



(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 I851297 B

(45)公告日：中華民國 113 (2024) 年 08 月 01 日

(21)申請案號：112123429

(22)申請日：中華民國 106 (2017) 年 10 月 30 日

(51)Int. Cl. : G06F7/22 (2006.01)

G06F7/00 (2006.01)

G06F9/46 (2006.01)

(30)優先權：2017/03/29 美國

15/472,932

(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 2010/0332909A1

US 2011/0047358A1

US 2011/0246521A1

US 2012/0095728A1

審查人員：許人偉

申請專利範圍項數：20 項 圖式數：7 共 73 頁

(54)名稱

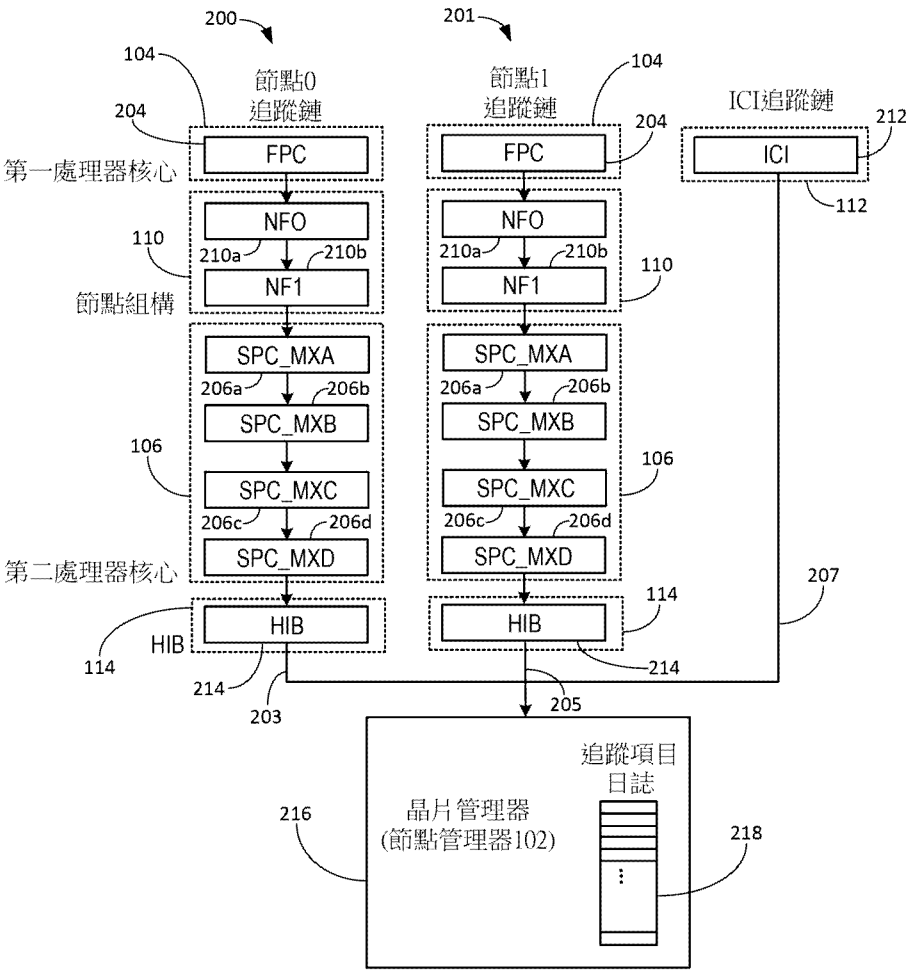
用於收集關於類神經網路運算之事件資料之方法、事件收集系統及非暫態電腦可讀儲存裝置

(57)摘要

本發明揭示一種電腦實施方法，其包含監測藉由第一處理器組件及第二處理器組件之程式碼之執行。一運算系統偵測藉由以下項滿足一觸發條件：i)識別該程式碼之一部分中之一運算元；或 ii) 判定該運算系統之一時脈之一當前時間指示一預定義時間值。該運算元及該預定義時間值係用於起始追蹤事件。當滿足該觸發條件時，該系統起始產生識別跨該運算系統發生之各自硬體事件之追蹤資料之追蹤事件。該系統使用該追蹤資料來產生一相關追蹤資料集。該相關追蹤資料指示該等各自硬體事件之一時間排序序列。該系統使用該相關追蹤資料集來分析該執行程式碼之效能。

A computer-implemented method that includes monitoring execution of program code by first and second processor components. A computing system detects that a trigger condition is satisfied by: i) identifying an operand in a portion of the program code; or ii) determining that a current time of a clock of the computing system indicates a predefined time value. The operand and the predefined time value are used to initiate trace events. When the trigger condition is satisfied the system initiates trace events that generate trace data identifying respective hardware events occurring across the computing system. The system uses the trace data to generate a correlated set of trace data. The correlated trace data indicates a time ordered sequence of the respective hardware events. The system uses the correlated set of trace data to analyze performance of the executing program code.

指定代表圖：



【圖2】

- 符號簡單說明：
- 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
 - 110:節點結構(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)
 - 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



I851297

【發明摘要】

【中文發明名稱】

用於收集關於類神經網路運算之事件資料之方法、事件收集系統及非暫態電腦可讀儲存裝置

【英文發明名稱】

METHOD, EVENT COLLECTION SYSTEM AND NON-TRANSITORY COMPUTER-READABLE STORAGE DEVICE FOR COLLECTING EVENT DATA ABOUT NEURAL NETWORK COMPUTATIONS

【中文】

本發明揭示一種電腦實施方法，其包含監測藉由第一處理器組件及第二處理器組件之程式碼之執行。一運算系統偵測藉由以下項滿足一觸發條件：i)識別該程式碼之一部分中之一運算元；或ii)判定該運算系統之一時脈之一當前時間指示一預定義時間值。該運算元及該預定義時間值係用於起始追蹤事件。當滿足該觸發條件時，該系統起始產生識別跨該運算系統發生之各自硬體事件之追蹤資料之追蹤事件。該系統使用該追蹤資料來產生一相關追蹤資料集。該相關追蹤資料指示該等各自硬體事件之一時間排序序列。該系統使用該相關追蹤資料集來分析該執行程式碼之效能。

【英文】

A computer-implemented method that includes monitoring execution of program code by first and second processor components. A computing system detects that a trigger condition is satisfied by: i) identifying an operand in a portion of the program code; or ii) determining that a current time of a clock of the computing

system indicates a predefined time value. The operand and the predefined time value are used to initiate trace events. When the trigger condition is satisfied the system initiates trace events that generate trace data identifying respective hardware events occurring across the computing system. The system uses the trace data to generate a correlated set of trace data. The correlated trace data indicates a time ordered sequence of the respective hardware events. The system uses the correlated set of trace data to analyze performance of the executing program code.

【指定代表圖】

圖2

【代表圖之符號簡單說明】

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

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)

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

【發明說明書】

【中文發明名稱】

用於收集關於類神經網路運算之事件資料之方法、事件收集系統及非暫態電腦可讀儲存裝置

【英文發明名稱】

METHOD, EVENT COLLECTION SYSTEM AND NON-TRANSITORY COMPUTER-READABLE STORAGE DEVICE FOR COLLECTING EVENT DATA ABOUT NEURAL NETWORK COMPUTATIONS

【技術領域】

【先前技術】

本說明書係關於分析程式碼之執行。

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新穎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一運算系統偵測基於以下項之至少一者滿足一觸發條件：i)識別至少該程式碼之該第一部分或該程式碼之該第二部分中之一運算元之出現，該運算元包含用於起始一或多個追蹤事件之一第一參數值；或ii)判定該運算系統之至少一時脈之一當前時間指示用於起始一或多個追蹤事件之一預定義時間值。

回應於偵測滿足觸發條件，方法包含藉由運算系統起始產生追蹤資料之至少一第一追蹤事件，該追蹤資料識別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之各自硬體事件。對於該等各自硬體事件之各者，追蹤資料至少包括一硬體事件時間戳記。方法進一步包含藉由運算系統使用該追蹤資料來產生指示在滿足觸發條件時產生之各自硬體事件之至少一時間排序序列之一相關追蹤資料集。

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，第一追蹤事件係產生識別跨分散式處理器單元發生之硬體事件之追蹤資料之一同步化追蹤事件，該追蹤資料亦識別用於各自硬體事件之一唯一追蹤識別符，且其中該等硬體事件包含複數個同步化硬體事件，且在該等事件共用一全域硬體事件時間戳記時使兩個硬體事件同步化。

在一些實施方案中，偵測滿足觸發條件包括：藉由運算系統偵測以下項之一者：i)運算元之第一參數值超過一暫存器之一第一限值，或ii)藉由當前時間指示之預定義時間值超過一暫存器之一第一限值；及回應於偵測滿足觸發條件，藉由運算系統起始產生追蹤資料之一第二追蹤事件，其中該追蹤資料識別在跨分散式處理器單元發生之各自硬體事件間共用之至

少一屬性。

在一些實施方案中，方法進一步包含：藉由運算系統偵測以下項之一者：運算元之一第二參數值超過暫存器之一第二限值，或藉由當前時間指示之一第二預定義時間值超過暫存器之一第二限值；及回應於偵測，在運算元之該第二參數值超過該第二限值時或在該第二預定義時間值超過該第二限值時藉由運算系統停止第二追蹤事件。

在一些實施方案中，運算元進一步包含以下項之至少一者：指示程式碼之一特定序列步驟之一參數；或指示分散式處理器單元之一特定效能狀態之一全域控制參數；且預定義時間值包含以下項之至少一者：藉由分散式處理器單元之一全域時脈指示之一特定時間值；或與該全域時脈相關聯之一預定義時間窗之一特定時間值。

在一些實施方案中，運算元具有一第一二進位資料結構且第一運算元之第一參數值對應於一追蹤標記，預定義時間值具有一第二二進位資料結構且當前時間係藉由一全域時脈指示，且其中由分散式處理器單元使用該全域時脈來產生一或多個硬體事件時間戳記。

在一些實施方案中，方法進一步包含：藉由運算系統之一編譯器將觸發條件之運算元插入至藉由第一處理器組件執行之程式碼之至少第一部分中；及藉由運算系統之該編譯器將觸發條件之至少一預定義時間值插入至藉由第二處理器組件執行之程式碼之至少第二部分中。

在一些實施方案中，起始第一追蹤事件或第二追蹤事件之至少一者包含：藉由運算系統產生藉由第一處理器核心之一第一計數暫存器接收之一第一控制信號，該第一控制信號引起與一第一硬體事件相關聯之資料儲存於該第一計數暫存器中；及藉由運算系統產生藉由第二處理器核心之一

第二計數暫存器接收之一第二控制信號，該第二控制信號引起與一第二硬體事件相關聯之資料儲存於該第二計數暫存器中。

在一些實施方案中，與第一硬體事件或第二硬體事件之一者相關聯之資料包括以下項之至少一者：寫入至分散式處理器單元之一特定處理器核心之一特定記憶體緩衝器之位元組之一數目；或藉由分散式處理器單元之一特定處理器核心執行之指令之一數目。

在一些實施方案中，方法進一步包含：藉由運算系統識別藉由第一處理器組件或第二處理器組件執行之程式碼之至少一部分中之一第二運算元之出現，該第二運算元包含一第二參數值；藉由運算系統判定基於該第二運算元之該第二參數值之以下情況之一者而滿足一篩選條件：超過一暫存器之一特定限值或低於該暫存器之一特定限值；及回應於判定滿足該篩選條件，藉由運算系統篩選一或多個追蹤事件，其中篩選該一或多個追蹤事件包括排除與一或多個硬體事件相關聯之追蹤資料之儲存。

