



(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 I815407 B

(45)公告日：中華民國 112 (2023) 年 09 月 11 日

(21)申請案號：111115169

(22)申請日：中華民國 111 (2022) 年 04 月 21 日

(51)Int. Cl. : G06F9/30 (2018.01)

G06N3/02 (2006.01)

(30)優先權：2021/06/17 美國

17/350,365

(71)申請人：美商萬國商業機器公司 (美國) INTERNATIONAL BUSINESS MACHINES CORPORATION (US)

美國

(72)發明人：衛蕭普特 賽門 WEISHAUP, SIMON (DE)；薩波里多 安東尼 SAPORITO, ANTHONY (US)；史洛歌 提摩西 SLEGEL, TIMOTHY (US)

(74)代理人：陳長文

(56)參考文獻：

CN 111126583A

CN 111353595A

US 2017/0103304A1

審查人員：曾祥峰

申請專利範圍項數：17 項 圖式數：11 共 113 頁

(54)名稱

用於促進一計算環境內之處理的電腦程式產品、電腦系統及電腦實施方法

(57)摘要

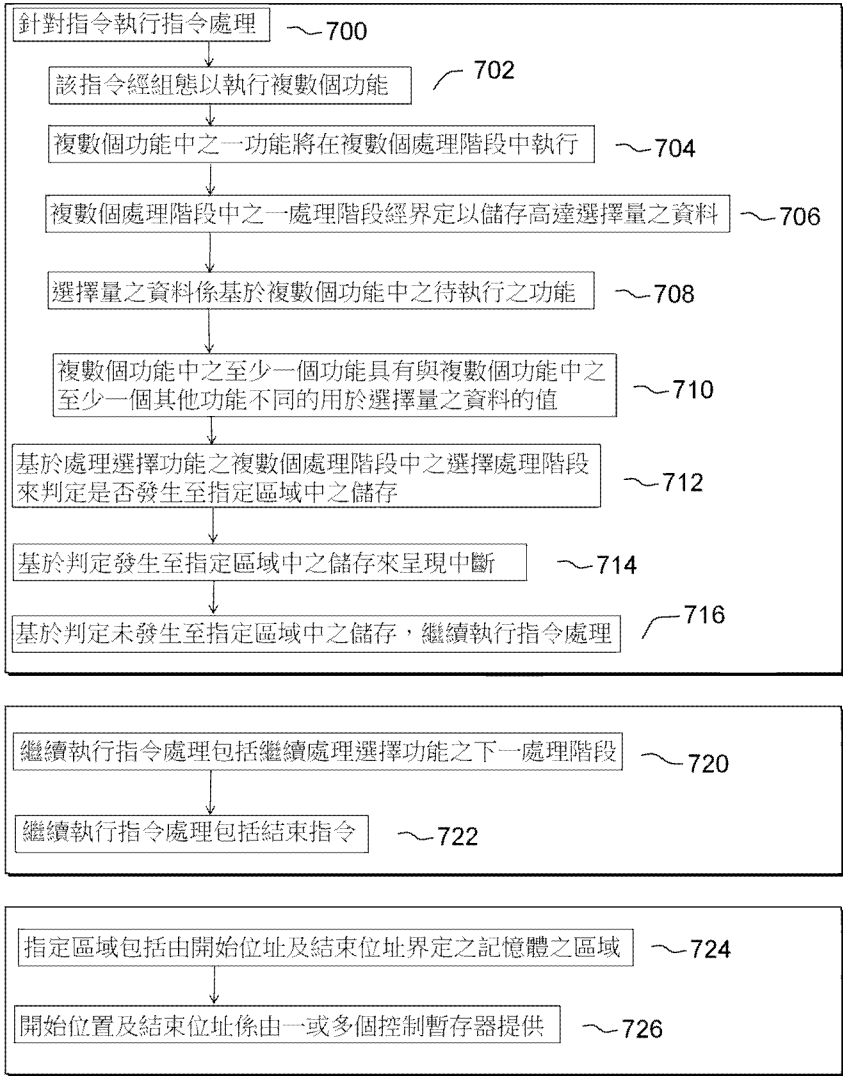
針對一指令執行指令處理。該指令經組態以執行複數個功能，其中該複數個功能中之一功能將在複數個處理階段中執行。一處理階段經界定以儲存高達一選擇量之資料。該選擇量之資料係基於待執行之該功能。該複數個功能中之至少一個功能具有與至少一個其他功能不同的一用於該選擇量之資料的值。基於處理一選擇功能之一選擇處理階段來判定是否發生至一指定區域中之一儲存。基於判定發生了至該指定區域中之該儲存，呈現一中斷，且基於判定並未發生至該指定區域中之該儲存，繼續指令處理。

Instruction processing is performed for an instruction. The instruction is configured to perform a plurality of functions, in which a function of the plurality of functions is to be performed in a plurality of processing phases. A processing phase is defined to store up to a select amount of data. The select amount of data is based on the function to be performed. At least one function of the plurality of functions has a different value for the select amount of data than at least one other function. A determination is made as to whether a store into a designated area occurred based on processing a select processing phase of a select function. Based on determining that the store into the designated area occurred, an interrupt is presented, and based on determining that the store into the designated area did not occur, instruction processing is continued.

指定代表圖：

符號簡單說明：

- 700:步驟
- 702:步驟
- 704:步驟
- 706:步驟
- 708:步驟
- 710:步驟
- 712:步驟
- 714:步驟
- 716:步驟
- 720:步驟
- 722:步驟
- 724:步驟
- 726:步驟



【圖7A】

**公告本**

I815407

【發明摘要】**【中文發明名稱】**

用於促進一計算環境內之處理的電腦程式產品、電腦系統及電腦實施方法

【英文發明名稱】

COMPUTER PROGRAM PRODUCT, COMPUTER SYSTEM, AND
COMPUTER-IMPLEMENTED METHOD FOR FACILITATING
PROCESSING WITHIN A COMPUTING ENVIRONMENT

【中文】

針對一指令執行指令處理。該指令經組態以執行複數個功能，其中該複數個功能中之一功能將在複數個處理階段中執行。一處理階段經界定以儲存高達一選擇量之資料。該選擇量之資料係基於待執行之該功能。該複數個功能中之至少一個功能具有與至少一個其他功能不同的一用於該選擇量之資料的值。基於處理一選擇功能之一選擇處理階段來判定是否發生至一指定區域中之一儲存。基於判定發生了至該指定區域中之該儲存，呈現一中斷，且基於判定並未發生至該指定區域中之該儲存，繼續指令處理。

【英文】

Instruction processing is performed for an instruction. The instruction is configured to perform a plurality of functions, in which a function of the plurality of functions is to be performed in a plurality of processing phases. A processing phase is defined to store up to a select amount of data. The select amount of data is based on the function to be performed. At least one function of the plurality of functions has a different value for the select amount of data than at least one other function. A

determination is made as to whether a store into a designated area occurred based on processing a select processing phase of a select function. Based on determining that the store into the designated area occurred, an interrupt is presented, and based on determining that the store into the designated area did not occur, instruction processing is continued.

【指定代表圖】

圖7A

【代表圖之符號簡單說明】

700: 步驟

702: 步驟

704: 步驟

706: 步驟

708: 步驟

710: 步驟

712: 步驟

714: 步驟

716: 步驟

720: 步驟

722: 步驟

724: 步驟

726: 步驟

【發明說明書】

【中文發明名稱】

用於促進一計算環境內之處理的電腦程式產品、電腦系統及電腦實施方法

【英文發明名稱】

COMPUTER PROGRAM PRODUCT, COMPUTER SYSTEM, AND
COMPUTER-IMPLEMENTED METHOD FOR FACILITATING
PROCESSING WITHIN A COMPUTING ENVIRONMENT

【技術領域】

【0001】 一或多個態樣大體上係關於促進計算環境內之處理，且特定言之，係關於改良此類處理。

【先前技術】

【0002】 為了增強資料及/或計算密集型計算環境中之處理，利用共處理器，諸如人工智慧加速器(亦被稱作神經網路處理器或神經網路加速器)。此類加速器提供大量計算能力，用於執行例如所涉及計算，諸如對矩陣或張量之計算。

【0003】 作為一實例，張量計算用於複雜處理，包括深度學習，其為機器學習之子集。深度學習或機器學習(人工智慧之態樣)用於各種技術中，包括但不限於工程化、製造、醫療技術、汽車技術、電腦處理等。

【0004】 在計算處理期間，若發生某些事件，則獲得與彼等事件及/或處理相關之資訊係有益的。因此，將在獲得此類資訊時取得進步。

【發明內容】

【0005】 經由提供用於促進計算環境內之處理的電腦程式產品來克服先前技術之缺點且提供額外優點。該電腦程式產品包括一或多個電腦可

讀儲存媒體及共同地儲存於該一或多個電腦可讀儲存媒體上以執行一方法之程式指令。該方法包括執行用於指令之指令處理。該指令經組態以執行複數個功能，且複數個功能中之一功能將在複數個處理階段中執行。複數個處理階段中之處理階段經定義以儲存高達選擇量之資料。選擇量之資料係基於複數個功能中之待執行之功能。複數個功能中之至少一個功能具有與複數個功能中之至少一個其他功能不同的用於選擇量之資料的值。基於處理選擇功能之複數個處理階段中之一選擇處理階段來判定是否發生至指定區域中之儲存。基於判定發生了至指定區域中之儲存，呈現中斷，且基於判定未發生至指定區域中之儲存，執行指令處理繼續。

【0006】 經組態以執行各種功能之單個架構化機器指令的使用藉由降低複雜度、減少資源使用及/或提高處理速度來改良計算環境內之效能。

【0007】 藉由檢查在處理階段之後是否發生了至指定區域中之儲存，僅在結束指令之前的單個階段內發生至指定區域中之儲存。由於在每一階段期間儲存選擇量之資料且選擇量可大於4K，因此測試是否存在至指定區域中之儲存(例如，是否處理多個連續4K頁面)可能需要縮減數目次檢查。

【0008】 在一個實例中，繼續執行指令處理包括進行至選擇功能之下一處理階段之處理。作為另一實例，繼續執行指令處理包括結束該指令。

【0009】 作為一實例，指定區域包括由開始位址及結束位址界定之記憶體之區域。開始位址及結束位址係由例如一或多個控制暫存器提供。

【0010】 在一個實施例中，判定是否發生至指定區域中之儲存包括

檢查資料經寫入至之記憶體位址是否處於指定區域中。基於該記憶體位址處於指定區域中，該判定指示存在至指定區域中之儲存。

【0011】 在一個實施例中，判定是否發生至指定區域中之儲存包括檢查待由選擇功能使用之參數區塊之至少一部分是否與指定區域重疊。基於參數區塊之至少一部分與指定區域重疊，該判定指示存在至指定區域中之儲存。

【0012】 用於參數區塊之重疊的檢查可在每一階段之後產生中斷，從而容易地提供關於記憶體存取之資訊。

【0013】 在一個實施例中，判定是否發生至指定區域中之儲存包括檢查待由選擇功能使用之功能特定保存區域之至少一部分是否與指定區域重疊。基於功能特定保存區域之至少一部分與指定區域重疊，該判定指示存在至指定區域中之儲存。

【0014】 若要使用針對功能特定保存區域之重疊的檢查，則可在每一階段之後產生中斷，從而容易地提供關於記憶體存取的資訊。

【0015】 在一或多個態樣中，可進行某些預測。舉例而言，經預測，選擇功能之複數個處理階段中之特定處理階段將產生中斷，其中該預測係基於選擇功能之記憶體存取模式。作為另一實例，基於判定待由選擇功能使用之參數區塊之至少一部分與指定區域重疊，預測將發生中斷。在另一實例中，基於判定待由選擇功能使用之功能特定保存區域之至少一部分與指定區域重疊，預測將發生中斷。

【0016】 至少，驗證程式的複雜性降低了，此可預測儲存改變中斷，從而提高驗證品質。舉例而言，基於對應於功能碼之記憶體存取模式，驗證程式可準確地預測指令將在某一階段之後中斷且呈現程式事件記

錄中斷；且驗證程式化可在程式事件記錄指定區域位於參數區塊及/或功能特定保存區域內之狀況下預測中斷，此係因為中斷將在每一階段之後呈現。

【0017】本文中亦描述及主張與一或多個態樣相關之電腦實施方法及系統。此外，本文中亦描述及可能主張與一或多個態樣相關之服務。

【0018】藉由本文中所描述之技術實現額外特徵及優勢。本文中詳細描述其他實施例及態樣且將其視為所主張態樣之部分。

【圖式簡單說明】

【0019】在本說明書之結尾處之申請專利範圍中作為實例特定地指出且清楚地主張一或多個態樣。一或多個態樣之前述內容以及物件、特徵及優點自結合隨附圖式進行之以下詳細描述顯而易見，其中：

圖1A描繪併有及使用本發明之一或多個態樣的計算環境之一個實例；

圖1B描繪根據本發明之一或多個態樣的圖1A之處理器的其他細節；

圖2A描繪根據本發明之一或多個態樣之關於與神經網路處理輔助指令相關聯的程式事件記錄之處理的一個實例；

圖2B描繪根據本發明之一或多個態樣之在神經網路處理輔助指令之功能之處理階段期間發生的儲存改變事件之一個實例；

圖3A描繪根據本發明之一或多個態樣的神經網路處理輔助指令之格式的一個實例；

圖3B描繪根據本發明之一或多個態樣的由神經網路處理輔助指令使用之通用暫存器的一個實例；

圖3C描繪根據本發明之一或多個態樣的由神經網路處理輔助指令支

援之功能碼的實例；

圖3D描繪根據本發明之一或多個態樣的由神經網路處理輔助指令使用之另一通用暫存器的一個實例；

圖3E描繪根據本發明之一或多個態樣的由神經網路處理輔助指令之查詢功能使用的參數區塊之一個實例；

圖3F描繪根據本發明之一或多個態樣的由神經網路處理輔助指令之一或多個非查詢功能使用的參數區塊之一個實例；

圖3G描繪根據本發明之一或多個態樣的由神經網路處理輔助指令使用之張量描述符的一個實例；

圖4描繪根據本發明之一或多個態樣的神經網路處理(NNP)資料類型-1資料類型之格式的一個實例；

圖5A至圖5C描繪根據本發明之一或多個態樣的由神經網路處理輔助指令使用之輸入資料佈局的實例；

圖6A至圖6C描繪根據本發明之一或多個態樣的對應於圖5A至圖5C之輸入資料佈局的實例輸出；

圖7A至圖7B描繪根據本發明之一或多個態樣之促進計算環境內之處理的一個實例；

圖8A描繪併有及使用本發明之一或多個態樣的計算環境之另一實例；

圖8B描繪根據本發明之一或多個態樣的圖8A之記憶體之其他細節的一個實例；

圖8C描繪根據本發明之一或多個態樣的圖8A之記憶體之其他細節的另一實例；

圖9A描繪併有及使用本發明之一或多個態樣的計算環境之又一實例；

圖9B描繪根據本發明之一或多個態樣之圖9A的記憶體之其他細節；

圖10描繪根據本發明之一或多個態樣的雲端計算環境之一個實施例；且

圖11描繪根據本發明之一或多個態樣的抽象模型層之一個實例。

【實施方式】

【0020】 根據本發明之一或多個態樣，提供一種促進計算環境內之處理的能力。作為一實例，提供促進獲得關於可在指令處理之階段期間發生的事件之資訊的能力。此資訊適用於判定在計算環境內執行之程式是否正如期望在執行及/或是否可執行最佳化；適用於縮減錯誤及/或複雜性；及/或適用於改良系統效能。此外，在一個實例中，該資訊藉由使得彼等程式能夠預測將發生事件(例如，產生中斷之儲存改變)來縮減驗證程式之複雜性，從而增加驗證品質、縮減錯誤及/或改良系統效能。該能力包括程式事件記錄設施，其用於例如已經針對某些類型的處理增強之除錯程式，該處理包括例如與由專用處理器執行之指令相關聯的處理，該等專用處理器諸如神經網路處理器(亦稱為神經網路加速器)。

【0021】 程式事件記錄已經增強之指令之一個實例為神經網路處理輔助指令，其為經組態以執行多個功能之單個指令(例如，硬體/軟體介面處之單個架構化硬體機器指令)。功能中之每一者經組態為單個指令(例如，單個架構化指令)之部分，從而縮減系統資源之使用並降低複雜度且改良系統效能。

【0022】 該指令可為由諸如通用處理器之處理器上之程式分派的通

用處理器指令集架構(ISA)之部分。其可由通用處理器執行，及/或該指令之一或多個功能可由諸如經組態以用於某些功能之共處理器之專用處理器執行，該專用處理器耦接至通用處理器或為該通用處理器之部分。其他變化亦係可能的。

【0023】 參看圖1A描述併有及使用本發明之一或多個態樣的計算環境之一個實施例。作為一實例，該計算環境係基於由紐約阿蒙克市之國際商業機器公司提供之z/Architecture®指令集架構。z/Architecture指令集架構之一個實施例描述於標題為「z/Architecture Principles of Operation」之公開案(IBM公開案第SA22-7832-12號，第十三版，2019年9月)中，該公開案特此以全文引用之方式併入本文中。然而，z/Architecture指令集架構僅為一個實例架構；國際商業機器公司及/或其他實體之其他架構及/或其他類型的計算環境可包括及/或使用本發明之一或多個態樣。z/Architecture及IBM為國際商業機器公司在至少一個管轄區域中之商標或註冊商標。

【0024】 參考圖1A，計算環境100包括例如以例如通用計算裝置之形式展示的電腦系統102。電腦系統102可包括但不限於經由一或多個匯流排及/或其他連接彼此耦接的一或多個通用處理器或處理單元104 (例如，中央處理單元(CPU))、諸如神經網路處理器105之至少一個專用處理器、記憶體106 (作為實例，亦稱為系統記憶體、主記憶體、主儲存器、中心儲存器或儲存器)及一或多個輸入/輸出(I/O)介面108。舉例而言，處理器104、105及記憶體106經由一或多個匯流排110耦接至I/O介面108，且處理器104、105經由一或多個匯流排111耦接至彼此。

【0025】 舉例而言，匯流排111為記憶體或快取一致性匯流排，且

匯流排110表示例如若干類型之匯流排結構中之任何一或多者，包括使用多種匯流排架構中之任一者的記憶體匯流排或記憶體控制器、周邊匯流排、加速圖形埠及處理器或區域匯流排。作為實例而非限制，此類架構包括工業標準架構(ISA)、微通道架構(MCA)、增強型ISA (EISA)、視訊電子標準協會(VESA)區域匯流排及周邊組件互連(PCI)。

【0026】 作為實例，一或多個專用處理器(例如，神經網路處理器)可與一或多個通用處理器分離但耦接至該一或多個通用處理器，及/或可嵌入於一或多個通用處理器內。許多變化係可能的。

【0027】 舉例而言，記憶體106可包括快取記憶體112，諸如共用快取記憶體，該快取記憶體可經由例如一或多個匯流排111耦接至處理器104之本端快取記憶體114及/或神經網路處理器105。另外，記憶體106可包括一或多個程式或應用程式116及至少一個作業系統118。實例作業系統包括由紐約阿蒙克市之國際商業機器公司提供之z/OS[®]作業系統。z/OS為國際商業機器公司在至少一個管轄區域中之商標或註冊商標。亦可使用由國際商業機器公司及/或其他實體供應的其他作業系統。記憶體106亦可包括一或多個電腦可讀程式指令120，該等指令可經組態以實行本發明之態樣之實施例的功能。

【0028】 此外，在一或多個實施例中，記憶體106包括處理器韌體122。處理器韌體包括例如處理器之微碼或毫碼。其包括例如用於實施較高階機器碼之硬體層級指令及/或資料結構。在一個實施例中，其包括例如專屬碼，該專屬碼通常作為包括受信任軟體之微碼或毫碼、特定於底層硬體之微碼或毫碼遞送，且控制對系統硬體之作業系統存取。

【0029】 電腦系統102可經由例如I/O介面108與一或多個外部裝置

130通信，該一或多個外部裝置諸如使用者終端機、磁帶機、指標裝置、顯示器及一或多個資料儲存裝置134等。資料儲存裝置134可儲存一或多個程式136、一或多個電腦可讀程式指令138及/或資料等。電腦可讀程式指令可經組態以實行本發明之態樣的實施例之功能。

【0030】 電腦系統102亦可經由例如I/O介面108與網路介面132通信，該網路介面使得電腦系統102能夠與諸如區域網路(LAN)、通用廣域網路(WAN)及/或公用網路(例如，網際網路)之一或多個網路通信，從而提供與其他計算裝置或系統之通信。

【0031】 電腦系統102可包括及/或耦接至抽取式/非抽取式、揮發性/非揮發性電腦系統儲存媒體。舉例而言，其可包括及/或耦接至非抽取式非揮發性磁性媒體(通常被稱作「硬碟機」)、用於自抽取式非揮發性磁碟(例如，「軟碟」)讀取及寫入至抽取式非揮發性磁碟(例如，「軟碟」)之磁碟機，及/或用於自諸如CD-ROM、DVD-ROM或其他光學媒體之抽取式非揮發性光碟讀取或寫入至抽取式非揮發性光碟之光碟機。應理解，可結合電腦系統102使用其他硬體及/或軟體組件。實例包括但不限於：微碼或毫碼、裝置驅動程式、冗餘處理單元、外部磁碟機陣列、RAID系統、磁帶機及資料存檔儲存系統等。

【0032】 電腦系統102可與眾多其他通用或專用計算系統環境或組態一起操作。可適合與電腦系統102一起使用之熟知計算系統、環境及/或組態之實例包括但不限於：個人電腦(PC)系統、伺服器電腦系統、精簡型用戶端、複雜型用戶端、手持型或膝上型電腦裝置、多處理器系統、基於微處理器之系統、機上盒、可程式化消費型電子裝置、網路PC、小型電腦系統、大型電腦系統及包括以上系統或裝置中之任一者的分散式雲端計

算環境，以及其類似者。

【0033】 在一個實例中，處理器(例如，處理器104及/或處理器105)包括用以執行指令之複數個功能組件(或其子集)。如圖1B中所描繪，此等功能組件包括例如：指令提取組件150，其用以提取將執行之指令；指令解碼單元152，其用以解碼所提取指令且獲得經解碼指令之運算元；一或多個指令執行組件154，其用以執行經解碼指令；記憶體存取組件156，其用以在必要時存取記憶體以執行指令；及寫回組件158，其用以提供所執行指令之結果。該等組件中之一或多者可在指令處理中存取及/或使用一或多個暫存器160。此外，組件中之一或多者可根據本發明之一或多個態樣包括用於執行例如神經網路處理輔助指令之程式事件記錄處理及/或神經網路處理輔助處理(或可使用本發明之一或多個態樣之其他處理)的一或多個其他組件之至少一部分或能夠存取該一或多個其他組件，如本文中所描述。一或多個其他組件可包括例如程式事件記錄組件170及/或神經網路處理輔助組件172 (及/或一或多個其他組件)。

【0034】 根據本發明之一或多個態樣，程式事件記錄經增強，使得由於經偵測事件引起之中斷可在指令完成之前且在選擇量之資料經寫入至選擇位置(例如，記憶體)之後發生，如由經執行指令之特定功能所界定。作為一實例，偵測一或多個選擇事件(例如，指定區域之儲存改變)可在經執行功能之複數個處理階段中之處理階段的處理之後發生。在處理階段中，高達選擇量之資料可儲存至記憶體，且該選擇量係基於經執行指令之功能。不同功能可具有可儲存於特定階段中之相同或不同選擇量之資料。若儲存選擇量之資料包括至指定區域中之儲存，則會偵測到選擇事件(例如，指定區域之儲存改變)且呈現中斷。

【0035】 參考圖2A至圖2B描述關於相關於選擇指令(諸如，神經網路處理輔助指令)之程式事件記錄處理的其他細節。起初參看圖2A，諸如神經網路處理輔助指令之指令係由例如通用處理器(諸如通用處理器104)起始(200)。通用處理器執行用於指令之指令處理，包括例如判定經組態以用於待執行之指令的複數個功能中之一功能(例如，非查詢功能)。

