

CHIP AND INTER-CORE DATA TRANSMISSION METHOD

TECHNICAL FIELD

The disclosure relates to the technical field of chips, in particular, relates to a chip
5 and an inter-core data transmission method.

BACKGROUND

With the development of science and technology, human society is rapidly entering
an era of intelligence. The important feature of the era of intelligence is that people may
acquire more and more types of data, the amount of data that people may acquire is
10 becoming larger and larger, and the demand for the data processing speed is becoming
higher and higher.

Chips are the cornerstone of data processing, which essentially determine the
ability to process data. From the perspective of application fields, the chips are mainly
developed in two routes. One is general-purpose chips, for example, central processing
15 units (CPU) and the like, which can provide great flexibility but have low effective
computing power when processing algorithms in a specific field; the other is special-
purpose chips, for example, tensor processing units (TPU) and the like, which have
high effective computing power in certain specific fields but are poor in processing
capacity or even have no processing capacity in flexible and universal fields.

20 Due to various data types and numerous data volumes in the era of intelligence,
the chips are required to not only process changing algorithms in different fields with
extremely high flexibility, but also rapidly process a great number of dramatically
increasing data volumes with extremely strong processing capacities.

SUMMARY

25 1. Objective

The present disclosure aims to provide a chip and an inter-core data transmission
method. The chip provided by an embodiment of the present disclosure includes a
plurality of interconnection structures, and data transmission among processing cores
in the same processing core group is realized through the interconnection structure
30 connected to the processing core group, so that the inter-core data transmissions of

different core groups can be processed in parallel, greatly improving the total data bandwidth and improving the performance of the chip.

2. Technical Solution

To solve the above problems, a first aspect of the present disclosure provides a
5 chip, including: a plurality of interconnection structures, wherein each of the
interconnection structures is connected to a processing core group, and all processing
cores connected to one interconnection structure are one processing core group;
wherein the processing cores belonging to a same processing core group perform data
transmission therebetween through the interconnection structure connected to the
10 processing core group; wherein, the plurality of the interconnection structures are not
directly connected to one another.

The chip provided by the embodiment of the present disclosure includes a plurality
of interconnection structures, and data transmission among the processing cores in the
same processing core group is realized through the interconnection structure connected
15 to the processing core group, so that the inter-core data transmissions of the different
processing core groups can be processed in parallel, greatly improving the total data
bandwidth and improving the performance of the chip.

Further, data is capable of being transmitted among processing cores belonging to
different processing core groups.

20 Further, amount of data transmitted between any two processing cores belonging
to one processing core group is greater than amount of data transmitted between
processing cores belonging to any two different processing core groups.

Further, the plurality of interconnection structures include a first interconnection
structure and a second interconnection structure; the first interconnection structure is
25 connected to a first processing core; and the first processing core is connected to the
second interconnection structure.

Further, the first interconnection structure is further connected to a second
processing core; the second interconnection structure is connected to a third processing
core; and the second processing core performs data transmission with the third
30 processing core through the first processing core.

Further, the plurality of interconnection structures include a third interconnection
structure; and all processing cores connected to the third interconnection structure are
not connected to the other interconnection structures.

Further, the processing cores connected to each interconnection structure are not connected to the other interconnection structures.

Further, each of the processing cores includes at least one transmission unit, and bandwidth of the interconnection structure meets bandwidth requirement of the transmission unit.

Further, the bandwidth of the interconnection structure meets bandwidth requirement of each of the processing cores in the processing core group connected to the interconnection structure.

Further, the bandwidth of the interconnection structure is greater than bandwidth requirement of each of the processing cores in the processing core group connected to the interconnection structure.

Further, the plurality of the interconnection structures are provided with at least two clock frequencies.

Further, the plurality of the interconnection structures are provided with at least two bit-width values.

Further, at least two processing core groups have different bandwidth requirements.

According to a second aspect of the present disclosure, there is provided a circuit board, including one or more chips provided by the first aspect.

According to a third aspect of the present disclosure, there is provided an electronic device, including one or more circuit boards provided by the second aspect.

