



(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 I526832 B

(45)公告日：中華民國 105 (2016) 年 03 月 21 日

(21)申請案號：103107028

(22)申請日：中華民國 103 (2014) 年 03 月 03 日

(51)Int. Cl. : G06F12/10 (2006.01)

(30)優先權：2013/03/05 美國 13/785,877

(71)申請人：高通公司(美國) QUALCOMM INCORPORATED (US)
美國

(72)發明人：曾湯瑪士 ZENG, THOMAS (US)；圖茲尼艾茲狄恩 TOUZNI, AZZEDINE (FR)；曾宗仁 TZENG, TZUNG REN (US)；波士黎斐爾 J BOSTLEY, PHIL J. (US)

(74)代理人：李世章

(56)參考文獻：

TW I275102

TW I307839

US 7386669B2

US 7434027B2

Barr et al. "pecTLB: A Mechanism for speculative Address Translation" ISCA'11, June 4-8, 2011, San Jose, California, USA 2011, p307-318

審查人員：何旭智

申請專利範圍項數：19 項 圖式數：6 共 35 頁

(54)名稱

用於減少執行硬體表搜尋 (HWTW) 所需的時間和計算資源量的方法和系統

METHODS AND SYSTEMS FOR REDUCING THE AMOUNT OF TIME AND COMPUTING RESOURCES THAT ARE REQUIRED TO PERFORM A HARDWARE TABLE WALK (HWTW)

(57)摘要

提供了在發生轉譯後備緩衝器(TLB)失敗的情況下減少用於執行硬體表搜尋(HWTW)所需的時間和計算資源量的電腦系統和方法。若在執行階段 2(S2)HWTW 以尋找儲存階段 1(S1)頁表的 PA 時發生 TLB 失敗，則 MMU 使用 IPA 來預測對應的 PA，由此避免執行任何 S2 查表的需要。此舉極大地降低了在執行該等類型的 HWTW 讀事務時需要執行的檢視的數目，其極大地減少了與執行該等類型的事務相關聯的處理管理負擔和效能懲罰。

A computer system and a method are provided that reduce the amount of time and computing resources that are required to perform a hardware table walk (HWTW) in the event that a translation lookaside buffer (TLB) miss occurs. If a TLB miss occurs when performing a stage 2 (S2) HWTW to find the PA at which a stage 1 (S1) page table is stored, the MMU uses the IP A to predict the corresponding PA, thereby avoiding the need to perform any of the S2 table lookups. This greatly reduces the number of lookups that need to be performed when performing these types of HWTW read transactions, which greatly reduces processing overhead and performance penalties associated with performing these types of transactions.

指定代表圖：

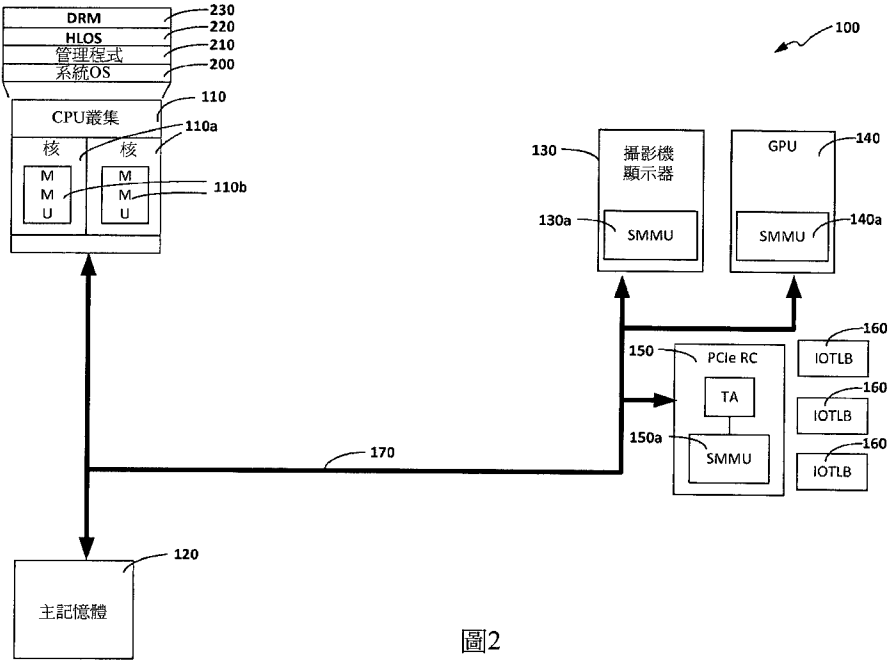


圖2

符號簡單說明：

- 100 . . . 電腦系統
- 110 . . . CPU 叢集
- 110a . . . CPU 核
- 110b . . . MMU
- 120 . . . 主記憶體
- 130 . . . 攝影機顯示器
- 130a . . . 系統 MMU(SMMU)
- 140 . . . 圖形處理單元(GPU)
- 140a . . . SMMU
- 150 . . . 快速周邊連接介面(PCIe)輸入/輸出(IO)設備
- 150a . . . SMMU
- 160 . . . IO TLB
- 170 . . . 系統匯流排
- 200 . . . 系統 OS
- 210 . . . 管理程式
- 220 . . . 訪客 HLOS
- 230 . . . 訪客數位版權管理器(DRM)

發明摘要

※ 申請案號：103107028

※ 申請日：2014 年 3 月 3 日

※IPC 分類：G06F 17/10 (2006.01)

【發明名稱】（中文/英文）

用於減少執行硬體表搜尋（HWTW）所需的時間和計算資源量的方法和系統

METHODS AND SYSTEMS FOR REDUCING THE AMOUNT OF TIME AND COMPUTING RESOURCES THAT ARE REQUIRED TO PERFORM A HARDWARE TABLE WALK (HWTW)

【中文】

提供了在發生轉譯後備緩衝器（TLB）失敗的情況下減少用於執行硬體表搜尋（HWTW）所需的時間和計算資源量的電腦系統和方法。若在執行階段2（S2）HWTW以尋找儲存階段1（S1）頁表的PA時發生TLB失敗，則MMU使用IPA來預測對應的PA，由此避免執行任何S2查表的需要。此舉極大地降低了在執行該等類型的HWTW讀事務時需要執行的檢視的數目，其極大地減少了與執行該等類型的事務相關聯的處理管理負擔和效能懲罰。

【英文】

A computer system and a method are provided that reduce the amount of time and computing resources that are required to perform a hardware table walk (HWTW) in the

event that a translation lookaside buffer (TLB) miss occurs. If a TLB miss occurs when performing a stage 2 (S2) HWTW to find the PA at which a stage 1 (SI) page table is stored, the MMU uses the IP A to predict the corresponding PA, thereby avoiding the need to perform any of the S2 table lookups. This greatly reduces the number of lookups that need to be performed when performing these types of HWTW read transactions, which greatly reduces processing overhead and performance penalties associated with performing these types of transactions.

【代表圖】

【本案指定代表圖】：第（ 2 ）圖。

【本代表圖之符號簡單說明】：

100 電腦系統

110 CPU 叢集

110a CPU 核

110b MMU

120 主記憶體

130 攝影機顯示器

130a 系統 MMU (SMMU)

140 圖形處理單元 (GPU)

140a SMMU

150 快速周邊連接介面 (PCIe) 輸入/輸出 (IO) 設備

150a SMMU

160 IO TLB

170 系統匯流排

200 系統 OS

210 管理程式

220 訪客 HLOS

230 訪客數位版權管理器 (DRM)

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

無

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)

用於減少執行硬體表搜尋 (HWTW) 所需的時間和計算資源量的方法和系統

METHODS AND SYSTEMS FOR REDUCING THE AMOUNT OF TIME AND COMPUTING RESOURCES THAT ARE REQUIRED TO PERFORM A HARDWARE TABLE WALK (HWTW)

【技術領域】

【0001】本發明係關於電腦系統，尤其係關於用於減少執行硬體表搜尋 (HWTW) 所需的時間和計算資源量的電腦系統和在電腦系統中使用的方法。

【先前技術】

【0002】現代電腦系統使用記憶體管理單元 (MMU) 來管理向一或多個實體記憶體設備 (諸如舉例而言固態記憶體設備) 寫入資料和從其讀取資料。電腦系統的MMU向該電腦系統的中央處理單元 (CPU) 提供虛擬記憶體，此允許CPU在每個應用程式自身的專用、毗連的虛擬記憶體位址空間中執行該應用程式，而非使所有的應用程式共用往往是成碎片的，或非毗連的實體記憶體位址空間。MMU的目的是為CPU將虛擬記憶體位址 (VA) 轉譯成實體記憶體位址 (PA)。CPU經由直接向MMU讀寫VA、而MMU將該等VA轉譯成PA並隨後寫或讀

