



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11)
CỤC SỞ HỮU TRÍ TUỆ



1-0050264

(51)^{2020.01} **G11C 17/16; G11C 11/54; G11C 17/18; (13) B**
G06N 3/063; G11C 11/56

(21) 1-2022-00844

(22) 26/08/2020

(86) PCT/US2020/047975 26/08/2020

(87) WO2021/041526 04/03/2021

(30) 16/556,505 30/08/2019 US

(45) 25/08/2025 449

(43) 25/05/2022 410A

(73) QUALCOMM INCORPORATED (US)

ATTN: International IP Administration, 5775 Morehouse Drive, San Diego, CA
92121-1714, United States of America

(72) YANG, Haining (US); CHIDAMBARAM, Periannan (US).

(74) Công ty TNHH Quốc tế D & N (D&N INTERNATIONAL CO.,LTD.)

(54) PHƯƠNG PHÁP VÀ MÁY ĐỂ LẬP TRÌNH THIẾT BỊ VÀ PHƯƠNG TIỆN BẮT
BIỂN ĐỌC ĐƯỢC BẰNG MÁY TÍNH

(21) 1-2022-00844

(57) Các khía cạnh nhất định của sáng chế đề xuất phương pháp và máy để lập trình thiết bị có một hoặc nhiều mạch lập trình được để triển khai, chẳng hạn, mô hình học máy. Một máy làm ví dụ thường bao gồm nhiều đường từ, nhiều đường bit và dây mạch lập trình được. Mỗi mạch lập trình được được ghép nối với đường từ tương ứng trong số nhiều đường từ và với đường bit tương ứng trong số nhiều đường bit và bao gồm: điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng, điện trở phụ, cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ, trong đó điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và mạch lập trình được tạo cấu hình để làm nổ có chọn lọc cầu chì bảo vệ. Ngoài ra, sáng chế còn đề cập đến phương tiện bất biến đọc được bằng máy tính.

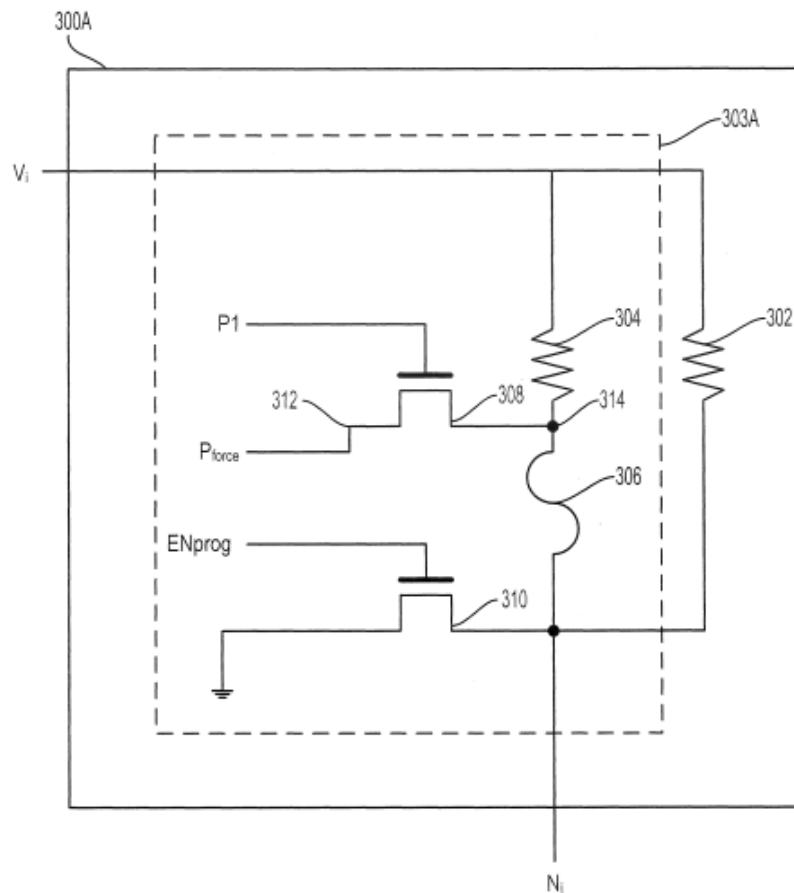


Fig.3A

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập chung đến các mạng nơron nhân tạo, và cụ thể hơn là đề cập đến các kỹ thuật và máy để thực hiện các hoạt động học máy trên các thiết bị biên.

Tình trạng kỹ thuật của sáng chế

Mạng nơron nhân tạo, mà có thể gồm có nhóm nơron nhân tạo (ví dụ, các mô hình nơron), là thiết bị máy tính hoặc tương trưng cho phương pháp do thiết bị máy tính thực hiện. Các mạng nơron này có thể được sử dụng cho các ứng dụng và/hoặc thiết bị khác nhau, như máy ảnh giao thức internet (Internet Protocol - IP), thiết bị internet vạn vật (Internet of Things - IoT), phương tiện giao thông tự vận hành, và/hoặc các robot dịch vụ.

Từng nút riêng trong mạng nơron nhân tạo có thể mô phỏng các nơron sinh học bằng cách nhận lấy dữ liệu đầu vào và thực hiện các hoạt động đơn giản trên dữ liệu này. Các kết quả của các hoạt động đơn giản được thực hiện trên dữ liệu được chuyển một cách chọn lọc đến các nơron khác. Các giá trị trọng số được kết hợp với mỗi vector và nút trong mạng, và các giá trị này ràng buộc việc dữ liệu đầu vào có liên quan với dữ liệu đầu ra như thế nào. Chẳng hạn, dữ liệu đầu vào của mỗi nút có thể được nhân với giá trị trọng số tương ứng, và các tích có thể được cộng lại. Tổng của các tích có thể được điều chỉnh bằng các độ lệch tùy ý, và hàm kích hoạt có thể được áp dụng cho kết quả này, tạo ra tín hiệu đầu ra của nút hoặc “kích hoạt đầu ra”. Giá trị trọng số ban đầu có thể được xác định bằng dòng lặp của dữ liệu huấn luyện qua mạng (ví dụ, giá trị trọng số được thiết lập trong pha huấn luyện trong đó mạng học cách xác định các lớp cụ thể bởi các đặc tính của dữ liệu đầu vào điển hình của chúng).

Có các loại mạng nơron nhân tạo khác nhau, như mạng nơron hồi quy (recurrent neural network - RNN), mạng nơron perceptron nhiều lớp (multilayer perceptron - MLP), mạng nơron tích chập (convolutional neural network - CNN), và các mạng tương tự. RNN làm việc trên nguyên lý tiết kiệm đầu ra của lớp và cấp đầu ra này trở lại đầu vào để giúp dự đoán kết quả của lớp. Trong các mạng nơron MLP, dữ liệu có thể được cấp vào lớp đầu vào, và một hoặc nhiều lớp ẩn tạo ra các mức độ trừu tượng hóa cho dữ

liệu này. Sau đó có thể tạo ra các dự đoán trên lớp đầu ra dựa trên dữ liệu được trừu tượng hóa này. Các MLP có thể đặc biệt thích hợp cho các vấn đề dự đoán phân loại khi các đầu vào được gán lớp hoặc nhãn. Các mạng nơron tích chập (convolutional neural network - CNN) là loại mạng nơron nhân tạo truyền thẳng. Các mạng nơron tích chập có thể bao gồm các tập hợp nơron nhân tạo mà mỗi tập hợp có một trường tiếp nhận (ví dụ, vùng được định vị trong không gian của không gian đầu vào) và cùng nhau lát kín không gian đầu vào. Các mạng nơron tích chập có nhiều ứng dụng. Cụ thể, các CNN đã được sử dụng rộng rãi trong lĩnh vực nhận diện và phân loại mẫu.

Trong các kiến trúc mạng nơron nhiều lớp, đầu ra của lớp nơron nhân tạo thứ nhất sẽ là đầu vào của lớp nơron nhân tạo thứ hai, đầu ra của lớp nơron nhân tạo thứ hai sẽ là đầu vào của lớp nơron nhân tạo thứ ba, và v.v.. Các mạng nơron tích chập có thể được huấn luyện để nhận diện hệ thống thứ bậc của các đặc điểm. Việc tính toán trong kiến trúc mạng nơron tích chập có thể được phân bổ trên tập hợp các nút xử lý, mà có thể được tạo cấu hình trong một hoặc nhiều chuỗi tính toán. Các kiến trúc nhiều lớp này có thể được huấn luyện một lớp tại một thời điểm và có thể được tinh chỉnh nhờ sử dụng sự lan truyền ngược.

Bản chất kỹ thuật của sáng chế

Các khía cạnh nhất định của sáng chế đề xuất máy trên đó các mô hình học máy có thể được thực thi. Nhìn chung, máy này bao gồm nhiều đường từ, nhiều đường bit và dây mạch lập trình được. Mỗi mạch lập trình được trong dây được ghép nối với đường từ tương ứng trong số nhiều đường từ và đường bit tương ứng trong số nhiều đường bit. Mỗi mạch lập trình được bao gồm điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng, điện trở phụ, cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ, trong đó điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và mạch lập trình được tạo cấu hình để làm nổ có chọn lọc cầu chì bảo vệ.

Các khía cạnh nhất định của sáng chế đề xuất phương pháp lập trình thiết bị có nhiều đường từ, nhiều đường bit và dây mạch lập trình được. Mỗi mạch lập trình được trong dây mạch lập trình được được ghép nối với đường từ tương ứng trong số nhiều đường từ và với đường bit tương ứng trong số nhiều đường bit. Phương pháp này bao gồm bước lập trình mạch lập trình được trong dây mạch lập trình được bằng cách làm nổ

có chọn lọc cầu chì bảo vệ (ví dụ, nhờ sử dụng mạch lập trình), trong đó cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ, điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

Các khía cạnh nhất định của sáng chế đề xuất máy để lập trình thiết bị. Máy này nhìn chung bao gồm nhiều đường từ, nhiều đường bit, dây mạch lập trình được, phương tiện để lập trình mạch lập trình được trong dây mạch lập trình được bằng cách làm nổ có chọn lọc cầu chì bảo vệ nhờ sử dụng phương tiện để lập trình, và phương tiện để vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình. Cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ, điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng.

Các khía cạnh nhất định của sáng chế đề xuất phương tiện bất biến đọc được bằng máy tính chứa các lệnh mà, khi được thực thi bởi ít nhất một bộ xử lý, khiến cho bộ xử lý thực hiện các hoạt động để lập trình thiết bị có nhiều đường từ, nhiều đường bit, và dây mạch lập trình được. Hoạt động này nhìn chung bao gồm lập trình mạch lập trình được trong dây mạch lập trình được bằng cách làm nổ có chọn lọc cầu chì bảo vệ (ví dụ, nhờ sử dụng mạch lập trình), trong đó cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ, điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

Các khía cạnh nhất định của sáng chế đề xuất phương pháp lập trình thiết bị để triển khai mô hình học máy. Phương pháp này nhìn chung bao gồm bước nhận, ở thiết bị, giá trị trọng số của mô hình học máy được huấn luyện cần được triển khai trên mạch lập trình được, lập trình mạch lập trình được bằng cách tách có chọn lọc một hoặc nhiều điện trở phụ trong mạch lập trình được ra khỏi trạng thái được ghép nối song song với điện trở chính, dựa trên giá trị trọng số của mô hình học máy được huấn luyện, và vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

Các khía cạnh nhất định của sáng chế đề xuất phương tiện bất biến đọc được bằng máy tính chứa các lệnh mà, khi được thực thi bởi ít nhất một bộ xử lý, khiến cho bộ xử lý

thực hiện các hoạt động để lập trình thiết bị có bộ xử lý và mạch lập trình được để triển khai mô hình học máy. Các hoạt động này nhìn chung bao gồm việc nhận, ở thiết bị, giá trị trọng số của mô hình học máy được huấn luyện cần được triển khai trên mạch lập trình được, lập trình mạch lập trình được bằng cách tách có chọn lọc một hoặc nhiều điện trở phụ trong mạch lập trình được ra khỏi trạng thái được ghép nối song song với điện trở chính, dựa trên giá trị trọng số của mô hình học máy được huấn luyện, và vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

Các khía cạnh nhất định của sáng chế đề xuất máy lập trình thiết bị để triển khai mô hình học máy. Máy này nhìn chung bao gồm dây mạch lập trình được, phương tiện để nhận, từ máy chủ, các giá trị trọng số của mô hình học máy được huấn luyện cần được triển khai trên dây mạch lập trình được, phương tiện để lập trình ít nhất một mạch lập trình được trong dây mạch lập trình được bằng cách tách có chọn lọc một hoặc nhiều điện trở phụ trong mạch lập trình được ra khỏi trạng thái được ghép nối song song với điện trở chính, dựa trên các giá trị trọng số của mô hình học máy được huấn luyện, và phương tiện để vận hành thiết bị nhờ sử dụng phương tiện mạch lập trình được đã được lập trình để nhận, ở thiết bị từ máy chủ, giá trị trọng số của mô hình học máy được huấn luyện cần được triển khai trên mạch lập trình được, phương tiện để lập trình mạch lập trình được bằng cách tách có chọn lọc một hoặc nhiều điện trở phụ trong mạch lập trình được ra khỏi trạng thái được ghép nối song song với điện trở chính, dựa trên các giá trị trọng số của mô hình học máy được huấn luyện, và phương tiện để vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

Các khía cạnh nhất định của sáng chế đề xuất máy lập trình thiết bị để triển khai mô hình học máy. Máy này bao gồm dây mạch lập trình được và ít nhất một bộ xử lý được ghép nối với dây mạch lập trình được. Ít nhất một bộ xử lý nhìn chung được tạo cấu hình để nhận các giá trị trọng số của mô hình học máy được huấn luyện cần được triển khai trên dây mạch lập trình được, lập trình ít nhất một mạch lập trình được trong dây mạch lập trình được bằng cách tách có chọn lọc một hoặc nhiều điện trở phụ trong mạch lập trình được ra khỏi trạng thái được ghép nối song song với điện trở chính, dựa trên các giá trị trọng số của mô hình học máy được huấn luyện, và vận hành máy nhờ sử dụng mạch lập trình được đã được lập trình.

Các khía cạnh, ưu điểm và dấu hiệu khác của sáng chế sẽ trở nên rõ ràng sau khi xem xét toàn bộ bản mô tả, bao gồm các phần sau: Mô tả vắn tắt các hình vẽ, Mô tả chi tiết sáng chế, và Yêu cầu bảo hộ.

