAPPING SCHEME INVOLVING PRE-ALLOCATION OF MEMORY AND APPARATUS FOR CONTROLLING DATA TRANSFER

【中文】 本發明概念是有關於使應用程式、作業系統、及硬體彼此通訊的由電腦實施的方法。所述方法使所述應用程式將應用程式層階虛擬位址轉換成物理位址並將所述物理位址傳送至作業系統。所述作業系統接著利用所述物理位址來確定作業系統層階虛擬位址並完成資料傳遞。

【英文】 The inventive concept pertains to a computer-implemented method by which an application, an operating system, and hardware communicate with one another. The method entails the application converting an application-level virtual address to a physical address and communicating the physical address to the operating system. The operating system then uses the physical address to determine OS-level virtual address and complete the data transfer.

【指定代表圖】 圖3。

【代表圖之符號簡單說明】

100：應用程式/應用程式使用者空間

102：應用程式-1

104：應用程式-2

106：偏移量

200：作業系統/核心

202：核心模組

204：物理位址

300：硬體

302：硬體元件

【特徵化學式】

無

【發明說明書】

【中文發明名稱】 使用伴隨記憶體預配置的緩衝器映射方案的電腦實施方法及控制資料傳遞的裝置

【英文發明名稱】 COMPUTER-IMPLEMENTED METHOD USING BUFFER MAPPING SCHEME INVOLVING PRE-ALLOCATION OF MEMORY AND APPARATUS FOR CONTROLLING DATA TRANSFER

[相關申請案的交叉參考]

【0001】 本申請案主張於 2016 年 4 月 22 日提出申請的美國臨時申請案第 62/326,537 號的權利，所述美國臨時申請案併入本案供參考。

【技術領域】

【0002】 本揭露大體而言是有關於記憶體緩衝器，且更具體而言是有關於一種伴隨記憶體預配置的緩衝器映射方案。

【先前技術】

【0003】 在以 UNIX 為基礎的儲存器/伺服器系統中，存在各自執行特定任務的各種應用程式及元件驅動器。為了使應用程式、作業系統（operating system，OS）/核心、及硬體有效地進行通訊，它們常會圍繞記憶體緩衝器進行傳遞。通常，在該些通訊期間，應用程式會將自己的應用程式層階虛擬位址傳送至作業系統/核心。記憶體緩衝器利用應用程式層階虛擬位址來調用驅動器，且驅動器將應用程式層階虛擬位址映射至作業系統/核心層階虛擬位

址。

【0004】 為使此種轉譯更加容易，虛擬記憶體及物理記憶體被劃分成大小方便的組塊（被稱作頁面（page））。在此種分頁的模型中，虛擬位址是由偏移量（offset）及虛擬頁面訊框號（virtual page frame number）構成。每當處理器遇到虛擬位址時，處理器會自虛擬位址提取偏移量及虛擬頁面訊框號。處理器接著會將虛擬頁面訊框號轉譯成物理頁面訊框號以存取在此一物理頁面中具有正確偏移量的位置。為了將虛擬位址轉譯成物理位址，處理器會首先算出虛擬位址頁面訊框號及在此一虛擬頁面內的偏移量。處理器利用虛擬頁面訊框號作為處理頁面表的索引以擷取所述處理頁面表的頁面表項。若處於所述偏移量的頁面表項有效，則處理器會自此表項取出物理頁面訊框號。被處理器用來將虛擬頁面訊框號轉換成物理頁面訊框號的表被稱作頁面表（page table）。

【0005】 虛擬位址是藉由將偏移量加至虛擬頁面號來計算。為了進一步加強保護，對於使用者空間應用程式及核心而言存在單獨的頁面表。為對使用者空間虛擬位址進行存取，核心層階軟體將使用者空間位址映射至核心位址空間。此過程涉及到為使用者空間位址創建核心頁面表項。

【0006】 至於硬體而言，作業系統/核心與硬體之間的連接是藉由直接記憶體存取（direct memory access，DMA）而發生。藉由利用直接記憶體存取，硬體元件可在無中央處理單元（central processing unit，CPU）參與的情況下往來於電腦的主記憶體而傳