本說明書中所描述之標的之另一新穎態樣可體現於一種硬體事件收集系統中，該硬體事件收集系統包含：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若干操作之指令，該等操作包括：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一運算系統偵測基於以下項之至少一者滿足一觸發條件：i)識別至少程式碼之第一部分或程式碼之第二部分中之一運算元之出現，該運算元包含用於起始一或多個追蹤事件之一第一參數值；或

ii)判定運算系統之至少一時脈之一當前時間指示用於起始一或多個追蹤事件之一預定義時間值。

回應於偵測滿足觸發條件，方法包含藉由運算系統起始產生追蹤資料之至少一第一追蹤事件，該追蹤資料識別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之各自硬體事件。對於該等各自硬體事件之各者，追蹤資料至少包括一硬體事件時間戳記。方法進一步包含藉由運算系統使用追蹤資料來產生指示在滿足觸發條件時產生之各自硬體事件之至少一時間排序序列之一相關追蹤資料集。

一般而言，本說明書中所描述之標的之一新穎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，監測藉由一處理器組件之程式碼之執行，該處理器組件經組態以執行該程式碼之至少一第一部分。

方法進一步包含藉由一運算系統偵測基於以下項之至少一者滿足一觸發條件：i)識別程式碼之至少第一部分中之一運算元之出現，該運算元包含用於起始一或多個追蹤事件之一第一參數值；或ii)判定運算系統之至少一時脈之一當前時間指示用於起始該一或多個追蹤事件之一預定義時間值。

回應於偵測滿足觸發條件，方法進一步包含：藉由運算系統產生藉由處理器組件之一計數暫存器接收之一控制信號，該控制信號引起與一硬體事件相關聯之計數資料儲存於該計數暫存器中；及藉由運算系統產生指示與執行程式碼相關聯之一或多個效能屬性之一資料結構，該資料結構係基於該所儲存計數資料之一或多個計數參數而產生。

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，計數暫存器係經組態以儲存關於運算系統之一或多個處理器組件之效能之計數資料之多個效能計數器之一者，且其中至少一效能計數器包含以下項之一者：一活動計數器、一停滯(stall)計數器、統計計數器或一取樣計數器。

在一些實施方案中，一或多個計數參數指示以下項之一者：i)藉由一特定處理器組件接收之指令之一數目；ii)藉由該特定處理器組件處理之指令之一數目；iii)藉由該特定處理器組件執行之指令之一數目；或iv)藉由該特定處理器組件執行之記憶體讀取之一數目或記憶體寫入之一數目。

在一些實施方案中，與執行程式碼相關聯之一效能屬性包括以下項之一者：i)執行該程式碼之一特定處理器組件之一停滯頻率；ii)該特定處理器組件之利用率低於一臨限利用率之一指示；或iii)藉由該特定處理器組件使用之一資料儲存佇列係處於或低於一佇列佔用臨限值之一指示。

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。觸發部分基於一全域時間計數器、嵌入式邏輯運算元及硬體暫存器而非純粹透過通常需要系統資源來監測及執行事件擷取之經寫碼序列來實現事件之同步化擷取。

同樣地，當硬體觸發經調諧以擷取同時發生而非作為不相關無效事

件擷取之短持續時間事件及同步化全域擷取時最佳化系統記憶體使用。用於較短持續時間同步化事件擷取之控制減輕資訊過載，藉此使一些記憶體資源能夠保持不使用且可用於其他系統程序。

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繪示用於一全域時間計數器(GTC)之一實例性資料結構及指示用於分散式硬體追蹤之一實例性運算系統對該GTC之使用案例之表格。

圖6繪示與用於分散式硬體追蹤之一實例性運算系統之組件相關聯之實例性時間計數器及硬體特徵之一方塊圖。

圖7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】

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第一處理器核心執行之程式碼之執行及藉由一

第二處理器核心執行之程式碼之執行。該運算系統將硬體事件之一時間線儲存於一記憶體緩衝器中。經儲存事件跨包含至少該第一處理器核心及該第二處理器核心之分散式處理器單元而發生。

時間線針對硬體事件之各者包含一事件時間戳記及特徵化該硬體事件之後設資料。系統自該時間線產生包含硬體事件之一資料結構。該系統將該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評估藉由第二處理器核心一或第二處理器核心執行之程式碼之效能。在事件追蹤之此背景內容中，本說明書描述圖1至圖4中所展示之用於一分散式運算系統中之硬體事件追蹤之方法。

本說明書進一步描述圖5至圖7中所展示之用於基於一或多個觸發機制之同步化硬體事件收集之方法。如下文將更詳細論述，本說明書中所描述之運算系統之態樣至少係關於追蹤資料及事件計數之協調/同步收集。特定言之，至少一態樣包含用於一獨立以及一分散式系統內之硬體效能計數器資料及追蹤事件資料之同步化收集之系統及方法。同步化事件收集增強對分散式程式碼之效能資料及除錯資料兩者之分析。部分透過回應於藉由系統100分析之經連接軟體組件/模組之執行而發生之事件之相互關聯實現該增強分析。

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一實例性處理器微晶片之組件及子組件內發生之事件的儲存。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之一集合，處理器微晶片之該集合協作以執行經組態以在處理器微晶片之該集合間分散式執行之一軟體/程式碼之各自部分。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、一節點結構(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128×128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 128×128 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、節點結構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，節點結構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、基於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中之一實例性追蹤標記暫存器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系統及方法有關之額外細節及描述係參考圖5至圖7之實施方案更詳細描述。

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。

如上文所提及，圖1至圖4已繪示用於一分散式運算系統中之硬體事件追蹤之方法。剩餘圖5至圖7描述用於一分散式運算系統中之同步化硬體事件收集之方法。硬體事件之該同步化收集可至少部分基於一或多個觸發機制。

圖5繪示一全域時間計數器(GTC) 502之一實例性資料結構及指示用於分散式硬體追蹤之一實例性運算系統(例如，系統100)使用該GTC 502之案例之表格504及506。在圖5之實施方案中，GTC 502係包含一60位元計數器及一4位元偏移之一64位元值。在替代實施方案中，GTC 502可為具有以位元為單位之在少於64個位元至多於64個位元之範圍內之一大小

之一固定大小之二進位資料段。

與GTC 502相關聯之資料結構應用於下文參考圖6之實施方案所描述之一全域主計數器以及本端時間計數器(LTC)。GTC 502之較高60個位元對應於每週期遞增一之一實際時間計數器，惟當「向上追趕(catching up)」或「向下追趕(catching down)」以解決GTC相位變動(下文進行描述)時除外。例如，在一本端GTC 502 (即，一LTC)之一更新期間，該本端GTC 502可不「以滴答聲記錄(tick)」達若干週期以「向下追趕(catching down)」一主全域GTC 502。在此等案例中，較低四個「偏移」位元係用於補償可在程式碼執行期間發生之GTC相位變動。

當解決GTC相位變動時，最後四個位元可在此等境況中遞增以部分維持針對本端硬體事件活動(其等之GTC原本係相同的)之時間戳記之間的區別。在大多數其他例項中，GTC 502之4位元偏移清除且因此未被使用。在一些實施方案中，偏移位元可計數高達15且接著停止。而在替代實施方案中，偏移位元可計數至如由一特定計數器設計所需要之任何整數值。在一些例項中，當GTC 502之較高60個位元再次開始以滴答聲記錄時清除偏移位元。

考慮GTC 502、602 (下文進行描述)之組態及管理，在一些實施方案中，經執行程式碼可藉由組態系統100之一或多個處理器晶片來組態或設置一GTC。組態該等處理器晶片可包含針對系統指定一全域主GTC以及針對系統100之各自節點或組件指定本端主GTC (即，LTC)。

組件可包含執行程式碼/軟體使用以運算及補償時脈延時及執行一或多個本端GTC (即，圖6之LTC)之手動調諧之硬體旋鈕/特徵。相對於主GTC預形成時脈延時及手動調諧的LTC補償以便最小化或減輕相位變動。

在一些實施方案中，系統100可包含其中組件硬體特徵藉由一「脈波-回波 (ping-and-echo)」機制自動判定連結延時之用於延時補償之一自動指示 (auto-pilot)模式。

在一些實施方案中，系統100可包含一連結/軟體堆疊，該連結/軟體堆疊包含用於調整一實例性控制/狀態暫存器(CSR)之參數值之程式碼之一部分或子集。在一些例項中，回應於該連結堆疊程式碼之執行而產生之控制信號引起一或多個參數值被寫入至該CSR。此等CSR參數值可用於組態 GTC 502、602或LTC之各種態樣。

例如，一或多個參數可包含：1)引起一主全域GTC、本端主GTC或一本端從GTC之選擇及一更新樣本大小之選擇之一GTC_config參數；2)用於重設一主GTC或清除一「脈波」量測之一GTC_sync參數；3)組態手動延時補償之一GTC_latency_compensation_control參數；4)指示一最小脈波延時之一GTC_link_ping參數；5)指示一實際GTC值之一GTC_count參數；及/或6)指示一本端GTC與主GTC之間的一最大觀測差異之一GTC_max_diff參數。

如上文所指示，可藉由系統100使用一全域主GTC 502及一或多個本端GTC 502以獲得事件時間戳記及判定在該系統內發生之硬體事件間之序列化排序。除了在活動間排序之外，GTC 502可用於判定活動之延時。在一些實施方案中，一特定GTC 502可呈現相位變動，因此，經判定延時可不準確達一或多個週期。

如上文所指示，一特定GTC 502遞增之速率可取決於一本端GTC是否不同於全域主GTC (即，當存在相位變動時)而改變。在一穩定狀態中，當不存在相位變動時，一主全域GTC 502每週期遞增一次且可正如一特定

組件LTC一樣好地用於計數(例如，一特定組件之)本端活動之延時及順序。

在一些實施方案中，一主GTC 502可比一特定LTC更慢地計數。在此等例項中，GTC 502之偏移位元繼續每本端時脈週期遞增一次。如藉由表格504、506所指示，給定GTC偏移位元之唯一功能性，一經擴展GTC(「GTC-ext」)可經定義以供系統100使用。在一些實施方案中，可藉由增加頂部60個位元(GTC:top)及底部四個偏移位元(GTC:offset)來建構GTC-ext，而在其他實施方案中，執行於一特定處理器中之程式碼可將GTC-ext運算為等於GTC:top加上一分數GTC:offset值。

在一些實施方案中且回應於偏移位元之遞增，一GTC-ext參數可用於判定事件排序以及本端操作之延時。表格504、506指示其中在判定硬體事件排序及事件操作之延時時該GTC-ext參數可證明為有用之使用案例。

表格504描述用於判定活動或硬體事件之間的順序之實例性GTC使用。在一些實施方案中，當使用一GTC 502來判定全域順序，例如，至少兩個相異運算節點(例如，各具有一主GTC之相異節點)之間的一或多個活動之順序時，GTC 502在一活動被分析之過程期間遞增之速率可不相關。然而，在一特定節點處之一本端GTC 502比節點主GTC慢時，GTC-ext可用於在兩個本端活動之間，或在兩個本端活動起始對遠端資源之寫入(如藉由表格504之特徵508所展示)時之間進行區別。

