I817920

(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 I817920 B**

(45)公告日： 中華民國 112 (2023) 年 10 月 01 日

(21)申請案號：112118986 (22)申請日： 中華民國 106 (2017) 年 10 月 31 日

(51)Int. Cl.： ***G06F13/14 (2006.01)*** ***G06F11/36 (2006.01)***
G06F11/30 (2006.01) ***G06F11/34 (2006.01)***
G06F16/48 (2019.01)

(30)優先權：2017/03/29 美國 15/473,101

(71)申請人：美商谷歌有限責任公司 (美國) GOOGLE LLC (US)
美國

(72)發明人：諾里 湯瑪士 NORRIE, THOMAS (US)；庫瑪 納文 KUMAR, NAVEEN (US)

(74)代理人：陳長文；簡秀如；金若芸

(56)參考文獻：

US 6530076B1 US 2012/0226839A1
US 2016/0070636A1 WO 2015/016920A1

審查人員：朱明宗

申請專利範圍項數：20 項 圖式數：5 共 47 頁

(54)名稱

用於分散式硬體追蹤之電腦實施方法、系統及非暫時性電腦儲存單元

(57)摘要

本發明揭示一種藉由一或多個處理器執行之電腦實施方法，該方法包含：監測藉由一第一處理器組件執行之程式碼之執行；及監測藉由一第二處理器組件執行之程式碼之執行。一運算系統將識別硬體事件之資料儲存於一記憶體緩衝器中。該等所儲存事件跨包含至少該第一處理器組件及該第二處理器組件之處理器單元發生。該等硬體事件各包含一事件時間戳記及特徵化該事件之後設資料。該系統產生識別該等硬體事件之一資料結構。該資料結構依一時間排序序列配置該等事件且使事件與至少該第一處理器組件或該第二處理器組件相關聯。該系統將該資料結構儲存於一主機裝置之一記憶體庫中及使用該資料結構來分析藉由該第一處理器組件或該第二處理器組件執行之該程式碼之效能。

A computer-implemented method executed by one or more processors, the method includes monitoring execution of program code executed by a first processor component; and monitoring execution of program code executed by a second processor component. A computing system stores data identifying hardware events in a memory buffer. The stored events occur across processor units that include at least the first and second processor components. The hardware events each include an event time stamp and metadata characterizing the event. The system generates a data structure identifying the hardware events. The data structure arranges the events in a time ordered sequence and associates events with at least the first or second processor components. The system stores the data structure in a memory bank of a host device and uses the data structure to analyze performance of the program code executed by the first or second processor components.

指定代表圖：

符號簡單說明：

500:程序

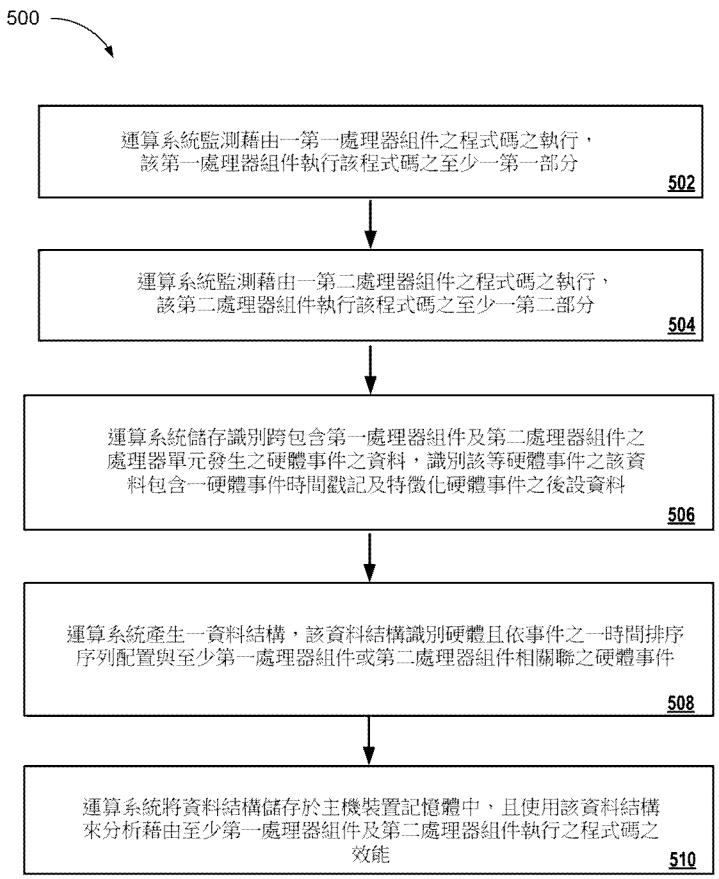
502:方塊

504:方塊

506:方塊

508:方塊

510:方塊



【圖5】

專利案號：112118986
原申請案號：110136035
公告本
I817920

【發明摘要】

【中文發明名稱】

用於分散式硬體追蹤之電腦實施方法、系統及非暫時性電腦儲存單元

【英文發明名稱】

A COMPUTER-IMPLEMENTED METHOD, SYSTEM, AND NON-TRANSITORY COMPUTER STORAGE UNIT FOR DISTRIBUTED HARDWARE TRACING

【中文】

本發明揭示一種藉由一或多個處理器執行之電腦實施方法，該方法包含：監測藉由一第一處理器組件執行之程式碼之執行；及監測藉由一第二處理器組件執行之程式碼之執行。一運算系統將識別硬體事件之資料儲存於一記憶體緩衝器中。該等所儲存事件跨包含至少該第一處理器組件及該第二處理器組件之處理器單元發生。該等硬體事件各包含一事件時間戳記及特徵化該事件之後設資料。該系統產生識別該等硬體事件之一資料結構。該資料結構依一時間排序序列配置該等事件且使事件與至少該第一處理器組件或該第二處理器組件相關聯。該系統將該資料結構儲存於一主機裝置之一記憶體庫中及使用該資料結構來分析藉由該第一處理器組件或該第二處理器組件執行之該程式碼之效能。

【英文】

A computer-implemented method executed by one or more processors, the method includes monitoring execution of program code executed by a first processor component; and monitoring execution of program code executed by a second processor component. A computing

system stores data identifying hardware events in a memory buffer. The stored events occur across processor units that include at least the first and second processor components. The hardware events each include an event time stamp and metadata characterizing the event. The system generates a data structure identifying the hardware events. The data structure arranges the events in a time ordered sequence and associates events with at least the first or second processor components. The system stores the data structure in a memory bank of a host device and uses the data structure to analyze performance of the program code executed by the first or second processor components.

【指定代表圖】

圖5

【代表圖之符號簡單說明】

500:程序

502:方塊

504:方塊

506:方塊

508:方塊

510:方塊

【發明說明書】

【中文發明名稱】

用於分散式硬體追蹤之電腦實施方法、系統及非暫時性電腦儲存單元

【英文發明名稱】

A COMPUTER-IMPLEMENTED METHOD, SYSTEM, AND NON-TRANSITORY COMPUTER STORAGE UNIT FOR DISTRIBUTED HARDWARE TRACING

【技術領域】

本說明書係關於分析程式碼之執行。

【先前技術】

執行於分散式硬體組件內之分散式軟體之有效效能分析可為一複雜任務。分散式硬體組件可為協作及互動以執行一較大軟體程式或程式碼之若干部分之兩個或兩個以上中央處理單元(CPU) (或圖形處理單元(GPU))之各自處理器核心。

從(例如，CPU或GPU內之)硬體角度來看，通常存在兩種類型之資訊或特徵可用於效能分析：1)硬體效能計數器；及2)硬體事件追蹤。

【發明內容】

一般而言，本說明書中所描述之標的之一態樣可體現於一種藉由一或多個處理器執行之電腦實施方法中。該方法包含：監測藉由一第一處理器組件之程式碼之執行，該第一處理器組件經組態以執行該程式碼之至少一第一部分；及監測藉由一第二處理器組件之該程式碼之執行，該第二處理器組件經組態以執行該程式碼之至少一第二部分。

該方法進一步包含藉由運算系統儲存跨包含第一處理器組件及第二

處理器組件之處理器單元發生之一或多個硬體事件之資料，且將其儲存於至少一記憶體緩衝器中。各硬體事件表示與程式碼之一記憶體存取操作、程式碼之一經發出指令或程式碼之一經執行指令相關聯之資料通信之至少一者。識別一或多個硬體事件之各者之資料包含一硬體事件時間戳記及特徵化該硬體事件之後設資料。方法包含藉由運算系統產生識別一或多個硬體事件之一資料結構，該資料結構經組態以依事件之一時間排序序列配置與至少第一處理器組件及第二處理器組件相關聯之該一或多個硬體事件。

方法進一步包含藉由運算系統將所產生資料結構儲存於一主機裝置之一記憶體庫中以用於分析藉由至少第一處理器組件或第二處理器組件執行之程式碼之效能。

此等及其他實施方案可各視需要包含以下特徵之一或多者。例如，在一些實施方案中，方法進一步包含：藉由運算系統偵測與藉由第一處理器組件或第二處理器組件之至少一者執行之程式碼之若干部分相關聯之一觸發功能；及回應於偵測該觸發功能，藉由運算系統起始引起與一或多個硬體事件相關聯之資料儲存於至少一記憶體緩衝器中之至少一追蹤事件。

在一些實施方案中，觸發功能對應於程式碼中之一特定序列步驟或藉由處理器單元所使用之一全域時脈指示之一特定時間參數之至少一者；且起始至少一追蹤事件包含判定一追蹤位元經設定為一特定值，該至少一追蹤事件與包含跨處理器單元發生之多個中間操作之一記憶體存取操作相關聯，且回應於判定該追蹤位元經設定為該特定值而將與該多個中間操作相關聯之資料儲存於一或多個記憶體緩衝器中。

在一些實施方案中，儲存識別一或多個硬體事件之資料進一步包含：將識別該一或多個硬體事件之若干硬體事件之一第一資料子集儲存於

第一處理器組件之一第一記憶體緩衝器中。該儲存回應於該第一處理器組件執行與程式碼之至少第一部分相關聯之一硬體追蹤指令而發生。

在一些實施方案中，儲存識別一或多個硬體事件之資料進一步包含：將識別該一或多個硬體事件之若干硬體事件之一第二資料子集儲存於第二處理器組件之一第二記憶體緩衝器中。該儲存回應於該第二處理器組件執行與程式碼之至少第二部分相關聯之一硬體追蹤指令而發生。

在一些實施方案中，產生資料結構進一步包含：藉由運算系統比較識別硬體事件之第一資料子集中之各自事件之至少硬體事件時間戳記與識別硬體事件之第二資料子集中之各自事件之至少硬體事件時間戳記；及藉由運算系統部分基於該第一子集中之該等各自事件與該第二子集中之該等各自事件之該比較而提供一相關硬體事件集，且提供其以呈現於資料結構中。

在一些實施方案中，所產生資料結構識別指示一特定硬體事件之一延時屬性之至少一參數，該延時屬性指示該特定硬體事件之至少一持續時間。在一些實施方案中，運算系統之至少一處理器係具有一或多個處理器組件之一多核心多節點處理器，且一或多個硬體事件部分對應於至少在一第一節點之第一處理器組件與一第二節點之第二處理器組件之間發生之資料傳送。

在一些實施方案中，第一處理器組件及第二處理器組件係以下項之一者：一處理器、一處理器核心、一記憶體存取引擎或運算系統之一硬體特徵，且一或多個硬體事件部分對應於一源與一目的地之間的資料封包之移動；且特徵化硬體事件之後設資料對應於一源記憶體位址、一目的地記憶體位址、一唯一追蹤識別號碼或與一直接記憶體存取(DMA)追蹤相關

聯之一大小參數之至少一者。

在一些實施方案中，一特定追蹤(ID)號碼係與跨處理器單元發生之多個硬體事件相關聯，且其中該多個硬體事件對應於一特定記憶體存取操作，且該特定追蹤ID號碼係用於使該多個硬體事件之一或多個硬體事件相互關聯且係用於基於該相互關聯判定該記憶體存取操作之一延時屬性。