PA來間接地讀寫PA。

【0003】爲了執行該等轉譯，MMU存取儲存於系統主記憶體中的頁表。頁表由頁表條目構成。頁表條目是由MMU用來將VA映射成PA的資訊。MMU通常包括轉譯後備緩衝器（TLB），該TLB是被用來快取最近使用過的映射的快取記憶體元件。當MMU需要將VA轉譯成PA時，MMU首先檢查TLB以決定是否存在關於該VA的匹配。若是，則MMU使用TLB中找到的映射來計算PA並隨後存取該PA（亦即，讀或寫該PA）。此被稱爲TLB「命中」。若MMU在TLB中未找到匹配，則此被稱作TLB「失敗」。

【0004】在TLB失敗的情形中，MMU執行所謂的硬體表搜尋（HWTW）。HWTW是耗時的且在計算上管理負擔很大的程序，其涉及執行「表搜尋（table walk）」以尋找MMU中的對應頁表並隨後讀取該頁表中的多個位置以尋找對應的VA到PA位址映射。MMU隨後使用該映射來計算對應的PA並將該映射寫回到TLB。

【0005】在實施作業系統（OS）虛擬化的電腦系統中，虛擬記憶體監視器（VMM）（通常亦被稱爲管理程式）被置於電腦系統的硬體與電腦系統的系統OS之間。管理程式以特權模式執行並且能夠主存一或多個訪客高級OS。在此類系統中，在OS上執行的應用程式使用第一層虛擬記憶體的VA來定址記憶體，並且在管理程式上執行的OS使用第二層虛擬記憶體的中間實體位址（IPA）來定址記憶體。在MMU中，執行階段1（S1）轉譯以將每個VA轉譯成IPA，並且執行階段2（S2）轉

譯以將每個IPA轉譯成PA。

【0006】若在執行此類轉譯時發生TLB失敗，則執行多級二維（2-D）HWTW以獲得用於計算對應的IPA和PA所需的表條目。執行該等多級2-D HWTW可導致MMU的相當大量的計算管理負擔，此通常導致效能懲罰。

【0007】圖1是已知的三級2-D HWTW的圖形圖示，該三級2-D HWTW在執行讀事務時發生TLB失敗時執行。圖1中示出的HWTW表示三級2-D HWTW的最壞情景，其需要執行15次查表來獲得資料被儲存在實體記憶體中的PA。對於此實例，電腦系統的MMU正執行管理程式，該管理程式主存至少一個訪客高級OS（HLOS），該HLOS進而正執行至少一個應用程式。在此類配置中，正由訪客HLOS分配的記憶體不是系統的實際實體記憶體，而是前述的中間實體記憶體。管理程式分配實際實體記憶體。因此，每個VA被轉譯成IPA，該IPA隨後被轉譯成實際實體記憶體的實際上儲存正被讀取的資料的PA。

【0008】該程序開始於MMU接收S1頁全域目錄（PGD）IPA 2。對於此最壞情景實例，將假定在MMU檢查TLB以尋找匹配時發生TLB失敗。由於此失敗，MMU必須執行HWTW。HWTW涉及執行三次S2查表3、4和5以獲得用於將IPA 2轉換成PA所需的映射並執行一次額外檢視6以讀取該PA。查表3、4和5分別涉及讀取S2 PGD、頁中間目錄（PMD）和頁表條目（PTE）。在檢視6時讀取PA向MMU提供了S1 PMD IPA 7。對於此最壞情景實例，將假定在MMU檢查TLB以尋找與S1 PMD IPA 7的匹配時發生TLB失敗。由於此失敗，MMU必須執行另一

HWTW。該HWTW涉及執行三次S2查表8、9和11以獲得用於將S1 PMD IPA 7轉換成PA所需的映射並執行一次額外檢視12以讀取該PA。查表8、9和11分別涉及讀取S2 PGD、PMD和PTE。在檢視12時讀取PA向MMU提供了S1 PTE IPA 13。

【0009】對於此最壞情景實例，將假定在MMU檢查TLB以尋找與S1 PTE IPA 13的匹配時發生TLB失敗。由於此失敗，MMU必須執行另一HWTW。該HWTW涉及執行三次S2查表14、15和16以獲得用於將S1 PTE IPA 13轉換成PA所需的映射並執行一次額外檢視17以讀取該PA。查表14、15和16分別涉及讀取S2 PGD、PMD和PTE。在檢視17時讀取PA向MMU提供了實際IPA 18。對於此最壞情景實例，將假定在MMU檢查TLB以尋找與實際IPA 18的匹配時發生TLB失敗。由於此失敗，MMU必須執行另一HWTW。該HWTW涉及執行三次S2查表19、21和22以獲得用於將實際IPA 18轉換成PA所需的映射。查表19、21和22分別涉及讀取S2 PGD、PMD和PTE。隨後讀取PA以獲得對應的讀取資料。在檢視18時讀取PA向MMU提供了S1 PTE IPA 13。

【0010】因此可以看出，在三級2-D HWTW的最壞情景中，執行了12次S2查表和3次S1查表，此是大量的計算管理負擔，其消耗了大量時間並導致效能懲罰。已使用各種技術和架構來減少在執行HWTW中所涉及的時間和處理管理負擔量，包括例如增加TLB的大小、使用多個TLB、使用平面嵌套的頁表、使用陰影分頁或推測陰影分頁、以及使用頁搜尋快取記憶體。儘管所有該等技術和架構能夠減少與執行HWTW相關聯

的處理管理負擔，但其往往在電腦系統的其他某處造成處理管理負擔的增加。

【0011】 因此，存在對減少執行HWTW所需的時間和計算資源量的電腦系統和方法的需要。

【發明內容】

【0012】 本發明針對用於減少執行HWTW所需的時間和計算資源量的電腦系統和在電腦系統中使用的方法。該電腦系統包括至少一個中央處理單元（CPU）、至少一個實體記憶體、至少一個TLB、以及至少一個MMU。CPU執行主機OS和管理程式。管理程式控制CPU上的至少第一訪客OS的執行。管理程式至少執行與第一訪客OS相關聯的第一VM。實體記憶體具有能藉由PA來定址的實體記憶體位置。至少一個頁表儲存在實體記憶體的實體記憶體位置處。頁表包括與用於將IPA映射成實體記憶體的實際PA的映射相對應的頁表條目。TLB儲存頁表條目的子集。在執行記憶體存取時，MMU決定與IPA相關聯的頁表條目是否儲存於TLB中。若與IPA相關聯的頁表條目未被儲存於TLB中，則已發生TLB失敗。若發生TLB失敗，則MMU預測實體記憶體的儲存與IPA相關聯的資料的PA，由此消除執行HWTW以計算PA的需要。

【0013】 該方法包括：

【0014】 在MMU中：

【0015】 決定與IPA相關聯的頁表條目是否儲存於TLB中；

【0016】 若決定與IPA相關聯的頁表條目未被儲存於TLB中，則判定已發生TLB失敗；及

【0017】若做出已發生TLB失敗的判定，則預測實體記憶體的儲存與IPA相關聯的資料的PA。

【0018】本發明亦提供了儲存電腦代碼的電腦可讀取媒體（CRM），該電腦代碼由一或多個處理器執行以用於減少與執行HWTW相關聯的處理管理負擔。該電腦代碼包括第一和第二代碼部分。第一代碼部分決定與IPA相關聯的頁表條目是否儲存於TLB中。若決定與該IPA相關聯的頁表條目未被儲存於TLB中，則第一代碼部分判定已發生TLB失敗。第二代碼部分在第一代碼部分判定已發生TLB失敗的情況下預測實體記憶體的儲存與該IPA相關聯的資料的PA。

【0019】該等以及其他特徵和優點從以下描述、附圖和請求項中將變得明顯。

【圖式簡單說明】

【0020】圖1是根據本發明的說明性實施例的電腦系統的方塊圖。

【0021】圖2圖示了根據說明性或示例性實施例的電腦系統的方塊圖，該電腦系統被配置成執行用於減少執行HWTW所需的時間和計算資源量的方法。

【0022】圖3是表示根據說明性實施例的由圖2中所示的管理程式執行以減少執行HWTW讀事務所需的時間和處理管理負擔量的方法的流程圖。

【0023】圖4是根據說明性實施例說明使用由圖3中所示的流程圖所表示的方法來執行HWTW讀事務的方式的圖形示圖。

【0024】圖5是根據說明性實施例的執行由圖3中所示的流程

圖表示的方法的硬體預測器的方塊圖。

【0025】圖6圖示了其中併入圖2中所示的電腦系統的行動智慧型電話的方塊圖。

【實施方式】

【0026】根據本文描述的說明性實施例，提供了用於減少執行HWTW所需的時間和計算資源量的電腦系統和在電腦系統中使用的方法。根據本文描述的實施例，當執行S2 HWTW以尋找儲存S1頁表的PA時發生TLB失敗時，MMU使用IPA來預測對應的PA，由此避免執行任何S2查表的需要。此舉極大地降低了在執行該等類型的HWTW讀事務時需要執行的檢視的數目，其極大地減少了與執行該等類型的事務相關聯的處理管理負擔和效能懲罰。