表格506描述用於運算本端活動/事件或跨越兩個或兩個以上節點之活動之延時之實例性GTC使用。在一些實施方案中，系統參數可能未指示一特定晶片處之一GTC 502是否按一正常速率(即，與一經指定主全域時脈同步)以滴答聲記錄或按與該經指定主全域時脈不同相之一速率以滴答

聲記錄。因此，在一些例項中，LTC可用於本端活動之事件延時判定，如藉由表格506之特徵510所展示。而在其他實施方案中，GTC或GTC-ext可用於量測一或多個操作之延時。

一般而言，在系統100之各種節點或組件處發生之硬體追蹤事件可包含一GTC參數。在一些實施方案中，當經追蹤硬體事件/活動跨越系統100之多個節點時，應經由GTC而非一LTC運算與本端活動相關聯之延時。在此等實施方案中，使用GTC或GTC-ext可為較佳計數器/時脈解決方案，此係因為跨兩個或兩個以上相異節點之LTC可能未同步化或按相同速率計時/以滴答聲記錄。

圖6繪示與用於分散式硬體追蹤之一實例性運算系統(例如，系統100)之各自組件相關聯之實例性時間計數器及一或多個計數器之一方塊圖。該等時間計數器可包含各與系統100之各自組件相關聯之全域時間計數器(GTC) 602及多個本端時間計數器(LTC)。

為清楚起見，儘管圖6描繪耦合至系統100之特定組件之一單個LTC區塊，然在替代實施方案中，各組件可包含一或多個本端時間計數器/LTC。此外，儘管圖6描繪一單個主GTC 602，然此特徵亦可對應於或表示經選定以充當全域主時脈之一特定LTC。例如，在一些實施方案中，GTC 602對應於藉由主機系統126實施以選擇一特定LTC充當一全域主時脈/計數器之計數器選擇邏輯。

在一些例項中，系統100可選擇ICI 112之一實例性節點來充當一全域主時脈且該實例性節點之LTC每時脈週期以滴答聲記錄達一。而在替代實施方案中，GTC 602係充當一全域主時脈且將一特定時間參數廣播至系統100之組件之一全域時間計數器。在任一實施方案中，GTC 602可為用

於使跨系統100發生之硬體事件相互關聯之一全域同步時脈。

在一些實施方案中，如上文所描述，GTC 602可具有跨系統100之小相位變動，但可經組態以避免長期時間漂移。例如，GTC 602可基於一特定處理器晶片之一實例性振盪器(例如，選定為全域主時脈之一振盪器)以滴答聲記錄或計數。一般而言，不存在特定於系統100中之一處理器晶片之該選定振盪器之長期漂移或相位變動，然而，隨時間漂移或變動可相對於相異處理器晶片之間的振盪器發生。因此，GTC 602可比一處理器晶片處之並非為全域主時脈之一本端時脈(LTC)更快或更慢地遞增(即，以滴答聲記錄)。

作為全域主時脈，GTC 602之一當前主值可向外廣播至ICI 112之所有節點。實例性ICI節點可對應於系統100之一或多個組件。接收節點可運算用於特定於特定組件/節點之本端操作之一實例性GTC之其數個本端值之間的一差。在一實例性取樣週期期間之經接收主值與本端GTC值之間的一最小差可用於更新本端GTC。隨著時間流逝，主值與本端GTC值之間的相位變動(若有)可基於使用上文論述之一或多個偏移位元實施之計數器調整來「向上追趕」。

如圖6之實施方案中所展示，系統100之組件可各包含一或多個效能計數器，且系統100內之效能計數器之集合組或子組在本文中被描述為多個計數器600。特定言之且特定於各組件，FPC 104可包含多個計數器604，SPC 106可包含多個計數器606，多工器記憶體(HBM) 108可包含多個計數器608，節點結構110可包含多個計數器610，ICI 112可包含多個計數器612且HIB 114可包含多個計數器614。

在系統100中，可基於與該系統之特定組件相關聯之一或多個硬體效

能計數器的使用來啟用/停用藉由一或多個處理器之程式碼執行之效能分析。此等效能計數器可對應於圖6之多個各自計數器且可包含以下項之至少一者：1)活動計數器；2)停滯計數器；3)統計計數器；及4)取樣計數器。

在一些實施方案中，系統100包含可經程式化以遞增及儲存與執行程式碼之技術態樣相關聯之指定計數資料之各種其他類型之效能計數器。此外，透過使用一或多個計數器，用於追蹤活動之硬體支援可包含(例如)與以下項有關之功能：指令追蹤事件資料的擷取；記憶體多工器追蹤事件資料的擷取；DMA追蹤事件資料的擷取；及追蹤緩衝器中之事件資料的擷取。

在一些實施方案中，多個計數器600之至少一子組可為可藉由主機系統126經由HIB 114存取之CSR或計數暫存器。此等計數暫存器/CSR可為經組態以儲存識別藉由一特定硬體組件接收、處理或執行之指令之各種類型之計數資料/資訊以及與執行程式碼之技術態樣相關聯之各種其他類型之資料/資訊之儲存裝置。在一些例項中，以與特定指令或處理器操作相關聯之計數或增值之形式(例如，一計數參數)儲存資料。在一些實施方案中，系統100之多個計數器600可對應於數百個或數千個效能計數器/計數暫存器。

關於上文提及之至少四種類型之效能計數器，在一些實施方案中，活動計數器可用於儲存及分析與針對不同硬體單元之利用率參數相關聯之資料。活動計數器亦可用於分析執行程式碼之部分或分析在程式碼執行期間經傳送之資料之數量或類型。例如，一些處理器核心可包含用於在各種指令串流中導出指令混合之活動計數器。例如，一活動計數器可經組態以

儲存與一實例性指令串流中之特定指令類型之執行相關聯之計數。

在一些實施方案中，活動計數器可包含計數器子組，諸如：在發出特定指令類型時遞增之發出計數器；可用於記憶體路徑上且回應於(例如)與(例如，介於多工器108與一VMEM 402之間的)記憶體傳送相關聯之讀取/寫入而遞增之記憶體活動計數器；及回應於指令之出現而遞增之一或多個其他一般活動計數器，諸如中斷、同步旗標及警報計數器。

關於停滯計數器，在一些實施方案中，此等計數器可提供指示歸因於一特定停滯原因使一特定硬體單元/組件停滯來代替進行與指令處理或資料運算相關聯之有用工作之週期之一數目之停滯計數。例如，當一特定硬體單元之利用率低於一臨限利用率，或程式碼未以一所要臨限效能位準執行時，可藉由系統100使用停滯計數器以獲得用於進行利用率及效能問題(例如，低於臨限利用率/效能)之根本原因分析之相關停滯資訊。

在一些實施方案中，停滯計數器可提供與以下項有關之相關資訊：基於經指示為不可用之特定正常可用記憶體路徑之偵測之停滯；基於一或多個同步旗標指令之延遲執行之停滯；及/或歸因於經指示為不可用之一或多個正常可用執行單元(例如，XU 122或RPU 124)之停滯。

在一些實施方案中，FPC 104可包含對藉由VPU 118執行之向量化運算提供純量數量之至少一純量單元。與該純量單元相關聯之停滯計數器可針對歸因於可在該純量單元內發生之一或多個危害之停滯提供相關資訊。在一些例項中，可發生各種危害類型且FPC 104可包含針對各危害類型之停滯計數器。實例性危害類型可包含DMA延遲危害、純量防禦延遲危害及/或純量延遲指令危害。

關於統計計數器，在一些實施方案中，系統100之組件可包含一或多

個資料佇列(例如，儲存媒體)且活動計數器可經組態以提供與佇列利用率有關之資訊。然而，在一些例項中，針對佇列之活動計數器可能不足以提供關於執行程式碼利用佇列之良好程度之所有資訊。在此等例項中，統計計數器可在一或多個組件中用於獲取特定資料計數，使得可判定與特定佇列中之平均佇列佔用率及所花費時間有關之運算。

在一些實施方案中，佇列統計計數器可提供與佇列佔用率、佇列插入狀態、佇列充滿狀態或處於臨限佔用狀態之佇列有關之統計計數。在一些例項中，當佇列統計計數器在作用中時，與一組件佇列相關聯之一佔用率計數器係在各處理器/指令週期中基於該佇列之當前佔用率遞增。

在其他例項中，當一作用中佇列統計計數器提供與佇列充滿狀態有關之計數時，該計數器在其中完全佔用一佇列之週期中遞增(例如，達一)經。此外，當一作用中佇列統計計數器提供與佇列插入狀態有關之計數時，該計數器回應於經放置於(即，佇列插入)該佇列中之資料塊而遞增。又此外，對於處於臨限佔用狀態之佇列，計數器可在其中當前佇列佔用率大於一特定佇列臨限佔用率之各週期中遞增(例如，達一)。

在一些實施方案中，系統100之組件可包含可自主機系統126週期性讀取以產生資料結構(例如，結構320或620)之取樣計數器組。該等資料結構可包含關於硬體事件、追蹤活動或針對發生於系統100之分散式處理器單元中之活動/事件之詳細資料處理資訊之經取樣統計資料。在一些實施方案中，基於經取樣統計資料產生之資料結構可包含，例如，結構320(上文進行描述)或結構620(下文進行描述)。在一些例項中，所產生資料結構係用於建置由主機系統126之監測引擎128使用以分析藉由系統100之組件執行之程式碼之效能之一或多個監測工具。

如上文所提及，系統100之組件可基於與一或多個觸發機制相關聯之全域控制起始事件追蹤。在一些實施方案中，可啟用、停用或以其他方式控制多個計數器600以起始引起追蹤點執行及產生包含與一或多個硬體事件有關之特定資料之追蹤項目之觸發機制。在一些實施方案中，與硬體事件有關之追蹤項目係填入有至少部分基於藉由上文所描述之多個計數器600之一或多者彙總之計數資料之事件資料。

全域控制大體上包含CSR控制(即，特定CSR硬體特徵/旋鈕)以及准許使計數器遞增及准許以一協調及同步化方式產生追蹤之觸發機制的使用。在一些實施方案中，協調及同步化控制可跨系統100之分散式處理器單元發生。如上文所論述，可部分透過使用用作一全域主時脈(例如，GTC 602)之一全域同步計數器來啟用同步化追蹤。

在一些實施方案中，系統100基於藉由分散式處理器單元之一全域時脈指示之一特定時間參數(例如，時間參數09:01.13)觸發同步化事件追蹤。例如，系統100可使用GTC 602之全域主時脈值作為一觸發以開始及停止所有計數器及追蹤精確達一特定操作之持續時間。

在一些例項中，系統100之一實例性軟體編譯器可將包含針對一觸發機制之運算元或參數值之指令嵌入或插入於執行程式碼內，該觸發機制係基於與GTC 602相關聯之一特定預定義時間窗。該預定義時間窗可包含具有一追蹤事件開始時間(例如，一第一GTC 602時間參數)及一追蹤事件結束時間(例如，在時間上遲於該第一GTC時間參數之一第二GTC 602時間參數)之一實例性追蹤持續時間。

如圖6之實施方案中所展示，系統100之組件可包含多個效能計數器600。由於其係關於觸發機制，所以效能計數器控制可包含用以啟用及停

用計數器以及清除計數器及暫停計數器之機制。此外，除了上文所描述之統計計數器功能之外，一些硬體單元可包含具有用於選擇該硬體單元內事件被計數之一硬體區塊子組之相關聯選擇器(例如，CSR)之統計計數器。