According to a fourth aspect of the present disclosure, there is provided an inter-core data transmission method, applicable to the chip provided by the first aspect, the method including performing data transmission among the processing cores belonging to the same processing core group through the interconnection structure connected to the processing core group; and performing data transmission between two processing cores belonging to two different processing core groups respectively through at least two of the interconnection structures.

Further, performing data transmission between the two processing cores belonging to two different processing core groups respectively through at least two of the interconnection structures, includes: performing data transmission between the two processing cores belonging to two different processing core groups respectively through two interconnection structures via a processing core that belongs to both of the two different processing core groups.

According to a fifth aspect of the present disclosure, there is provided a computer storage medium storing a computer program, which, upon execution by a processor, implements the steps of the inter-core data transmission method of the fourth aspect.

5 According to a sixth aspect of the present disclosure, there is provided an electronic device, including a memory, a processor, and a computer program stored in the memory and executable by the processor, wherein the processor executing the program implements the steps of the inter-core data transmission method of the fourth aspect.

10 According to a seventh aspect of the present disclosure, there is provided a computer program product, including computer instructions, wherein, when the computer instructions are executed by a computing device, the computing device performs the steps of the inter-core data transmission method of the fourth aspect.

3. Beneficial Effects

15 The technical solution of the present disclosure has the following beneficial technical effects:

(1) The chip provided by the embodiment of the present disclosure is provided with a plurality of interconnection structures, and the data transmission among the processing cores of the same processing core group is realized through the interconnection structure connected to the processing core group, so that the inter-core data transmissions of different processing core groups can be processed in parallel. On the one hand, the total data bandwidth is greatly improved, the data transmission efficiency among cores is improved, the power consumption is reduced, and the performance of the chip is improved; on the other hand, since the data transmissions of different groups can be processed in parallel, the data transmissions of the different groups will not interfere with each other, thus avoiding the occurrence of data congestion and improving the performance of the chip.

20 (2) In the chip provided by the embodiment of the present disclosure, when data is transmitted between cores belonging to the same group, only the data formats of the cores in the same group need to be considered, without the need to consider the data formats of other cores in other groups of the chip, reducing the complexity of the circuitry in the chip.

30 (3) Since an interconnection structure is used for data transmission within a processing core group, and different interconnection structures can be set with different

clock frequencies and different bit widths, the corresponding bit width may be set according to the requirements of data transmission of each group. Compared with all cores in a chip using one interconnection structure for data transmission, this reduces power consumption, improves the performance of the entire chip, and increases the flexibility of the chip design, and it can reduce the chip area by determining the appropriate bit widths and frequencies of the interconnection structures according to the actual bandwidth requirements of the different processing core groups.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic structural diagram of a type of chip.

FIG. 2 is a structural diagram of a chip according to an embodiment of the present disclosure.

FIG. 3 is a structural diagram of a chip according to an embodiment of the present disclosure.

FIG. 4 is a schematic flowchart of an inter-core data transmission method according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

In order to make the objective, technical solution and advantages of the present disclosure clearer, the present disclosure is further described in detail below in conjunction with specific embodiments and with reference to the accompanying drawings. It should be understood that these descriptions are exemplary only and are not intended to limit the scope of the present disclosure. In addition, in the following illustration, descriptions of well-known structures and techniques are omitted to avoid unnecessary confusion about the concepts of the present disclosure.

In a multi-core (or many-core) chip, all cores may be isomorphic, that is, each core is of the same structure, and heterogeneous cores may also exist in all cores, that is, there are at least two types of cores with different structures. In a multi-core (or many-core) architecture, there are often multiple cores involving in the execution of a certain task at the same time. At this time, there is data transmission among the cores. Therefore, for a multi-core chip, which architecture to use and how to perform data transmission, such that the homogeneous or heterogeneous cores methodically form a chip with excellent performance, are crucial.

FIG. 1 is a schematic structural diagram of a type of chip.

In the chip structure shown in FIG. 1, the chip has only one network on chip, and all cores are connected with each other through a network on chip (NoC) to exchange data.

