**Tối ưu hóa năng lượng**

**của các bộ xử lý mạng cảm biến**

**hoạt động dưới ngưỡng điện áp**

1. **Bài toán**:

Các ứng dụng mạng cảm biến (sensor network application) xuất hiện ở nhiều lĩnh vực, từ các ứng dụng giám sát y khoa, kiểm tra chất lượng, giám sát quân sự. Các ứng dụng này có tính chất là nhỏ, hạn chế năng lượng, nhu cầu tính toán không cao. Chúng thường là các thiết bị nhỏ, được cung cấp năng lượng bằng pin hay có thêm các thiết bị tổng hợp năng lượng (như tế bào quan điện). Các thiết bị này được đặt vào môi trường mà con người muốn giám sát như giám sát hoạt động của nhãn cầu thì thiết bị phải đủ nhỏ để có thể đặt vào lớp biểu bì của nhãn cầu. Các bộ xử lý của ứng dụng mạng cảm biến không đòi hỏi phải có khả năng tính toán cao. Do đó, bài toán mà bài báo giải quyết là đưa ra các thiết kế tối ưu hóa năng lượng cho bộ xử lý mạng cảm biến.

1. **Cách nhóm nghiên cứu giải quyết bài toán**:

Đọc bài báo, em thấy nhóm nghiên cứu đã đi từ nghiên cứu rút ra kết luận và kiểm nghiệm các kết luận bằng các mạch mẫu. Cụ thể gồm các bước sau:

* Phân tích bài toán
* Thiết kế mạch hoạt động dưới ngưỡng điện áp
* Phân tích sự đánh đổi trong kiến trúc
  1. **Phân tích bài toán**:
     1. Nhóm nghiên cứu phân tích bài toán mà họ gặp phải, bao gồm:
        + Các yêu cầu mà ứng dụng mạng cảm biến phải đảm bảo.

Hiệu suất tính toán của các bộ xử lý phải đáp ứng nhu cầu của ứng dụng

Do yếu tố hạn chế năng lượng nên bộ xử lý mạng cảm biến này phải tiêu thụ năng lượng ít nhất có thể.

* + - * Các yếu tố có thể dẫn đến lời giải của bài toán.

Có thể giảm nhu cầu năng lượng của ứng dùng bằng cách giảm điện áp, các thiết bị này sẽ hoạt động ở điện áp dưới ngưỡng (từ 400mV trở xuống).

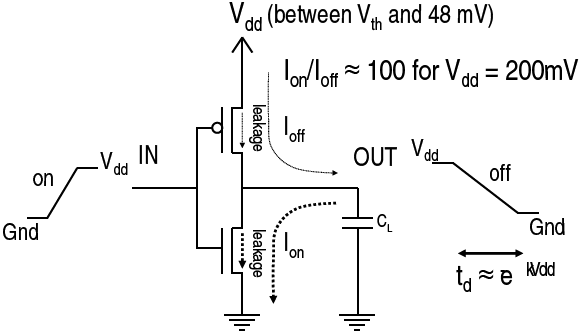
Khả năng tính toán của bộ xử lý không đòi hỏi cao, do đó có thể giảm tần số xung clock xuống, như vậy sẽ tiết kiệm năng lượng hơn.

Nghiên cứu tập trung xem xyét để giảm năng lượng mỗi câu lệnh tiêu thụ. Nhóm nghiên cứu xem các bộ xử lý mạng cảm biến này đều đã ứng dụng các kỹ thuật tiết kiệm năng lượng trong thời gian mà nó không hoạt động (ở trạng thái chờ).

* + 1. Ngoài ra, nhằm mục đích đánh giá hiệu quả thực thi, hiệu suất, khả năng tính toán và nhu cầu năng lượng của các bộ xử lý mạng cảm biến, họ cũng đưa ra tập hợp điểm chuẩn (benchmark test). Các điểm chuẩn này được chia thành 3 loại: các giải thuật giao tiếp (***Communication Algorithms***), xử lý tính toán (***Computational processing***) và các giải thuật cảm biến (***Sensing Algorithms***). Chúng ta có thể xem chi tiết của nó ở Table 1 trong bài báo.
  1. **Thiết kế mạch hoạt động dưới ngưỡng điện áp**:

Trong bài báo này, nhóm nghiên cứu tập trung thảo luận trong việc thiết kế kiến trúc hiệu quả năng lượng ở những mức điện áp dưới ngưỡng. Có 2 vấn đề mà họ gặp phải là:

* + - * ***Xác định điện áp hoạt động tối ưu năng lượng***: ở điện áp trên ngưỡng thì giảm điện áp lương cải thiện được hiệu quả năng lượng. Còn ở điện áp dưới ngưỡng, điều này không còn đúng nữa, rò rĩ năng lượng tăng tỷ lệ với điện áp. Do đó, nguồn năng lượng cung cấp tồn tại khi năng lượng mỗi lệnh tiêu thụ là nhỏ nhất.
      * ***Xác định các thông số thiết kế***, xác định một bản thiết kế hiệu quả năng lượng khi xử lý với nguồn cung cấp tối ưu năng lượng. Hiểu được các tham số này sẽ là chìa khóa cho việc thiết kế các kiến trúc tối ưu hóa năng lượng hoạt động dưới ngưỡng. Bài báo cáo trình bày cách mà các tham số này khác với các yếu tố quan trọng khác đã được cân nhắc trong xử lý trên ngưỡng điện áp.
    1. **Hoạt động mạch ở điện áp dưới ngưỡng**: được trình bày chi tiết trong phần 3.1 của bài báo.



**Hình 2**: Inverter ở điện áp dưới ngưỡng.

* + - * Vấn đề với inverter là khi điện áp trên ngưỡng thì inverter vẫn hoạt động bình thường; nhưng với điện áp dưới ngưởng thì đầu ra của inverter đều ở trang thái đóng 0V, bất kể giá trị logic của đầu vào inverter. Nhóm nghiên cứu phát hiện được sự khác biệt về dòng rò rỉ dưới ngưỡng của 2 transistor NMOS & PMOS tạo ra sự phóng dung lượng để tạo nên kết quả chuyển đổi tín hiệu. Cụ thể là dòng rò rỉ của transistor NMOS bằng khoảng 100 lần dòng rò rỉ của transistor PMOS. Ngoài ra, tỷ lệ Ion/Ioff là khoảng 100, vẫn còn đủ cao để có được biến đổi điện áp đầu ra có nghĩa.
      * Nếu chúng ta giảm điện áp cung cấp từ 200mV xuống 100mV, khi đầu vào là Vdd thì dòng rò rỉ của transistor NMOS giảm theo hàm mũ, chỉ còn khoảng 10 lần so với các transistor PMOS. Do đó, độ trể của inverter tăng 10 lần nếu điện áp cung cấp giảm 2 lần. Các phụ thuộc hàm mũ dòng rò rỉ Vgs dẩn đến sự phụ thuộc hàm mũ của độ trể mạch trên điện áp cung cấp. Ta có biểu thức:

Với k là một hằng số phụ thuộc nhiệt độ & công nghệ.

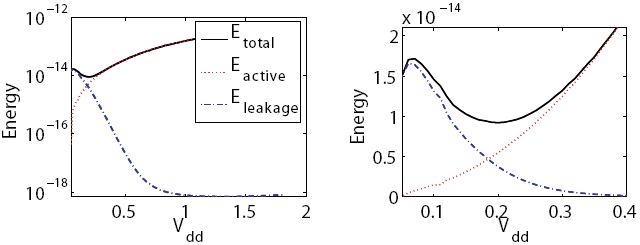
* + - * Ngoài ra, giảm điện áp cung cấp xuống 100mV đã giảm tỷ lệ Ion/Ioff chỉ còn 10, kết quả là điện áp đầu ra giao động. Khi nguồn điện áp giảm hơn nữa, sự giao động điện áp đầu ra sẽ giảm đến độ không thể xác định được giá trị logic. Công nghệ hiện tại thì điện áp cung cấp tối thiểu là 48mV.
    1. **Kiến trúc tối ưu năng lượng**:
       - Năng lượng tiêu tốn cho mỗi câu lệnh có thể được biểu diển như sau:

Với Ecycle là năng lượng tiêu thụ trung bình cho mỗi chu kỳ; CPI là số chu kỳ trung bình của mỗi lệnh. Rỏ ràng là CPI độc lập với điện áp cung cấp, nhưng nó rất quan trọng khi tạo nên cân bằng kiến trúc.

* + - * Năng lượng tổng thể của mỗi chu trình biểu diễn bằng tổng năng lượng động & năng lượng rò rỉ, như sau:

Với α là yếu tố hoạt động; Cs là tổng điện dung của mạch; Vdd là điện áp cung cấp; Ileak là dòng rò rỉ; tclk là thời gian chu kỳ xung clock.