遞資料。為使直接記憶體存取發揮作用，記憶體緩衝器被頻繁地映射至對硬體元件而言可見的位址範圍。此位址範圍被稱作輸入輸出（input/output，IO）虛擬位址。視架構而定，此可涉及在輸入輸出虛擬位址與電腦主記憶體的物理位址之間設置轉譯。通常，此會在利用輸入輸出記憶體管理單元（IO memory management unit，IOMMU）硬體時發生。對於某些架構，輸入輸出虛擬位址可能與電腦主記憶體的物理位址相同。

【0007】 上述映射方案對作業系統/核心造成很重的負擔，作業系統/核心需要首先藉由設置頁面表項而將應用程式層階虛擬位址轉譯成作業系統層階虛擬位址。相似地，應為每一次直接記憶體存取傳遞建立直接記憶體存取映射。需要一種使作業系統、應用程式、及硬體更高效地進行通訊的方法。

【發明內容】

【0008】 在一個態樣中，本發明概念是有關於一種使作業系統與應用程式進行通訊的由電腦實施的方法。所述方法涉及自應用程式直接接收物理位址的作業系統以及基於記憶體的所述物理位址來確定作業系統層階虛擬位址。

【0009】 在另一態樣中，本發明概念是有關於一種使單一節點中的應用程式、作業系統、及硬體彼此通訊的由電腦實施的方法。所述方法使所述應用程式將應用程式層階虛擬位址轉換成物理位址，並將所述物理位址傳送至所述作業系統。所述作業系統接著利用所述物理位址確定作業系統層階虛擬位址。

【0010】 在又一態樣中，本發明概念是有關於一種控制資料傳遞的裝置，所述裝置包括記憶體映射器，所述記憶體映射器使應用程式能夠將所述應用程式的應用程式層階虛擬位址轉換成物理位址並將所述物理位址傳送至作業系統。

【圖式簡單說明】

【0011】

圖 1 是根據一個實施例的在單一節點中的提供一個可能環境的使用者空間、核心空間、及硬體的概念圖。

圖 2A 是示出根據一個實施例的應用程式、作業系統、及硬體之間的通訊的示意圖。

圖 2B 是根據一個實施例的另一示意圖，其示出使用者空間中的多個應用程式及核心中的不同虛擬位址指向共用物理位址進而達成緩衝器共享。

圖 3 是說明根據一個實施例的應用程式、作業系統、及硬體之間的通訊方法的圖。

【實施方式】

【0012】 本系統使得作業系統不必每當通過記憶體緩衝器時便設置基於核心層階頁面表的轉譯。在本揭露中，應用程式將物理位址傳送至核心。根據一個實施例，所述核心具有此緩衝器所需的映射。因此，核心可計算虛擬位址且無需每次均執行映射操作。由於所有核心模組共享同一虛擬位址空間，因此任何作業系統模組（不僅是對記憶體進行配置的作業系統模組）皆可利用物理位

址獲得虛擬位址並對緩衝器進行操作。

【0013】 不同應用程式之間利用緩衝器偏移量進行通訊。應用程式利用虛擬位址對緩衝器進行操作。所述應用程式可藉由僅將偏移量加至緩衝器起始處的虛擬位址來計算其有的虛擬位址。

【0014】 應用程式可藉由僅將偏移量加至緩衝器起始處的直接記憶體存取位址而確定所述偏移量的直接記憶體存取位置。應用程式可在無核心參與的情況下將緩衝器位址直接傳遞至硬體元件。

【0015】 儘管本揭露是在單一節點的上下文中書寫，然而此並非是對揭露的限制。

【0016】 圖 1 是根據一個實施例在單一節點中的提供一個可能環境的使用者空間、核心空間、及硬體的概念圖。如圖所示，形成使用者空間的應用程式 100、作業系統（OS）/核心 200、及硬體 300 彼此通訊以接收並執行使用者請求。硬體 300 包括各種元件、中央處理單元、及系統記憶體。作業系統 200 在使用者空間與硬體 300 之間進行介接，並使得應用程式 100 能夠存取系統記憶體等。元件驅動器通常是作業系統 200 的一部分。記憶體映射器將影像檔案及資料檔案映射至使用者空間中的應用程式中。在記憶體映射中，檔案的內容被鏈結至虛擬位址。