5 As shown in FIG. 1, the chip is provided with cores of various structures, including $K_1_C_1$ 、 $K_2_C_1$ and $K_M_C_1$, where M denotes that there are M types of cores. There are multiple cores for each type. For example, the first type of cores K_1 includes $K_1_C_1$, $K_1_C_2$ and $K_1_C_{NM}$, where NM denotes that the number of cores of the type is NM. That is, $K_M_C_{NM}$ represents the NMth core of the Mth structure.

10 Since there are many types of cores (M types) in the chip, and different types of cores use one same NoC for data exchange, each core has to know the data formats (bit widths) of all the other cores that need to exchange data. Since the same NoC is used among the cores, there are relatively high requirements for the format of the data packets sent or received, that is, the sending core encodes the data format into the data
15 format of the receiving core, or the receiving core decodes the data and re-encodes it into the data format of the receiving core according to the type of the sending core. This results in high complexity of the data packet routing algorithm and heavy burden on each core.

 In addition, since all cores share one NoC, due to the different amounts of data
20 exchanged among cores, when a pair of cores are performing data transmission, the other cores cannot use the NoC for data transmission, and they have to wait until the pair of cores have transmitted the data to perform data transmission. This results in a delay in data transmission, and leads to the degradation of the chip performance. For example, the bandwidth of data transmission between the first core and the second core
25 is very large, and during the data transmission between the first and second cores, the third core generates data that requires to be sent to the fourth core. The amount of this data is relatively small, but it has high real-time requirements. Since all cores share one NoC, the third core needs to wait until the data transmission between the first and second cores is completed, and then performs data transmission. In such cases, the data
30 with high real-time requirements is not transmitted in a timely manner, affecting the performance of the chip.

 In order to solve the above problems, the technical solution of the present disclosure is provided.

The chip provided by an embodiment of the present disclosure will be described in detail below. In the description of the present disclosure, it should be noted that the terms "first", "second", "third" and "fourth" are only used for descriptive purposes and cannot be construed as indicating or implying relative importance. In addition, the technical features involved in the different embodiments of the present disclosure described below can be combined with each other as long as they do not conflict with each other.

FIG. 2 is a structural diagram of a chip according to an embodiment of the present disclosure.

As shown in FIG. 2, the chip includes: a plurality of interconnection structures, wherein each of the interconnection structures is connected to a processing core group, and all processing cores connected to one interconnection structure are one processing core group; wherein the processing cores belonging to a same processing core group perform data transmission therebetween through the interconnection structure connected to the processing core group; wherein, the plurality of the interconnection structures are not directly connected to one another.

In the embodiment shown in FIG. 2, the chip includes o interconnection structures, where the interconnection structure may be an inter-core interconnection structure Fabric, a network on chip (NoC), a bus, or a switch. The interconnection structure of the present disclosure takes the network on chip (NoC) as an example, but is not limited thereto. The o interconnection structures are NoC₁, NoC₂ ... NoC _{o} , respectively, and each NoC connects to a processing core group. The processing cores in the processing core group perform data transmission therebetween through the NoC connected to the processing core group.

The chip includes M types of cores, and each type of cores has NM cores. For example, K₂_C_{N2} denotes the N₂th core of the second type of the cores, and K _{M} _C_{NM} denotes the NMth core of the Mth type of the cores.

In the embodiment shown in FIG. 2, the first processing core group includes core K₁_C₁, core K₁_C₂...core K₁_C_{N1}, core K₂_C₁, and core K₂_C₂. The first network on chip NoC₁ connects to the first processing core group, i.e., each processing core in the first processing core group is connected to NoC₁. The data transmission among the cores of the first processing core group is realized through the first network on chip NoC₁. For example, when core K₁_C₁ has data that needs to be transmitted to core

K₁_C₂, core K₁_C₁ transmits the data to the first network on chip NoC₁, which then transmits the data to core K₁_C₂.

