

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
09 September 2022 (09.09.2022)



(10) International Publication Number
WO 2022/186763 A1

(51) International Patent Classification:
G06N 3/063 (2006.01) *G06F 7/53* (2006.01)

(21) International Application Number:
PCT/SG2021/050107

(22) International Filing Date:
03 March 2021 (03.03.2021)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 1 Fusionopolis Way, #20-10 Connexis North Tower, Singapore 138632 (SG).

(72) Inventors: **PARAMASIVAM, Vishnu**; c/o Industry Development, Institute of Microelectronics, 2 Fusionopolis Way, #08-02 Innovis Tower, Singapore 138634 (SG). **JIANG, Wenyu**; c/o PKM, Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632 (SG). **DO, Anh Tuan**; c/o Industry Development, Institute of Microelectronics, 2 Fusionopolis Way, #08-02 Innovis Tower, Singapore 138634 (SG). **WONG, Ming Ming**; c/o Industry Development, Institute of Microelectronics, 2 Fusionopolis Way, #08-02 Innovis Tower, Singapore 138634 (SG). **MANI, Aarthi**; c/o In-

dustry Development, Institute of Microelectronics, 2 Fusionopolis Way, #08-02 Innovis Tower, Singapore 138634 (SG). **SHRESTHA, Sumit Bam**; c/o PKM, Institute for Infocomm Research, 1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632 (SG).

(74) Agent: **VIERING, JENTSCHURA & PARTNER LLP**; P.O. Box 1088, Rochor Post Office, Rochor Road, Singapore 911833 (SG).

(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, 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,

(54) Title: NEURAL NETWORK PROCESSOR SYSTEM WITH RECONFIGURABLE NEURAL PROCESSING UNIT, AND METHOD OF OPERATING AND METHOD OF FORMING THEREOF

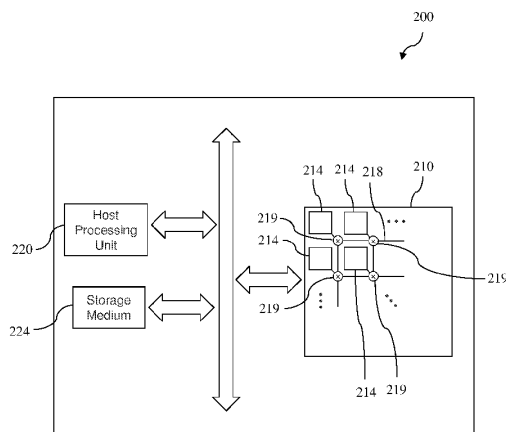


FIG. 2

(57) Abstract: There is provided a neural network processor system including: a neural processing unit including a plurality of neural processing cores; a router network including a plurality of routers communicatively coupled to the plurality of neural processing cores, respectively; and a host processing unit communicatively coupled to the neural processing unit based on the router network and configured to coordinate the neural processing unit for performing neural network computations. Each neural processing core includes: a control register block configured to receive and store partial sum configuration information from the host processing unit; and a partial sum interface communicatively coupled to the control register block and configured to transmit a first partial sum neural packet generated by the neural processing core to a first another neural processing core of the plurality of neural processing cores and/or receive a second partial sum neural packet generated by a second another neural processing core of the plurality of neural processing cores, based on the partial sum configuration information stored in the neural processing core. Furthermore, a first set of neural processing cores of the plurality of neural processing cores are combinable based on the partial sum configuration information respectively stored therein to form a first neurosynaptic column chain. There is also provided a corresponding method of operating and a corresponding method of forming the neural network processor system.

[Continued on next page]

WO 2022/186763 A1

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))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*