【0017】 圖 2A 是示出根據一個實施例的應用程式、作業系統、及硬體之間的通訊的示意圖。就應用程式 100 而言，示出有兩個應用程式（應用程式 X 及應用程式 Y）。所述兩個應用程式中的每一者均具有其自己的虛擬記憶體及其自己的虛擬位址集合（在圖 2A

中示作 VM-X 及 VM-Y)。每一應用程式亦具有其自己的頁面表 110，頁面表 110 將所述每一應用程式自己的相應虛擬頁面映射成記憶體中的物理頁面。舉例而言，如圖所繪示，將應用程式 X 的虛擬頁面訊框號 0 (VPFN 0) 映射成記憶體中的物理頁面訊框號 1 (PFN1)，且將應用程式 Y 的虛擬頁面訊框號 1 (VPFN1) 映射成記憶體中的物理頁面訊框號 4 (PFN4)。

【0018】 利用虛擬頁面訊框號作為偏移量對頁面表 110 進行存取。為了將虛擬位址轉譯成物理位址，會首先確定虛擬位址頁面訊框號及在虛擬頁面內的偏移量。若虛擬記憶體位址有效且表項有效，則處理器會取出物理頁面訊框號並將其乘以頁面大小以得到物理記憶體中的頁面的基點 (base) 的位址。接著，加上所述偏移量。

【0019】 舉例而言，在圖 2A 所說明的情形中，假定頁面大小為 0×2000 。對於 VM-Y 中為 0×2194 的位址而言，處理器會將此一位址轉譯成 0×194 的偏移量且在虛擬頁面訊框號 1 中。此虛擬頁面訊框號 1 被映射至以 0×8000 (4×2000) 開始的物理頁面訊框號 4。將 0×194 的偏移量加至物理頁面訊框號會產生為 0×8194 的最終物理位址。儘管各應用程式僅利用基點的虛擬位址及相對於虛擬位址的偏移量來彼此通訊，然而本揭露的系統使得各應用程式能夠利用物理位址來與核心進行通訊。如圖所示，核心轉譯表 210 用於將物理位址轉譯成作業系統層階虛擬位址。核心轉譯表 210 使得能夠將物理位址轉譯成虛擬位址，且核心轉譯表 210

可為作業系統專用的。

【0020】 根據一個實施例，記憶體被預配置且與應用程式 100 共享，使得應用程式 100 及作業系統 200 可均對物理位址表進行存取。本文所述「預配置」意指在對緩衝器進行任何使用以在應用程式/核心/硬體域之間傳遞資料之前進行的配置。此外，作業系統 200 中的不同模組（其中所有模組共享同一虛擬位址空間）將物理位址轉換成所述作業系統自己的作業系統層階虛擬位址。被每一作業系統用來將物理位址轉譯成虛擬位址的方法取決於每一作業系統的架構。舉例而言，Linux 作業系統可針對某些位址範圍利用簡單算術法將物理位址轉譯成虛擬位址。當實作於 Linux 中時，本揭露系統的經預配置的緩衝器落於其中利用簡單算術法來達到物理位址的此位址範圍中。某些其他作業系統可以不同的機制來達成此結果。

【0021】 應用程式可藉由僅將偏移量加至緩衝器起始處的直接記憶體存取位址來計算所述偏移量的直接記憶體存取位址。在此方式，應用程式可在無作業系統 200 參與的情況下將緩衝器位址直接傳遞至硬體 300 元件。

【0022】 圖 2B 是根據一個實施例的另一示意圖，其示出使用者空間中的多個應用程式及核心中的不同虛擬位址指向同一物理位址進而達成緩衝器共享。圖 2B 示出應用程式使用者空間 100 中的應用程式 X 及應用程式 Y。在應用程式 X 中，被標記為「緩衝器-1」的資料儲存於應用程式 X 位址空間中的 0×3000 處。此資料例如