【0027】圖2圖示了根據說明性或示例性實施例的電腦系統100的方塊圖，其被配置成執行用於減少執行S2 HWTW以尋找儲存S1頁表的PA所需的時間和計算資源量的方法。圖2中所示的電腦系統100的實例包括CPU叢集110、主記憶體120、攝影機顯示器130、圖形處理單元（GPU）140、快速周邊連接介面（PCIe）輸入/輸出（IO）設備150、多個IO TLB（IOTLB）160和系統匯流排170。CPU叢集110具有多個CPU核110a，每個CPU核110a具有MMU 110b。每個CPU核110a可以是微處理器或任何其他合適的處理器。攝影機顯示器130具有系統MMU（SMMU）130a。GPU 140具有其自身的SMMU 140a。同樣，PCIe IO設備150具有其自身的SMMU 150a。

【0028】處理器核110a的MMU 110b被配置成執行將VA轉譯

成IPA以及將IPA轉譯成PA的任務。頁表被儲存於主記憶體120中。MMU 110b以及SMMU 130a、140a和150a中的每一者具有其自身的TLB（出於清楚目的而未圖示），該TLB儲存被儲存在主記憶體120中的頁表的子集。根據本說明性實施例，在發生TLB失敗之後，MMU 110b執行處理IPA以預測PA的預測演算法。該預測演算法可被數學地表達為：

$$PA=f(IPA), \quad (\text{式1})$$

【0029】其中f表示數學函數。可被用於此目的的函數f在以下參照圖5詳細描述。短語「預測」當該短語在本文中使用时意指「決定」，並且並不暗示隨機或機率性決定，儘管隨機或機率性決定未必從本發明的範疇中排除。由預測演算法作出的預測通常是（但並非必然是）決定性的。

【0030】CPU叢集110執行系統OS 200以及虛擬機器監視器（VMM），或即管理程式210。管理程式210管理轉譯任務，該等轉譯任務除了執行轉譯以外亦包括更新儲存於MMU 110b以及SMMU 130a、140a和150a中的頁表。管理程式210亦執行訪客HLOS 220及/或訪客數位版權管理器（DRM）230。HLOS 220可以與攝影機顯示器130相關聯，並且DRM 230可以與GPU 140相關聯。管理程式210管理HLOS 220和DRM 230。

【0031】在發生TLB失敗之後，管理程式210配置MMU 110b以及SMMU 130a、140a和150a以執行將IPA轉換成PA的預測演算法。在此類情形中，針對與TLB失敗相關聯的VA的起始IPA是按S1轉譯正常開始的典型方式從CPU叢集110的硬體基址暫存器（出於清楚目的而未圖示）獲得的。預測演算法隨後

根據式1預測PA，如將在以下更詳細描述的。爲了管理和更新SMMU 130a、140a和150a，CPU MMU 110b在匯流排170上向SMMU 130a、140a和150a發送分散式虛擬記憶體（DVM）訊息。MMU 110b以及SMMU 130a、140a和150a存取主記憶體120以執行HWTW。

【0032】根據說明性實施例，CPU MMU 110b將MMU訊務分類成三種事務類，即：(1)用於尋找儲存S1頁表的PA的S2 HWTW讀事務；(2)客戶機事務；及(3)位址故障（AF）/髒標誌寫事務。根據該說明性實施例，預測演算法僅針對類1事務（亦即，HWTW讀事務）將IPA轉換成PA。對於所有其他事務類，根據本說明性實施例，MMU 110b以及SMMU 130a、140a和150a按典型方式執行所有其他轉譯（例如，S1和客戶機事務S2轉譯）。

【0033】圖3是表示根據說明性實施例的由CPU MMU 110b執行以減少執行HWTW讀事務所需的時間和處理管理負擔量的方法的流程圖。方框301表示方法開始，其通常發生在CPU叢集110啟動並開始執行系統OS 200和管理程式210時。MMU 110b將訊務分類成前述的事務類(1)、(2)和(3)，如由方框302所指示的。分類程序可將事務分類成多於或少於該三個類別，但至少一個類別將是類(1)事務，亦即，用於尋找儲存S1頁表的PA的S2 HWTW讀事務。在由方框303表示的步驟處，決定在執行類(1)事務時是否已發生TLB失敗。若否，則該方法行進至方框306，在此MMU 110b或SMMU 130a、140a或150a以正常方式執行HWTW。

【0034】 若在由方框303表示的步驟處CPU MMU 110b決定在執行類(1)事務時發生了失敗，則該方法行進至由方框305表示的步驟。在由方框305表示的步驟處，執行前述預測演算法以將IPA轉換或轉譯成PA。

【0035】 圖4是根據說明性實施例說明執行HWTW讀事務的方式的圖形示圖。對於此說明性實施例，為了示例性目的，假定頁表是三級頁表並且HWTW是2-D HWTW。該實例亦假定TLB失敗最壞情景。該程序開始於MMU接收VA並隨後從控制暫存器（出於清楚目的而未圖示）取得S1 PGD IPA 401。MMU隨後檢查TLB以尋找與S1 PGD IPA 401的匹配。對於此最壞情景實例，將假定在MMU檢查TLB以尋找匹配時發生TLB失敗。由於該失敗，MMU執行預測演算法以將S1 PGD IPA 401轉換成儲存S1 PMD IPA 403的PA 402。因此，使用單次檢視來將S1 PGD IPA 401轉換成PA 402

【0036】 對於此最壞情景實例，將假定在MMU檢查TLB以尋找與S1 PMD IPA 403的匹配時發生TLB失敗。由於該失敗，MMU執行預測演算法以將S1 PMD IPA 403轉換成儲存S1 PTE IPA 405的PA 404。因此，使用單次檢視來將S1 PMD IPA 403轉換成PA 404對於此最壞情景實例，將假定在MMU檢查TLB以尋找與S1 PTE IPA 405的匹配時發生TLB失敗。由於該失敗，MMU執行預測演算法以將S1 PTE IPA 405轉換成儲存IPA1 407的PA 406。一旦IPA1 407已被獲得，就執行三次檢視408、409和411以獲得儲存要讀取的資料的最終PA 412。

【0037】 因此，根據該實施例可以看出，檢視的總次數已從

15（圖1）減少到6，此表示在處理管理負擔方面減少了60%。當然，本發明不限於具有特定級數或特定HWTW維數的MMU配置。本領域技藝人士將理解，本發明的概念和原理無論頁表的配置如何均是適用的。此外，儘管方法和系統在本文是參照IPA到PA轉換來描述的，但其在不使用IPA的系統中亦可等同地適用於直接VA到PA轉換。

【0038】圖5是執行預測演算法的預測器500的說明性實施例的方塊圖。預測器500通常在MMU 110b中並在SMMU 130a、140a和150a中實施。如以上所指示的，根據說明性實施例，該預測演算法僅在執行類1讀事務時執行。圖5中示出的預測器500的配置是允許針對類1事務啓用預測器500並且針對所有其他事務類（包括類2和3事務）禁用預測器500的一種配置的實例。

【0039】圖5中示出的預測器500的配置亦允許預測器500選擇以上在式1中用於基於IPA來計算PA的函數f。每個虛擬機器（VM）可以使用不同的函數f集，因此被使用的函數集確保在IPA範疇上在IPA與PA之間存在一對一映射是重要的。管理程式210可以管理多個HLOS或DRM，其中的每一個將具有在管理程式210中執行的對應VM。被使用的該等函數集確保所預測的PA與分配給另一VM的所預測PA不交疊。

【0040】函數f的實例是：

【0041】 $PA=IPA$ ；

【0042】 $PA=IPA+Offset_function(VMID)$ ，其中VMID是跨所有VM識別與HWTW讀事務相關聯的VM的唯一性識別符，並

且 Offset_function (偏移_函數) 是具有基於與 VMID 相關聯的特定偏移值來選擇的輸出的函數；及

【0043】 $PA = IPA \text{ XOR } \text{Extended_VMID}$ ，其中 XOR 表示異或運算並且 Extended_VMID 是擴展 VMID。管理程式 210 選擇函數 f 以使得避免各 VM 之間的衝突。

【0044】在圖 5 中，假定函數 f 是多項式，並且管理程式 210 從多個多項式中選擇要被用作函數 f 的多項式。被選擇的多項式可以基於例如正為其執行 HWTW 讀事務的 VM 的 VMID。預測器 500 的配置暫存器 510 保持一或多個預測賦能位元 510a 以及一或多個多項式選擇位元 510b。預測器 500 的多項式計算硬體 520 包括基於從暫存器 510 接收的多項式選擇位元 510b 的值來選擇多項式函數的硬體。多項式計算硬體 520 亦接收 IPA 到 PA 轉譯請求並根據所選擇的多項式函數來處理該請求以產生所預測的 PA。