The second processing core group includes: core K₁_C₁, core K₁_C_{N1}, core K₂_C₂...core K₂_C_{N2}, core K_M_C₁, and core K_M_C₂. Each processing core in the second
5 processing core group is connected to the second network on chip NoC₂. The data transmission among the cores of the second processing core group is realized through the second network on chip NoC₂.

The oth processing core group includes core K₁_C_{N1}, core K₂_C₂, core K_M_C₂...core K_M_C_{NM}. Each core in the oth processing core group is connected to the
10 oth network on chip NoC_o. The data transmission among the cores of the oth processing core group is realized through the oth network on chip NoC_o.

The chip provided by the embodiment of the present disclosure is provided with a plurality of interconnection structures, and each of the interconnection structures is connected to a processing core group, and data transmission among the processing cores
15 in the same processing core group is realized through the interconnection structure connected to the processing core group, so that the inter-core data transmissions of different groups can be processed in parallel. On the one hand, when data is transmitted between cores belonging to the same group, only the data formats of the cores in the same group need to be considered, without the need to consider the data formats of
20 other cores in other groups of the chip, which reduces the complexity of the circuitry in the chip, improves the data transmission efficiency among cores, reduces the power consumption, and improves the performance of the chip; on the other hand, since the data transmissions of different groups can be processed in parallel, the data transmissions of the different groups will not interfere with each other, thus avoiding
25 the occurrence of data congestion and improving the performance of the chip.

In an embodiment, data is capable of being transmitted among processing cores belonging to different processing core groups.

In a preferred embodiment, the plurality of interconnection structures include a first interconnection structure and a second interconnection structure; the first
30 interconnection structure is connected to a first processing core; and the first processing core is connected to the second interconnection structure.

Specifically, in the embodiment, the same processing core may be provided in two different processing core groups at the same time. For example, in the embodiment

shown in FIG. 2, the processing core $K_1_C_1$ is set in both the first processing core group and the second processing core group, i.e., the processing core $K_1_C_1$ is connected to both the first network on chip NoC_1 and the second network on chip NoC_2 .

5 In an embodiment, the amount of data transmitted between any two processing cores belonging to the same processing core group is greater than the amount of data transmitted between processing cores belonging to any two different processing core groups.

Specifically, in the embodiment, if there is a relatively large amount of data transmission between certain two cores, the two cores will be assigned to the same group.
10

In an embodiment, the first interconnection structure is further connected to a second processing core; the second interconnection structure is connected to a third processing core; and the second processing core performs data transmission with the third processing core through the first processing core.

15 Specifically, in the embodiment, if data needs to be transmitted between two cores that belong to two groups respectively, generally speaking, the amount of data to be transmitted between these two cores is relatively small. If there is only a small amount of data to be exchanged between certain two cores, the data may be exchanged through some bridging transit cores, i.e., data transmission is achieved through the cores located in both groups at the same time. For example, core $K_2_C_1$ connected to NoC_1 needs to send data to core $K_M_C_1$ connected to NoC_2 . Since the two cores are not in one processing core group, $K_2_C_1$ may perform the transmission through a core that is connected to both NoC_1 and NoC_2 , such as $K_1_C_1$ or $K_2_C_2$. That is, $K_2_C_1$ sends data to $K_1_C_1$ through NoC_1 , and $K_1_C_1$ sends the data to $K_M_C_1$ through NoC_2 , thereby
20 realizing data transmission.
25

For example, in the embodiment shown in FIG. 2, core $K_1_C_2$ of the first processing core group needs to send data to core $K_M_C_2$ of the second processing core group, it may be achieved by processing cores $K_1_C_1$, $K_1_C_{N1}$, or $K_2_C_2$.

Specifically, the processing core $K_1_C_2$ sends data to the processing core $K_1_C_1$ through the first network on chip NoC_1 , and the processing core $K_1_C_1$ sends the data to the processing core $K_M_C_2$ through the second network on chip NoC_2 .
30

In an embodiment, the plurality of interconnection structures include a third interconnection structure; and all processing cores connected to the third interconnection structure are not connected to the other interconnection structures.