106-07-03

利用圖 2A 中所示處理而轉譯至核心位址空間 0×5000 。同一資料（緩衝器-1）對應於應用程式-Y 位址空間中的位址 0×1000 處的資料，但應用程式 X 與應用程式 Y 二者均能夠利用物理位址而指向同一資料。被標記為「緩衝器-2」的資料儲存於應用程式 Y 的虛擬位址 0×4000 處且對應於核心位址空間 0×7000 。由於各核心模組共享同一虛擬位址空間此一事實，任何作業系統模組均可利用圖 2B 中所示核心位址來獲取虛擬位址。

【0023】圖 3 是說明根據一個實施例的應用程式、作業系統、及硬體之間的通訊方法的圖。更具體而言，圖 3 所示實施例示出使用者空間（應用程式 100）中的應用程式-1 102 與應用程式-2 104 利用偏移量 106 彼此通訊、以及應用程式-2 104 利用物理位址 204 而與核心模組 202 進行通訊。所述應用程式（在此實例中為應用程式-2 104）亦可利用藉由利用偏移量而產生的直接記憶體存取位址而與硬體元件 302 直接通訊。

【0024】根據一個實施例，本揭露的系統包括一種其儲存有電腦程式的機器可讀取儲存器（machine-readable storage），所述電腦程式具有可由機器執行的至少一個碼段（code section），藉此使所述機器執行如上所述的各步驟。

【0025】根據一個實施例，本揭露的系統可達成於硬體、軟體、或硬體與軟體的組合中。儘管本揭露著重於涉及一個電腦系統的單一節點實施方案，然而本揭露亦可適用於其中不同組件橫跨多個互連的電腦系統的分佈式方式。適於施行本文所述方法的任何

種類的電腦系統或裝置均適合。硬體與軟體的典型組合可為具有電腦程式的通用電腦系統，所述電腦程式在被加載及執行時控制電腦系統以使所述電腦系統施行本文所述方法。

【0026】 本揭露的系統可嵌置於電腦程式產品中，所述電腦程式產品包括能夠達成本文所述方法的所有特徵且在被加載於電腦系統中時能夠施行該些方法。在本上下文中，「電腦程式」意指旨在使具有資訊處理能力的系統直接執行特定功能或在以下操作中的一者或兩者之後執行特定功能的呈現任何語言、代碼、或記法（notation）的表達形式的一組指令：轉換成另一語言、代碼、或記法；以不同材料形式進行再製（reproduction）。

【0027】 儘管已參照某些實施例闡述了本揭露，然而熟習此項技術者應理解，可作出各種改變且可對等效形式進行替代，而此並不背離發明的範圍。另外，可在不背離本揭露的範圍的條件下作出諸多潤飾以使特定情景或材料適應於本揭露的教示內容。因此，旨在使本揭露不僅限於所揭露的特定實施例，而是本揭露將包括落於隨附申請專利範圍的範圍內的所有實施例。

【符號說明】

【0028】

100：應用程式/應用程式使用者空間

102：應用程式-1

104：應用程式-2

106：偏移量

110：頁面表

200：作業系統/核心

202：核心模組

204：物理位址

210：核心轉譯表

300：硬體

302：硬體元件

VM-X、VM-Y：虛擬位址集合

【發明申請專利範圍】

【第1項】 一種使作業系統與第一應用程式及第二應用程式進行通訊的由電腦實施的方法，所述方法包括：

直接自所述第一應用程式接收第一物理位址，其中所述第一應用程式具有第一頁面表，並且藉由利用所述第一頁面表將所述第一應用程式的虛擬頁面映射至記憶體的物理頁面來確定所述第一物理位址，以及利用第一物理頁面訊框號及第一偏移量來確定與所述第一應用程式對應的所述第一物理位址；

利用應用程式偏移量使所述第二應用程式與所述第一應用程式通訊，其中所述應用程式偏移量被應用於所述第一物理位址來確定與所述第二應用程式對應的第二物理位址；以及

基於所述第一物理位址確定作業系統層階（OS 層階）虛擬位址，以達成資料傳遞，其中僅基於所述第一物理位址藉由核心轉譯表確定所述作業系統層階虛擬位址，且其中所述作業系統僅通過所述第一物理位址與所述第一應用程式及所述第二應用程式進行通訊。