如上文參考追蹤多工器控制論述，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。在一些例項中，可藉由系統100基於FPC 104 (或核心116)中之一實例性追蹤標記(TM)暫存器之值，或基於SPC 106中之一實例性TM暫存器之值而動態開始及停止追蹤。

在一些實施方案中，一實例性追蹤控制參數可用於定義一特定追蹤窗。例如，該追蹤控制參數可准許依據GTC 602之下限及上限(例如，開始時間及停止時間時脈值)及FPC 104中之一實例性追蹤標記暫存器或SPC 106中之一實例性追蹤標記暫存器之下限及上限來定義追蹤窗。

在一些實施方案中，執行於系統100內之程式碼可包含用於在FPC 104或SPC 106之一者內在各處理器週期觸發追蹤事件之觸發處理邏輯。在一些例項中，觸發邏輯係與可藉由執行軟體/程式碼插入至一或多個指令串流中之一特殊「設定追蹤標記」指令。

例如，且關於用於FPC 104之觸發邏輯，當在FPC 104內發出設定一追蹤標記之一實例性指令時，觸發邏輯可包含分析該指令之一追蹤標記運算元且比較該運算元之一參數值與一實例性追蹤標記窗之值。此分析及比較係用於判定該運算元參數值是等於一下限值還是落在一特定追蹤標記窗之該下限值及一上限值範圍內。

若運算元之一參數值落在追蹤標記窗之界限內，則可激發一觸發(例如，滿足觸發條件)且FPC 104可進入一追蹤或計數模式且啟用多個計數器604之一或多個計數器。在一些實施方案中，當一後續「設定追蹤標

記」指令包含一運算元參數值(該運算元參數值引起一停止計數觸發基於該運算元值等於或超出一實例性追蹤標記窗之上限而激發)時，FPC 104退出追蹤/計數模式。

在一些實施方案中，系統100可偵測基於以下項之至少一者滿足一觸發條件：i)識別執行程式碼之至少一第一部分中之一實例性追蹤標記運算元之出現或ii)判定系統100之至少一時脈之一當前時間指示用於起始一或多個追蹤事件之一預定義時間值。該追蹤標記運算元可包含用於起始一或多個追蹤事件及/或效能計數之一參數值(例如，一程式碼步驟序列/值)。系統100亦可偵測基於一當前時間GTC 602指示用於起始一或多個追蹤事件及/或效能計數之一預定義時間值之一判定滿足一觸發條件。

回應於偵測滿足觸發條件，系統100之一處理器組件可產生藉由該處理器組件之一計數暫存器接收之一控制信號。該控制信號可引起與一或多個硬體事件相關聯之各種計數資料儲存於該計數暫存器中。在一些實施方案中，系統100可經組態或程式化以(例如)經由主機系統126產生指示與執行程式碼相關聯之一或多個效能屬性之一資料結構。例如，該資料結構可基於所儲存計數資料之一或多個計數參數而產生。在一些實施方案中，計數暫存器係經組態以儲存關於系統100之一或多個處理器組件之效能之計數資料之多個效能計數器600之一者。

為進一步繪示追蹤標記觸發，在一實例性程式碼序列中，設定追蹤標記指令可包含識別一分散式軟體建構之一32位元觸發運算元(例如，一實例性類神經網路推理工作負載中之一運算步驟)。FPC 104可包含在程式碼中被識別為「tracemark_limits」之一專用CSR。儘管32位元運算元係在此實例性程式碼序列中進行描述，然在替代實施方案中，觸發運算元或

暫存器限制運算元/參數值可為具有少於32位元或多於32位元之二進位資料結構。

此專用CSR可編碼至少兩個32位元值。一第一編碼32位元值可對應於用於開始追蹤/計數之一追蹤標記下限，而一第二編碼32位元值可對應於用於停止追蹤/計數之一追蹤標記上限。在一些實施方案中，第一編碼值及第二編碼值對應於一分散式軟體程式/建構(諸如一實例性類神經網路推理工作負載或任何其他分散式程式碼)中之運算步驟。

再次參考實例性程式碼序列，當執行「設定追蹤標記」指令時，比較32位元運算元值與

與追蹤標記觸發非常相似，當將GTC 602用作(例如，針對FPC 104或SPC 106之)一觸發時，可針對GTC 602之一當前值評估用於一特定組件之實例性GTC/LTC觸發暫存器之上限及下限。因此，可比較GTC 602之一當前時間參數或值與一實例性預定義時間窗之一下限以使用一GTC觸發執行追蹤活動。類似地，在執行GTC觸發以開始計數活動之後，可比較GTC 602之一當前時間參數或值與GTC觸發暫存器之上限以判定GTC 602是否匹配上限值或超出該上限值範圍。回應於GTC 602至少匹配該上限值，一停止計數觸發將激發以引起組件退出計數模式。

在一些實施方案中，SPC 106包含FPC 104之設定追蹤標記指令之一對應物。因此，用於SPC 106之觸發處理邏輯可以與FPC 104中相同或實

質上類似之方式起作用。例如，可包含一追蹤標記參數作為針對藉由SPC 106發出之一或多個指令之一運算元。在一些例項中，運算元之觸發評估在各指令發出時發生，且評估邏輯可實質上類似於在FPC 104中啟用計數之方式(例如，使用一追蹤標記運算元)。例如，在SPC 106中，當包含於一設定追蹤標記指令中之一追蹤標記運算元係在一實例性追蹤窗之下限及上限內時，啟用追蹤。同樣地，當包含於該追蹤標記指令中之該追蹤標記運算元超出該追蹤窗之下限及上限時，停用追蹤。

如上文所論述，除了使用GTC 602或追蹤標記器之外，系統100亦可包含一「所有人」追蹤模式。一般而言，此所有人追蹤模式可引起跨一整個處理器晶片追蹤由(例如)由FPC 104或SPC 106之一組件產生之一控制信號加以控制。在一些實施方案中，該控制信號可為指示系統100之分散式處理器單元之一特定效能狀態之一全域控制參數。在一些例項中，該特定效能狀態對應於「所有人」追蹤模式。

在一些實施方案中，一或多個硬體組件(例如，HIB 114或ICI 112)可經程式化或經組態以回應於「所有人」追蹤模式之激活而起始事件追蹤，或簡單忽略指示該所有人追蹤模式之一傳播控制信號。因此，儘管系統100可包含用來傳播一所有人追蹤控制信號之一傳訊機構，然系統100亦可包含用以引起至少一硬體組件忽視該控制信號之機構。

例如，系統100之一使用者(例如，一效能分析員)可能想要使用HIB 114來追蹤指示所有主機通信之資料封包。該使用者可(例如)經由一外部編譯器插入一實例性程式指令以引起HIB 114進入一「始終追蹤」模式，同時亦使用所有人追蹤模式來擷取來自系統100之其他硬體組件之資料之特定單個步驟序列。因此，當經程式化以在始終追蹤模式中執行時，HIB

114可忽略用於產生來自其他硬體組件之單個步驟序列追蹤資料之所有人追蹤控制信號。

在一些實施方案中，追蹤點可回應於與藉由系統100之組件之程式碼之執行有關之一特定動作或條件之發生而觸發。例如，一分散式軟體程式可包含各包含共用或重疊程式碼序列或步驟之多個軟體模組。例如，處理器核心104、106可各接收及執行一分散式軟體程式之部分且各程式部分可包含一實例性程式碼序列1-100。

如上文所指示，在一些實施方案中，系統100可經組態以包含具有或基於分散式程式碼之一特定序列步驟之一或多個觸發條件/機制。因此，系統100可經組態以回應於處理器核心104、106之各者在程式碼執行期間到達一特定經寫碼序列/步驟(例如，程式碼序列步驟33)而觸發一追蹤點。在一些例項中，經執行之分散式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。

例如，如上文所提及，系統100可經組態以偵測基於識別藉由FPC 104或SPC 106之一者執行之程式碼之至少一部分中之運算元之出現滿足一或多個觸發條件。回應於偵測滿足至少一觸發條件，系統100可起始產生追蹤項目/追蹤資料或計數資料之一或多個追蹤事件。在一些實施方案中，所產生追蹤項目可包含在跨分散式處理器單元發生之各自硬體事件間共用之至少一屬性(例如，一共用時間戳記或程式碼序列步驟)。

例如，且如藉由資料結構620所展示，所產生追蹤項目622可對應於一同步化追蹤事件且追蹤資料可包含針對各自硬體事件之唯一追蹤識別符及在至少兩個各自硬體事件間共用之一全域硬體事件時間戳記(例如，09:01.13)。因此，對於追蹤項目622，在各自硬體事件間共用之至少一屬

性可為硬體事件時間戳記。同樣地，當系統100基於一特定程式碼序列觸發追蹤點時，對於追蹤項目624而言，在各自硬體事件間共用之至少一屬性可為程式碼序列或追蹤標記步驟。

在一些實施方案中，追蹤項目可包含一時間戳記標頭或欄位。該時間戳記標頭可包含識別產生追蹤項目之時間之一48位元時間戳記。在一些例項中，時間戳記標頭可對應於GTC 502之低48個位元。

如上文所指示，在一些實施方案中，系統100之一編譯器可將針對多個觸發條件之一或多個運算元插入至藉由系統100之實例性處理器核心執行之程式碼之部分中。例如，與FPC 104相關聯之一編譯器可將針對觸發條件之一或多個運算元插入或嵌入至藉由FPC 104執行之程式碼之部分中。同樣地，與SPC 106相關聯之一編譯器可將針對觸發條件之一或多個運算元插入或嵌入至藉由SPC 106執行之程式碼之部分中。

一般而言，且如上文所提及，系統100經組態以偵測一或多個觸發條件，且回應於偵測該等條件起始至少一追蹤事件。在一些實施方案中，起始該至少一追蹤事件可包含系統100將(例如)一第一控制信號提供至FPC 104之一第一效能計數器/暫存器(例如，一CSR，諸如一活動計數器)。此第一控制信號可引起與至少一硬體事件相關聯之資料儲存於該第一效能計數器中。

同樣地，在一些實施方案中，起始至少一追蹤事件可進一步包含系統100將(例如)一第二控制信號提供至SPC 106之一第二效能計數器/暫存器(例如，一CSR，諸如一統計計數器)。此第二控制信號可引起與至少一硬體事件相關聯之資料儲存於該第二效能計數器中。

在一些實施方案中，與至少一硬體事件相關聯之資料可包含以下項

之一者：1)寫入至系統100之分散式處理器單元之一特定處理器核心之一特定記憶體緩衝器之位元組之一數目；或2)藉由系統100之分散式處理器單元之一特定處理器核心執行之指令之一數目。

一般而言，所描述之追蹤系統包含基於硬體之觸發作用機制。觸發可早期經組態以基於軟體語意激發。例如，針對觸發條件之運算元可插入或嵌入於軟體組件之前序碼及收尾碼處(例如，一分散式軟體系統中之函數、迴圈標頭(loop-header)或其他適當位置處)。如上文所提及，在一些例項中，藉由系統100中之一編譯器完成針對觸發條件之運算元之插入，而在其他例項中，可藉由程式設計員在藉由系統100之程式碼執行之前嵌入/插入觸發及觸發運算元，或可藉由主機系統126在即時程式碼執行期間嵌入觸發及觸發運算元。