【0036】 根據本發明之一態樣，該功能將在處理階段中執行，其中在每一處理階段期間，高達選擇量之資料可儲存至選擇位置(例如，記憶體)。待在一階段內處理之選擇量之資料係取決於該功能。下文提供可基於神經網路處理輔助指令之功能碼在處理階段期間儲存至例如目標記憶體之實例功能及實例最大量的資料(可在其他實施例中提供額外、較少及/或其他功能、功能碼及/或最大量等)：

【0037】 NNPA-ADD (功能碼16)：每階段65536個目標位元組(16個連續4K頁面)

【0038】 NNPA-SUB (功能碼17)：每階段65536個目標位元組(16個連續4K頁面)

【0039】 NNPA-MUL (功能碼18)：每階段65536個目標位元組(16個連續4K頁面)

【0040】 NNPA-DIV (功能碼19)：每階段65536個目標位元組(16個連續4K頁面)

【0041】 NNPA-MIN (功能碼20)：每階段65536個目標位元組(16個連續4K頁面)

【0042】 NNPA-MAX (功能碼21)：每階段65536個目標位元組(16個連續4K頁面)

【0043】 NNPA-LOG (功能碼32)：每階段65536個目標位元組(16個連續4K頁面)

【0044】 NNPA-EXP (功能碼33)：每階段65536個目標位元組(16個連續4K頁面)

【0045】 NNPA-RELU (功能碼49)：每階段65536個目標位元組(16個連續4K頁面)

【0046】 NNPA-TANH (功能碼50)：每階段65536個目標位元組(16個連續4K頁面)

【0047】 NNPA-SIGMOID (功能碼51)：每階段65536個目標位元組(16個連續4K頁面)

【0048】 NNPA-SOFTMAX (功能碼52)：每階段65536個目標位元組(16個非連續4K頁面)；此外2個連續4K頁面被稱作「功能特定保存區域」

【0049】 NNPA-BATCHNORM (功能碼64)：每階段65536個目標位元組(16個非連續4K頁面)

【0050】 NNPA-MAXPOOL2D (功能碼80)：每階段131072個目標位元組(32個非連續4K頁面)

【0051】 NNPA-AVGPOOL2D (功能碼81)：每階段131072個目標位元組(32個非連續4K頁面)

【0052】 NNPA-LSTM (功能碼96)：每階段16384個目標位元組(4個非連續4K頁面)

【0053】 NNPA-GRU (功能碼97)：每階段8192個目標位元組(2個非連續4K頁面)

【0054】 NNPA-CONVOLUTION (功能碼112)：每階段131072個目標位元組(32個非連續4K頁面)

【0055】 NNPA-MATMUL-OP (功能碼113)：每階段131072個目標位元組(32個非連續4K頁面)

【0056】 NNPA-MATMUL-OP-BCAST23 (功能碼114)：每階段131072個目標位元組(32個非連續4K頁面)

【0057】 如所指示，在處理階段中執行指令(例如，神經網路處理輔助指令)之功能。因此，基於指令之開始，功能之處理階段之處理開始202。在一個實施例中，經選擇以執行之功能(在本文中被稱作選擇功能)係由專用處理器(例如，神經網路處理器105)執行，且因此，作為初始處理階段之部分，通用處理器將待執行之選擇功能之指示提供至專用處理器。此外，在一個實例中，處理階段之處理包括藉由通用處理器(例如，通用處理器104之韌體122)將用於在處理階段期間執行選擇功能之記憶體位址發送至專用處理器。在一個實例中，該韌體亦將處理階段之結束的指示提供至專用處理器。在一個實例中，作為將記憶體位址發送至專用處理器之部分，通用處理器(例如，韌體122)執行一或多個記憶體相關檢查，包括與程式事件記錄相關的記憶體相關檢查。舉例而言，該韌體在提供用於給定處理階段之位址之前判定經呈現用於處理階段之記憶體位址是否屬於程式事件記錄指定區域。若一或多個位址將屬於指定區域，則該韌體向專用處理器指示處理將在處理此階段之後停止。

【0058】 專用處理器(例如，神經網路處理器105)接收位址，以及關於階段結束及是否存在例如程式事件記錄事件之資訊，且對輸入資料(例如，輸入張量資料)執行選擇功能以提供輸出資料(例如，輸出張量資

料)。在處理階段完成時，專用處理器向韌體傳遞在當前階段結束時增加之計數器，且停止處理直至其接收用於下一階段之額外記憶體位址。

【0059】 通用處理器(例如，韌體122)接收最後一個階段之處理完成之指示，且關於是否已經處理所有資料做出判定204。若已經處理所有資料，則指令完成206。然而，若尚未處理所有資料，則(例如，藉由韌體)關於是否是最後一個階段中之寫入至記憶體期間存在至指定區域中之儲存進行檢查208。作為一實例，對於程式事件記錄，指定區域係經由例如控制暫存器(諸如控制暫存器10)界定，該控制暫存器10提供指定區域及控制暫存器11之開始位址，該控制暫存器11提供指定區域之結束位址。在其他實例中，可使用其他控制暫存器及/或其他機構可用於提供開始及結束位址。若資料經寫入至與指定區域重疊之一或多個記憶體位址，則判定存在至指定區域中之儲存，且偵測到程式事件記錄儲存改變事件，假設程式事件記錄儲存改變經啟用。在一個實例中，當儲存改變事件控制(例如，選擇控制暫存器(例如，控制暫存器9)之選擇位元(例如，位元34))經設定(例如，為一)且儲存改變空間控制(例如，選擇控制暫存器(例如，控制暫存器9)之選擇位元(例如，位元42))經設定(例如，為零)或位址空間控制元素之選擇位元(例如，位元56)經設定為例如一時，啟用程式事件記錄儲存改變。

【0060】 此外，在一個實例中，若由功能使用之參數區塊之至少一部分及/或功能特定保存區域之至少一部分(若待由功能使用)與指定區域重疊，則在一個實施例中，判定存在至指定區域之儲存(例如，儲存改變事件)，與是否發生此儲存無關。在另一實例中，判定僅在存在至與指定區域重疊之部分中之儲存的情況下，才發生至指定區域之儲存。在又一實例

中，判定，若整個功能保存區域(若待由功能使用)或若整個參數區塊與指定區域重疊，則發生至指定區域之儲存，與是否更新了參數區塊或功能特定保存區域無關。其他變化為可能的。

【0061】 若判定不存在至指定區域之儲存，則處理繼續至下一處理階段202，其中通用處理器將額外記憶體位址提供至專用處理器，以用於在下一階段中進一步處理功能。然而，若判定在最後一個階段期間發生了至指定區域之儲存，則例如韌體呈現中斷210，且指令結束206。無需等到處理了所有資料，呈現此中斷。此外，根據本發明之一態樣，大於4K的資料量可在處理階段期間且在中斷之前儲存。舉例而言，在處理階段期間產生之所得輸出資料量取決於功能碼，其中不同功能碼可具有不同資料量。另外，資料可儲存至若干4K頁面中，該等4K頁面在處理階段期間在邏輯位址空間內可或可不為連續的；可在功能之執行期間更新參數區塊並且可多次更新參數區塊；且可基於經執行之功能更新功能特定保存區域並且可多次更新功能特定保存區域。存在其他可能性。

【0062】 在一個實施例中，當指令結束時，亦可判定最後一個階段之處理是否引起至指定區域中之儲存。此資訊可用於例如除錯目的。

【0063】 圖2B中展示程式事件記錄儲存改變事件之描繪之一個實例。如所描繪，在一個實例中，神經網路處理輔助指令之選擇功能之階段0在未儲存至指定區域中之情況下經處理；然而，在階段1之處理期間，執行寫入至指定區域中250。因此，在該階段結束時，發生程式事件記錄儲存改變事件252，從而導致指令中斷(例如，指令以例如3之條件碼結束)。除非且直至指令之再起動，才執行階段2。

【0064】 如本文中所描述，在一個實例中，經處理之指令為神經網

路處理輔助指令。在一個實例中，指令在通用處理器(例如，處理器104)上起始且由指令指定之功能取決於該功能而在通用處理器及/或專用處理器(例如，神經網路處理器105)上執行。舉例而言，在一個實例中，在通用處理器上執行神經網路處理輔助指令之查詢功能且在專用處理器上執行非查詢功能。然而，其他變化係可能的。若功能待在專用處理器上執行(例如，其為非查詢功能，或在另一實例中，為一或多個選定功能)，則資訊例如藉由通用處理器提供至專用處理器以用於執行該功能，諸如與待用於神經網路計算中之張量資料相關的記憶體位址資訊。專用處理器獲得資訊且執行功能。在完成執行功能之後，處理返回至通用處理器，該通用處理器完成指令。在其他實例中，該指令在一或多個通用處理器或一或多個專用處理器上起始、執行及完成。其他變化係可能的。

【0065】 在一個實例中，參看圖3A，神經網路處理輔助指令300具有RRE格式，其表示具有延伸作業碼(opcode)之暫存器及暫存器操作。如圖3A中所展示，在一個實例中，神經網路處理輔助指令300包括指示神經網路處理輔助操作之作業碼(opcode)欄位302 (例如，位元0至15)。在一個實例中，指令之位元16至31被保留且將含有零。在本文中對指令及/或指令功能之描述中，指示特定位置、特定欄位及/或特定欄位大小(例如，特定位元組及/或位元)。然而，可提供其他位置、欄位及/或大小。另外，儘管可指定將位元設定為例如一或零之特定值，但此僅為實例。在其他實例中，若設定，則可將位元設定為不同值，諸如相反值或另一值。許多變化係可能的。

【0066】 在一個實例中，該指令使用由該指令隱含地指定之複數個通用暫存器。舉例而言，神經網路處理輔助指令300使用隱含的暫存器：

通用暫存器0及通用暫存器1，其實例係分別參看圖3B及圖3D進行描述。

【0067】 參看圖3B，在一個實例中，通用暫存器0包括功能碼欄位及可在指令完成後更新之狀態欄位。作為實例，通用暫存器0包括回應碼欄位310 (例如，位元0至15)、異常旗標欄位312 (例如，位元24至31)及功能碼欄位314 (例如，位元56至63)。此外，在一個實例中，保留通用暫存器0之位元16至23及32至55，且其將含有零。一或多個欄位供由指令執行之特定功能使用。在一個實例中，並非所有欄位均由所有功能使用。在下文描述欄位中之每一者：

【0068】 回應碼(RC) 310：此欄位(例如，位元位置0至15)含有回應碼。當神經網路處理輔助指令之執行以例如一之條件碼完成時，儲存回應碼。當遇到無效輸入條件時，將非零值儲存至回應碼欄位，其指示在執行期間辨識到無效輸入條件之原因，且設定選定條件碼，例如1。在一個實例中，如下定義儲存至回應碼欄位之碼：

【0069】 回應碼		含義
【0070】	0001	模型不支援如由參數區塊版本號碼指定之參數區塊的格式。
【0071】	0002	機器上未定義或安裝指定功能。
【0072】	0010	不支援指定張量資料佈局格式。
【0073】	0011	不支援指定張量資料類型。
【0074】	0012	指定單個張量維度大於最大維度索引大小。
【0075】	0013	指定張量之大小大於最大張量大小。
【0076】	0014	指定張量位址未在4K位元組邊界上對準。
【0077】	0015	功能特定保存區域位址未在4K位元組邊界上

對準。

【0078】 F000-FFFF功能特定回應碼。針對某些功能定義此等回應碼。

【0079】 異常旗標(EF) 312：此欄位(例如，位元位置24至31)包括異常旗標。若在指令執行期間偵測到異常條件，則對應異常旗標控制項(例如，位元)將被設定為例如一；否則，控制項保持不變。在第一次調用指令之前，將異常旗標欄位初始化為零。在指令執行期間，保留旗標不變。在一個實例中，如下定義儲存至異常旗標欄位之旗標：

【0080】 EF (位元) 含義

【0081】 0 範圍違規。當在輸入張量中偵測到非數值或非數值經儲存至輸出張量時，設定此旗標。此旗標例如僅在該指令運用例如0之條件碼完成時係有效的。

【0082】 1-7 保留。

【0083】 功能碼(FC) 314：此欄位(例如，位元位置56至63)包括功能碼。為神經網路處理輔助指令指派之功能碼的實例描繪於圖3C中。未指派所有其他功能碼。若指定未指派或未安裝之功能碼，則設定例如0002 hex之回應碼及例如1之選擇條件碼。此欄位在執行期間不會被修改。

【0084】 如所指示，除通用暫存器0以外，神經網路處理輔助指令亦使用通用暫存器1，其實例描繪於圖3D中。作為實例，24位元定址模式中之位元40至63、31位元定址模式中之位元33至63或64位元定址模式中之位元0至63包括參數區塊320之位址。舉例而言，通用暫存器1之內容指定儲存器中之參數區塊的最左位元組之邏輯位址。參數區塊待指明於雙字

邊界上；否則，辨識到規格異常。對於所有功能，通用暫存器1之內容不會被修改。

【0085】 在存取暫存器模式中，作為實例，存取暫存器1指定含有參數區塊、輸入張量、輸出張量及功能特定保存區域之位址空間。

【0086】 在一個實例中，取決於由待執行之指令指定的功能，參數區塊可具有不同格式。舉例而言，該指令之查詢功能具有一個格式之參數區塊，且該指令之其他功能具有另一格式之參數區塊。在另一實例中，所有功能均使用相同的參數區塊格式。其他變化亦係可能的。

【0087】 參看圖3E描述由諸如NNPA查詢可用功能(QAF)操作之查詢功能使用的參數區塊之一個實例。如所展示，在一個實例中，NNPA查詢可用功能參數區塊330包括例如：

【0088】 已安裝功能向量332：參數區塊之此欄位(例如，位元組0至31)包括已安裝功能向量。在一個實例中，已安裝功能向量之位元0至255分別對應於神經網路處理輔助指令之功能碼0至255。當位元為例如一時，安裝對應功能；否則，不安裝功能。

【0089】 已安裝參數區塊格式向量334：參數區塊之此欄位(例如，位元組32至47)包括已安裝參數區塊格式向量。在一個實例中，已安裝參數區塊格式向量之位元0至127對應於用於神經網路處理輔助指令之非查詢功能的參數區塊格式0至127。當位元為例如一時，安裝對應參數區塊格式；否則，不安裝該參數區塊格式。

【0090】 已安裝資料類型336：參數區塊之此欄位(例如，位元組48至49)包括已安裝資料類型向量。在一個實例中，已安裝資料類型向量之位元0至15對應於正安裝之資料類型。當位元為例如一時，安裝對應資料

類型；否則，不安裝資料類型。實例資料類型包括(額外、更少及/或其他資料類型係可能的)：

- 【0091】 位元 料類型
- 【0092】 0 NNP資料類型1
- 【0093】 1-15 保留

【0094】 已安裝資料佈局格式338：參數區塊之此欄位(例如，位元組52至55)包括已安裝資料佈局格式向量。在一個實例中，已安裝資料佈局格式向量之位元0至31對應於正安裝之資料佈局格式。當位元為例如一時，安裝對應資料佈局格式；否則，不安裝該資料佈局格式。實例資料佈局格式包括(額外、更少及/或其他資料佈局格式係可能的)：

- 【0095】 位元 資料佈局格式
- 【0096】 0 4D-特徵張量
- 【0097】 1 4D-內核張量
- 【0098】 2-31 保留

【0099】 最大維度索引大小340：參數區塊之此欄位(例如，位元組60至63)包括例如32位元無正負號二進位整數，其指定任何指定張量之指定維度索引大小中的元素之最大數目。在另一實例中，最大維度索引大小指定任何指定張量之指定維度索引大小中的位元組之最大數目。其他實例亦為可能的。

【0100】 最大張量大小342：參數區塊之此欄位(例如，位元組64至71)包括例如32位元無正負號二進位整數，其指定包括張量格式所需之任何填補位元組的任何指定張量中之位元組的最大數目。在另一實例中，最大張量大小指定包括張量格式所需之任何填補的任何指定張量中之總元素

的最大數目。其他實例亦為可能的。

【0101】 已安裝NNP資料類型1轉換向量344：參數區塊之此欄位(例如，位元組72至73)包括已安裝NNP資料類型1轉換向量。在一個實例中，已安裝NNP資料類型1轉換向量之位元0至15對應於自/至NNP資料類型1格式之已安裝資料類型轉換。當位元為一時，安裝對應轉換；否則，不安裝轉換。可指定額外、更少及/或其他轉換。

【0102】	位元	資料類型
【0103】	0	保留
【0104】	1	BFP微格式
【0105】	2	BFP短格式
【0106】	3-15	保留

【0107】 儘管參看圖3E描述用於查詢功能之參數區塊的一個實例，但可使用用於查詢功能之參數區塊的其他格式，包括NNPA查詢可用功能操作。在一個實例中，該格式可取決於待執行之查詢功能之類型。另外，參數區塊及/或參數區塊之每一欄位可包括額外、更少及/或其他資訊。

【0108】 除用於查詢功能之參數區塊以外，在一個實例中，亦存在用於非查詢功能之參數區塊格式，諸如神經網路處理輔助指令之非查詢功能。參看圖3F描述由非查詢功能使用之參數區塊的一個實例，諸如神經網路處理輔助指令之非查詢功能。

【0109】 如所展示，在一個實例中，由例如神經網路處理輔助指令之非查詢功能使用的參數區塊350包括例如：

【0110】 參數區塊版本號碼352：參數區塊之此欄位(例如，位元組0至1)指定參數區塊之版本及大小。在一個實例中，參數區塊版本號碼之

位元0至8被保留且將含有零，且參數區塊版本號碼之位元9至15含有指定參數區塊之格式的無正負號二進位整數。查詢功能提供指示可用之參數區塊格式的機制。當模型不支援指定參數區塊之大小或格式時，例如0001 hex之回應碼儲存於通用暫存器0中，且指令藉由設定條件碼(例如，條件碼1)而完成。參數區塊版本號碼係由程式指定且在指令執行期間不會被修改。

【0111】 模型版本號碼354：參數區塊之此欄位(例如，位元組2)為識別執行指令(例如，特定非查詢功能)之模型的無正負號二進位整數。當接續旗標(下文所描述)為一時，模型版本號碼可為操作之輸入，以便解譯參數區塊之接續狀態緩衝器欄位(下文所描述)的內容，以重新繼續操作。

【0112】 接續旗標356：參數區塊之此欄位(例如，位元63)當為例如一時指示操作部分完成且接續狀態緩衝器之內容可用以重新繼續操作。程式將接續旗標初始化為零且在出於重新繼續操作之目的而重新執行指令的情況下不修改接續旗標；否則，結果不可預測。

【0113】 若在操作開始時設定接續旗標且參數區塊之內容自初始調用起已改變，則結果不可預測。

【0114】 功能特定保存區域位址358：參數區塊之此欄位(例如，位元組56至63)包括功能特定保存區域之邏輯位址。在一個實例中，功能特定保存區域位址待在4K位元組邊界上對準；否則，例如0015 hex之回應碼設定於通用暫存器0中且指令以例如1之條件碼完成。位址以當前定址模式為準。功能特定保存區域之大小取決於功能碼。

【0115】 當整個功能特定保存區域與指明的程式事件記錄(PER)儲存區域重疊時，在適用時針對功能特定保存區域而辨識PER儲存區改變事

件。當僅功能特定保存區域之一部分與指明的PER儲存區域重疊時，會發生以下情況中之哪一種為模型相依的：

【0116】 * 在適用時針對整個功能特定保存區域而辨識PER儲存區改變事件。

【0117】 * 在適用時針對所儲存之功能特定保存區域的部分而辨識PER儲存區改變事件。

【0118】 當整個參數區塊與指定的PER儲存區域重疊時，適當時針對參數區塊辨識到PER儲存區改變事件。當僅參數區塊之一部分與指定的PER儲存區域重疊時，會發生以下情況中之哪一種為模型相依的：

【0119】 * 在適用時針對整個參數區塊而辨識PER儲存區改變事件。

【0120】 * 適當時針對參數區塊中經儲存部分辨識到PER儲存器改變事件。

【0121】 在適用時針對參數區塊而辨識PER零位址偵測事件。在一個實例中，零位址偵測不適用於張量位址或功能特定保存區域位址。

【0122】 輸出張量描述符(例如，1至2) 360/輸入張量描述符(例如，1至3) 365：參看圖3G描述張量描述符之一個實例。在一個實例中，張量描述符360、365包括：

【0123】 資料佈局格式382：張量描述符之此欄位(例如，位元組0)指定資料佈局格式。舉例而言，有效資料佈局格式包括(額外、更少及/或其他資料佈局格式係可能的)：

【0124】 格式	描述	對準(位元組)
-----------	----	---------

【0125】 0	4D-特徵張量	4096
----------	---------	------

【0126】 1 4D-內核張量 4096

【0127】 2-255 保留 --

【0128】 若指定不支援或保留的資料佈局格式，則例如0010 hex之回應碼儲存於通用暫存器0中，且指令藉由設定例如1之條件碼而完成。

【0129】 資料類型384：此欄位(例如，位元組1)指定張量之資料類型。下文描述所支援之資料類型的實例(額外、更少及/或其他資料類型係可能的)：

【0130】 值	資料類型	資料大小(位元)
----------	------	----------

【0131】 0	NNP資料類型1	16
----------	----------	----

【0132】 1-255	保留	--
--------------	----	----

【0133】 若指定不支援或保留的資料類型，則例如0011 hex之回應碼儲存於通用暫存器0中，且指令藉由設定例如1之條件碼而完成。

【0134】 維度1至4索引大小386：總體而言，維度索引大小一至四指定4D張量之形狀。每一維度索引大小將大於零且小於或等於最大維度索引大小(340，圖3E)；否則，例如0012 hex之回應碼儲存於通用暫存器0中且指令藉由設定例如1之條件碼而完成。總張量大小將小於或等於最大張量大小(342，圖3E)；否則，例如0013 hex之回應碼儲存於通用暫存器0中且指令藉由設定例如1之條件碼而完成。

【0135】 在一個實例中，為判定具有NNPA資料類型1之元素的4D特徵張量中的位元組之數目(亦即，總張量大小)，使用以下公式：維度索引4*維度索引3*ceil (維度索引2/32)*32*ceil (維度索引1/64)*64*2。

【0136】 張量位址388：張量描述符之此欄位(例如，位元組24至31)包括張量之最左位元組的邏輯位址。位址以當前定址模式為準。

【0137】 若位址在相關聯的資料佈局格式的邊界上未對準，則例如 0014 hex之回應碼儲存於通用暫存器0中且指令藉由設定例如1之條件碼而完成。

【0138】 在存取暫存器模式中，存取暫存器1指定儲存器中含所有作用中輸入及輸出張量之位址空間。

【0139】 返回圖3F，在一個實例中，參數區塊350進一步包括可由特定功能使用之功能特定參數1至5 (370)，如本文中所描述。

【0140】 此外，在一個實例中，參數區塊350包括接續狀態緩衝器欄位375，該欄位包括在此指令之操作待重新繼續的情況下使用的資料(或資料之位置)。

【0141】 作為操作之輸入，參數區塊之保留欄位應含有零。當操作結束時，保留欄位可儲存為零或保持不變。

【0142】 儘管參看圖3F描述用於非查詢功能之參數區塊的一個實例，但可使用用於非查詢功能之參數區塊的其他格式，包括神經網路處理輔助指令之非查詢功能。在一個實例中，該格式可取決於待執行之功能之類型。此外，儘管參看圖3G描述張量描述符之一個實例，但亦可使用其他格式。此外，可使用用於輸入及輸出張量之不同格式。其他變化為可能的。

【0143】 下文描述關於由神經網路處理輔助指令之一個實施例支援之各種功能的其他細節：

【0144】 功能碼0：NNPA-QAF (查詢可用功能)

