Internal distribution code:

(A) [-] Publication in OJ
(B) [-] To Chairmen and Members
(C) [-] To Chairmen
(D) [X] No distribution

Datasheet for the decision of 5 August 2024

Case Number: T 0099/22 - 3.5.06

Application Number: 19175038.9

Publication Number: 3564813

IPC: G06F9/30, G06F9/38, G06F9/46, G06T1/20, G06N3/063

Language of the proceedings: EN

Title of invention:
COMPUTE OPTIMIZATION MECHANISM

Applicant:
Intel Corporation

Headword:
Mixed precision FMAC/INTEL

Relevant legal provisions:
EPC Art. 56

Keyword:
Inventive step - (no)

Decisions cited:

Catchword:

Europäisches Patentamt
European Patent Office
Office européen des brevets

Beschwerdekammern

Boards of Appeal

Chambres de recours

Boards of Appeal of the
European Patent Office
Richard-Reitzner-Allee 8
85540 Haar
GERMANY
Tel. +49 (0)89 2399-0

Case Number: T 0099/22 - 3.5.06

D E C I S I O N
of Technical Board of Appeal 3.5.06
of 5 August 2024

Appellant: (Applicant)	Intel Corporation 2200 Mission College Boulevard Santa Clara, CA 95054 (US)
Representative:	Samson & Partner Patentanwälte mbB Widenmayerstraße 6 80538 München (DE)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 2 July 2021 refusing European patent application No. 19175038.9 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman	M. Müller
Members:	G. Zucka
	B. Müller

Summary of Facts and Submissions

I. The appeal is against the decision by the examining division, dispatched with reasons on 2 July 2021, to refuse European patent application 19175038.9, on the basis that none of the three requests satisfied the requirements of Article 56 EPC, in view of the following document:

D3: N. Brunie *et al.*: "Mixed-precision Fused Multiply and Add", made available on the Internet on 17 November 2011 at the following URL: https://hal-ens-lyon.archives-ouvertes.fr/ensl-00642157, XP055596852.

It is noted that the above bibliographic data for D3 differ from those in the appealed decision. The latter are incorrect and relate to a different version of this document. The references to D3 in the appealed decision (see reasons 15.1.1) refer to the version indicated above, as is apparent from the fact that the decision refers to pages 1-6 as opposed to pages 165-169 of the said different version. The board has no reason to assume that the reasoning of the examining division is affected by this confusion.

II. A notice of appeal was received on 9 August 2021, the appeal fee being paid on the same date. A statement of grounds of appeal was received on 12 November 2021.

III. The appellant requested that the decision under appeal be set aside and a patent granted on the basis of the claims of a main or a first or second auxiliary request . The appellant made a conditional request for oral proceedings.

IV. The board issued a summons to oral proceedings. In an annex to the summons, the board set out its preliminary opinion on the appeal, according to which the appealed decision should be upheld.

V. On 10 June 2024, the appellant filed claims of a new main and auxiliary request, replacing all previous requests.

VI. The appellant requests that the decision under appeal be set aside and a patent be granted on the basis of claims 1 to 3 of the main or the auxiliary request filed on 10 June 2024.

The further text on file is:

description pages
2 to 86 as originally filed,
1 and 87 received on 28 April 2020;

drawing sheets
1 to 38 as originally filed.

VII. Claim 1 of the main request reads as follows:

"A graphics processor hosting a compute mechanism, wherein the compute mechanism is configured to support mixed precision fused multiply-accumulate, FMAC, operations, wherein the FMAC operations include an arithmetic-logic unit, ALU, operation of D=A*B+C, wherein A and B and C include different precision and format."

VIII. Claim 1 of the auxiliary request reads as follows:

"A graphics processor comprising an arithmetic-logic unit, ALU, and hosting a compute mechanism, wherein the compute mechanism is configured to support mixed precision fused multiply-accumulate, FMAC, operations, wherein the FMAC operations include an FMAC ALU operation of D=A*B+C, wherein the operands A and B and C include different precision and format,

wherein an ALU instruction is provided that includes a 16-bit attribute specifying a different format for each of the operands A, B, and C and a required format for the result D, wherein the ALU executes the ALU instruction by parsing the 16-bit attribute to determine the different format for each operand A, B, and C, converting each operand A, B, and C in the required format and executing the FMAC ALU operation."

IX. The wording of the other claims of the main and the auxiliary request is not relevant for the present decision.

X. At the end of the oral proceedings, the chairman announced the board's decision.

Reasons for the Decision

1. *The invention*

The application relates to data processing using a general-purpose graphics processing unit (GPGPU, see description [0001]).