Mô tả vắn tắt các hình vẽ

Fig.1 minh họa phương án thực hiện ví dụ của hệ thống trên chip (system-on-a-chip - SOC).

Fig.2 minh họa ví dụ về dây mạch lập trình được mà có thể được lập trình để thực hiện hoạt động học máy, theo các khía cạnh của sáng chế.

Fig.3A minh họa ví dụ về mạch lập trình được bao gồm điện trở chính và điện trở phụ được sử dụng để triển khai tham số trọng số của mô hình học máy, theo các khía cạnh của sáng chế.

Fig.3B minh họa ví dụ về mạch lập trình được bao gồm điện trở chính và nhiều điện trở phụ được sử dụng để triển khai tham số trọng số của mô hình học máy, theo các khía cạnh của sáng chế.

Fig.4 minh họa các hoạt động làm ví dụ để thực hiện các hoạt động học máy trên thiết bị có mạch lập trình được, theo các khía cạnh của sáng chế.

Fig.5 minh họa các hoạt động làm ví dụ để lập trình mạch lập trình được nhằm điều chỉnh điện trở của mạch lập trình được, theo các khía cạnh của sáng chế.

Fig.6 minh họa ví dụ về môi trường mạng trong đó mạch lập trình được ở thiết bị biên được lập trình để triển khai mô hình học máy được huấn luyện, theo các khía cạnh của sáng chế.

Fig.7 thể hiện sơ đồ khối minh họa kiến trúc phần mềm làm ví dụ để lập mô hình các chức năng trí tuệ nhân tạo (artificial intelligence - AI), theo các khía cạnh của sáng chế.

Mô tả chi tiết sáng chế

Các khía cạnh nhất định của sáng chế đề xuất các phương pháp và máy để thực hiện các hoạt động học máy trên thiết bị có mạch lập trình được.

Với sự tham chiếu tới các hình vẽ, một số khía cạnh làm ví dụ của sáng chế được mô tả. Thuật ngữ “làm ví dụ” được sử dụng ở đây có nghĩa là “có vai trò làm ví dụ, mẫu

hoặc minh họa”. Khía cạnh bất kỳ được mô tả ở đây là “làm ví dụ” không nhất thiết được hiểu là được ưu tiên hay có lợi so với các khía cạnh khác.

Các mạng nơron nhân tạo làm ví dụ

Fig.1 minh họa phương án thực hiện ví dụ của hệ thống trên chip (system-on-a-chip - SOC) 100, hệ thống này có thể bao gồm bộ xử lý trung tâm (central processing unit - CPU) 102 hoặc CPU nhiều lõi được tạo cấu hình để thực hiện hàm dropout (bỏ học) Monte Carlo song song, theo các phương án nhất định của sáng chế. Các biến số (ví dụ, tín hiệu nơron và các trọng số liên kết), các tham số hệ thống kết hợp với thiết bị máy tính (ví dụ, mạng nơron với các trọng số), độ trễ, thông tin bin tần số, và thông tin tác vụ có thể được lưu trữ trong khối bộ nhớ kết hợp với đơn vị xử lý thần kinh (neural processing unit - NPU) 108, trong khối bộ nhớ kết hợp với CPU 102, trong khối bộ nhớ kết hợp với đơn vị xử lý đồ họa (graphics processing unit - GPU) 104, trong khối bộ nhớ kết hợp với bộ xử lý tín hiệu số (digital signal processor - DSP) 106, trong khối bộ nhớ 118, hoặc có thể được phân bổ qua nhiều khối. Các lệnh được thực thi ở CPU 102 có thể được tải từ bộ nhớ chương trình kết hợp với CPU 102 hoặc có thể được tải từ khối bộ nhớ 118.

SOC 100 cũng có thể bao gồm các khối xử lý bổ sung được tạo riêng theo các hàm chuyên biệt, như GPU 104, DSP 106, khối kết nối 110, mà có thể bao gồm kết nối thế hệ thứ năm (fifth generation - 5G), kết nối tiến hóa dài hạn thế hệ thứ tư (fourth generation long term evolution - 4G LTE), kết nối Wi-Fi, kết nối USB, kết nối Bluetooth, và tương tự, và bộ xử lý đa phương tiện 112 mà có thể, chẳng hạn, phát hiện và nhận diện các cử chỉ. Theo một phương án thực hiện, NPU được triển khai trong CPU 102, DSP 106, và/hoặc GPU 104. SOC 100 cũng có thể bao gồm bộ xử lý cảm biến 114, bộ xử lý tín hiệu ảnh (image signal processor - ISP) 116, và/hoặc module dẫn đường 120, có thể bao gồm hệ thống định vị toàn cầu.

SOC 100 có thể được dựa trên tập lệnh ARM. Theo một khía cạnh của sáng chế, các lệnh được tải vào CPU 102 có thể bao gồm mã để tra cứu kết quả phép nhân được lưu trong bảng tìm kiếm (lookup table - LUT) tương ứng với tích của giá trị đầu vào và trọng số bộ lọc. Các lệnh được tải vào CPU 102 cũng có thể bao gồm mã để bắt hoạt bộ nhân trong phép toán nhân của tích khi phát hiện thấy tích này so khớp trùng bảng tìm kiếm. Ngoài ra, các lệnh được tải vào CPU 102 có thể bao gồm mã để lưu

trữ tích được tính toán của giá trị đầu vào và trọng số bộ lọc khi phát hiện thấy tích này không trùng bằng tìm kiếm.

SOC 100 có thể còn bao gồm dây mạch lập trình được. Như được đề cập chi tiết hơn dưới đây, mỗi mạch trong dây mạch lập trình được có thể là ô bao gồm nhiều điện trở. Mỗi mạch có thể được lập trình để ghép nối hoặc khử phép có chọn lọc điện trở chính với một hoặc nhiều điện trở phụ để thực hiện tham số trọng số trong mô hình học máy. Dây mạch lập trình được có thể được cập nhật để phản ánh các tham số trọng số mới cho mô hình học máy hoặc có thể được lập trình lại để triển khai mô hình học máy khác, nếu cần.

SOC 100 và/hoặc các thành phần của nó có thể được tạo cấu hình để thực hiện các loại tác vụ học máy hoặc tác vụ học sâu ở thiết bị biên, như xe cơ giới tự vận hành (ví dụ, xe tự lái) hoặc thiết bị máy tính di động. Các thiết bị biên này có thể có tài nguyên tính toán hạn chế nhưng có thể hoạt động trong các môi trường trong đó mong muốn có thời gian chờ giảm và chi phí truyền thông dữ liệu (ví dụ, khi thực hiện các tác vụ nhận diện đối tượng từ dữ liệu video được thu theo thời gian thực, thực hiện các tác vụ nhận diện cục bộ để giảm thiểu thời gian chờ liên quan đến việc cung cấp dữ liệu đến hệ thống ở xa để phân tích, v.v.). Hơn nữa, các thiết bị biên này có thể có các ràng buộc bảo đảm mà có thể đặt ra các giới hạn lên mức độ dữ liệu có thể được bộc lộ ra ngoài của thiết bị biên. Bằng cách sử dụng dây mạch lập trình được để triển khai mô hình học máy hoặc kiến trúc học sâu, các khía cạnh của sáng chế có thể cho phép thực thi có hiệu quả về công suất các hàm học máy trên thiết bị biên.

Các kiến trúc học sâu có thể thực hiện tác vụ nhận diện đối tượng bằng cách học để biểu diễn các đầu vào ở các mức trừu tượng cao hơn kế tiếp trong mỗi lớp, nhờ đó tạo nên biểu diễn đặc trưng hữu ích của dữ liệu đầu vào. Theo cách này, kỹ thuật học sâu giải quyết điểm bế tắc chính của học máy thông thường. Trước khi kỹ thuật học sâu ra đời, cách tiếp cận học máy đối với vấn đề nhận dạng đối tượng có thể đã phụ thuộc rất nhiều vào các đặc điểm do con người thiết kế, có lẽ là kết hợp với bộ phân loại nông. Bộ phân loại nông có thể là bộ phân loại tuyến tính hai lớp, chẳng hạn, trong đó tổng có trọng số của các giá trị đầu vào (ví dụ, các thành phần vectơ đầu vào) có thể được so sánh với ngưỡng để dự đoán đầu vào này thuộc lớp nào. Các đặc trưng do con người thiết kế có thể là các mẫu hoặc nhân được các kỹ sư có chuyên môn về miền điều chỉnh

cho phù hợp với miền vấn đề cụ thể. Các kiến trúc học sâu, trái lại, có thể học để biểu diễn các đặc điểm mà tương tự với thứ mà người kỹ sư có thể thiết kế, nhưng thông qua việc huấn luyện. Hơn thế nữa, mạng sâu có thể học để biểu diễn và nhận diện các loại đặc điểm mới mà con người có thể chưa nghĩ đến.

Kiến trúc học sâu có thể học hệ thống thứ bậc các đặc điểm. Nếu được thể hiện bằng dữ liệu thị giác, chẳng hạn, lớp thứ nhất có thể học để nhận diện các đặc điểm tương đối đơn giản, như các biên, trong dòng đầu vào. Theo một ví dụ khác, nếu được thể hiện bằng dữ liệu thính giác, lớp thứ nhất có thể học để nhận diện công suất phổ ở các tần số chuyên biệt. Lớp thứ hai, lấy đầu ra của lớp thứ nhất làm đầu vào, có thể học để nhận diện các tổ hợp đặc điểm, như các hình dạng đơn giản đối với dữ liệu thị giác hoặc các tổ hợp âm thanh đối với dữ liệu thính giác. Chẳng hạn, các lớp cao hơn có thể học để biểu diễn các hình dạng phức tạp trong dữ liệu thị giác hoặc các từ trong dữ liệu thính giác. Ngoài ra, các lớp cao hơn có thể học để nhận diện các đối tượng thị giác hoặc các cụm từ được nói thông thường.

Các kiến trúc học sâu có thể hoạt động đặc biệt tốt khi áp dụng cho các vấn đề có cấu trúc phân cấp tự nhiên. Ví dụ, việc phân loại các phương tiện cơ giới có thể được hưởng lợi từ việc học thứ nhất để nhận dạng bánh xe, kính chắn gió và các đặc trưng khác. Các đặc điểm này có thể được kết hợp ở các lớp cao hơn theo các cách khác nhau để nhận diện xe hơi, xe tải, và máy bay.

Các mạng nơron có thể được thiết kế với các loại mẫu kết nối khác nhau. Trong các mạng truyền thẳng, thông tin được truyền từ các lớp thấp hơn đến các lớp cao hơn, với mỗi nơron trong lớp cho trước truyền thông đến các nơron ở các lớp cao hơn. Biểu diễn có thứ bậc có thể được xây dựng ở các lớp kế tiếp của mạng truyền thẳng, như được mô tả ở trên. Các mạng nơron cũng có thể có các kết nối hồi quy hoặc phản hồi (còn gọi là từ trên xuống dưới). Trong kết nối hồi quy, đầu ra từ nơron ở lớp đã cho có thể được truyền thông với một nơron khác trong cùng lớp. Kiến trúc hồi quy có thể có thể hữu ích trong việc nhận diện các mẫu bất qua nhiều hơn một trong các khúc dữ liệu đầu vào mà được phân phối đến mạng nơron trong chuỗi. Kết nối từ nơron ở lớp đã cho đến nơron ở lớp thấp hơn được gọi là kết nối phản hồi (hoặc từ trên xuống dưới). Mạng có nhiều kết nối phản hồi có thể hữu ích khi sự nhận diện khái niệm bậc cao có thể hỗ trợ thêm việc phân biệt các đặc điểm bậc thấp cụ thể của đầu vào.

Mạng tin cậy sâu (deep belief network - DBN) là mô hình xác suất bao gồm nhiều lớp của nút ẩn. Các DBN có thể được sử dụng để trích xuất biểu diễn có thứ bậc của các tập dữ liệu huấn luyện. Có thể thu được DBN bằng cách xếp chồng các lớp máy Boltzmann hạn chế (Restricted Boltzmann Machine - RBM). RBM là loại mạng nơron nhân tạo mà có thể học phân bố xác suất trên tập các đầu vào. Vì các RBM có thể học phân bố xác suất khi không có thông tin về lớp mà mỗi đầu vào nên được phân loại vào đó, các RBM thường được sử dụng trong việc học không có giám sát. Bằng cách sử dụng mô hình kết hợp không có giám sát và có giám sát, các RBM bên dưới của DBN có thể được huấn luyện theo cách không có giám sát và có thể đóng vai trò là bộ trích xuất đặc trưng, và RBM phía trên có thể được huấn luyện theo cách có giám sát (trên phân phối chung các đầu vào từ lớp trước và các lớp đích) và có thể đóng vai trò như bộ phân loại.

Mạng tích chập sâu (deep convolutional network - DCN) là mạng gồm các mạng tích chập, được tạo cấu hình với các lớp tổng hợp và chuẩn hóa bổ sung. Các DCN đã đạt được hiệu suất cao nhất trong nhiều tác vụ. Các DCN có thể được huấn luyện bằng cách sử dụng phương pháp học có giám sát, trong đó cả mục tiêu đầu vào và đầu ra đều được biết đến đối với nhiều hình mẫu và được sử dụng để sửa đổi các trọng số của mạng bằng cách sử dụng phương pháp giảm dần gradient.

Các DCN có thể là các mạng truyền thẳng. Ngoài ra, như được mô tả ở trên, các kết nối từ nơron ở lớp thứ nhất của DCN đến nhóm nơron ở lớp cao hơn tiếp theo được dùng chung qua các nơron trong lớp thứ nhất. Các kết nối truyền thẳng và dùng chung của các DCN có thể được khai thác để xử lý nhanh. Gánh nặng tính toán của DCN có thể ít hơn nhiều, ví dụ, khi so với mạng nơron có kích thước tương tự bao gồm các kết nối lặp lại hoặc phản hồi.