【0145】 神經網路處理輔助(NNPA)查詢功能提供指示諸如以下各者之選定資訊的機制：已安裝功能之可用性；已安裝參數區塊格式；已安

裝資料類型；已安裝資料佈局格式；最大維度索引大小及最大張量大小。獲得資訊且將其置放於諸如參數區塊(例如，參數區塊330)之選定位置中。當操作結束時，參數區塊之保留欄位可儲存為零或可保持不變。

【0146】 在執行查詢功能之一個實施例時，諸如通用處理器104之處理器獲得與選定處理器之特定模型相關之資訊，諸如神經網路處理器(諸如，神經網路處理器105)之特定模型。處理器或機器之特定模型具有某些能力。處理器或機器之另一模型可具有額外、更少及/或不同能力及/或屬於具有額外、更少及/或不同能力之不同代(例如，當代或未來一代)。將所獲得資訊置放於參數區塊(例如，參數區塊330)或可由一或多個應用程式存取及/或供一或多個應用程式使用之其他結構中，該一或多個應用程式可在進一步處理中使用此資訊。在一個實例中，參數區塊及/或參數區塊之資訊維持於記憶體中。在其他實施例中，可在一或多個硬體暫存器中維護參數區塊及/或資訊。作為另一實例，查詢功能可為由作業系統執行之特權操作，其使應用程式設計介面可用以使此資訊可用於應用程式或非特權程式。在又另一實例中，該查詢功能係由諸如神經網路處理器105之專用處理器執行。其他變化為可能的。

【0147】 資訊係例如藉由執行查詢功能之處理器的韌體獲得。韌體知曉特定處理器(例如，神經網路處理器)之特定模型的屬性。此資訊可儲存於例如控制區塊、暫存器及/或記憶體中及/或以其他方式可由執行查詢功能之處理器存取。

【0148】 舉例而言，所獲得資訊包括關於特定處理器之至少一或多個資料屬性的模型相依詳細資訊，包括例如特定處理器之選定模型的一或多個已安裝或支援的資料類型、一或多個已安裝或支援的資料佈局格式及

/或一或多個已安裝或支援的資料大小。當查詢功能(例如，NNPA-QAF功能)之執行完成時，作為實例，設定條件碼0。在一個實例中，條件碼1、2及3不適用於查詢功能。下文描述與所獲得資訊相關之其他資訊。

【0149】 如所指示，在一個實例中，所獲得資訊包括關於例如神經網路處理器之特定模型之一或多個資料屬性的模型相依資訊。資料屬性之一個實例為神經網路處理器之已安裝資料類型。舉例而言，作為實例，神經網路處理器(或其他處理器)之特定模型可支援一或多個資料類型，諸如NNP資料類型1資料類型(亦被稱作神經網路處理資料類型1資料類型)及/或其他資料類型。NNP資料類型1資料類型為16位元浮點格式，其為深度學習訓練及推斷計算提供數個優點，包括例如：保持深度學習網路之準確度；消除簡化捨入模式及處置極端狀況之次正常格式；自動捨入至最接近的算術運算值；及將無窮大且非數字(NaN)之特殊實體組合成一個值(NINF)，其由算術運算接受及處置。NINF提供指數溢出及無效運算(諸如，除以零)之較佳預設。此允許許多程式在不隱藏此類錯誤之情況下及在不使用特殊化異常處理常式之情況下繼續執行。其他模型相依資料類型亦係可能的。

【0150】 NNP資料類型1資料類型之格式的一個實例描繪於圖4中。如所描繪，在一個實例中，NNP資料類型1資料可以格式400表示，其包括例如正負號402 (例如，位元0)、指數+31 404(例如，位元1至6)及分數406 (例如，位元7至15)。

【0151】 下文描繪NNP資料類型1格式之實例性質：

【0152】 性質	NNP資料類型1
【0153】 格式長度(位元)	16個位元

【0154】	有偏指數長度(位元)	6個位元
【0155】	分數長度(位元)	9個位元
【0156】	精確度(P)	10個位元
【0157】	最大左單位視圖指數(E _{max})	32
【0158】	最小左單位視圖指數(E _{min})	-31
【0159】	左單位視圖(LUV)偏差	31
【0160】	N _{max}	$(1-2^{-9}) \times 2^{33} \approx 8.6 \times 10^9$
【0161】	N _{min}	$(1+2^{-9}) \times 2^{-31} \approx 4.6 \times 10^{-10}$
【0162】	D _{min}	---

其中 $\approx$ 指示值為近似值，N_{max}為最大(在量值上)可表示的有限數，且N_{min}為最小(在量值上)可表示的數字。

【0163】 下文描述與NNP資料類型1資料類型相關之其他細節：

【0164】 有偏指數：上文展示了用以允許指數表達為無正負號數之偏差。有偏指數類似於二進位浮點格式之特性，除了對全零及全一之有偏指數未添加特殊含義，如下文參考NNP資料類型1資料類型之類別所描述。

【0165】 有效位：NNP資料類型1數之二進位小數點被視為在最左分數位元之左方。二進位小數點之左邊存在暗示單位位元，其對於正常數被視為一且對於零被視為零。左方附加有隱含的單位位元之分數為數字之有效位。

【0166】 正常NNP資料類型1之值為有效位乘以基數2之無偏指數冪。

【0167】 非零數之值：下文展示非零數之值：

0時，分數為非零。當有偏指數為全一時，分數並非全一。其他有偏指數值可具有任何分數值。對於所有正常數，隱含的單位位元為一。

【0181】 NINF：NINF由全一之有偏指數及全一之分數表示。NINF表示不在NNP資料類型1（亦即，經設計用於深度學習之16位元浮點，其具有6個指數位元及9個分數位元）中之可表示值之範圍內的值。通常，NINF僅在計算期間傳播使得其在最後保持可見。

【0182】 儘管在一個實例中支援NNP資料類型1資料類型，但可支援其他模型相依、特殊的或非標準的資料類型以及一或多個標準資料類型，包括但不限於：僅舉幾例，IEEE 754短精度、二進位浮點16位元、IEEE半精度浮點、8位元浮點、4位元整數格式及/或8位元整數格式。此等資料格式對於神經網路處理具有不同品質。作為實例，較小資料類型（例如，較少位元）可被較快地處理且使用較少快取記憶體/記憶體，且較大資料類型在神經網路中提供較高的結果準確度。待支援之資料類型可在查詢參數區塊中（例如，在參數區塊330之已安裝資料類型欄位336中）具有一或多個指派位元。舉例而言，在已安裝資料類型欄位中指示由特定處理器支援之特殊或非標準資料類型，但不指示標準資料類型。在其他實施例中，亦指示一或多個標準資料類型。其他變化為可能的。

【0183】 在一個特定實例中，為NNP資料類型1資料類型保留已安裝資料類型欄位336之位元0，且當將其設定為例如1時，其指示處理器支援NNP資料類型1。作為一實例，已安裝資料類型之位元向量經組態以表示多達16個資料類型，其中將一位元指派給每一資料類型。然而，在其他實施例中，位元向量可支援更多或更少資料類型。另外，可組態向量，其中將一或多個位元指派給一資料類型。許多實例係可能的及/或可在向量

中支援及/或指示額外、更少及/或其他資料類型。

【0184】 在一個實例中，查詢功能獲得安裝於模型相依處理器上之資料類型的指示，且藉由例如設定參數區塊330之已安裝資料類型欄位336中之一或多個位元而將指示置放於參數區塊中。另外，在一個實例中，查詢功能獲得已安裝資料佈局格式(另一資料屬性)之指示，且藉由例如設定已安裝資料佈局格式欄位338中之一或多個位元而將資訊置放於參數區塊中。舉例而言，實例資料佈局格式包括4D特徵張量佈局及4D內核張量佈局。在一個實例中，本文中所描述之功能使用4D特徵張量佈局，且在一個實例中，卷積功能使用4D內核張量佈局。此等資料佈局格式以提高執行神經網路處理輔助指令之功能之處理效率的方式配置儲存器中用於張量之資料。舉例而言，為高效地操作，神經網路處理輔助指令使用以特定資料佈局格式提供之輸入張量。儘管提供實例佈局，但可針對本文中所描述之功能及/或其他功能提供額外、更少及/或其他佈局。

【0185】 特定處理器模型之佈局的使用或可用性由已安裝資料佈局格式(例如，參數區塊330之欄位338)之向量提供。舉例而言，該向量為允許CPU向應用程式傳達支援哪些佈局之已安裝資料佈局格式之位元向量。舉例而言，為4D特徵張量佈局保留位元0，且當將其設定為例如1時，其指示處理器支援4D特徵張量佈局；且為4D內核張量佈局保留位元1，且當將其設定為例如1時，其指示處理器支援4D內核張量佈局。在一個實例中，已安裝資料佈局格式之位元向量經組態以表示多達16個資料佈局，其中將一位元指派給每一資料佈局。然而，在其他實施例中，一位元向量可支援更多或更少資料佈局。另外，可組態向量，其中將一或多個位元指派給資料佈局。許多實例係可能的。下文描述關於4D特徵張量佈局及4D內

核張量佈局之其他細節。同樣，現在或在未來可使用其他佈局以最佳化效能。

【0186】 在一個實例中，神經網路處理輔助指令使用4D張量進行操作，亦即，具有4個維度之張量。此等4D張量係例如以列優先方式自本文中所描述之通用輸入張量獲得，亦即，當以遞增的記憶體位址次序列舉張量元素時，稱為E1之內部維度將首先經由以0開始至E1索引大小-1之E1索引大小值遞增，之後將增加E2維度之索引且重複E1維度之步進。最後增加稱為E4維度之外部維度的索引。

【0187】 具有較低數目個維度之張量(例如，3D或1D張量)將表示為4D張量，其中該4D張量之一或多個維度超過設定為1之原始張量維度。

【0188】 本文中描述將具有維度E4、E3、E2、E1之列優先通用4D張量變換成4D特徵張量佈局(在本文中亦被稱作NNPA資料佈局格式0 4D特徵張量)：

【0189】 舉例而言，所得張量可表示為例如64元素向量之4D張量或具有如下維度之5D張量：

【0190】 $E4, [E1/64], E3, [E2/32] * 32, 64$ ， $\lceil \cdot \rceil$ 係指向上取整函數。(換言之： $E4 * E3 * \text{ceil}(E2/32) * 32 * \text{ceil}(E1/64) * 64$ 個元素。)

【0191】 通用張量之元素 $[e4][e3][e2][e1]$ 可映射至所得5D張量之以下元素：

【0192】 $[e4][\lfloor e1/64 \rfloor][e3][e2][e1 \bmod 64]$ ，其中 $\lfloor \cdot \rfloor$ 為向下取整函數且mod係模數。(換言之：元素 $(E3 * e2_limit * e1_limit * e4x) + (e2_limit * e3x * 64) + (e2x * 64) + (\lfloor e1x/64 \rfloor * e2_limit * E3 * 64) + (e1x \bmod 64)$ ，其中 $e2_limit = [E2/32] * 32$ 且 $e1_limit = [E1/64] * 64$ 。)

【0193】 所得張量可大於通用張量。在通用張量中不具有對應元素之所得張量之元素稱為填補元素。

【0194】 考慮64元素向量之NNPA資料佈局格式0 4D特徵張量的元素 $[fe4][fe1][fe3][fe2][fe0]$ 或其等效表示為元素之5D張量。此元素為填補元素或其在通用4D張量中之對應元素，其中維度E4、E3、E2、E1可藉由以下公式判定：

【0195】 · 若 $fe2 \geq E2$ ，則此為E2 (或頁面)填補元素

【0196】 · 否則，若 $fe1 * 64 + fe0 \geq E1$ ，則此為E1 (或列)填補元素

【0197】 · 否則，通用4D張量中之對應元素為：

【0198】 $[fe4][fe3][fe2][fe1 * 64 + fe0]$

【0199】 對於基於卷積神經網路之人工智慧模型，特徵張量之4個維度的含義通常可映射至：

【0200】 · E4：N-小批次之大小

【0201】 · E3：H-3D張量/影像之高度

【0202】 · E2：W-3D張量/影像之寬度

【0203】 · E1：C-3D張量之通道或類別

【0204】 對於基於機器學習或遞迴神經網路之人工智慧模型，4D特徵張量之4個維度的含義通常可映射至：

【0205】 · E4：T-時間步驟或模型之數目

【0206】 · E3：保留，通常設定為1

【0207】 · E2：N_{mb}-小批次大小

【0208】 · E1：L-特徵

【0209】 NNPA資料佈局格式0提供例如具有4k位元組資料區塊(頁

面)之二維資料局部性以及用於所產生張量之外部維度的4k位元組區塊資料對準。

【0210】對於輸入張量，填補元素位元組被忽略，且對於輸出張量，填補元素位元組不可預測。填補位元組上之PER儲存區改變不可預測。

【0211】具有維度E1、E2、E3及E4之4D特徵張量佈局的輸入資料佈局之一個實例展示於圖5A至圖5C中，且4D特徵張量佈局之實例輸出描繪於圖6A至圖6C中。參看圖5A，展示3D張量500，其具有維度E1、E2及E3。在一個實例中，每一3D張量包括複數個2D張量502。每一2D張量502中之數字描述其元素中之每一者將在記憶體中之位置的記憶體偏移。輸入用以將原始張量(例如，圖5A至圖5C之原始4D張量)之資料佈置在記憶體中，如對應於圖5A至圖5C之圖6A至圖6C中所展示。

【0212】在圖6A中，作為實例，記憶體600之單元(例如，記憶體頁面)包括預選數目(例如，32)個列602，其中之每一者由例如e2_page_idx識別；且每一列具有預選數目(例如，64)個元素604，其各自由例如e1_page_idx識別。若一列不包括預選數目個元素，則其被填補606，被稱作列或E1填補；且若記憶體單元不具有預選數目個列，則其被填補608，被稱作頁面或E2填補。作為實例，列填補為例如零或其他值等，且頁面填補為例如現有值、零或其他值。

【0213】在一個實例中，列之輸出元素係基於其對應輸入在E1方向上之元素位置而提供於記憶體中(例如，頁面中)。舉例而言，參看圖5A，所展示之三個矩陣的元素位置0、1及2(例如，在每一矩陣之同一位置處的元素位置)展示於圖6A之頁面0的列0中，等等。在此實例中，4D張量為

小的，且表示4D張量之每一2D張量的所有元素皆適配於一個頁面中。然而，此情形僅為一個實例。2D張量可包括一或多個頁面。若2D張量係基於4D張量之重新格式化而產生，則2D張量之頁面數目係基於4D張量之大小。在一個實例中，一或多個向上取整函數用以判定2D張量中之列數目及每一列中之元素數目，其將指示待使用多少頁面。其他變化為可能的。

【0214】除4D特徵張量佈局以外，在一個實例中，神經網路處理器亦可支援4D內核張量，其重新配置4D張量之元素以在執行諸如卷積之某些人工智慧(例如，神經網路處理輔助)操作時減小記憶體存取及資料搜集步驟之數目。作為實例，將具有維度E4、E3、E2、E1之列優先通用4D張量變換成NNPA資料佈局格式1 4D內核張量(4D內核張量)，如本文中所描述：

【0215】所得張量可表示為例如64元素向量之4D張量或具有如下維度之5D張量：

【0216】 $\lceil E1/64 \rceil, E4, E3, \lceil E2/32 \rceil * 32, 64$ ， $\lceil \cdot \rceil$ 係指向上取整函數。(換言之： $E4 * E3 * \lceil E2/32 \rceil * 32 * \lceil E1/64 \rceil * 64$ 個元素。)

【0217】通用張量之元素 $[e4][e3][e2][e1]$ 可映射至所得5D張量之以下元素：

【0218】 $\lfloor e1/64 \rfloor [e4][e3][e2][e1 \bmod 64]$ ，其中 $\lfloor \cdot \rfloor$ 係指向下取整函數且mod係模數。換言之：元素 $(\lfloor e1x/64 \rfloor * E4 * E3 * e2_limit * 64) + (e4x * E3 * e2_limit * 64) + (e3x * e2_limit * 64) + (e2x * 64) + (e1x \bmod 64)$ ，其中 $e2_limit = \lceil E2/32 \rceil * 32$ 且 $e1_limit = \lceil E1/64 \rceil * 64$ 。

【0219】所得張量可大於通用張量。在通用張量中不具有對應元素之所得張量之元素稱為填補元素。

【0220】 考慮64元素向量之NNPA資料佈局格式1 4D特徵張量的元素[fe1][fe4][fe3][fe2][fe0]或其等效表示為元素之5D張量。此元素為填補元素或其在通用4D張量中之對應元素，其中維度E4、E3、E2、E1可藉由以下公式判定：

【0221】 · 若 $fe2 \geq E2$ ，則此為E2 (或頁面)填補元素

【0222】 · 否則，若 $fe1 * 64 + fe0 \geq E1$ ，則此為E1 (或列)填補元素

【0223】 · 否則，通用4D張量中之對應元素為

【0224】 [fe4][fe3][fe2][fe1*64+fe0]

【0225】 對於基於卷積神經網路之人工智慧模型，核心張量之4個維度的含義通常可映射至：

【0226】 · E4：H-3D張量/影像之高度

【0227】 · E3：W-3D張量/影像之寬度

【0228】 · E2：C-3D張量之通道數目

【0229】 · E1：K-內核數目

【0230】 NNPA資料佈局格式1提供例如4k位元組資料區塊(頁面)內之二維內核平行度以及用於產生張量之外部維度的4k位元組區塊資料對準，以實現高效處理。

【0231】 對於輸入張量，忽略填補位元組。填補位元組上之PER儲存區改變不可預測。

【0232】 同樣，儘管實例資料佈局格式包括4D特徵張量佈局及4D內核張量佈局，但處理器(例如，神經網路處理器105)可支援其他資料佈局格式。獲得所支援資料佈局之指示且藉由設定例如欄位338中之一或多個位元而將其置放於查詢參數區塊中。

【0233】 根據本發明之一或多個態樣，查詢參數區塊亦包括其他資料屬性資訊，其包括例如資料之所支援大小資訊。諸如神經網路處理器之處理器通常具有基於內部緩衝器大小、處理單元、資料匯流排結構、記憶體限制等之限制，其可限制張量維度之最大大小及/或張量之總大小。因此，查詢功能提供欄位以將此等限制傳達至應用程式。舉例而言，基於執行查詢功能，處理器獲得各種資料大小，諸如最大維度索引大小(例如，65,536個元素)及最大張量大小(例如，8 GB)，且將此資訊分別包括於參數區塊(例如，參數區塊330)之欄位340及342中。額外、更少及/或其他大小資訊亦可由處理器(例如，神經網路處理器105)支援，且因此獲得並置放於參數區塊中，例如欄位340、342及/或其他欄位中。在其他實施例中，限制可更小或更大，及/或大小可為其他單位，諸如位元組而非元素、元素而非位元組等。此外，其他實施例允許每一維度具有不同最大大小，而非所有維度具有相同最大值。許多變化係可能的。

【0234】 根據本發明之一或多個態樣，提供查詢功能，該查詢功能傳達與選定處理器(例如，神經網路處理器105)之特定模型相關的詳細資訊。舉例而言，詳細資訊包括與特定處理器相關之模型相依資訊。(處理器亦可支援標準資料屬性，諸如標準資料類型、標準資料佈局等，該等屬性為隱含的且未必由查詢功能呈現；但在其他實施例中，查詢功能可指示資料屬性之所有或各種選定子集等。)儘管提供實例資訊，但在其他實施例中，可提供其他資訊。針對處理器之不同模型及/或不同處理器可能不同的所獲得資訊用以執行人工智慧及/或其他處理。人工智慧及/或其他處理可使用例如神經網路處理輔助指令之一或多個非查詢功能。藉由一或多次執行神經網路處理輔助指令及指定非查詢特定功能來執行用於處理中之

特定非查詢功能。

【0235】 下文描述由神經網路處理輔助指令支援之實例非查詢功能的其他細節(在其他實施例中，可支援額外、更少及/或其他功能)：

【0236】 功能碼16：NNPA-ADD (加法)

【0237】 當指定NNPA-ADD功能時，將由張量描述符1所描述之輸入張量1的每一元素加至由張量描述符2所描述之輸入張量2的對應元素，且將所得總和置放於由輸出張量描述符所描述之輸出張量的對應元素中。

【0238】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0239】 在一個實例中，輸入張量1、輸入張量2及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0240】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0241】 功能碼17：NNPA-SUB (減法)

【0242】 當指定NNPA-SUB功能時，自藉由張量描述符1所描述之輸入張量1的對應元素減去由張量描述符2所描述之輸入張量2的每一元素，且將所得差置放於輸出張量之對應元素中。

【0243】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例

如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0244】 在一個實例中，輸入張量1、輸入張量2及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0245】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0246】 功能碼18：NNPA-MUL (乘法)

【0247】 當指定NNPA-MUL功能時，將由張量描述符1所描述之輸入張量1 (乘數)之每一元素與由張量描述符2所描述之輸入張量2 (被乘數)之對應元素的乘積置放於輸出張量之對應元素中。

【0248】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0249】 在一個實例中，輸入張量1、輸入張量2及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0250】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0251】 功能碼19：NNPA-DIV (除法)

【0252】 當指定NNPA-DIV功能時，將由張量描述符1所描述之輸入張量1 (被除數)的每一元素除以由張量描述符2所描述之輸入張量2 (除數)的對應元素，且將商置放於輸出張量之對應元素中。

【0253】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0254】 在一個實例中，輸入張量1、輸入張量2及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0255】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0256】 功能碼20：NNPA-MIN (最小值)

【0257】 當指定NNPA-MIN功能時，將由張量描述符1所描述之輸入張量1的每一元素與由張量描述符2所描述之輸入張量2的對應元素進行比較。將兩個值中之較小值置放於輸出張量描述符之對應元素中。若兩個值相等，則將值置放於輸出張量之對應元素中。

【0258】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0259】 在一個實例中，輸入張量1、輸入張量2及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0260】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0261】 功能碼21：NNPA-MAX (最大值)

【0262】 當指定NNPA-MAX功能時，將由張量描述符1所描述之輸入張量1的每一元素與由張量描述符2所描述之輸入張量2的對應元素進行比較。將兩個值中之較大值置放於輸出張量描述符之對應元素中。若兩個值相同，則將值置放於輸出張量之對應元素中。

【0263】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0264】 在一個實例中，輸入張量1、輸入張量2及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0265】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0266】 功能碼32：NNPA-LOG (自然對數)

【0267】 當指定NNPA-LOG功能時，對於由張量描述符1所描述之輸入張量的每一元素，若彼元素大於零，則由輸出張量描述符所描述之輸出張量中的對應元素為彼元素之自然對數。否則，輸出張量中之對應元素無法用數值表示，且儲存與目標資料類型中之負無窮大相關聯的值。

【0268】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條

件碼完成。

【0269】 在一個實例中，輸入張量1及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0270】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0271】 功能碼33：NNPA-EXP (指數)

【0272】 當指定NNPA-EXP功能時，對於由張量描述符1所描述之輸入張量的每一元素，由輸出張量描述符所描述之輸出張量中的對應元素為彼元素之指數。

【0273】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0274】 在一個實例中，輸入張量1及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0275】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0276】 功能碼49：NNPA-RELU (經整流線性單元)

【0277】 當指定NNPA-RELU功能時，對於由張量描述符1所描述之輸入張量的每一元素，若彼元素小於或等於零，則由輸出張量描述符所描

述之輸出張量中的對應元素為零。否則，輸出張量中之對應元素為輸入張量中之元素與在功能特定參數1中指定之限幅值中的最小值。

【0278】 作為實例，功能特定參數1定義RELU操作之限幅值。舉例而言，限幅值在功能特定參數1之位元16至31中。限幅值指定於例如NNPA資料類型1格式中。限幅值零指示使用最大正值；換言之，不執行限幅。若指定負值，則辨識到一般運算元資料異常。

【0279】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0280】 在一個實例中，輸入張量1及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0281】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符2、輸入張量描述符3及功能特定保存區域位址欄位。在一個實例中，功能特定參數2至5將含有零。