【第2項】 如申請專利範圍第1項所述的方法，更包括：

在所述第一應用程式或所述第二應用程式接收使用者輸入之前，執行記憶體配置並與所述第一應用程式及所述第二應用程式共享所述記憶體配置。

【第3項】 如申請專利範圍第1項所述的方法，其中在所述作業系統中存在多個模組，所述模組中的每一者均具有其自己的作業系

統層階虛擬記憶體，所述方法更包括：容許所有所述模組利用所述第一物理位址而與所述第一應用程式直接通訊。

【第4項】 一種使單一節點中的第一應用程式、第二應用程式、作業系統、及硬體利用緩衝器而彼此通訊的由電腦實施的方法，所述方法包括：

所述第一應用程式將第一應用程式層階虛擬位址轉換成與所述第一應用程式對應的第一物理位址，並將所述第一物理位址傳送至所述作業系統；

利用應用程式偏移量使所述第二應用程式與所述第一應用程式通訊，其中所述應用程式偏移量被應用於所述第一物理位址來確定與所述第二應用程式對應的第二物理位址；以及

所述作業系統利用所述第一物理位址確定作業系統層階虛擬位址，其中僅基於所述第一物理位址藉由核心轉譯表確定所述作業系統層階虛擬位址，且其中所述作業系統僅通過所述第一物理位址與所述第一應用程式及所述第二應用程式進行通訊。

【第5項】 如申請專利範圍第4項所述的方法，其中所述第一應用程式利用直接記憶體存取（DMA）位址在無所述作業系統參與的情況下與所述硬體進行通訊，所述直接記憶體存取位址是利用硬體偏移量計算出。

【第6項】 如申請專利範圍第4項所述的方法，其中所述作業系統在接收使用者輸入之前預配置記憶體緩衝器並提供將物理位址映射至應用程式的途徑。

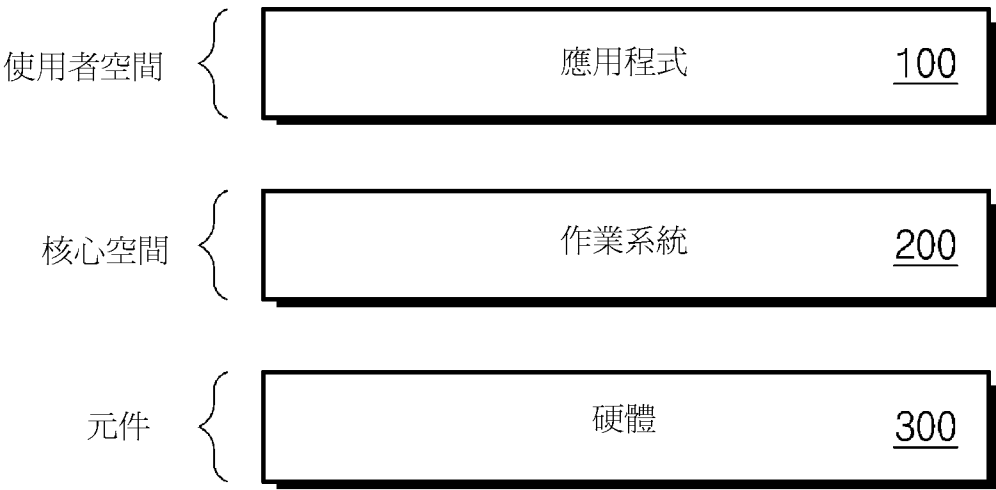
【第7項】 如申請專利範圍第4項所述的方法，其中在利用所述緩衝器在所述第一應用程式、第二應用程式、所述作業系統、及所述硬體之間傳遞資料之前，所述第一應用程式得到記憶體的所述第一物理位址及直接記憶體存取位址。

【第8項】 如申請專利範圍第4項所述的方法，其中在所述作業系統中存在多個模組，且其中所述模組中的每一者利用所述第一物理位址而與所述第一應用程式直接通訊並確定自己的作業系統層階虛擬位址。