Các mạch lập trình được làm ví dụ để thực hiện các hoạt động học máy trên các thiết bị biên

Các hoạt động học máy thường có thể được thực hiện trên các đơn vị xử lý trung tâm (central processing unit - CPU) hiệu năng cao hoặc các bộ xử lý chuyên dụng, như đơn vị xử lý tensor (tensor processor unit - TPU) hoặc các đơn vị xử lý đồ họa (graphics processing unit - GPU) lập trình được. Các hoạt động học sâu có thể tốn kém về tính toán và bộ nhớ. Hơn nữa, vì chi phí tính toán cho các hoạt động này, các bộ xử lý mà trên đó các hoạt động học sâu thực thi có thể tiêu thụ mức công suất đáng kể (ví dụ, công

suất 200 oát). Tuy nhiên, việc sử dụng điện năng của các CPU hoặc bộ xử lý chuyên dụng mà trên đó các hoạt động học sâu thực thi, có thể hạn chế khả năng thực hiện các hoạt động học sâu trên các thiết bị biên khác nhau, như xe cơ giới hoặc điện thoại di động.

Bộ nhớ điện tĩnh (non-volatile memory - NVM) và các điện trở bộ nhớ đã được đề xuất để sử dụng trong các ứng dụng học máy năng lượng thấp. Các điện trở bộ nhớ thường là các thiết bị bộ nhớ có thể thay đổi các giá trị lưu trong các thiết bị bộ nhớ bằng cách tăng hoặc giảm điện trở của mạch. Trong ứng dụng học máy, các dây điện trở bộ nhớ được đề xuất để thực hiện các phép nhân ma trận-vector và/hoặc ma trận-ma trận. Chẳng hạn, dây điện trở bộ nhớ thanh ngang mật độ cao có thể được sử dụng trong đó điện trở bộ nhớ kết nối đường từ và đường bit ở mỗi mỗi chỗ giao của dây, sao cho vector đầu vào của tập điện thể được nhập vào đồng thời vào mỗi trong số các đường từ trong dây. Trong các ví dụ này, NVM có thể được sử dụng dưới dạng các xinap để điều chỉnh tỷ lệ tín hiệu, hoặc các trọng số được sử dụng trong mô hình học máy. Việc học có thể được thực hiện thông qua các thay đổi đối với các trọng số sử dụng trong mô hình học máy. Các dây phân tử và/hoặc điện trở bộ nhớ của NVM có thể cho phép tính toán trong bộ nhớ (ví dụ, hiệu năng của các phép toán logic khác nhau trong bộ nhớ) và dòng dữ liệu cố định trọng số (ví dụ, hiệu năng của các hoạt động khác nhau mà không cần di chuyển dữ liệu vào và ra khỏi bộ nhớ, trái ngược với các phép toán nhân-và-cộng dồn trong các CPU hoặc GPU phổ biến mà đòi hỏi ba phép toán đọc bộ nhớ và một phép toán ghi bộ nhớ), mà có thể làm giảm việc sử dụng năng lượng từ việc di chuyển dữ liệu trong hệ thống. Vì NVM và các điện trở bộ nhớ có thể cho phép việc tính toán có lợi về công suất, NVM và điện trở bộ nhớ có thể thích hợp để thực thi việc học sâu, hoặc suy luận, các ứng dụng trên thiết bị biên trong khi không cần huấn luyện và triển khai mô hình tốn kém về tính toán ở các máy chủ từ xa.

Tuy nhiên, điện trở bộ nhớ có thể có các đặc tính khác mà làm giảm tính thích hợp của các thiết bị này để triển khai ở các thiết bị biên. Chẳng hạn, điện trở bộ nhớ, như bộ nhớ truy cập ngẫu nhiên điện trở (Resistive Random Access Memory - RRAM), bộ nhớ thay đổi pha (phase change memory - PCM), hoặc bộ nhớ truy cập ngẫu nhiên từ (magnetic random access memory - MRAM) truyền mômen spin (spin-transfer torque - STT), có thể có sự sai khác từ thiết bị thiết bị này đến thiết bị khác được tạo ra vào thời điểm sản xuất và có thể có sự phụ thuộc dòng điện của chương trình sau khi lập trình. Độ

lệch chuẩn của điện tích dẫn PCM có thể vượt quá mức thay đổi trung bình. RRAM có thể cần lượng điện năng lớn (ví dụ, lớn hơn 2,5 vôn) để lập trình và có thể có sự sai khác lớn về điện trở. Cuối cùng, STT MRAM có thể có các ràng buộc kích thước làm giới hạn sự tích hợp của STT MRAM với các mạch khác. Chẳng hạn, STT MRAM có thể được sản xuất trên quy trình 28 nm có các bóng bán dẫn lập trình lớn và mạch ngoại vi, trong khi các mạch tích hợp khác ở thiết bị biên (ví dụ, CPU, GPU, SOC, v.v.) có thể được sản xuất trên các quy trình quang khắc nhỏ hơn.

Ngoài ra, sự sai khác giữa các chip khác có thể tồn tại trong quá trình sản xuất chất bán dẫn. Các sai khác có thể được tạo ra từ các thiết bị đầu dây chuyền (front-end-of-line - FEOL) được sử dụng để làm mẫu cho sai khác bán dẫn và/hoặc quy trình gia công kim loại cuối dây chuyền (back-end-of-line - BEOL). Vì sai khác của quá trình sản xuất và giữa các thiết bị, có thể huấn luyện số lượng chip giới hạn có điện trở bộ nhớ cho các ứng dụng học máy ở đám mây; tuy nhiên, có thể không thể huấn luyện số lượng lớn chip có điện trở bộ nhớ cho các ứng dụng học máy trên các thiết bị di động. Chẳng hạn, các khiếm khuyết thiết bị có thể dẫn đến một hoặc nhiều điện trở bộ nhớ bị hỏng, như điện trở bộ nhớ bị đóng băng ở trạng thái dẫn cao (“kẹt (stuck-on)”) hoặc trạng thái dẫn thấp (“bị loại ra (stuck-off)”). Mỗi chip có thể có số lượng điện trở bộ nhớ hỏng khác nhau do các khiếm khuyết thiết bị, mà có thể tạo ra phương sai ở đầu ra đối với đầu vào cho trước bất kỳ ở các chip khác nhau triển khai cùng một mô hình. Cuối cùng, vì NVM hoặc các thiết bị điện trở bộ nhớ có thể sử dụng các quy trình sản xuất khác nhau từ tập các quy trình sản xuất được sử dụng trong sản xuất các mạch tích hợp khác cho các thiết bị biên (ví dụ, CPU, GPU, SOC, v.v.), sự tích hợp của NVM hoặc các điện trở bộ nhớ vào các bộ xử lý sử dụng trong các thiết bị biên có thể bị giới hạn.

Để cho phép các hoạt động học máy tiết kiệm điện năng trên thiết bị biên, các khía cạnh của sáng chế đề xuất mạch lập trình được bao gồm dây điện trở mà có thể được sử dụng để thực hiện các tham số trọng số của mô hình học máy được huấn luyện, như mạng nơ-ron sâu. Hệ thống máy tính hoặc cụm hệ thống máy tính ở xa thiết bị biên (ví dụ, các máy chủ dành riêng hoặc các hệ thống điện toán đám mây) có thể huấn luyện mô hình học máy (ví dụ, mạng nơ-ron sâu) để xác định các trọng số và độ lệch liên kết và có thể đưa ra các trọng số và độ lệch liên kết xác định cho thiết bị biên, và thiết bị biên có thể lập trình mạch lập trình được theo đó để triển khai mô hình học máy được huấn luyện. Theo một số khía cạnh, mạch lập trình được có thể được thử nghiệm về tác vụ

chuyên biệt để xác định có nên cập nhật các trọng số và độ lệch liên kết để đáp ứng đặc tả hay không. Hơn nữa, thiết bị biên có thể được cập nhật trong khi được triển khai để thực thi các tác vụ mới hoặc để phản ánh việc học mới nhận được từ hệ thống máy tính từ xa hoặc cụm hệ thống máy tính.

Bằng cách sử dụng mạch lập trình được bao gồm các dây điện trở để thực hiện các hoạt động học máy, các khía cạnh của sáng chế có thể cho phép việc thực hiện các mô hình học máy trên mạch lập trình được mà giải quyết vấn đề về hiệu suất điện năng và tốc độ của chip dựa trên điện trở bộ nhớ. Các mạch lập trình được này có thể tránh được việc sử dụng điện trở bộ nhớ và thay vào đó sử dụng các thành phần được chế tạo trên công nghệ xử lý sẵn có, mà có thể cho phép việc sản xuất lượng lớn mạch lập trình được này. Hơn nữa, bằng cách sử dụng các thành phần điện thông thường, các chip có thể được chế tạo với các đặc tính phù hợp (ví dụ, số lượng điện trở lập trình được trong dây), có thể giảm đến mức thấp nhất tính biến thiên trên các chip khác nhau, và các chip có thể được điều chỉnh phù hợp để làm giảm bớt tính biến thiên trong quá trình sản xuất từ các chênh lệch về điện trở dây kim loại, sự sai khác của bóng bán dẫn, và sự sai khác của điện trở. Hơn nữa, khả năng điều chỉnh phù hợp của các mạch lập trình được, cho phép các cập nhật đối với các trọng số và độ lệch liên kết trong mạch, có thể cho phép các khả năng học tập mới được triển khai cho các thiết bị biên trong lĩnh vực này.

Fig.2 minh họa dây 200 của các mạch lập trình được mà có thể được lập trình để thực hiện hoạt động học máy, theo các khía cạnh của sáng chế. Như được minh họa, dây 200 có thể bao gồm nhiều mạch lập trình được 202, mà có thể tương ứng với mạch lập trình được 300 được minh họa trên các hình vẽ Fig.3A và Fig.3B. 3A và Fig.3B. Dây 200 có thể bao gồm tập n dây đầu vào $V_i, V_{i+1}, V_{i+2}, \dots, V_{i+n}$ và tập m đường đầu ra $N_j, N_{j+1}, N_{j+2}, \dots, N_{j+m}$, trong đó n và m là các số nguyên dương. Các dây đầu vào V có thể là các đường từ mà qua đó các đầu vào tương tự biểu diễn vector đặc trưng được cung cấp cho dây 200, và các đường đầu ra N có thể là các đường bit biểu diễn tích của phân tử vector đặc trưng và trọng số liên kết được biểu diễn bởi giá trị điện trở của mạch lập trình được 202 trong dây 200. Mỗi mạch lập trình được 202 trong dây 200 có thể được định vị ở chỗ giao của dây đầu vào V và đường đầu ra N sao cho mỗi mạch lập trình được tạo ra giá trị bằng cách xử lý đầu vào điện thế nhận được (ví dụ, từ dây đầu vào V) thông qua điện trở chính và điện trở phụ bất kỳ được ghép nối điện với điện trở chính theo cách song song (ví dụ, điện trở phụ mà cầu chì bảo vệ lập trình được tương ứng cho nó chưa

được làm nổi). Ví dụ, như được minh họa trên Fig.2, mạch lập trình được 202a, 202b, và 202c có thể lần lượt được định vị ở chỗ giao nhau của dây đầu vào V_i và các đường đầu ra N_j , N_{j+1} , và N_{j+2} , sao cho mạch lập trình được 202a, 202b, và 202c nhận đầu vào từ dây đầu vào V_i để xử lý và tạo các đầu ra trên các đường đầu ra tương ứng N_j , N_{j+1} , và N_{j+2} . Tương tự, mạch lập trình được 202d, 202e, và 202f có thể lần lượt được định vị ở chỗ giao nhau của dây đầu vào V_{i+1} và các đường đầu ra N_j , N_{j+1} , và N_{j+2} , để xử lý đầu vào từ dây đầu vào V_{i+1} , mạch lập trình được 202g, 202h, và 202i có thể lần lượt được định vị ở chỗ giao nhau của dây đầu vào V_{i+2} và các đường đầu ra N_j , N_{j+1} , và N_{j+2} , để xử lý đầu vào từ dây đầu vào V_{i+2} , và v.v..

Mỗi mạch lập trình được 202 trong dãy 200 có thể tạo đầu ra (ví dụ, dòng điện đầu ra) dưới dạng sự kết hợp của giá trị đầu vào đi qua điện trở chính của mạch lập trình được 202 và giá trị đầu vào đi qua một hoặc nhiều điện trở phụ được ghép nối chọn lọc với điện trở chính. Các đầu ra tạo ra bởi nhiều mạch lập trình được có thể được kết hợp (ví dụ, phép lấy tổng dòng điện) làm đầu ra của một đường đầu ra đơn nhất trong dãy 200. Chẳng hạn, các đầu ra tạo ra bởi mạch lập trình được 202a, 202d, và 202g có thể được kết hợp vào đầu ra tạo ra trên đường đầu ra N_j , các đầu ra tạo ra bởi mạch lập trình được 202b, 202e, và 202h có thể được kết hợp vào đầu ra tạo ra trên đường đầu ra N_{j+1} , các đầu ra tạo ra bởi mạch lập trình được 202c, 202f, và 202i có thể được kết hợp vào đầu ra tạo ra trên đường đầu ra N_{j+2} , và v.v.. Nhìn chung, đầu ra tạo ra trên đường đầu ra N từ mỗi mạch lập trình được 202 có thể là tích của phần tử vector đặc trưng và trọng số liên kết được biểu diễn bởi điện trở tổng lập trình được của mỗi mạch lập trình được trên đường đầu ra. Trong quá trình tạo đầu ra thông qua nhiều mạch lập trình được 202, như đã đề cập ở đây, các phép toán nhân và cộng dồn có thể được tích hợp vào dãy 200, mà có thể làm giảm mức tiêu thụ năng lượng trong các hoạt động học máy. Đối với các khía cạnh nhất định, dòng điện kết hợp tạo ra bởi mạch lập trình được trên đường đầu ra N đã cho có thể được cung cấp dưới dạng tín hiệu tương tự đầu vào cho giai đoạn tiếp theo (ví dụ, một mạch tương tự khác). Đối với các khía cạnh khác, dòng điện đường bit kết hợp có thể được sử dụng để nạp điện cho tụ điện để tạo các xung nhọn đầu ra. Trong trường hợp này, tụ điện có thể được phóng điện để tạo ra xung nhọn khi tụ điện được khởi động bằng kỹ thuật bất kỳ trong số các kỹ thuật thích hợp (ví dụ, mạch khởi động xung nhọn khi điện thế của tụ điện đạt điện thế định trước). Đối với các phương án khác nữa, dòng

điện đường bit kết hợp có thể được chuyển hóa thành tín hiệu số bằng bộ chuyển đổi tương tự-số (analog-to-digital converter - ADC).