【0282】 功能碼50：NNPA-TANH

【0283】 當指定NNPA-TANH功能時，對於由張量描述符1所描述之輸入張量的每一元素，由輸出張量描述符所描述之輸出張量中的對應元素值為彼元素之雙曲正切。

【0284】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例

如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0285】 在一個實例中，輸入張量1及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0286】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0287】 功能碼51：NNPA-SIGMOID

【0288】 當指定NNPA-SIGMOID功能時，對於由張量描述符1所描述之輸入張量的每一元素，由輸出張量描述符所描述之輸出張量中的對應元素為彼元素之S型。

【0289】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0290】 在一個實例中，輸入張量1及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0291】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符2、輸入張量描述符3、功能特定參數1至5及功能特定保存區域位址欄位。

【0292】 功能碼52：NNPA-SOFTMAX

【0293】 當指定NNPA-SOFTMAX功能時，對於輸入張量1之維度1

中的每一向量，計算輸出張量中之對應向量，如下文所描述：

【0294】 * 計算向量之最大值。

【0295】 * 計算向量之維度1中的每一元素與上文計算之最大值之間的差之指數的總和。若輸入向量之維度1中的元素及上文計算之最大值兩者為數值且差為非數值，則彼元素之指數的結果強制為零。

【0296】 * 對於向量中之每一元素，中間商由元素與上文計算之最大值之間的差之指數除以上文計算之總和形成。將視情況選用之激勵函數應用於此中間商以形成輸出向量中之對應元素。

【0297】 對於例如維度1中的所有維度4索引大小x維度3索引大小x維度2索引大小個向量，重複此程序。

【0298】 在一個實例中，NNPA-SOFTMAX函數特定參數1控制激勵函數。作為一實例，功能特定參數1之ACT欄位(例如，位元28至31)指定激勵函數。實例激勵函數包括：

【0299】 ACT	激勵函數
------------	------

【0300】 0	不執行激勵函數
----------	---------

【0301】 1	LOG
----------	-----

【0302】 2-15	保留
-------------	----

【0303】 若為ACT欄位指定保留值，則報告例如F001 hex之回應碼且操作以例如1之條件碼完成。

【0304】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條

件碼完成。

【0305】 在一個實例中，若輸入張量之維度3索引大小不等於一，則儲存例如F000 hex之回應碼且指令以例如1之條件碼完成。

【0306】 在一個實例中，輸入張量1及輸出張量之形狀、資料佈局及資料類型將相同；否則，辨識到一般運算元資料異常。

【0307】 在一個實例中，忽略輸出張量描述符2、輸入張量描述符2及輸入張量描述符3。在一個實例中，功能特定參數2至5將含有零。

【0308】 此功能可使用8K位元組功能特定保存區域。

【0309】 在一個實施例中，當獲得維度1中之向量時，取決於指定資料佈局格式，元素在記憶體中可能不連續。若輸入張量1之維度1向量的所有元素含有可在指定資料類型中表示之最大量值負數，則結果可能不太準確。

【0310】 功能碼64：NNPA-BATCHNORM (批次正規化)

【0311】 當指定NNPA-BATCHNORM功能時，對於輸入1張量之維度1中的每一向量，藉由將向量中之每一元素乘以構成輸入2張量之維度1向量中之對應元素來計算輸出張量之維度1中的對應向量。接著將全精度乘積加至構成輸入3張量之維度1向量中的對應元素，且接著捨入至輸出張量之指定資料類型的精度。對於例如維度1中的所有維度4索引大小×維度3索引大小×維度2索引大小個向量，重複此程序。

【0312】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條

件碼完成。

【0313】 在一個實例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0314】 * 輸入張量1及輸出張量之形狀及資料佈局將相同。

【0315】 * 輸入張量及輸出張量之資料類型將相同。

【0316】 * 輸入張量1、2、3及輸出張量之維度1索引大小將相同。

【0317】 * 輸入張量2及3之維度2、3及4索引大小將為一。

【0318】 在一個實例中，忽略輸出張量描述符2及功能特定保存區域位址欄位。在一個實例中，功能特定參數2至5將含有零。

【0319】 **功能碼80：NNPA-MAXPOOL2D**

功能碼81：NNPA-AVGPOOL2D

【0320】 當指定NNPA-MAXPOOL2D或NNPA-AVGPOOL2D功能時，藉由指定操作減少由輸入張量1描述符所描述之輸入張量1以彙總輸入窗口。藉由在維度索引2及3上移動2D滑動窗口來選擇輸入窗口。窗口之摘要為輸出張量中之元素。滑動窗口維度係由例如功能特定參數4及功能特定參數5描述。當計算鄰近輸出張量元素時滑動窗口在輸入1張量上移動的量稱為步幅。滑動窗口步幅係由例如功能特定參數2及功能特定參數3指定。當指定NNPA-MAXPOOL2D操作時，在窗口上執行下文所定義之Max運算。當指定NNPA-AVGPOOL2D操作時，在窗口上執行下文所描述之AVG運算。若指定填補類型為有效(Valid)，則將窗口中之所有元素添加至用以計算所得輸出元素之集合。若指定填補類型為相同(Same)，則取決於窗口之位置，僅來自窗口之元素之子集可添加至用以計算所得輸出元

素之集合。

【0321】 在一個實例中，CollectElements操作將元素添加至元素之集合且遞增集合中元素之數目。每當窗口開始位置移動時，集合被清空。是否存取執行操作不需要之元素為不可預測的。

【0322】 Max運算：在一個實例中，藉由將集合中之所有元素彼此進行比較及傳回最大值來計算窗口中之元素的集合中之最大值。

【0323】 Avg (求平均)運算：在一個實例中，將窗口中之元素的集合之平均值計算為集合中之所有元素的總和除以集合中元素之數目。

【0324】 在一個實例中，如下分配欄位：

【0325】 * 集用功能特定參數1控制填補類型。舉例而言，功能特定參數1之位元29至31包括指定填補類型之填補(PAD)欄位。舉例而言，實例類型包括：

【0326】	填補	填補類型
【0327】	0	有效
【0328】	1	相同
【0329】	2-7	保留

【0330】 若為填補欄位指定保留值，則報告例如F000 hex之回應碼且操作以例如1之條件碼完成。

【0331】 在一個實例中，保留功能特定參數1之位元位置0至28，且其將含有零。

【0332】 * 功能特定參數2含有例如32位元無正負號二進位整數，其指定維度2步幅(D2S)，該步幅指定滑動窗口在維度2中移動之元素數目。

【0333】 * 功能特定參數3含有例如32位元無正負號二進位整數，其指定維度3步幅(D3S)，該步幅指定滑動窗口在維度3中移動之元素數目。

【0334】 * 功能特定參數4含有例如32位元無正負號二進位整數，其指定維度2窗口大小(D2WS)，該大小指定滑動窗口含有之維度2中之元素數目。

【0335】 * 功能特定參數5含有例如32位元無正負號二進位整數，其指定維度3窗口大小(D3WS)，該大小指定滑動窗口含有之維度3中之元素數目。

【0336】 在一個實例中，功能特定參數2至5中之指定值小於或等於最大維度索引大小，且功能特定參數4至5中之指定值大於零；否則，報告例如0012 hex之回應碼且操作以例如1之條件碼完成。

【0337】 若維度2步幅及維度3步幅兩者均為零且維度2窗口大小或維度3窗口大小大於例如1024，則儲存例如F001 hex之回應碼。若維度2步幅及維度3步幅兩者均大於例如零且維度2窗口大小或維度3窗口大小大於例如64，則儲存例如F002 hex之回應碼。若維度2步幅及維度3步幅兩者均大於例如零且維度2步幅或維度3步幅大於例如30，則儲存例如F003 hex之回應碼。若維度2步幅及維度3步幅兩者均大於例如零且輸入張量維度2索引大小或輸入張量維度3索引大小大於例如1024，則儲存例如F004 hex之回應碼。對於所有以上條件，指令以例如1之條件碼完成。

【0338】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例

如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0339】 在一個實例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0340】 * 輸入張量及輸出張量之維度4索引大小及維度1索引大小將相同。

【0341】 * 輸入張量及輸出張量之資料佈局及資料類型將相同。

【0342】 * 在一個實例中，若維度2步幅及維度3步幅兩者均為零，則以下額外條件將為真：

【0343】 * 輸入張量維度2索引大小將等於維度2窗口大小。

【0344】 * 輸入張量之輸入張量維度3索引大小將等於維度3窗口大小。

【0345】 * 輸出張量之維度2索引大小及維度3索引大小將為一。

【0346】 * 指定填補將為有效。

【0347】 * 在一個實例中，若維度2步幅或維度3步幅為非零，則兩個步幅將為非零。

【0348】 * 在一個實例中，若維度2步幅及維度3步幅兩者均大於零，則以下額外條件將為真：

【0349】 * 當指定填補為有效時，維度2窗口大小將小於或等於輸入張量之維度2索引大小。

【0350】 * 當指定填補為有效時，維度3窗口大小將小於或等於輸入張量之維度3索引大小。

【0351】 * 當指定填補為相同時，將滿足輸入及輸出張量之維度2

索引大小與維度3索引大小之間的以下關係(集用相同填補)：

$$\text{【0352】 } O1D2IS = \left\lceil \frac{I1D2IS}{D2S} \right\rceil$$

$$\text{【0353】 } O1D3IS = \left\lceil \frac{I1D3IS}{D3S} \right\rceil$$

【0354】 其中：

【0355】 $IxDyIS$ 定義於張量描述符x中之輸入張量x的維度y索引大小。

【0356】 $OxDyIS$ 定義於張量描述符x中之輸出張量x的維度y索引大小。

【0357】 $D2S$ 維度2步幅。

【0358】 $D3S$ 維度3步幅。

【0359】 * 當指定填補為有效時，將滿足輸入及輸出張量之維度2索引大小與維度3索引大小之間的以下關係(集用相同填補)：

$$\text{【0360】 } O1D2IS = \left\lceil \frac{(I1D2IS - D2WS + 1)}{D2S} \right\rceil$$

$$\text{【0361】 } O1D3IS = \left\lceil \frac{(I1D3IS - D3WS +)}{D3S} \right\rceil$$

【0362】 其中 $D2WS$ 為維度2窗口大小且 $D3WS$ 為維度3窗口大小。

【0363】 忽略輸出張量描述符2、輸入張量描述符2及3以及功能特定保存區域位址欄位。

【0364】 **功能碼96：NNPA-LSTMACT (長短期記憶體激勵)**

【0365】 當指定NNPA-LSTMACT功能時，由輸入張量1描述符所描述、針對每一維度4索引值而分裂成四個子張量的輸入張量1連同由輸入張量2描述符所描述、針對每一維度4索引值而分裂成四個子張量的輸入張量2以及由輸入張量3描述符所描述之輸入張量3為LSTMACT操作之輸入。在LSTMACT操作結束時，將結果寫入至由輸出張量1描述符所描述之輸出張量1及由輸出張量2描述符所描述之輸出張量2。

【0366】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼0010 hex或0011 hex分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0367】 在一個實施例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0368】 * 輸入張量3以及輸出張量1及2之維度4索引大小將等於例如一。

【0369】 * 輸入張量1及輸入張量2之維度4索引大小將等於例如四。

【0370】 * 例如所有輸入張量及兩個輸出張量之維度3索引大小將等於例如一。

【0371】 * 例如所有輸入張量及兩個輸出張量之資料佈局及資料類型將相同。

【0372】 * 例如所有輸入張量及兩個輸出張量之維度1索引大小將相同。

【0373】 * 例如所有輸入張量及兩個輸出張量之維度2索引大小將相同。

【0374】 在一個實例中，忽略功能特定保存區域位址欄位。在一個實例中，功能特定參數1至5將含有零。

【0375】 **功能碼97：NNPA-GRUACT (閘控遞迴單元激勵)**

【0376】 當指定NNPA-GRUACT功能時，由輸入張量1描述符所描

述、針對每一維度4索引值而分裂成三個子張量的輸入張量1連同由輸入張量2描述符所描述、針對每一維度4索引值而分裂成三個子張量的輸入張量2以及由輸入張量3描述符所描述之輸入張量3為GRUACT操作之輸入。在GRUACT操作結束時，儲存由輸出張量描述符所描述之輸出張量。

【0377】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0378】 在一個實施例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0379】 * 輸出張量及輸入張量3之維度4索引大小將等於例如一。

【0380】 * 輸入張量1及輸入張量2之維度4索引大小將等於例如三。

【0381】 * 例如所有輸入張量及輸出張量之維度3索引大小將等於例如一。

【0382】 * 例如所有輸入張量及輸出張量之維度1索引大小將相同。

【0383】 * 例如所有輸入張量及輸出張量之維度2索引大小將相同。

【0384】 * 例如所有輸入張量及輸出張量之資料佈局及資料類型將相同。

【0385】 在一個實例中，忽略輸出張量描述符2及功能特定保存區

域位址欄位。在一個實例中，功能特定參數2至5將含有零。

【0386】 功能碼112：NNPA-CONVOLUTION

【0387】 當指定NNPA-CONVOLUTION功能時，對於由輸出張量1描述符所描述之輸出張量中的每一輸出元素，自藉由輸入張量1描述符所描述之輸入張量1選擇由維度索引3、2及1組成之3維輸入1窗口。自藉由輸入張量2描述符所描述之張量2選擇由維度索引4、3及2組成的相同大小之3維輸入2窗口。將輸入1窗口中之元素乘以輸入2窗口中之對應元素，且將所有乘積加在一起以產生初始總和。將此初始總和加至輸入張量3之對應元素以計算中間總和值。輸出張量之元素為對中間總和執行之指定激勵函數的結果。若未指定激勵函數，則輸出元素等於中間總和。

【0388】 若指定填補類型為有效，則窗口中之所有元素用以計算所得初始總和。若指定填補類型為相同，則當計算所得初始總和時，取決於窗口之位置，輸入1窗口之一些元素可能隱含為零。

【0389】 是否存取執行操作不需要之元素為不可預測的。

【0390】 在一個實例中，如下分配由卷積函數使用之功能特定參數的欄位：

【0391】 * NNPA-CONVOLUTION功能特定參數1控制填補類型及激勵函數。在一個實例中，功能特定參數1之位元29至31包括指定填補類型之填補欄位。實例類型如下：

【0392】 PAD	填補類型
【0393】 0	有效
【0394】 1	相同
【0395】 2-7	保留

【0396】 若為填補欄位指定保留值，則報告例如F000 hex之回應碼且操作以例如1之條件碼完成。

【0397】 此外，在一個實例中，NNPA-CONVOLUTION功能特定參數1之位元24至27包括指定激勵函數之激勵欄位。實例函數如下：

【0398】 ACT 激勵函數

【0399】 0 不執行激勵函數

【0400】 1 RELU

【0401】 2-15 保留

【0402】 若指定RELU之激勵函數，則如下判定所得輸出元素值：若中間總和值小於或等於零，則輸出張量中之對應元素為零；否則，輸出張量中之對應元素為中間總和值與在功能特定參數4中指定之限幅值中的最小值。

【0403】 若為ACT欄位指定保留值，則報告例如F001 hex之回應碼且操作以例如1之條件碼完成。

【0404】 * 功能特定參數2含有例如32位元無正負號二進位整數，其指定維度2步幅(D2S)，該步幅指定滑動窗口在維度2中移動之元素數目。

【0405】 * 功能特定參數3含有例如32位元無正負號二進位整數，其指定維度3步幅(D3S)，該步幅指定滑動窗口在維度3中移動之元素數目。

【0406】 功能特定參數2至3中之指定值將小於最大維度索引大小；否則，報告例如0012 hex之回應碼且操作以例如1之條件碼完成。

【0407】 * 功能特定參數4定義用於視情況選用之RELU操作的限幅

值。在一個實例中，限幅值在功能特定參數4之位元16至31中。

【0408】 在一個實例中，若ACT欄位為零，則忽略此欄位。若ACT欄位指定RELU，則限幅值指定於NNP資料類型1格式中。限幅值零指示使用最大正值；換言之，不執行限幅。若指定非零，則辨識到一般運算元資料異常。

【0409】 在一個實例中，若除輸入張量2以外的指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若輸入張量2中之指定資料佈局未指定4D內核張量(例如，資料佈局=1)，則回應碼，例如0010 hex設定於通用暫存器0中，且指令以例如1之條件碼完成。在一個實例中，若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則例如0011 hex之回應碼設定於通用暫存器0中且指令以例如1之條件碼完成。

【0410】 若維度2步幅及維度3步幅兩者均為零且輸入張量2之維度3索引大小或維度4索引大小大於例如448，則儲存例如F002 hex之回應碼。若維度2步幅及維度3步幅兩者均大於零且輸入張量2之維度3索引大小或維度4索引大小大於例如64，則儲存例如F003 hex之回應碼且操作以例如1之條件碼完成。若維度2步幅或維度3步幅大於例如13，則儲存例如F004 hex之回應碼且操作以例如1之條件碼完成。

【0411】 在一個實例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0412】 * 輸入張量1、輸入張量3及輸出張量之資料佈局將相同。

【0413】 * 所有輸入張量及輸出張量之資料類型將相同。

【0414】 * 輸入3張量之維度2、維度3及維度4索引大小將為1。

【0415】 * 輸出張量之維度4索引大小將等於輸入1張量之維度4索引大小。

【0416】 * 輸出張量之維度1索引大小將等於輸入2張量之維度1索引大小及輸入3張量之維度1索引大小。

【0417】 * 輸入1張量之維度1索引大小將等於輸入2張量之維度2索引大小。

【0418】 * 在一個實例中，若維度2步幅及維度3步幅兩者均為零，則以下額外條件將為真：

【0419】 * 輸入1張量維度2索引大小將等於輸入2張量之維度3索引大小。

【0420】 * 輸入張量之輸入1張量維度3索引大小將等於輸入2張量之維度4索引大小。

【0421】 * 輸出張量之維度2索引大小及維度3索引大小將為一。

【0422】 * 指定填補將為有效。

【0423】 * 若維度2步幅或維度3步幅為非零，則兩個步幅將為非零。

【0424】 * 在一個實例中，若維度2步幅及維度3步幅兩者均大於零，則以下額外條件將為真：

【0425】 * 當指定填補為有效時，輸入1張量之維度2索引大小將大於或等於輸入張量2之維度3索引大小。

【0426】 * 當指定填補為有效時，輸入1張量之維度3索引大小將大於或等於輸入2張量之維度4索引大小。

【0427】 * 在一個實例(卷積相同填補)中，當指定填補為相同時，

將滿足輸入1張量及輸出張量之維度2索引大小與維度3索引大小之間的以下關係：

$$\text{【0428】 } O1D2IS = \left\lceil \frac{I1D2IS}{D2S} \right\rceil$$

$$\text{【0429】 } O1D3IS = \left\lceil \frac{I1D3IS}{D3S} \right\rceil$$

【0430】 其中：

【0431】 $O1D2IS$ 輸出張量之維度2索引大小。

【0432】 $O1D3IS$ 輸出張量之維度3索引大小。

【0433】 $I1D2IS$ 輸入1張量之維度2索引大小。

【0434】 $I1D3IS$ 輸入1張量之維度3索引大小。

【0435】 $D2S$ 維度2步幅。

【0436】 $D3S$ 維度3步幅。

【0437】 * 在一個實例(卷積有效填補)中，當指定填補為有效時，將滿足輸入1張量之維度2索引大小及維度3索引大小、輸入2張量及輸出張量之維度3索引大小及維度4索引大小之間的以下關係：

$$\text{【0438】 } O1D2IS = \left\lceil \frac{(I1D2IS - I2D3IS + 1)}{D2S} \right\rceil$$

$$\text{【0439】 } O1D3IS = \left\lceil \frac{(I1D3IS - I2D4IS)}{D3S} \right\rceil$$

【0440】 其中：

【0441】 $O1D2IS$ 輸出張量之維度2索引大小。

【0442】 $O1D3IS$ 輸出張量之維度3索引大小。

【0443】 $I1D2IS$ 輸入1張量之維度2索引大小。

【0444】 $I1D3IS$ 輸入1張量之維度3索引大小。

【0445】 $I2D3IS$ 輸入2張量之維度3索引大小。

【0446】 $I2D4IS$ 輸入2張量之維度4索引大小。

【0447】 D2S 維度2步幅。

【0448】 D3S 維度3步幅。

【0449】 在一個實例中，忽略輸出張量描述符2及功能特定保存以及位址欄位。在一個實例中，功能特定參數5將含有零。

【0450】 **功能碼113：NNPA-MATMUL-OP (矩陣乘法運算)**

【0451】 在一個實例中，當指定NNPA-MATMUL-OP功能時，如下文所描述計算由輸出張量描述符所描述之輸出張量中的每一元素：

【0452】 * 使用下文所描述之獲得維度1向量(get-dimension-1-vector)操作自藉由輸入張量1描述符所描述之輸入張量1選擇維度1向量。

【0453】 * 使用下文所描述之獲得維度2向量操作自藉由輸入張量2描述符所描述之輸入張量2選擇維度2向量。

【0454】 * 使用下文所描述之點積運算來計算維度1向量及維度2向量之中間點積。

【0455】 * 對中間點積及由輸入張量3描述符所描述之輸入張量3之元素執行運算，其中維度索引4及維度索引1值與輸出張量元素相同。所得元素儲存於輸出張量中。藉由功能特定參數1判定且下文描述融合運算。

【0456】 獲得維度1向量操作：對於指定輸出元素，自輸入1張量選擇維度1向量，其中輸入維度4索引為輸出維度4索引，輸入維度3索引為輸出維度3索引，且輸入維度2索引為輸出維度2索引。

【0457】 獲得維度2向量操作：對於指定輸出元素，自輸入2張量選擇維度2向量，其中輸入維度4索引為輸出維度4索引，輸入維度3索引為輸出維度3索引，且輸入維度1索引為輸出維度1索引。

【0458】 點積運算：相同大小及資料類型之兩個向量的中間點積經

計算為輸入向量1中之每一元素與輸入向量2之對應元素的乘積之總和。

【0459】 融合運算：功能特定參數1控制對中間點積及來自輸入張量3之對應元素執行的運算。在一個實例中，NNPA-MATMUL-OP功能特定參數1包括例如位元24至31中之運算欄位。運算欄位指定經執行運算。下文指示實例運算：

【0460】 運算	運算類型
【0461】 0	加法
【0462】 1	比較點積是否為高的
【0463】 2	比較點積是否不低
【0464】 3	比較點積及元素是否相等
【0465】 4	比較點積及元素是否不相等
【0466】 5	比較點積是否不高
【0467】 6	比較點積是否為低的

【0468】 在一個實例中，對於加法運算類型，將輸入張量3元素加至中間點積。對於比較運算類型，將中間點積與輸入張量3元素進行比較，且若比較為真，則將結果設定為例如值+1；否則，在為輸出張量指定之資料類型中，將其設定為例如值+0。

【0469】 在一個實例中，保留運算(OPERATION)欄位之所有其他值。若為運算欄位指定保留值，則報告例如F000 hex之回應碼且操作以例如1之條件碼完成。

【0470】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例

如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0471】 在一個實施例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0472】 * 所有輸入張量及輸出張量之維度4索引大小將相同。