例如，在一些實施方案中，可跨一分散式硬體系統(例如，在系統100之各晶片/節點中)藉由執行於主機系統126上之效能分析軟體來設置指示何時激發觸發之一條件。當觸發激發時，在一特定硬體單元中執行硬體追蹤及效能計數，且設置類似觸發條件以停止追蹤資料收集。

在其他實施方案中，系統100中之一觸發可為作為針對一指令之一運算元之一單調增加值。可藉由一處理器核心之一實例性編譯器將該指令插入於重要軟體建構之前序碼及收尾碼處。一般而言，對於執行程式碼而言，使指令中之觸發運算元始終存在的效能損失是微不足道的，此係因為在一觸發激發前不收集效能資料。

例如，一觸發條件可藉由寫入至一硬體組件中之一實例性暫存器來設置。為了評估或偵測觸發，每當遇到一觸發指令時，硬體組件可比較此暫存器中之值與該指令中之觸發條件之運算元值。一般而言，應遠在特定

經寫碼之序列步驟之前設置一觸發條件，使得分散式系統中之處理器晶片/節點皆不會過早地達到該觸發條件。

在一些例項中，硬體觸發實施方案需要在一實例性處理器晶片/節點之不同組件內傳播一觸發信號。觸發作用機制之至少一優點使一分散式軟體程式能夠在與另一(第二)節點處之一相關活動(例如，程式碼序列步驟33)截然不同之一時間在一第一節點上執行一活動(例如，程式碼序列步驟33)。在此實施方案中，系統100可維持針對藉由不同硬體組件或處理器節點在不同時間段執行之相同或實質上相關活動之硬體效能資料之準確收集。

在一些實施方案中，一實例性觸發使用情況可包含系統100在無需重新程式化、重設或重新插入一特定觸發條件之情況下收集多個追蹤資料集。例如，執行程式碼可包含產生其中設定一追蹤標記之相同運算元多次出現之一指令序列之軟體建構。在一些例項中，當系統100之一處理器在一迴圈中執行程式碼之一特定部分時，相同設定追蹤標記運算元可多次出現。在一程式迴圈之執行期間，事件追蹤可根據該程式迴圈之反覆開始及結束。

例如，可基於在一迴圈反覆開始時出現之一運算元值滿足一觸發條件，例如，觸發激發及事件追蹤開始。同樣地，基於在該迴圈反覆結束時出現之一運算元值可不再滿足一觸發條件，例如，事件追蹤停止。一般而言，一迴圈反覆之開始及一迴圈反覆之結束可對應於針對一給定軟體建構之一程式迴圈之邊界。因此，一追蹤工作階段可在該軟體建構之各執行(例如，各迴圈反覆)時開始及結束。

在一些實施方案中，可依以下方式判定適當觸發運算元值(例如，

GTC或追蹤標記)。如上文所提及，一分散式軟體程式可包含軟體模組，該等軟體模組各包含共用或重疊程式碼序列或步驟。在系統100內之分散式程式碼之執行期間，一效能分析員可經由系統100之一運算控制台讀取或分析一追蹤標記暫存器之一CSR值以判定(例如)一當前程式碼序列或步驟數。

在一些例項中，分析員重複讀取CSR值以判定序列或步驟之一平均增加速率。該經判定之增加速率可用於基於(例如)一追蹤標記運算元值之出現而預測性設定一或多個觸發條件。在一些實施方案中，當使用GTC時，一效能分析員可讀取或分析系統/機器(或硬體組件)時脈速度以預測或判定GTC增加之一速率。該分析員可接著使用該經判定速率以基於一特定GTC值之出現而設定一或多個觸發條件。

圖7係使用運算系統100及系統100之一或多個節點200、201進行分散式硬體追蹤之一實例性程序700之一程序流程圖。因此，可使用包含節點200、201之資源之系統100之以上提及之運算資源之一或多者實施程序700。

程序700以方塊702開始且包含運算系統100監測藉由一第一處理器核心執行之程式碼之執行。在一些實施方案中，該第一處理器核心經組態以執行經監測之該程式碼之至少一第一部分。在程序700之方塊704，系統100監測藉由一第二處理器核心執行之程式碼之執行。在一些實施方案中，該第二處理器核心經組態以執行經監測之該程式碼之至少一第二部分。

在方塊706，系統100偵測滿足至少一觸發條件。在一些實施方案中，系統100偵測基於識別藉由第一處理器核心或第二處理器核心之一者

執行之程式碼之至少一部分中之一運算元之出現滿足觸發條件。在一些實施方案中，執行於系統100內之程式碼包含用於觸發追蹤事件之觸發處理邏輯。

在一些例項中，觸發邏輯係與插入至一或多個指令串流中之一「設定追蹤標記」指令相關聯。該觸發邏輯可包含識別該指令之一追蹤標記運算元之出現，分析該追蹤標記運算元及比較該運算元與一實例性追蹤標記窗之值。在一些例項中，此分析及比較係用於偵測滿足第一觸發條件。

在其他例項中，觸發邏輯係與可藉由系統100之一外部使用者或藉由該系統之一實例性編譯器插入至一或多個指令串流中之一預定義時間值相關聯。該觸發邏輯可包含判定運算系統之至少一時脈之一當前時間指示用於起始一或多個追蹤事件之一預定義時間值。在一些實施方案中，判定該當前時間指示一預定義時間值可包含接收該預定義時間值及比較該預定義時間值與一實例性預定義時間窗之時間值。

在方塊708，回應於偵測滿足觸發條件，系統100起始產生追蹤資料之至少一第一追蹤事件。該所產生追蹤資料識別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各自硬體事件。

例如，回應於偵測滿足觸發條件，系統100可起始產生識別跨FPC 104及SPC 106發生之硬體事件之追蹤資料之一同步化追蹤事件。在一些實施方案中，追蹤資料識別針對各自硬體事件之一唯一追蹤識別符。在一些例項中，硬體事件包含多個同步化硬體事件，且在該等事件共用一全域硬體事件時間戳記時使該多個同步化事件之至少兩個硬體事件同步化。

在一些實施方案中，多個觸發條件可與執行程式碼相關聯。在一些例項中，可基於識別藉由至少第一處理器核心或第二處理器核心執行之程

式碼之至少一部分中之一特定運算元之出現來滿足一特定觸發條件。該特定運算元可包含可各用於起始一或多個追蹤事件之一第二追蹤事件之多個參數值之一者。在此實施方案中，偵測滿足該特定觸發條件可包含系統100偵測一特定參數值超過一暫存器之一第一限值。

在一些實施方案中，回應於偵測滿足特定觸發條件，系統100起始產生識別在跨分散式處理器單元發生之各自硬體事件間共用之至少一屬性之追蹤資料之一追蹤事件。在一些實施方案中，該等各自硬體事件之一屬性可包含至少一源記憶體位址、一目的地記憶體位址或一序列/程式運算步驟。然而至少一共用屬性可包含程式碼之一特定序列步驟或程式步驟。

在一些實施方案中，一或多個觸發條件可與以下項之至少一者相關聯：1)程式碼之一特定序列步驟；2)指示分散式處理器單元之一特定效能狀態之一全域控制參數；3)藉由該等分散式處理器單元之一全域時脈/計數器指示之一當前時間或特定時間參數；或4)與該全域時脈相關聯之一預定義時間窗。

在一些實施方案中，系統100亦可偵測特定運算元之特定參數值，或一特定預定義時間值超過暫存器之一第二限值。在一些例項中，回應於此偵測，在一特定運算元之特定參數值，或一特定預定義時間值超過暫存器之該第二限值時，系統100可停止第二追蹤事件。

在程序700之方塊710，系統100向一主機裝置提供包含各自硬體事件之一相關追蹤資料集。在一些實施方案中，該相關追蹤資料集指示在滿足第一觸發條件時產生之該等各自硬體事件之至少一時間排序序列。對於各自硬體事件之各者，追蹤資料至少包括針對硬體事件之一硬體事件時間戳記及特徵化該硬體事件之後設資料。在一些例項中，相關追蹤資料集至少

部分對應於資料結構620。

在方塊712，運算系統100使用相關追蹤資料集來分析藉由至少第一處理器核心及第二處理器核心執行之程式碼之效能。在一些實施方案中，藉由主機系統126使用資料結構620 (即，對應於相關追蹤資料)來分析藉由至少處理器核心104及106執行之程式碼之效能。同樣地，可藉由主機系統126使用資料結構620來分析系統100之至少一組件之效能。

例如，主機系統126或一實例性使用者可分析資料結構620以偵測或判定是否存在與程式碼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能在一經分配之執行時間窗內或低於一臨限延時完成特定指令串流之執行。

此外，使用者或主機裝置/系統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產生超出可接受臨限結果參數範圍之結果資料。在一些實施方案中，所產生結果資料可與藉由系統100之執行實質上類似指令/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。一相關追蹤資料集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，相關追蹤項目之一實例性資料結構可指示第二組件之超出可接受參數範圍之結果參數值。此等結果可指示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

502:全域時間計數器(GTC)/本端全域時間計數器(GTC)/主全域全域
時間計數器(GTC)/全域主全域時間計數器(GTC)/主全域時間計數器(GTC)

504:表格

506:表格

508:特徵

510:特徵

600:計數器/效能計數器

602:全域時間計數器(GTC)/主全域時間計數器(GTC)

604:計數器

606:計數器

608:計數器

610:計數器

612:計數器

614:計數器

620:結構/資料結構

622:追蹤項目

624:追蹤項目

700:程序

702:方塊

704:方塊

706:方塊

708:方塊

710:方塊

712:方塊

【發明申請專利範圍】

【請求項1】

一種用於收集事件資料之方法，其包括：

執行一程式碼序列以使用一處理器之多個組件來執行多個操作，該等組件之各組件具有指示一時間值之一時間計數器，其中各時間計數器包括用以補償當執行該等操作時該等時間計數器之間之相位變動之一個別偏移位元，其中該程式碼序列包括作為在執行該等操作時用於觸發跨該處理器之兩個或兩個以上組件之一追蹤事件之一觸發條件之一第一時間參數，且其中該等時間計數器之一者經選擇為一全域時間計數器；

判定由該全域時間計數器所代表之一全域時間值；

基於該全域時間值使用該個別偏移位元至少更新由未經選擇為該全域時間計數器之一對應時間計數器所代表之時間值；

藉由將該時間值與該程式碼序列中之該第一時間參數進行比較，以判定是否滿足該觸發條件；

回應於判定滿足該觸發條件，觸發該追蹤事件以產生用於該處理器之該等組件之該兩個或兩個以上者之事件資料，其中該事件資料基於該全域時間計數器被同步化；及

將該事件資料提供至用於分析該程式碼序列之一效能之一主機。

【請求項2】

如請求項1之方法，其中基於該全域時間值使用該個別偏移位元至少更新由未經選擇為該全域時間計數器之該對應時間計數器所代表之該時間值包括：

廣播由該全域時間計數器所代表之該全域時間值至未經選擇為該全

域時間計數器之該對應時間計數器；

計算由該對應時間計數器所代表之該時間值與該全域時間直之間之一差異；及

使用該個別偏移位元調整該對應時間計數器以減少該差異。

【請求項3】

如請求項1之方法，其中各時間計數器包括一固定大小二進位資料，該固定大小二進位資料包括一多位元計數器及該個別偏移位元，其中該個別偏移位元包括四個位元。

【請求項4】

如請求項1之方法，其中該程式碼序列進一步包括一第二時間參數；

其中該第二時間參數指示該第一時間參數之一較晚時間；及

其中該觸發條件包括一預定義時間窗，該預定義時間窗包括基於該第一時間參數用於觸發該追蹤事件之一開始時間，及基於該第二時間參數用於停止該追蹤事件之一結束時間。

【請求項5】

如請求項1之方法，其中該等組件之各組件包括一或多個效能計數器，該一或多個效能計數器經組態以收集用於分析該程式碼序列之該效能之計數資料，該一或多個效能計數器包括一活動計數器、一停滯(stall)計數器、一統計計數器或一取樣計數器之至少一者。

