(12) UK Patent Application (19) GB (11) 2610975 (13) A

(43) Date of Reproduction by UK Office 22.03.2023

(21) Application No: **2218461.8**

(22) Date of Filing: **14.04.2021**

Date Lodged: **08.12.2022**

(30) Priority Data:
(31) **15929618** (32) **13.05.2020** (33) **US**

(86) International Application Data:
PCT/CN2021/087267 En 14.04.2021

(87) International Publication Data:
WO2021/227757 En 18.11.2021

(71) Applicant(s):
International Business Machines Corporation
(Incorporated in USA - New York)
New Orchard Road, Armonk, New York 10504,
United States of America

(72) Inventor(s):
Ashish Ranjan
Arvind Kumar
Carl Radens

(74) Agent and/or Address for Service:
Elkington and Fife LLP
Prospect House, 8 Pembroke Road, SEVENOAKS,
Kent, TN13 1XR, United Kingdom

(51) INT CL:
G06F 12/02 (2006.01)

(56) Documents Cited:
WO 2019/157599 A1 **CN 103678143 A**
US 20190164037 A1

(58) Field of Search:
INT CL **G06F**
Other: **CNKI, DWPI, SIPOABS, CNABS**

(54) Title of the Invention: **Optimal placement of data structures in a hybrid memory based inference computing platform**
Abstract Title: **Optimal placement of data structures in a hybrid memory based inference computing platform**

(57) Various embodiments are provided for optimized placement of data structures in memory in a computing environment. One or more data structures may be dynamically allocated in a hybrid memory according to weights and activations of the or more data structures in a deep neural network ("DNN"). The hybrid memory includes at least a first memory and a second memory that differ according to write endurance attributes.

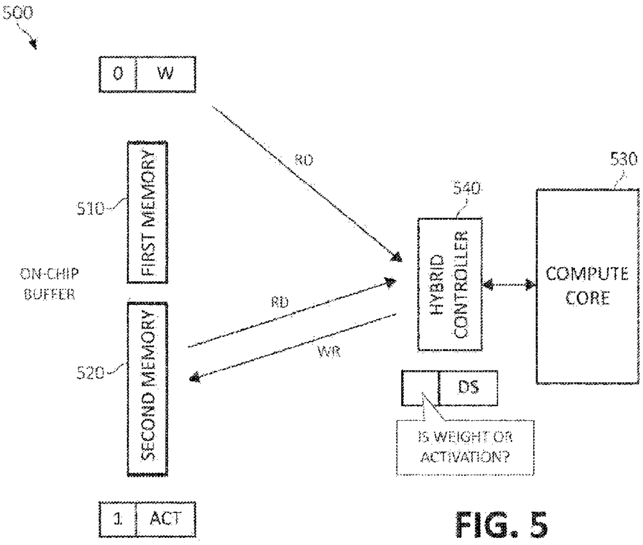

FIG. 5