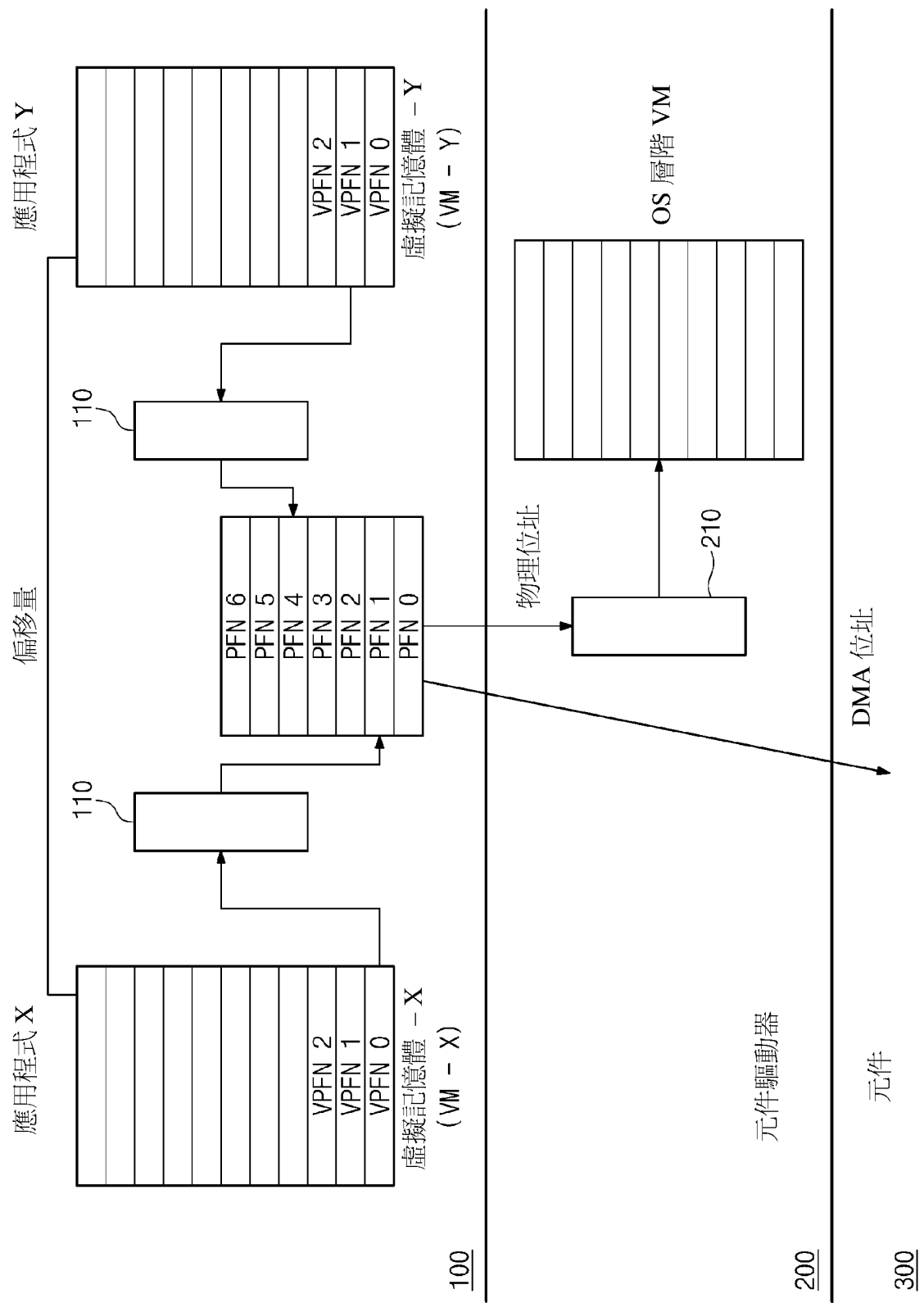
【第9項】 一種控制資料傳遞的裝置，所述裝置包括記憶體映射器，所述記憶體映射器使第一應用程式能夠將第一應用程式層階虛擬位址轉換成與所述第一應用程式對應的第一物理位址並將所述第一物理位址傳送至作業系統，並且利用應用程式偏移量使第二應用程式與所述第一應用程式通訊，其中所述應用程式偏移量被應用於所述第一物理位址來確定與所述第二應用程式對應的第二物理位址，且其中核心轉譯表用以僅基於所述第一物理位址確定作業系統層階虛擬位址，且其中所述作業系統僅通過所述第一物理位址與所述第一應用程式及所述第二應用程式進行通訊。