【請求項6】

如請求項5之方法，其中該一或多個效能計數器經組態以收集該計數資料包括：

收集包括指令追蹤事件資料的擷取、記憶體多工器追蹤事件資料的

擷取或直接記憶體存取追蹤事件資料之至少一者之該計數資料；

將該計數資料儲存於一或多個計數暫存器或控制/狀態暫存器。

【請求項7】

如請求項5之方法，其中：

該事件資料之一子集描述由該處理器實施之一或多個記憶體存取操作；及

該一或多個記憶體存取操作包括自一第一組件之一第一記憶體移動向量元素至該處理器之一第二組件之一第二記憶體。

【請求項8】

一種用於收集事件資料之事件收集系統，該系統包括：

一或多個處理裝置；及

一或多個非暫時性機器可讀儲存媒體，其或其等用於儲存可藉由該一或多個處理裝置執行以引起包括以下項之操作之效能之指令：

執行一程式碼序列以使用一處理器之多個組件來執行多個操作，該等組件之各組件具有指示一時間值之一時間計數器，其中各時間計數器包括用以補償當執行該等操作時該等時間計數器之間之相位變動之一個別偏移位元，其中該程式碼序列包括作為在執行該等操作時用於觸發跨該處理器之兩個或兩個以上組件之一追蹤事件之一觸發條件之一第一時間參數，其中該等時間計數器之一者經選擇為一全域時間計數器；

判定由該全域時間計數器所代表之一全域時間值；

基於該全域時間值使用該個別偏移位元至少更新由未經選擇為該全域時間計數器之一對應時間計數器所代表之時間值；

藉由將該時間值與該程式碼序列中之該第一時間參數進行比較，以判定是否滿足該觸發條件；

回應於判定滿足該觸發條件，觸發該追蹤事件以產生用於該處理器之該等組件之該兩個或兩個以上者之事件資料，其中該事件資料基於該全域時間計數器被同步化；及

將該事件資料提供至用於分析該程式碼序列之該效能之一主機。

【請求項9】

如請求項8之系統，其中基於該全域時間值使用該個別偏移位元至少更新由未經選擇為該全域時間計數器之該對應時間計數器所代表之該時間值包括：

廣播由該全域時間計數器所代表之該全域時間值至未經選擇為該全域時間計數器之該對應時間計數器；

計算由該對應時間計數器所代表之該時間值與該全域時間直之間之一差異；及

使用該個別偏移位元調整該對應時間計數器以減少該差異。

【請求項10】

如請求項8之系統，其中各時間計數器包括一固定大小二進位資料，該固定大小二進位資料包括一多位元計數器及該個別偏移位元，其中該個別偏移位元包括四個位元。

【請求項11】

如請求項8之系統，其中該程式碼序列進一步包括一第二時間參數；其中該第二時間參數指示該第一時間參數之一較晚時間；及

其中該觸發條件包括一預定義時間窗，該預定義時間窗包括基於該第一時間參數用於觸發該追蹤事件之一開始時間，及基於該第二時間參數用於停止該追蹤事件之一結束時間。

【請求項12】

如請求項8之系統，其中該等組件之各組件包括一或多個效能計數器，該一或多個效能計數器經組態以收集用於分析該程式碼序列之該效能之計數資料，該一或多個效能計數器包括一活動計數器、一停滯計數器、一統計計數器或一取樣計數器之至少一者。

【請求項13】

如請求項12之系統，其中該一或多個效能計數器經組態以收集該計數資料包括：

收集包括指令追蹤事件資料的擷取、記憶體多工器追蹤事件資料的擷取或直接記憶體存取追蹤事件資料之至少一者之該計數資料；

將該計數資料儲存於一或多個計數暫存器或控制/狀態暫存器。

【請求項14】

如請求項12之系統，其中：

該事件資料之一子集描述由該處理器實施之一或多個記憶體存取操作；及

該一或多個記憶體存取操作包括自一第一組件之一第一記憶體移動向量元素至該處理器之一第二組件之一第二記憶體。

【請求項15】

一或多個非暫時性電腦可讀儲存媒體，其或其等用於儲存可藉由一或多個處理裝置執行以引起用於收集事件資料之操作之效能之指令，該等

操作包括：

執行一程式碼序列以使用一處理器之多個組件來執行多個操作，該等組件之各組件具有指示一時間值之一時間計數器，其中各時間計數器包括用以補償當執行該等操作時該等時間計數器之間之相位變動之一個別偏移位元，其中該程式碼序列包括作為在執行該等操作時用於觸發跨該處理器之兩個或兩個以上組件之一追蹤事件之一觸發條件之一第一時間參數，其中該等時間計數器之一者經選擇為一全域時間計數器；

判定由該全域時間計數器所代表之一全域時間值；

基於該全域時間值使用該個別偏移位元至少更新由未經選擇為該全域時間計數器之一對應時間計數器所代表之時間值；

藉由將該時間值與該程式碼序列中之該第一時間參數進行比較，以判定是否滿足該觸發條件；

回應於判定滿足該觸發條件，觸發該追蹤事件以產生用於該處理器之該等組件之該兩個或兩個以上者之事件資料，其中該事件資料基於該全域時間計數器被同步化；及

將該事件資料提供至被用於分析該程式碼序列之該效能之一主機。

【請求項16】

如請求項15之儲存媒體，其中基於該全域時間值使用該個別偏移位元至少更新由未經選擇為該全域時間計數器之該對應時間計數器所代表之該時間值包括：

廣播由該全域時間計數器所代表之該全域時間值至未經選擇為該全域時間計數器之該對應時間計數器；

計算由該對應時間計數器所代表之該時間值與該全域時間直之間之

一差異；及

使用該個別偏移位元調整該對應時間計數器以減少該差異。

【請求項17】

如請求項15之儲存媒體，其中各時間計數器包括一固定大小二進位資料，該固定大小二進位資料包括一多位元計數器及該個別偏移位元，其中該個別偏移位元包括四個位元。

【請求項18】

如請求項15之儲存媒體，其中該程式碼序列進一步包括一第二時間參數；

其中該第二時間參數指示該第一時間參數之一較晚時間；及

其中該觸發條件包括一預定義時間窗，該預定義時間窗包括基於該第一時間參數用於觸發該追蹤事件之一開始時間，及基於該第二時間參數用於停止該追蹤事件之一結束時間。

【請求項19】

如請求項15之儲存媒體，其中該等組件之各組件包括一或多個效能計數器，該一或多個效能計數器經組態以收集用於分析該程式碼序列之該效能之計數資料，該一或多個效能計數器包括一活動計數器、一停滯計數器、一統計計數器或一取樣計數器之至少一者。