Theo một số khía cạnh, dây 200 có thể được thử nghiệm sau khi mạch lập trình được 202 trong dây 200 được lập trình để xác nhận rằng dây 200 hoạt động trong phạm vi đặc tả (ví dụ, tạo ra các kết quả mong đợi của quá trình học máy đối với đầu vào đã cho). Nếu dây 200 không tạo ra các kết quả được kỳ vọng (ví dụ, do phương sai hoặc sai số trong sản xuất ở một hoặc nhiều phần tử trong mạch lập trình được 202), thì dây 200 có thể được thử nghiệm thêm để xác định một hoặc nhiều mạch lập trình được 202 trong dây mà không phù hợp với đặc tả và khiến cho dây 200 tạo ra kết quả không chính xác. Cầu chì bảo vệ lập trình được trong một hoặc nhiều mạch lập trình được 202 được xác định có thể được làm nổ để điều chỉnh trọng số liên kết của mạch lập trình được 202 hoặc để loại bỏ một cách hiệu quả một hoặc nhiều mạch lập trình được 202 xác định ra khỏi dây 200.

Fig.3A minh họa ví dụ về mạch lập trình được 300A sử dụng dây điện trở có điện trở chính và điện trở phụ để thực hiện giá trị trọng số cho mô hình học máy. Mạch lập trình được 300A có thể tương ứng với mạch lập trình được 202 bất kỳ được minh họa trên Fig.2. Mạch lập trình được 300A có thể sử dụng các thành phần điện chuẩn, như các điện trở kim loại, cầu chì bảo vệ lập trình được, và bóng bán dẫn, mà, như đã đề cập ở đây, có thể tạo thuận lợi cho việc thực hiện mạch lập trình được 300A (hoặc các dây của chúng) trong các mạch tích hợp khác, như CPU hoặc SOC. Như được minh họa, mạch lập trình được 300A bao gồm điện trở chính 302 và mạch điện trở phụ 303A. Như được minh họa, mạch điện trở phụ 303A có thể bao gồm điện trở phụ 304, cầu chì bảo vệ lập trình được 306 (ví dụ, eFuse), công tắc lập trình cầu chì bảo vệ 308, và công tắc cho phép lập trình 310. Như được minh họa chi tiết hơn dưới đây, mạch điện trở phụ 303A có thể được lập trình có chọn lọc để thực hiện việc thay đổi giá trị trọng số trong mô hình học máy dưới dạng sự kết hợp điện trở của điện trở chính 302 và điện trở phụ 304.

Điện trở chính 302 và điện trở phụ 304 được bao gồm dưới dạng một phần của mạch điện trở phụ 303A có thể được thiết lập song song để sử dụng điện trở kết hợp của điện trở chính 302 và điện trở phụ 304 để thực hiện giá trị trọng số trong mô hình học máy. Điện trở chính 302 có thể, theo một số khía cạnh, bao gồm điện trở có điện trở cao, và điện trở phụ 304 có thể bao gồm điện trở có điện trở cao hoặc thấp. Theo một số khía

cạnh, điện trở chính 302 và điện trở phụ 304 có thể bao gồm các điện trở có điện trở gần bằng nhau. Việc sử dụng các điện trở có điện trở gần bằng nhau cho cả điện trở chính 302 và điện trở phụ 304 tạo ra khoảng điều chỉnh tăng ở các trọng số được lập trình vào mỗi mạch lập trình được 300A, so với các điện trở có điện trở lệch nhau theo một hoặc nhiều bậc độ lớn. Điện trở chính 302 và/hoặc điện trở phụ 304 có thể được thực hiện dưới dạng các điện trở kim loại chính xác.

Mạch lập trình được 300A thường tạo ra dòng điện đầu ra là tổng của các dòng điện đi qua điện trở chính 302 và có thể là qua điện trở phụ 304. Đối với giá trị trọng số mà không biến đổi đáng kể giá trị của đầu vào, mạch lập trình được 300A có thể được lập trình để tách điện trở phụ 304 khỏi điện trở chính 302 sao cho dòng điện đầu ra của mạch lập trình được 300A không thay đổi hoặc thay đổi tối thiểu, do việc là điện trở tổng của mạch lập trình được 300A trong trường hợp này bằng với điện trở của điện trở chính 302. Đối với giá trị trọng số mà biến đổi giá trị của đầu vào, mạch lập trình được 300A có thể được lập trình sao cho điện trở chính 302 và điện trở phụ 304 được ghép nối điện theo cách song song để tạo giá trị đầu ra. Bằng cách ghép nối điện trở chính 302 và điện trở phụ 304 theo cách song song, dòng điện đầu ra N_j có thể được tính theo phương trình:
$$N_j = \frac{V_i}{total\ resistance},$$
 trong đó điện trở tổng là sự kết hợp của các điện trở của điện trở chính 302 và điện trở phụ 304, có giá trị nhỏ hơn điện trở riêng rẽ trong tổ hợp song song này và do đó dẫn đến dòng điện đầu ra lớn hơn tương đối N_j .

Để lập trình mạch lập trình được 300A, công tắc cho phép lập trình 310 được chuyển về trạng thái lập trình (ví dụ, được kích hoạt bằng công tắc đóng), và công tắc lập trình cầu chì bảo vệ 308 có thể được thiết lập để đặt điện thế lên cầu chì bảo vệ lập trình được 306 bên trong mạch điện trở phụ 303A. Công tắc cho phép lập trình 310 có thể được chuyển sang trạng thái lập trình, chẳng hạn, bằng các nhập tín hiệu điều khiển ENprog vào công tắc cho phép lập trình 310 để cho phép có chọn lọc việc lập trình của mạch điện trở phụ 303A. Công tắc lập trình cầu chì bảo vệ 308 có thể được ghép nối giữa nút thứ nhất 312 và nút thứ hai 314. Để lập trình cầu chì bảo vệ lập trình được 306 (ví dụ, làm nổi có chọn lọc cầu chì bảo vệ 306 và tách điện trở phụ 304 khỏi mạch lập trình được 300A), tín hiệu điều khiển P1 có thể được nhập vào công tắc lập trình cầu chì bảo vệ 308 để cho phép đặt điện thế P_{force} lên cầu chì bảo vệ lập trình được 306. Điện thế P_{force} có thể được nhập vào công tắc lập trình cầu chì bảo vệ 308 thông qua nút thứ nhất

312 và được đặt có chọn lọc lên cầu chì bảo vệ lập trình được 306 thông qua nút thứ hai 314. Nhìn chung, việc đặt công tắc cho phép lập trình 310 ở trạng thái lập trình và bất tín hiệu điều khiển P1 cho phép dòng điện lập trình chạy qua công tắc lập trình cầu chì bảo vệ 308, cầu chì bảo vệ lập trình được 306, và công tắc cho phép lập trình 310 ở ngoài nút điện thế tham chiếu (ví dụ, nối đất) cho mạch lập trình được 300A. Điện thế P_{force} , khi được đặt lên cầu chì bảo vệ lập trình được 306 thông qua công tắc lập trình cầu chì bảo vệ 308, có thể đủ để làm nổ cầu chì bảo vệ lập trình được 306 và tách điện cho điện trở phụ 304 khỏi điện trở chính 302. Điện thế P_{force} có thể là điện thế của nguồn cấp điện, chẳng hạn. Theo một số khía cạnh, điện thế cũng có thể được đặt lên cầu chì bảo vệ lập trình được 306, thông qua công tắc lập trình cầu chì bảo vệ 308, để hủy bỏ việc làm nổ cầu chì bảo vệ và ghép nối lại điện trở phụ 304 vào điện trở chính 302.

Fig.3B minh họa ví dụ về mạch lập trình được 300B bao gồm nhiều điện trở phụ trong mạch điện trở phụ 303B. Mạch lập trình được 300B có thể tương ứng với mạch lập trình được 202 bất kỳ được minh họa trên Fig.2. Như được minh họa, công tắc cho phép lập trình 310 và các công tắc lập trình cầu chì bảo vệ 308A và 308B có thể có tác dụng lập trình mạch lập trình được 300B bằng cách lần lượt ghép nối hoặc tách nối điện cho điện trở phụ 304A và 304B, với điện trở chính 302. Mặc dù chỉ có hai điện trở phụ 304A và 304B được minh họa trên Fig.3B, người đọc sẽ hiểu rằng một số khía cạnh có thể bao gồm nhiều hơn hai điện trở phụ được ghép nối chọn lọc theo cách song song. Tuy nhiên, để dễ mô tả và minh họa các khái niệm, phần mô tả sau sẽ sử dụng hai điện trở phụ.

Khi công tắc cho phép lập trình 310 được chuyển sang trạng thái lập trình (ví dụ, thông qua việc áp dụng tín hiệu điều khiển của ENprog cho công tắc cho phép lập trình 310), công tắc lập trình cầu chì bảo vệ 308A có thể được thiết lập (ví dụ, thông qua việc áp dụng tín hiệu P1 cho công tắc lập trình cầu chì bảo vệ 308A) để đặt điện thế P_{force_1} ở nút 312A lên cầu chì bảo vệ 306A thông qua nút thứ hai 314A. Tách riêng với hoạt động trên, công tắc lập trình cầu chì bảo vệ 308B có thể được thiết lập (ví dụ, thông qua việc áp dụng tín hiệu P2 lên công tắc lập trình cầu chì bảo vệ 308B) để đặt điện thế P_{force_2} ở nút 312B lên cầu chì bảo vệ 306B thông qua nút thứ hai 314B. Như đã đề cập, các điện thế P_{force_1} và P_{force_2} có thể lần lượt được sử dụng để làm nổ cầu chì bảo vệ 306A và 306B, mà có thể tách điện cho điện trở phụ 304 trong mạch điện trở phụ 303B khỏi điện trở chính 302, hoặc có thể được sử dụng để hủy bỏ việc làm nổ cầu chì bảo vệ 306A và 306B để ghép nối điện cho điện trở phụ 304A và 304B trong mạch điện trở phụ 303B

theo cách song song với điện trở chính 302. Mặc dù Fig.3B minh họa rằng các cầu chì bảo vệ 306A và 306B có thể được làm nổ độc lập có chọn lọc bằng cách bật công tắc cho phép lập trình 310 và công tắc lập trình cầu chì bảo vệ tương ứng 308A và/hoặc 308B, cần hiểu rằng cầu chì bảo vệ lập trình được 306 có thể được làm nổ cùng nhau nhờ sử dụng một công tắc lập trình cầu chì bảo vệ đơn nhất 308 (ví dụ, để tách tất cả trong số điện trở phụ 304 trong mạch điện trở phụ 303B khỏi điện trở chính 302 dựa trên giá trị trọng số để được thực hiện bởi mạch lập trình được 300B).

Theo một số khía cạnh, số lượng điện trở phụ 304 được ghép nối điện vào điện trở chính 302 có thể tương ứng với số lượng bit được sử dụng cho giá trị trọng số. Chẳng hạn, giá trị trọng số hai bit có thể được thực hiện trong mạch lập trình được 300A nhờ sử dụng điện trở chính 302 và một điện trở phụ đơn nhất 304, giá trị trọng số ba bit có thể được thực hiện trong mạch lập trình được 300B nhờ sử dụng điện trở chính 302 và hai điện trở phụ 304A và 304B, và v.v..

Để vận hành mạch lập trình được 300 để thực hiện hoạt động học máy, công tắc lập trình cầu chì bảo vệ 308 và công tắc cho phép lập trình 310 có thể được tắt (ví dụ, được giải hoạt bằng công tắc mở). Việc tắt công tắc lập trình cầu chì bảo vệ 308 và công tắc cho phép lập trình 310 cho phép mạch lập trình được 300 nhận nhận đầu vào điện biểu diễn vector đặc trưng hoặc đầu vào khác mà hoạt động học máy sẽ được thực hiện trên đó, chuyển đầu vào điện qua điện trở chính 302 và điện trở phụ được ghép nối 304 bất kỳ, và tạo giá trị đầu ra là sự kết hợp của đầu ra của điện trở chính 302 và điện trở phụ 304 bất kỳ được ghép nối song song với điện trở chính 302.

Fig.4 minh họa các hoạt động làm ví dụ 400 mà có thể được thực hiện bởi thiết bị có mạch lập trình được để thực hiện các hoạt động học máy, theo các khía cạnh của sáng chế. Như được minh họa, các hoạt động 400 bắt đầu ở khối 402, khi thiết bị nhận, từ máy chủ, giá trị trọng số của mô hình học máy được huấn luyện cần được triển khai trên mạch lập trình được. Như đã đề cập, giá trị trọng số có thể được tạo ra bởi máy chủ bằng cách xử lý tập dữ liệu huấn luyện trên máy chủ sao cho mô hình học máy có thể được triển khai trên thiết bị biên mà không kéo theo thiết bị biên trong việc huấn luyện mô hình học máy.

Ở khối 404, thiết bị lập trình mạch lập trình được bằng cách tách có chọn lọc một hoặc nhiều điện trở phụ ra khỏi trạng thái được ghép nối song song với điện trở chính,

dựa trên giá trị trọng số của mô hình học máy được huấn luyện. Để tách có chọn lọc một hoặc nhiều điện trở phụ, thiết bị có thể kích hoạt công tắc cho phép lập trình trong mạch lập trình được để chuyển mạch về chế độ ghi và kích hoạt công tắc lập trình cầu chì bảo vệ trong mạch lập trình được để cho phép điện thế được áp vào cầu chì bảo vệ lập trình được mà đủ để làm nổ cầu chì bảo vệ lập trình được và khiến cho điện trở phụ mà cầu chì bảo vệ lập trình được được ghép nối vào đó được tách điện khỏi điện trở chính trong mạch lập trình được. Khi mạch lập trình được được lập trình, thiết bị có thể thoát ra khỏi chế độ lập trình bằng cách mở (tắt) công tắc cho phép lập trình và công tắc lập trình cầu chì bảo vệ, cho phép các đầu vào khác chảy vào mạch lập trình được để xử lý.

Ở khối 406, hệ thống vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình. Để vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình, hệ thống có thể nhận dữ liệu để thực hiện hoạt động học máy trên đó. Dữ liệu có thể, chẳng hạn, là luồng bit tương ứng với hình ảnh chụp được, dữ liệu nguyên bản, v.v.. Hệ thống có thể chuyển đổi luồng bit thành tín hiệu tương tự, nhập tín hiệu tương tự này vào mạch lập trình được đã được lập trình, và nhận tín hiệu đầu ra từ mạch lập trình được đã được lập trình biểu diễn các kết quả của việc thực hiện hoạt động học máy trên tín hiệu đầu vào này.

