



## (51) International Patent Classification:

G06F 12/0862 (2016.01)

G06F 12/0813 (2016.01)

G06F 12/084 (2016.01)

## (21) International Application Number:

PCT/GB2021/053222

## (22) International Filing Date:

09 December 2021 (09.12.2021)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

17/123,527

16 December 2020 (16.12.2020) US

(71) Applicant: **ARM LIMITED** [GB/GB]; 110 Fulbourn Road, Cambridge, Cambridgeshire CB1 9NJ (GB).

(72) Inventors: **DUNHAM, Curtis Glenn**; Arm Limited, 110 Fulbourn Road, Cambridge, Cambridgeshire CB1 9NJ (GB). **BEARD, Jonathan Curtis**; Arm Limited, 110 Fulbourn Road, Cambridge, Cambridgeshire CB1 9NJ (GB).

(74) Agent: **TLIP, LTD.** et al.; Association No 587, 14 King Street, Leeds, Yorkshire LS1 2HL (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

## Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

## Published:

- with international search report (Art. 21(3))

(54) Title: CACHE STASH RELAY

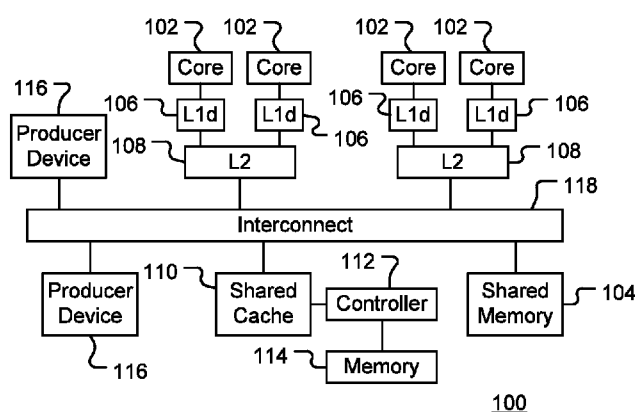


FIG. 1

(57) Abstract: In a cache stash relay, first data, from a producer device, is stashed in a shared cache of a data processing system. The first data is associated with first data addresses in a shared memory of the data processing system. An address pattern of the first data addresses is identified. When a request for second data, associated with a second data address, is received from a processing unit of the data processing system, any data associated with data addresses in the identified address pattern are relayed from the shared cache to a local cache of the processing unit if the second data address is in the identified address pattern. The relaying may include pushing the data from the shared cache to the local cache or a pre-fetcher of the processing unit pulling the data from the shared cache to the local cache in response to a message.



## CACHE STASH RELAY

## BACKGROUND

[0001] In most data processing systems, processing speeds are much faster than memory access speeds. Consequently, the availability of data limits the performance of a system. High speed caches can provide significant improvement, provided that data is in the cache when needed by a processor core. Caches work well, for example, when data at a particular address is loaded into the cache when first used and then reused before being evicted from the cache. However, no improvement is provided when data is not reused before eviction.

[0002] Caches also provide benefit when data usage can be predicted and the data pre-fetched from memory. However, prediction errors, due to program branching for example, result in unnecessary memory read operations.

[0003] Caches also provide performance benefits when data from an input device can be injected directly into a cache of a target processor core. This approach is also called cache stashing. However, in a multiprocessor system, cache stashing requires that the input device knows the core identifier of the target processor. This information may not be available or it may be undesirable to provide the information to an input device.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0004] The accompanying drawings provide visual representations which will be used to more fully describe various representative embodiments and can be used by those skilled in the art to better understand the representative embodiments disclosed and their inherent advantages. In these drawings, like reference numerals identify corresponding or analogous elements.

[0005] **FIG. 1** is a block diagram of a data processing system, in accordance with various representative embodiments.

[0006] **FIG. 2** is a transaction diagram showing information flow in a data processing system, in accordance with various representative embodiments.

[0007] **FIG. 3** is a flow chart of a method of data movement in a data processing system, in accordance with various representative embodiments.

[0008] **FIG. 4** is a flow chart of a method of data movement in a data processing system, in accordance with various representative embodiments.

[0009] **FIG. 5** is a diagram of a logical arrangement of information in a shared cache of a data processing system.

[0010] **FIG. 6** is a diagram of a logical arrangement of address pattern information in a data processing system, in accordance with various representative embodiments.

[0011] **FIG. 7** is a diagram of a logical arrangement of stream information in a data processing system, in accordance with various representative embodiments.

[0012] **FIG. 8** is a transaction diagram showing information flow in a data processing system, in accordance with various representative embodiments.

## DETAILED DESCRIPTION

[0013] The various apparatus and devices described herein provide mechanisms for cache data injection in a data processing system.

[0014] While this present disclosure is susceptible of embodiment in many different forms, there is shown in the drawings and will herein be described in detail specific embodiments, with the understanding that the embodiments shown and described herein should be considered as providing examples of the principles of the present disclosure and are not intended to limit the present disclosure to the specific embodiments shown and described. In the description below, like reference numerals are used to describe the same, similar or corresponding parts in the several views of the drawings. For simplicity and clarity of illustration, reference numerals may be repeated among the figures to indicate corresponding or analogous elements. In this document, relational terms such as first and second, top and bottom, and the like may be used solely to distinguish one entity or action from another entity or action without necessarily requiring or implying any actual such relationship or order between such entities or actions. The terms “comprises,” “comprising,” “includes,” “including,” “has,” “having,” or any other variations thereof, are intended to cover a non-exclusive inclusion, such that a process, method, article, or apparatus that comprises a list of elements does not include only those elements but may include other elements not expressly listed or inherent to such process, method, article, or apparatus. An element preceded by “comprises ...a” does not, without more constraints, preclude the existence of additional identical elements in the process, method, article, or apparatus that comprises the element.

[0015] Reference throughout this document to “one embodiment,” “certain embodiments,” “an embodiment,” “implementation(s),” “aspect(s),” or similar terms means that a particular feature, structure, or characteristic described in connection with the embodiment is included

in at least one embodiment of the present disclosure. Thus, the appearances of such phrases or in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments without limitation.

**[0016]** The term “or”, as used herein, is to be interpreted as an inclusive or meaning any one or any combination. Therefore, “A, B or C” means “any of the following: A; B; C; A and B; A and C; B and C; A, B and C”. An exception to this definition will occur only when a combination of elements, functions, steps or acts are in some way inherently mutually exclusive.