【0473】 * 所有輸入張量及輸出張量之維度3索引大小將等於一。

【0474】 * 輸入張量3之維度2索引大小將等於一。

【0475】 * 輸入張量1及輸出張量之維度2索引大小將相同。

【0476】 * 輸入張量1之維度1索引大小及輸入張量2之維度2索引大小將相同。

【0477】 * 輸入張量2、輸入張量3及輸出張量之維度1索引大小將相同。

【0478】 * 所有輸入張量及輸出張量之資料佈局及資料類型將相同。

【0479】 在一個實施例中，忽略輸出張量描述符2及功能特定保存區域位址欄位。在一實例中，功能特定參數2至5將含有零。

【0480】 **功能碼114：NNPA-MATMUL-OP-BCAST23 (矩陣乘法運算-廣播23)**

【0481】 在一個實例中，當指定NNPA-MATMUL-OP-BCAST23功能時，如下文所描述計算由輸出張量描述符所描述之輸出張量中的每一元素：

【0482】 * 使用下文所描述之獲得維度1向量操作自藉由輸入張量1描述符所描述之輸入張量1選擇維度1向量。

【0483】 * 使用下文所描述之獲得維度2向量操作自藉由輸入張量2描述符所描述之輸入張量2選擇維度2向量。

【0484】 * 使用下文所描述之點積運算來計算維度1向量及維度2向量之點積。

【0485】 * 將具有與輸出張量元素相同之維度索引1值的由輸入張量3描述符所描述之輸入張量3的元素加至先前計算之點積且儲存於輸出張量中。

【0486】 獲得維度1向量操作：對於指定輸出元素，自輸入1張量選擇維度1向量，其中輸入維度4索引為輸出維度4索引，輸入維度3索引為輸出維度3索引，且輸入維度2索引為輸出維度2索引。

【0487】 獲得維度2向量操作：對於指定輸出元素，自輸入2張量選擇維度2向量，其中輸入維度4索引為一，輸入維度3索引為輸出維度3索引，且輸入維度1索引為輸出維度1索引。

【0488】 點積運算：相同大小及資料類型之兩個向量的中間乘積經計算為輸入向量1中之每一元素與輸入向量2之對應元素的乘積之總和。

【0489】 在一個實例中，若指定張量描述符中之任一者中的指定資料佈局未指定4D特徵張量(例如，資料佈局=0)或若任何指定張量描述符中之資料類型未指定NNP資料類型1 (例如，資料類型=0)，則回應碼，例如0010 hex或0011 hex，分別設定於通用暫存器0中，且指令以例如1之條件碼完成。

【0490】 在一個實施例中，以下條件將為真，否則，辨識到一般運算元資料異常：

【0491】 * 輸入張量1及輸出張量之維度4索引大小將相同。

【0492】 * 輸入張量2及輸入張量3之維度4索引大小將等於一。

【0493】 * 所有輸入張量及輸出張量之維度3索引大小將等於一。

【0494】 * 輸入張量3之維度2索引大小將等於一。

【0495】 * 輸入張量1及輸出張量之維度2索引大小將相同。

【0496】 * 輸入張量1之維度1索引大小及輸入張量2之維度2索引大小將相同。

【0497】 * 輸入張量2、輸入張量3及輸出張量之維度1索引大小將相同。

【0498】 * 所有輸入張量及輸出張量之資料佈局及資料類型將相同。

【0499】 在一個實施例中，忽略輸出張量描述符2及功能特定保存區域位址欄位。在一個實例中，功能特定參數1至5將含有零。

【0500】 對於神經網路處理輔助指令，在一個實施例中，若輸出張量與任何輸入張量或參數區塊重疊，則結果不可預測。

【0501】 作為實例，當嘗試執行神經網路處理輔助指令且未在例如雙字邊界上指明參數區塊時，辨識到規格異常。

【0502】 當嘗試執行神經網路處理輔助指令且存在例如張量描述符不一致時，辨識到一般運算元資料異常。

【0503】 神經網路處理輔助指令之所得條件碼包括例如：0-正常完成；1-設定回應碼；2--；3-CPU判定之處理資料量。

【0504】 在一個實施例中，神經網路處理輔助指令之執行優先順序包括例如：

【0505】 1.-7. 優先順序與一般狀況之程式中斷條件之優先順序

相同的異常。

【0506】 8.A 由於指定未指派或未安功能碼之條件碼1。

【0507】 8.B 由於未在雙字邊界上指定參數區塊之規格異常。

【0508】 9. 存取參數區塊之存取異常。

【0509】 10. 由於模型不支援參數區塊之指定格式的條件碼1。

【0510】 11.A 由於不支援指定張量資料佈局之條件碼1。

【0511】 11.B 由於張量描述符之間的不同資料佈局之一般運算元資料異常。

【0512】 12.A 由於除包括於以上條項8.A、10及11.A以及以下12.B.1中之彼等條件以外的條件的條件碼1。

【0513】 12.B.1 由於NNPA-RELU及NNPA-CONVOLUTION之無效輸出張量資料類型的條件碼1。

【0514】 12.B.2 NNPA-RELU 功能特定參數1及NNPA-CONVOLUTION功能特定參數4之無效值的一般運算元資料異常。

【0515】 13.A 存取輸出張量之存取異常。

【0516】 13.B 存取輸入張量之存取異常。

【0517】 13.C 存取功能特定保存區域之存取異常。

【0518】 14. 條件碼0。

【0519】 如本文中所描述，單個指令(例如，神經網路處理輔助指令)經組態以執行複數個功能，包括一查詢功能及複數個非查詢功能。每一非查詢功能可在大量資料上操作，且因此，在處理階段中執行功能。在每一處理階段期間，高達選擇量之資料可經儲存至選擇位置(例如，記憶體)中。選擇量之資料取決於經執行功能。若在特定階段期間，判定發生

至指定區域之儲存(例如，發生至指定區域之實際儲存；參數區塊及/或功能特定保存區域(若經界定為由功能使用)與指定區域完全地重疊；參數區塊及/或功能保存區域(若經界定為由功能使用)與指定區域部分地重疊且存在於重疊部分中之儲存；及/或參數區塊及/或功能保存區域(若經界定為由功能使用)與指定區域部分地重疊，不管是否存在於重疊部分中之儲存)，中斷發生，且指令以例如3之條件碼完成。藉由檢查在處理階段之後是否發生了至指定區域中之儲存，僅在結束指令之前的單個階段內發生至指定區域中之儲存。由於在每一階段期間儲存選擇量之資料且選擇量可大於4K，因此測試是否存在至指定區域中之儲存(例如，是否處理多個連續4K頁面)可能需要縮減數目次檢查。舉例而言，不需要每一4K頁面之後的測試。作為一實例，神經網路處理輔助指令能夠在中斷指示事件之前將超過4K的額外位元組儲存至與指定程式事件記錄儲存區域相交之每一運算元位置。

【0520】 本發明之一或多個態樣不可避免地與電腦技術相關且促進電腦內之處理，從而改良其效能。經組態以執行各種功能之單個架構化機器指令的使用藉由降低複雜度、減少資源使用及提高處理速度來改良計算環境內之效能。此外，藉由增強程式事件記錄設施，促進程式之除錯，且降低程式複雜性及/或程式化錯誤，從而改良系統效能。作為實例，藉由以下操作來增強程式事件記錄：允許指令在中斷指示事件之前將超過4K的額外位元組儲存至與指定程式事件記錄儲存區域相交之每一運算元位置；藉由僅在中斷之前的單個處理階段內儲存至指定區域中；及/或藉由縮減程式事件記錄檢查之數目，此係由於可在檢查之前儲存超過4K頁面。此外，在一或多個實施例中，降低例如驗證程式之複雜性，此可預測

儲存改變中斷，從而增加驗證品質。舉例而言，基於對應於功能碼之記憶體存取模式，驗證程式可準確地預測指令將在某一階段之後呈現程式事件記錄中斷；且驗證程式化可在程式事件記錄指定區域位於參數區塊及/或功能特定保存區域內之狀況下預測中斷，此係因為中斷將在每一階段之後呈現。程式可利用程式事件記錄儲存改變以識別例如非既定記憶體更新。經獲得資訊、程式事件記錄設施及/或指令可用於許多技術領域中，諸如電腦處理、醫療處理、工程化、汽車技術、製造等。藉由在獲得資訊及/或增強程式事件記錄方面提供最佳化，藉由例如縮減錯誤及/或執行時間來改良此等技術領域。

【0521】參看圖7A及圖7B描述促進計算環境內之處理的一個實施例之其他細節，此係因為該計算環境與本發明之一或多個態樣相關。

【0522】參看圖7A，在一個實施例中，針對指令執行指令處理700。該指令經組態以執行複數個功能702。複數個功能中之功能將在複數個處理階段中執行704，其中複數個處理階段中之處理階段經界定以儲存高達選擇量之資料706。選擇量之資料係基於複數個功能中之待執行之功能708。複數個功能中之至少一個功能具有與複數個功能中之至少一個其他功能不同的用於選擇量之資料的值710。基於處理選擇功能之複數個處理階段中之一選擇處理階段來判定是否發生至指定區域中之儲存712。基於判定發生了至指定區域中之儲存，呈現中斷714，且基於判定未發生至指定區域中之儲存，執行指令處理繼續716。

【0523】經組態以執行各種功能之單個架構化機器指令的使用藉由降低複雜度、縮減資源使用及/或提高處理速度來改良計算環境內之效能。

【0524】藉由檢查在處理階段之後是否發生了至指定區域中之儲存，僅在結束指令之前的單個階段內發生至指定區域中之儲存。由於在每一階段期間儲存選擇量之資料且選擇量可大於4K，因此測試是否存在至指定區域中之儲存(例如，是否處理多個連續4K頁面)可能需要縮減數目次檢查。

【0525】在一個實例中，繼續執行指令處理包括進行至選擇功能之下一處理階段之處理720。作為另一實例，繼續執行指令處理包括結束該指令722。

【0526】作為一實例，指定區域包括由開始位址及結束位址界定之記憶體之區域724。開始位址及結束位址係由例如一或多個控制暫存器提供726。

【0527】在一個實施例中，參看圖7B，判定是否發生至指定區域中之儲存包括檢查資料經寫入至之記憶體位址是否處於指定區域中730，且基於記憶體位址處於指定區域中，該判定指示存在至指定區域中之儲存732。

【0528】在一個實施例中，判定是否發生至指定區域中之儲存包括檢查待由選擇功能使用之參數區塊之至少一部分是否與指定區域重疊734，且基於與指定區域重疊之參數區塊之至少一部分，該判定指示存在至指定區域中之儲存736。

【0529】用於參數區塊之重疊的檢查可在每一階段之後產生中斷，從而容易地提供關於記憶體存取之資訊。

【0530】在一個實施例中，判定是否發生至指定區域中之儲存包括檢查待由功能使用之功能特定保存區域之至少一部分是否與指定區域重疊

738，且基於與指定區域重疊之功能特定保存區域之至少一部分，該判定指示存在至指定區域中之儲存740。

【0531】若要使用針對功能特定保存區域之重疊的檢查，則可在每一階段之後產生中斷，從而容易地提供關於記憶體存取的資訊。

【0532】在一或多個態樣中，可進行某些預測。舉例而言，經預測，選擇功能之複數個處理階段中之特定處理階段將產生中斷750，其中該預測係基於經執行之選擇功能之記憶體存取模式752。作為另一實例，經預測，將基於判定待由選擇功能使用之參數區塊之至少一部分與指定區域重疊而發生中斷754，及/或經預測，將基於判定待由選擇功能使用之功能特定保存區域之至少一部分與指定區域重疊而發生中斷756。

【0533】至少，驗證程式的複雜性降低了，此可預測儲存改變中斷，從而提高驗證品質。舉例而言，基於對應於功能碼之記憶體存取模式，驗證程式可準確地預測指令將在某一階段之後中斷且呈現程式事件記錄中斷；且驗證程式化可在程式事件記錄指定區域位於參數區塊及/或功能特定保存區域內之狀況下預測中斷，此係因為中斷將在每一階段之後呈現。

【0534】其他變化及實施例為可能的。

【0535】本發明之態樣可由許多類型之計算環境使用。參考圖8A描述併有及使用本發明之一或多個態樣的計算環境之另一實例。作為實例，圖8A之計算環境係基於由紐約阿蒙克市之國際商業機器公司供應之z/Architecture®指令集架構。然而，z/Architecture指令集架構僅為一個實例架構。再次，計算環境可基於其他架構，包括但不限於Intel® x86架構、國際商業機器公司之其他架構及/或其他公司之架構。Intel為英特爾

公司(Intel Corporation)或其子公司在美國及其他國家之商標或註冊商標。

【0536】 在一個實例中，計算環境10包括中央電子複合體(CEC) 11。中央電子複合體11包括複數個組件，諸如記憶體12 (亦稱為系統記憶體、主記憶體、主儲存器、中央儲存器、儲存器)，該記憶體耦接至一或多個處理器，諸如一或多個通用處理器(亦稱為中央處理單元(CPU) 13)及一或多個專用處理器(例如，神經網路處理器31)，且耦接至輸入/輸出(I/O)子系統14。

【0537】 作為實例，一或多個專用處理器可與一或多個通用處理器分離及/或至少一個專用處理器可嵌入於至少一個通用處理器內。其他變化亦係可能的。

【0538】 I/O子系統14可為中央電子複合體之部分或與其分開。其導引主儲存器12與耦接至中央電子複合體之輸入/輸出控制單元15及輸入/輸出(I/O)裝置16之間的資訊流。

【0539】 可使用許多類型之I/O裝置。一個特定類型為資料儲存裝置17。資料儲存裝置17可儲存一或多個程式18、一或多個電腦可讀程式指令19，及/或資料等等。電腦可讀程式指令可經組態以進行本發明之態樣之實施例的功能。

【0540】 中央電子複合體11可包括及/或耦接至抽取式/非抽取式、揮發性/非揮發性電腦系統儲存媒體。舉例而言，其可包括及/或耦接至非抽取式非揮發性磁性媒體(通常被稱作「硬碟機」)、用於自抽取式非揮發性磁碟(例如，「軟碟」)讀取及寫入至抽取式非揮發性磁碟之磁碟機，及/或用於自諸如CD-ROM、DVD-ROM或其他光學媒體之抽取式非揮發性光

碟讀取或寫入至抽取式非揮發性光碟之光碟機。應理解，可結合中央電子複合體11使用其他硬體及/或軟體組件。實例包括但不限於：微碼或毫碼、裝置驅動程式、冗餘處理單元、外部磁碟機陣列、RAID系統、磁帶機及資料存檔儲存系統等。

【0541】 此外，中央電子複合體11可與眾多其他通用或專用計算系統環境或組態一起操作。可適合與中央電子複合體11一起使用之熟知計算系統、環境及/或組態之實例包括但不限於：個人電腦(PC)系統、伺服器電腦系統、精簡型用戶端、複雜型用戶端、手持型或膝上型電腦裝置、多處理器系統、基於微處理器之系統、機上盒、可程式化消費型電子裝置、網路PC、小型電腦系統、大型電腦系統及包括以上系統或裝置中之任一者的分散式雲端計算環境，以及其類似者。

【0542】 在一或多個實施例中，中央電子複合體11提供邏輯分割及/或虛擬化支援。在一個實施例中，如圖8B中所展示，記憶體12包括例如一或多個邏輯分割區20、管理邏輯分割區之超管理器21，及處理器韌體22。超管理器21之一個實例為由紐約阿蒙克市之國際商業機器公司提供的處理器資源/系統管理器(PR/SM™)。PR/SM為國際商業機器公司在至少一個管轄區域中之商標或註冊商標。

【0543】 每一邏輯分割區20能夠充當單獨的系統。亦即，每一邏輯分割區可獨立地重設、運行客體作業系統23 (諸如由紐約阿蒙克市的國際商業機器公司提供的z/OS®作業系統)或其他控制碼24 (諸如耦接設施控制碼(CFCC))，且用不同程式25操作。在邏輯分割區中運行之作業系統或應用程式呈現為能夠存取完整的系統，但實際上，僅其一部分可用。儘管供應z/OS作業系統作為一實例，但可根據本發明之一或多個態樣使用由國

際商業機器公司及/或其他公司供應之其他作業系統。

【0544】 記憶體12耦接至例如CPU 13 (圖8A)，其為可分配至邏輯分割區之實體處理器資源。舉例而言，邏輯分割區20可包括一或多個邏輯處理器，其中之每一者表示可動態地分配至邏輯分割區之實體處理器資源13中的全部或一部分。

【0545】 在又一實施例中，中央電子複合體提供虛擬機支援(具有或不具有邏輯分割支援)。如圖8C中所展示，中央電子複合體11之記憶體12包括例如一或多個虛擬機26、管理虛擬機之諸如超管理器27的虛擬機管理器，及處理器韌體28。超管理器27之一個實例為由紐約阿蒙克市之國際商業機器公司提供的z/VM[®]超管理器。超管理器有時稱為主機。z/VM為國際商業機器公司在至少一個管轄區域中之商標或註冊商標。

【0546】 中央電子複合體之虛擬機支援提供操作大量虛擬機26之能力，該等虛擬機各自能夠用不同程式29操作且運行客體作業系統30，諸如Linux[®]作業系統。每一虛擬機26能夠充當單獨的系統。亦即，每一虛擬機可獨立地進行重設，運行客體作業系統，且藉由不同程式操作。在虛擬機中運行之作業系統或應用程式呈現為能夠存取完整系統，但實際上，僅其一部分可用。儘管供應z/VM及Linux作為實例，但可根據本發明之一或多個態樣使用其他虛擬機管理器及/或作業系統。註冊商標Linux[®]係依照Linux基金會(Linux Foundation)的分許可而使用，Linux基金會為該商標在全球範圍內的所有者Linus Torvalds的獨家被授權人。

【0547】 參考圖9A描述併有及使用本發明之一或多個態樣的計算環境之另一實施例。在此實例中，計算環境36包括例如原生中央處理單元(CPU) 37、記憶體38及一或多個輸入/輸出裝置及/或介面39，前述各者經

由例如一或多個匯流排40及/或其他連接而彼此耦接。作為實例，計算環境36可包括：由紐約阿蒙克市之國際商業機器公司提供之PowerPC[®]處理器；由加州帕洛阿爾托(Palo Alto, California)之惠普公司(Hewlett Packard Co.)提供的具有Intel[®] Itanium[®] II處理器之HP Superdome；及/或基於由國際商業機器公司、惠普公司、英特爾公司(Intel Corporation)、甲骨文公司(Oracle)及/或其他公司提供之架構的其他機器。PowerPC為國際商業機器公司在至少一個管轄區域中之商標或註冊商標。Itanium為英特爾公司或其子公司在美國及其他國家之商標或註冊商標。

【0548】 原生中央處理單元37包括一或多個原生暫存器41，諸如在環境內之處理期間使用的一或多個通用暫存器及/或一或多個專用暫存器。此等暫存器包括表示在任何特定時間點處之環境狀態之資訊。

【0549】 此外，原生中央處理單元37執行儲存於記憶體38中之指令及程式碼。在一個特定實例中，中央處理單元執行儲存於記憶體38中之仿真器程式碼42。此程式碼使得在一個架構中組態之計算環境能夠仿真另一架構。舉例而言，仿真器程式碼42允許基於除z/Architecture指令集架構以外之架構的機器，諸如PowerPC處理器、HP Superdome伺服器或其他者，能夠仿真z/Architecture指令集架構且執行基於z/Architecture指令集架構開發之軟體及指令。

【0550】 參考9B描述與仿真器程式碼42有關之其他細節。儲存於記憶體38中之客體指令43包含經開發以在除原生CPU 37之架構以外的架構中執行之軟體指令(例如，與機器指令相關)。舉例而言，客體指令43可能已經設計以在基於z/Architecture指令集架構之處理器上執行，但替代地，在可為例如Intel Itanium II處理器之原生CPU 37上仿真。在一個實

例中，仿真器程式碼42包括指令提取常式44以自記憶體38獲得一或多個客體指令43，且視情況提供對所獲得指令之本端緩衝。其亦包括指令轉譯常式45以判定已經獲得的客體指令之類型且將客體指令轉譯成一或多個對應的原生指令46。此轉譯包括例如識別待由客體指令執行之功能及挑選原生指令執行彼功能。

【0551】此外，仿真器程式碼42包括仿真控制常式47以使得執行原生指令。仿真控制常式47可使原生CPU 37執行仿真一或多個先前所獲得之客體指令之原生指令的常式且在此執行完結時，將控制傳回至指令提取常式以仿真獲得下一客體指令或一組客體指令。原生指令46之執行可包括將資料自記憶體38加載至暫存器中；將資料自暫存器儲存回至記憶體；或執行某一類型之算術或邏輯運算(如藉由轉譯常式判定)。

【0552】每一常式例如實施於軟體中，該軟體儲存於記憶體中且藉由原生中央處理單元37執行。在其他實例中，常式或操作中之一或多者實施於韌體、硬體、軟體或其某一組合中。可使用原生CPU之暫存器41或藉由使用記憶體38中之位置來仿真所仿真處理器之暫存器。在實施例中，客體指令43、原生指令46及仿真器程式碼42可駐留於同一記憶體中或可分配於不同記憶體裝置當中。

【0553】根據本發明之一或多個態樣，可仿真之指令包括本文中所描述的神經網路輔助處理指令。此外，根據本發明之一或多個態樣，可仿真其他指令及/或程式事件記錄設施之一或多個態樣。

【0554】上文所描述之計算環境僅為可使用之計算環境的實例。可使用其他環境，包括但不限於未經分割之環境、經分割之環境、雲端環境及/或仿真環境；實施例不限於任一種環境。儘管本文中描述計算環境之

各種實例，但本發明之一或多個態樣可與許多類型之環境一起使用。本文中所提供之計算環境僅為實例。

【0555】 每一計算環境能夠經組態以包括本發明之一或多個態樣。

【0556】 一或多個態樣可係關於雲端計算。

【0557】 應理解，儘管本發明包括關於雲端計算之詳細描述，但本文中所敘述之教示的實施不限於雲端計算環境。實情為，本發明之實施例能夠結合現在已知或稍後開發之任何其他類型之計算環境來實施。

【0558】 雲端計算為用於實現對可快速佈建並以最小管理工作釋放或與服務之提供者互動的可組態計算資源(例如，網路、網路頻寬、伺服器、處理、記憶體、儲存器、應用程式、虛擬機及服務)之共用集區的便利隨選網路存取的服務遞送之模型。此雲端模型可包括至少五個特性、至少三個服務模型及至少四個部署模型。

【0559】 特性如下：

【0560】 按需自助服務：雲端消費者可視需要自動地單向布建計算能力(諸如，伺服器時間及網路儲存器)，而無需與服務提供者之人為互動。

【0561】 寬頻網路存取：可經由網路獲得能力及經由標準機制存取能力，該等標準機制藉由異質精簡型或複雜型用戶端平台(例如，行動電話、膝上型電腦及PDA)促進使用。

【0562】 資源集用：提供者之計算資源經集用以使用多租戶模型為多個消費者服務，其中根據需要動態指派及重新指派不同實體及虛擬資源。存在位置獨立性之意義，此係因為消費者通常不具有對所提供之資源之確切位置的控制或瞭解，但可能能夠按較高抽象層級(例如，國家、州

或資料中心)指定位置。

【0563】 快速彈性：可快速地且彈性地佈建能力，在一些狀況下自動地佈建能力，以迅速地向外延展，且可快速地釋放能力以迅速地向內延展。在消費者看來，可用於佈建之能力常常看起來為無限的且可在任何時間以任何量來購買。

【0564】 所量測服務：雲端系統藉由在適於服務類型(例如，儲存、處理、頻寬及作用中使用者帳戶)之某一抽象層級下充分利用計量能力而自動控制及最佳化資源使用。可監視、控制及報告資源使用，從而為所利用服務之提供者及消費者兩者提供透明度。

【0565】 服務模型如下：

【0566】 軟體即服務(SaaS)：提供給消費者之能力係使用在雲端基礎架構上運行之提供者之應用程式。可經由諸如網頁瀏覽器(例如，基於網頁之電子郵件)之精簡型用戶端介面自各種用戶端裝置獲取應用程式。消費者並不管理或控制包括網路、伺服器、作業系統、儲存器或甚至個別應用程式能力之基礎雲端基礎架構，其中可能的異常為有限的使用者特定應用程式組態設定。

【0567】 平台即服務(PaaS)：提供給消費者之能力係將使用由提供者所支援之程式化語言及工具建立的消費者建立或獲取之應用程式部署至雲端基礎架構上。消費者並不管理或控制包括網路、伺服器、作業系統或儲存器之底層雲端基礎架構，但控制所部署之應用程式及可能的代管環境組態之應用程式。