本說明書中所描述之標的之另一態樣可體現於一種分散式硬體追蹤系統中，該分散式硬體追蹤系統包含：一或多個處理器，其或其等包含一或多個處理器核心；一或多個機器可讀儲存單元，其或其等用於儲存可藉由該一或多個處理器執行以執行包含以下項之操作之指令：監測藉由一第一處理器組件之程式碼之執行，該第一處理器組件經組態以執行該程式碼之至少一第一部分；及監測藉由一第二處理器組件之該程式碼之執行，該第二處理器組件經組態以執行該程式碼之至少一第二部分。

方法進一步包含藉由運算系統儲存識別跨包含第一處理器組件及第二處理器組件之處理器單元發生之一或多個硬體事件之資料，且將其儲存於至少一記憶體緩衝器中。各硬體事件表示與程式碼之一記憶體存取操作、程式碼之一經發出指令或程式碼之一經執行指令相關聯之資料通信之至少一者。識別一或多個硬體事件之各者之資料包含一硬體事件時間戳記及特徵化該硬體事件之後設資料。方法包含藉由運算系統產生識別一或多個硬體事件之一資料結構，該資料結構經組態以依事件之一時間排序序列配置與至少第一處理器組件及第二處理器組件相關聯之該一或多個硬體事件。

方法進一步包含藉由運算系統將所產生資料結構儲存於一主機裝置之一記憶體庫中以用於分析藉由至少第一處理器組件或第二處理器組件執

行之程式碼之效能。

此態樣及其他態樣之其他實施方案包含經組態以執行方法之動作之對應系統、設備及在電腦儲存裝置上編碼之電腦程式。一或多個電腦之一系統可藉由安裝於該系統上之進行操作以使該系統執行動作之軟體、韌體、硬體或其等之一組合而如此組態。一或多個電腦程式可藉由具有在藉由資料處理設備執行時使該設備執行動作之指令而如此組態。

可在特定實施例中實施本說明書中所描述之標的以便實現以下優點之一或多者。所描述之硬體追蹤系統實現在藉由包含多節點多核心處理器之分散式處理單元執行一分散式軟體程式期間發生之硬體事件之有效相互關聯。該所描述之硬體追蹤系統進一步包含依多個跨節點組態實現硬體事件/追蹤資料之收集及相互關聯之機構。

硬體追蹤系統藉由使用透過硬體旋鈕/特徵執行之動態觸發增強運算效率。此外，硬體事件可運用諸如唯一追蹤識別符、事件時間戳記、事件源位址及事件目的地位址之事件描述符以一序列化方式進行時間排序。此等描述符幫助軟體程式設計員及處理器設計工程師對可能在源程式碼執行期間出現之軟體及硬體效能問題進行有效除錯及分析。

在隨附圖式及下文描述中闡述本說明書中所描述之標的之一或多個實施方案之細節。將自描述、圖式及發明申請專利範圍明白標的之其他潛在特徵、態樣及優點。

【圖式簡單說明】

圖1繪示用於分散式硬體追蹤之一實例性運算系統之一方塊圖。

圖2繪示用於分散式硬體追蹤之一實例性運算系統之追蹤鏈及各自節點之一方塊圖。

圖3繪示一實例性追蹤多工器設計架構及一實例性資料結構之一方塊圖。

圖4係指示用於藉由用於分散式硬體追蹤之一實例性運算系統執行之一直接記憶體存取追蹤事件之追蹤活動之一方塊圖。

圖5係用於分散式硬體追蹤之一實例性程序之一程序流程圖。

各種圖式中之相同元件符號及名稱指示相同元件。

【實施方式】

相關申請案之交叉參考

本申請案係關於於2017年3月29日申請且代理人檔案號碼為16113-8129001之標題為「Synchronous Hardware Event Collection」之美國專利申請案第15/472,932號。美國專利申請案第15/472,932號之全部揭示內容以其全文引用方式明確併入本文中。

本說明書中所描述之標的大體上係關於分散式硬體追蹤。特定言之，一運算系統監測藉由一或多個處理器核心執行之程式碼之執行。例如，該運算系統可監測藉由一第一處理器核心執行之程式碼之執行及藉由至少一第二處理器核心執行之程式碼之執行。該運算系統將識別一或多個硬體事件之資料儲存於一記憶體緩衝器中。識別事件之該所儲存資料對應於跨包含至少第一處理器核心及第二處理器核心之分散式處理器單元發生之事件。

對於各硬體事件，所儲存資料包含一事件時間戳記及特徵化該硬體事件之後設資料。系統產生識別硬體事件之一資料結構。該資料結構依一時間排序序列配置事件且使事件與至少第一處理器核心或第二處理器核心相關聯。系統將資料結構儲存於一主機裝置之一記憶體庫中且使用該資料

結構來分析藉由第一處理器核心或第二處理器核心執行之程式碼之效能。

圖1繪示用於分散式硬體追蹤之一實例性運算系統100之一方塊圖。如本說明書中所使用，分散式硬體系統追蹤對應於識別一實例性處理器微晶片之組件及子組件內發生之事件之資料的儲存。此外，如本文中所使用，一分散式硬體系統(或追蹤系統)對應於處理器微晶片或處理單元之一集合，處理器微晶片或處理單元之該集合協作以執行經組態以在處理器微晶片或分散式處理單元之該集合間分散式執行之一軟體/程式碼之各自部分。

系統100可為具有一或多個處理器或處理單元之一分散式處理系統，該一或多個處理器或處理單元以一分散式方式執行一軟體程式，即，藉由在系統100之不同處理單元上執行程式碼之不同部分(part或portion)。處理單元可包含兩個或兩個以上處理器、處理器微晶片或處理單元，例如，至少一第一處理單元及一第二處理單元。

在一些實施方案中，在第一處理單元接收及執行一分散式軟體程式之程式碼之一第一部分時且在第二處理單元接收及執行相同分散式軟體程式之程式碼之一第二部分時，兩個或兩個以上處理單元可為分散式處理單元。

在一些實施方案中，系統100之不同處理器晶片可形成分散式硬體系統之各自節點。在替代實施方案中，一單個處理器晶片可包含可各形成該處理器晶片之各自節點之一或多個處理器核心及硬體特徵。

例如，在一中央處理單元(CPU)之背景內容中，一處理器晶片可包含至少兩個節點且各節點可為該CPU之一各自核心。替代性地，在一圖形處理器單元(GPU)之背景內容中，一處理器晶片可包含至少兩個節點且各節

點可為該GPU之一各自串流多處理器。運算系統100可包含多個處理器組件。在一些實施方案中，該等處理器組件可為一處理器晶片、一處理器核心、一記憶體存取引擎或整個運算系統100之至少一硬體組件之至少一者。

在一些例項中，諸如一處理器核心之一處理器組件可為經組態以基於執行程式碼之至少一經發出指令執行至少一特定操作之一固定功能組件。在其他例項中，諸如一記憶體存取引擎(MAE)之一處理器組件可經組態以依比藉由系統100之其他處理器組件執行之程式碼低的細節或粒度等級執行程式碼。

例如，藉由一處理器核心執行之程式碼可引起一MAE描述符產生並傳輸/發送至該MAE。在接收該描述符之後，MAE可基於該MAE描述符執行一資料傳送操作。在一些實施方案中，藉由MAE執行之資料傳送可包含(例如)經由系統100之特定資料路徑或介面組件往返於該系統之特定組件移動資料，或將資料請求發出至系統100之一實例性組態匯流排上。

在一些實施方案中，系統100之一實例性處理器晶片之各張量節點可具有可為處理程式指令之硬體區塊/特徵之至少兩個「前端」。如下文更詳細論述，一第一前端可對應於第一處理器核心104，而一第二前端可對應於第二處理器核心106。因此，該第一處理器核心及該第二處理器核心亦可在本文中描述為第一前端104及第二前端106。

如本說明書中所使用，一追蹤鏈可為其上可放置追蹤項目以傳輸至系統100內之一實例性晶片管理器之一特定實體資料通信匯流排。經接收之追蹤項目可為包含多個位元組及多個二進位值或數字之資料字組/結構。因此，描述符「字組」指示可作為一單位藉由一實例性處理器核心之

硬體裝置處置之一固定大小之二進位資料段。

在一些實施方案中，分散式硬體追蹤系統之處理器晶片係各執行該晶片之各自核心中之程式碼之若干部分之多核心處理器(即，具有多個核心)。在一些實施方案中，程式碼之若干部分可對應於用於一實例性多層類神經網路之推理工作負載之向量化運算。而在替代實施方案中，程式碼之若干部分可大體上對應於與習知程式設計語言相關聯之軟體模組。

運算系統100大體上包含一節點管理器102、一第一處理器核心(FPC) 104、一第二處理器核心(SPC) 106、一節點組構(fabric) (NF) 110、一資料路由器112及一主機介面區塊(HIB) 114。在一些實施方案中，系統100可包含經組態以執行信號切換、多工及解多工功能之一記憶體多工器108。系統100進一步包含一張量核心116，該張量核心116包含安置於其中之FPC 104。張量核心116可為經組態以對多維資料陣列執行向量化運算之一實例性運算裝置。張量核心116可包含與一矩陣單元(MXU) 120、轉置單元(XU) 122及縮減及排列單元(RPU) 124互動之一向量處理單元(VPU) 118。在一些實施方案中，運算系統100可包含一習知CPU或GPU之一或多個執行單元，諸如載入/儲存單元、算術邏輯單元(ALU)及向量單元。

系統100之組件共同包含一大組硬體效能計數器以及促成該等組件內之追蹤活動的完成之支援硬體。如下文更詳細描述，藉由系統100之各自處理器核心執行之程式碼可包含用於在程式碼執行期間同時啟用多個效能計數器之嵌入式觸發。一般而言，經偵測觸發引起針對一或多個追蹤事件產生追蹤資料。該追蹤資料可對應於儲存於計數器中且可經分析以辨別程式碼之效能特性之遞增參數計數。針對各自追蹤事件之資料可儲存於一實

例性儲存媒體(例如，一硬體緩衝器)中且可包含回應於觸發之偵測而產生之一時間戳記。

此外，可針對在系統100之硬體組件內發生之各種事件產生追蹤資料。實例性事件可包含節點間及跨節點通信操作，諸如直接記憶體存取(DMA)操作及同步旗標更新(各在下文更詳細描述)。在一些實施方案中，系統100可包含大體上被稱為全域時間計數器(「GTC」)之一全域同步時間戳記計數器。在其他實施方案中，系統100可包含其他類型之全域時脈，諸如一Lamport時脈。

GTC可用於執行於一分散式處理環境中之軟體/程式碼之程式碼執行與效能之精確相互關聯。此外且與GTC部分有關，在一些實施方案中系統100可包含藉由分散式軟體程式使用以依一高度協調方式開始及停止一分散式系統中之資料追蹤之一或多個觸發機制。

在一些實施方案中，一主機系統126編譯可包含在偵測時觸發以引起與硬體事件相關聯之追蹤資料之擷取及儲存之嵌入式運算元之程式碼。在一些實施方案中，主機系統126將該經編譯之程式碼提供至系統100之一或多個處理器晶片。在替代實施方案中，可藉由一實例性外部編譯器編譯程式碼(具有嵌入式觸發)且將該程式碼載入至系統100之一或多個處理器晶片。在一些例項中，編譯器可設定與嵌入於軟體指令之部分中之特定觸發相關聯之一或多個追蹤位元(下文進行論述)。經編譯之程式碼可為藉由系統100之一或多個組件執行之一分散式軟體程式。

主機系統126可包含經組態以監測藉由系統100之一或多個組件之程式碼之執行之一監測引擎128。在一些實施方案中，監測引擎128使主機系統126能夠監測藉由至少FPC 104及SPC 106執行之程式碼之執行。例