**[0017]** As used herein, the term “configured to”, when applied to an element, means that the element may be designed or constructed to perform a designated function, or that the element has the required structure to enable it to be reconfigured or adapted to perform that function.

**[0018]** **FIG. 1** is a block diagram of a data processing system 100, in accordance with various representative embodiments. System 100 includes a number of processing cores 102 configured to access a shared memory 104. Memory access speeds are much slower than processor speeds, so system 100 also includes one or more high speed caches. The caches may be arranged in a hierarchy, the highest level (level one or ‘L1’) being the fastest and closest to the processing cores. In the example shown, each processing core 102 has a local or private L1 cache 106. This may be reserved for data (L1d) with another L1 cache (not shown) reserved for instructions. Alternatively, the L1 cache could be used for both instructions and data. A larger, slower level two (L2) cache 108 may be shared by several processing cores and feeds data to the L1 caches.

**[0019]** In addition, system 100 includes shared cache 110. The shared cache is the lowest level of the cache hierarchy and is referred to as a last level cache (LLC) or system cache (SC). Shared cache 110 is controlled by a controller 112 that has access to a table memory 114. These will be discussed in more detail below.

**[0020]** Producer devices 116 provide data to be processed or consumed by the processing cores. Example producer devices include network interface cards (NICs), other processors, input devices, Direct Memory Access (DMA) controllers, etc.

**[0021]** Coherent interconnect circuitry 118 enables information, such as messages and data, to be passed between the various elements of the data processing system.

**[0022]** In a conventional data processing system, data received from a producer device 116 is stored in a designated buffer in the shared memory 104.

**[0023]** FIG. 2 is a transaction diagram 200 showing information flow in a data processing system, in accordance with various representative embodiments. The diagram shows timeline 202 for a producer device, timeline 204 for a shared cache, timeline 206 for interconnect circuitry, timeline 208 for a local cache of a consumer processor core and timeline 210 for the consumer processor core. In each case, time flows from top to bottom in the diagram.

**[0024]** At time (1) the producer device starts stashing or injecting data into the shared cache. The producer device may be provided with one or more target addresses. The target address may be the address of a pre-allocated data block in the shared cache, for example. The data may also be sent to a memory. Alternatively, the data in the shared cache may be flagged as having a 'dirty' or 'modified' coherency status to indicate that it should be written to memory when evicted from the cache.

**[0025]** In accordance with an embodiment of the present disclosure, at time (2) the controller of the shared cache recognizes an address pattern of the stashed data. For example, a sequence of consecutive addresses may be detected. If this is a new address pattern, an entry for the pattern is created in a prediction table. The entry contains descriptors of the address pattern, such as start address and end address, or a start address and size (number of addresses or number of cache lines), for example. If the pattern is not new, the existing entry may be updated. For example, the entry could be updated with a new end address or size.

**[0026]** At time (3), a consumer core issues a load request for data associated with an address in the identified address pattern. First, the local cache is checked to see if a copy of the data is stored there. If not, a cache miss occurs and a read request is sent to home agent associated with the data address. The read request is sent via the interconnect circuitry, which introduced a first latency. The home agent checks for the requested data in the shared cache, introducing a second latency and, finding the data, returns the data via the interconnect circuitry, introducing a third latency. Finally, the data may pass through a local cache hierarchy (such as L2 and L1 caches) before finally reaching the consumer core at time (4). Thus, while the latency of a memory read operation has been avoided by stashing the data in the shared cache, there is still considerable latency in the read operation.

**[0027]** In accordance with an embodiment of the present disclosure, at time (5) the controller of the shared cache recognizes that the consumer core has accessed data associated with an identified address pattern and adds an identifier of the consumer core in the prediction table entry for that address pattern. The identifier indicates that the consumer core is a predicted target for the data associated with the pattern. Now that the consumer core has

been identified as a target, data associated with other addresses in the address pattern are pushed to consumer core at time (6). Thus, the data is first stashed in the shared cache and then relayed to the local cache of the consumer core when that core is predicted as a target. On receiving these data values the consumer core may select to stash the data in a local cache or discard them.

**[0028]** At time (7), the consumer core issues a load request for data associated with another address in the identified address pattern. This time, the requested data is found in the local cache, having been relayed to the cache at time (6). The data is loaded at time (8). Thus, the latency of requesting the data, via the interconnect circuitry, from the shared cache has been avoided. This provides a considerable performance advantage.

**[0029]** At time (9), additional or new data is stashed in the shared cache by the producer device. The prediction table is checked at time (10) and reveals the data to be associated with an address in an identified address pattern. The prediction table identifies the consumer core as a potential target for the data and, at time (11), the new data is relayed to the identified consumer core.

**[0030]** At time (12), the consumer core issues a load request for new data. The requested data is found in the local cache, having been relayed to the cache at time (11). The data is loaded at time (13). Again, the latency of retrieving the data, via the interconnect circuitry, from the shared cache has been avoided, providing a considerable performance advantage.

**[0031]** **FIG. 3** is a flow chart of a method 300 of data movement in a data processing system, in accordance with various representative embodiments. Following start block 302, new data is stashed or injected into a shared cache of the data processing system at block 304. At decision block 306 it is determined, by controller circuitry associated with the shared cache, if the address associated with the new data, together with addresses associated with previously stashed data, form a pattern of addresses. If not, as depicted by the negative branch from decision block 306, stream statistics may be updated at block 308. For example, if the previous address was part of a stream of consecutive address, but the new address is not, the length of the terminated stream may be used to update a counter of streams of that length or similar length. If the address is part of a pattern of addresses, as depicted by the positive branch from decision block 306, flow continues to decision block 310. If the address pattern is a newly identified pattern, as depicted by the positive branch from decision block 310, a new entry is added to the prediction table for the pattern. The entry includes descriptors of the pattern, such as start address, stop address, stride, etc. If the identified

pattern is not new, the corresponding entry in the prediction table is accessed to determine if a target consumer core is listed in the table as a predicted target. If a target consumer core is identified, as depicted by the positive branch from decision block 314, the data is relayed to a local cache of that target consumer core at block 316.