【0045】預測賦能位元 510a 和類 1 賦能位元在及閘 530 的輸入端被接收。當在執行類 1 讀事務時已發生失敗時，類 1 賦能位元被確定。預測器 500 的多工器 (MUX) 540 在 MUX 540 的選擇器埠處接收及閘 530 的輸出並接收所預測的 PA 和以正常方式獲得的 IPA 到 PA 轉譯結果。當預測賦能位元 510a 和類 1 賦能位元兩者均被確定時，S2 搜尋控制邏輯和狀態機 550 被禁用並且 MUX 540 將所預測的 PA 選擇為來自 MUX 540 的輸出。

【0046】當預測賦能位元 510a 及/或類 1 賦能位元被否定時，S2 搜尋控制邏輯和狀態機 550 被啟用。當 S2 搜尋控制邏輯和狀態機 550 被啟用時，可由 S2 搜尋控制邏輯和狀態機 550 在主記

憶體 120 中執行其他類型的 S2 搜尋（例如，類 2 和類 3）。因此，當 S2 搜尋控制邏輯和狀態機 550 被啓用時，MUX 540 輸出從 S2 搜尋控制邏輯和狀態機 550 輸出的 IPA 到 PA 轉譯結果。

【0047】應注意，預測器 500 可具有許多不同的配置。圖 5 中所示的預測器 500 的配置僅僅是用於執行預測演算法的許多合適的配置中的一種。本領域技藝人士將理解，不同於圖 5 中所示的配置的許多配置可被用來執行預測演算法。

【0048】圖 2 中所示的電腦系統 100 可在其中執行記憶體虛擬化的任何類型的系統中實施，該系統例如包括桌上型電腦、伺服器 and 行動智慧型電話。圖 6 圖示了其中併入電腦系統 100 的行動智慧型電話 600 的方塊圖。智慧型電話 600 除了其必須能夠執行本文所描述的方法以外不限於任何特定類型的智慧型電話或具有任何特定配置。此外，圖 6 中圖示的智慧型電話 600 意欲作為具有上下文感知和用於執行本文描述的方法的處理能力的蜂巢式電話的簡化實例。本領域一般技藝人士將理解智慧型電話的操作和構造，由此實施細節已被省略。

【0049】根據本說明性實施例，智慧型電話 600 包括在系統匯流排 612 上連接在一起的基頻子系統 610 和射頻（RF）子系統 620。系統匯流排 612 通常包括實體和邏輯連接，其將上述元件耦合在一起並賦能其互用性。RF 子系統 620 可以是無線收發機。儘管為了清楚起見未圖示細節，但 RF 子系統 620 一般包括發射（Tx）模組 630（其具有用於準備基頻資訊信號以供傳輸的調制、升頻轉換和放大電路系統），包括接收（Rx）模組 640（其具有用於接收 RF 信號並將該 RF 信號降頻轉換到基頻資訊

信號以恢復資料的放大、濾波和降頻轉換電路系統)，並且包括前端模組（FEM）650（其包括能將發射信號從接收信號中分離出來的雙訊器電路系統、雙工器電路系統，或任何其他電路系統，如本領域技藝人士已知的）。天線660被連接到FEM 650。

【0050】基頻子系統610一般包括電腦系統100、類比電路元件616和數位電路元件618，其經由系統匯流排612電耦合在一起。系統匯流排612通常包括實體和邏輯連接以將上述元件耦合在一起並賦能其互用性。

【0051】輸入/輸出（I/O）元件621經由連接624被連接到基頻子系統610。I/O元件621通常包括例如話筒、小鍵盤、揚聲器、指標設備、使用者介面控制元件、以及允許使用者提供輸入命令並從智慧型電話600接收輸出的任何其他設備或系統。記憶體628經由連接629被連接至基頻子系統610。記憶體628可以是任何類型的揮發性或非揮發性記憶體。記憶體628可以永久安裝在智慧型電話600中，或者可以是可移除式記憶體元件，諸如可移除式記憶卡。

【0052】類比電路系統616和數位電路系統618包括信號處理、信號轉換和將由I/O元件621提供的輸入信號轉換成要傳送的資訊信號的邏輯。類似地，類比電路系統616和數位電路系統618包括用於從收到信號產生包含已恢復資訊的資訊信號的信號處理元件。數位電路系統618可包括例如數位訊號處理器（DSP）、現場可程式設計閘陣列（FPGA），或任何其他處理設備。因為基頻子系統610既包括類比元件又包括數位元件，

所以其可被稱為混合信號設備（MSD）。

【0053】智慧型電話600可包括各種感測器中的一或多個，諸如舉例而言，相機661、話筒662、全球定位系統（GPS）感測器663、加速計665、陀螺儀667和數位羅盤668。該等感測器經由匯流排612與基頻子系統610通訊。

【0054】使電腦系統100嵌入在智慧型電話600中允許多個OS和多個相應VM在智慧型電話600上執行。在此環境中，電腦系統100的管理程式210（圖2）提供了智慧型電話600的硬體與正由VM執行的應用軟體之間的安全分隔。

【0055】以上參照圖3描述的方法可單獨在硬體中實施，或者在硬體與軟體的組合或硬體與韌體的組合中實施。同樣，圖2中所示的電腦系統100的許多元件可單獨在硬體中實施，或者在硬體與軟體或韌體的組合中實施。例如，管理程式210可單獨在硬體中實施，或者在硬體與軟體或韌體的組合中實施。在方法或電腦系統100的元件在軟體或韌體中實施的情形中，對應代碼被儲存在主記憶體120（圖2）中，該主記憶體120是電腦可讀取媒體。主記憶體120通常是固態電腦可讀取媒體，諸如非揮發性隨機存取記憶體（RAM）、唯讀記憶體（ROM）設備、可程式設計ROM（PROM）、可抹除PROM（EPROM）等等。然而，其他類型的電腦可讀取媒體可被用於儲存代碼，諸如舉例而言磁性或光學儲存設備。

【0056】亦應注意，可對以上參照圖2-6描述的方法做出許多變動而不偏離本發明的範疇。例如，圖2中所示的電腦系統100的配置可按數種方式修改，如本領域技藝人士將理解的。此

外，圖6中所示的智慧型電話600僅僅是具有用於執行該方法的合適配置和功能性的行動設備的一個實例。本領域技藝人士鑒於本文提供的描述將理解，可對圖6中所示的智慧型電話600作出許多變動而不偏離本發明的範疇。該等及其他變動均落在本發明的範疇之內。本文描述的說明性實施例意欲說明本發明的原理和概念，但本發明並不限於該等實施例，如本領域技藝人士將理解的。

【符號說明】

【 0057】

2 S1 頁全域目錄（PGD）IPA

3 S2 查表

4 S2 查表

5 S2 查表

6 額外檢視

7 S1 PMD IPA

8 S2 查表

9 S2 查表

11 S2 查表

12 額外檢視

13 S1 PTE IPA

14 S2 查表

15 S2 查表

16 S2 查表

17 額外檢視

- 18 實際 IPA
- 19 S2 查表
- 21 S2 查表
- 22 S2 查表
- 100 電腦系統
- 110 CPU 叢集
- 110a CPU 核
- 110b MMU
- 120 主記憶體
- 130 攝影機顯示器
- 130a 系統 MMU (SMMU)
- 140 圖形處理單元 (GPU)
- 140a SMMU
- 150 快速周邊連接介面 (PCIe) 輸入/輸出 (IO) 設備
- 150a SMMU
- 160 IO TLB
- 170 系統匯流排
- 200 系統 OS
- 210 管理程式
- 220 訪客 HLOS
- 230 訪客數位版權管理器 (DRM)
- 301 方框
- 302 方框
- 303 方框

- 305 方框
- 306 方框
- 401 S1 PGD IPA
- 402 PA
- 403 S1 PMD IPA
- 404 PA
- 405 S1 PTE IPA
- 406 PA
- 407 IPA1
- 408 檢視
- 409 檢視
- 411 檢視
- 412 最終 PA
- 500 預測器
- 510a 預測賦能位元
- 510b 多項式選擇位元
- 520 多項式計算硬體
- 530 及閘
- 540 多工器 (MUX)
- 550 S2 搜尋控制邏輯和狀態機
- 610 基頻子系統
- 612 系統匯流排
- 616 類比電路元件
- 618 數位電路元件

620 射頻 (RF) 子系統

621 輸入/輸出 (I/O) 元件

624 連接

628 記憶體

629 連接

630 發射 (Tx) 模組

640 接收 (Rx) 模組

650 前端模組 (FEM)

