

I. GIỚI THIỆU

1 BỐI CẢNH

Đô thị hóa nhanh chóng và thách thức về phát triển bền vững đã cho thấy được tầm quan trọng của việc xây dựng “Thành phố Xanh và Thông minh” trong việc nâng cao chất lượng cuộc sống. Trí tuệ Nhân tạo (AI), đóng vai trò quan trọng trong việc phân tích dữ liệu chuỗi thời gian thu thập từ mạng lưới cảm biến dày đặc trong thành phố (ô nhiễm tiếng ồn, chất lượng không khí, mực nước, tiêu thụ năng lượng ...). Việc xử lý lượng dữ liệu khổng lồ này đòi hỏi các giải pháp tính toán hiệu năng cao, tiêu thụ năng lượng thấp và có khả năng triển khai tại biên giúp giảm độ trễ, giảm tải cho hệ thống máy chủ, tiết kiệm băng thông, tăng cường tính riêng tư, bảo mật.

2 MỤC TIÊU

Đề tài này nhằm phát triển FlexAI-1D, một bộ tăng tốc AI trên FPGA, với các mục tiêu cụ thể đóng góp vào việc xây dựng Thành phố Xanh và Thông minh:

- Khả năng cấu hình & Hỗ trợ đa dạng Kiến trúc
- Tiêu thụ năng lượng thấp
- Tốc độ xử lý cao
- Tối ưu cho 1D-CNN trong giám sát đô thị
- Tích hợp SoC
- Hiệu quả chi phí