如，在程式碼執行期間，主機系統126可至少藉由接收基於所產生之追蹤資料之硬體事件之週期時間線而經由監測引擎128監測該執行程式碼之效能。儘管針對主機系統126展示一單個區塊，然在一些實施方案中，系統126可包含與系統100之多個處理器晶片或晶片核心相關聯之多個主機(或主機子系統)。

在其他實施方案中，涉及至少三個處理器核心之跨節點通信可使主機系統126在資料訊務橫穿FPC 104與一實例性第三處理器核心/節點之間的一通信路徑時監測一或多個中間「躍點(hop)」處之資料訊務。例如，FPC 104及該第三處理器核心可為在給定時間段執行程式碼之僅有核心。因此，自FPC 104至第三處理器核心之一資料傳送可在將資料自FPC 104傳送至第三處理器核心時針對SPC 106處之一中間躍點產生追蹤資料。換言之，在系統100中之資料路由期間，自一第一處理器晶片去往一第三處理器晶片之資料可需要橫穿一第二處理器晶片，且因此該資料路由操作之執行可引起在該第二晶片中針對路由活動產生追蹤項目。

在執行經編譯之程式碼時，系統100之組件可互動以產生在一分散式電腦系統中發生之硬體事件之時間線。該等硬體事件可包含節點內及跨節點通信事件。一分散式硬體系統之實例性節點及其等之相關聯通信在下文參考圖2更詳細描述。在一些實施方案中，產生針對至少一硬體事件時間線識別一硬體事件集合之一資料結構。該時間線實現分散式系統中發生之事件的重建。在一些實施方案中，事件重建可包含基於在一特定事件之發生期間產生之時間戳記之分析之正確事件排序。

一般而言，一實例性分散式硬體追蹤系統可包含系統100之上述組件以及與一主機系統126相關聯之至少一主機控制器。當(例如)以一時間排

序或序列化方式使事件資料相互關聯時，自一分散式追蹤系統獲得之資料之效能或除錯可係有用的。在一些實施方案中，在對應於經連接之軟體模組之多個所儲存硬體事件經儲存且接著序列化以藉由主機系統126進行結構化分析時可發生資料相互關聯。對於包含多個主機系統之實施方案，可(例如)藉由主機控制器執行經由不同主機獲得之資料之相互關聯。

在一些實施方案中，FPC 104及SPC 106各為一多核心處理器晶片之相異核心；而在其他實施方案中，FPC 104及SPC 106係相異多核心處理器晶片之各自核心。如上文所指示，系統100可包含至少具有FPC 104及SPC 106之分散式處理器單元。在一些實施方案中，系統100之分散式處理器單元可包含經組態以執行一較大分散式軟體程式或程式碼之至少一部分之一或多個硬體或軟體組件。

資料路由器112係提供系統100之組件之間的資料通信路徑之一晶片間互連件(ICI)。特定言之，路由器112可提供FPC 104與SPC 106之間及與核心104、106相關聯之各自組件之間的通信耦合或連接。節點組構110與資料路由器112互動以在系統100之分散式硬體組件及子組件內移動資料封包。

節點管理器102係管理多節點處理器晶片中之低階節點功能之一高階裝置。如下文更詳細論述，一處理器晶片之一或多個節點可包含藉由節點管理器102控制以管理及儲存本端項目日誌中之硬體事件資料之晶片管理器。記憶體多工器108係可對提供至一實例性外部高頻寬記憶體(HBM)之資料信號或自該外部HBM接收之資料信號執行切換、多工及解多工操作之一多工裝置。

在一些實施方案中，在多工器108在FPC 104與SPC 106之間切換時

可藉由多工器108產生一實例性追蹤項目(下文進行描述)。記憶體多工器108可潛在影響無法存取多工器108之一特定處理器核心104、106之效能。因此，藉由多工器108產生之追蹤項目資料可幫助理解與各自核心104、106相關聯之特定系統活動之延時中之所得尖峰。在一些實施方案中，可在一實例性硬體事件時間線中對起始於多工器108內之硬體事件資料(例如，下文所論述之追蹤點)連同針對節點組構110之事件資料進行分組。在特定追蹤活動引起針對多個硬體組件之事件資料儲存於一實例性硬體緩衝器(例如，下文論述之追蹤項目日誌218)中時，可發生事件分組。

在系統100中，效能分析硬體涵蓋FPC 104、SPC 106、多工器108、節點組構110、資料路由器112及HIB 114。此等硬體組件或單元之各者包含硬體效能計數器以及硬體事件追蹤設施及功能。在一些實施方案中，VPU 118、MXU 120、XU 122及RPU 124並不包含其等自身專用效能硬體。而是，在此等實施方案中，FPC 104可經組態以對VPU 118、MXU 120、XU 122及RPU 124提供所需計數器。

VPU 118可包含支援與一實例性矩陣向量處理器之向量元素相關聯之局部高頻寬資料處理及算術運算之一內部設計架構。MXU 120係經組態以對被乘數之向量資料集執行(例如)高達128x128矩陣乘法之一矩陣相乘單元。

XU 122係經組態以對與矩陣乘法運算相關聯之向量資料執行(例如)高達128x128矩陣轉置操作之一轉置單元。RPU 124可包含一積分(sigma)單元及一排列單元。該積分單元對與矩陣乘法運算相關聯之向量資料執行循序縮減。該等縮減可包含求和及各種類型之比較運算。該排列單元可充分排列或複製與矩陣乘法運算相關聯之向量資料之所有元素。

在一些實施方案中，藉由系統100之組件執行之程式碼可表示機器學習、類神經網路推理運算及/或一或多個直接記憶體存取功能。系統100之組件可經組態以執行包含引起該系統之一(若干)處理單元或裝置執行一或多個功能之指令之一或多個軟體程式。術語「組件」意欲包含任何資料處理裝置或儲存裝置(諸如控制狀態暫存器)或能夠處理及儲存資料之任何其他裝置。

系統100可大體上包含多個處理單元或裝置，該等處理單元或裝置可包含一或多個處理器(例如，微處理器或中央處理單元(CPU))、圖形處理單元(GPU)、特定應用積體電路(ASIC)或不同處理器之一組合。在替代實施例中，系統100可各包含提供用於執行與本說明書中所描述之硬體追蹤功能有關之運算之額外處理選項之其他運算資源/裝置(例如，基於雲端之伺服器)。

處理單元或裝置可進一步包含一或多個記憶體單元或記憶體庫(例如，暫存器/計數器)。在一些實施方案中，處理單元執行儲存於系統100之裝置之記憶體中之程式設計指令以執行本說明書中所描述之一或多個功能。記憶體單元/記憶體庫可包含一或多個非暫時性機器可讀儲存媒體。該非暫時性機器可讀儲存媒體可包含固態記憶體、磁碟及光碟、一隨機存取記憶體(RAM)、一唯讀記憶體(ROM)、一可擦除可程式化唯讀記憶體(例如，EPROM、EEPROM或快閃記憶體)或能夠儲存資訊之任何其他有形媒體。

圖2繪示用於藉由系統100執行之分散式硬體追蹤之實例性追蹤鏈及各自實例性節點200、201之一方塊圖。在一些實施方案中，系統100之該等節點200、201可為一單個多核心處理器內之不同節點。在其他實施方

案中，節點200可為一第一多核心處理器晶片中之一第一節點且節點201可為一第二多核心處理器晶片中之一第二節點。

儘管在圖2之實施方案中描繪兩個節點，但在替代實施方案中，系統100可包含多個節點。對於涉及多個節點之實施方案，跨節點資料傳送可在橫穿多個節點之沿著一實例性資料路徑之中間躍點處產生追蹤資料。例如，中間躍點可對應於通過一特定資料傳送路徑中之相異節點之資料傳送。在一些例項中，可針對在通過一或多個節點之跨節點資料傳送期間發生之一或多個中間躍點產生與ICI追蹤/硬體事件相關聯之追蹤資料。

在一些實施方案中，節點0及節點1係用於與針對推理工作負載之程式碼之部分相關聯之向量化運算之張量節點。如本說明書中所使用，一張量係一多維幾何物件且實例性多維幾何物件包含矩陣及資料陣列。

如圖2之實施方案中所展示，節點200包含與系統100之組件之至少一子組互動之一追蹤鏈203。同樣地，節點201包含與系統100之組件之至少一子組互動之一追蹤鏈205。在一些實施方案中，節點200、201係相同組件子組之實例性節點，而在其他實施方案中，節點200、201係相異組件子組之各自節點。資料路由器/ICI 112包含大體上與追蹤鏈203及205會聚以將追蹤資料提供至晶片管理器216之一追蹤鏈207。

在圖2之實施方案中，節點200、201可各包含至少具有FPC 104、SPC 106、節點組構110及HIB 114之各自組件子組。節點200、201之各組件包含經組態以將藉由節點之一特定組件產生之追蹤點(下文進行描述)進行分組之一或多個追蹤多工器。FPC 104包含一追蹤多工器204，節點組構110包含追蹤多工器210a/b，SPC 106包含追蹤多工器206a/b/c/d，HIB 214包含追蹤多工器214且ICI 212包含追蹤多工器212。在一些實施方案中，用

於各追蹤多工器之一追蹤控制暫存器容許啟用及停用個別追蹤點。在一些例項中，對於一或多個追蹤多工器，其等對應追蹤控制暫存器可包含個別啟用位元以及更廣泛追蹤多工器控制。

一般而言，追蹤控制暫存器可為接收及儲存追蹤指令資料之習知控制狀態暫存器(CSR)。關於更廣泛追蹤多工器控制，在一些實施方案中，可基於藉由系統100執行之CSR寫入來啟用及停用追蹤。在一些實施方案中，可藉由系統100基於一全域時間計數器(GTC)之值、FPC 104 (或核心116)中之一實例性追蹤標記暫存器之值，或基於SPC 106中之一步驟標記之值而動態開始及停止追蹤。

與用於動態開始及停止追蹤活動以及用於同步化硬體事件收集之運算系統及電腦實施方法有關之細節及描述係描述於2017年3月29日申請且代理人檔案號碼為16113-8129001之標題為「Synchronous Hardware Event Collection」之相關美國專利申請案第15/472,932號中。美國專利申請案第15/472,932號之全部揭示內容以其全文引用方式明確併入本文中。

在一些實施方案中，對於核心116，FPC 104可使用一追蹤控制參數來定義與發生於核心116內之事件活動相關聯之一追蹤窗。該追蹤控制參數容許依據GTC之下限及上限以及追蹤標記暫存器之下限及上限來定義該追蹤窗。

在一些實施方案中，系統100可包含實現縮減所產生之追蹤項目之數目之功能(諸如追蹤事件篩選特徵)。例如，FPC 104及SPC 106可各包含限制各核心在一實例性所產生追蹤描述符(下文進行描述)中設定一追蹤位元之速率之篩選特徵。HIB 114可包含類似篩選特徵，諸如限制與特定DMA追蹤事件之擷取相關聯之追蹤位元之一實例性DMA速率限制器。此

外，HIB 114可包含(例如，經由一啟用位元之)用於限制哪些佇列供給DMA追蹤項目之控制。

在一些實施方案中，用於一DMA操作之一描述符可具有藉由主機系統126之一實例性編譯器設定之一追蹤位元。當設定該追蹤位元時，判定及產生追蹤資料之硬體特徵/旋鈕係用於完成一實例性追蹤事件。在一些例項中，DMA中之一最終追蹤位元可為藉由編譯器靜態插入之一追蹤位元與藉由一特定硬體組件動態判定之一追蹤位元之間的一邏輯OR運算。因此，在一些例項中，除篩選之外，編譯器產生之追蹤位元可提供用以縮減所產生之追蹤資料之總量之一機制。