For example, in the embodiment shown in FIG. 2, the chip further includes the
5 third network on chip, and the third network on chip is connected to a third processing core group, which includes core $K_1_C_3$ and core $K_3_C_1$. These two cores are not connected to the other interconnection structures.

In an embodiment of the present disclosure, the processing cores connected to each interconnection structure in the chip are not connected to the other interconnection
10 structures. In the embodiment, the processing cores connected to each interconnection structure are not connected to the other interconnection structures, i.e., no data is transmitted between two processing cores belonging to two processing core groups respectively.

In an embodiment, each of the processing cores includes at least one transmission
15 unit, and the bandwidth of the interconnection structure meets the bandwidth requirement of the transmission unit. For example, the bandwidth of the interconnection structure is equal to the bandwidth requirement of the transmission unit, so that the processing core is able to send data to the interconnection structure.

The bandwidth requirement of the transmission unit may be the bandwidth
20 required by the transmission unit to send a certain amount of data within a certain period of time.

In an embodiment, the bandwidth of the interconnection structure meets the bandwidth requirement of each of the processing cores in the processing core group connected to the interconnection structure.

25 The bandwidth requirement of the processing core may be the bandwidth required by the processing core to send a certain amount of data within a certain period of time.

Preferably, the bandwidth of the interconnection structure is greater than the bandwidth requirement of each of the processing cores in the processing core group connected to the interconnection structure. In this way, each interconnection structure
30 can best match the structures of all processing cores in the processing core group connected to the interconnection structure, which, on the one hand, maximizes the area saved, and on the other hand, makes the chip have better performance.

In an embodiment, the bandwidth requirements of at least two processing core groups in the chip are different. For example, the bandwidth requirement of the first group is 1MBps, and the bandwidth requirement of the second group is 10MBps. In this way, the Noc having a corresponding structure may be provided based on the different requirements of the processing core groups, and thus reducing the area of the Noc to the greatest extent.

FIG. 3 is a schematic structural diagram of a chip according to an embodiment of the present disclosure.

As shown in FIG. 3, the chip includes 6 cores, which include 3 types of cores with different structures, namely the first type of cores K_1 , the second type of cores K_2 and the third type of cores K_3 . Each type has two cores.

The two cores of the first type of the cores K_1 are core K_{1_C1} and core K_{1_C2} ; the two cores of the second type of the cores K_2 are core K_{2_C1} and core K_{2_C2} ; and the two cores of the third type of the cores K_3 are core K_{3_C1} and core K_{3_C2} .

The six cores in the chip are divided into three groups.

The first processing core group consists of core K_{1_C1} , core K_{1_C2} , core K_{2_C1} and core K_{2_C2} . Each core in the first processing core group is connected to the first interconnection structure NoC_1 .

The second processing core group consists of K_{1_C1} , K_{2_C2} , K_{3_C1} , and K_{3_C2} . Each core in the second processing core group is connected to the second interconnection structure NoC_2 .

The third processing core group consists of K_{1_C2} , K_{2_C2} , and K_{3_C2} . Each core in the third processing core group is connected to the third interconnection structure NoC_3 .

In a preferred embodiment, each core in the chip includes at least one computing unit.

The computing unit may be an execution unit (EU), or a processing unit (PU).

The chip includes a plurality of interconnection structures, and each interconnection structure connects to a processing core group. The processing cores in the same group performs data transmission therebetween through the interconnection structure connected to the processing core group. The interconnection structure will design data packets and corresponding protocols dedicated to its corresponding group

of cores based on the bandwidth required by the cores in the group and the data bit width used by the cores.

In an embodiment, the plurality of the interconnection structures in the chip are provided with at least two clock frequencies. Since the chip is provided with the plurality of the interconnection structures, each interconnection structure connecting to a processing core group, and the data transmission among the processing cores in each processing core group may be processed in parallel, the plurality of the interconnection structures in the chip can be provided with different clock frequencies.

Optionally, the plurality of the interconnection structures in the chip may be provided with the same clock frequency.