* + - * Từ biểu thức trên, rỏ ràng năng lượng động giảm theo hàm bình phương ở điện áp trên & dưới ngưỡng. Tuy nhiên, hành vi rò rỉ năng lượng là khác nhau ở các khoảng xử lý superthreshold và subthreshold.

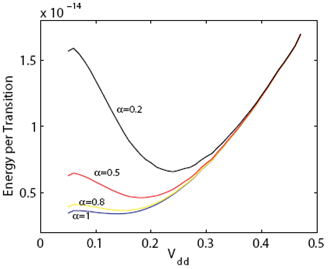


**Hình 3**: Biểu đồ hàm năng lượng theo điện áp

Ở điện áp cung cấp superthreshold, thời gian chu kỳ xung clock tclk tăng tuyến tính khi điện áp cung cấp giảm dần. Cùng thời gian đó, dòng rò rỉ giảm gần như tuyến tính. Năng lượng rò rỉ vẫn gần như không đổi. Do đó, giảm điện áp cung cấp cải thiện hiệu quả năng lượng tổng thể do năng lượng động giảm. Hình 3a cho thấy kết quả mô phỏng SPICE cho một chuỗi 20 trạng thái trong công nghệ 0.18μm.

Tuy nhiên, ở mức điện áp subthreshold, chu kỳ xung clock tăng theo hàm cấp số nhân của điện áp; trong khi dòng rò rỉ vẫn tiếp tục giảm gần như tuyến tính. Do đó, năng lượng rò rỉ sẽ tăng khi điện áp cung cấp giảm. Năng lượng động giảm là kết quả của điện áp cung cấp đã tối ưu hóa năng lượng. Điều này đã được trình bày trong phần 3b. Ở mức điện áp tối ưu năng lượng, năng lượng rò & năng lượng động gần như cân bằng. Việc giảm thêm điện áp cung cấp sẽ không tạo thêm hiệu quả năng lượng nữa. Lý do là vì sự mất cân đối tăng lên trong năng lượng rò rỉ. Điều này cho thấy điện áp tối ưu năng lượng độc lập với nhiệt độ xử lý & điện áp ngưỡng transistor. Bởi vì chúng tác động đối lập nhau giửa thời gian 1 chu kỳ & dòng rò rỉ, do đó chúng triệt tiêu nhau.

* + - * Các phân tích ở trên cho thấy rằng thiết kế có giới hạn đến hiệu suất năng lượng của nó mà không phụ thuộc tần số mà nó hoạt động. Hiệu quả năng lượng tối đa đạt được bản thiết kế hoạt động ở điện áp tối ưu Vmin của nó. Khi một điện áp thấp hơn Vmin cho hiệu quả năng lượng tốt hơn thì xác định những yếu tố đã ảnh hướng đến Vmin và đánh đổi Vmin nhưng vẫn đảm bảo các ràng buộc về hiệu suất. Một thiết kế với tỷ lệ năng lượng động/năng lượng rò rỉ cao sẽ có Vmin thấp hơn. Điều này được minh họa trong hình 4, mô phỏng các yếu tố hoạt động α khác nhau.