Theo một số khía cạnh, việc tách có chọn lọc một hoặc nhiều điện trở phụ trong mạch lập trình được bao gồm làm nổ có chọn lọc một hoặc nhiều cầu chì bảo vệ lập trình được để tách điện một hoặc nhiều điện trở phụ trong mạch lập trình được khỏi điện trở chính trong mạch lập trình được. Mỗi của cầu chì bảo vệ lập trình được có thể được ghép nối nối tiếp với điện trở phụ của một hoặc nhiều điện trở phụ (ví dụ, trong mạch điện trở phụ 303A). Theo một số khía cạnh, làm nổ có chọn lọc một hoặc nhiều cầu chì bảo vệ lập trình được có thể bao gồm việc đóng có chọn lọc (kích hoạt hoặc bật) một hoặc nhiều công tắc, mỗi công tắc được ghép nối giữa nút thứ nhất được ghép nối vào nguồn điện thế và nút thứ hai được ghép nối vào điện trở phụ và cầu chì bảo vệ lập trình được.

Theo một số khía cạnh, hệ thống có thể còn lập trình lại mạch lập trình được dựa trên giá trị trọng số được cập nhật cho mô hình học máy được huấn luyện.

Theo một số khía cạnh, hệ thống có thể còn lập trình lại mạch lập trình được để triển khai mô hình học máy khác.

Theo một số khía cạnh, việc vận hành thiết bị nhờ sử dụng mạch lập trình được bao gồm bước nhận đầu vào bao gồm vector đặc trưng được biểu diễn dưới dạng đầu vào tương tự. Đầu ra có thể được tạo ra dựa trên tích của phần tử của vector đặc trưng và điện trở của điện trở chính và một hoặc nhiều điện trở phụ được ghép nối vào điện trở chính.

Theo một số khía cạnh, số lượng của một hoặc nhiều điện trở phụ trong mạch tương ứng với số lượng bit chính xác kết hợp với giá trị trọng số.

Theo một số khía cạnh, việc vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình bao gồm sử dụng đầu ra của mạch lập trình được kết hợp với đầu ra của một hoặc nhiều mạch lập trình được khác trong phép toán nhân và cộng dồn.

Theo một số khía cạnh, mô hình học máy được huấn luyện được dựa trên mạng noron sâu được huấn luyện để thực hiện tác vụ chuyên biệt.

Theo một số khía cạnh, thiết bị có thể bao gồm thiết bị máy tính di động. Theo một số khía cạnh, thiết bị có thể bao gồm xe có động cơ tự động.

Fig.5 minh họa các hoạt động làm ví dụ 500 mà có thể được thực hiện để lập trình thiết bị, theo các khía cạnh của sáng chế. Như đã đề cập, thiết bị có thể có nhiều đường từ, nhiều đường bit, và dây mạch lập trình được. Mỗi mạch lập trình được trong dây mạch lập trình được có thể được ghép nối với đường từ tương ứng trong số nhiều đường từ và đường bit tương ứng trong số nhiều đường bit. Việc lập trình mạch lập trình được có thể biến đổi điện trở tổng của mạch lập trình được bằng cách ghép nối hoặc tách có chọn lọc điện trở phụ nối tiếp với điện trở chính.

Như được minh họa, các hoạt động 500 bắt đầu ở khối 502, ở đó thiết bị lập trình mạch lập trình được trong dây mạch lập trình được bằng cách làm nổ có chọn lọc cầu chì bảo vệ nhờ sử dụng mạch lập trình. Cầu chì bảo vệ có thể được ghép nối nối tiếp với điện trở phụ, và điện trở phụ và cầu chì bảo vệ có thể được ghép nối giữa đường từ tương ứng và đường bit tương ứng. Điện trở chính cũng có thể được ghép nối giữa đường từ tương ứng và đường bit tương ứng. Để lập trình mạch lập trình được, thiết bị làm nổ có chọn lọc cầu chì bảo vệ nhờ sử dụng mạch lập trình. Bằng cách làm nổ có chọn lọc cầu chì bảo vệ được ghép nối với điện trở phụ, thiết bị tách điện trở phụ khỏi điện trở chính trong mạch lập trình được. Khi được tách, điện trở phụ có thể không còn được ghép nối giữa đường từ và đường bit, nhờ đó điều chỉnh điện trở tổng của mạch lập trình được thành tổ

hợp của điện trở của điện trở chính và điện trở phụ (nếu có) mà vẫn được ghép nối nối tiếp với điện trở chính.

Ở khối 504, hệ thống vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình. Để vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình, hệ thống có thể nhận dữ liệu để thực hiện hoạt động học máy trên đó. Dữ liệu có thể, chẳng hạn, là luồng bit tương ứng với hình ảnh chụp được, âm thanh, dữ liệu dạng văn bản, v.v.. Hệ thống có thể chuyển đổi luồng bit thành tín hiệu tương tự, nhập tín hiệu tương tự này vào mạch lập trình được đã được lập trình, và nhận tín hiệu đầu ra từ mạch lập trình được đã được lập trình biểu diễn các kết quả của việc thực hiện hoạt động học máy trên tín hiệu đầu vào.

Fig.6 minh họa ví dụ về môi trường mạng trong đó thiết bị biên có dây mạch lập trình được có thể được lập trình để triển khai mô hình học máy bởi máy chủ. Như được minh họa, môi trường mạng 600 bao gồm thiết bị biên 620, máy chủ huấn luyện 630, và kho dữ liệu huấn luyện 640, ít nhất là thiết bị biên 620 và máy chủ huấn luyện 630 được kết nối qua mạng 610.

Thiết bị biên 620 có thể đại diện cho thiết bị trên đó hoạt động học máy, như nhận diện đối tượng từ ảnh hoặc chuỗi ảnh chụp, sẽ được thực hiện. Thiết bị biên 620 có thể bao gồm, chẳng hạn, điện thoại di động, máy tính bảng, máy tính xách tay hoặc máy tính để bàn, xe có động cơ tự động, hoặc thiết bị khác mà các hoạt động học máy có thể được thực thi trên đó. Vì các thiết bị biên 620 có thể bị ràng buộc về công suất hoặc tài nguyên tính toán, có thể không thực tế khi huấn luyện mô hình học máy trên các thiết bị này; tuy nhiên, mô hình học máy được huấn luyện có thể được thực thi trên các thiết bị này.

Như được minh họa, thiết bị biên 620 bao gồm bộ thu phát 622, bộ xử lý 624, và mạch lập trình được 626. Bộ thu phát 622 nói chung nhận thông tin từ máy chủ huấn luyện 630 bao gồm các tham số trọng số mô hình cần được triển khai trong mạch lập trình được 626 nhằm thực thi hoạt động học máy trên thiết bị biên 620. Bộ thu phát 622 chuyển các tham số đến bộ xử lý 624, mà có thể tạo tín hiệu lập trình và nhập tín hiệu lập trình này vào mạch lập trình được 626 (ví dụ, được thực hiện như được thể hiện trong mạch lập trình được 300A trên Fig.3A). Như đã đề cập ở trên, tín hiệu lập trình có thể bao gồm các điện thế được áp cho công tắc lập trình cầu chì bảo vệ và công tắc cho phép lập trình để chuyển đổi mạch lập trình được 626 thành chế độ lập trình và tách có chọn

lọc một hoặc nhiều điện trở phụ khỏi điện trở chính trong mạch lập trình được 626. Sau khi mạch lập trình được lập trình, bộ xử lý 624 có thể chuyển đổi mạch lập trình được 626 về chế độ hoạt động, thu nhận dữ liệu trên đó để thực hiện hoạt động học máy, và chuyển dữ liệu thu nhận được đến mạch lập trình được đã được lập trình 626 để xử lý.

Máy chủ huấn luyện 630 nhìn chung đại diện cho máy chủ, cụm máy chủ, các thực thể điện toán đám mây, hoặc các thiết bị máy tính khác mà trên đó mô hình học máy có thể được huấn luyện để triển khai trên thiết bị biên 620. Như được minh họa, máy chủ huấn luyện 630 bao gồm bộ huấn luyện mô hình 632 và bộ thu phát 634. Bộ huấn luyện mô hình 632 có thể được tạo cấu hình để huấn luyện mô hình học máy nhờ sử dụng các kỹ thuật học máy không được giám sát, bán giám sát hoặc được giám sát hoàn toàn dựa trên tập dữ liệu truy vấn từ kho dữ liệu huấn luyện 640. Các kỹ thuật học không được giám sát có thể được thực thi trên tập dữ liệu huấn luyện thô, không có chú thích từ kho dữ liệu huấn luyện 640. Các kỹ thuật học bán giám sát có thể được thực thi trên tập dữ liệu huấn luyện từ kho dữ liệu huấn luyện 640 bao gồm tập dữ liệu có gắn nhãn và tập dữ liệu không có gắn nhãn mà có thể được sử dụng để kiểm tra mô hình học máy. Các kỹ thuật học có giám sát có thể được thực thi trên tập dữ liệu có gắn nhãn từ kho dữ liệu huấn luyện 640.

Sau khi bộ huấn luyện mô hình 632 huấn luyện mô hình học máy (ví dụ, xác định các trọng số để áp dụng cho dữ liệu nhập vào mô hình học máy), máy chủ huấn luyện 630 có thể xuất ra thông tin về mô hình học máy qua bộ thu phát 634 đến một hoặc nhiều thiết bị biên 620 và lệnh cho các thiết bị biên 620 lập trình mạch lập trình được 626 theo thông tin đầu ra này. Thông tin bổ sung có thể được xuất ra qua bộ thu phát 634 theo thời gian, chẳng hạn, để cập nhật mô hình học máy triển khai trên các thiết bị biên 620 và/hoặc thực hiện các mô hình học máy mới trên các thiết bị biên 620.

Fig.7 là sơ đồ khối minh họa kiến trúc phần mềm làm ví dụ 700 mà có thể tạo mô hình các hàm trí tuệ nhân tạo (artificial intelligence - AI). Nhờ sử dụng kiến trúc 700, các ứng dụng có thể được thiết kế mà có thể khiến các khối xử lý khác nhau của SOC 720 (chẳng hạn CPU 722, DSP 724, GPU 726, và/hoặc NPU 728) hỗ trợ việc nén và/hoặc giải nén video nhờ sử dụng các mô hình sinh liệu nhiều tầng, theo các khía cạnh của sáng chế.

Ứng dụng AI 702 có thể được tạo cấu hình để gọi các hàm được định rõ trong không gian người dùng 704 mà có thể, chẳng hạn, nén và/hoặc giải nén tín hiệu video (hoặc các phiên bản mã hóa của chúng) nhờ sử dụng các mô hình sinh liệu nhiều tầng. Ứng dụng AI 702 có thể, chẳng hạn, tạo cấu hình micro và máy ảnh phụ thuộc theo cách khác nhau vào việc liệu cảnh được nhận diện là văn phòng, hội trường, nhà hàng hay khung cảnh bên ngoài như hồ nước. Ứng dụng AI 702 có thể lập yêu cầu biên soạn mã chương trình kết hợp với thư viện được định rõ trong giao diện lập trình ứng dụng (application programming interface - API) của hàm AI 706. Yêu cầu này sau cùng có thể dựa vào đầu ra của mạng nơron sâu được tạo cấu hình để đưa ra hồi đáp suy luận dựa trên dữ liệu video và dữ liệu định vị, chẳng hạn.

Máy thời gian chạy (run-time) 708, mà có thể là mã biên soạn của khung thời gian chạy, có thể dễ tiếp cận hơn cho ứng dụng AI 702. Ứng dụng AI 702 có thể khiến máy thời gian chạy, chẳng hạn, yêu cầu suy luận ở khoảng thời gian cụ thể hoặc được khởi động bởi sự kiện phát hiện bởi giao diện người dùng của ứng dụng. Khi được khiến cho đưa ra hồi đáp suy luận, máy thời gian chạy sau đó có thể gửi tín hiệu đến hệ điều hành trong không gian hệ điều hành (operating system - OS) 710, như Linux Kernel 712, chạy trên SOC 720. Do đó, hệ điều hành có thể gây ra việc nén và/hoặc giải nén video nhờ sử dụng các mô hình sinh liệu nhiều tầng được thực hiện trên CPU 722, DSP 724, GPU 726, NPU 728, hoặc tổ hợp nào đó của chúng. CPU 722 có thể được truy cập trực tiếp bởi hệ điều hành, và các khối xử lý khác có thể được truy cập thông qua trình điều khiển, như trình điều khiển 714, 716, hoặc 718 lần lượt đối với DSP 724, GPU 726, hoặc NPU 728. Theo ví dụ minh họa, mạng nơron sâu có thể được tạo cấu hình để chạy trên tổ hợp của các khối xử lý, như CPU 722, DSP 724, và GPU 726, hoặc có thể chạy trên NPU 728.

Phương tiện để nhận có thể bao gồm, chẳng hạn, bộ thu phát (ví dụ, khối kết nối 110 được minh họa trên Fig.1 và/hoặc bộ thu phát 522 được minh họa trên Fig.5). Phương tiện để lập trình và phương tiện để lập trình lại có thể bao gồm bộ xử lý—như CPU 102, GPU 104, DSP 106, hoặc NPU 108 được minh họa trên Fig.1 và/hoặc bộ xử lý 624 được minh họa trên Fig.6—và/hoặc các công tắc, như công tắc 308, 308A, 308B, và/hoặc 310 trên các hình vẽ Fig.3A và Fig.3B. Phương tiện để vận hành có thể bao gồm bộ xử lý, như CPU 102, GPU 104, DSP 106, hoặc NPU 108 được minh họa trên Fig.1 và/hoặc bộ xử lý 624 được minh họa trên Fig.6. Phương tiện để tạo có thể bao gồm mạch

lập trình được 300A hoặc 300B được minh họa trên Fig.3A hoặc 3B và/hoặc dây 200 của mạch lập trình được minh họa trên Fig.2.