In an embodiment, the plurality of the interconnection structures in the chip are provided with at least two bit-width values. Since the chip is provided with the plurality of the interconnection structures, each interconnection structure connecting to a processing core group, and the data transmission among the processing cores in each processing core group may be processed in parallel, the plurality of the interconnection structures in the chip can be provided with different bit-width values.

Optionally, the plurality of the interconnection structures in the chip may be provided with the same bit-width value.

It should be noted that $\text{bandwidth} = \text{bit width} * \text{clock frequency}$. If a higher bandwidth is required for data transmission among the processing cores of a processing core group, the bandwidth of the NoC can be determined (for example, determining the appropriate clock frequency and bit width) according to the bandwidth requirements of the processing core group, so that the bandwidth of the NoC meets the bandwidth requirement of each processing core in the processing core group. For example, a higher bit width of the NoC may be set. In addition, since the data transmissions of the processing core groups are independent of each other, if the other processing core groups do not need higher bandwidth, the appropriate bit width of the NoC can be set according to their needs, which can reduce the area of the chip, and thus saves power consumption.

Further, in the chip according to the embodiment of the present disclosure, since an NoC is used for data transmission within one processing core group, and different NoCs can be set with different clock frequencies and different bit widths, the corresponding bit width can be set according to the requirements of data transmission

of each group. Compared with all cores in a chip using one interconnection structure for data transmission, this reduces power consumption, improves the performance of the entire chip, increases the flexibility of the chip design, and may reduce chip area.

5 The technical solution of the present disclosure has the following beneficial technical effects:

(1) The chip provided by the embodiment of the present disclosure is provided with a plurality of interconnection structures, and the data transmission within the same processing core group is realized through the interconnection structure connected to the processing core group, so that the inter-core data transmissions of different processing
10 core groups can be processed in parallel. On the one hand, the total data bandwidth is greatly improved, the data transmission efficiency among cores is improved, the power consumption is reduced, and the performance of the chip is improved; on the other hand, since the data transmissions of different groups can be processed in parallel, the data transmissions of the different groups will not interfere with each other, thus avoiding
15 the occurrence of data congestion and improving the performance of the chip.

(2) In the chip provided by the embodiment of the present disclosure, when data is transmitted between cores belonging to the same group, only the data formats of the cores in the same group need to be considered, without the need to consider the data formats of other cores in other groups of the chip, reducing the complexity of the
20 circuitry in the chip.

An embodiment of the present disclosure further provides a circuit board, including one or more chips provided by the above embodiments.

An embodiment of the present disclosure further provides an electronic device, including one or more circuit boards provided by the above embodiments.

25 FIG. 4 is a schematic flowchart of an inter-core data transmission method according to an embodiment of the present disclosure.

As shown in FIG. 4, the method includes step S101~step S102:

Step S101: performing data transmission among the processing cores belonging to the same processing core group through the interconnection structure connected to the
30 processing core group.

Step S102: performing data transmission between two processing cores belonging to two different processing core groups respectively through at least two of the interconnection structures.

In an embodiment, the step of performing data transmission between two processing cores belonging to two different processing core groups respectively through at least two of the interconnection includes:

performing data transmission between the two processing cores belonging to two
5 different processing core groups respectively through two interconnection structures
via a processing core that belongs to both of the two different processing core groups.

For example, in the embodiment shown in FIG. 3, core K₂_C₁ connected to NoC₁
needs to send data to core K₃_C₁ connected to NoC₂. Since the two cores are not in one
processing core group, K₂_C₁ may perform the transmission through a core that is
10 connected to both NoC₁ and NoC₂, such as K₁_C₁ or K₂_C₂. That is, K₂_C₁ sends data
to K₁_C₁ through NoC₁, and K₁_C₁ sends the data to K₃_C₁ through NoC₂, thereby
realizing data transmission.

It should be noted that data transmission between two processing cores belonging
to two different processing core groups respectively may be performed through the
15 plurality of the interconnection structures.