**[0032]** FIG. 4 is a flow chart of a method 400 of data movement in a data processing system, in accordance with various representative embodiments. Following start block 402, a controller associated with a shared cache receives a new request, from a consumer core, for data associated with a designated data address at block 404. At block 406, the requested data is returned to the core, having been retrieved from the shared cache or, if not present in the shared cache, from the designated address in a memory. A decision block 408, a prediction table is searched to determine if the designated address is in an identified address pattern. If not, as depicted by the negative branch from decision block 408, flow returns to block 404. If the designated address is part of a previously identified address pattern, as depicted by the positive branch from decision block 408, an identifier of the consumer core is added to the entry in the prediction table for the identified address pattern at block 410, indicating the core to be a target for data associated with the address pattern. At block 412, any other data in the shared cache that is associated with addresses in the identified address pattern is relayed to a local cache of the core. In this way, data associated with the address pattern is placed in local cache of the core in anticipation that the core, having accessed one address in the pattern, with likely request access to data associated with other addresses in the pattern. When this access is requested, the data will be available in the local cache of the core, thereby avoiding the delay associated with requesting the data via the interconnect circuitry.

**[0033]** FIG. 5 is a diagram of a logical arrangement of information 500 in a shared cache of a data processing system. The information includes data array 502 for storing cached data. The data is logically organized as a number of lines, each containing data associated with multiple addresses. Corresponding address tags are stored in tag array 504. The tag array enables the controller to identify which line of data in data array 502 contains data associated with a given address. When data is stashed in the shared cache by a producer device, its address tag is added to address tag array 504 and the data is stored in a corresponding line of data array 502. The status of a line of data is stored in status array 506. This may indicate, for example, if a line of data is valid or invalid. In addition, the status may indicate whether or not the line was stashed or is associated with an identified address pattern. This would reduce the number of accesses to the prediction table.

[0034] **FIG. 6** is a diagram of a prediction table 600, in accordance with various representative embodiments. Each logical row, 602a, 602b, 602c, etc., of the prediction table includes a number of descriptors, 604a, 604b, 604c, etc., of an associated identified address pattern. An address pattern may have any number of descriptors. For example, in row 602b, descriptor 604a is a starting address and descriptor 604b is an end address of a pattern comprising a stream of consecutive addresses. In row 602c, descriptor 604a is a starting address and descriptor 604b is a size (such as a number of addresses, a number of cache lines or a number of data blocks). Descriptor 604c is a 'stride' of a pattern, indicating the number of elements between successive elements in a data stream.

[0035] An additional entry 606 in each row identifies consumer cores that have accessed data associated with an address in the address pattern. This may be, for example, a list of explicit identifiers or a bit pattern in which each bit position corresponds to consumer core in the data processing system. Other logical arrangements or formats may be used without departing from the present disclosure.

[0036] The prediction table may be reset at intervals to remove outdated information. In addition, a consumer core may provide feedback to the controller as to the utilization of relayed data. This feedback may be used to update the prediction table. For example, a core identifier may be removed from an entry in the table if the corresponding relayed data was found to be not useful.

[0037] As described above, the controller associated with the shared cache identifies a stream of addresses associated with stashed data as an address pattern. However, since space in the local caches is limited, it may be advantageous to select which address patterns should be relayed to a target consumer core. For example, stashing data associated with longer streams may provide greater performance benefits. In accordance with an embodiment of the disclosure, statistics relating to the length of data streams are maintained. A number of stream length ranges are designated and, for each length range, a counter value is incremented each time a stream of the corresponding length is completed. Based on this information, the controller may select a threshold on the stream length. For example, the controller could select to only relay data associated with stream lengths of 10 or more.

[0038] The table may be reset periodically. Alternatively, the counter values may be decremented at designated time intervals so that they reflect more recent core behavior.

[0039] **FIG. 7** is a diagram of a logical arrangement of stream information 700 in a data processing system in accordance with embodiments of the disclosure. As described above, the controller associated with the shared cache identifies a stream of addresses associated

with stashed data as an address pattern. However, since space in the local caches is limited, it may be advantageous to select which address patterns should be relayed to a target consumer core. For example, stashing data associated with longer streams may provide greater performance benefits. In accordance with an embodiment of the disclosure, statistics relating to the length of data streams are maintained. In the example shown, a number of stream length ranges are designated. For each length range, a counter value is incremented each time a stream of the corresponding length is completed. In the example shown, streams with length 2-4 (cache lines or address) have been detected 23 times, streams with length 5-8 have been detected 43 times, streams with length 9-16 have been detected 121 times, etc. Based on this information, the controller may select a threshold on the stream length. For example, the controller could select to only relay data associated with stream lengths of 9 or more.

[0040] The table may be reset periodically. Alternatively, the counter values may be decremented at designated time intervals so that they reflect more recent core behavior.

[0041] **FIG. 8** is a transaction diagram 800 showing information flow in a data processing system, in accordance with various representative embodiments. The diagram shows timeline 802 for a producer device, timeline 804 for a shared cache, timeline 806 for interconnect circuitry, timeline 808 for a local cache of a consumer processor core and timeline 810 for the consumer processor core. In each case, time flows from top to bottom in the diagram.

[0042] As in the embodiment shown in **FIG. 2** and described above, at time (1) the producer device begins stashing data into the shared cache. The producer device may be provided with one or more target addresses. The target address may be a pre-allocated data block in the shared cache, for example. At time (2) the controller of the shared cache recognizes an address pattern of the stashed data. For example, a sequence of consecutive addresses may be detected. If this is a new address pattern, an entry for the pattern is created in a prediction table. At time (3), a consumer core issues a load request, via the interconnect circuitry, for data associated with an address in the identified address pattern consumer core. At time (4) the data is received at the core. At time (5) the controller of the shared cache recognizes that the consumer core has accessed data associated with an identified address pattern and stores an identifier of that consumer core in the prediction table entry for the address pattern. This indicates that the consumer core is a predicted target for data associated with the pattern.

[0043] In accordance with an embodiment of the disclosure, at time (6) the controller sends a message, via the interconnect circuitry, to a pre-fetcher of the requesting consumer

core, indicating that data associated with the address pattern is available in the shared cache. At time (7), the pre-fetcher of the in the consumer core issues a request to pre-fetch data associated with the address pattern. At time (8), the requested data is relayed from the shared cache to a local cache of the consumer core. When the consumer core executes a load instruction for the data, at time (9), the data is available in the local cache and can be loaded, with minimum latency, at time (10).

**[0044]** Numerous details have been set forth to provide an understanding of the embodiments described herein. The embodiments may be practiced without these details. In other instances, well-known methods, procedures, and components have not been described in detail to avoid obscuring the embodiments described. The disclosure is not to be considered as limited to the scope of the embodiments described herein.

