



(51) International Patent Classification:

G06N 3/063 (2006.01) G06F 13/28 (2006.01)
G06N 3/04 (2006.01) G06F 9/30 (2018.01)

(21) International Application Number:

PCT/US2017/046963

(22) International Filing Date:

15 August 2017 (15.08.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

15/335,769 27 October 2016 (27.10.2016) US

(71) Applicant: **GOOGLE LLC** [US/US]; 1600 Amphitheatre
Parkway, Mountain View, California 94043 (US).

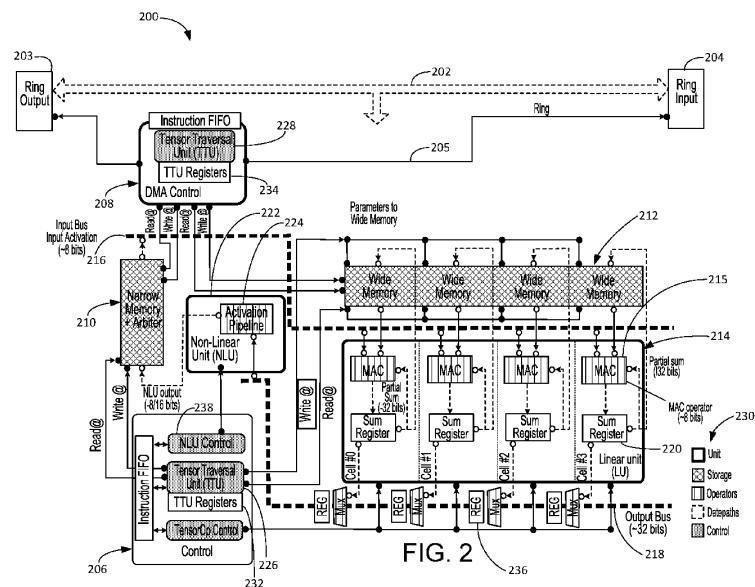
(72) Inventors: **TEMAM, Olivier**; 8 Rue de Londres, 75009
Paris (FR). **NARAYANASWAMI, Ravi**; 1600 Amphithe-

atre Parkway, Mountain View, California 94043 (US).
KHAITAN, Harshit; 1600 Amphitheatre Parkway, Mountain
View, California 94043 (US). **WOO, Dong Hyuk**;
1600 Amphitheatre Parkway, Mountain View, California
94043 (US).

(74) Agent: **HENRY, Joel** et al.; Fish & Richardson P.C., P.O.
Box 1022, Minneapolis, Minnesota 55440-1022 (US).

(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, 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,

(54) Title: NEURAL NETWORK COMPUTE TILE



(57) Abstract: A computing unit is disclosed, comprising a first memory bank for storing input activations and a second memory bank for storing parameters used in performing computations. The computing unit includes at least one multiply accumulate ("MAC") operator that receives parameters from the second memory bank and performs computations. The computing unit further includes a first traversal unit that provides a control signal to the first memory bank to cause an input activation to be provided to a data bus accessible by the MAC operator. The computing unit performs one or more computations associated with at least one element of a data array, the one or more computations being performed by the MAC operator and comprising, in part, a multiply operation of the input activation received from the data bus and a parameter received from the second memory bank.

[Continued on next page]

SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*