Các mạch minh họa khác nhau mô tả liên quan đến các khía cạnh được mô tả ở đây có thể được thực hiện trong hoặc với mạch tích hợp (integrated circuit - IC), như bộ xử lý, bộ xử lý tín hiệu số (digital signal processor - DSP), mạch tích hợp chuyên dụng (application-specific integrated circuit - ASIC), mảng cổng lập trình được theo trường (field-programmable gate array - FPGA), hoặc thiết bị logic lập trình được khác. Bộ xử lý có thể là bộ vi xử lý, nhưng theo cách khác, bộ xử lý có thể là bộ xử lý, bộ điều khiển, bộ vi điều khiển, hoặc máy trạng thái thông thường bất kỳ. Bộ xử lý có thể được triển khai dưới dạng kết hợp của thiết bị tính toán (ví dụ, kết hợp của DSP và bộ vi xử lý, nhiều bộ vi xử lý, một hoặc nhiều bộ vi xử lý kết hợp với lõi DSP, hoặc bất kỳ cấu hình tương tự khác).

Cũng lưu ý rằng các bước hoạt động được mô tả theo khía cạnh bất kỳ trong số các khía cạnh ví dụ ở đây được mô tả để làm ví dụ. Các hoạt động được mô tả có thể được thực hiện theo nhiều trình tự khác với trình tự được minh họa. Hơn thế nữa, các hoạt động được mô tả ở một bước hoạt động duy nhất có thể thực tế được thực hiện ở nhiều bước khác nhau. Ngoài ra, một hoặc nhiều bước hoạt động mô tả theo các khía cạnh minh họa có thể được kết hợp. Cần hiểu rằng các bước hoạt động được minh họa trên các biểu đồ lưu đồ có thể trải qua rất nhiều cải biến khác nhau như sẽ hiển nhiên đối với người có hiểu biết trung bình trong lĩnh vực kỹ thuật này. Người có hiểu biết trung bình trong lĩnh vực này sẽ hiểu rằng thông tin và tín hiệu có thể được biểu diễn bằng cách sử dụng các công nghệ và kỹ thuật khác nhau. Chẳng hạn, dữ liệu, chỉ lệnh, lệnh, thông tin, tín hiệu, bit, ký hiệu, và chip mà có thể được nói đến trong suốt bản mô tả ở trên có thể được biểu diễn bằng điện áp, dòng điện, sóng điện từ, trường hoặc các hạt từ, trường hoặc các hạt quang, hoặc kết hợp bất kỳ của chúng.

Như được sử dụng ở đây, cụm từ đề cập đến “ít nhất một trong số” danh sách các mục đề cập đến tổ hợp bất kỳ của các mục đó, bao gồm cả các thành phần riêng lẻ. Ví dụ, “ít nhất một trong số: a, b, hoặc c” được dự tính bao gồm a, b, c, a-b, a-c, b-c, và a-b-c, cũng như tổ hợp bất kỳ có nhiều phần tử giống nhau (ví dụ, a-a, a-a-a, a-a-b, a-a-c, a-b-b, a-c-c, b-b, b-b-b, b-b-c, c-c, và c-c-c hoặc thứ tự khác bất kỳ của a, b, và c).

Sáng chế được đưa ra để cho phép người có hiểu biết trung bình bất kỳ trong lĩnh vực này thực hành hoặc sử dụng các khía cạnh của sáng chế. Có nhiều dạng cải biến đối với các phương án của sáng chế sẽ được người có hiểu biết trung bình về lĩnh vực kỹ thuật này tìm ra một cách dễ dàng, và các nguyên lý chung đã nêu ra trong sáng chế có thể được áp dụng cho những phương án khác mà không bị coi là vượt ra ngoài phạm vi của sáng chế. Do đó, sáng chế không dự định bị hạn chế ở các ví dụ và phương án được mô tả ở đây mà sẽ có được phạm vi bảo hộ rộng nhất phù hợp với các nguyên tắc và dấu hiệu mới bộc lộ ở đây.

YÊU CẦU BẢO HỘ

1. Máy để lập trình thiết bị, máy này bao gồm:

nhiều đường từ;

nhiều đường bit; và

dây mạch lập trình được, trong đó mỗi mạch lập trình được được ghép nối với đường từ tương ứng trong số nhiều đường từ và với đường bit tương ứng trong số nhiều đường bit và bao gồm:

điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng,

điện trở phụ,

cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ, trong đó điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và

mạch lập trình được tạo cấu hình để làm nổ có chọn lọc cầu chì bảo vệ.

2. Máy theo điểm 1, trong đó:

mạch lập trình được còn bao gồm ít nhất một điện trở phụ bổ sung được ghép nối vào ít nhất một cầu chì bảo vệ bổ sung, ít nhất một điện trở phụ bổ sung và ít nhất một cầu chì bảo vệ bổ sung này được kết nối nối tiếp, và

mạch lập trình được tạo cấu hình để làm nổ có chọn lọc một hoặc nhiều trong số cầu chì bảo vệ và ít nhất một cầu chì bảo vệ bổ sung.

3. Máy theo điểm 1, trong đó mạch lập trình được còn bao gồm một hoặc nhiều công tắc, mỗi công tắc được ghép nối giữa nút thứ nhất được ghép nối vào nguồn điện thế và nút thứ hai được ghép nối vào điện trở phụ và cầu chì bảo vệ.

4. Máy theo điểm 1, trong đó máy này còn bao gồm bộ xử lý được tạo cấu hình để:

lập trình lại ít nhất một mạch trong số các mạch lập trình được trong dãy dựa trên giá trị trọng số được cập nhật cho mô hình học máy được huấn luyện.

5. Máy theo điểm 1, trong đó máy này còn bao gồm bộ xử lý được tạo cấu hình để:

lập trình lại ít nhất một mạch trong số các mạch lập trình được trong dãy để triển khai mô hình học máy khác.

6. Máy theo điểm 1, trong đó số lượng điện trở phụ trong mỗi mạch lập trình được tương ứng với số lượng bit chính xác kết hợp với các giá trị trọng số của mô hình học máy được huấn luyện.
7. Máy theo điểm 1, trong đó đầu ra của mỗi mạch lập trình được trong dãy mạch lập trình được sử dụng trong phép toán nhân và cộng dồn.
8. Máy theo điểm 1, trong đó dãy mạch lập trình được được tạo cấu hình để thực thi mô hình học máy được huấn luyện, và trong đó mô hình học máy được huấn luyện được dựa trên mạng nơ-ron sâu được huấn luyện để thực hiện tác vụ chuyên biệt.
9. Máy theo điểm 1, trong đó máy này là thiết bị máy tính di động.
10. Máy theo điểm 1, trong đó máy là xe có động cơ tự động.
11. Phương pháp lập trình thiết bị có nhiều đường từ, nhiều đường bit, và dãy mạch lập trình được, trong đó mỗi mạch lập trình được được ghép nối với đường từ tương ứng trong số nhiều đường từ và với đường bit tương ứng trong số nhiều đường bit, phương pháp này bao gồm các bước:

lập trình mạch lập trình được trong dãy mạch lập trình được bằng cách làm nổ có chọn lọc cầu chì bảo vệ, trong đó:

cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ,

điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và

điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng; và

vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

12. Phương pháp theo điểm 11, trong đó việc làm nổ có chọn lọc cầu chì bảo vệ bao gồm làm nổ có chọn lọc một hoặc nhiều trong số cầu chì bảo vệ được ghép nối với điện trở phụ và ít nhất một cầu chì bảo vệ bổ sung được ghép nối vào ít nhất một điện trở phụ bổ sung, ít nhất một cầu chì bảo vệ bổ sung và ít nhất một điện trở phụ bổ sung này được kết nối nối tiếp.

13. Phương pháp theo điểm 12, trong đó việc làm nỏ có chọn lọc cầu chì bảo vệ bao gồm việc đóng có chọn lọc một hoặc nhiều công tắc, mỗi công tắc được ghép nối giữa nút thứ nhất được ghép nối vào nguồn điện thế và nút thứ hai được ghép nối vào điện trở phụ và cầu chì bảo vệ.
14. Phương pháp theo điểm 11, trong đó mạch lập trình được lập trình để triển khai mô hình học máy được huấn luyện và được lập trình dựa trên giá trị trọng số cho mô hình học máy được huấn luyện.
15. Phương pháp theo điểm 14, còn bao gồm bước lập trình lại mạch lập trình được dựa trên giá trị trọng số được cập nhật cho mô hình học máy được huấn luyện.
16. Phương pháp theo điểm 14, còn bao gồm bước lập trình lại mạch lập trình được để triển khai mô hình học máy khác.
17. Phương pháp theo điểm 14, trong đó mô hình học máy được huấn luyện được dựa trên mạng nơron sâu được huấn luyện để thực hiện tác vụ chuyên biệt.
18. Phương pháp theo điểm 14, trong đó số lượng điện trở phụ trong mạch lập trình được tương ứng với số lượng bit chính xác kết hợp với giá trị trọng số của mô hình học máy được huấn luyện.
19. Phương pháp theo điểm 11, trong đó việc vận hành thiết bị nhờ sử dụng mạch lập trình được bao gồm:
- nhận đầu vào bao gồm vector đặc trưng được biểu diễn dưới dạng đầu vào tương tự; và
- tạo đầu ra dựa trên tích của phần tử của vector đặc trưng và điện trở của điện trở chính và điện trở phụ.
20. Phương pháp theo điểm 11, trong đó việc vận hành bao gồm sử dụng đầu ra của mạch lập trình được kết hợp với đầu ra của một hoặc nhiều mạch lập trình được khác trong dãy mạch lập trình được trong phép toán nhân và cộng dồn.
21. Phương pháp theo điểm 11, trong đó thiết bị bao gồm một trong số thiết bị máy tính di động hoặc xe có động cơ tự động.
22. Máy để lập trình thiết bị, máy này bao gồm:
- nhiều đường từ;

nhiều đường bit;

dây mạch lập trình được, trong đó mỗi mạch lập trình được được ghép nối với đường từ tương ứng trong số nhiều đường từ và với đường bit tương ứng trong số nhiều đường bit;

phương tiện để lập trình mạch lập trình được trong dây mạch lập trình được bằng cách làm nổ có chọn lọc cầu chì bảo vệ nhờ sử dụng phương tiện để lập trình, trong đó:

cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ,

điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và

điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng; và

phương tiện để vận hành máy nhờ sử dụng mạch lập trình được đã được lập trình.

23. Máy theo điểm 22, trong đó phương tiện để lập trình được tạo cấu hình để làm nổ có chọn lọc một hoặc nhiều trong số cầu chì bảo vệ được ghép nối với điện trở phụ và ít nhất một cầu chì bảo vệ bổ sung được ghép nối vào ít nhất một điện trở phụ bổ sung, ít nhất một cầu chì bảo vệ bổ sung và ít nhất một điện trở phụ bổ sung này được kết nối nối tiếp.

24. Máy theo điểm 23, trong đó phương tiện để lập trình được tạo cấu hình để làm nổ có chọn lọc cầu chì bảo vệ bằng cách đóng một hoặc nhiều công tắc, mỗi công tắc được ghép nối giữa nút thứ nhất được ghép nối vào nguồn điện thế và nút thứ hai được ghép nối vào điện trở phụ và cầu chì bảo vệ.

25. Máy theo điểm 22, trong đó mạch lập trình được được lập trình để triển khai mô hình học máy được huấn luyện, trong đó mạch lập trình được được lập trình dựa trên giá trị trọng số cho mô hình học máy được huấn luyện, và trong đó máy còn bao gồm phương tiện để lập trình lại mạch lập trình được dựa trên giá trị trọng số được cập nhật cho mô hình học máy được huấn luyện.

26. Máy theo điểm 22, trong đó mạch lập trình được được lập trình để triển khai mô hình học máy được huấn luyện và trong đó máy còn bao gồm phương tiện để lập trình lại mạch lập trình được để triển khai mô hình học máy khác.

27. Máy theo điểm 22, trong đó phương tiện để vận hành máy nhờ sử dụng mạch lập trình được đã được lập trình bao gồm:

phương tiện để nhận đầu vào bao gồm vector đặc trưng được biểu diễn dưới dạng đầu vào tương tự; và

phương tiện để tạo đầu ra dựa trên tích của phần tử của vector đặc trưng và điện trở của điện trở chính và điện trở phụ.

28. Máy theo điểm 22, trong đó mạch lập trình được lập trình để triển khai mô hình học máy được huấn luyện, trong đó mạch lập trình được lập trình dựa trên giá trị trọng số cho mô hình học máy được huấn luyện, và trong đó số lượng điện trở phụ trong mạch lập trình được tương ứng với số lượng bit chính xác kết hợp với giá trị trọng số cho mô hình học máy được huấn luyện.

29. Máy theo điểm 22, trong đó phương tiện để vận hành bao gồm phương tiện để sử dụng đầu ra của mạch lập trình được kết hợp với đầu ra của một hoặc nhiều mạch lập trình được khác trong dãy mạch lập trình được trong phép toán nhân và cộng dồn.

