(19) SUOMI - FINLAND
(FI)

PATENTTI- JA REKISTERIHALLITUS
PATENT- OCH REGISTERSTYRELSEN
FINNISH PATENT AND REGISTRATION OFFICE

(10) **FI/EP4002105 T3**

(12) **EUROOPPAPATENTIN KÄÄNNÖS**
ÖVERSÄTTNING AV EUROPEISKT PATENT
TRANSLATION OF EUROPEAN PATENT SPECIFICATION

(45) Käännöksen kuulutuspäivä - Kungörelsedag av översättning - Translation available to the public **21.11.2024**

(97) Eurooppapatentin myöntämispäivä - Meddelandedatum för det europeiska patentet - Date of grant of European patent **04.09.2024**

(51) Kansainvälinen patenttiluokitus - Internationell patentklassificering - International patent classification
G06F 9/30 (2018 . 01)

(96) Eurooppapatenttihakemus - Europeisk patentansökan - European patent application EP21217772.9

(22) Tekemispäivä - Ingivningsdag - Filing date **08.10.2019**

(97) Patenttihakemuksen julkiseksitulopäivä - Patentansökans publiceringsdag - Patent application available to the public **25.05.2022**

(30) Etuoikeus - Prioritet - Priority

09.11.2018 US US201816186387

(73) Haltija - Innehavare - Holder

1• Intel Corporation , 2200 Mission College Boulevard , Santa Clara, CA 95054 , (US)

(72) Keksijä - Uppfinnare - Inventor

1• Heinecke, Alexander F. , San Jose CA, 95134 , (US)
2• Valentine, Robert , 36054 Kiryat Tivon HA , (IL)
3• Charney, Mark J. , Lexington MA, 02421 , (US)
4• Sade, Raanan , 3658900 Kibutz Sarid HA , (IL)
5• Adelman, Menachem , 3229624 Haifa HA , (IL)
6• Sperber, Zeev , 3092832 Zichron Yackov , (IL)
7• Gradstein, Amit , 3052316 Binyamina HA , (IL)
8• Rubanovich, Simon , 34792 Haifa HA , (IL)

(74) Asiamies - Ombud - Agent

Papula Oy , P.O.Box 981 , 00101 Helsinki , (FI)

(54) Keksinnön nimitys - Uppfinningens benämning - Title of the invention

JÄRJESTELMÄT JA MENETELMÄT 16-BITTISTEN LIUKULUKUMUOTOISTEN MATRIISIPISTETULOKÄSKYJEN SUORITTAMISEEN

SYSTEMS AND METHODS FOR PERFORMING 16-BIT FLOATING-POINT MATRIX DOT PRODUCT INSTRUCTIONS

PATENTTIVAATIMUKSET

1. Prosessointiyksikkö, joka käsittää:

noutopiirin, joka noutaa käskyn;

dekoodauspiirin käskyn dekoodaamiseksi, ja käskyllä on op-koodi, ensimmäinen kenttä, jolla määritetään useiden dataelementtien ensimmäinen säilytyspaikka, joka vastaa ensimmäistä matriisia, jossa on M riviä ja N saraketta 32-bittisiä yhden tarkkuuksisia liukulukutietoelementtejä, toinen kenttä, jolla määritetään useiden dataelementtien toinen säilytyspaikka, joka vastaa toista matriisia, jossa on M riviä kertaa K saraketta 16-bittisiä liukulukutietoelementtejä, joilla on bfloat16-muoto, ja kolmas kenttä, jolla määritetään useiden dataelementtien kolmas säilytyspaikka, joka vastaa kolmatta matriisia, jossa on K riviä kertaa N saraketta 16-bittisiä liukulukutietoelementtejä, joilla on bfloat16-muoto; ja

dekoodauspiirin kanssa kytketty suorituspiiri, joka suorittaa käskyä vastaavat operaatiot toisen matriisin M rivin jokaiselle riville m ja kolmannen matriisin N sarakkeen jokaiselle sarakkeelle n:

tuottaa pistetuoton toisen matriisin riviä m vastaavista K 16-bittisistä liukulukutietoelementeistä ja kolmannen matriisin saraketta n vastaavista K 16-bittisistä liukulukutietoelementeistä;

kerryttää pistetuoton 32-bittisen yhden tarkkuuksisen liukulukutietoelementin kanssa, joka vastaa M rivin riviä m ja N sarakkeen saraketta n, ensimmäisen matriisin 32-bittisen yhden tarkkuuksisen liukulukutietoelementin tuloksen tuottamiseksi; ja

tallentaa tuloksena olevan 32-bittisen yhden tarkkuuksisen liukulukutietoelementin ensimmäisen tallennuspaikan paikkaan, joka vastaa ensimmäisen matriisin riviä m ja saraketta n, jossa suorituspiiri suorittaa käskyä vastaavat operaatiot käsitellä toisen ja kolmannen matriisin denormaalit arvot nollana.

2. Patenttivaatimuksen 1 mukainen prosessointiyksikkö, joka sisältää lisäksi ohjausrekisterin pyöristystavan määrittämiseksi, jolloin pistetulon tuottaminen ja pistetulojen kasaaminen sisältävät kumpikin yhden pyöristystavan soveltamisen riippumatta ohjausrekisterin määrittämästä pyöristystavasta.

3. Patenttivaatimuksen 2 mukainen prosessointiyksikkö, jossa prosessointiyksikön ei tarvitse konsultoida ohjausrekisteriä ohjeen vuoksi.

4. Jonkin patenttivaatimuksen 1-3 mukainen prosessointiyksikkö, jossa prosessointiyksikön ei ole tarkoitus päivittää ohjausrekisteriä käskyn vuoksi.