例如，主機系統126之一編譯器可決定僅對於一或多個遠端DMA操作(例如，跨至少兩個節點之一DMA)設定追蹤位元且對於一或多個本端DMA操作(例如，諸如節點200之一特定張量節點內之一DMA)清除追蹤位元。以此方式，可基於限於跨節點(即，遠端) DMA操作之追蹤活動而非包含跨節點及本端DMA操作兩者之追蹤活動來縮減所產生之追蹤資料量。

在一些實施方案中，藉由系統100起始之至少一追蹤事件可與包含跨系統100發生之多個中間操作之一記憶體存取操作相關聯。用於該記憶體存取操作之一描述符(例如，一MAE描述符)可包含引起與該多個中間操作相關聯之資料儲存於一或多個記憶體緩衝器中之一追蹤位元。因此，該追蹤位元可用於給中間記憶體操作「加標籤」且在資料封包橫穿系統100時，在DMA操作之中間躍點處產生多個追蹤事件。

在一些實施方案中，ICI 112可包含針對節點200、201之一特定組件之各入口埠及出口埠提供控制功能性之一組啟用位元及一組封包篩選器。

此等啟用位元及封包篩選器容許ICI 112啟用及停用與節點200、201之特定組件相關聯之追蹤點。除了啟用及停用追蹤點之外，ICI 112亦可經組態以基於事件源、事件目的地及追蹤事件封包類型而篩選追蹤資料。

在一些實施方案中，除了使用步驟標記器、GTC或追蹤標記器之外，用於處理器核心104、106及HIB 114之各追蹤控制暫存器亦可包含一「所有人(everyone)」追蹤模式。此「所有人」追蹤模式可實現藉由追蹤多工器204或追蹤多工器206a控制跨一整個處理器晶片之追蹤。在該所有人追蹤模式中，追蹤多工器204及206a可發送指定特定追蹤多工器(多工器204或多工器206a)是否處於一追蹤窗內之一「窗內」追蹤控制信號。

可將該窗內追蹤控制信號廣播或普遍傳輸至(例如)一處理器晶片內或跨多個處理器晶片之所有其他追蹤多工器。至其他追蹤多工器的廣播可使得在多工器204或多工器206a執行追蹤活動時實現所有追蹤。在一些實施方案中，與處理器核心104、106及HIB 114相關聯之追蹤多工器各包含指定何時及/或如何產生「所有人追蹤」控制信號之一追蹤窗控制暫存器。

在一些實施方案中，大體上基於是否在針對DMA操作或控制訊息之資料字組中設定橫穿ICI/資料路由器112之一追蹤位元來啟用追蹤多工器210a/b及追蹤多工器212中之追蹤活動。DMA操作或控制訊息可為固定大小二進位資料結構，該等固定大小二進位資料結構在二進位資料封包內可具有基於特定境況或軟體條件設定之一追蹤位元。

例如，在FPC 104 (或SPC 106)中運用一追蹤類型DMA指令起始一DMA操作且起始器(處理器核心104或106)係處於一追蹤窗內時，將在該特定DMA中設定追蹤位元。在另一實例中，對於FPC 104，若FPC 104處於一追蹤窗內且啟用引起追蹤資料被儲存之一追蹤點，則用於將資料寫入

至系統100內之另一組件之控制訊息將使追蹤位元被設定。

在一些實施方案中，零長度DMA操作提供系統100內之一更廣泛DMA實施方案之一實例。例如，一些DMA操作可在系統100內產生非DMA活動。亦可追蹤該非DMA活動之執行(例如，產生追蹤資料)，如同該非DMA活動係一DMA操作般(例如，包含非零長度操作之DMA活動)。例如，在一源位置處起始但未發送或傳送任何資料(例如，零長度)之一DMA操作可代替性地發送一控制訊息至目的地位置。該控制訊息將指示在目的地處不存在待接收或處理之資料，且該控制訊息本身將如同一非零長度DMA操作將被追蹤般由系統100追蹤。

在一些例項中，對於SPC 106，零長度DMA操作可產生一控制訊息，且與該訊息相關聯之一追蹤位元僅在DMA使該追蹤位元被設定的情況下(即，在該控制訊息不具有一零長度的情況下)被設定。一般而言，若HIB 114處於一追蹤窗內，則自主機系統126起始之DMA操作將使追蹤位元被設定。

在圖2之實施方案中，追蹤鏈203接收針對與節點0對準之組件子組之追蹤項目資料，而追蹤鏈205接收針對與節點1對準之組件子組之追蹤項目資料。各追蹤鏈203、205、207係由各自節點200、201及ICI 112使用以提供追蹤項目資料至一晶片管理器216之一實例性追蹤項目資料日誌218之相異資料通信路徑。因此，追蹤鏈203、205、207之端點係其中追蹤事件可儲存於實例性記憶體單元中之晶片管理器216。

在一些實施方案中，晶片管理器216之至少一記憶體單元可為128位元寬且可具有至少20,000個追蹤項目之一記憶體深度。在替代實施方案中，至少一記憶體單元可具有一更大或更小位元寬度且可具有能夠儲存更

多或更少個項目之一記憶體深度。

在一些實施方案中，晶片管理器216可包含執行指令以管理所接收之追蹤項目資料之至少一處理裝置。例如，晶片管理器216可執行指令以掃描/分析針對經由追蹤鏈203、205、207接收之追蹤資料之各自硬體事件之時間戳記資料。基於該分析，晶片管理器216可填入追蹤項目日誌218以包含可用於識別(或產生)硬體追蹤事件之一時間排序序列之資料。當系統100之處理單元執行一實例性分散式軟體程式時，硬體追蹤事件可對應於在組件及子組件級發生之資料封包之移動。

在一些實施方案中，系統100之硬體單元可產生以一非時間排序方式(即，無序)填入一實例性硬體追蹤緩衝器之追蹤項目(及對應時間戳記)。例如，晶片管理器216可使具有所產生之時間戳記之多個追蹤項目被插入至項目日誌218中。該多個經插入追蹤項目之各自追蹤項目可未相對於彼此按時間排序。在此實施方案中，可藉由主機系統126之一實例性主機緩衝器接收非時間排序追蹤項目。在藉由該主機緩衝器接收時，主機系統126可執行與效能分析/監測軟體有關之指令以掃描/分析各自追蹤項目之時間戳記資料。該等經執行指令可用於將追蹤項目分類及建構/產生硬體追蹤事件之一時間線。

在一些實施方案中，可在經由一主機DMA操作之一追蹤工作階段期間自項目日誌218移除追蹤項目。在一些例項中，主機系統126無法如將項目增加至日誌一樣快地將項目DMA出追蹤項目日誌218。在其他實施方案中，項目日誌218可包含一預定義記憶體深度。若達到項目日誌218之記憶體深度限制，則額外追蹤項目可能丟失。為控制丟失哪些追蹤項目，項目日誌218可在先進先出(FIFO)模式中或替代性地在一覆寫記錄模式中

操作。

在一些實施方案中，可藉由系統100使用覆寫記錄模式來支援與事後剖析除錯相關聯之效能分析。例如，可執行程式碼達其中啟用追蹤活動及啟用覆寫記錄模式之一特定時間段。回應於系統100內之一事後剖析軟體事件(例如，一程式崩潰)，監測藉由主機系統126執行之軟體可分析一實例性硬體追蹤緩衝器之資料內容以深入瞭解在該程式崩潰之前發生之硬體事件。如本說明書中所使用，事後剖析除錯係關於在程式碼已崩潰或已大體上未能如預期般執行/操作之後之程式碼之分析或除錯。

在FIFO模式中，若項目日誌218係滿的，且若主機系統126確實移除一特定時間框內之經保存日誌項目以節省記憶體資源，則可不將新追蹤項目保存至晶片管理器216之一記憶體單元。而在覆寫記錄模式中，若項目日誌218因為主機系統126確實移除一特定時間框內之經保存日誌項目以節省記憶體資源而係滿的，則新追蹤項目可覆寫儲存於項目日誌218內之最舊追蹤項目。在一些實施方案中，回應於一DMA操作使用HIB 114之處理特徵將追蹤項目移動至主機系統126之一記憶體中。

如本說明書中所使用，一追蹤點係一追蹤項目及藉由晶片管理器216接收且儲存於追蹤項目日誌218中之與該追蹤項目相關聯之資料之產生器。在一些實施方案中，一多核心多節點處理器微晶片可包含該晶片內之三條追蹤鏈，使得一第一追蹤鏈自一晶片節點0接收追蹤項目，一第二追蹤鏈自一晶片節點1接收追蹤項目且一第三追蹤鏈自該晶片之一ICI路由器接收追蹤項目。

各追蹤點在其追蹤鏈內具有其插入至追蹤項目之標頭中之一唯一追蹤識別號碼。在一些實施方案中，各追蹤項目在藉由資料字組之一或多個

位元組/位元指示之一標頭中識別其起始於之追蹤鏈。例如，各追蹤項目可包含具有傳達關於一特定追蹤事件之資訊之經定義之欄位格式(例如，標頭、酬載等)之一資料結構。一追蹤項目中之各欄位對應於可應用於產生該追蹤項目之追蹤點之有用資料。

如上文所指示，可將各追蹤項目寫入至或儲存於與追蹤項目日誌218相關聯之晶片管理器216之一記憶體單元內。在一些實施方案中，可個別啟用或停用追蹤點且多個追蹤點儘管具有不同追蹤點識別符仍可產生相同類型之追蹤項目。

在一些實施方案中，各追蹤項目類型可包含一追蹤名稱、追蹤描述及識別針對追蹤項目內之特定欄位及/或一欄位集合之編碼之一標頭。名稱、描述及標頭共同提供追蹤項目所表示內容之一描述。從晶片管理器216的角度來看，此描述亦可識別一特定追蹤項目在其上進入一特定處理器晶片內之特定追蹤鏈203、205、207。因此，一追蹤項目內之欄位表示與該描述有關之資料段(例如，以位元組/位元為單位)且可為用於判定哪一追蹤點產生一特定追蹤項目之一追蹤項目識別符。

在一些實施方案中，與所儲存硬體事件之一或多者相關聯之追蹤項目資料可部分對應於a)至少在一節點0與節點1之間；b)至少在節點0內之組件之間；及c)至少在節點1內之組件之間發生的資料通信。例如，所儲存硬體事件可部分對應於1)節點0之FPC 104與節點1之FPC 104；節點0之FPC 104與節點0之SPC 106；2)節點1之SPC 106與節點1之SPC 106之至少一者之間發生的資料通信。

圖3繪示一實例性追蹤多工器設計架構300及一實例性資料結構320之一方塊圖。追蹤多工器設計300大體上包含一追蹤匯流排輸入302、一匯

流排仲裁器304及一本端追蹤點仲裁器306、一匯流排FIFO 308、至少一本端追蹤事件佇列310、一共用追蹤事件FIFO 312及一追蹤匯流排輸出314。

多工器設計300對應於安置於系統100之一組件內之一實例性追蹤多工器。多工器設計300可包含以下功能性。匯流排輸入302可與本端追蹤點資料有關，該本端追蹤點資料暫時儲存於匯流排FIFO 308內，直至仲裁邏輯(例如，仲裁器304)可引起該追蹤資料被放置於一實例性追蹤鏈上之時為止。針對一組件之一或多個追蹤點可將追蹤事件資料插入於至少一本端追蹤事件佇列310中。仲裁器306提供一級仲裁且實現自儲存於佇列310內之本端追蹤事件間選擇事件。選定事件被放置於亦充當一儲存佇列之共用追蹤事件FIFO 312中。

仲裁器304提供自FIFO佇列312接收本端追蹤事件及經由追蹤匯流排輸出314將該等本端追蹤事件合併至一特定追蹤鏈203、205、207上之二級仲裁。在一些實施方案中，追蹤項目可比其等可被合併至共用FIFO 312上更快地被推入本端佇列310中，或替代性地，追蹤項目可比其等可被合併至追蹤匯流排314上更快地被推入共用FIFO 312中。當此等案例發生時，各自佇列310及312將變得充滿追蹤資料。