According to the statement of grounds of appeal (section III.1 "Explanation of the invention"), the claimed subject-matter provides an improvement of fused multiply-accumulate (FMAC) operations. In contrast to what is said to be the case in much of the prior art, it supports a single instruction for FMAC operations with operands of different precision.

2. *Main request - inventive step; Article 56 EPC*

2.1 The document D3 is considered to be a suitable starting point for an inventive step analysis of the subject-matter of claim 1 of the main request.

2.2 D3 discloses a compute mechanism which could be hosted for instance on a graphics processor (see section I "Introduction", first two paragraphs).

2.3 The compute mechanism is configured to support mixed-precision fused multiply-accumulate, FMAC, operations, wherein the FMAC operations include an arithmetic-logic unit, ALU, operation of D=A*B+C, wherein C has a different precision than A and B (*ibid.*, third paragraph: the MPFMA*k* computes R=A*B+C (rounded), where A and B are binary*k* numbers, and C is a binary2*k* number).

2.4 Claim 1 of the main request specifies that "A and B and C include different precision and format". One might question whether this formulation is satisfied by the specific situation in D3 in which C has a different precision than A and B, or whether it implies that all three operands may have different precision and format at the same time. However, the board is satisfied that the latter is the intended meaning, as is made clear in claim 1 of the auxiliary request which provides that "a

different format for each of the operands A, B and C" is specified in the claimed "16-bit attribute". For this reason, the board assumes to the appellant's benefit that claim 1 of both requests implies that the claimed compute mechanism must be prepared to handle three operands with different precision and format each.

2.5 Accordingly, the difference between the subject-matter of claim 1 of the main request and the apparatus of D3 is that in the claim all three operands may have a different precision and format, whereas, in D3, A and B have the same precision and format.

2.6 This difference solves the problem of dealing with a situation where the precision and format of the operands is not fixed in advance. Following the arguments made by the appellant during the oral proceedings, said features allow for a greater flexibility.

2.7 According to the board, however, the skilled person is bound to face such situation at some point and will want to solve it. One of the most straightforward solutions would for instance be to apply suitable conversions before the MPFMA operation is carried out.

In the interest of integration, which is part of normal technical evolution, the skilled person would consider applying this conversion already as part of the ALU operation or operations. The FMAC operation would then include an ALU operation of D=A*B+C, wherein A and B and C include different precision and format.

2.8 The skilled person would thereby arrive at the subject-matter of claim 1 of the main request without the need for an inventive step.

The board therefore concludes that the main request does not satisfy the requirements of Article 56 EPC.

3. *Auxiliary request*

3.1 Claim 1 of the auxiliary request now specifies that the claimed graphics processor comprises an ALU. This is not considered to constitute a difference with the main request, as the presence of an ALU was already implied by the presence of ALU operations.

3.2 According to claim 1 of the auxiliary request, an ALU instruction is now provided that includes a 16-bit attribute specifying a different format for each of the operands A, B, and C and a required format for the result D, wherein the ALU executes the ALU instruction by parsing the 16-bit attribute to determine the different format for each operand A, B, and C, converting each operand A, B, and C in the required format and executing the FMAC ALU operation.

3.3 In view of the board, it goes without saying that, if the format of the FMAC operands A, B and C is not fixed, the FMAC operation should in some way be informed about those respective formats. One natural way to do this is to use a 16-bit attribute, given the common occurrence and use of 16-bit values in ALUs. Such a choice in itself provides no particular advantage.

In addition, the board notes that the required number of bits will depend on the (unspecified) number of

formats which each of A, B, C and D is allowed to have. For certain numbers of allowed formats, the number 16 will also be a convenient, and thus obvious, choice for this further reason.

The ALU would then execute the ALU instruction by parsing the 16-bit attribute to determine the different format for each operand A, B, and C.

3.4 It would make perfect sense to the skilled person to use the highest possible precision that can be expressed by the 16-bit attribute for the result of the FMAC operation, viz. D. Claim 1 of the main request does not exclude this choice. It is noted that also in D3 the result has the highest precision.

The precision of A, B and C is not known in advance. It would therefore also make sense not to provide several calculation methods for different combinations of precisions, but to have a single such method working with the highest possible precision for the three operands, and to convert the operands to this highest possible precision before the calculation method is applied.

3.5 The skilled person would thus arrive at the subject-matter of claim 1 of the auxiliary request without the need for an inventive step.

3.6 The appellant submitted during the oral proceedings before the board that the claimed apparatus would allow to choose a less precise "required format" D, and that converting each operand to that smaller precision before the FMAC operation would lead to a faster

execution. The claim is however not limited to such a scenario.

3.7 The board therefore concludes that the auxiliary request does not satisfy the requirements of Article 56 EPC.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chairman:

L. Stridde Martin Müller

Decision electronically authenticated