【0568】 基礎架構即服務(IaaS)：提供給消費者之能力係佈建處理、儲存、網路及其他基礎計算資源，其中消費者能夠部署及運行可包括

作業系統及應用程式之任意軟體。消費者並不管理或控制基礎雲端基礎架構，但具有對作業系統、儲存器、所部署應用程式之控制，及可能的對選擇網路連接組件(例如，主機防火牆)之有限控制。

【0569】 部署模型如下：

【0570】 私用雲端：僅針對組織操作雲端基礎架構。私用雲端可由組織或第三方來管理且可存在於內部部署或外部部署。

【0571】 社群雲端：該雲端基礎架構由若干組織共用且支援具有共用關注事項(例如，任務、安全要求、策略及合規性考量)之特定社群。社群雲端可由組織或第三方來管理且可存在於內部部署或外部部署。

【0572】 公用雲端：可使得雲端基礎架構可用於公眾或大型工業集團且為出售雲端服務之組織所擁有。

【0573】 混合雲端：該雲端基礎架構為兩個或多於兩個雲端(私用、社群或公用)之組合物，該等雲端保持獨特實體但藉由實現資料及應用程式攜帶性(例如，用於在雲端之間實現負載平衡之雲端爆裂)之標準化或專屬技術繫結在一起。

【0574】 藉由集中於無狀態性、低耦合、模組化及語義互操作性對雲端計算環境進行服務定向。雲端計算之關鍵為包括互連節點之網路的基礎架構。

【0575】 現參考圖10，描繪說明性雲端計算環境50。如所展示，雲端計算環境50包括一或多個雲端計算節點52，雲端消費者所使用之諸如個人數位助理(PDA)或蜂巢式電話54A、桌上型電腦54B、膝上型電腦54C及/或汽車電腦系統54N的本端計算裝置可與該一或多個雲端計算節點進行通信。節點52可彼此通信。可在一或多個網路(諸如，如上文所描述

之私用、社群、公用或混合雲端或其組合)中將該等節點實體地或虛擬地分組(未圖示)。此情形允許雲端計算環境50提供基礎架構、平台及/或軟體作為服務，針對該等服務，雲端消費者不需要在本端計算裝置上維護資源。應理解，圖10中所展示之計算裝置54A至54N之類型意欲僅為說明性的，且計算節點52及雲端計算環境50可經由任何類型之網路及/或網路可定址連接(例如，使用網頁瀏覽器)與任何類型之電腦化裝置通信。

【0576】 現參考圖11，展示由雲端計算環境50 (圖10)提供之一組功能抽象層。事先應理解，圖11中所展示之組件、層及功能意欲僅為說明性的且本發明之實施例不限於此。如所描繪，提供以下層及對應功能：

【0577】 硬體及軟體層60包括硬體及軟體組件。硬體組件之實例包括：大型電腦61；基於精簡指令集電腦(IRISC)架構之伺服器62；伺服器63；刀片伺服器64；儲存裝置65；以及網路及網路連接組件66。在一些實施例中，軟體組件包括網路應用程式伺服器軟體67及資料庫軟體68。

【0578】 虛擬化層70提供抽象層，可自該抽象層提供虛擬實體之以下實例：虛擬伺服器71；虛擬儲存器72；虛擬網路73，包括虛擬私用網路；虛擬應用程式及作業系統74；及虛擬用戶端75。

【0579】 在一個實例中，管理層80可提供下文所描述之功能。資源佈建81提供計算資源及用以執行雲端計算環境內之任務之其他資源的動態採購。當在雲端計算環境內利用資源時，計量及定價82提供成本追蹤，及對此等資源之消耗之帳務處理及發票開立。在一個實例中，此等資源可包括應用程式軟體授權。安全性為雲端消費者及任務提供身分驗證，以及對資料及其他資源之保護。使用者入口網站83為消費者及系統管理者提供對雲端計算環境之存取。服務等級管理84提供雲端計算資源分配及管理使得

滿足所需服務等級。服務等級協議(SLA)規劃及實現85提供雲端計算資源之預先配置及採購，針對雲端計算資源之未來要求係根據SLA來預期。

【0580】 工作負載層90提供功能性之實例，可針對該功能性利用雲端計算環境。可自此層提供之工作負載及功能的實例包括：地圖測繪及導航91；軟體開發及生命週期管理92；虛擬教室教育遞送93；資料分析處理94；異動處理95；及程式事件記錄及/或神經網路處理輔助處理96。

【0581】 本發明之態樣可為在任何可能之技術細節整合層級處的系統、方法及/或電腦程式產品。電腦程式產品可包括電腦可讀儲存媒體(或多個媒體)，其上具有電腦可讀程式指令以使得處理器實行本發明之態樣。

【0582】 電腦可讀儲存媒體可為有形裝置，其可保持及儲存指令以供指令執行裝置使用。電腦可讀儲存媒體可為(例如但不限於)：電子儲存裝置、磁性儲存裝置、光學儲存裝置、電磁儲存裝置、半導體儲存裝置或前述各者之任何合適組合。電腦可讀儲存媒體之更具體實例之非窮盡性清單包括以下各者：攜帶型電腦磁片、硬碟、隨機存取記憶體(RAM)、唯讀記憶體(ROM)、可擦除可程式化唯讀記憶體(EPROM或快閃記憶體)、靜態隨機存取記憶體(SRAM)、攜帶型緊密光碟唯讀記憶體(CD-ROM)、數位通用光碟(DVD)、記憶卡、軟性磁碟、機械編碼裝置(諸如其上記錄有指令之凹槽中之打孔卡片或凸起結構)及前述各者之任何合適組合。如本文中所使用，不將電腦可讀儲存媒體本身理解為暫時信號，諸如無線電波或其他自由傳播之電磁波、經由波導或其他傳輸媒體傳播之電磁波(例如，經由光纖電纜傳遞之光脈衝)，或經由導線傳輸之電信號。

【0583】 本文中所描述之電腦可讀程式指令可自電腦可讀儲存媒體

下載至各別計算/處理裝置或經由網路(例如，網際網路、區域網路、廣域網路及/或無線網路)下載至外部電腦或外部儲存裝置。網路可包含銅傳輸纜線、光傳輸光纖、無線傳輸、路由器、防火牆、交換器、閘道器電腦及/或邊緣伺服器。每一計算/處理裝置中之網路配接卡或網路介面自網路接收電腦可讀程式指令且轉遞電腦可讀程式指令以用於儲存於各別計算/處理裝置內之電腦可讀儲存媒體中。

【0584】 用於實行本發明之操作的電腦可讀程式指令可為以一或多個程式設計語言之任何組合撰寫的組譯器指令、指令集架構(ISA)指令、機器指令、機器相關指令、微碼、韌體指令、狀態設定資料、積體電路系統之組態資料或原始碼或目標碼，該一或多個程式設計語言包括諸如 Smalltalk、C++或類似者的物件導向式程式設計語言，及諸如「C」程式設計語言的程序程式設計語言或類似程式設計語言。電腦可讀程式指令可完全在使用者之電腦上、部分在使用者之電腦上、作為獨立套裝軟體、部分在使用者之電腦上且部分在遠端電腦上或完全在遠端電腦或伺服器上而執行。在後一情境下，遠端電腦可經由任何類型之網路連接至使用者電腦，網路類型包括區域網路(LAN)或廣域網路(WAN)，或可連接至外部電腦(例如，經由網際網路使用網際網路服務提供者)。在一些實施例中，電子電路系統(包括例如可程式化邏輯電路系統、場可程式化閘陣列(FPGA)或可程式化邏輯陣列(PLA))可藉由利用電腦可讀程式指令之狀態資訊來個人化電子電路系統而執行電腦可讀程式指令，以便執行本發明之態樣。

【0585】 本文參考根據本發明之實施例之方法、設備(系統)及電腦程式產品之流程圖說明及/或方塊圖來描述本發明之態樣。應理解，可藉由電腦可讀程式指令實施流程圖說明及/或方塊圖中之每一區塊，及流程

圖說明及/或方塊圖中的區塊之組合。

【0586】 可將此等電腦可讀程式指令提供至電腦或其他可程式化資料處理設備之處理器以產生一機器，以使得經由該電腦或其他可程式化資料處理設備之處理器執行之指令建立用於實施一或多個流程圖及/或方塊圖區塊中所指定之功能/動作之手段。亦可將此等電腦可讀程式指令儲存於電腦可讀儲存媒體中，其可指導電腦、可程式化資料處理設備及/或其他裝置以特定方式起作用，使得儲存有指令之電腦可讀儲存媒體包含製品，該製品包括實施在該一或多個流程圖及/或方塊圖區塊中指定之功能/動作之態樣。

【0587】 電腦可讀程式指令亦可載入至電腦、其他可程式化資料處理設備或其他裝置上，以使一系列操作步驟在該電腦、其他可程式化設備或其他裝置上執行以產生電腦實施程序，使得在該電腦、其他可程式化設備或其他裝置上執行之指令實施該一或多個流程圖及/或方塊圖區塊中所指定之功能/動作。

【0588】 諸圖中之流程圖及方塊圖說明根據本發明之各種實施例之系統、方法及電腦程式產品之可能實施之架構、功能性及操作。就此而言，流程圖或方塊圖中之每一區塊可表示指令之模組、區段或部分，其包含用於實施經指定邏輯功能之一或多個可執行指令。在一些替代實施中，區塊中所指出的功能可不按圖式中所指出的次序發生。舉例而言，取決於所涉及之功能性，連續展示之兩個區塊實際上可實現為一個步驟，同時、實質上同時、以部分或完全在時間上重疊之方式執行，或該等區塊有時可以相反次序執行。亦將注意，可藉由執行指定功能或動作或實行專用硬體及電腦指令之組合的基於專用硬體之系統實施方塊圖及/或流程圖說明之

每一區塊，及方塊圖及/或流程圖說明中之區塊之組合。

【0589】 除上述以外，可藉由供應客戶環境之管理之服務提供者提供、供應、部署、管理、服務一或多個態樣等。舉例而言，服務提供者可建立、維持、支援(等)電腦程式碼及/或執行用於一或多個客戶之一或多個態樣的電腦基礎架構。作為回報，服務提供者可根據訂用及/或收費協議接收來自客戶之付款(作為實例)。另外或替代地，服務提供者可接收來自向一或多個第三方出售廣告內容之付款。

【0590】 在一個態樣中，可部署一應用程式用於執行一或多個實施例。作為一個實例，應用程式之部署包含提供可操作以執行一或多個實施例之電腦基礎架構。

【0591】 作為另一態樣，可部署計算基礎架構，包含將電腦可讀程式碼整合至計算系統中，其中程式碼結合計算系統能夠執行一或多個實施例。

【0592】 作為又一態樣，可提供一種用於整合計算基礎架構之程序，包含將電腦可讀程式碼整合至電腦系統中。電腦系統包含電腦可讀媒體，其中電腦媒體包含一或多個實施例。程式碼結合電腦系統能夠執行一或多個實施例。

【0593】 儘管上文描述各種實施例，但其僅為實例。舉例而言，其他架構之計算環境可用以併有及/或使用一或多個態樣。此外，可使用不同指令或操作。另外，可使用不同類型之暫存器及/或不同暫存器。此外，可支援其他資料格式、資料佈局及/或資料大小。在一或多個實施例中，可使用一或多個通用處理器、一或多個專用處理器或通用處理器與專用處理器之組合。許多變化係可能的。

【0594】本文中描述各種態樣。此外，在不脫離本發明之態樣之精神的情況下，許多變化係可能的。應注意，除非不一致，否則本文所描述之每一態樣或特徵及其變體可與任何其他態樣或特徵組合。

【0595】此外，其他類型之計算環境可係有益的且可被使用。作為一實例，可使用適合於儲存及/或執行程式碼之資料處理系統，其包括直接或經由系統匯流排間接地耦接至記憶體元件之至少兩個處理器。記憶體元件包括(例如)在實際執行程式碼期間使用之本端記憶體、大容量儲存器，及提供至少某一程式碼之臨時儲存以便減少在執行期間必須自大容量儲存器擷取程式碼之次數的快取記憶體。

【0596】輸入/輸出或I/O裝置(包括但不限於鍵盤、顯示器、指標裝置、DASD、磁帶、CD、DVD、隨身碟(Thumb Drive)及其他記憶體媒體等)可直接或經由介入之I/O控制器耦接至系統。網路配接器亦可耦接至系統以使得資料處理系統能夠變成經由介入的私人網路或公用網路耦接至其他資料處理系統或遠端印表機或儲存裝置。數據機、纜線數據機及乙太網卡僅為幾個可用類型之網路配接器。

【0597】本文中所使用之術語僅出於描述特定實施例之目的且並不意欲為限制性的。如本文中所使用，除非上下文另外明確指示，否則單數形式「一(/an)及「該」意欲亦包括複數形式。應進一步理解，術語「包含(comprises及/或comprising)」在用於本說明書中時指定所陳述特徵、整數、步驟、操作、元件及/或組件之存在，但不排除一或多個其他特徵、整數、步驟、操作、元件、組件及/或其群組之存在或添加。

【0598】以下申請專利範圍中之所有構件或步驟加功能元件之對應結構、材料、動作及等效物(若存在)意欲包括用於結合如特定主張之其他

所主張元件來執行功能的任何結構、材料或動作。已出於說明及描述之目的呈現一或多個實施例之描述，但其不意欲為窮盡性的或限於所揭示之形式。對於一般熟習此項技術者而言，許多修改及變化將為顯而易見的。選取及描述實施例以便最佳地解釋各種態樣及實際應用，且使得一般熟習此項技術者能夠理解具有如適於所預期之特定用途之各種修改的各種實施例。

【符號說明】

【0599】

- 0: 通用暫存器
- 1: 通用暫存器
- 10: 計算環境
- 11: 中央電子複合體(CEC)
- 12: 記憶體
- 13: 中央處理單元(CPU)
- 14: 輸入/輸出(I/O)子系統
- 15: 輸入/輸出控制單元
- 16: 輸入/輸出(I/O)裝置
- 17: 資料儲存裝置
- 18: 程式
- 19: 電腦可讀程式指令
- 20: 邏輯分割區
- 21: 超管理器
- 22: 處理器韌體

- 23: 客體作業系統
- 24: 控制碼
- 25: 不同程式
- 26: 虛擬機
- 27: 超管理器
- 28: 處理器韌體
- 29: 不同程式
- 30: 客體作業系統
- 31: 神經網路處理器
- 34: 位元
- 36: 計算環境
- 37: 原生中央處理單元(CPU)
- 38: 記憶體
- 39: 輸入/輸出裝置及/或介面
- 40: 匯流排
- 41: 原生暫存器
- 42: 位元
- 43: 客體指令
- 44: 指令提取常式
- 45: 指令轉譯常式
- 46: 原生指令
- 47: 仿真控制常式
- 50: 說明性雲端計算環境

- 52: 雲端計算節點
- 54A: 個人數位助理(PDA)或蜂巢式電話
- 54B: 桌上型電腦
- 54C: 膝上型電腦
- 54N: 汽車電腦系統
- 56: 位元
- 60: 硬體及軟體層
- 61: 大型電腦
- 62: 基於精簡指令集電腦(IRISC)架構之伺服器
- 63: 伺服器
- 64: 刀片伺服器
- 65: 儲存裝置
- 66: 網路及網路連接組件
- 67: 網路應用程式伺服器軟體
- 68: 資料庫軟體
- 70: 虛擬化層
- 71: 虛擬伺服器
- 72: 虛擬儲存器
- 73: 虛擬網路
- 74: 虛擬應用程式及作業系統
- 75: 虛擬用戶端
- 80: 管理層
- 81: 資源佈建

- 82: 計量及定價
- 83: 使用者入口網站
- 84: 服務等級管理
- 85: 服務等級協定(SLA)規劃及實現
- 90: 工作負載層
- 91: 地圖測繪及導航
- 92: 軟體開發及生命週期管理
- 93: 虛擬教室教育遞送
- 94: 資料分析處理
- 95: 異動處理
- 96: 神經網路處理輔助處理
- 100: 計算環境
- 102: 電腦系統
- 104: 通用處理器
- 105: 神經網路處理器
- 106: 記憶體
- 108: 輸入/輸出(I/O)介面
- 110: 匯流排
- 111: 匯流排
- 112: 快取記憶體
- 114: 本端快取記憶體
- 116: 程式或應用程式
- 118: 至少一個作業系統

120: 電腦可讀程式指令

122: 韌體

130: 外部裝置

132: 網路介面

134: 資料儲存裝置

136: 程式

138: 電腦可讀程式指令

150: 指令提取組件

152: 指令解碼單元

154: 指令執行組件

156: 記憶體存取組件

158: 寫回組件

160: 暫存器

170: 程式事件記錄組件

172: 神經網路處理輔助組件

200: 步驟

202: 步驟

204: 步驟

206: 步驟

208: 步驟

210: 步驟

250: 步驟

252: 步驟

- 300: 神經網路處理輔助指令
- 302: 作業碼(opcode)欄位
- 310: 回應碼欄位
- 312: 異常旗標欄位
- 314: 功能碼欄位
- 320: 參數區塊
- 330: NNPA查詢可用功能參數區塊
- 332: 已安裝功能向量
- 334: 已安裝參數區塊格式向量
- 336: 已安裝資料類型
- 338: 已安裝資料佈局格式
- 340: 最大維度索引大小
- 342: 最大張量大小
- 344: 已安裝NNP資料類型1轉換向量
- 350: 參數區塊
- 352: 參數區塊版本號碼
- 354: 模型版本號碼
- 356: 接續旗標
- 358: 功能特定保存區域位址
- 360: 輸出張量描述符
- 365: 輸入張量描述符
- 370: 功能特定參數
- 375: 接續狀態緩衝器欄位

382: 資料佈局格式

384: 資料類型

386: 維度1至4索引大小

388: 張量位址

400: 格式

402: 正負號

404: 指數+31

406: 分數

500: 3D張量

502: 2D張量

600: 記憶體單元

602: 預選數目個列

604: 預選數目個元素

606: 步驟

608: 步驟

700: 步驟

702: 步驟

704: 步驟

706: 步驟

708: 步驟

710: 步驟

712: 步驟

714: 步驟

- 716: 步驟
- 720: 步驟
- 722: 步驟
- 724: 步驟
- 726: 步驟
- 730: 步驟
- 732: 步驟
- 734: 步驟
- 736: 步驟
- 738: 步驟
- 740: 步驟
- 750: 步驟
- 752: 步驟
- 754: 步驟
- 756: 步驟
- E1: 維度
- E2: 維度
- E3: 維度
- E4: 維度

【發明申請專利範圍】

【請求項1】

一種用於促進一計算環境內之處理的電腦程式產品，該電腦程式產品包含：

一或多個電腦可讀儲存媒體及共同地儲存於該一或多個電腦可讀儲存媒體上以執行一方法之程式指令，該方法包含：

針對一指令執行指令處理，該指令經組態以執行複數個功能，其中該複數個功能中之一功能將在複數個處理階段中執行，其中該複數個處理階段中之一處理階段經界定以儲存高達一選擇量之資料，該選擇量之資料係基於該複數個功能中之待執行之該功能，其中該複數個功能中之至少一個功能具有與該複數個功能中之至少一個其他功能不同的一用於該選擇量之資料的值；

預測該選擇功能之該複數個處理階段中之一特定處理階段將產生一中斷，該預測係基於該選擇功能之一記憶體存取模式；

基於處理一選擇功能之該複數個處理階段中之一選擇處理階段來判定是否發生至一指定區域中之一儲存；

基於判定發生了至該指定區域中之該儲存，呈現該中斷；及

基於判定並未發生至該指定區域中之該儲存，繼續執行指令處理。

【請求項2】

如請求項1之電腦程式產品，其中該繼續執行指令處理包括進行該選擇功能之一下一處理階段之處理。

【請求項3】

如請求項1之電腦程式產品，其中該繼續執行指令處理包含結束該指令。

【請求項4】

如請求項1之電腦程式產品，其中該指定區域包含由一開始位址及一結束位址界定之記憶體之一區域，其中該開始位址及該結束位址係由一或多個控制暫存器提供。

【請求項5】

如請求項1之電腦程式產品，其中該判定是否發生至該指定區域中之該儲存包含檢查資料經寫入至之一記憶體位址是否處於該指定區域中，其中基於該記憶體位址處於該指定區域中，該判定指示存在至該指定區域中之一儲存。

【請求項6】

如請求項1之電腦程式產品，其中該判定是否發生至該指定區域中之該儲存包含檢查待由該選擇功能使用之一參數區塊之至少一部分是否與該指定區域重疊，其中基於該參數區塊之該至少一部分與該指定區域重疊，該判定指示存在至該指定區域中之一儲存。

【請求項7】

如請求項1之電腦程式產品，其中該判定是否發生至該指定區域中之該儲存包含檢查待由該選擇功能使用之一功能特定保存區域之至少一部分是否與該指定區域重疊，其中基於該功能特定保存區域之該至少一部分與該指定區域重疊，該判定指示存在至該指定區域中之一儲存。

【請求項8】

如請求項1之電腦程式產品，其中該方法進一步包含基於判定待由該

選擇功能使用之一參數區塊之至少一部分與該指定區域重疊，預測將發生一中斷。

【請求項9】

如請求項1之電腦程式產品，其中該方法進一步包含基於判定待由該功能使用之一功能特定保存區域之至少一部分與該指定區域重疊，預測將發生一中斷。

【請求項10】

一種用於促進一計算環境內之處理的電腦系統，該電腦系統包含：

一記憶體；及

至少一個處理器，其與該記憶體進行通信，其中該電腦系統經組態以執行一方法，該方法包含：

針對一指令執行指令處理，該指令經組態以執行複數個功能，其中該複數個功能中之一功能將在複數個處理階段中執行，其中該複數個處理階段中之一處理階段經界定以儲存高達一選擇量之資料，該選擇量之資料係基於該複數個功能中之待執行之該功能，其中該複數個功能中之至少一個功能具有與該複數個功能中之至少一個其他功能不同的一用於該選擇量之資料的值；

預測該選擇功能之該複數個處理階段中之一特定處理階段將產生一中斷，該預測係基於該選擇功能之一記憶體存取模式；

基於處理一選擇功能之該複數個處理階段中之一選擇處理階段來判定是否發生至一指定區域中之一儲存；

基於判定發生了至該指定區域中之該儲存，呈現該中斷；及

基於判定並未發生至該指定區域中之該儲存，繼續執行指令處

理。

【請求項11】

如請求項10之電腦系統，其中該判定是否發生至該指定區域中之該儲存包含檢查資料經寫入至之一記憶體位址是否處於該指定區域中，其中基於該記憶體位址處於該指定區域中，該判定指示存在至該指定區域中之一儲存。

【請求項12】

如請求項10之電腦系統，其中該判定是否發生至該指定區域中之該儲存包含檢查待由該選擇功能使用之一參數區塊之至少一部分是否與該指定區域重疊，其中基於該參數區塊之該至少一部分與該指定區域重疊，該判定指示存在至該指定區域中之一儲存。

【請求項13】

如請求項10之電腦系統，其中該判定是否發生至該指定區域中之該儲存包含檢查待由該選擇功能使用之一功能特定保存區域之至少一部分是否與該指定區域重疊，其中基於該功能特定保存區域之該至少一部分與該指定區域重疊，該判定指示存在至該指定區域中之一儲存。

【請求項14】

一種促進一計算環境內之處理的電腦實施方法，該電腦實施方法包含：

針對一指令執行指令處理，該指令經組態以執行複數個功能，其中該複數個功能中之一功能將在複數個處理階段中執行，其中該複數個處理階段中之一處理階段經界定以儲存高達一選擇量之資料，該選擇量之資料係基於該複數個功能中之待執行之該功能，其中該複數個功能中之至少一