**[0045]** Those skilled in the art will recognize that the present disclosure has been described by means of examples. The present disclosure could be implemented using hardware component equivalents such as special purpose hardware and/or dedicated processors which are equivalents to the present disclosure as described and claimed. Similarly, dedicated processors and/or dedicated hard wired logic may be used to construct alternative equivalent embodiments of the present disclosure.

**[0046]** Dedicated or reconfigurable hardware components used to implement the disclosed mechanisms may be described, for example, by instructions of a hardware description language (HDL), such as VHDL, Verilog or RTL (Register Transfer Language), or by a netlist of components and connectivity. The instructions may be at a functional level or a logical level or a combination thereof. The instructions or netlist may be input to an automated design or fabrication process (sometimes referred to as high-level synthesis) that interprets the instructions and creates digital hardware that implements the described functionality or logic.

**[0047]** The HDL instructions or the netlist may be stored on non-transient computer readable medium such as Electrically Erasable Programmable Read Only Memory (EEPROM); non-volatile memory (NVM); mass storage such as a hard disc drive, floppy disc drive, optical disc drive; optical storage elements, magnetic storage elements, magneto-optical storage elements, flash memory, core memory and/or other equivalent storage technologies without departing from the present disclosure. Such alternative storage devices should be considered equivalents.

**[0048]** Various embodiments described herein are implemented using dedicated hardware, configurable hardware or programmed processors executing programming instructions that

are broadly described in flow chart form that can be stored on any suitable electronic storage medium or transmitted over any suitable electronic communication medium. A combination of these elements may be used. Those skilled in the art will appreciate that the processes and mechanisms described above can be implemented in any number of variations without departing from the present disclosure. For example, the order of certain operations carried out can often be varied, additional operations can be added or operations can be deleted without departing from the present disclosure. Such variations are contemplated and considered equivalent.

**[0049]** The various representative embodiments, which have been described in detail herein, have been presented by way of example and not by way of limitation. It will be understood by those skilled in the art that various changes may be made in the form and details of the described embodiments resulting in equivalent embodiments that remain within the scope of the appended claims.

**[0050]** In one embodiment, a method for cache stashing is provided. First data, from a producer device of a data processing system, is stashed in a shared cache of the data processing system. The first data is associated with a plurality of first data addresses in a shared memory of the data processing system and an address pattern of the plurality of first data addresses is identified. In response to receiving a request, from a processing unit of the data processing system, for second data associated with a second data address, the second data address is determined to be in the identified address pattern and data associated with data addresses in the identified address pattern is relayed from the shared cache to a local cache of the processing unit.

**[0051]** The identified address pattern may include a range of consecutive data addresses.

**[0052]** In a further embodiment, relaying the data associated with data addresses in the identified address pattern includes pushing the data from the shared cache to the local cache of the processing unit.

**[0053]** In a further embodiment, relaying the data associated with data addresses in the identified address pattern includes sending a message to inform a pre-fetcher of the processing unit of the identified address pattern and the pre-fetcher of the processing unit pulling the data from the shared cache to the local cache of the processing unit.

**[0054]** In further embodiment, third data from the producer device is stashed in the shared cache, the third data associated with a third data address in the shared memory. When the third data address is determined to be in the identified address pattern, the processing unit

is predicted to be a target for data in the identified address pattern and the third data is relayed from the shared cache and stashed in the local cache of the processing unit.

**[0055]** In a further embodiment, descriptors of the identified address pattern are stored in an entry of a prediction table and an identifier of the processing unit is stored in the entry of the prediction table as a target for the data associated with the identified address pattern.

**[0056]** In a further embodiment, third data is relayed from the producer device and stashed in the shared cache of the data processing system, the third data associated with a third data address in the shared memory. When, from the prediction table, the third data address is determined to be in the identified address pattern, it is determined, from the prediction table, that the processing unit is the target for data in the identified address pattern and the third data is relayed from the shared cache and stashed in the local cache of the processing unit.

**[0057]** The shared cache may be a system level cache or a last level cache of the data processing system and the local cache of the processing unit may be a level one (L1) cache or a level two (L2) cache of the processing unit.

**[0058]** In a further embodiment, a data processing system is provided. The data processing system including a first processing unit, one or more second processing units, a shared cache, a producer device and a controller. The first processing unit has a local cache and the shared cache is accessible by the first processing unit. The stashed data is associated with a plurality of first data addresses in a shared memory of the data processing system. The producer device is configured to stash data in the shared cache. The controller is configured to identify an address pattern of the plurality of the first data addresses. Responsive to receiving a request, from the first processing unit, for second data associated with a second data address, the controller determines the second data address to be in the identified address pattern and data associated with data addresses in the identified address pattern is relayed from the shared cache to a local cache of the first processing unit.

**[0059]** In a further embodiment, the controller is configured to push the data associated with data addresses in the identified address pattern from the shared cache to the local cache of the first processing unit.

**[0060]** In a still further embodiment, the controller is configured to send a message to a pre-fetcher of the first processing unit when data stashed in the shared cache is determined to be an identified address pattern. The pre-fetcher of the first processing unit then pulls the data associated with data addresses in the identified address pattern from the shared cache to the local cache of the first processing unit in response to the message.

**[0061]** The shared cache may be a system cache or a last level cache and the local cache of the first processing unit may be a level one (L1) cache or a level two (L2) cache, for example.

**[0062]** The identified address pattern may be a range of consecutive data addresses, for example.

**[0063]** In a further embodiment, the data processing system includes a memory and the controller is configured to store descriptors of the identified address pattern in an entry of a prediction table stored in the memory.

**[0064]** The controller may be configured to store an identifier of the first processing unit in the entry of the prediction table as a target for the data associated with the identified address pattern. In addition, the controller may be further configured to determine, from the prediction table, that a third data address, associated with third data stashed in the shared cache by the producer device, is in the identified address pattern and determine, from the prediction table, the first processing unit to be target for data associated with the identified address pattern. In this case, the third data is relayed from the shared cache to the local cache of the first processing unit.

**[0065]** The producer device may be, for example, a network interface card, an input device, a camera, a hardware accelerator or a direct memory access (DMA) device.

**[0066]** The producer device may be a processing unit of the one or more second processing units or some other device.

**[0067]** In a still further embodiment, the data processing system includes interconnect circuitry, where the first processing unit and the shared cache are operationally coupled via the interconnect circuitry.

