

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
14 October 2021 (14.10.2021)



(10) International Publication Number
WO 2021/204435 A1

(51) International Patent Classification:

G06N 3/063 (2006.01) G06F 17/16 (2006.01)
G06F 17/15 (2006.01)

(21) International Application Number:

PCT/EP2021/051906

(22) International Filing Date:

28 January 2021 (28.01.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

16/842,035 07 April 2020 (07.04.2020) US

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York, 10504 (US).

(71) Applicant (for MG only): **IBM UNITED KINGDOM LIMITED** [GB/GB]; PO Box 41, North Harbour, Portsmouth Hampshire PO6 3AU (GB).

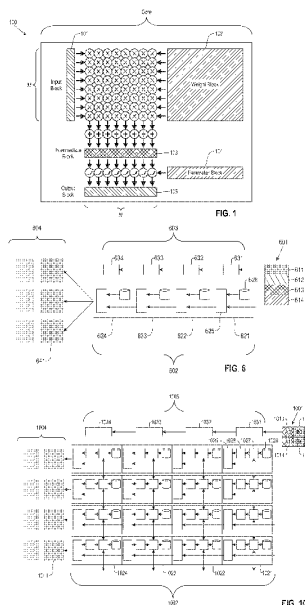
(72) Inventors: **SAWADA, Jun**; IBM Corporation, 11501 Burnet Road, Austin, Texas 78758-3400 (US). **MODHA, Dharmendra**; IBM Corporation, 650 Harry Road, Almaden Research Center, San Jose, California 95120-6099 (US). **CASSIDY, Andrew, Stephen**; IBM Corporation,

11501 Burnet Road, Austin, Texas 78758-3400 (US). **ARTHUR, John, Vernon**; IBM Corporation, 650 Harry Road, Almaden Research Center, San Jose, California 95120-6099 (US). **NAYAK, Tapan**; IBM Corporation, 650 Harry Road, Almaden Research Center, San Jose, California 95120-6099 (US). **ORTEGA OTERO, Carlos**; IBM Corporation, 650 Harry Road, Almaden Research Center, San Jose, California 95120-6099 (US). **TABA, Brian, Seisho**; IBM Corporation, 650 Harry Road, Almaden Research Center, San Jose, California 95120-6099 (US). **AKOPYAN, Filipp**; IBM Corporation, 1101 Kitchawan Road, PO Box 218, Yorktown Heights, New York 10598-0218 (US). **DATTA, Pallab**; IBM Corporation, 650 Harry Road, Almaden Research Center, San Jose, California 95120-6099 (US).

(74) Agent: **ROBERTSON, Tracey**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Hursley, Winchester Hampshire SO21 2JN (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,

(54) Title: NEURAL NETWORK WEIGHT DISTRIBUTION FROM A GRID OF MEMORY ELEMENTS



(57) Abstract: Neural inference chips for computing neural activations are provided. In various embodiments, a neural inference chip comprises at least one neural core, a memory array, an instruction buffer, and an instruction memory. The instruction buffer has a position corresponding to each of a plurality of elements of the memory array. The instruction memory provides at least one instruction to the instruction buffer. The instruction buffer advances the at least one instruction between positions in the instruction buffer. The instruction buffer provides the at least one instruction to at least one of the plurality of elements of the memory array from its associated position in the instruction buffer when the memory of the at least one of the plurality of elements contains data associated with the at least one instruction. Each element of the memory array provides a data block from its memory to its horizontal buffer in response to the arrival of an associated instruction from the instruction buffer. The horizontal buffer of each element of the memory array provides a data block to the horizontal buffer of another of the elements of the memory array or to the at least one neural core.

WO 2021/204435 A1

[Continued on next page]

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).

Published:

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