For example, in the embodiment shown in FIG. 3, the processing core K₂_C₁ may
send data to core K₃_C₁ through NoC₁, NoC₂ and NoC₃. Specifically, the processing
core K₂_C₁ sends data to the processing core K₁_C₂ through the NoC₁, the processing
core K₁_C₂ sends the data to the processing core K₃_C₂ through NoC₃, and the
20 processing core K₃_C₂ sends the data to the processing core K₃_C₁ through NoC₂.

In the inter-core data transmission method provided by the embodiment of the
present disclosure, the processing cores belonging to the same processing core group
perform data transmission therebetween through the interconnection structure
connected to the processing core group, so that the inter-core data transmissions of
25 different processing core groups can be processed in parallel. On the one hand, the total
data bandwidth is greatly improved, the data transmission efficiency among cores is
improved, the power consumption is reduced, and the performance of the chip is
improved; on the other hand, since the data transmissions of different groups can be
processed in parallel, the data transmissions of the different groups will not interfere
30 with each other, thus avoiding the occurrence of data congestion and improving the
performance of the chip.

According to an embodiment of the present disclosure, there is provided a
computer storage medium storing a computer program, which, upon execution by a

processor, implements the steps of the inter-core data transmission method provided by the above embodiments.

According to an embodiment of the present disclosure, there is provided an electronic device, including a memory, a processor, and a computer program stored in the memory and executable by the processor, wherein the processor executing the program implements the steps of the inter-core data transmission method provided by the above embodiments.

According to an embodiment of the present disclosure, there is provided a computer program product, including computer instructions, wherein, when the computer instructions are executed by a computing device, the computing device performs the steps of the inter-core data transmission method provided by the above embodiments.

The flowcharts and block diagrams in the accompanying drawings of the present disclosure show the possible architecture, functions, and operations of a system, a method, and a computer program product according to various embodiments of the present disclosure. In this regard, each block in the flowcharts or block diagrams may represent a module, a program segment or a part of codes, and the module, the program segment or the part of the codes contains one or more executable instructions for implementing the defined logical functions. It should also be noted that in some implementations as alternatives, the functions labeled in the blocks may occur in an order different from the order labeled in the accompanying drawings. For example, two sequentially shown blocks may be substantially executed in parallel in fact, and they sometimes may also be executed in a reverse order, depending on related functions. It should also be noted that each block in the block diagrams and/or the flowcharts and the combination of the blocks in the block diagrams and/or the flowcharts may be implemented by a dedicated system based on hardware for executing defined functions or operations, or may be implemented by a combination of the dedicated hardware and computer instructions.

The units described in the embodiments of the present disclosure may be implemented in a software fashion or may be implemented in a hardware fashion. The names of the units do not constitute a limitation to the units in some cases.

The functions described above herein may be performed, at least in part, by one or more hardware logic components. For example, non-restrictively, exemplary types

of hardware logic components that can be used include: a field programmable gate array (FPGA), an application-specific integrated circuit (ASIC), an application-specific standard product (ASSP), a system on chip (SOC), a complex programmable logic device (CPLD), and the like.

5 In the context of the present disclosure, a machine-readable medium may be a tangible medium that may contain or store a program for use by or in conjunction with an instruction execution system, apparatus or device. The machine-readable medium may be a machine-readable signal medium or a machine-readable storage medium. The machine-readable medium may include, but not limited to, an electronic, magnetic,
10 optical, electromagnetic, infrared or semiconductor system, apparatus or device, or any suitable combination of the foregoing. A more specific example of the machine-readable storage medium includes an electrical connection based on one or more wires, a portable computer disk, a hard disk, a random-access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (an EPROM or a flash
15 memory), an optical fiber, a portable compact disk read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the above content.

It should be understood that the above specific embodiments of the present disclosure are used only for exemplary illustration or explanation of the principle of the
20 present disclosure, but not to limit the present disclosure. Therefore, any modifications, equivalent replacements, improvements, etc. made without departing from the spirit and scope of the present disclosure should be included within the protection scope of the present disclosure. Furthermore, the appended claims of the present disclosure are intended to cover all variations and modifications that fall within the scope and
25 boundaries of the appended claims, or the equivalents of such scope and boundaries.