5. Jonkin patenttivaatimuksen 1-4 mukainen prosessointiyksikkö, jossa prosessointiyksikkö ei merkitse poikkeuksia käskyn vuoksi.

6. Jonkin patenttivaatimuksen 1-5 prosessointiyksikkö, jossa suorituspiiri suorittaa käskyä vastaavan operaation ajaa denormaalit arvot nollaan.

7. Jonkin patenttivaatimuksen 1-6 prosessointiyksikkö, jossa toisen matriisin M riviä ja kolmannen matriisin N saraketta on yhtä monta.

8. Jonkin patenttivaatimuksen 1-7 prosessointiyksikkö, jossa ensimmäinen, toinen ja kolmas säilytyspaikka ovat 128-bittisissä vektorirekistereissä.

9. Jonkin patenttivaatimuksen 1-8 prosessointiyksikkö, jossa jokainen pistetulo on 32-bittinen yhden tarkkuuksinen liukulukutulo.

10. Jonkin patenttivaatimuksen 1-9 mukainen prosessointiyksikkö, joka lisäksi sisältää:

haaraennustepiirin;
rekisterin uudelleennimityspiiri; ja
aikataulutuspiirin, joka aikatauluttaa dekoodatun käskyn suoritettavaksi.

11. Jonkin patenttivaatimuksen 1-10 prosessointiyksikkö, jossa prosessointiyksikkö on yleiskäyttöinen keskusyksikkö, CPU-ydin.

12. Jonkin patenttivaatimuksen 1-11 prosessointiyksikkö, jossa prosessointiyksikkö on RISC-ydin (Reduced Instruction Set Computing).

13. Patenttivaatimuksen 2 mukainen prosessointiyksikkö, jossa yhden kierroksen tila on kiinteä käskylle.

14. System-on-a-chip, SoC, joka käsittää:

muistinohjaimen; ja
yleiskäyttöisen keskusyksikön, CPU:n, ydin, joka on kytketty muistinohjaimeen, ja yleiskäyttöinen CPU-ydin on jonkin patenttivaatimuksen 1-13 mukainen prosessointiyksikkö.

15. Patenttivaatimuksen 14 SoC, joka sisältää lisäksi yhden tai useamman seuraavista: verkkoprosessori, joka on kytketty yleiskäyttöisen CPU-ytimen kanssa, apuprosessori, joka on kytketty yleiskäyttöisen CPU-ytimen kanssa, ja kuvaprosessori, joka on kytketty yleiskäyttöisen CPU-ytimen kanssa.

16. Järjestelmä, joka käsittää:

muistin; ja

yleiskäyttöisen keskusyksikön, CPU-ydin, joka on kytketty muistiin, ja yleiskäyttöinen CPU-ydin on jonkin patenttivaatimuksen 1-13 mukainen prosessoriyksikkö.

17. Patenttivaatimuksen 16 mukainen järjestelmä, joka lisäksi sisältää yhden tai useamman seuraavista: yleiskäyttöiseen CPU-ytimeen kytketty massamuistilaite, yleiskäyttöiseen CPU-ytimeen kytketty Peripheral Component Interconnect, PCI, Express -väylä, yleiskäyttöiseen CPU-ytimeen kytketty viestintälaite.

18. Menetelmä tietojen käsittelyä varten prosessointiyksikössä, joka käsittää:

haetaan käsky;

dekoodataan käsky, jossa käskyllä on op-koodi, ensimmäinen kenttä, joka määrittelee useiden dataelementtien ensimmäisen säilytyspaikan, joka vastaa ensimmäistä matriisia, jossa on M riviä ja N saraketta 32-bittisiä yhden tarkkuuksisia liukulukutietoelementtejä, toinen kenttä, joka määrittelee useiden dataelementtien toisen säilytyspaikan, joka vastaa toista matriisia, jossa on M riviä kertaa K saraketta 16-bittisiä liukulukutietoelementtejä, joilla on bfloat16-muoto, ja kolmas kenttä, joka määrittelee useiden dataelementtien kolmannen säilytyspaikan, joka vastaa kolmatta matriisia, jossa on K riviä kertaa N saraketta 16-bittisiä liukulukutietoelementtejä, joilla on bfloat16-muoto;

suoritetaan käskyä vastaavat operaatiot, mukaan lukien toisen matriisin M rivin jokaiselle riville m ja kolmannen matriisin N sarakkeen jokaiselle sarakkeelle n:

tuotetaan pistetulo toisen matriisin riviä m vastaavista K 16-bittisistä liukulukutietoelementeistä ja kolmannen matriisin saraketta n vastaavista K 16-bittisistä liukulukutietoelementeistä;

kerrytetään pistetulo 32-bittisen yhden tarkkuuksisen liukulukutietoelementin kanssa, joka vastaa M rivin riviä m ja N sarakkeen saraketta n ensimmäisen matriisin sarakkeista; ja

tallennetaan tuloksena saatu 32-bittinen yhden tarkkuuksisen liukulukutietoelementti ensimmäisen tallennuspaikan paikkaan, joka vastaa ensimmäisen matriisin riviä m ja saraketta n, jossa operaatioiden suorittaminen käsittää sellaisen operaation suorittamisen, joka vastaa käskyä käsitellä toisen ja kolmannen matriisin denormaalit arvot nollana.

19. Tietokoneohjelma, joka käsittää ohjeita, jotka prosessorin suorittamana saavat prosessorin suorittamaan patenttivaatimuksen 18 mukaisen menetelmän.

20. Tietokoneella luettava tietoväline, johon on tallennettu patenttivaatimuksen 19 tietokoneohjelma.