660 天線

661 相機

662 話筒

663 全球定位系統 (GPS) 感測器

665 加速計

【生物材料寄存】

國內寄存資訊【請依寄存機構、日期、號碼順序註記】

無

國外寄存資訊【請依寄存國家、機構、日期、號碼順序註記】

無

【序列表】(請換頁單獨記載)

無

申請專利範圍

1. 一種減少與執行一硬體表搜尋（HWTW）相關聯的處理管理負擔的電腦系統，該系統包括：

至少一個中央處理單元（CPU），該CPU執行一主機作業系統（OS）和一管理程式，該管理程式控制該CPU上的至少第一訪客OS的執行，該管理程式至少執行與該第一訪客OS相關聯的一第一虛擬機器（VM）；

與該CPU處於通訊的一實體記憶體，該實體記憶體具有能藉由實體位址（PA）來定址的實體記憶體位置，其中至少一個頁表儲存在該實體記憶體的實體記憶體位置處，該頁表包括與用於將一中間實體位址（IPA）映射成該實體記憶體的一實際PA的映射相對應的頁表條目；

至少一個轉譯後備緩衝器（TLB），其儲存該等頁表條目的一子集；及

至少一個記憶體管理單元（MMU），其與該CPU、與該實體記憶體並與該TLB處於通訊，其中該MMU決定與一IPA相關聯的的頁表條目是否儲存於該TLB中，其中若與該IPA相關聯的頁表條目未被儲存於該TLB中，則已發生一TLB失敗，並且其中若發生一TLB失敗，則該MMU預測儲存著與該IPA相關聯的資料之該實體記憶體的一PA；

其中該MMU將該PA預測為該IPA的一函數 f ： $PA=f(IPA)$

；

其中該函數 f 選自多個函數，並且其中該多個函數中的每一個函數提供該IPA與所預測的PA之間的一對一映射；及

其中該管理程式正至少執行與一數位版權管理器（DRM）電腦程式相關聯的一第二VM，並且其中該函數f是IPA_Offset_function(VMID)，其中VMID是跨該第一VM和該第二VM的識別與該TLB失敗相關聯的VM的一唯一性識別符，並且其中IPA_Offset_function是具有一輸出的一函數，該輸出是基於與在發生該TLB失敗時正使用該IPA存取記憶體體的該第一VM或該第二VM的VMID相關聯的一特定偏移值來選擇的，並且其中該所預測的PA被預測為：

● PA=IPA_Offset_function(VMID)。

2. 如請求項1述及之電腦系統，其中該函數f是一多項式。

3. 如請求項1述及之電腦系統，其中該函數f是一單位函數以使得PA=IPA。

● 4. 一種減少與執行一硬體表搜尋（HWTW）相關聯的處理管理負擔的電腦系統，該系統包括：

至少一個中央處理單元（CPU），該CPU執行一主機作業系統（OS）和一管理程式，該管理程式控制該CPU上的至少一第一訪客OS的執行，該管理程式至少執行與該第一訪客OS相關聯的一第一虛擬機器（VM）；

與該CPU處於通訊的一實體記憶體，該實體記憶體具有能藉由實體位址（PA）來定址的實體記憶體位置，其中至少一個頁表儲存在該實體記憶體的實體記憶體位置處，該頁表

包括與用於將一中間實體位址（IPA）映射成該實體記憶體的一實際PA的映射相對應的頁表條目；

至少一個轉譯後備緩衝器（TLB），其儲存該等頁表條目的一子集；及

至少一個記憶體管理單元（MMU），該MMU與該CPU、與該實體記憶體並與該TLB處於通訊，其中該MMU決定與一IPA相關聯的的頁表條目是否儲存於該TLB中，其中若與該IPA相關聯的頁表條目未被儲存於該TLB中，則已發生一TLB失敗，並且其中若發生一TLB失敗，則該MMU預測儲存著與該IPA相關聯的資料之該實體記憶體的一PA；

其中該MMU將該PA預測為該IPA的一函數 f ： $PA=f(IPA)$

；

其中該函數 f 選自多個函數，並且其中該多個函數中的每一個函數提供該IPA與所預測的PA之間的一對一映射；及

其中該管理程式正至少執行與一數位版權管理器（DRM）電腦程式相關聯的一第二VM，並且其中該函數 f 是 $IPA \text{ XOR } Extended_VMID$ ，其中XOR表示異或運算並且 $Extended_VMID$ 是一擴展VMID，並且其中該所預測的PA被預測為：

$PA=IPA \text{ XOR } Extended_VMID$ 。

5. 如請求項1述及之電腦系統，其中該電腦系統是一行動設備的一部分。

6. 如請求項5述及之電腦系統，其中該行動設備是一行動電

話。

7. 如請求項6述及之電腦系統，其中該行動電話是一智慧型電話。

8. 一種減少與執行一硬體表搜尋（HWTW）相關聯的處理管理負擔的方法，該方法包括以下步驟：

提供至少一個中央處理單元（CPU）、至少一個實體記憶體、至少一個轉譯後備緩衝器（TLB）、以及至少一個記憶體管理單元（MMU），該CPU、該實體記憶體、該TLB、以及該MMU彼此處於通訊，該CPU執行一主機作業系統（OS）和一管理程式，該管理程式控制該CPU上的至少一第一訪客OS的執行，該管理程式至少執行與該第一訪客OS相關聯的一第一虛擬機器（VM），該實體記憶體具有能藉由實體位址（PA）來定址的實體記憶體位置，其中至少一個頁表儲存在該實體記憶體的實體記憶體位置處，該頁表包括與用於將一中間實體位址（IPA）映射成該實體記憶體的一實際PA的映射相對應的頁表條目，該TLB儲存該等頁表條目的一子集；及

在該MMU中：

決定與一IPA相關聯的頁表條目是否儲存於該TLB中，

若決定與該IPA相關聯的頁表條目未被儲存於該TLB中，則判定已發生一TLB失敗，以及

若做出已發生一TLB失敗的一判定，則預測儲存著與該IPA相關聯的資料之該實體記憶體的一PA；

其中該MMU將該PA預測為該IPA的一函數 f ： $PA=f(IPA)$

其中該函數 f 選自多個函數，並且其中該多個函數中的每一個函數提供該IPA與該所預測的PA之間的一對一映射；及

其中該管理程式正至少執行與一數位版權管理器（DRM）電腦程式相關聯的一第二VM，並且其中該函數 f 是 $IPA_Offset_function(VMID)$ ，其中VMID是跨該第一VM和該第二VM的識別與該TLB失敗相關聯的該VM的一唯一性識別符，並且其中 $IPA_Offset_function$ 是具有一輸出的一函數，該輸出是基於與在發生該TLB失敗時正使用該IPA存取記憶體的該第一VM或該第二VM的該VMID相關聯的一特定偏移值來選擇的，並且其中該所預測的PA被預測為：

$PA=IPA_Offset_function(VMID)$ 。

9. 如請求項8述及之方法，其中該函數 f 是一多項式。

10. 如請求項8述及之方法，其中該函數 f 是一單位函數以使得 $PA=IPA$ 。

11. 一種減少與執行一硬體表搜尋（HWTW）相關聯的處理管理負擔的方法，該方法包括以下步驟：

提供至少一個中央處理單元（CPU）、至少一個實體記憶體、至少一個轉譯後備緩衝器（TLB）、以及至少一個記憶體管理單元（MMU），該CPU、該實體記憶體、該TLB、以及該MMU彼此處於通訊，該CPU執行一主機作業系統（OS）和一

管理程式，該管理程式控制該CPU上的至少一第一訪客OS的執行，該管理程式至少執行與該第一訪客OS相關聯的一第一虛擬機器（VM），該實體記憶體具有能藉由實體位址（PA）來定址的實體記憶體位置，其中至少一個頁表儲存在該實體記憶體的實體記憶體位置處，該頁表包括與用於將一中間實體位址（IPA）映射成該實體記憶體的一實際PA的映射相對應的頁表條目，該TLB儲存該等頁表條目的一子集；及

在該MMU中：

● 決定與一IPA相關聯的頁表條目是否儲存於該TLB中，
若決定與該IPA相關聯的頁表條目未被儲存於該TLB中，則判定已發生一TLB失敗，以及

若做出已發生一TLB失敗的一判定，則預測儲存著與該IPA相關聯的資料之該實體記憶體的一PA；

其中該MMU將該PA預測為該IPA的一函數 f ： $PA=f(IPA)$

其中該函數 f 選自多個函數，並且其中該多個函數中的每一個函數提供該IPA與該所預測的PA之間的一對一映射；及

● 其中該管理程式正至少執行與一數位版權管理器（DRM）電腦程式相關聯的一第二VM，並且其中該函數 f 是 $IPA \text{ XOR } Extended_VMID$ ，其中XOR表示異或運算並且 $Extended_VMID$ 是一擴展VMID，並且其中該所預測的PA被預測為： $PA=IPA \text{ XOR } Extended_VMID$ 。