在一些實施方案中，在佇列310或312變得充滿追蹤資料時，系統100可經組態使得最新追蹤項目被捨棄且未儲存至或合併至一特定佇列。在其他實施方案中，取代捨棄追蹤項目，當特定佇列(例如，佇列310、312)填滿時，系統100可經組態以使一實例性處理管線停滯，直至經填充之佇列再次具有可用於接收項目之佇列空間。

例如，可使使用佇列310、312之一處理管線停滯直至將一足夠或臨

限數目個追蹤項目合併至追蹤匯流排314上。該足夠或臨限數目可對應於導致供一或多個追蹤項目被佇列310、312接收之可用佇列空間之一特定數目個經合併追蹤項目。其中使處理管線停滯直至下游佇列空間變得可用之實施方案可基於特定追蹤項目被保留而非被捨棄而提供較高保真度追蹤資料。

在一些實施方案中，本端追蹤佇列係與追蹤項目所需一樣寬，使得各追蹤項目在本端佇列310中僅佔一位點。然而，共用追蹤FIFO佇列312可使用一唯一追蹤項目線編碼，使得一些追蹤項目可在共用佇列312中佔據兩個位置。在一些實施方案中，當捨棄一追蹤封包之任何資料時，捨棄完整封包使得在追蹤項目日誌218中不會出現部分封包。

一般而言，一追蹤係與系統100之一特定組件相關聯之活動或硬體事件之一時間線。不像作為彙總資料之效能計數器(下文進行描述)，追蹤含有提供對在一指定追蹤窗期間發生之硬體活動深入瞭解之詳細事件資料。所描述之硬體系統實現對分散式硬體追蹤之廣泛支援，包含追蹤項目的產生、追蹤項目在硬體管理之緩衝器中的暫時儲存、一或多個追蹤類型的靜態及動態啟用及追蹤項目資料至主機系統126的串流。

在一些實施方案中，可針對藉由系統100之組件執行之硬體事件產生追蹤，該等硬體事件諸如產生一DMA操作、執行一DMA操作、發出/執行特定指令或更新同步旗標。在一些例項中，追蹤活動可用於追蹤透過系統進行之DMA，或追蹤在一特定處理器核心上執行之指令。

系統100可經組態以產生自硬體事件之一時間線識別一或多個硬體事件322、324之至少一資料結構320。在一些實施方案中，資料結構320依事件之一時間排序序列配置與至少FPC 104及SPC 106相關聯之一或多個

硬體事件322、324。在一些例項中，系統100可將資料結構320儲存於主機系統126之一主機控制裝置之一記憶體庫中。資料結構320可用於評估藉由至少處理器核心104及106執行之程式碼之效能。

如藉由硬體事件324所展示，在一些實施方案中，一特定追蹤識別(ID)號碼(例如，追蹤ID ‘003)可與跨分散式處理器單元發生之多個硬體事件相關聯。該多個硬體事件可對應於一特定記憶體存取操作(例如，一DMA)，且該特定追蹤ID號碼係用於使一或多個硬體事件相互關聯。

例如，如藉由事件324所指示，針對一DMA操作之一單個追蹤ID可包含對應於該DMA中之多個不同點之多個時距(time step)。在一些例項中，追蹤ID ‘003可具有經識別為彼此相隔一段時間之一「經發出」事件、一「經執行」事件及一「經完成」事件。因此，在此方面，追蹤ID可進一步用於基於相互關聯及參考時距判定記憶體存取操作之一延時屬性。

在一些實施方案中，產生資料結構320可包含(例如)系統100比較一第一硬體事件子集中之各自事件之事件時間戳記與一第二硬體事件子集中之各自事件之事件時間戳記。產生資料結構320可進一步包含系統100部分基於該第一事件子集與該第二事件子集之間的該比較而提供一相關硬體事件集以呈現於資料結構中。

如圖3中所展示，資料結構320可識別指示一特定硬體事件322、324之一延時屬性之至少一參數。該延時屬性可指示該特定硬體事件之至少一持續時間。在一些實施方案中，藉由由主機系統126之一控制裝置執行之軟體指令產生資料結構320。在一些例項中，可回應於該控制裝置將追蹤項目資料儲存至主機系統126之一記憶體磁碟/單元而產生結構320。

圖4係指示針對藉由系統100執行之一直接記憶體存取(DMA)追蹤事

件之實例性追蹤活動之一方塊圖400。對於DMA追蹤，源自一第一處理器節點至一第二處理器節點之一實例性DMA操作之資料可經由ICI 112行進且可產生沿著資料路徑之中間ICI/路由器躍點。當該DMA操作橫穿ICI 112時，該DMA操作將在一處理器晶片內之各節點處及沿著各躍點產生追蹤項目。藉由此等所產生追蹤項目之各者擷取資訊以沿著節點及躍點重建DMA操作之一時間進展。

一實例性DMA操作可與圖4之實施方案中所描繪之程序步驟相關聯。對於此操作而言，一本端DMA將資料自與處理器核心104、106之至少一者相關聯之一虛擬記憶體402 (vmem 402)傳送至HBM 108。圖400中所描繪之編號對應於表格404之步驟且大體上表示節點組構110中之活動或藉由節點組構110起始之活動。

表格404之步驟大體上描述相關聯追蹤點。實例性操作將針對此DMA產生六個追蹤項目。步驟一包含自處理器核心至節點組構110之在該節點組構中產生一追蹤點之初始DMA請求。步驟二包含其中節點組構110要求處理器核心傳送在節點組構110中產生另一追蹤點之資料之一讀取命令。當vmem 402完成節點組構110之讀取時，該實例性操作並不具有針對步驟三之一追蹤項目。

步驟四包含節點組構110執行一讀取資源更新以引起處理器核心中之在該處理器核心中產生一追蹤點之一同步旗標更新。步驟五包含其中節點組構110向記憶體多工器108通知待寫入至HBM之即將來臨資料之一寫入命令。經由該寫入命令之該通知在節點組構110中產生一追蹤點，而在步驟六，寫入至HBM的完成亦在節點組構110中產生一追蹤點。在步驟七，節點組構110執行一寫入資源更新以引起處理器核心中之在該處理器核心

中(例如，在FPC 104中)產生一追蹤點之一同步旗標更新。除了該寫入資源更新之外，節點組構110亦可執行其中將針對DMA操作之資料完成傳訊回至處理器核心之一確認更新(「ack更新」)。該ack更新可產生類似於藉由寫入資源更新產生之追蹤項目之追蹤項目。

在另一實例性DMA操作中，當在起始節點之一節點組構110中發出一DMA指令時，產生一第一追蹤項目。可在節點組構110中產生額外追蹤項目以擷取用於讀取DMA之資料及將該資料寫入至外傳佇列之時間。在一些實施方案中，節點組構110可將DMA資料封包化成較小資料塊。對於封包化成較小塊之資料，可針對一第一資料塊及一最後資料塊產生讀取及寫入追蹤項目。視需要，除了該第一資料塊及該最後資料塊之外，可設定所有資料塊以產生追蹤項目。

對於可需要ICI躍點之遠端/非本端DMA操作，第一資料塊及最後資料塊可在沿著ICI/路由器112之各中間躍點中之入口點及出口點處產生額外追蹤項目。當DMA資料到達一目的地節點時，在該目的地節點處產生類似於先前節點組構110項目之追蹤項目(例如，讀取/寫入第一資料塊及最後資料塊)。在一些實施方案中，DMA操作之一最終步驟可包含與該DMA相關聯之經執行指令引起在目的地節點處對一同步旗標之一更新。當更新該同步旗標時，可產生指示DMA操作的完成之一追蹤項目。

在一些實施方案中，在各組件處於追蹤模式中時，藉由FPC 104、SPC 106或HIB 114起始DMA追蹤，使得追蹤點可被執行。系統100之組件可經由一觸發機制基於FPC 104或SPC 106中之全域控制而進入追蹤模式。追蹤點回應於與藉由系統100之組件之程式碼之執行相關聯之一特定動作或條件之發生而觸發。例如，程式碼之部分可包含可藉由系統100之

至少一硬體組件偵測之嵌入式觸發功能。

系統100之組件可經組態以偵測與藉由FPC 104或SPC 106之至少一者執行之程式碼之部分相關聯之一觸發功能。在一些例項中，該觸發功能可對應於以下項之至少一者：1)經執行之程式碼之一部分或模組中之一特定序列步驟；或2)藉由系統100之分散式處理器單元所使用之GTC指示之一特定時間參數。

回應於偵測觸發功能，系統100之一特定組件可起始、觸發或執行引起與一或多個硬體事件相關聯之追蹤項目資料儲存於該硬體組件之至少一記憶體緩衝器中之至少一追蹤點(例如，一追蹤事件)。如上文所提及，可接著藉由至少一追蹤鏈203、205、207將所儲存追蹤資料提供至晶片管理器216。

圖5係用於使用系統100之組件特徵及系統100之一或多個節點200、201之分散式硬體追蹤之一實例性程序500之一程序流程圖。因此，可使用包含節點200、201之系統100之以上提及之運算資源之一或多者實施程序500。

程序500以方塊502開始且包含運算系統100監測藉由一或多個處理器組件(包含至少FPC 104及SPC 106)執行之程式碼之執行。在一些實施方案中，可至少部分藉由多個主機系統或一單個主機系統之子系統監測產生追蹤活動之程式碼之執行。因此，在此等實施方案中，系統100可執行與針對跨分散式處理單元發生之硬體事件之追蹤活動之分析有關之多個程序500。

在一些實施方案中，一第一處理器組件經組態以執行經監測之程式碼之至少一第一部分。在方塊504，程序500包含運算系統100監測藉由一

第二處理器組件執行之程式碼之執行。在一些實施方案中，該第二處理器組件經組態以執行經監測之程式碼之至少一第二部分。

運算系統100之組件可各包含至少一記憶體緩衝器。程序500之方塊506包含系統100將識別一或多個硬體事件之資料儲存於一特定組件之至少一記憶體緩衝器中。在一些實施方案中，硬體事件跨包含至少第一處理器組件及第二處理器組件之分散式處理器單元發生。識別硬體事件之所儲存資料可各包含一硬體事件時間戳記及特徵化硬體事件之後設資料。在一些實施方案中，一硬體事件集合對應於一時間線事件。

例如，系統100可儲存識別部分對應於系統100內之一源硬體組件與系統100內之一目的地硬體組件之間的資料封包之移動之一或多個硬體事件之資料。在一些實施方案中，特徵化硬體事件之所儲存後設資料可對應於以下項之至少一者：1)一源記憶體位址、2)一目的地記憶體位址、3)與引起硬體事件被儲存之一追蹤項目有關之一唯一追蹤識別號碼、或4)與一直接記憶體存取(DMA)追蹤項目相關聯之一大小參數。

在一些實施方案中，儲存識別一硬體事件集合之資料包含將事件資料儲存於FPC 104及/或SPC 106之(例如)對應於至少一本端追蹤事件佇列310之一記憶體緩衝器中。該所儲存事件資料可指示可用於產生硬體事件之一較大時間線之硬體事件資料子集。在一些實施方案中，事件資料的儲存回應於FPC 104或SPC 106之至少一者執行與藉由系統100之組件執行之程式碼之部分相關聯之硬體追蹤指令而發生。

在程序500之方塊508，系統100產生自硬體事件集合識別一或多個硬體事件之一資料結構(諸如結構320)。該資料結構可依事件之一時間排序序列配置與至少第一處理器組件及第二處理器組件相關聯之一或多個硬體

事件。在一些實施方案中，資料結構識別針對一特定追蹤事件之一硬體事件時間戳記、與該追蹤事件相關聯之一源位址或與該追蹤事件相關聯之一記憶體位址。

在程序500之方塊510，系統100將所產生資料結構儲存於與主機系統126相關聯之一主機裝置之一記憶體庫中。在一些實施方案中，可藉由主機系統126使用所儲存資料結構以分析藉由至少第一處理器組件或第二處理器組件執行之程式碼之效能。同樣地，可藉由主機系統126使用所儲存資料結構以分析系統100之至少一組件之效能。