【第10項】 如申請專利範圍第9項所述的裝置，其中所述記憶體映射器使所述第一應用程式能夠利用直接記憶體存取位址及硬體偏移量來與硬體直接通訊。

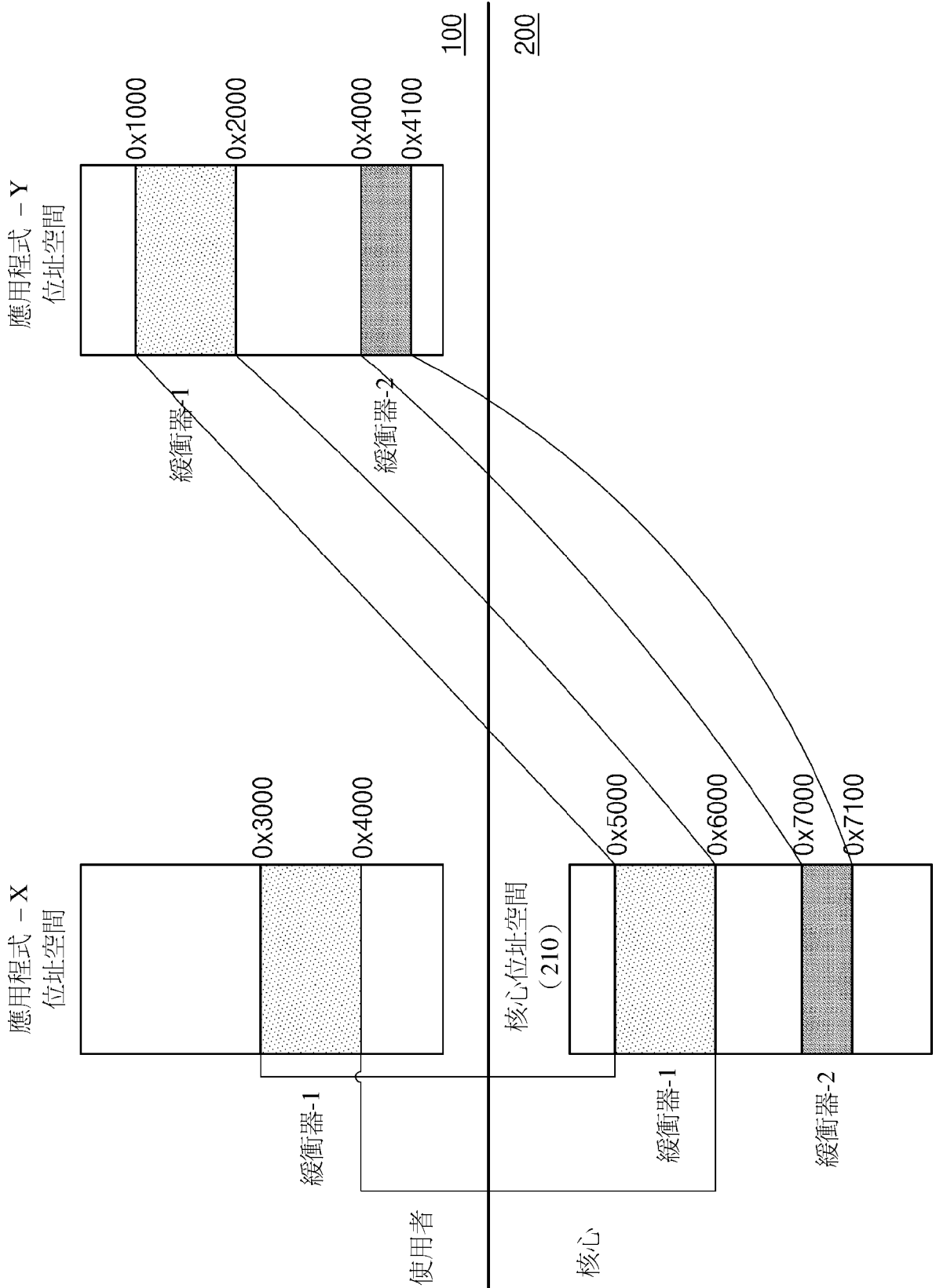