What is claimed:

1. A method comprising:

stashing first data, from a producer device of a data processing system, in a shared cache of the data processing system, the first data associated with a plurality of first data addresses in a shared memory of the data processing system;

identifying an address pattern of the plurality of first data addresses;

responsive to receiving a request, from a processing unit of the data processing system, for second data associated with a second data address:

determining the second data address to be in the identified address pattern; and

relaying data associated with data addresses in the identified address pattern from the shared cache to a local cache of the processing unit.

2. The method of claim 1, where the identified address pattern comprises a range of consecutive data addresses.

3. The method of claim 1, where relaying the data associated with data addresses in the identified address pattern includes pushing the data from the shared cache to the local cache of the processing unit.

4. The method of claim 1, where relaying the data associated with data addresses in the identified address pattern includes a pre-fetcher of the processing unit pulling the data from the shared cache to the local cache of the processing unit.

5. The method of claim 4, further comprising sending a message to inform the pre-fetcher of the processing unit of the identified address pattern.

6. The method of claim 1, further comprising:

responsive to stashing third data from the producer device in the shared cache, the third data associated with a third data address in the shared memory:

determining the third data address to be in the identified address pattern;

predicting the processing unit to be a target for data in the identified address pattern; and

stashing the third data from the shared cache in the local cache of the processing unit.

7. The method of claim 1, further comprising:

storing descriptors of the identified address pattern in an entry of a prediction table;  
and  
storing an identifier of the processing unit in the entry of the prediction table as a target for the data associated with the identified address pattern.

8. The method of claim 7, further comprising:

responsive to stashing third data from the producer device in the shared cache of the data processing system, the third data associated with a third data address in the shared memory:

determining, from the prediction table, the third data address to be in the identified address pattern;

determining, from the prediction table, the processing unit to be target for data in the identified address pattern; and

stashing the third data from the shared cache in the local cache of the processing unit.

9. The method of claim 1, where the shared cache comprises a system level cache or a last level cache of the data processing system and the local cache of the processing unit comprises a level one (L1) cache or a level two (L2) cache of the processing unit.

10. A data processing system comprising:

a first processing unit having a local cache;

a shared cache accessible by the first processing unit and one or more second processing units of the data processing system;

a producer device configured to stash data in the shared cache, the stashed data associated with a plurality of first data addresses in a shared memory of the data processing system; and

a controller configured to:

identify an address pattern of the plurality of the first data addresses;  
responsive to receiving a request, from the first processing unit, for second data associated with a second data address:

determine the second data address to be in the identified address pattern; and

relay data associated with data addresses in the identified address pattern from the shared cache to a local cache of the first processing unit.

11. The data processing system of claim 10, where the controller is further configured to push the data associated with data addresses in the identified address pattern from the shared cache to the local cache of the first processing unit.

12. The data processing system of claim 10, where the controller is further configured to send a message to a pre-fetcher of the first processing unit when data stashed in the shared cache is determined to be an identified address pattern, and the pre-fetcher of the first processing unit is configured to pull the data associated with data addresses in the identified address pattern from the shared cache to the local cache of the first processing unit.

13. The data processing system of claim 10, where the shared cache comprises a system cache or a last level cache and the local cache of the first processing unit comprises a level one (L1) cache or a level two (L2) cache.

14. The data processing system of claim 10, where the identified address pattern comprises a range of consecutive data addresses.

15. The data processing system of claim 10, further comprising a memory, where the controller is configured to store descriptors of the identified address pattern in an entry of a prediction table stored in the memory.

16. The data processing system of claim 15, where the controller is further configured to store an identifier of the first processing unit in the entry of the prediction table as a target for the data associated with the identified address pattern.

17. The data processing system of claim 16, where the controller is further configured to:

determine, from the prediction table, that a third data address, associated with third data stashed in the shared cache by the producer device, is in the identified address pattern;

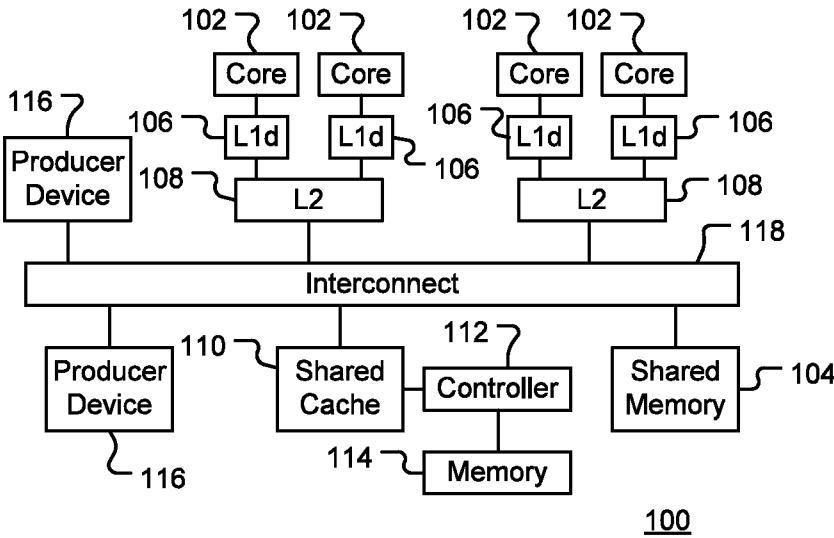
determine, from the prediction table, the first processing unit to be target for data associated with the identified address pattern; and

relay the third data from the shared cache to the local cache of the first processing unit.

18. The data processing system of claim 10, where the producer device comprises a network interface card, an input device, a camera, a hardware accelerator or a direct memory access (DMA) device.