例如，使用者或主機系統126可分析資料結構以偵測或判定是否存在與程式碼內之一特定軟體模組之執行相關聯之一效能問題。一實例性問題可包含該軟體模組未在一經分配之執行時窗內完成執行。

此外，使用者或主機裝置126可偵測或判定系統100之一特定組件是否高於或低於一臨限效能位準操作。與組件效能有關之一實例性問題可包含一特定硬體組件執行特定事件，但產生超出結果資料之可接受參數範圍之結果資料。在一些實施方案中，結果資料可能與藉由系統100之執行實質上類似操作之其他相關組件產生之結果資料不一致。

例如，在程式碼的執行期間，可需要系統100之一第一組件完成一操作及產生一結果。同樣地，可需要系統100之一第二組件完成一實質上類似操作及產生一實質上類似結果。所產生資料結構之分析可指示該第二組件產生與藉由該第一組件產生之結果顯著不同之一結果。同樣地，該資料結構可指示第二組件之明顯超出可接受結果參數之一範圍之一結果參數值。此等結果可能指示系統100之第二組件之一潛在效能問題。

可在數位電子電路、有形體現之電腦軟體或韌體、電腦硬體(包含本

說明書中所揭示之結構及其等之結構等效物)或其等之一或多者之組合中實施本說明書中所描述之標的及功能操作之實施例。本說明書中所描述之標的之實施例可實施為一或多個電腦程式，即，在一有形非暫時性程式載體上編碼以藉由資料處理設備執行或控制資料處理設備之操作之電腦程式指令之一或多個模組。替代性地或此外，程式指令可在一人工產生之傳播信號(例如，一機器產生之電、光學或電磁信號)上予以編碼，該傳播信號經產生以編碼資訊用於傳輸至合適接收器設備以藉由一資料處理設備執行。電腦儲存媒體可為一機器可讀儲存裝置、一機器可讀儲存基板、一隨機或串列存取記憶體裝置或其等之一或多者之一組合。

可藉由執行一或多個電腦程式以藉由對輸入資料操作及產生(若干)輸出來執行功能之一或多個可程式化電腦來執行本說明書中所描述之程序及邏輯流程。亦可藉由專用邏輯電路(例如，一FPGA (場可程式化閘陣列)、一ASIC (特定應用積體電路)或一GPGPU (通用圖形處理單元))來執行該等程序及邏輯流程，且設備亦可實施為該專用邏輯電路。

適用於一電腦程式的執行之電腦包含(舉例而言，可基於)通用微處理器或專用微處理器或兩者或任何其他種類之中央處理單元。一般而言，一中央處理單元將自一唯讀記憶體或一隨機存取記憶體或兩者接收指令及資料。一電腦之基本元件係用於執行(performing或executing)指令之一中央處理單元及用於儲存指令及資料之一或多個記憶體裝置。一般而言，一電腦將亦包含用於儲存資料之一或多個大容量儲存裝置(例如，磁碟、磁光碟或光碟)，或可操作耦合以自該一或多個大容量儲存裝置接收資料或將資料傳送至該一或多個大容量儲存裝置，或該兩種情況。然而，一電腦未必具有此等裝置。

適用於儲存電腦程式指令及資料之電腦可讀媒體包含所有形式的非揮發性記憶體、媒體及記憶體裝置，舉例而言，包含：半導體記憶體裝置，例如，EPROM、EEPROM，及快閃記憶體裝置；磁碟，例如，內部硬碟或隨身碟。處理器及記憶體可藉由專用邏輯電路增補或併入專用邏輯電路中。

雖然本說明書含有許多特定實施方案細節，但此等細節不應被理解為限制任何發明或可主張之內容之範疇，而是被理解為描述可特定於特定發明之特定實施例之特徵。本說明書中在分離實施例之背景內容中所描述之特定特徵亦可組合實施於一單個實施例中。相反地，在一單個實施例之背景內容中描述之各種特徵亦可分別實施於多個實施例中或以任何合適子組合實施。此外，儘管特徵在上文可被描述為依特定組合起作用且甚至最初如此主張，然來自一所主張之組合之一或多個特徵在一些情況中可自該組合免除，且該所主張之組合可係關於一子組合或一子組合之變型。

類似地，雖然在圖式中依一特定順序描繪操作，但此不應被理解為需要依所展示之該特定順序或依循序順序來執行此等操作或需要執行所有經繪示之操作以達成所要結果。在特定境況中，多任務處理及平行處理可為有利的。此外，上文所描述之實施例中之各種系統模組及組件之分離不應被理解為在所有實施例中需要此分離，且應理解，所描述之程式組件及系統可大體上一起整合於一單個軟體產品中或封裝於多個軟體產品中。

已描述標的之特定實施例。其他實施例係在以下發明申請專利範圍之範疇內。例如，在發明申請專利範圍中敘述之動作可依不同順序執行且仍達成所要結果。作為一實例，附圖中所描繪之程序並不一定需要所展示之特定順序或循序順序來達成所要結果。在某些實施方案中，多任務處理

及平行處理可為有利的。

【符號說明】

100:運算系統/系統

102:節點管理器

104:第一處理器核心(FPC)/第一前端/核心/處理器核心

106:第二處理器核心(SPC)/第二前端/核心/處理器核心

108:記憶體多工器/多工器/高頻寬記憶體(HBM)

110:節點組構(NF)

112:資料路由器/路由器/晶片間互連件(ICI)

114:主機介面區塊(HIB)

116:張量核心/核心

118:向量處理單元(VPU)

120:矩陣單元(MXU)

122:轉置單元(XU)

124:縮減及排列單元(RPU)

126:主機系統/主機裝置

128:監測引擎

200:節點

201:節點

203:追蹤鏈

204:追蹤多工器/多工器

205:追蹤鏈

206a:追蹤多工器/多工器

206b:追蹤多工器

206c:追蹤多工器

206d:追蹤多工器

207:追蹤鏈

210a:追蹤多工器

210b:追蹤多工器

212:晶片間互連件(ICI)/追蹤多工器

214:主機介面區塊(HIB)/追蹤多工器

216:晶片管理器

218:追蹤項目日誌/追蹤項目資料日誌/項目日誌

300:追蹤多工器設計架構/追蹤多工器設計/多工器設計

302:追蹤匯流排輸入/匯流排輸入

304:匯流排仲裁器/仲裁器

306:本端追蹤點仲裁器/仲裁器

308:匯流排先進先出(FIFO)

310:本端追蹤事件佇列/佇列/本端佇列

312:共用追蹤事件先進先出(FIFO)/先進先出(FIFO)佇列/共用先進先出(FIFO)/佇列/共用追蹤先進先出(FIFO)佇列/共用佇列

314:追蹤匯流排輸出/追蹤匯流排

320:資料結構/結構

322:硬體事件

324:硬體事件/事件

400:方塊圖

402:虛擬記憶體

404:表格

500:程序

502:方塊

504:方塊

506:方塊

508:方塊

510:方塊

【發明申請專利範圍】

【請求項1】

一種由一硬體追蹤系統實施用於擷取描述硬體事件之事件資料(capturing event data describing hardware events)之方法，其中該硬體追蹤系統包括複數個硬體組件，該方法包括：

藉由該硬體追蹤系統之一硬體組件偵測一觸發(trigger)，其觸發了描述該等硬體事件之該事件資料之擷取，其中該觸發係被滿足以作為由該硬體追蹤系統中之至少一硬件組件執行一程式碼之一結果；

回應於偵測到該觸發係被滿足，將該事件資料儲存在該硬體追蹤系統之一儲存媒體中，其中該事件資料包括該等硬體事件之一時間線(timeline)，且該事件資料係特定於包括經由特定資料路徑在該硬體追蹤系統之一或多個硬體組件之間之一資料傳送(data transfer)之硬體操作；及

將用於藉由一主機分析該程式碼之執行之該事件資料提供至該主機。

【請求項2】

如請求項1之方法，其中該觸發係經編碼作為在該程式碼中使用與該硬體組件相關聯之一參數值之一嵌入式運算元(embedded operand)。

【請求項3】

如請求項1之方法，其中該硬體追蹤系統包括複數個硬體效能計數器，其中儲存該事件資料包括：

將該事件資料儲存至該複數個硬體效能計數器之一者，其中該經儲存之事件資料包括描述特定硬體事件之各別遞增參數計數

(incremental parameter counts)。

【請求項4】

如請求項1之方法，其中該事件資料包括一或多個追蹤點，其中該一或多個追蹤點之各者係經組態以產生與一特定硬體事件相關聯之一追蹤項目(trace entry)。

【請求項5】

如請求項1之方法，其中該觸發係與一或多個追蹤位元(trace bit)相關聯，其中該一或多個追蹤位元之至少一追蹤位元係藉由該主機而被插入(inserted)至該程式碼中或者藉由該硬體追蹤系統中之一特定硬體組件所動態判定。

【請求項6】

如請求項5之方法，其中該一或多個追蹤位元係經組態以提供用以縮減所擷取及儲存之事件資料之總量之一機制，以回應於偵測到該觸發係被滿足。

【請求項7】

一種硬體追蹤系統，其係用於擷取描述硬體事件之事件資料，該硬體追蹤系統包括複數個硬體組件，其包括：

一或多個處理裝置；及

一或多個非暫態機器可讀儲存裝置，其用於儲存可由該一或多個處理裝置執行之指令以引起操作之實施，其包括：

藉由該硬體追蹤系統之一硬體組件偵測一觸發，其觸發了描述該等硬體事件之該事件資料之擷取，其中該觸發係被滿足以作為由該硬體追蹤系統中之至少一硬件組件執行一程式碼之一結果；

回應於偵測到該觸發係被滿足，將該事件資料儲存在該硬體追蹤系統之一儲存媒體中，其中該事件資料包括該等硬體事件之一時間線，且該事件資料係特定於包括經由特定資料路徑在該硬體追蹤系統之一或多個硬體組件之間之一資料傳送之硬體操作；及

將用於藉由一主機分析該程式碼之執行之該事件資料提供至該主機。

【請求項8】

如請求項7之硬體追蹤系統，其中該觸發係經編碼作為在該程式碼中使用與該硬體組件相關聯之一參數值之一嵌入式運算元。

【請求項9】

如請求項7之硬體追蹤系統，其進一步包括複數個硬體效能計數器，其中儲存該事件資料包括：

將該事件資料儲存至該複數個硬體效能計數器之一者，其中該經儲存之事件資料包括描述特定硬體事件之各別遞增參數計數。

【請求項10】

如請求項7之硬體追蹤系統，其中該事件資料包括一或多個追蹤點，其中該一或多個追蹤點之各者係經組態以產生與一特定硬體事件相關聯之一追蹤項目。

【請求項11】

如請求項7之硬體追蹤系統，其中該觸發係與一或多個追蹤位元相關聯，其中該一或多個追蹤位元之至少一追蹤位元係藉由該主機而被插入至該程式碼中或者藉由該硬體追蹤系統中之一特定硬體組件所動態判定。

【請求項12】

如請求項11之硬體追蹤系統，其中該一或多個追蹤位元係經組態以提供用以縮減所擷取及儲存之事件資料之總量之一機制，以回應於偵測到該觸發係被滿足。

【請求項13】

如請求項7之硬體追蹤系統，其中該硬體追蹤系統進一步包括一追蹤事件篩選特徵(trace event filtering feature)，其經組態以縮減待由限制一追蹤位元之一值所產生之追蹤項目之數目。

【請求項14】

一或多個非暫態機器可讀儲存媒體，其用於儲存用以擷取描述在一硬體追蹤系統中之硬體事件之事件資料之指令，其中該硬體追蹤系統包括複數個硬體組件，該等指令可由一或多個電腦執行以引起操作之實施，該等操作之實施包括：