30. Phương tiện bất biến đọc được bằng máy tính có các lệnh được lưu trữ trên đó, mà khi được thực thi bởi bộ xử lý, thực hiện hoạt động để lập trình thiết bị bao gồm bộ xử lý, nhiều đường từ, nhiều đường bit, và dãy mạch lập trình được, trong đó mỗi mạch lập trình được được ghép nối với đường từ tương ứng trong số nhiều đường từ và với đường bit tương ứng trong số nhiều đường bit, hoạt động này bao gồm:

lập trình mạch lập trình được trong dãy mạch lập trình được bằng cách làm nổ có chọn lọc cầu chì bảo vệ, trong đó:

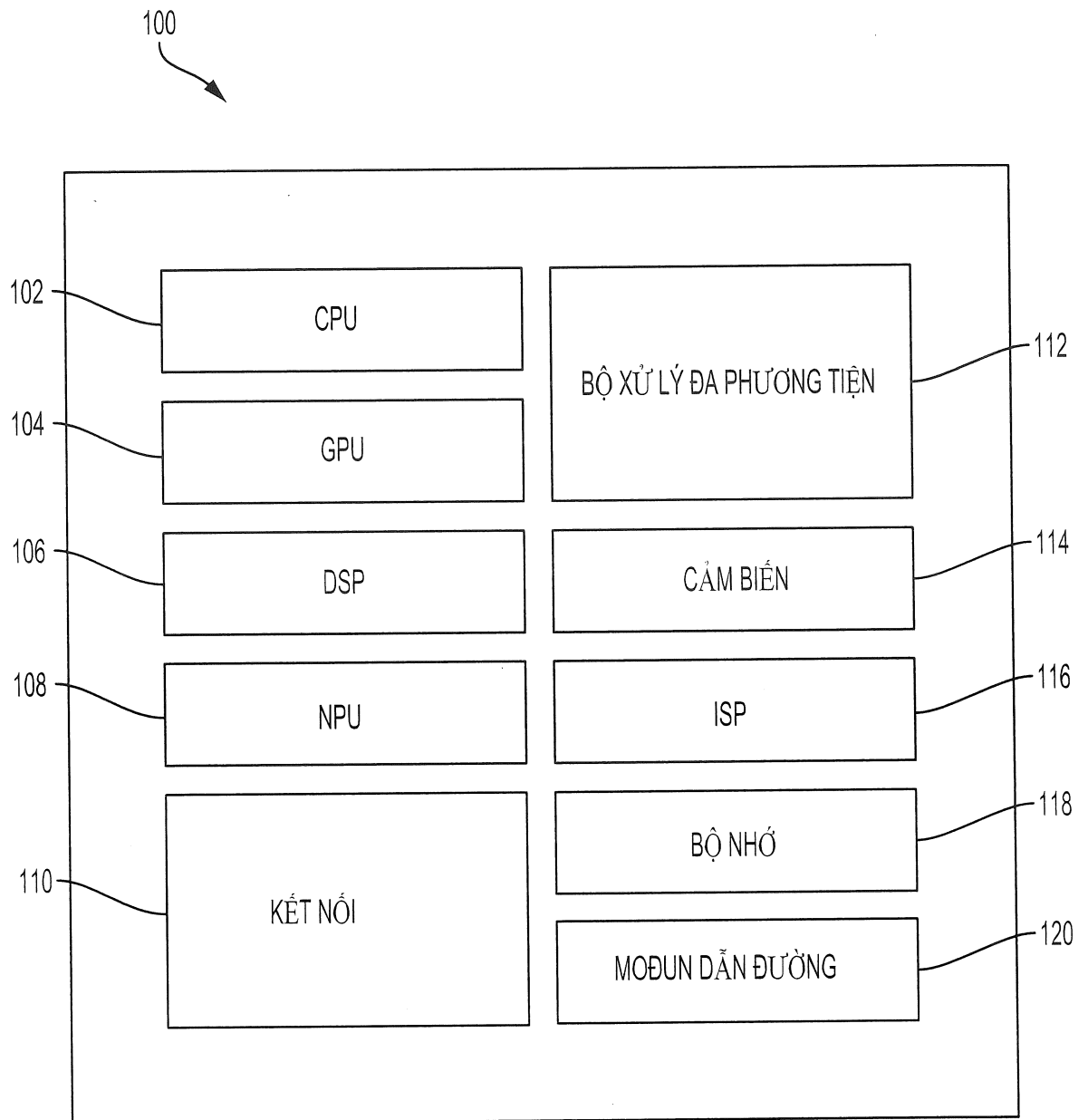
cầu chì bảo vệ được ghép nối nối tiếp với điện trở phụ,

điện trở phụ và cầu chì bảo vệ được ghép nối giữa đường từ tương ứng và đường bit tương ứng, và

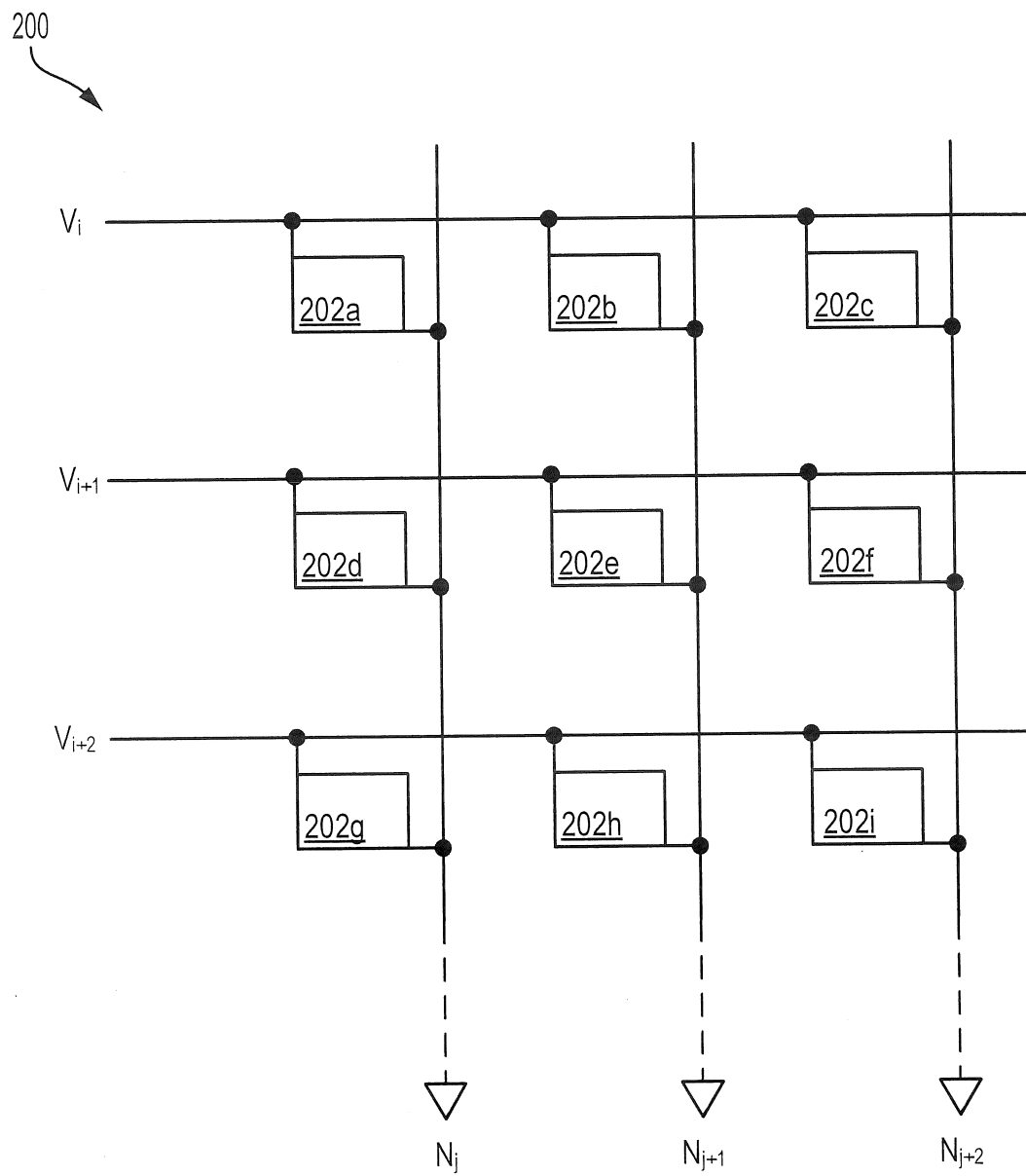
điện trở chính được ghép nối giữa đường từ tương ứng và đường bit tương ứng; và

vận hành thiết bị nhờ sử dụng mạch lập trình được đã được lập trình.

