

TÓM TẮT NHỮNG ĐÓNG GÓP MỚI CỦA LUẬN ÁN

Họ & tên NCS : Lê Minh MSNCS: 2027005

Thuộc chuyên ngành : Kỹ thuật Điện tử Khoá: 2020

Tên luận án : Nghiên cứu tối ưu và thực thi hệ điện toán mô phỏng
hệ thần kinh dùng điện trở nhớ

Người hướng dẫn chính: PGS. TS Trương Ngọc Sơn

Người hướng dẫn phụ :

Tóm tắt những đóng góp mới về lý luận và học thuật của luận án:

Hệ điện toán mô phỏng hệ thần kinh (neuromorphic computing system) hướng đến việc triển khai các mạng nơ-ron nhân tạo, các mô hình học sâu trực tiếp trên vi mạch, mô phỏng cách não bộ xử lý thông tin, thay vì dựa vào kiến trúc phần mềm truyền thống. Các hệ điện toán mô phỏng hệ thần kinh đã được triển khai khá thành công trên các nền tảng VLSI với công nghệ CMOS, tuy nhiên công nghệ này hiện tại đang tiến dần đến tới hạn về kích thước. Việc nghiên cứu các linh kiện mới, công nghệ mới cho các thiết kế hệ thống mô phỏng hệ thần kinh đang là một hướng nghiên cứu được quan tâm. Điện trở nhớ (memristor) có đặc tính là giá trị điện trở thay đổi khi dòng điện đủ lớn qua nó và khả năng lưu trữ giá trị điện trở khi mất điện, các đặc tính này giống với đặc tính của khớp nối thần kinh (synapse) trong mô hình hóa não bộ sinh học. Do đó điện trở nhớ được ứng dụng phổ biến trong thiết kế các hệ điện toán mô phỏng hệ thần kinh, thay cho công nghệ CMOS truyền thống. Ngoài khả năng có thể lập trình và lưu trữ, các mảng điện trở nhớ còn cho phép thực thi các phép tính chủ yếu trong mạng nơ-ron nhân tạo (phép nhân, phép cộng) ngay bên trong mảng điện trở nhớ ở mức mạch điện (in-memory computing), cho phép tăng tốc các phép toán xử lý trong mạng nơ-ron nhân tạo và thay cho việc thiết kế các vi mạch tính toán trên VLSI. Việc thực thi các tập trọng số và các phép tính toán trên bộ nhớ thường sử dụng 2 mảng điện trở nhớ bù nhau (Complementary crossbar array). Đây là một kiến trúc phổ biến và được minh chứng là hiệu quả trong thực thi hệ điện toán hệ thần kinh, cho phép lưu trữ và tính toán các phép tính của các trọng số có dấu hoặc thực hiện các hàm tính điểm giống nhau dựa trên phép toán XNOR. Mảng điện trở nhớ đồng nhất (Twin crrossbar array) sau đó được công bố cho thấy khả năng tiết kiệm được công suất đối với dữ liệu là các ảnh nhị phân có mật độ bit 1 thấp. Nhằm giảm số lượng điện trở nhớ để giảm các yếu tố như công suất và ảnh hưởng của sự biến thiên linh kiện trong quá trình chế tạo (process-variation), mảng điện trở nhớ đơn (Single crossbar array) được giới thiệu dựa trên

việc tối giản hóa hàm XNOR và được xem như là một kiến trúc tiềm năng cho các thiết kế hệ điện toán mô phỏng hệ thần kinh.

Các kiến trúc mảng điện trở nhớ đã được chứng minh là hiệu quả trong thiết kế hệ điện toán mô phỏng hệ thần kinh, tuy nhiên, các kiến trúc được đánh giá với các điều kiện và ràng buộc khác nhau. Để có thể lựa chọn một kiến trúc tối ưu, cần một phân tích và đánh giá tổng quan trên cùng một số điều kiện. Luận án thực hiện nghiên cứu, phân tích và đánh giá 3 kiến trúc mảng điện trở nhớ (kiến trúc mảng bù, kiến trúc mảng đồng nhất, kiến trúc mảng đơn) với các điều kiện như biến thiên điện trở nhớ, nhiễu trên dữ liệu, công suất tiêu thụ và số lượng điện trở nhớ thực hiện một khớp nối. Đóng góp đầu tiên của luận án là đưa ra được kết quả phân tích, đánh giá tổng quát 3 kiến trúc mảng điện trở nhớ được sử dụng để thực thi hệ điện toán mô phỏng hệ thần kinh nhằm xác định kiến trúc tiềm năng đáp ứng được nhiều yếu tố thiết kế.

Từ kết quả phân tích và đánh giá, luận án cho thấy mảng điện trở nhớ đơn có nhiều ưu điểm hơn so với 2 mảng còn lại, tuy nhiên, song song đó nó cũng có hạn chế đối với dữ liệu nhị phân có mật độ dữ liệu thấp. Luận án đã đề xuất một giải pháp để tối ưu kiến trúc mảng điện trở nhớ đơn thông qua việc phân tích mô hình toán học của hàm XNOR. Các thiết kế và mô phỏng ở mức mạch điện cho thấy kiến trúc tối ưu này đã khắc phục được ảnh hưởng của mật độ dữ liệu đến hoạt động của mảng điện trở nhớ, đồng thời giữ được những ưu điểm vượt trội như sự tiết kiệm công suất và số lượng điện trở nhớ trên mảng, tính kháng nhiễu và sai số điện trở. Đây cũng là một đóng góp thứ 2 của luận án và có ý nghĩa trong việc áp dụng kiến trúc mảng điện trở nhớ trong thực thi hệ điện toán mô phỏng hệ thần kinh.

Kiến trúc mảng điện trở nhớ tối ưu sau đó được ứng dụng để thiết kế một mạng nơ-ron XNOR (XNOR-Net) nhiều lớp. Mạng nơ-ron XNOR là một mô hình mạng nơ-ron nhị phân được triển khai hiệu quả trên các thiết bị có tài nguyên giới hạn. Trong kiến trúc này, luận án đã ứng dụng kiến trúc mảng điện trở nhớ đã được tối ưu cho các lớp ẩn. Đối với lớp ngõ ra, bằng lập luận phân tích, luận án cho thấy có thể sử dụng kiến trúc mảng điện trở đơn tối giản nhằm giảm thiểu được mạch ngoại vi, tiết kiệm công suất và diện tích mà không làm ảnh hưởng đến sai số nhận dạng của mạch. Đây cũng là một đóng góp của luận án trong ứng dụng kiến trúc mảng điện trở nhớ thiết kế các mô hình học sâu trên nền tảng vi mạch.

Tp. Hồ Chí Minh, ngày tháng năm

Nghiên cứu sinh

(Ký và ghi rõ họ tên)

Lê Minh