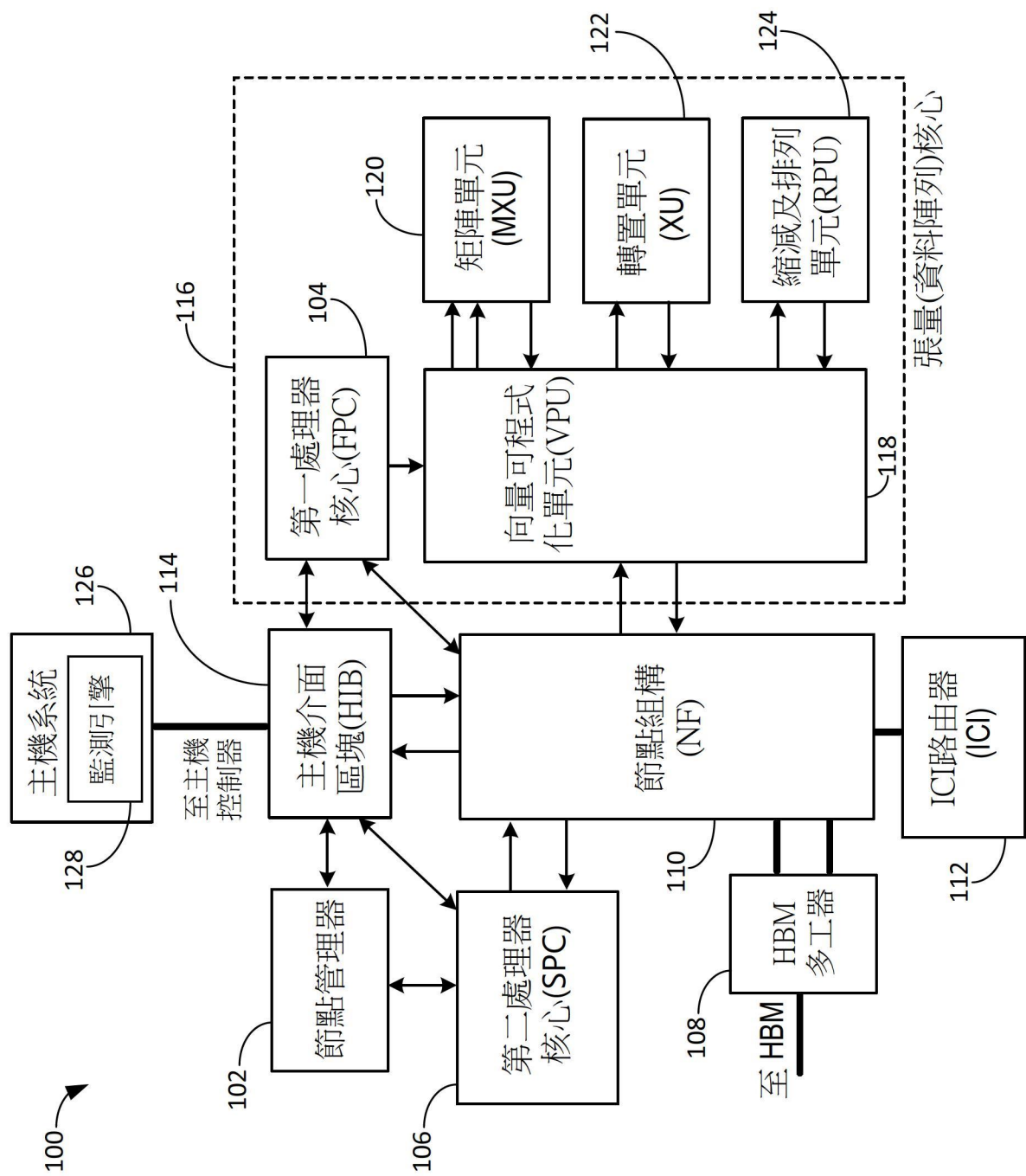
【請求項20】

如請求項19之儲存媒體，其中該一或多個效能計數器經組態以收集該計數資料包括：

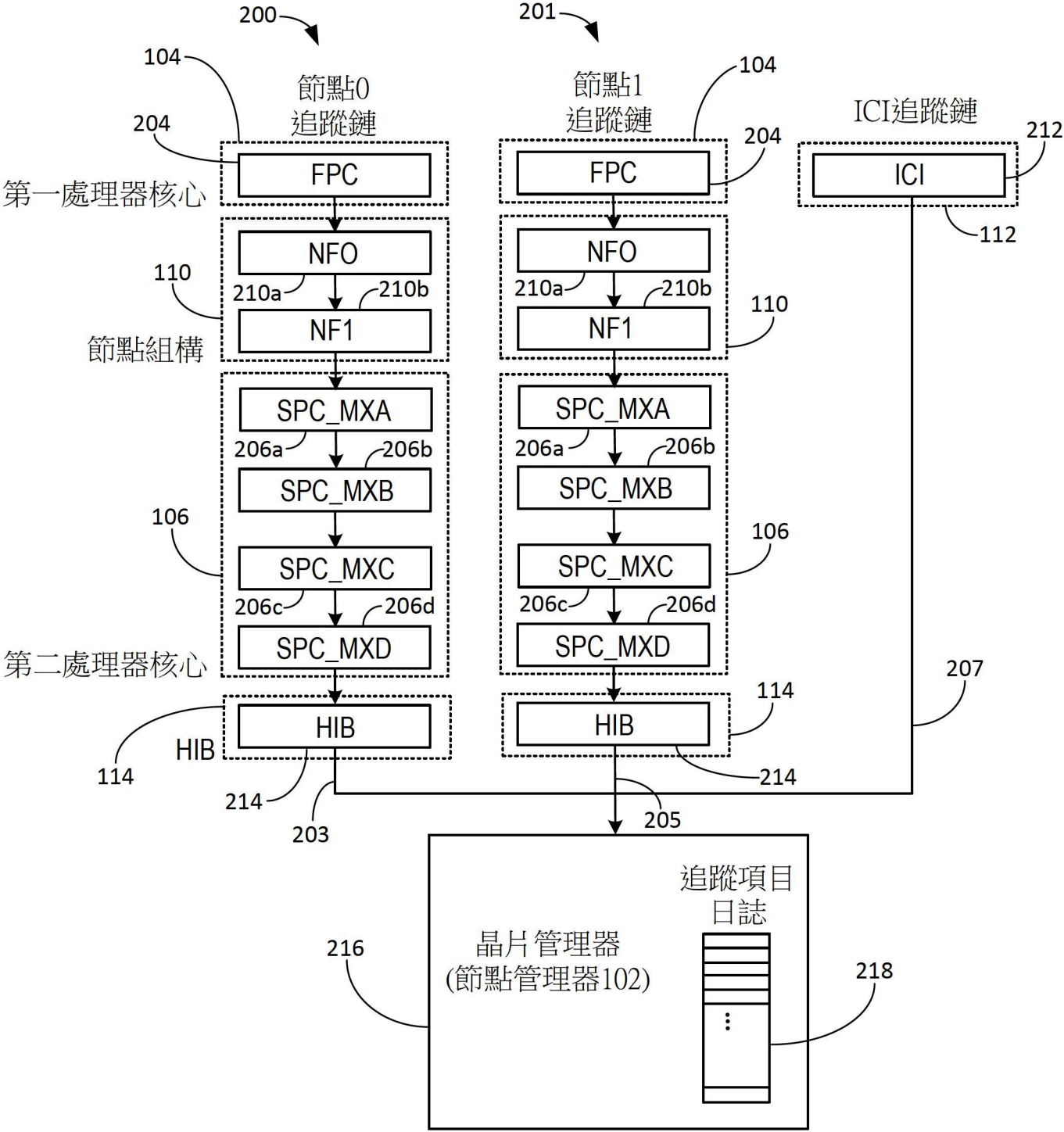
收集包括指令追蹤事件資料的擷取、記憶體多工器追蹤事件資料的擷取或直接記憶體存取追蹤事件資料之至少一者之該計數資料；