II. KIẾN TRÚC ĐỀ XUẤT

Kiến trúc 1D-CNN phức tạp như mạng dựa trên khối Inception

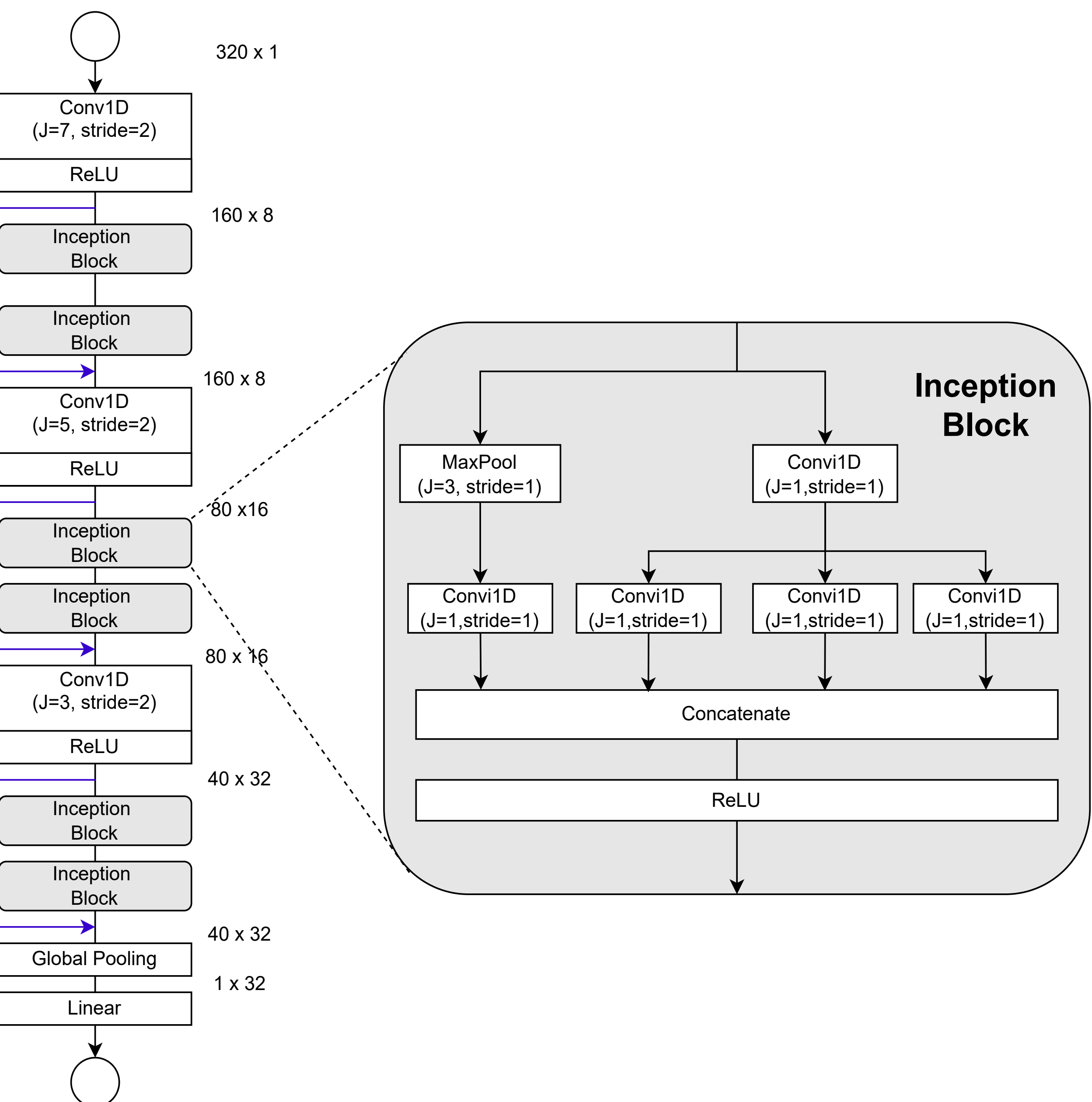


Fig 4. Kiến trúc tham khảo Inception

Bên cạnh khối chính, có các khối phụ trợ sau:

- Khối giao tiếp: giao tiếp với PS và bộ nhớ ngoài, truyền dữ liệu và nhận lệnh điều khiển/cấu hình từ CPU.
- Khối Điều khiển Trung tâm: giải mã Tập lệnh Cấu hình chuyên dụng, quản lý các luồng dữ liệu phức tạp cho các cấu trúc liên kết mạng đa dạng
- Bộ nhớ Ngủ cảnh (Context Memory): lưu trữ chuỗi các lệnh cấu hình đã được biên dịch
- Bộ nhớ Trọng số/Độ lệch (Weights/Bias Memory) SRAM/BRAM trên chip dùng để lưu trữ trọng số và các giá trị độ lệch của mô hình CNN đang được thực thi

Kiến trúc FlexAI-1D khi tích hợp vào SoC với các thành phần được liệt kê bên trên