【發明圖式】



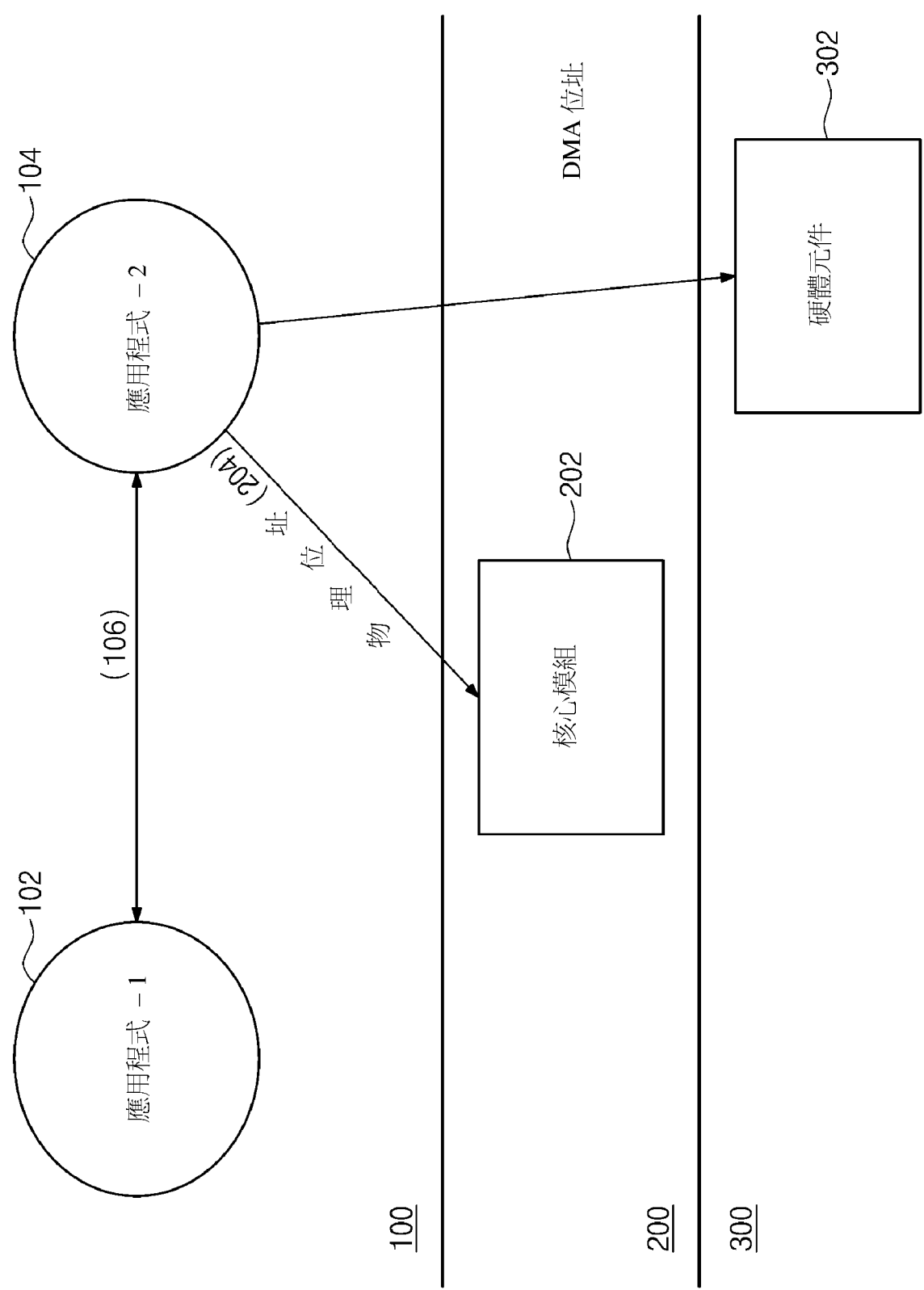
【圖1】



【圖2A】



【圖2B】



【圖3】