藉由該硬體追蹤系統之一硬體組件偵測一觸發，其觸發了描述該等硬體事件之該事件資料之擷取，其中該觸發係被滿足以作為由該硬體追蹤系統中之至少一硬件組件執行一程式碼之一結果；

回應於偵測到該觸發係被滿足，將該事件資料儲存在該硬體追蹤系統之一儲存媒體中，其中該事件資料包括該等硬體事件之一時間線，且該事件資料係特定於包括經由特定資料路徑在該硬體追蹤系統之一或多個硬體組件之間之一資料傳送之硬體操作；及

將用於藉由一主機分析該程式碼之執行之該事件資料提供至該主機。

【請求項15】

如請求項14之一或多個非暫態機器可讀儲存媒體，其中該觸發係經

編碼作為在該程式碼中使用與該硬體組件相關聯之一參數值之一嵌入式運算元。

【請求項16】

如請求項14之一或多個非暫態機器可讀儲存媒體，其中該硬體追蹤系統包括複數個硬體效能計數器，其中儲存該事件資料包括：

將該事件資料儲存至該複數個硬體效能計數器之一者，其中該經儲存之事件資料包括描述特定硬體事件之各別遞增參數計數。

【請求項17】

如請求項14之一或多個非暫態機器可讀儲存媒體，其中該事件資料包括一或多個追蹤點，其中該一或多個追蹤點之各者係經組態以產生與一特定硬體事件相關聯之一追蹤項目。

【請求項18】

如請求項14之一或多個非暫態機器可讀儲存媒體，其中該觸發係與一或多個追蹤位元相關聯，其中該一或多個追蹤位元之至少一追蹤位元係藉由該主機而被插入至該程式碼中或者藉由該硬體追蹤系統中之一特定硬體組件所動態判定。