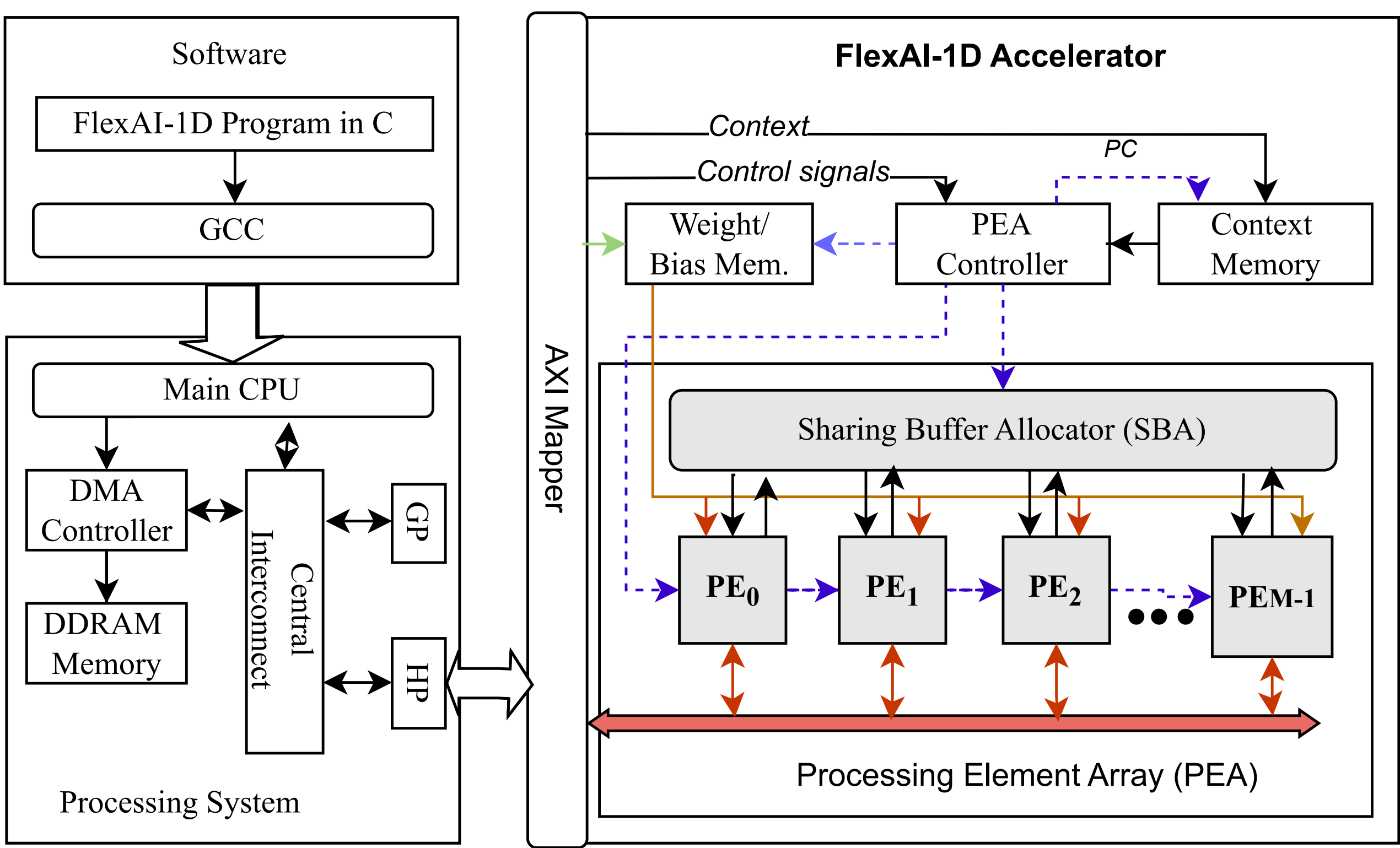


Fig 2. 1D-CNN trong SoC

Để cung cấp dữ liệu cho các Khối Xử lý song song, khối Phân bổ Bộ đệm Chia sẻ (SBA) được sử dụng để định tuyến các sample dữ liệu đầu vào từ bộ đệm chung đến các PE cụ thể, giảm thiểu truy cập bộ nhớ ngoài.

