PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/EP2020/087570	22.12.2020	15.01.2020

International Patent Classification (IPC) or both national classification and IPC
INV. G06F15/173 G06N3/063

Applicant
GRAPHCORE LIMITED

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☒ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk - Pays Bas Tel. +31 70 340 - 2040 Fax: +31 70 340 - 3016	see form PCT/ISA/210	De Poy, Iker Telephone No. +31 70 340-0

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/EP2020/087570

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	2-6, 8-17
	No:	Claims	1, 7, 18
Inventive step (IS)	Yes:	Claims	
	No:	Claims	1-18
Industrial applicability (IA)	Yes:	Claims	1-18
	No:	Claims	

2. Citations and explanations

see separate sheet

Box No. VII Certain defects in the international application

The following defects in the form or contents of the international application have been noted:

see separate sheet

1 **Re Item V**

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

Reference is made to the following documents:

D1 CN 109 409 512 A (UNIV XI AN JIAOTONG) 1 March 2019 (2019-03-01)

D2 US 2018/322386 A1 (SRIDHARAN SRINIVAS [IN] ET AL) 8 November 2018 (2018-11-08)

In this communication the machine translation of D1 into English, which is annexed, is used.

The subject-matter of claims 1 to 18 addresses a problem in the field of computer architecture and therefore claims 1 to 18 have industrial applicability.

1.1 **Independent claims**

The present application does not meet the criteria of Article 33(1) PCT, because the subject-matter of claims 1 and 18 is not new in the sense of Article 33(2) PCT.

1.1.1 Independent claim 1

Document D1 is considered to be the prior art closest to the subject-matter of claim 1, and discloses (the references in parenthesis applying to document D1):

> **A data processing system** (figure 4) **comprising a plurality of processors** (figure 4, computing array comprising PEs)**, wherein each of the processors comprises at least one circuit configured to perform data transfer operations during each of at least some of a plurality of exchange stages to transfer data determined in dependence upon data received at the respective processor in a preceding one of the exchange stages from at least one other of the processors, each of the data transfer operations being for transfer of data to another one of the plurality of processors** (paragraph [0035], "A computing array is generated by instantiating a plurality of configurable computing units, and the computing array is divided into regions, and different regions can provide different convolution layer parameters to complete parallel computing of different types of convolution modes"")**, wherein each at least one circuit is configured to:**

perform data transfer operations to transfer outgoing data to one or more others of the processors during a first of the exchange stages;

receive incoming data from the one or more others of the processors during the first of the exchange stages;

determine further outgoing data in dependence upon at least part of the incoming data;

(paragraph [0057], "Further, according to hardware resources and the computing performance requirements of the system, multiple configurable computing units can be instantiated and connected to each other to generate a convolution calculation array. Convolution calculations for different types of convolution layers can be completed by this array; for some networks In the model, there are two or more sizes of convolution kernels in the same convolutional layer, which can divide the array and provide different convolution parameters for different areas. In order to ensure the synchronization of the output results of all areas of the calculation array, By calculating the difference between different convolution kernel sizes, the time difference between the calculation units in different regions can be found to produce the output result. The calculation unit in the less calculation area will wait for the calculation unit in the more calculation area until the time difference is zero. Then start the calculation to ensure the synchronization of the output results of the array and complete the parallel calculation of different types of convolution methods.")

count an amount of at least part of the incoming data received during the first of the exchange stages from the one or more others of the processors; and

in response to determining that the amount of the at least part of the incoming data received has reached a predefined amount, perform data transfer operations to transfer the further outgoing data to the one or more others of the processors during a second of the exchange stages.

(paragraph [0056], "The control module configures the upper limit value for each counter, the upper limit value of the input data counter and the output data counter is configured as k*a, and the upper limit value of the input weight counter is configured as k*a*b, The upper limit value of the output channel number counter is configured as b, and the upper limit value of the output characteristic map size counter is configured as h; when the input

counters are all at the upper limit value, the calculation unit starts to perform calculation, and each output counter performs corresponding counting and The jump of the state machine is controlled; when each output counter is at the upper limit value, it indicates that the convolution calculation of some or all of the output channels of the convolutional layer has been completed.")