The present disclosure has been described above with reference to the embodiments of the present disclosure. However, these examples are for illustrative purposes only, and are not intended to limit the scope of the present disclosure. The scope of the disclosure is defined by the appended claims and their equivalents. Without
30 departing from the scope of the present disclosure, those skilled in the art can make various substitutions and modifications, and these substitutions and modifications should all fall within the scope of the present disclosure.

Obviously, the above embodiments are only examples for clear illustration, and are not intended to limit the manner of implementation. For those of ordinary skill in the art, there are other variations or changes that can be made in different forms based on the above description. It is not necessary or possible to exhaust all embodiments
5 here. The obvious variations or changes derived therefrom are still within the protection scope of the present disclosure.

CLAIMS

1. A chip, comprising: a plurality of interconnection structures, wherein each of the interconnection structures is connected to at least two processing cores, and all processing cores connected to one interconnection structure are one processing core group;
- 5 wherein the processing cores belonging to a same processing core group perform data transmission therebetween through the interconnection structure connected to the processing core group;
- wherein, the plurality of the interconnection structures are not directly connected
- 10 to one another.
2. The chip according to claim 1, wherein data is capable of being transmitted among processing cores belonging to different processing core groups.
3. The chip according to claim 2, wherein
- amount of data transmitted between any two processing cores belonging to one
- 15 processing core group is greater than amount of data transmitted between processing cores belonging to any two different processing core groups.
4. The chip according to any of claim 1-3, wherein
- the plurality of interconnection structures include a first interconnection structure and a second interconnection structure;
- 20 the first interconnection structure is connected to a first processing core; and
- the first processing core is connected to the second interconnection structure.
5. The chip according to claim 4, wherein
- the first interconnection structure is further connected to a second processing core;
- the second interconnection structure is connected to a third processing core; and
- 25 the second processing core performs data transmission with the third processing core through the first processing core.
6. The chip according to claim 4 or 5, wherein
- the plurality of interconnection structures include a third interconnection structure;
- and
- 30 all processing cores connected to the third interconnection structure are not connected to the other interconnection structures.

7. The chip according to any one of claims 1-6, wherein bandwidth of the interconnection structure meets bandwidth requirement of each of the processing cores in the processing core group connected to the interconnection structure.

5 8. The chip according to claim 7, wherein each of the processing cores comprises at least one transmission unit, and bandwidth of the interconnection structure meets bandwidth requirement of the transmission unit of each processing core connected to the interconnection structure.

9. The chip according to any one of claims 1-8, wherein the plurality of the interconnection structures are provided with at least two clock frequencies; and/or
10 the plurality of the interconnection structures are provided with at least two bit-width values.

10. A circuit board, comprising one or more chips according to any one of claims 1-9.

11. An electronic device, comprising one or more circuit boards according to claim
15 10.

12. An inter-core data transmission method, applicable to the chip according to any one of claims 1-9, comprising:

performing data transmission among the processing cores belonging to the same processing core group through the interconnection structure connected to the processing
20 core group; and

performing data transmission between two processing cores belonging to two different processing core groups respectively through at least two of the interconnection structures.

13. The method according to claim 12, wherein performing data transmission
25 between the two processing cores belonging to two different processing core groups respectively through at least two of the interconnection structures, comprises:

performing data transmission between the two processing cores belonging to two different processing core groups respectively through two interconnection structures via a processing core that belongs to both of the two different processing core groups.

30 14. A computer storage medium storing a computer program, which, upon execution by a processor, implements the steps of the inter-core data transmission method according to claim 12 or 13.

15. An electronic device, comprising a memory, a processor, and a computer program stored in the memory and executable by the processor, wherein the processor executing the program implements the steps of the inter-core data transmission method according to claim 12 or 13.

- 5 16. A computer program product, comprising computer instructions, wherein, when the computer instructions are executed by a computing device, the computing device performs the steps of the inter-core data transmission method according to claim 12 or 13.