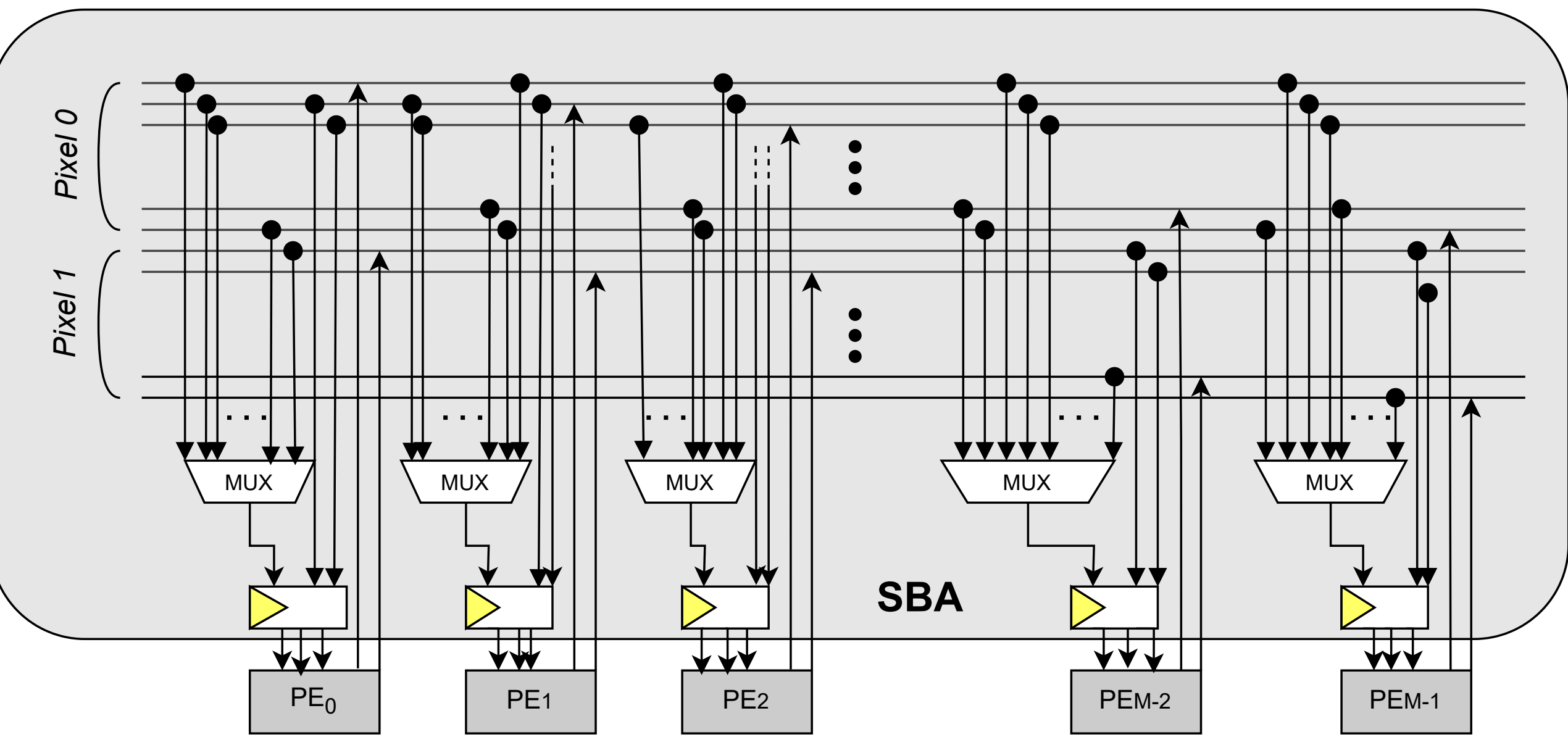


Fig 3. Kiến trúc tổng quan khối SBA

Khả năng tính toán dựa trên Mạng các Khối Xử lý (PEA), trong đó mỗi Khối Xử lý (PE) có kiến trúc như 4. Mỗi PE chứa các bộ nhớ cục bộ để lưu trữ một phần dữ liệu, giảm thiểu việc truy cập vào bộ nhớ đệm chung.