The subject-matter of said claim is therefore not new (Article 33(2) PCT).

It is furthermore noted that the subject-matter of independent claim 1 cannot be considered to involve an inventive step over the disclosure of document D2 (see passages cited in the International Search Report), taken in combination with the general knowledge of the skilled person.

1.1.2 Independent claim 18

The reasoning above applies, mutatis mutandis, to said corresponding claim 18, the subject matter of which is therefore also not new (Article 33(2) PCT).

1.2 **Dependent claims**

Dependent claim 2-17 do not appear to contain any additional features which, in combination with the features of any claim to which they refer, meet the requirements of the PCT in respect of novelty or inventive step, (Article 33(2) and (3) PCT).

- claims 2 and 3: Performing partial operations before the amount of incoming data received has reached the predefined amount relates to a mere implementation detail from which the skilled person would select, in accordance with the circumstances, without the exercise of inventive skill.

- claims 4-6: Using different locations in the buffer and the number of processors used relate to mere implementation details from which the skilled person would select, in accordance with the circumstances, without the exercise of inventive skill.

- claim 7: D1 further discloses that the plurality of processing units receive part of the incoming data from the one or more others of the processors; and send part of the outgoing data to the one or more others of the processors; wherein the steps of counting the amount of incoming data received and determining that the amount of the incoming data received has reached the predefined amount are performed by one or more of the plurality of processing units of a first type. (paragraphs [0056-0057], "The control module configures the upper limit value for each counter, the upper limit value of the input data counter and the output data counter is configured as k*a, and the upper limit value of the

input weight counter is configured as k*a*b, The upper limit value of the output channel number counter is configured as b, and the upper limit value of the output characteristic map size counter is configured as h; when the input counters are all at the upper limit value, the calculation unit starts to perform calculation, and each output counter performs corresponding counting and The jump of the state machine is controlled; when each output counter is at the upper limit value, it indicates that the convolution calculation of some or all of the output channels of the convolutional layer has been completed.")

- claim 8-10: D1 discloses that any processor of the processor array is able to count the number of input data (paragraphs [0056-0057]). Although D1 is silent regarding that the PEs may be a combination of different types, this difference relates to a mere implementation detail from which the skilled person would select, in accordance with the circumstances, without the exercise of inventive skill.

- claim 11: D1 discloses that the data comprise a set of gradients for weights of a machine learning model (abstract).

- claim 12: Polling the the counting circuitry to determine the amount of the incoming data received seems to be merely one of several straightforward possibilities from which the skilled person would select, in accordance with circumstances, without the exercise of inventive skill.

- claims 13, 14 and 17: Using a RDMA to transmit data, the type of network topology and the type of circuit (e.g ASICS or FPGA) merely represent customary design choices that the person skilled in the art would apply according to circumstances without exercising any inventive step activity.

- claims 15-16: Although D1 is silent regarding that it performs a reduce-scatter collective, such operation is well known in the field of neural networks, as seen foe example in D2 (figure 14E).

2 **Re Item VII**

Certain defects in the international application

2.1 Independent claims 1 and 18 are not in the two-part form in accordance with Rule 6.3(b) PCT, which in the present case would be appropriate, with those features known in combination from the prior art (document D1) being placed in the preamble (Rule 6.3(b) (I) PCT) and with the remaining features being included in the characterising part (Rule 6.3(b) (ii) PCT).

2.2 To meet the requirements of Rule 5.1 (a) (ii) PCT, documents D1 and D2 should be identified in the description and the relevant background art disclosed therein should be briefly discussed.

2.3 The claims should include reference signs in order to meet the requirements of Rule 6.2(b) PCT.