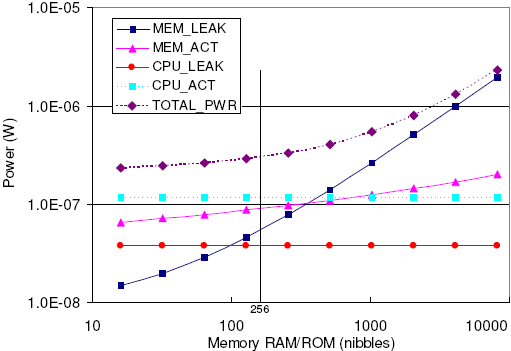


**Hình 4**: Những điểm hoạt động tối ưu năng lượng.

* + - * Trong hình 4, Vmin tăng khi yếu tố hoạt động giảm từ 1 xuống 0.2 transistor/chu kỳ. Bởi vì tỷ lệ dòng năng lượng động/rò rỉ là tỷ lệ thuận với yếu tố hoạt động
      * Tương tự, tỷ lệ của năng lượng động/rò rỉ tỷ lệ nghịch với thời gian chu kỳ xung clock, bởi vì năng lượng rò rỉ tăng tuyến tính với thời gian chu kỳ xung clock:
      * Có thể tính toán giá trị Vmin gần đúng bằng công thức sau:
      * Nhóm nghiên cứu rút ra kết luận là Cs cần phải là hằng hay giảm xuống, α phải được tăng lên nhằm mục đích tối ưu hóa năng lượng. Giá trị α cao tương ứng với hiệu dụng của transistor cao. Có nghĩa là một phần của mạch hay những cổng không hoạt động của mạch sẽ giảm. Xem xét 2 thiết kế với số lượng thiết bị bằng nhau, chúng bằng nhau về hiệu suất tính toán (chúng yêu cầu 1 số lượng giống hệt nhau các thiết bị chuyển mạch để hoàn tất 1 câu lệnh). Việc thiết kế với α cao hơn sẽ hiệu quả năng lượng hơn vì một số lý do. Thứ nhất, α cao hơn cho phép một Vmin và do đó năng lượng động ít hơn. Thứ 2, tại mọi thời điểm sẽ có ít thiết bị tạo ra sự rò rỉ, năng lượng rò rỉ giảm xuống. Cuối cùng, vì số lượng trung bình các switches/cycles cao, nên nó tốn ít thời gian để hoàn tất tính toán, do đó cũng giảm năng lượng rò rỉ trên mỗi câu lệnh.
      * CPI tỷ lệ nghịch với α. Tối ưu CPI trở nên rất quan trọng trong thiết kế vi xử lý vì nó không chỉ giảm rò rỉ bằng cách loại bỏ các thiết bị nhàn rỗi nhưng lại tác động năng lượng động thông qua việc giảm Vmin.
      * Tóm lại, việc tối ưu xử lý dưới ngưỡng điện áp là rất phức tạp, bởi vì nó phụ thuộc rất lớn đến 4 nhân tố: CPI, Cs, α, tclk. Hơn nữa, phụ thuộc Cs, α, tclk vào hiện thực vật lý làm nó trở nên khó để xác định hiệu quả năng lượng dưới ngưỡng nếu không nghiên cứu chi tiết hiện thực của một bản thiết kế. Nghiên cứu những thiết kế subthreshold hiệu quả năng lượng do đó phải bao gồm sự so sánh chi tiết của những hiện thực vật lý.
  1. **Phân tích sự đánh đổi trong kiến trúc**:

Để đưa ra kiến trúc tốt, nhóm nghiên cứu có thể sẽ đánh đổi một tính chất này để có được tính chất khác nhằm đạt được hiệu quả năng lượng. Cụ thể là họ xác định kiến trúc tập lệnh (ISA) và những đặc tính vi kiến trúc làm việc một cách tốt nhất để giảm nặng lượng tiêu thụ ở những điện áp dưới ngưỡng. Trước hết, họ kiểm tra sự cân bằng giữa tính diển đạt của tập lệnh (điều này làm cho chiều dài của mã nhỏ gọn) và độ phức tạp điều khiển logic (được giảm với các lệnh đơn giản). Ngoài ra, họ sẽ kiểm tra 21 mẫu thiết kế bộ xử lý mạng cảm biến, chúng đều được hiên thực bằng công nghệ 0.13μm của IBM.