12. 如請求項8述及之方法，其中該管理程式控制該CPU上至少第一和第二訪客OS的執行，並且其中該管理程式亦至少在

執行與該第二訪客 OS 相關聯的一第二 VM，並且其中由該 MMU 用於預測 PA 的該函數 f 針對與該第一 VM 相關聯的一失敗來預測在第一 PA 範圍中的 PA 並針對與該第二 VM 相關聯的一失敗來預測在第二 PA 範圍中的 PA，並且其中該第一 PA 範圍和該第二 PA 範圍彼此不同。

13. 如請求項 8 述及之方法，其中該方法由一行動設備的一電腦系統執行。

14. 如請求項 13 述及之方法，其中該行動設備是一行動電話。

15. 如請求項 14 述及之方法，其中該行動電話是一智慧型電話。

16. 一種其上儲存有一電腦代碼的非瞬態電腦可讀取媒體（CRM），該電腦代碼由一或多個處理器執行以用於減少與執行一硬體表搜尋（HWTW）相關聯的處理管理負擔，該電腦代碼包括：

用於決定與一中間實體位址（IPA）相關聯的頁表條目是否儲存於一 TLB 中的一第一代碼部分，其中若決定與該 IPA 相關聯的頁表條目未被儲存於該 TLB 中，則該第一代碼部分判定已發生一 TLB 失敗；及

用於在該第一代碼部分判定已發生一 TLB 失敗的情況下

預測儲存著與該IPA相關聯的資料之一實體記憶體的一實體位址（PA）的第二代碼部分

其中該第二代碼部分將該PA預測為該IPA的一函數f：
 $PA=f(IPA)$ ；

其中該第二代碼部分從多個函數中選擇該函數f，並且其中該多個函數中的每一個函數提供該IPA與該所預測的PA之間的一對一映射；及

其中該函數f是IPA_Offset_function(VMID)，其中VMID是跨第一虛擬機器（VM）和第二VM的將該第一VM和該第二VM之一識別為與該TLB失敗相關聯的一VM的一唯一性識別符，並且其中IPA_Offset_function是具有一輸出的一函數，該輸出是基於與在發生該TLB失敗時正使用該IPA存取記憶體的該第一VM或該第二VM的該VMID相關聯的一特定偏移值來選擇的，並且其中該所預測的PA被預測為：

$PA=IPA_Offset_function(VMID)$ 。

17. 如請求項16述及之非瞬態CRM，其中該函數f是一多項式。

18. 如請求項16述及之非瞬態CRM，其中該函數f是一單位函數以使得 $PA=IPA$ 。

19. 一種其上儲存有一電腦代碼的非瞬態電腦可讀取媒體（CRM），該電腦代碼由一或多個處理器執行以用於減少與執行

一硬體表搜尋（HWTW）相關聯的處理管理負擔，該電腦代碼包括：

用於決定與一中間實體位址（IPA）相關聯的頁表條目是否儲存於一TLB中的一第一代碼部分，其中若決定與該IPA相關聯的頁表條目未被儲存於該TLB中，則該第一代碼部分判定已發生一TLB失敗；及

用於在該第一代碼部分判定已發生一TLB失敗的情況下預測儲存著與該IPA相關聯的資料一實體記憶體的一實體位址（PA）的一第二代碼部分；

其中該第二代碼部分將該PA預測為該IPA的一函數f：
 $PA=f(IPA)$ ；

其中該第二代碼部分從多個函數中選擇該函數f，並且其中該多個函數中的每一個函數提供該IPA與該所預測的PA之間的一對一映射；及

其中該函數f是IPA XOR Extended_VMID，其中XOR表示異或運算，並且Extended_VMID是一擴展VMID，並且其中該預測PA被預測為： $PA=IPA \text{ XOR } Extended_VMID$ 。