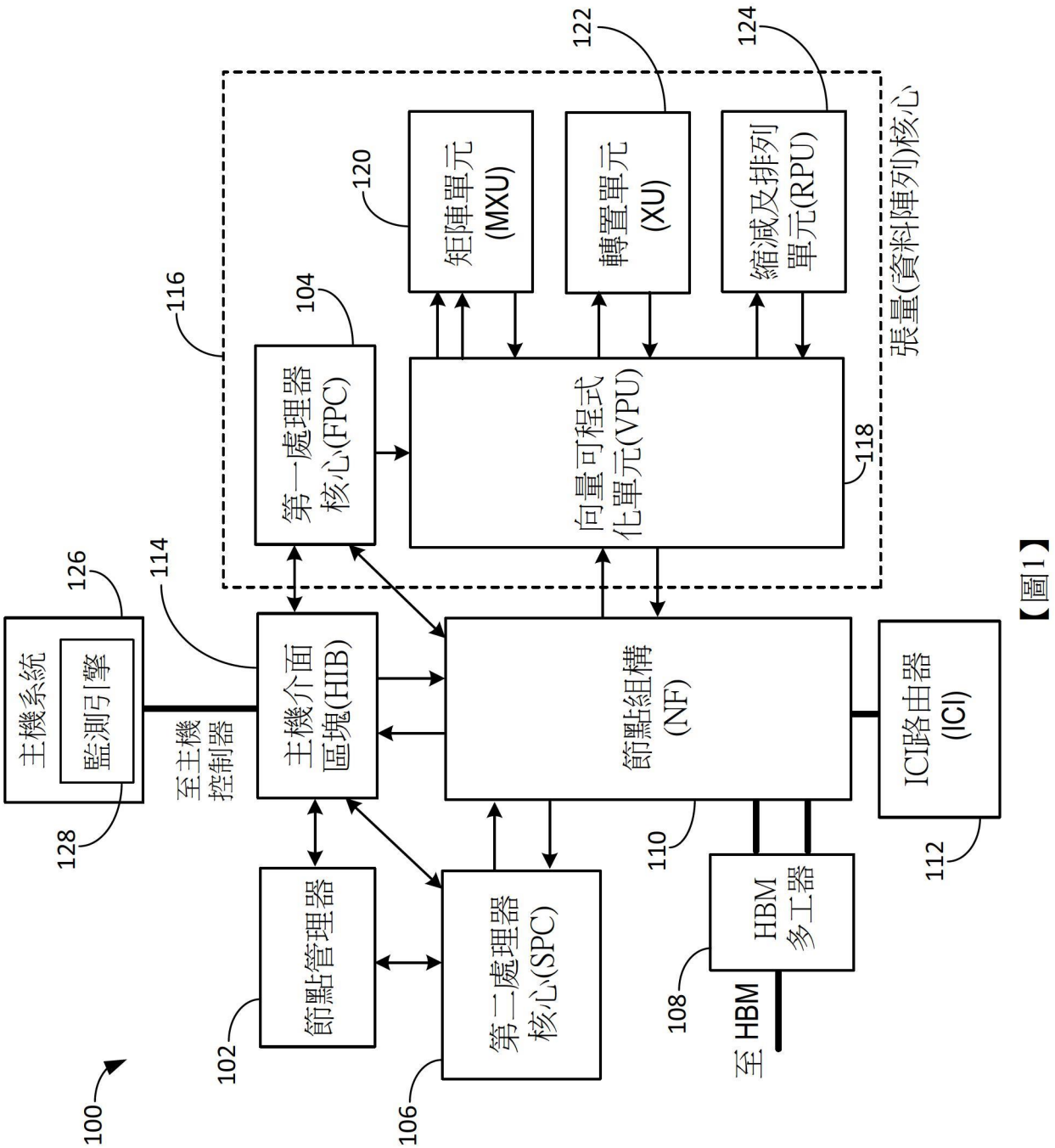
【請求項19】

如請求項18之一或多個非暫態機器可讀儲存媒體，其中該一或多個追蹤位元係經組態以提供用以縮減所擷取及儲存之事件資料之總量之一機制，以回應於偵測到該觸發係被滿足。

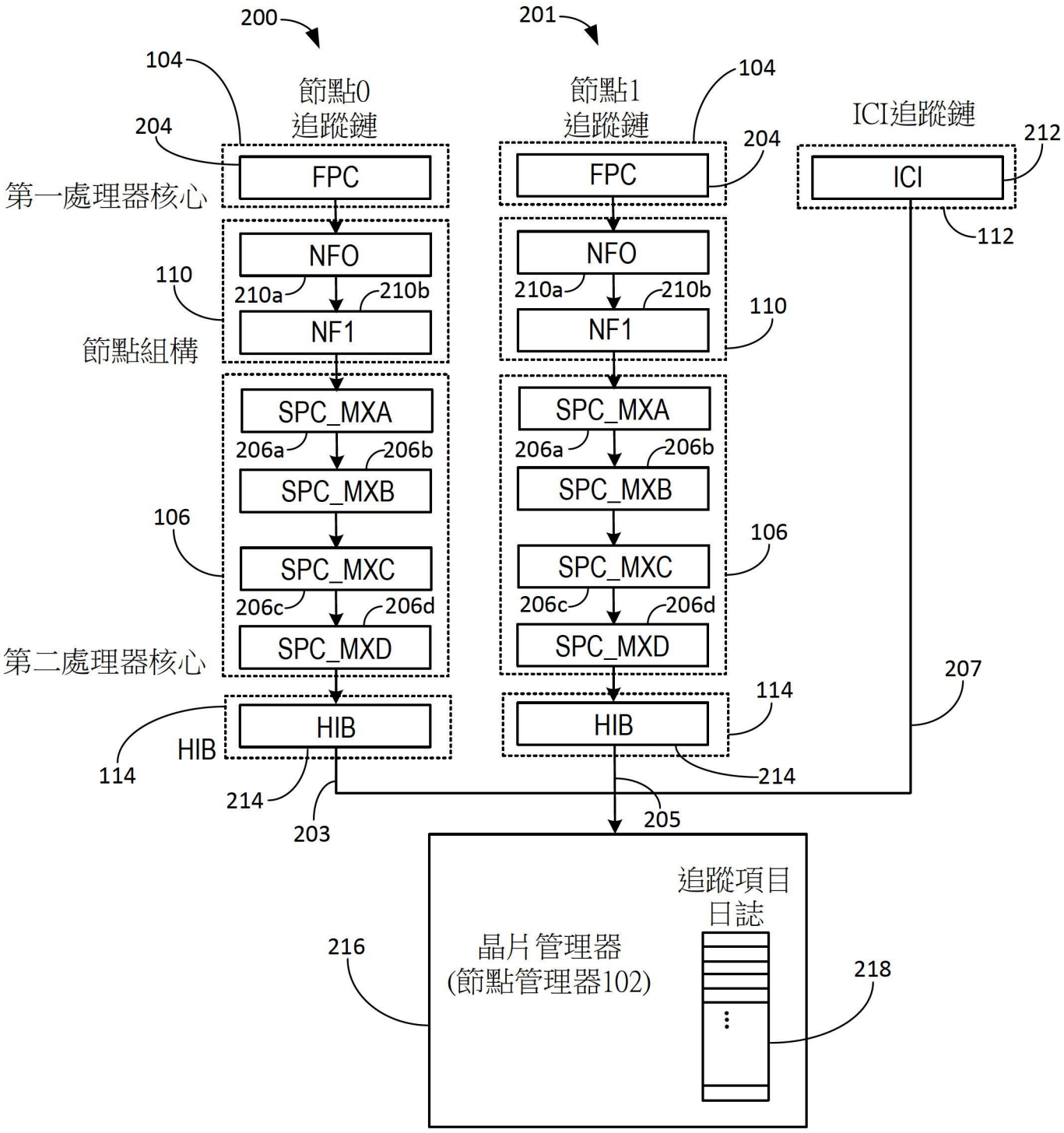
【請求項20】

如請求項14之一或多個非暫態機器可讀儲存媒體，其中該硬體追蹤系統進一步包括一追蹤事件篩選特徵，其經組態以縮減待由限制一追蹤位元之一值所產生之追蹤項目之數目。

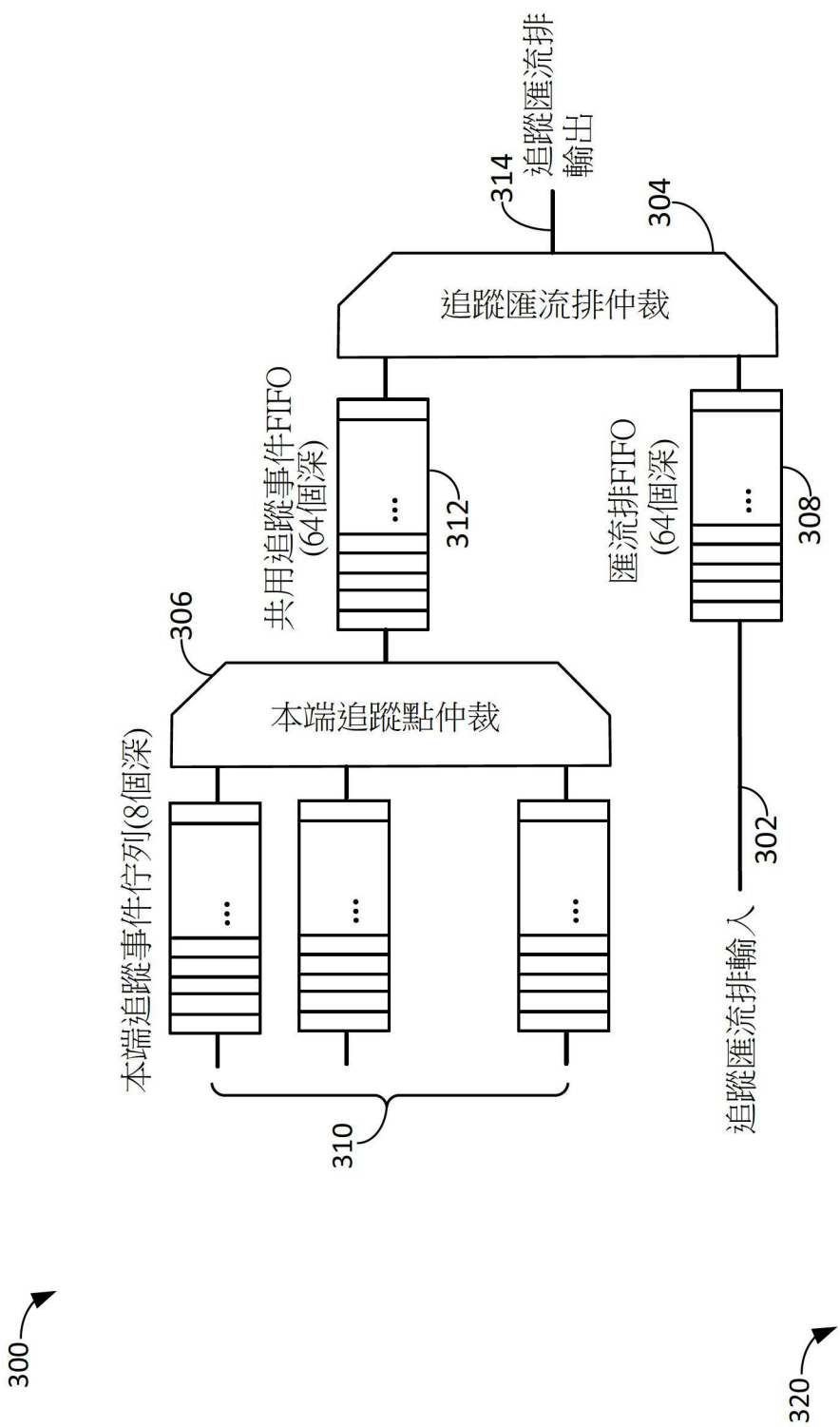
【發明圖式】


主機系統
監測引擎
至主機控制器
主機介面區塊(HIB)
第一處理器核心(FPC)
矩陣單元(MXU)
轉置單元(XU)
縮減及排列單元(RPU)
向量可程式化單元(VPU)
張量(資料陣列)核心
節點組構(NF)
ICI路由器(ICI)
節點管理器
第二處理器核心(SPC)
HBM多工器
至 HBM
100
102
104
106
108
110
112
114
116
118
120
122
124
126
128

【圖1】


200
201
104
節點0
追蹤鏈
104
節點1
追蹤鏈
ICI追蹤鏈
212
204
FPC
204
FPC
ICI
第一處理器核心
112
NFO
NFO
110
210a
210b
210a
210b
NF1
NF1
110
節點組構
SPC_MXA
SPC_MXA
206a
206b
206a
206b
SPC_MXB
SPC_MXB
106
106
SPC_MXC
SPC_MXC
206c
206d
206c
206d
SPC_MXD
SPC_MXD
第二處理器核心
207
114
HIB
HIB
HIB
114
214
214
203
205
追蹤項目
日誌
216
晶片管理器
(節點管理器102)
218

【圖2】

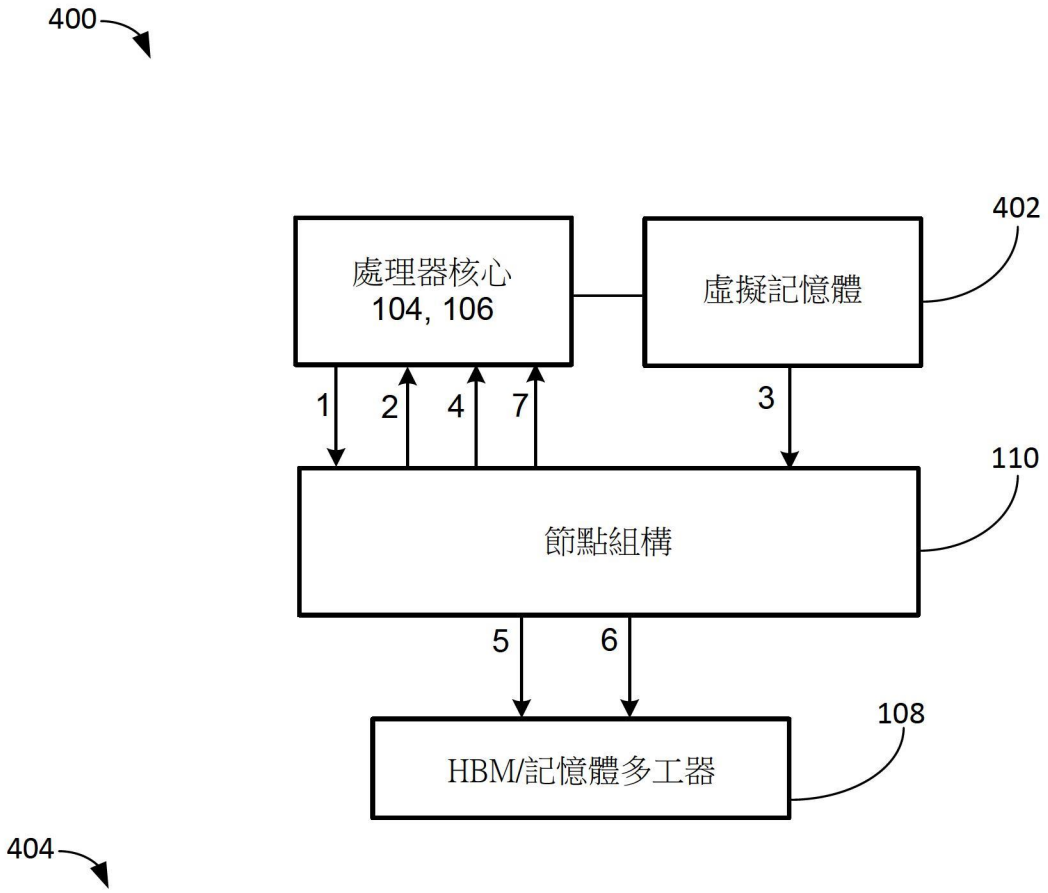


時間	追蹤ID	源	目的地	(若干)組件	追蹤事件	大小	延時
09:01.13	000001	0x00F100	0x00F200	FPC→ FPC	DMA	128個位元組	9μs
09:05.27	000002	0x00F300	0x00F400	FPC→ SPC	DMA	256個位元組	.023ms
09:12.35	000003	0x00F500	0x00F600	SPC→ NF	DMA (經發出)	512個位元組	.xxx secs
09:12.49	000003	0x00F600	0x00F700	NF →HBM	DMA (經執行)	512個位元組	.xxx secs
09:12.58	000003	0x00F700	0x00F800	NF →FPC	DMA (經完成)	512個位元組	.xxx secs

322

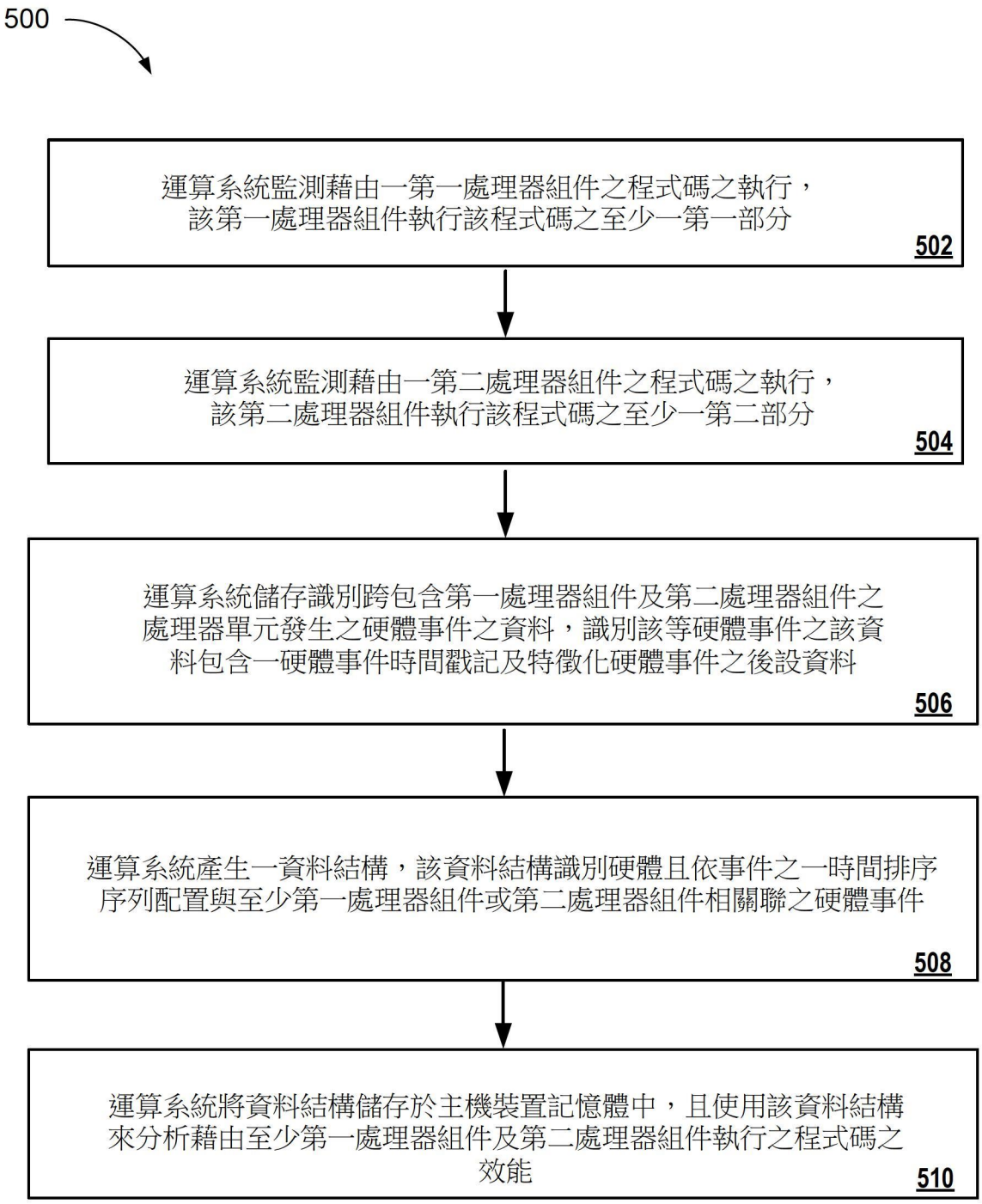
324

【圖3】



步驟	操作
1	初始DMA請求；節點組構中之追蹤點
2	讀取CMD：NF要求核心傳送資料；NF中之追蹤點
3	讀取完成：此時在NF中無追蹤點！
4	讀取資源更新：核心中之同步旗標更新；FPC中之追蹤點
5	寫入CMD：NF通知HBM；NF中之追蹤點
6	寫入完成：NF中之追蹤點
7	寫入資源更新：FPC中之同步旗標更新：FPC中之追蹤點

【圖4】


500
運算系統監測藉由一第一處理器組件之程式碼之執行，該第一處理器組件執行該程式碼之至少一第一部分
502
運算系統監測藉由一第二處理器組件之程式碼之執行，該第二處理器組件執行該程式碼之至少一第二部分
504
運算系統儲存識別跨包含第一處理器組件及第二處理器組件之處理器單元發生之硬體事件之資料，識別該等硬體事件之該資料包含一硬體事件時間戳記及特徵化硬體事件之後設資料
506
運算系統產生一資料結構，該資料結構識別硬體且依事件之一時間排序序列配置與至少第一處理器組件或第二處理器組件相關聯之硬體事件
508
運算系統將資料結構儲存於主機裝置記憶體中，且使用該資料結構來分析藉由至少第一處理器組件及第二處理器組件執行之程式碼之效能
510

【圖5】