將該計數資料儲存於一或多個計數暫存器或控制/狀態暫存器。

【發明圖式】

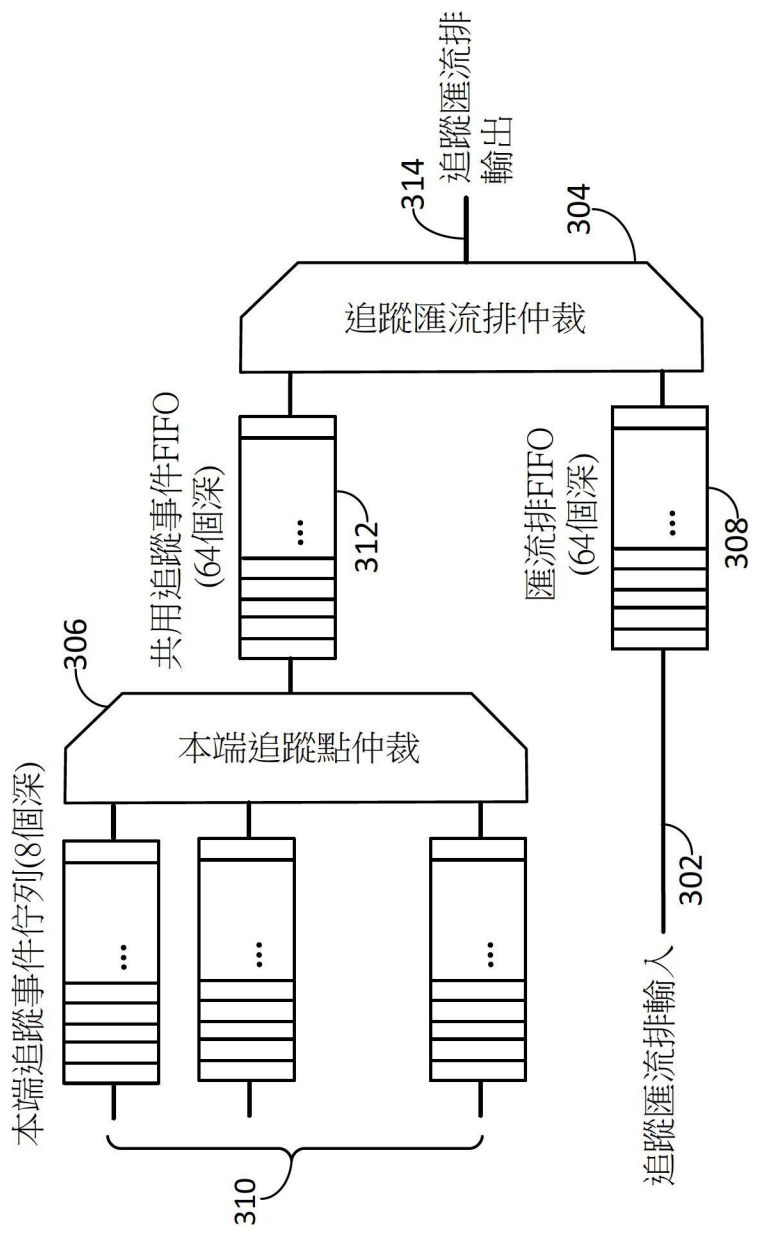


【圖1】



【圖2】

300

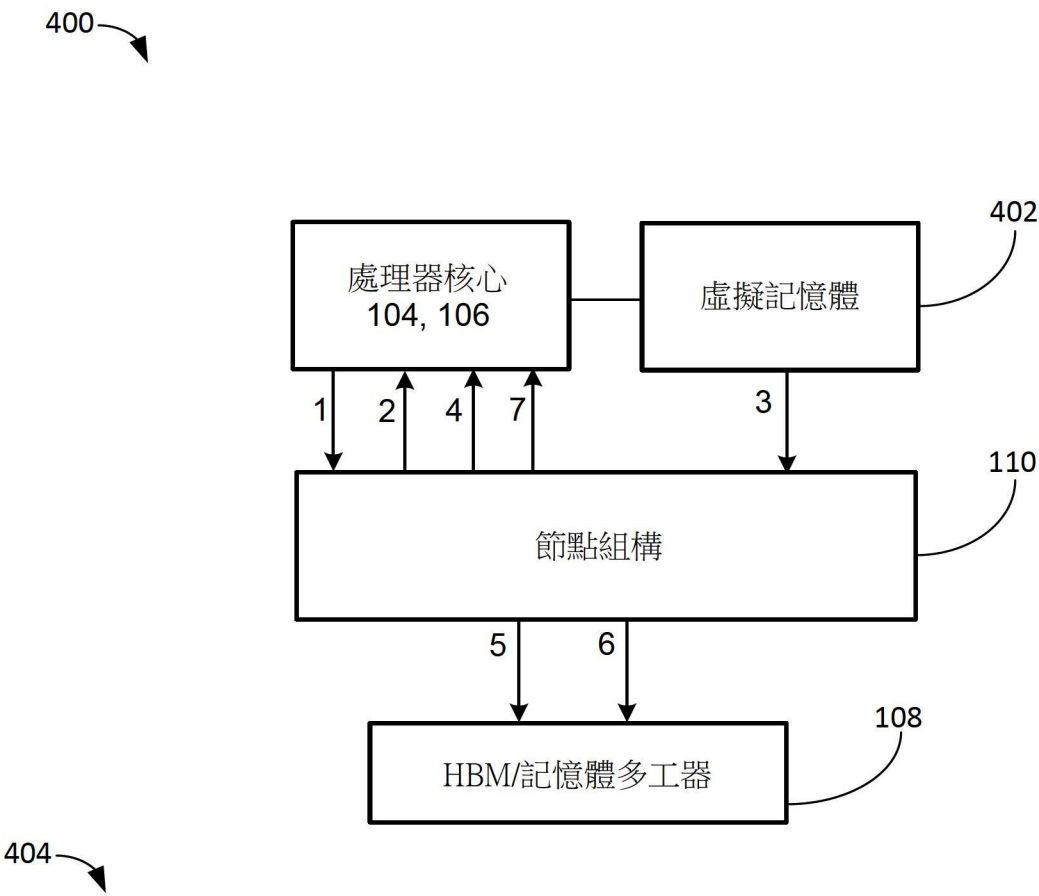


320

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

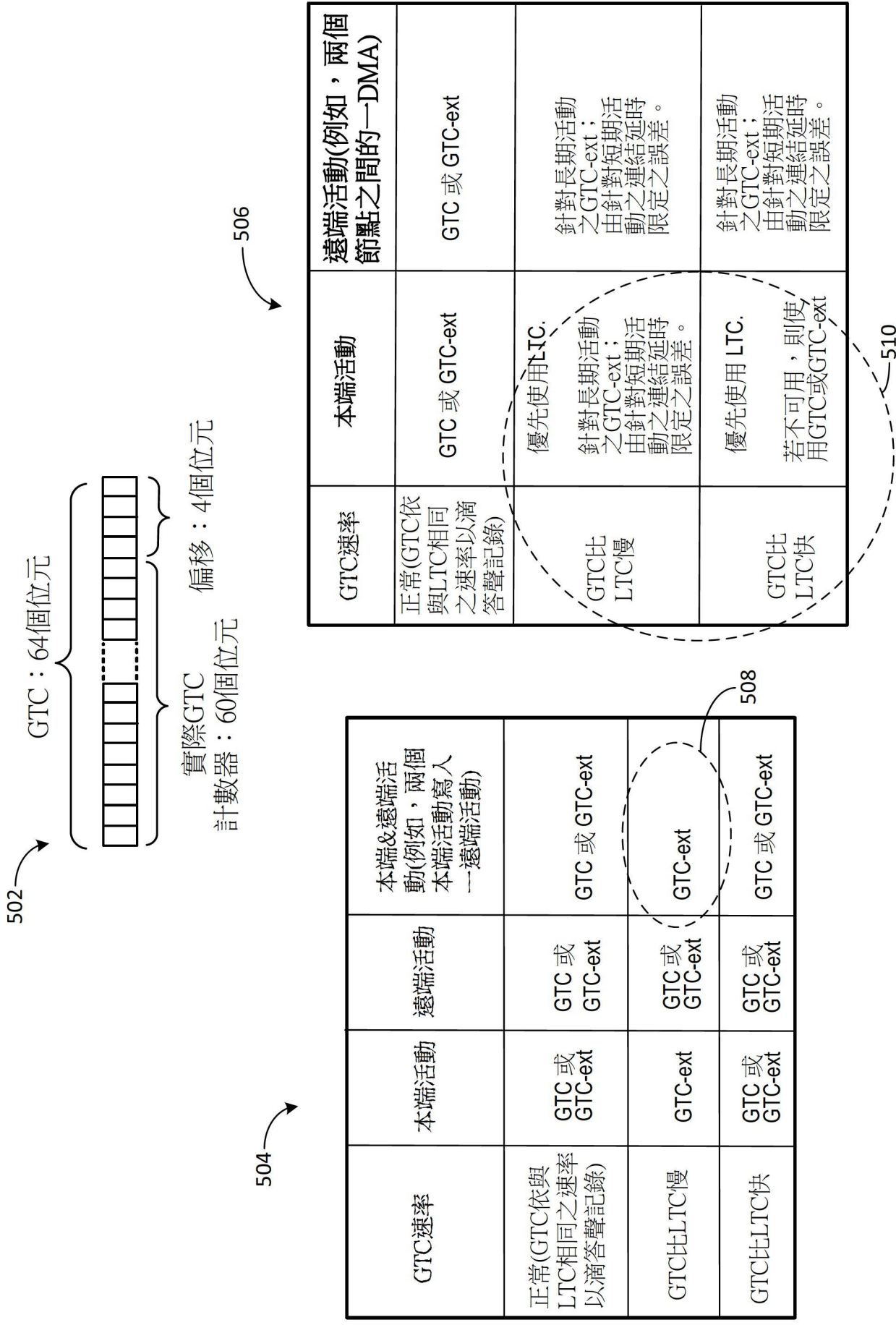
322
324

【圖3】

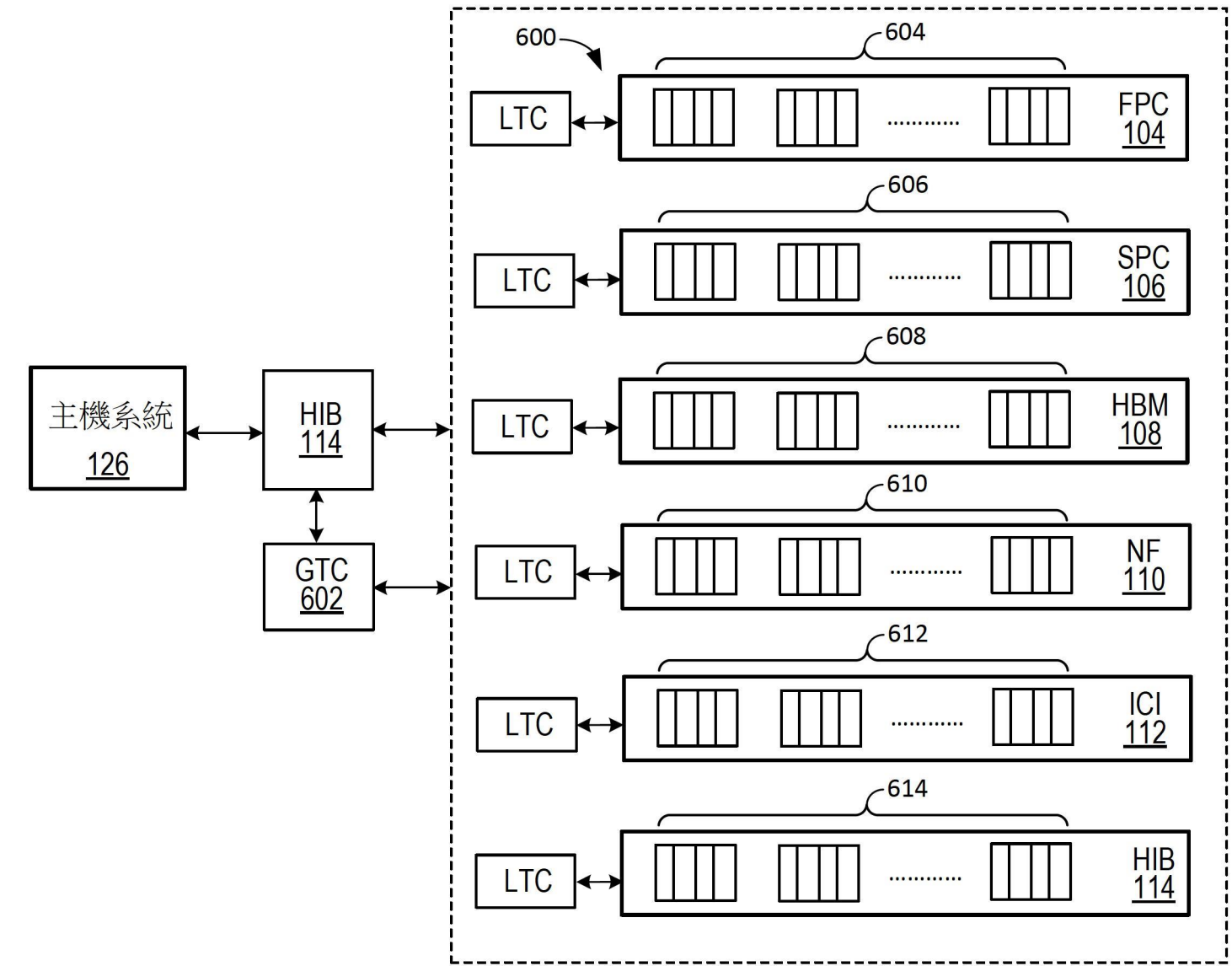


步驟	操作
1	初始DMA請求；節點組構中之追蹤點
2	讀取CMD：NF要求核心傳送資料；NF中之追蹤點
3	讀取完成：此時在NF中無追蹤點！
4	讀取資源更新：核心中之同步旗標更新；FPC中之追蹤點
5	寫入CMD：NF通知HBM；NF中之追蹤點
6	寫入完成：NF中之追蹤點
7	寫入資源更新：FPC中之同步旗標更新；FPC中之追蹤點

【圖4】



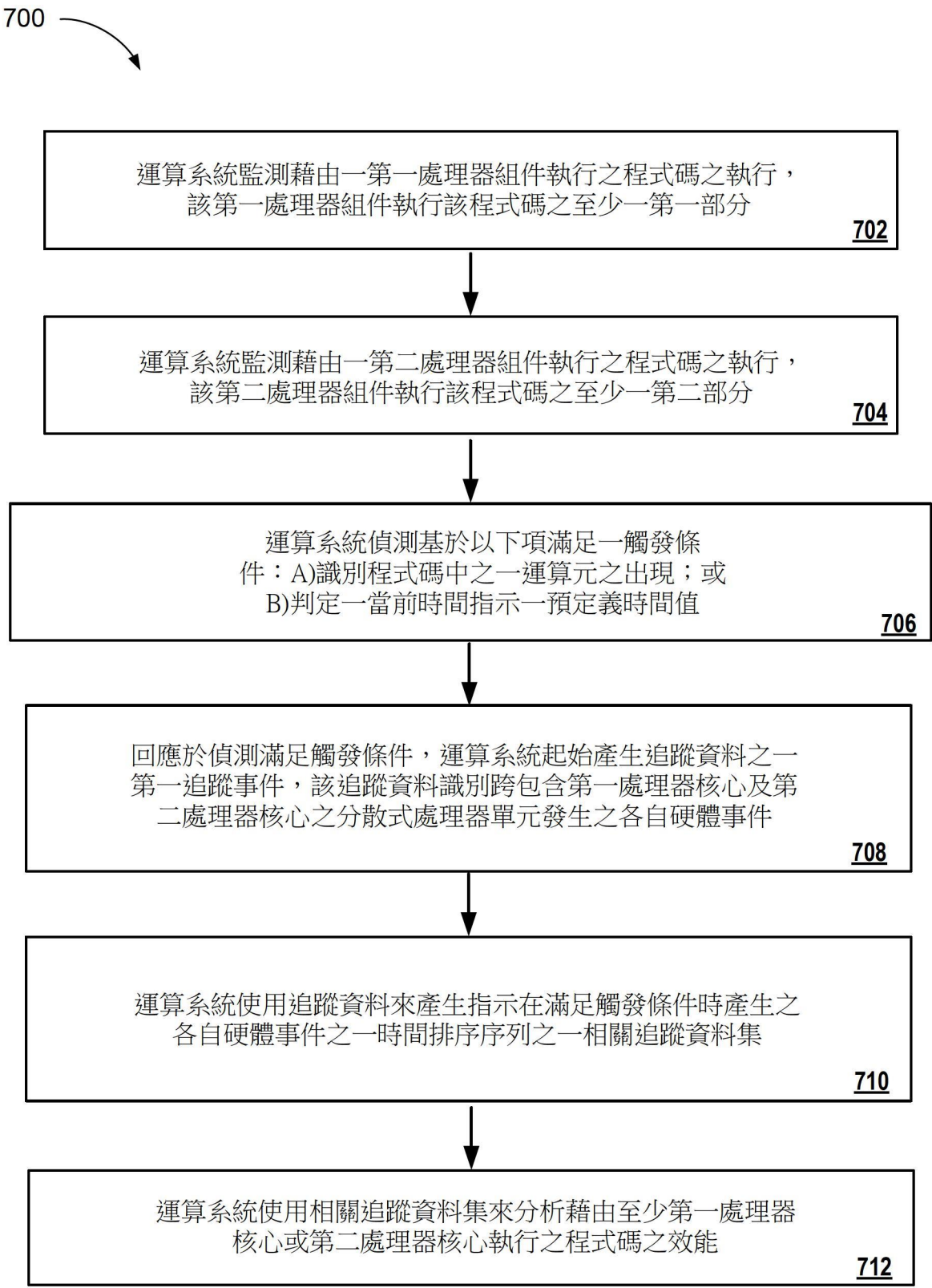
【圖5】



620

時間	追蹤ID	源	目的地	(若干)組件	追蹤事件	序列或TM	延時
09:01.13	000001	0x00F100	0x00F200	FPC → FPC	DMA	程式碼序列56	9μs
09:01.13	000002	0x00F300	0x00F400	FPC → SPC	DMA	程式碼序列75	.023ms
09:15.36	000003	0x00F500	0x00F600	FPC	DMA/指令	程式碼序列33	.041ms
09:26.19	000004	0x00F700	0x00F800	SPC	DMA/指令	程式碼序列33	.062ms

【圖6】



【圖7】