個功能具有與該複數個功能中之至少一個其他功能不同的一用於該選擇量之資料的值；

預測該選擇功能之該複數個處理階段中之一特定處理階段將產生一中斷，該預測係基於該選擇功能之一記憶體存取模式；

基於處理一選擇功能之該複數個處理階段中之一選擇處理階段來判定是否發生至一指定區域中之一儲存；

基於判定發生了至該指定區域中之該儲存，呈現該中斷；及

基於判定並未發生至該指定區域中之該儲存，繼續執行指令處理。

【請求項15】

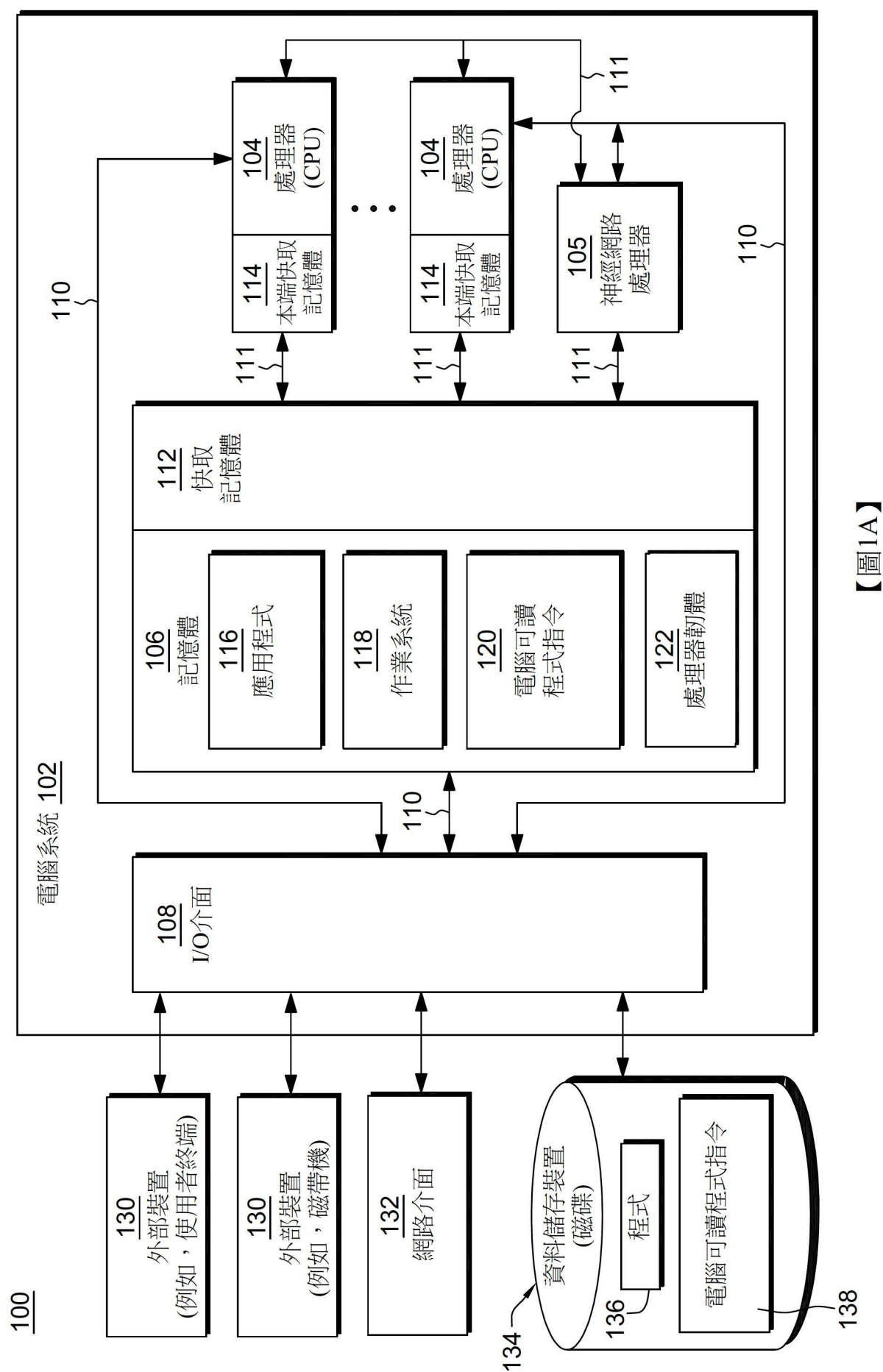
如請求項14之電腦實施方法，其中該判定是否發生至該指定區域中之該儲存包含檢查資料經寫入至之一記憶體位址是否處於該指定區域中，其中基於該記憶體位址處於該指定區域中，該判定指示存在至該指定區域中之一儲存。

【請求項16】

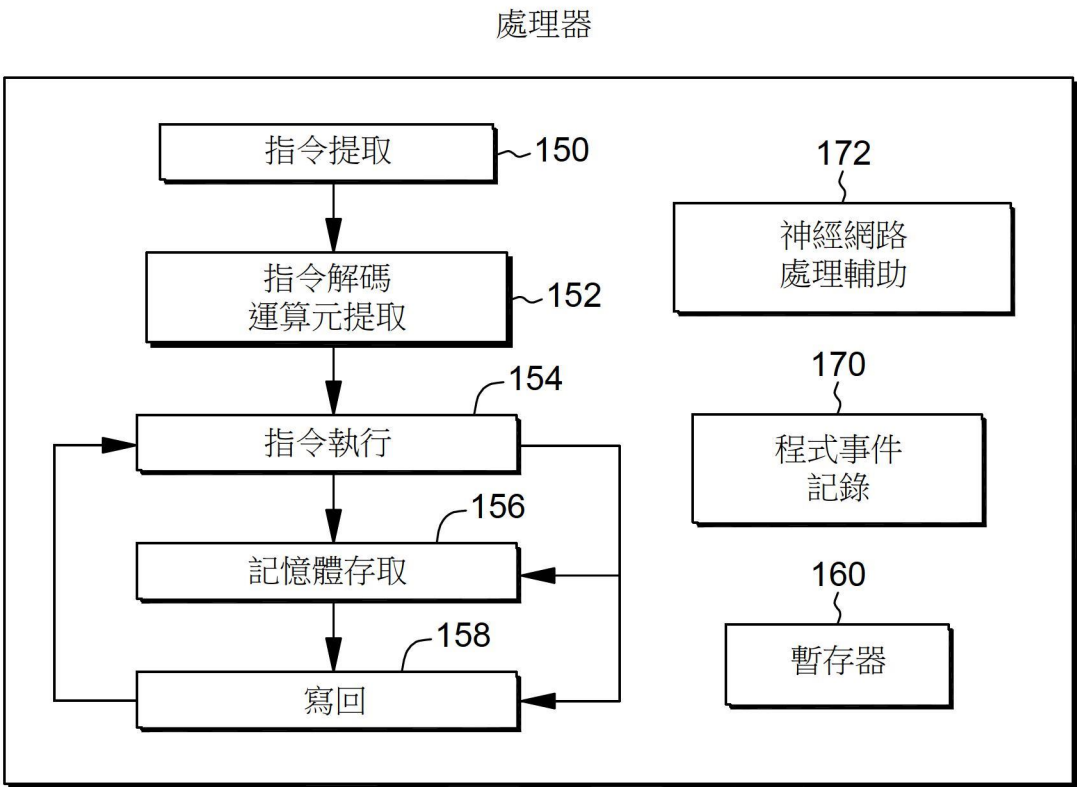
如請求項14之電腦實施方法，其中該判定是否發生至該指定區域中之該儲存包含檢查待由該選擇功能使用之一參數區塊之至少一部分是否與該指定區域重疊，其中基於該參數區塊之該至少一部分與該指定區域重疊，該判定指示存在至該指定區域中之一儲存。

【請求項17】

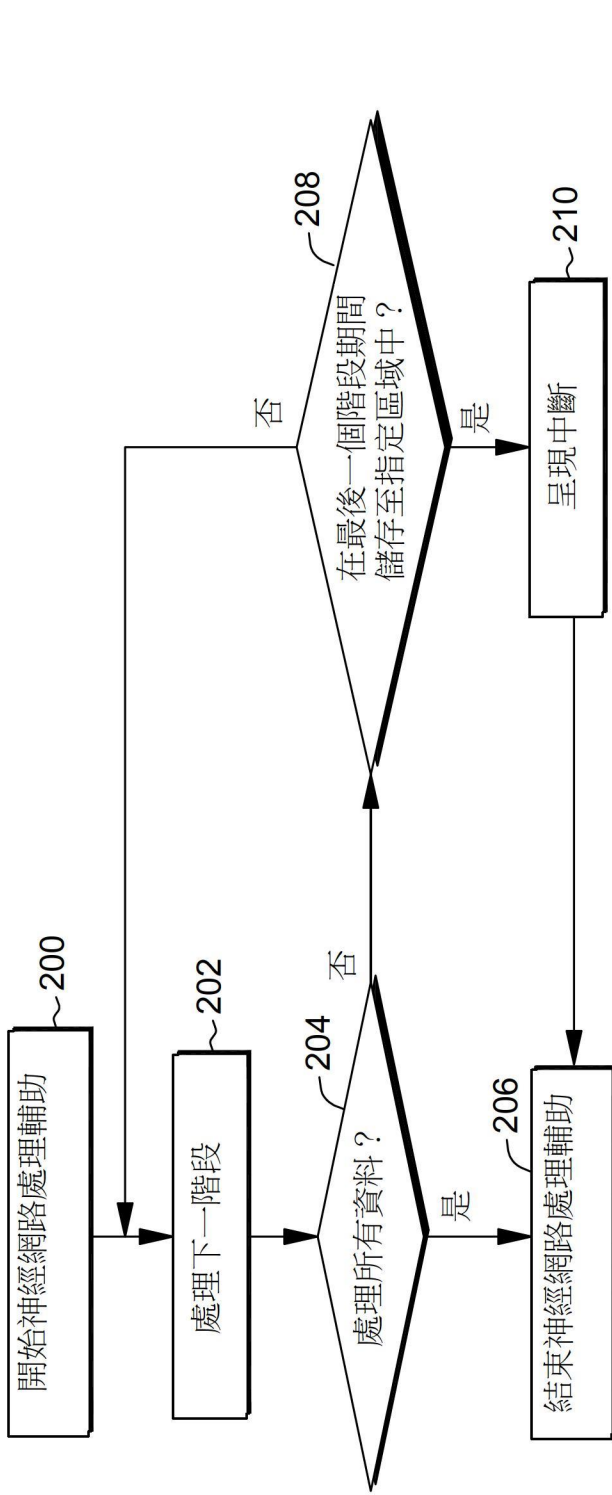
如請求項14之電腦實施方法，其中該判定是否發生至該指定區域中之該儲存包含檢查待由該選擇功能使用之一功能特定保存區域之至少一部分是否與該指定區域重疊，其中基於該功能特定保存區域之該至少一部分與該指定區域重疊，該判定指示存在至該指定區域中之一儲存。



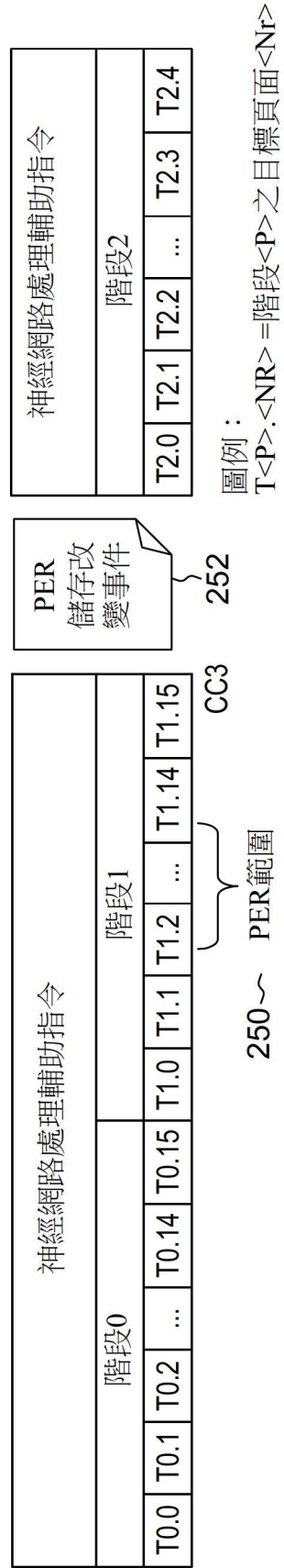
【圖1A】



【圖1B】



【圖2A】



【圖2B】



【圖3A】

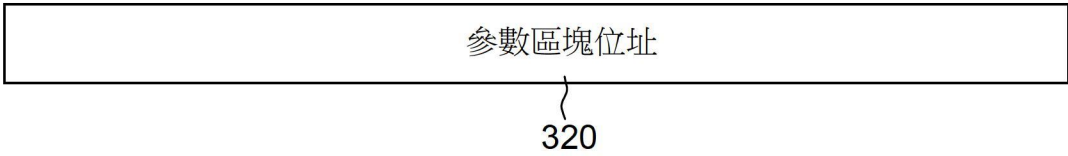


【圖3B】

碼 (DEC)	碼 (HEX)	功能	參數區塊大小 (位元組)
0	0	NNPA - QAF	256
16	10	NNPA - ADD	4096
17	11	NNPA - SUB	4096
18	12	NNPA - MUL	4096
19	13	NNPA - DIV	4096
20	14	NNPA - MIN	4096
21	15	NNPA - MAX	4096
32	20	NNPA - LOG	4096
33	21	NNPA - EXP	4096
49	31	NNPA - RELU	4096
50	32	NNPA - TANH	4096
51	33	NNPA - SIGMOID	4096
52	34	NNPA - SOFTMAX	4096
64	40	NNPA - BATCHNORM	4096
80	50	NNPA - MAXPOOL2D	4096
81	51	NNPA - AVGPOOL2D	4096
96	60	NNPA - LSTMACT	4096
97	61	NNPA - GRUACT	4096
112	70	NNPA - CONVOLUTION	4096
113	71	NNPA - MATMUL-OP	4096
114	72	NNPA - MATMUL-OP-BCAST23	4096