圖式

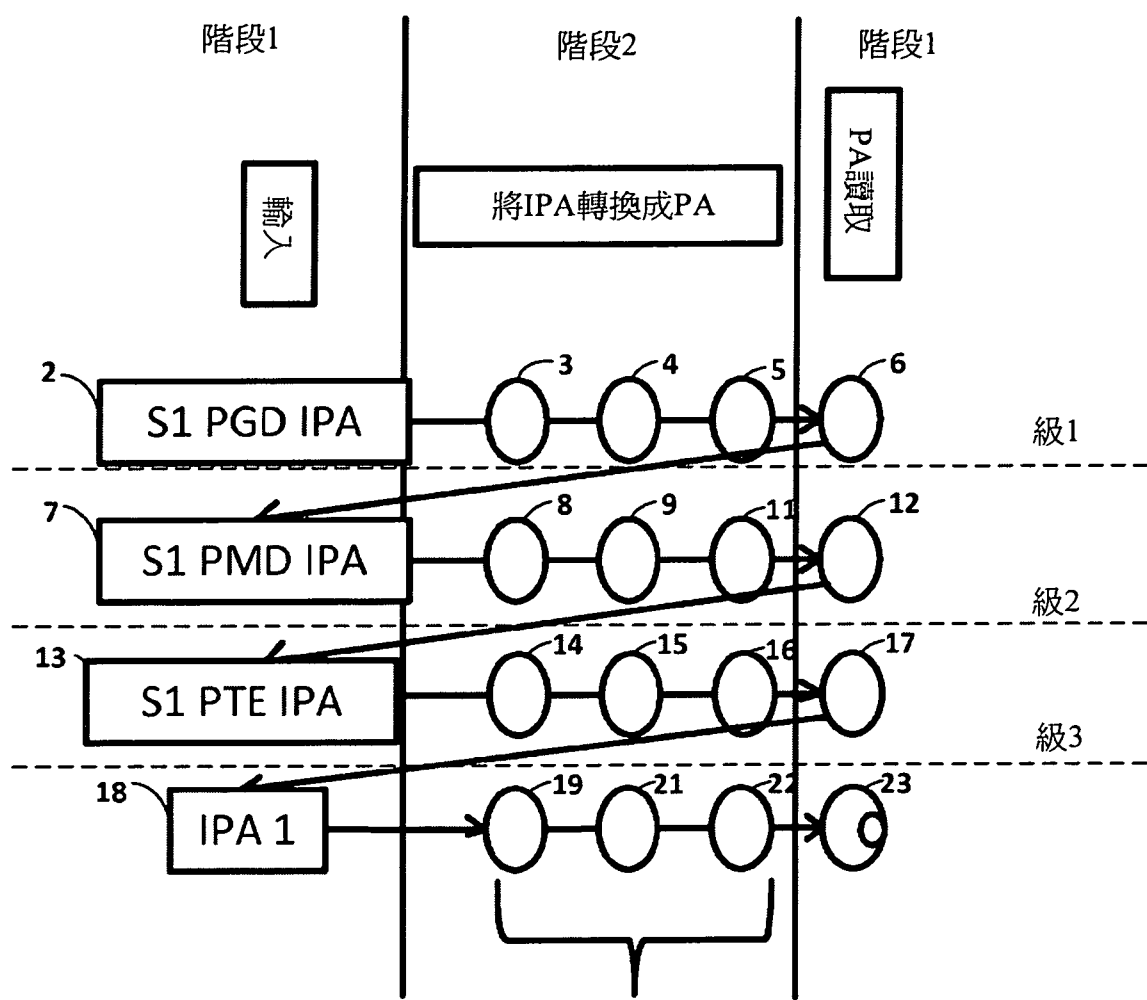


圖1
先前技術

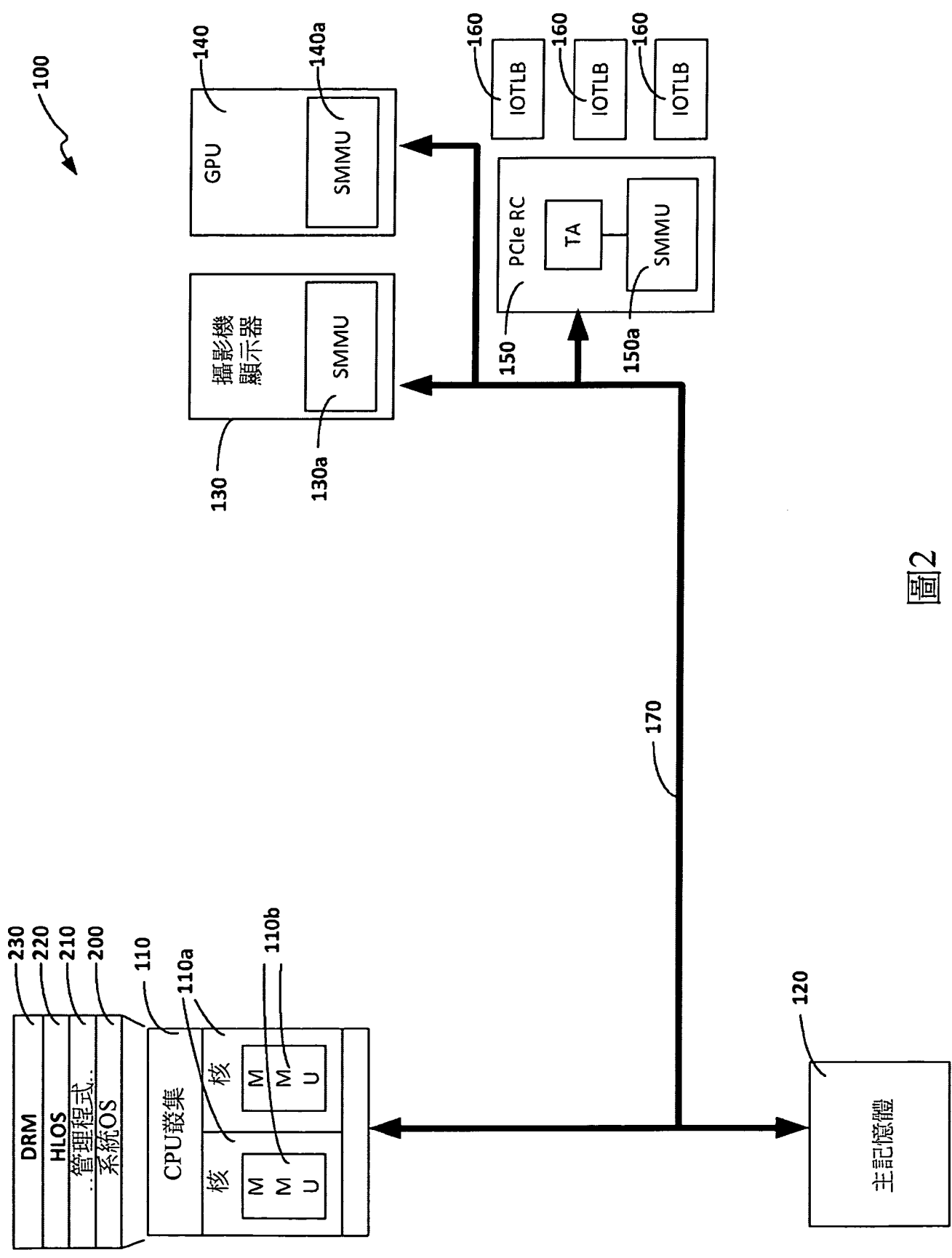


圖2

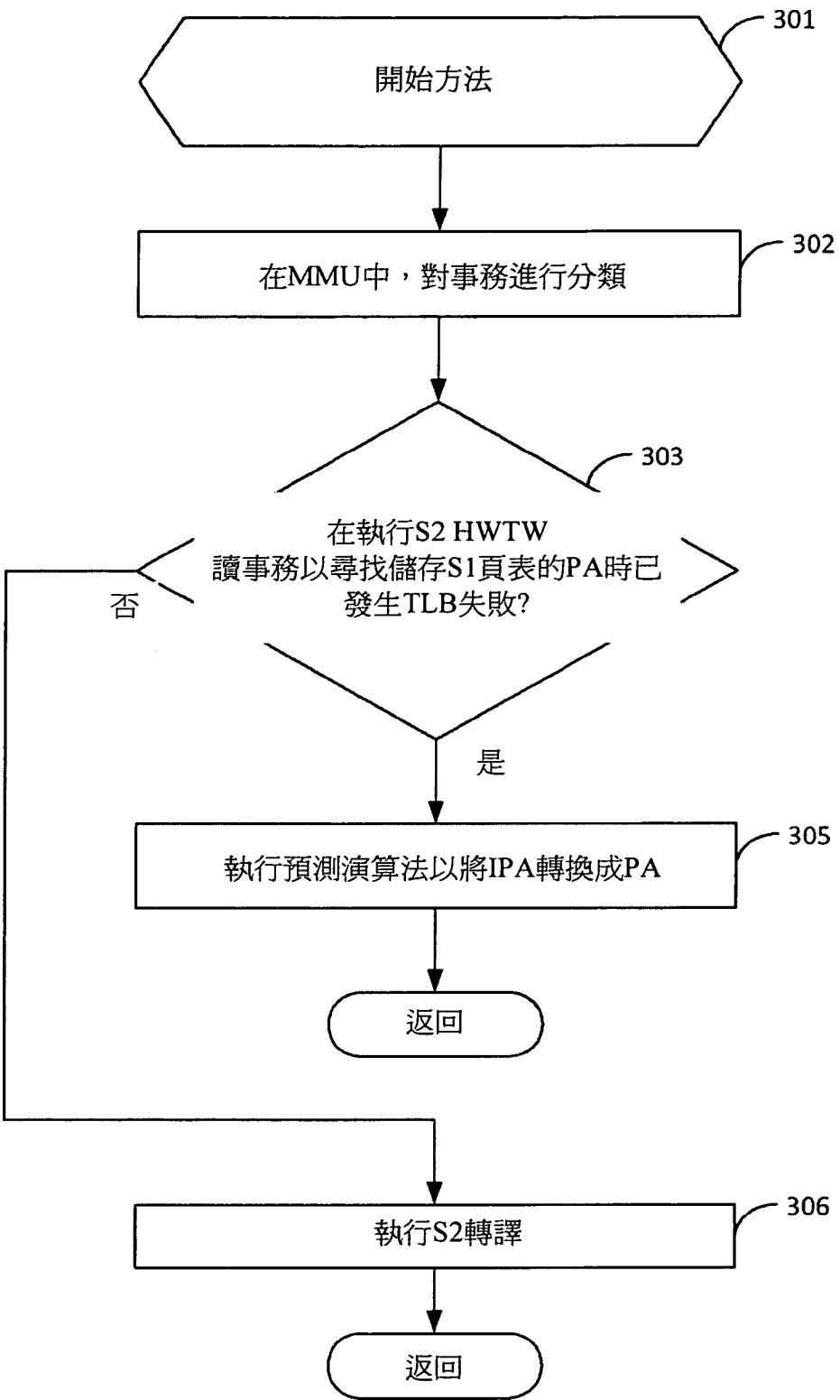