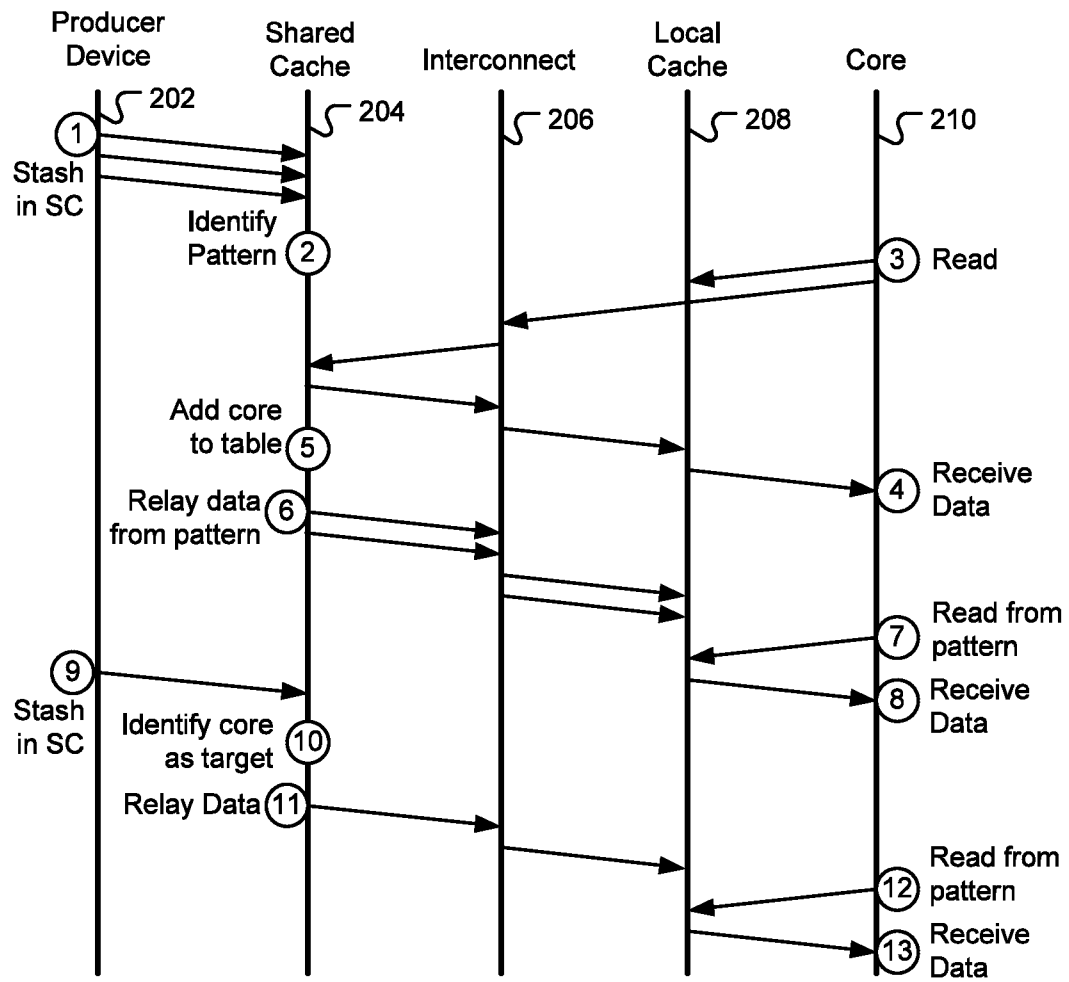
19. The data processing system of claim 10, where the producer device comprises a processing unit of the one or more second processing units.

20. The data processing system of claim 10, further comprising interconnect circuitry, where the first processing unit and the shared cache are operationally coupled via the interconnect circuitry.