【圖3C】

通用暫存器1



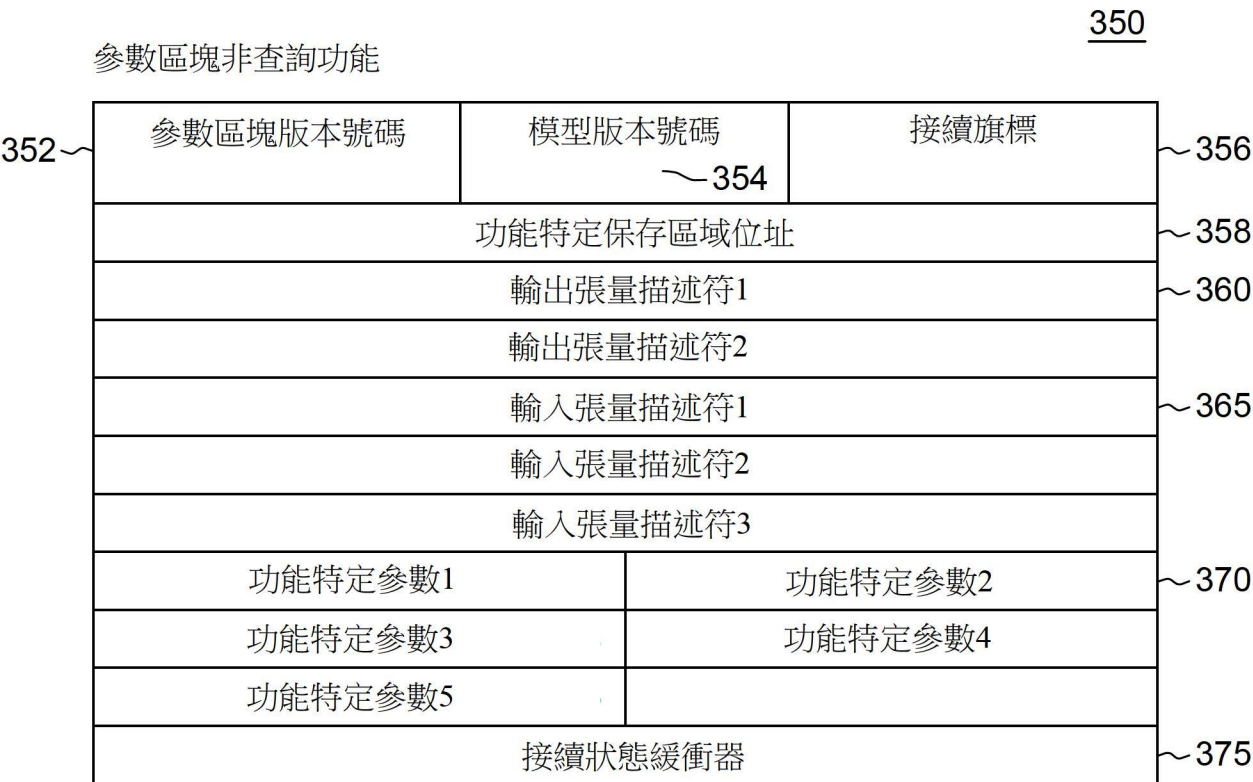
【圖3D】

330

參數區塊-查詢功能



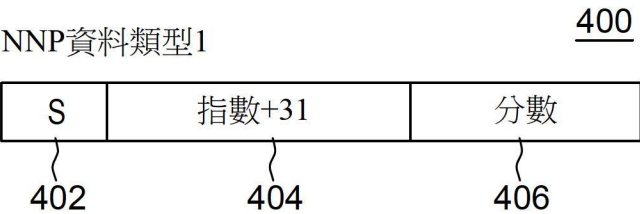
【圖3E】



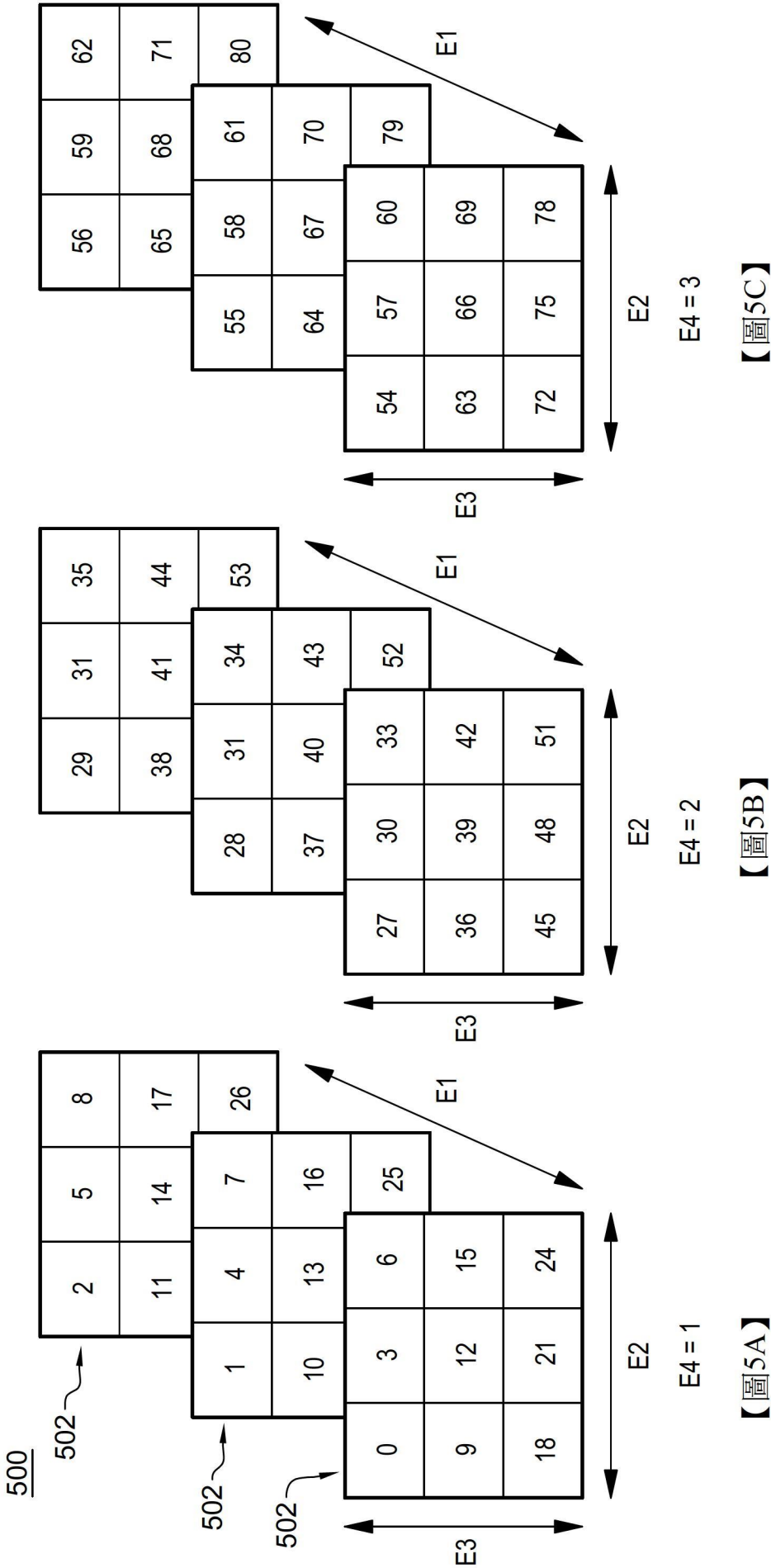
【圖3F】



【圖3G】

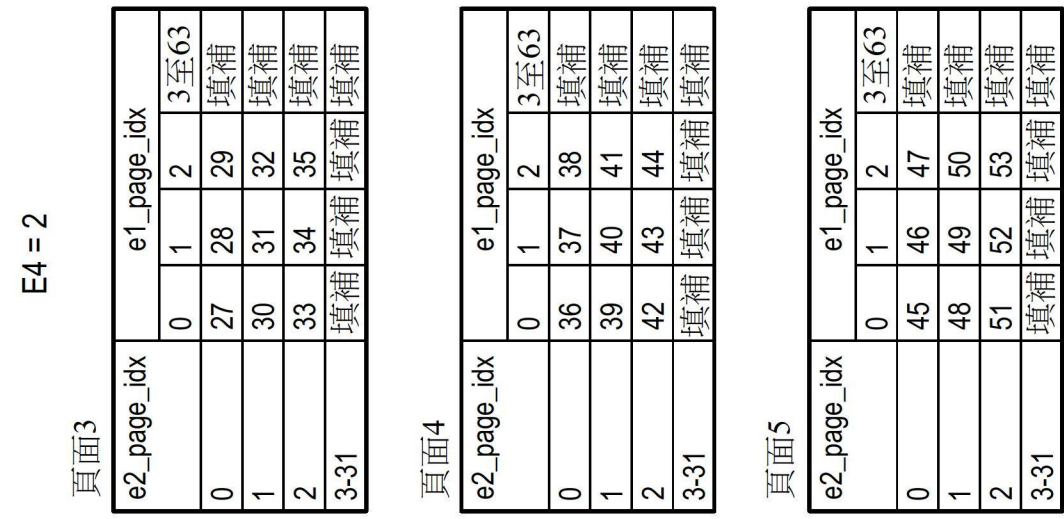


【圖4】

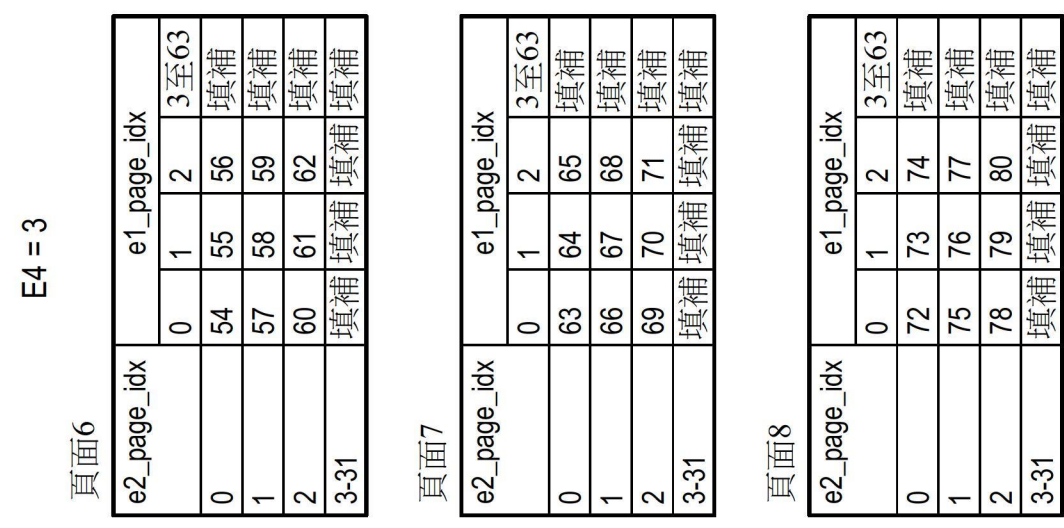




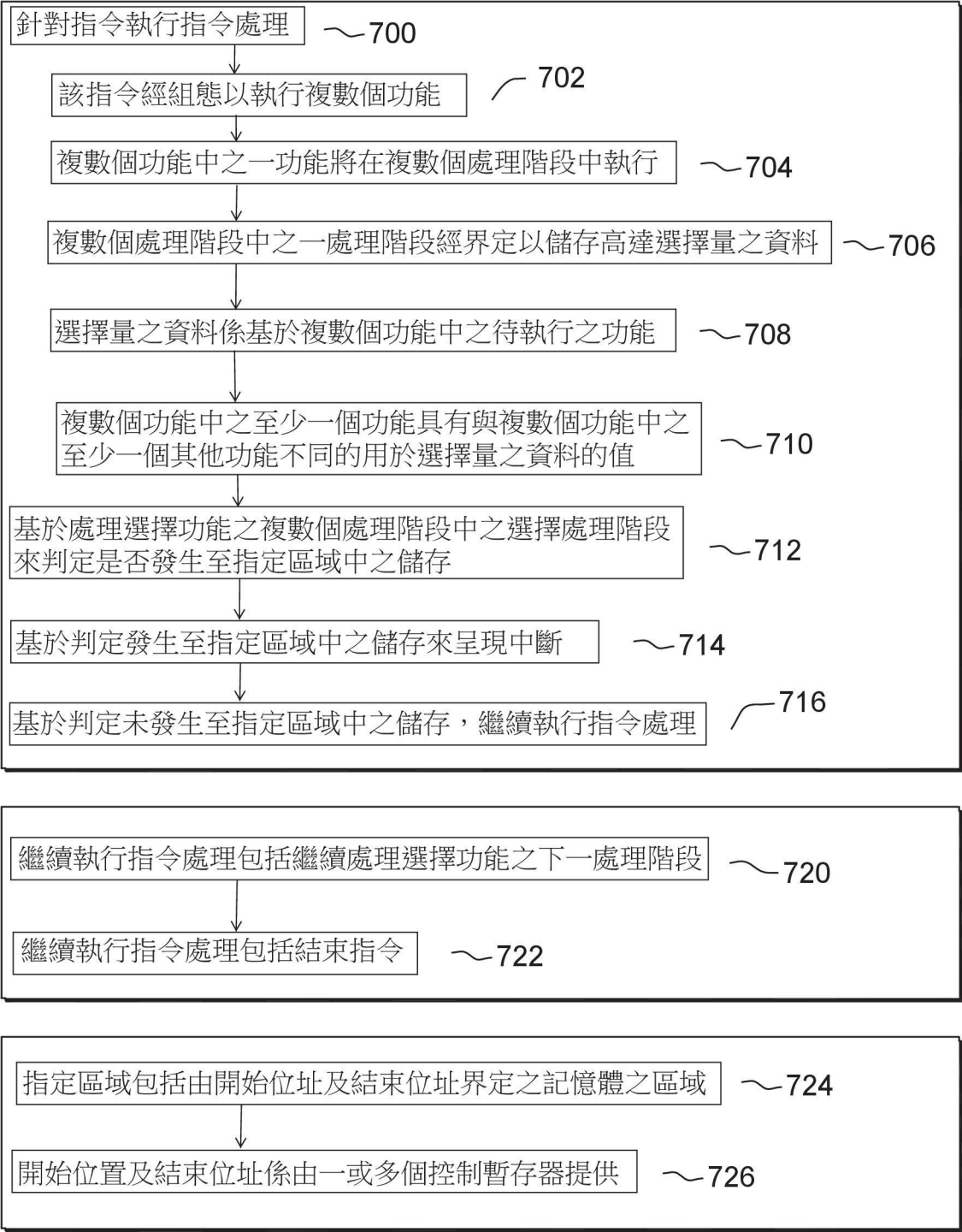
【圖6A】



【圖6B】



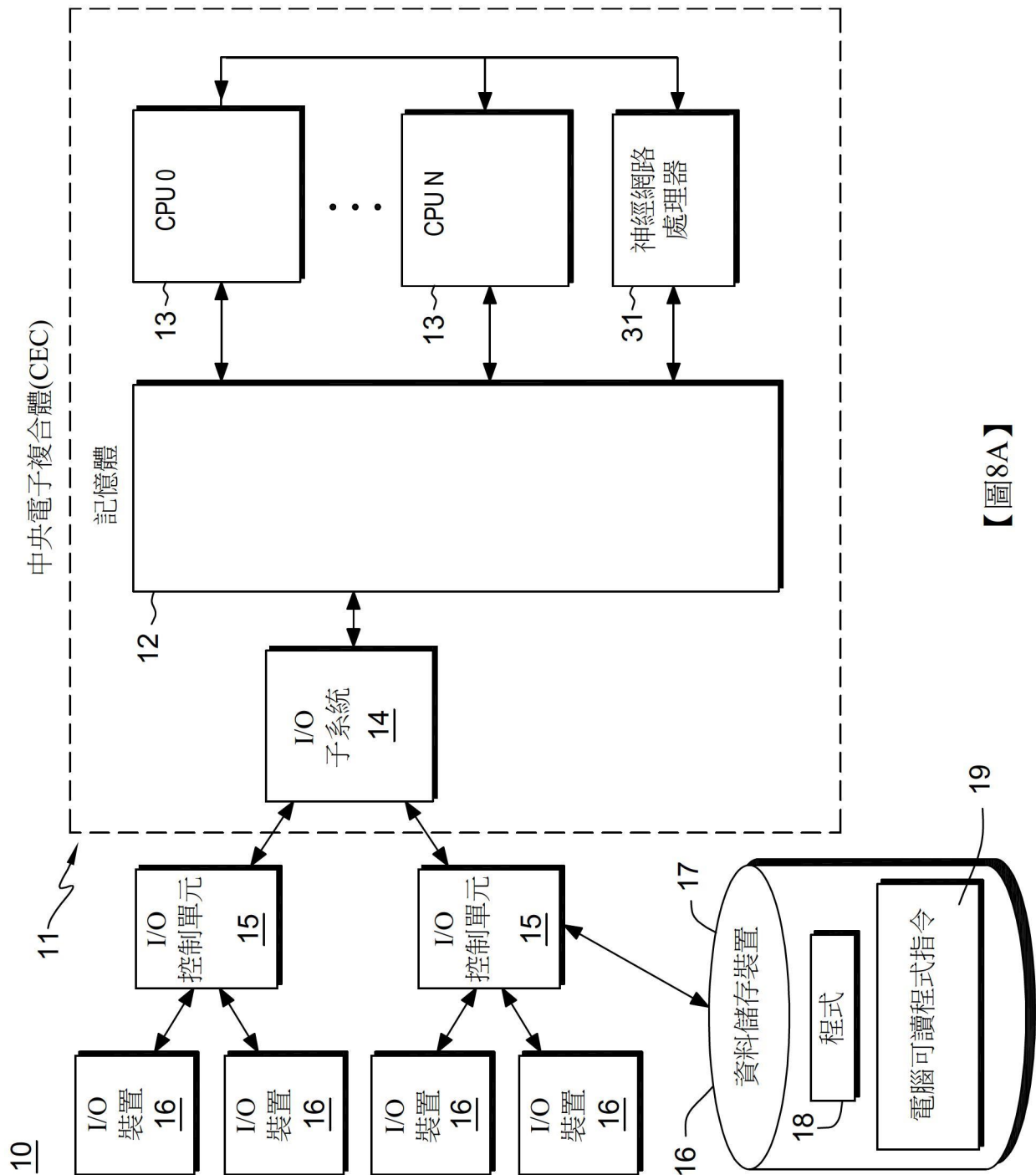
【圖6C】



【圖7A】

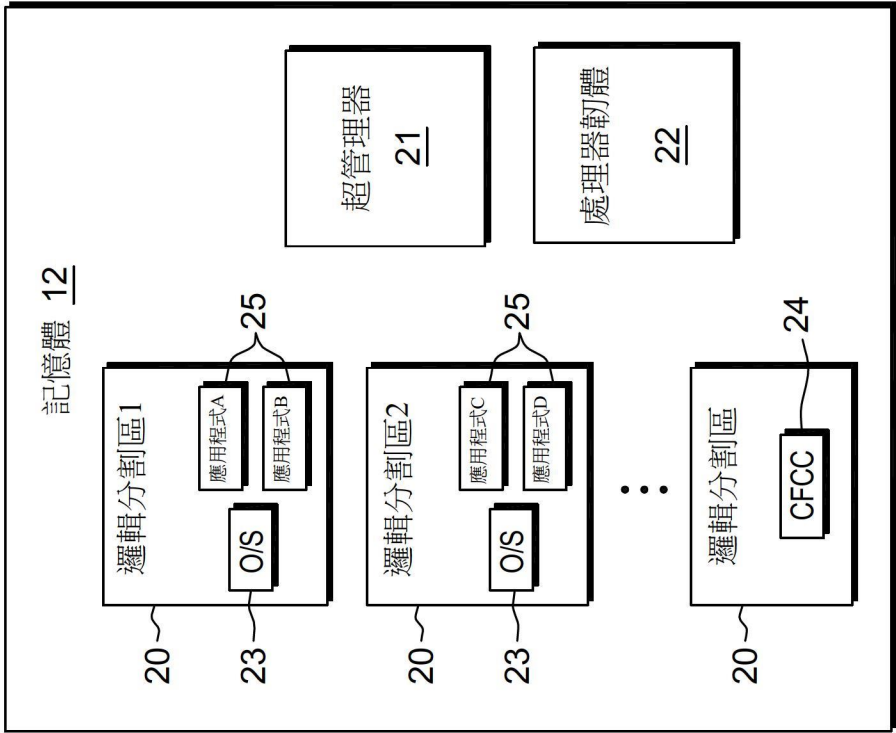


【圖7B】



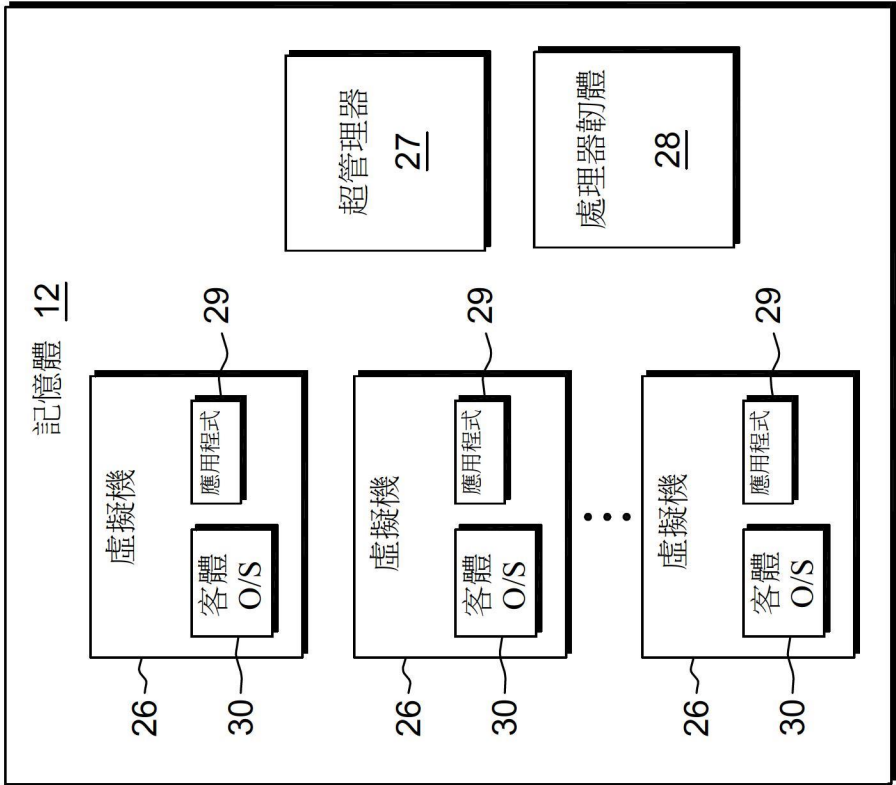
【圖8A】

11



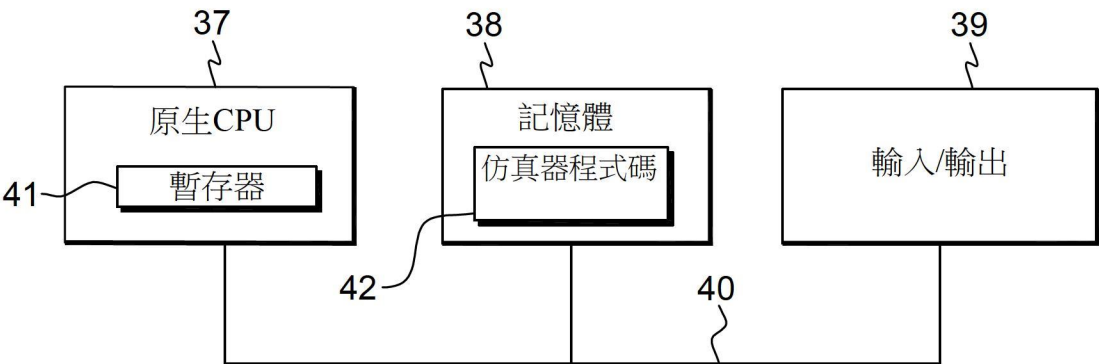
【圖8B】

11

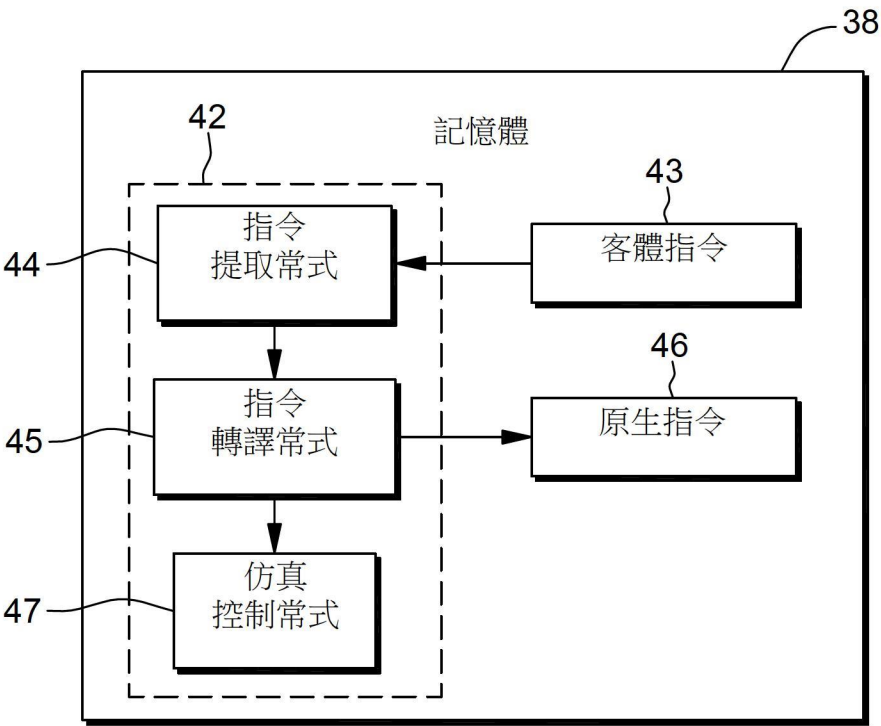


【圖8C】

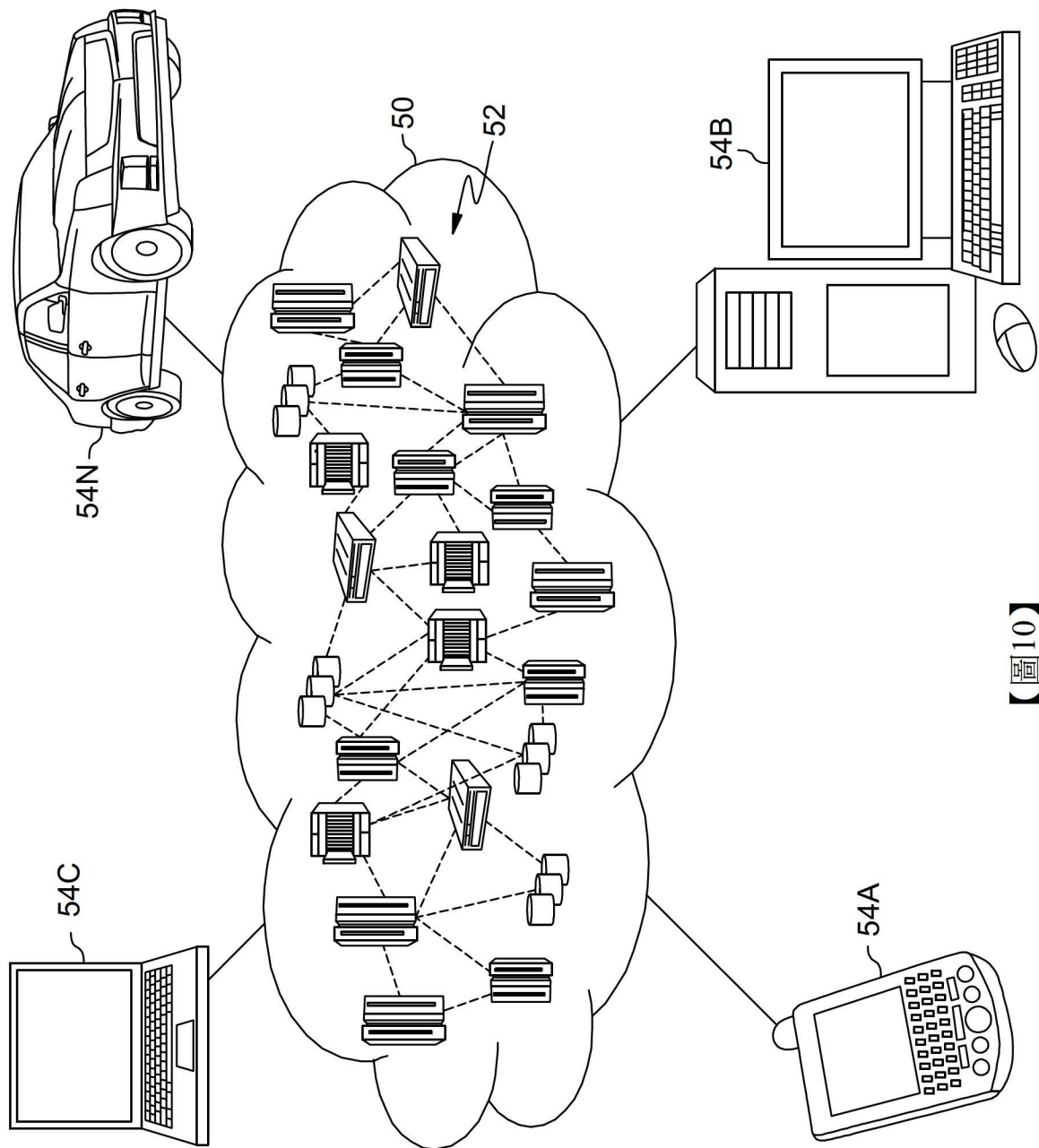
36



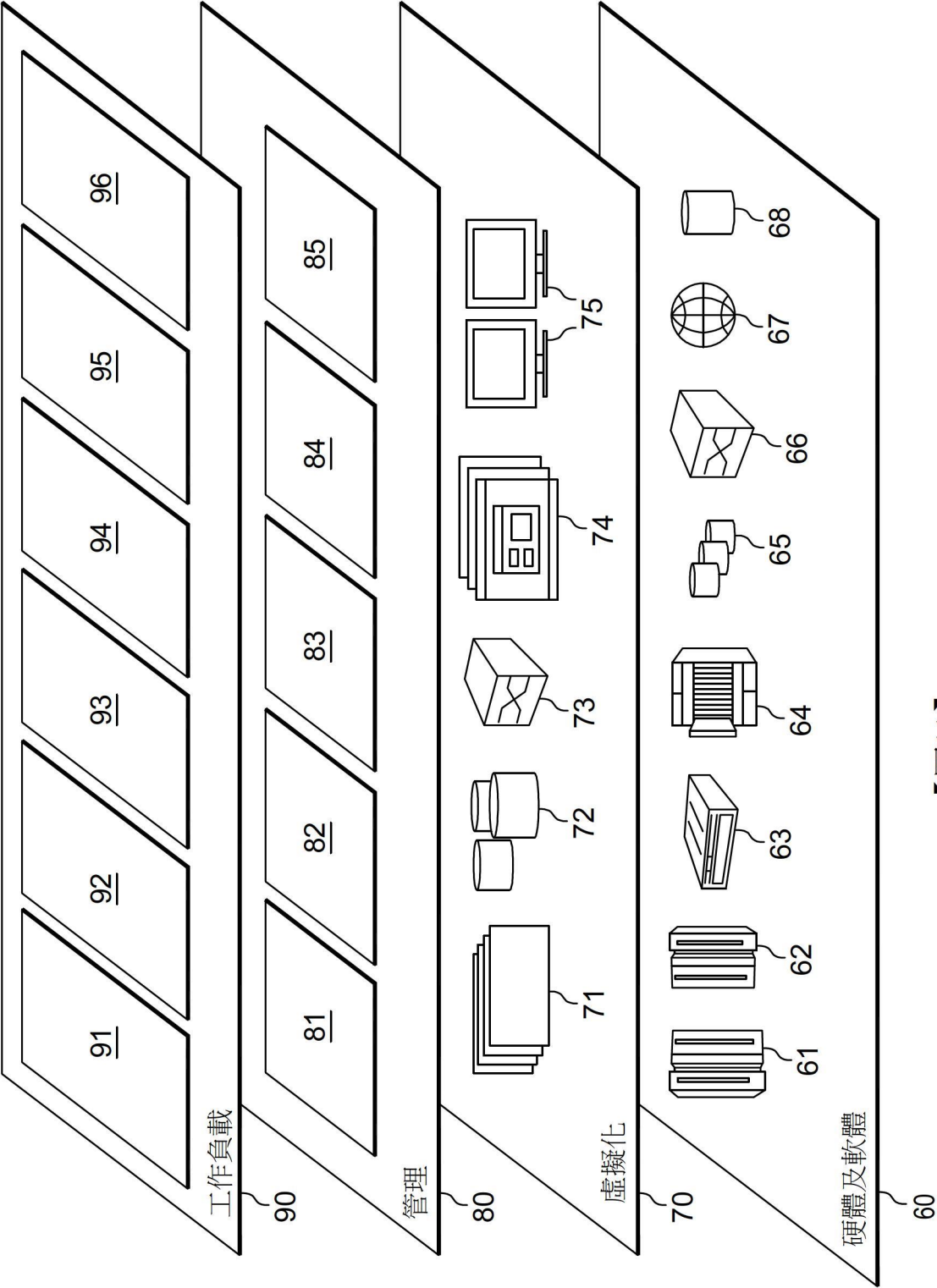
【圖9A】



【圖9B】



【圖10】



【圖11】