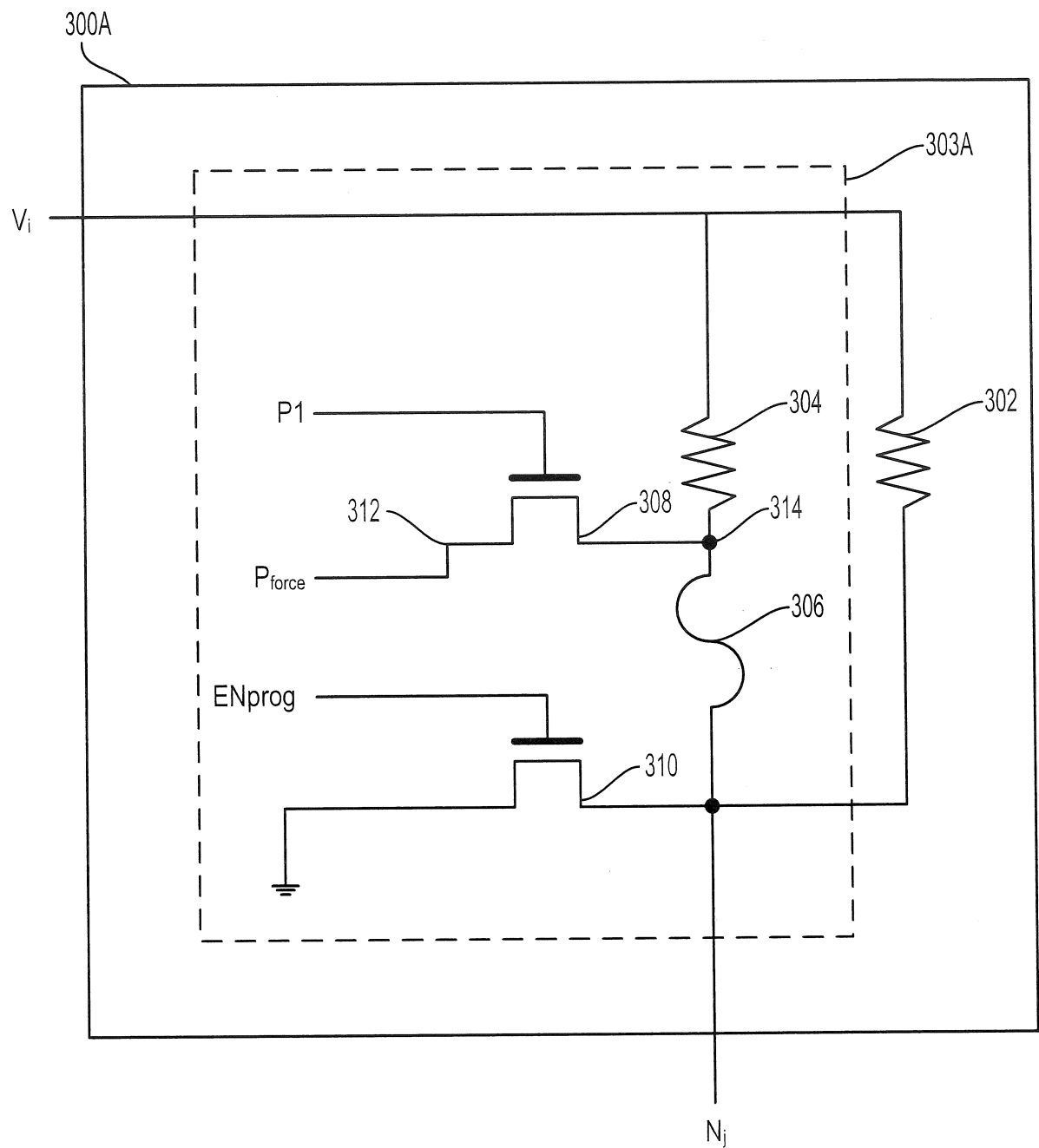
1/8

*Fig.1*

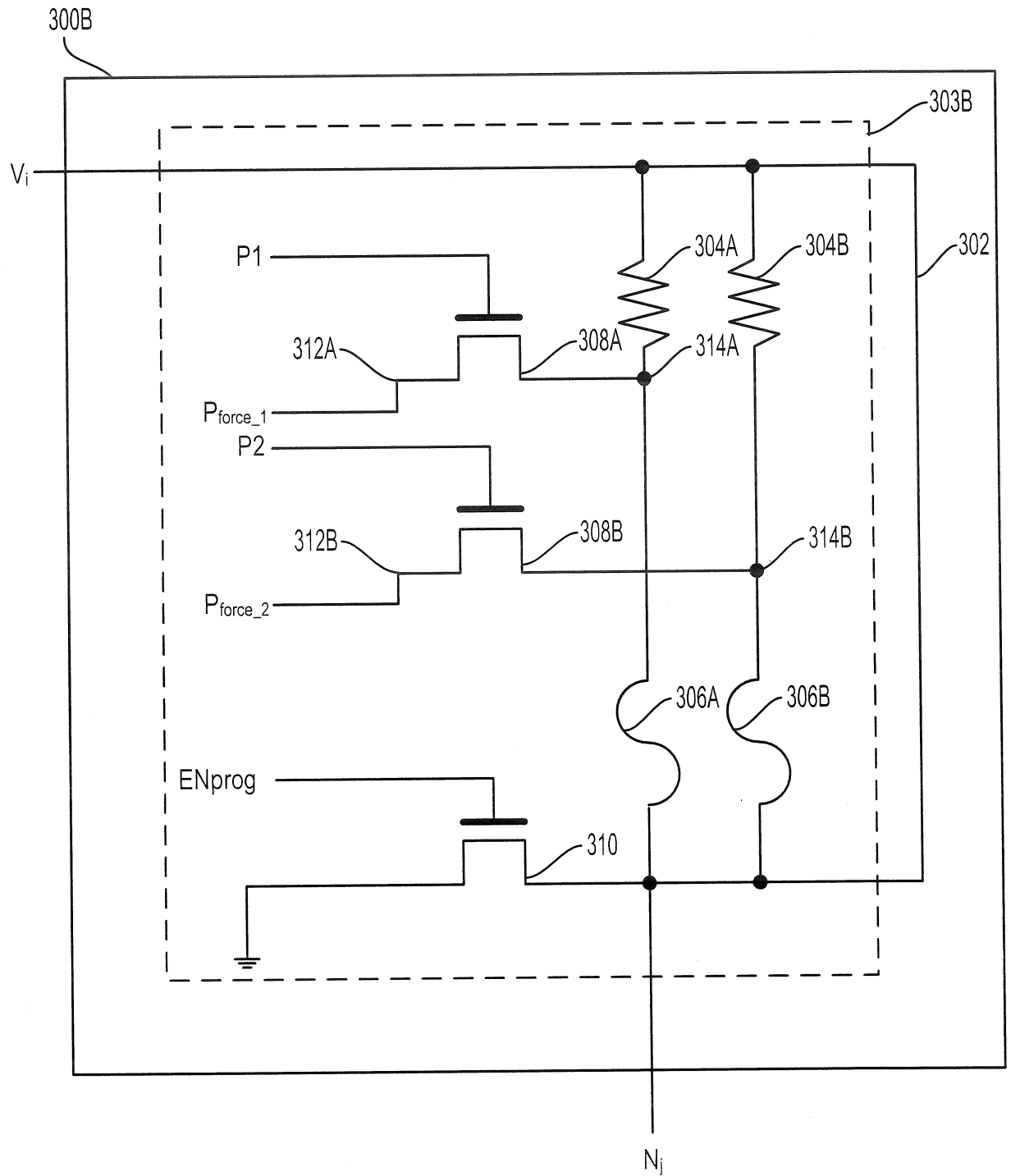
2/8

**Fig.2**

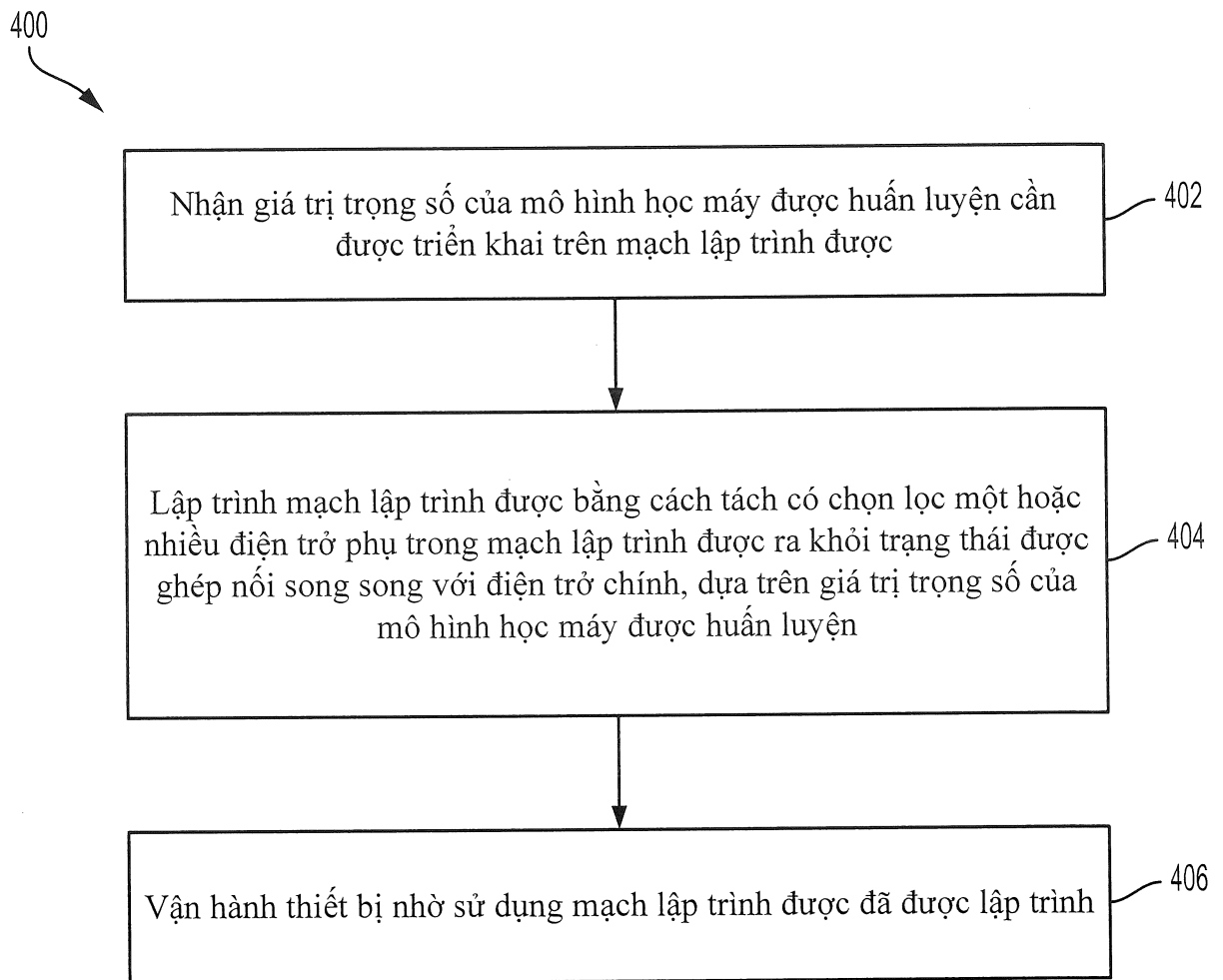
3/8

*Fig.3A*

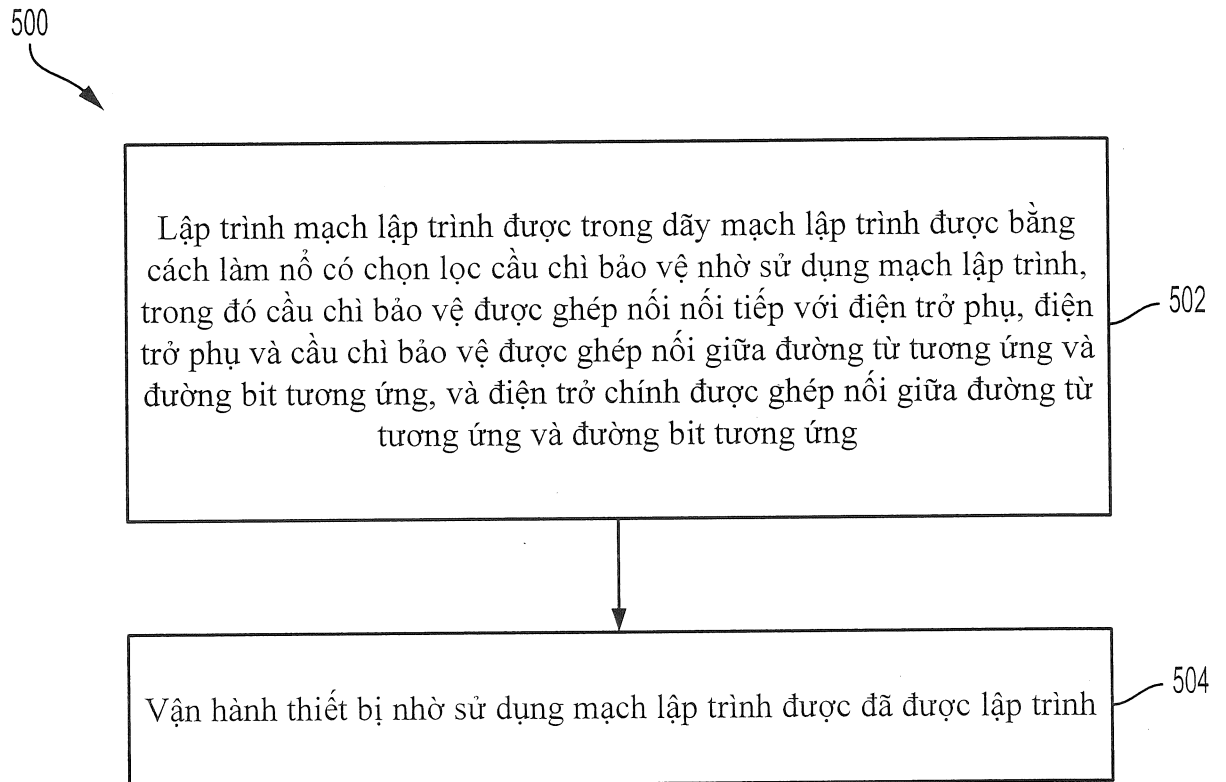
4/8

**Fig.3B**

5/8

**Fig.4**

6/8

*Fig.5*

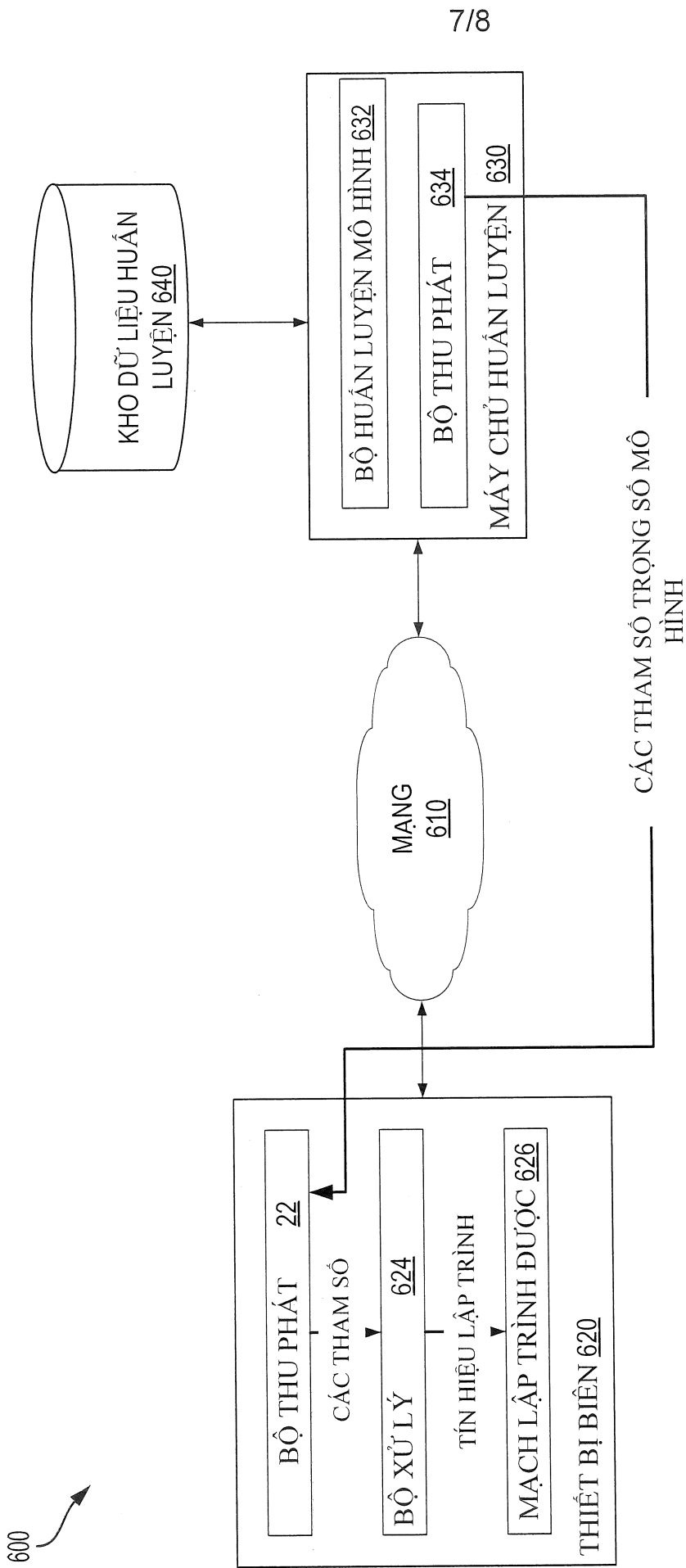


Fig.6

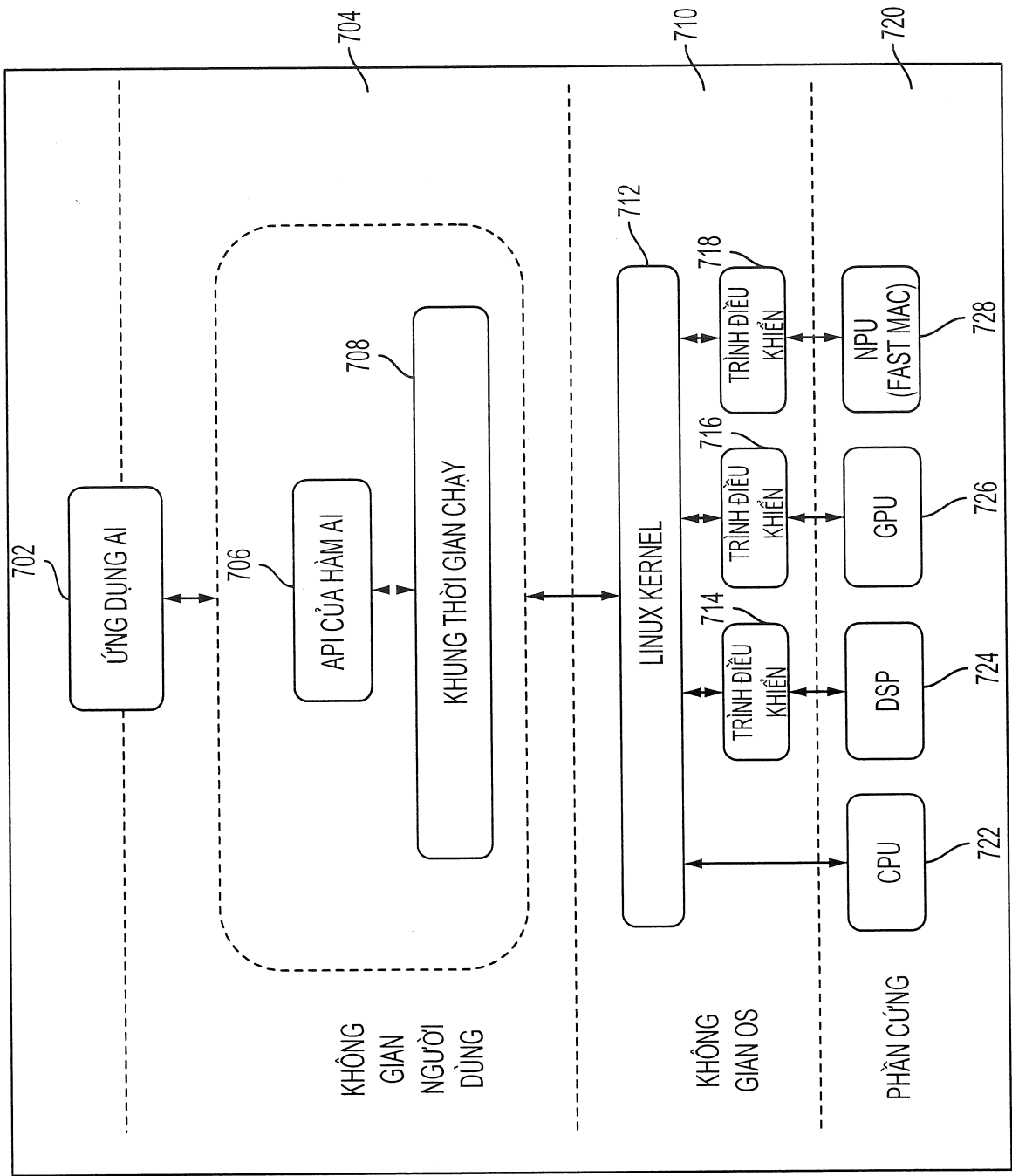


Fig. 7