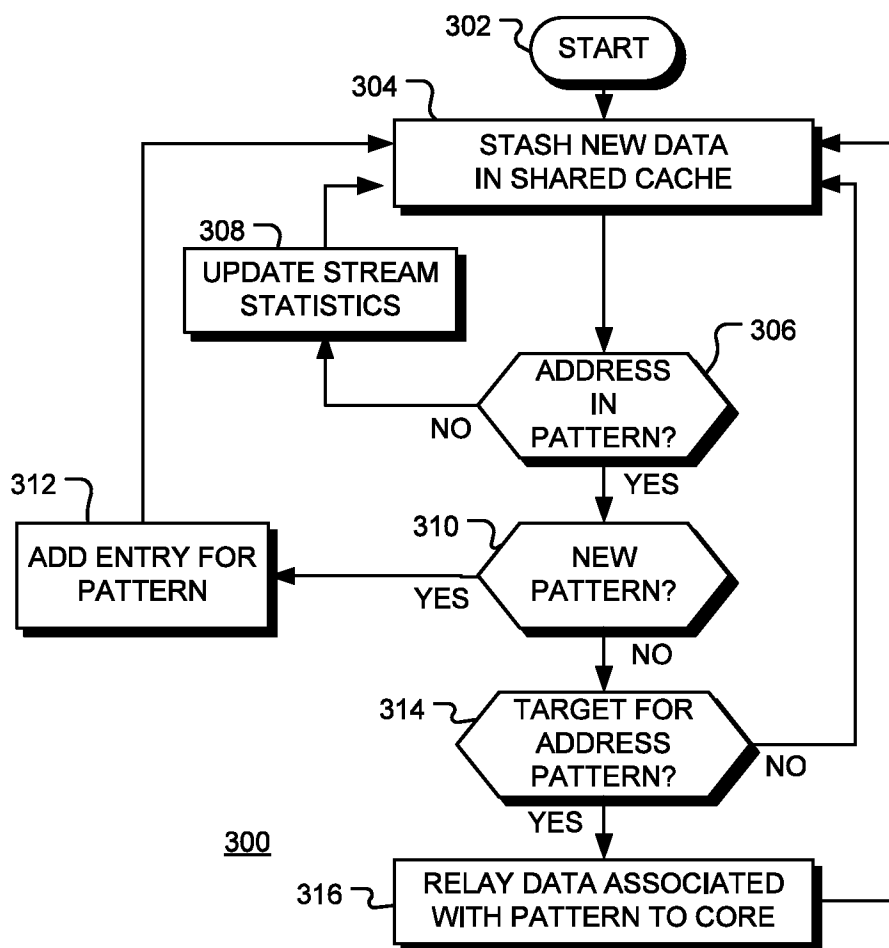


**FIG. 1**

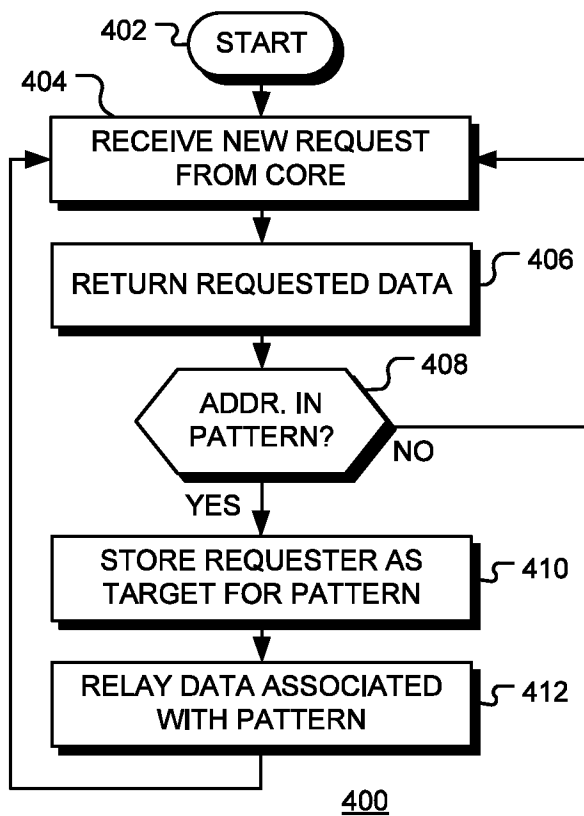
2/6

**FIG. 2**200

3/6

**FIG. 3**

4/6

**FIG. 4**

5/6

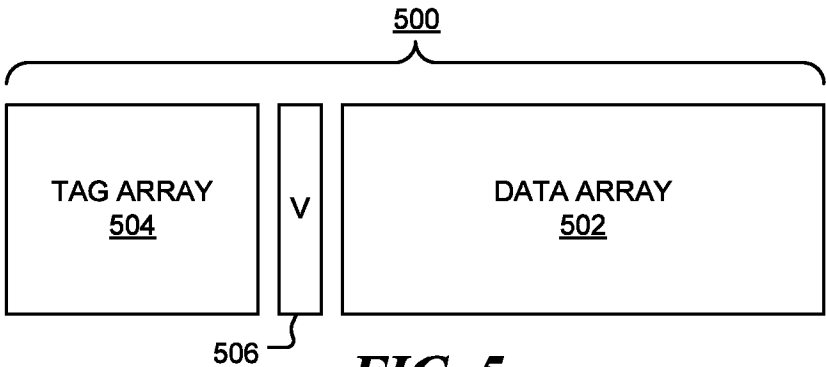


FIG. 5

|      |              |              |              |            |
|------|--------------|--------------|--------------|------------|
|      | 604a         | 604b         | 604c         | 606        |
| 602a | DESCRIPTOR 1 | DESCRIPTOR 2 | DESCRIPTOR 3 | TARGET IDs |
| 602b | Start Addr.  | End Addr.    |              | TARGET IDs |
| 602c | Start Addr.  | Size         | Stride       | TARGET IDs |
|      | ⋮            | ⋮            | ⋮            | ⋮          |

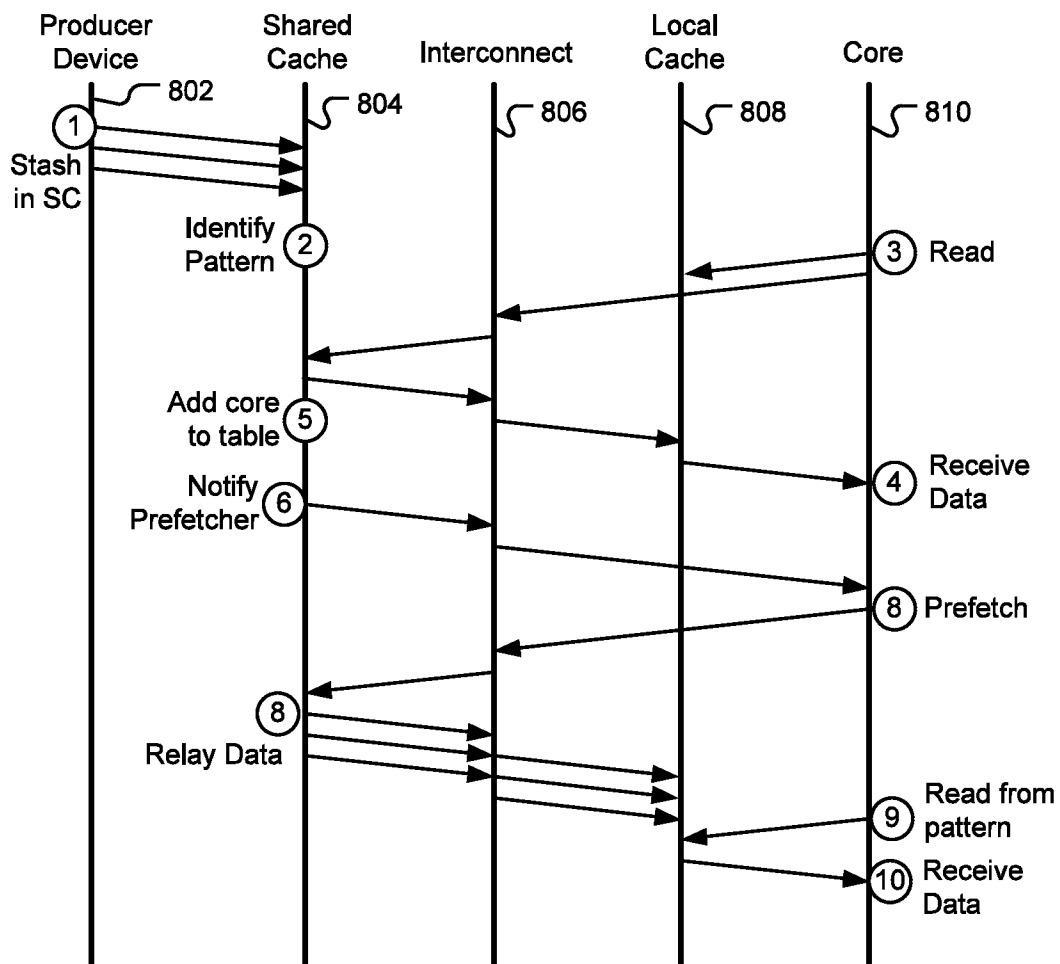
600

FIG. 6

|               |     |     |      |       |  |
|---------------|-----|-----|------|-------|--|
|               | 700 |     |      |       |  |
| Stream Length | 2-4 | 5-8 | 9-16 | 17-32 |  |
| Count         | 23  | 43  | 121  | 54    |  |