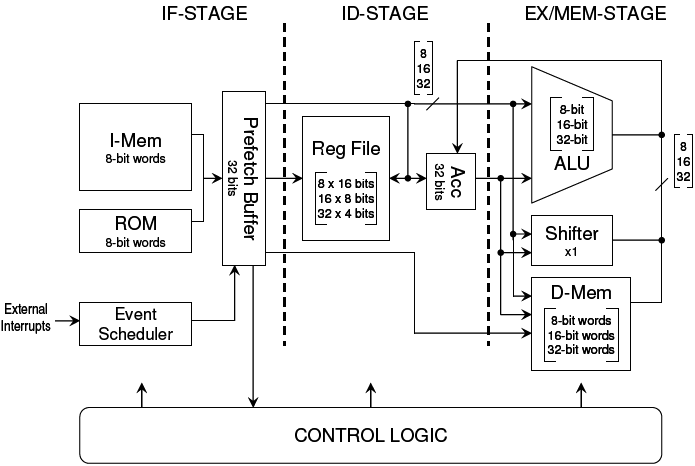
* + 1. **Nền tảng thực nghiệm**:
       - Nhóm nghiên cứu dùng các kỹ thuật, công nghệ để hổ trợ họ trong lúc thiết kế & đánh giá các bộ xử lý mẫu mà họ đã phát triển. Họ dùng các công nghệ này vào việc thiết kế các bộ xử lý & bộ nhớ. Sau đó, họ mô phòng, dùng PrimPower để tính năng lượng hoạt động & năng lượng rò rỉ của bộ xử lý; dùng PrimPower xác định sức mạnh của bộ giải mã & MUX; dùng Spice xác định năng lượng động & rò rỉ tổng thể của bộ nhớ.
       - Tổng năng lượng rò rỉ & hoạt động (trên mỗi chu kỳ của tất cả CPU & những thiết kế bộ nhớ) được tính dựa vào những đường cong SPICE ở các mức điện áp từ 100mV đến 600mV để xác định điểm điện áp tối ưu. Đây là điện áp làm thiết kế sử dụng hiệu quả năng lượng & ít tốn năng lượng cho mỗi chu kỳ nhất. Cuối cùng, tính số lượng năng lượng tiêu tốn cho mỗi câu lệnh, trung bình CPI được dùng (xác định khi ứng dụng được mô phỏng).
       - Chi tiết phần này được trình bày trong phần 4.1 của bài báo.
    2. **Tối ưu kiến trúc tập lệnh**:
       - Vì bộ nhớ & ROM dùng lưu trữ các câu lệnh & bộ điều khiển logic dùng để hiện thực các câu lệnh sẽ làm tốn năng lượng tĩnh & động. Do đó, kiến trúc tập lệnh là yếu tố rất quan trọng trong việc phát triển bộ xử lý mạng cảm biến. Kích thước bộ nhớ & kích thước điều khiển luận lý là yếu tố tác động đến thiết kế tập lệnh. Với 1 tập lệnh đơn giản, kích thước code sẽ tăng trong khi kích thước điều khiển vẫn nhỏ. Ngược lại, với một tập lệnh phức tạp, để kích thước code giảm phải trả giá bằng điều khiển logic phức tạp.



**Hình 5**: Đánh đổi giữa năng lượng cho việc xử lý & cho bộ nhớ.

(năng lượng cần thiết cho bộ xử lý & các thành phần bộ nhớ có kích thước khác nhau)

* + - * Tính chất tối thiểu hóa bộ nhớ & ROM rất quan trọng, nó được minh họa trong hình 5. Đồ thị cho thấy năng lượng rò rỉ (LEAK) & năng lượng hoạt động (ACT) trung bình của các thành phần; các thành phần này được kết hợp từ các bộ nhớ khác nhau với bộ xử lý mạng cảm biến hiệu quả năng lượng nhất. Những kiến trúc bộ nhớ được tạo thành từ 1/2 RAM & 1/2 ROM.
      * Vì lý do trên, nhu cầu bộ nhớ cần phải giảm & chú ý đến giảm thiểu rò rỉ năng lượng ở các ô bộ nhớ. Nhóm nghiên cứu tìm cách mã hóa tập lệnh dày đặt (a dense instruction set encoding). Cuối cùng, nhóm nghiên cứu đưa ra tập lệnh dùng cho bộ xử lý mạng cảm biến. Bảng 3 trong bài báo trình bày một danh sách rút gọn tập lệnh này.
      * Công việc tiếp theo mà nhóm nghiên cứu thực hiện là tiến hành các tính toán, thử nghiệm để có được ISA tối ưu. Các tối ư ISA này là sự đánh đổi giửa kích thước mã và độ phức tạp điều khiển. Phần này đã được trình bày trong phần cuối 4.2 của bài báo. Kết quả các thử nghiệm được thể hiện trong hình 6.
    1. **Phân tích kích thước thiết kế vi kiến trúc**: (Microarchitectual Design Space Analysis):