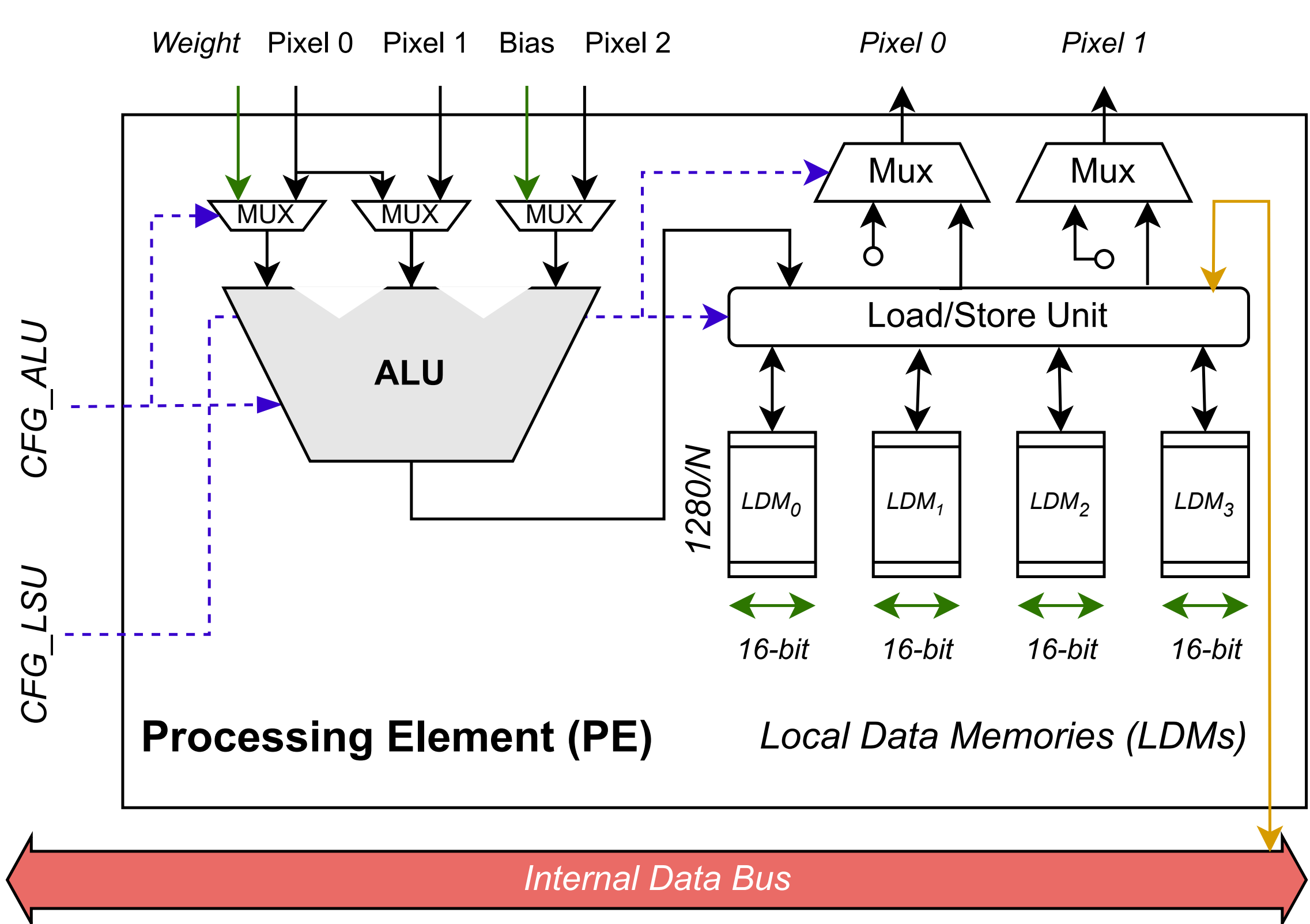


Fig 4. Kiến trúc tổng quan khối PE

Bên trong mỗi PE sử dụng ALU chuyên biệt cho 1D-CNN

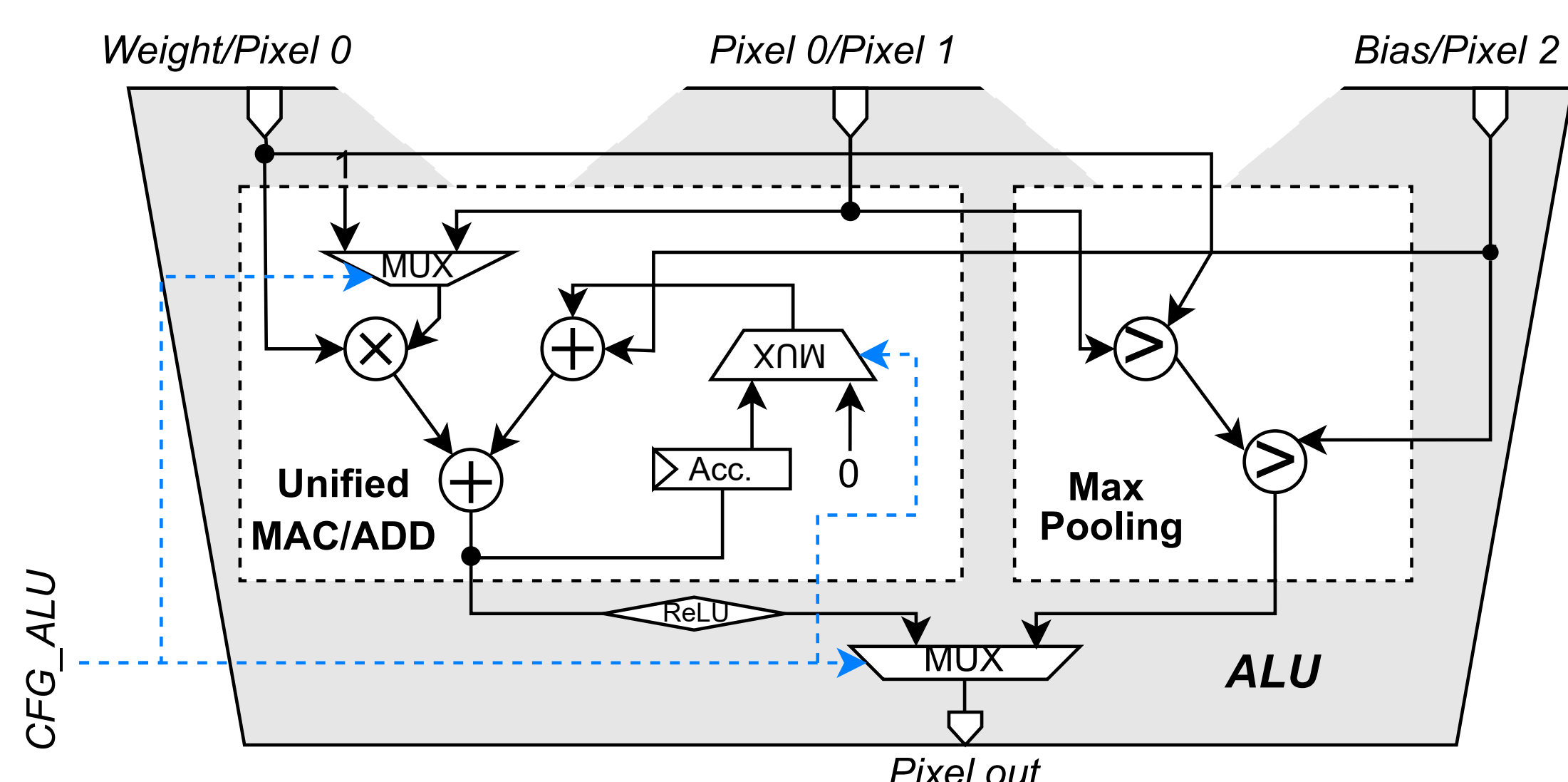


Fig 4. Kiến trúc tổng quan khối ALU

III. ĐẶC TÍNH VÀ ƯU ĐIỂM

1. Đặc tính

- Khả năng cấu hình toàn diện và hỗ trợ kiến trúc tổng quát
- Tiêu thụ năng lượng thấp
- Tốc độ xử lý cao
- Tối ưu tài nguyên FPGA
- Độ chính xác cao

2. Ưu điểm

- **Hiệu quả chi phí cho hạ tầng đô thị:** So với ASIC, FlexAI-1D trên FPGA có chi phí thấp hơn để triển khai các hệ thống giám sát đa dạng, cần cập nhật thường xuyên trong môi trường đô thị. So với CPU/GPU, FlexAI-1D cung cấp hiệu năng/điện năng vượt trội góp phần vào tính bền vững của thành phố.
- **Nâng cao hiệu quả giám sát và quản lý đô thị:** Khả năng xử lý tại biên giúp các hệ thống hoạt động hiệu quả hơn, phản ứng nhanh hơn với các sự kiện.
- **Thúc đẩy đổi mới trong Thành phố Thông minh:** Khả năng thích ứng cao cho phép thử nghiệm và triển khai nhanh chóng các mô hình AI mới.