FIG. 7

6/6



800

**FIG. 8**

# INTERNATIONAL SEARCH REPORT

International application No  
**PCT/GB2021/053222**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                           |                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|
| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br><b>INV. G06F12/0862 G06F12/084 G06F12/0813</b><br><b>ADD.</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                           |                                                                                                                    |
| According to International Patent Classification (IPC) or to both national classification and IPC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                           |                                                                                                                    |
| <b>B. FIELDS SEARCHED</b><br>Minimum documentation searched (classification system followed by classification symbols)<br><b>G06F</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                           |                                                                                                                    |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                           |                                                                                                                    |
| Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br><b>EPO-Internal, WPI Data</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                           |                                                                                                                    |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                           |                                                                                                                    |
| Category*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                        | Relevant to claim No.                                                                                              |
| <b>Y</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>US 2017/010966 A1 (MITTAL MILLIND [US])</b><br><b>12 January 2017 (2017-01-12)</b>                                                                                                     | <b>1-4, 6,</b><br><b>9-11, 13,</b><br><b>14, 18-20</b>                                                             |
| <b>A</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>paragraphs [0007] - [0010], [0017] -</b><br><b>[0060]; claims 1-18; figures 1-7</b><br>-----                                                                                           | <b>5, 7, 8,</b><br><b>12, 15-17</b>                                                                                |
| <b>Y</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>US 2007/174555 A1 (BURTSCHER MARTIN [US])</b><br><b>ET AL) 26 July 2007 (2007-07-26)</b>                                                                                               | <b>1-4, 6,</b><br><b>9-11, 13,</b><br><b>14, 18-20</b>                                                             |
| <b>A</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>paragraphs [0013] - [0026] - paragraphs</b><br><b>[0034] - [0052]; claims 1-7; figures 1-3</b><br>-----                                                                                | <b>5, 7, 8,</b><br><b>12, 15-17</b>                                                                                |
| <b>A</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <b>EP 3 550 437 A1 (INTEL CORP [US])</b><br><b>9 October 2019 (2019-10-09)</b><br><b>paragraphs [0129] - [0140]; figures 15-18</b><br>-----<br><div style="text-align: center;">-/-</div> | <b>1-20</b>                                                                                                        |
| <div style="display: flex; justify-content: space-between;"> <div> <input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.         </div> <div> <input checked="" type="checkbox"/> See patent family annex.         </div> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                           |                                                                                                                    |
| <div style="display: flex;"> <div style="flex: 1;"> <p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="flex: 1;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p> </div> </div> |                                                                                                                                                                                           |                                                                                                                    |
| Date of the actual completion of the international search<br><br><div style="text-align: center;"><b>8 March 2022</b></div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                           | Date of mailing of the international search report<br><br><div style="text-align: center;"><b>18/03/2022</b></div> |
| Name and mailing address of the ISA/<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                           | Authorized officer<br><br><div style="text-align: center;"><b>Jardon, Stéphan</b></div>                            |

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/GB2021/053222

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                    |                       |
|------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                 | Relevant to claim No. |
| A                                                    | WO 2019/085649 A1 (HUAWEI TECH CO LTD [CN]) 9 May 2019 (2019-05-09)<br>paragraphs [0004], [0013], [0052] - [0089] - paragraphs [0113] - [0140];<br>figures 1-3<br>-----            | 1-20                  |
| A                                                    | US 2020/334153 A1 (ZHANG CHUNHUI [US] ET AL) 22 October 2020 (2020-10-22)<br>paragraphs [0099] - [0100], [0109] - [0154]; claims 1-20; figure 4<br>-----                           | 1-20                  |
| A                                                    | WO 2019/211610 A1 (ADVANCED RISC MACH LTD [GB]) 7 November 2019 (2019-11-07)<br>paragraphs [0002] - [0003], [0015] - [0017] - paragraphs [0026] - [0030];<br>figures 1, 3<br>----- | 1-20                  |
| A                                                    | US 2019/129871 A1 (CARRO ALEJANDRO RICO [US] ET AL) 2 May 2019 (2019-05-02)<br>paragraphs [0025] - [0029], [0035] - [0045], [0046] - [0050]; figures 1, 2, 5, 6<br>-----           | 1-20                  |
| A                                                    | US 2004/162949 A1 (SCOTT STEVEN L [US] ET AL) 19 August 2004 (2004-08-19)<br>paragraphs [0007], [0013] - [0073];<br>claims 1-22; figures 1, 3-5<br>-----                           | 1-20                  |
| A                                                    | US 2014/351519 A1 (MUNOZ ROBERT J [US]) 27 November 2014 (2014-11-27)<br>paragraphs [0016] - [0023]; claims 1-20;<br>figures 1, 2<br>-----                                         | 1-20                  |
| A                                                    | US 2019/114262 A1 (DIMOND ROBERT GWILYM [GB] ET AL) 18 April 2019 (2019-04-18)<br>paragraphs [0031] - [0033], [0057] - [0062]; claims 1-10; figures 1, 2A, 2B, 3, 4A, 4B<br>-----  | 1-20                  |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2021/053222

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s)                                                                                   | Publication<br>date                                                              |
|-------------------------------------------|---------------------|--------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| US 2017010966 A1                          | 12-01-2017          | US 2017010966 A1<br>WO 2017011021 A1                                                                         | 12-01-2017<br>19-01-2017                                                         |
| US 2007174555 A1                          | 26-07-2007          | NONE                                                                                                         |                                                                                  |
| EP 3550437 A1                             | 09-10-2019          | CN 110347431 A<br>EP 3550437 A1<br>US 2019310853 A1                                                          | 18-10-2019<br>09-10-2019<br>10-10-2019                                           |
| WO 2019085649 A1                          | 09-05-2019          | CN 109753445 A<br>WO 2019085649 A1                                                                           | 14-05-2019<br>09-05-2019                                                         |
| US 2020334153 A1                          | 22-10-2020          | CN 110865968 A<br>US 2020334153 A1                                                                           | 06-03-2020<br>22-10-2020                                                         |
| WO 2019211610 A1                          | 07-11-2019          | CN 112106032 A<br>EP 3788495 A1<br>JP 2021522611 A<br>KR 20210005193 A<br>US 10452593 B1<br>WO 2019211610 A1 | 18-12-2020<br>10-03-2021<br>30-08-2021<br>13-01-2021<br>22-10-2019<br>07-11-2019 |
| US 2019129871 A1                          | 02-05-2019          | EP 3704589 A1<br>US 2019129871 A1<br>WO 2019086839 A1                                                        | 09-09-2020<br>02-05-2019<br>09-05-2019                                           |
| US 2004162949 A1                          | 19-08-2004          | US 2004162949 A1<br>US 2006248286 A1                                                                         | 19-08-2004<br>02-11-2006                                                         |
| US 2014351519 A1                          | 27-11-2014          | NONE                                                                                                         |                                                                                  |
| US 2019114262 A1                          | 18-04-2019          | GB 2567466 A<br>US 2019114262 A1                                                                             | 17-04-2019<br>18-04-2019                                                         |