**Hình 7**: Tổng quan vi kiến trúc của bộ xử lý mạng cảm biến.

* + - * Hình 7 minh họa vi kiến trúc bộ xử lý mạng cảm biến. Hình này trình bày toàn bộ kiến trúc đã được nghiên cứu. Có nhiều biến thể khác nhau chỉ bao gồm một tập con các thuộc tính được trình bày trong hình này.
      * Bộ xử lý chứa 3 ống pipeline. IF-STAGE chứa bộ nhớ và ROM lệnh, và một prefletch buffer. Prefetch buffer là một buffer 32 bit chứa tối đa 4 lệnh. Nó làm đầy từ bộ nhớ lệnh mỗi khi bộ giải mã lệnh phát hiện nó không chứa một lệnh hoàn tất. ID-STAGE chứa tập thanh ghi (register file). Tập thanh ghi này (register file) có 4 phần tử 32-bit. Giá trị từ tập thanh ghi được gởi đến bộ tích lũy, là một thanh ghi 32 bit. Bộ tích lũy là nơi duy nhất kết quả lệnh được lưu trữ. Tùy chọn, một dường dữ liệu tồn tại giữa bộ tích lũy và tập tin thanh ghi (register file), đường dữ liệu này cho phép những giá trị của bộ tích lũy này được ghi trở lại (write back) vào tập thanh ghi. Trạng thái EX-STAGE chứa các đơn vị chức năng & bộ nhớ dữ liệu.
      * Những sự kiện bên ngoài (từ các bộ cảm biến) được lập lịch theo cơ chế sự kiện. Scheduler có 2 sự kiện đầu vào được đánh độ ưu tiên cao/thấp. Những sự kiện có độ ưu tiên thấp được xử lý theo thứ tự nó đến bộ xử lý mạng cảm biến. Những sự kiện có độ ưu tiên cao cũng được xử lý theo thứ tự nhưng chúng có quyền ưu tiên được xử lý trước các lệnh có độ ưu tiên thấp.
  1. **Phân tích kết quả thực nghiệm**:

Từ các kết quả thực nghiệm, nhóm nghiên cứu dùng biểu đồ Pareto (20-80) phân tích kết quả để xác thực những nghiên cứu và rút ra kinh nghiệm cho các thiết kế sau này. Trong hình 8 của bài báo, những thiết kế gần với đường cong sẽ nhanh hơn & hiệu quả năng lượng hơn các thiết kế ở xa. Những thiết kế này thỏa nghiên tắc tối ưu Pareto (20-80) tiêu biểu cho các thiết kế tốt nhất đã được phát triển. Chúng có sự đánh đổi giữa năng lượng & hiệu suất khác nhau và đạt được sự cân bằng ở những mức điện áp dưới ngưỡng để đạt được cùng lúc 3 tính chất:

* + 1. Hiệu quả CPI
    2. Tiết kiệm diện tích
    3. Hiệu dụng transistor cao nhất

1. **Kết quả đạt được & công việc trong tương lai**:

Trong bài báo này, nhóm nghiên cứu vấn đề của các bộ xử lý cảm biến tối ưu hóa năng lượng. Các bộ xử lý mạng cảm biến bị hạn chế năng lượng do có kích thước nhỏ. Do các ứng dụng mạng cảm biến có nhu cầu hiệu suất rất thấp cho nên có thể tối ưu năng lượng bằng cách cho các bộ xử lý hoạt động ở điện áp dưới ngưỡng. Thực nghiệm cho thấy để thiết kế bộ xử lý này thì phải tập trung vào việc cân bằng 3 yếu tố:

* + 1. Giảm diện tích tổng thể
    2. Tăng hiệu quả các transistor
    3. Duy trì hiệu quả CPI.

Ngoài ra, họ cũng xem xét sự đánh đổi giữa tập lệnh phức tạp (dẫn đến kích thước mã nhỏ gọn) và độ phức tạp điểu khiển luận lý (giảm khi tập lệnh đơn giản). Kết quả cho thấy giảm kích thước code luôn luôn làm tăng kích thước khối điều khiển luận lý, ngay cả với những chương trình đơn giản. Do đó, những thiết kế có ISA nhỏ gọn lại rất thích hợp với các nhu cầu giảm bộ nhớ.

Vẫn còn cơ hội để tìm ra các bản thiết kế hiệu quả năng lượng tốt hơn. Trong tương lại, nhóm nghiên cứu sẽ tiếp tục giảm Vdd thêm nữa, tìm cách tối ưu vi kiến trúc hơn nữa để cải thiện CPI, giảm diện tích & hoạt động. Nếu chỉ tập trung vào các ứng dụng cảm biến băng thông thấp, có nhiều khả năng để khám phá thêm các thiết kế với các điểm hoạt động điện áp nhỏ hơn điện áp Vdd tối tiểu. Tuy nhiên, để tìm được các mẫu thiết kế này, chúng chắc chắn phải được thiết kế để có hiệu dụng transistor cực kỳ cao, diện tích nhỏ và CPIs hiệu quả.