圖3

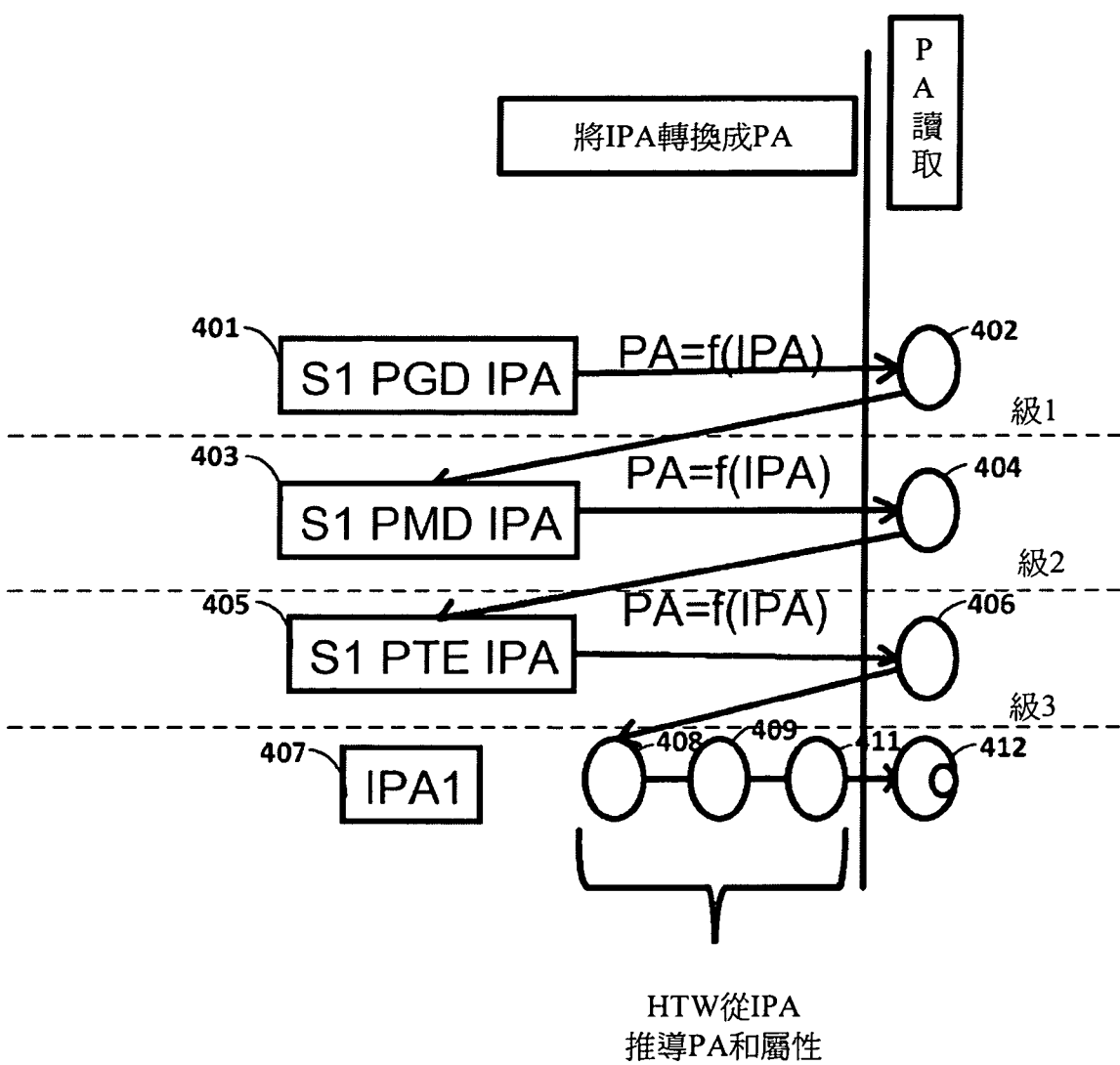


圖4

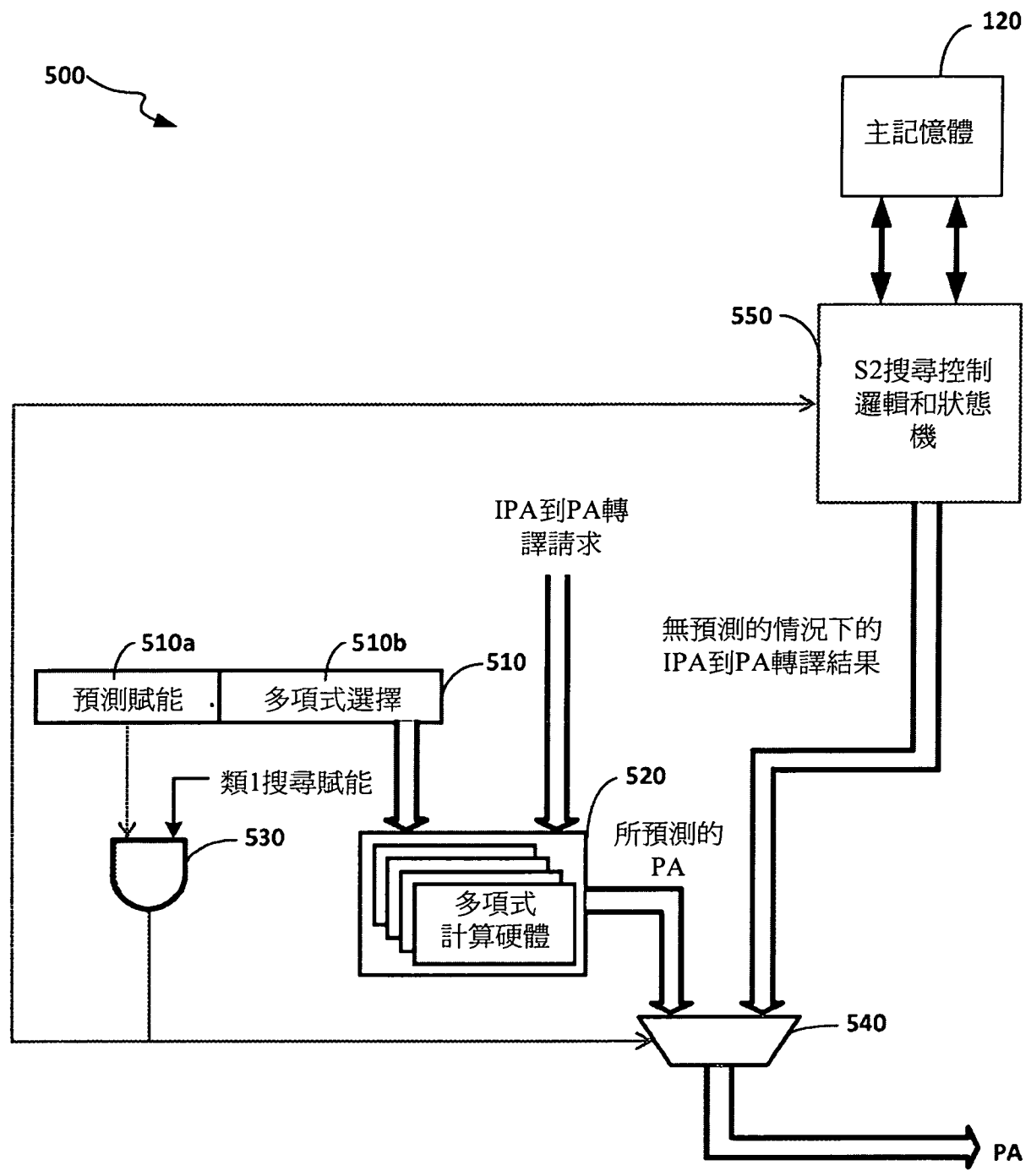


圖5

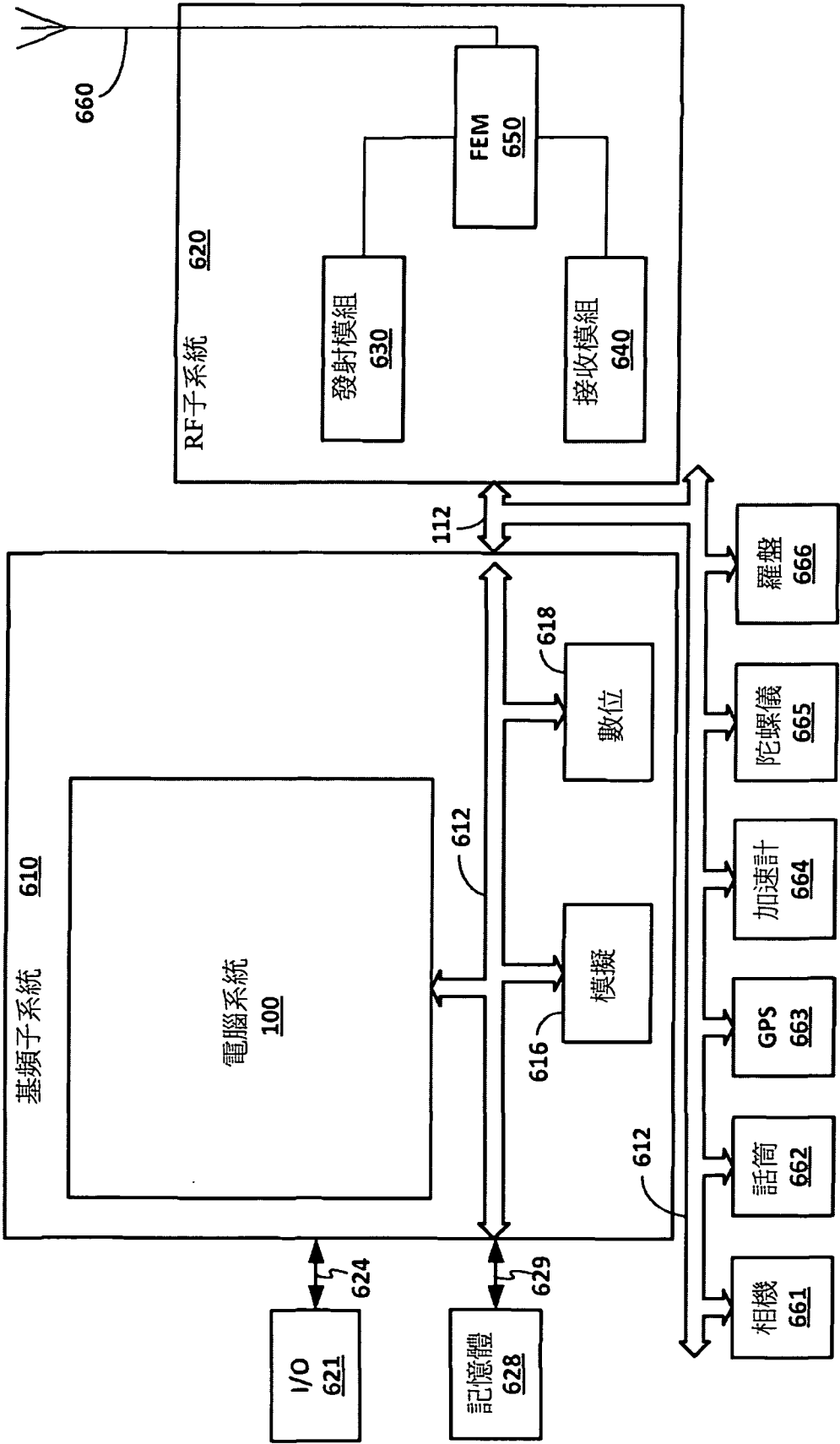


圖6