IV. PHẠM VI ỨNG DỤNG

- **Thích ứng với các thuật toán mới:** FlexAI-1D có thể dễ dàng cấu hình để chạy các mô hình mới. Điều này đặc biệt quan trọng vì lĩnh vực AI phát triển rất nhanh.
- **Giải quyết các bài toán chuyên biệt cho từng đô thị:** FlexAI-1D cung cấp nền tảng phần cứng linh hoạt để triển khai nhanh chóng các giải pháp AI phù hợp với nhu cầu của từng đô thị.
- **Giám sát Môi trường Toàn diện (Xanh):** phân tích tiếng ồn đô thị, giám sát chất lượng không khí và nước ...
- **Quản lý Năng lượng Thông minh và Tối ưu hóa (Xanh & Thông minh):** phân tích dữ liệu từ đồng hồ điện thông minh
- **Giao thông Thông minh và An toàn (Thông minh):** Phân tích dữ liệu từ cảm biến vòng từ, camera hoặc cảm biến rung động trên cầu đường
- **An ninh và An toàn Công cộng Nâng cao:** phân tích âm thanh đô thị
- **Nông nghiệp Đô thị Thông minh và Bền vững (Xanh):** Phân tích dữ liệu từ cảm biến độ ẩm đất, nhiệt độ, ánh sáng, và các chỉ số khác trong các hệ thống nông nghiệp đô thị