Energy-Efficient Histogram Equalization on FPGA

Andrea Sanny, Yi-Hua Yang and Viktor K. Prasanna

Ming Hsieh Department of Electrical Engineering

University of Southern California

Los Angeles, CA 90089

Outline

- Histogram Equalization
- Throughput Optimizations
- Memory Optimizations
- Off-Chip Memory
- Architecture Details
- Experimental Results
- Conclusion

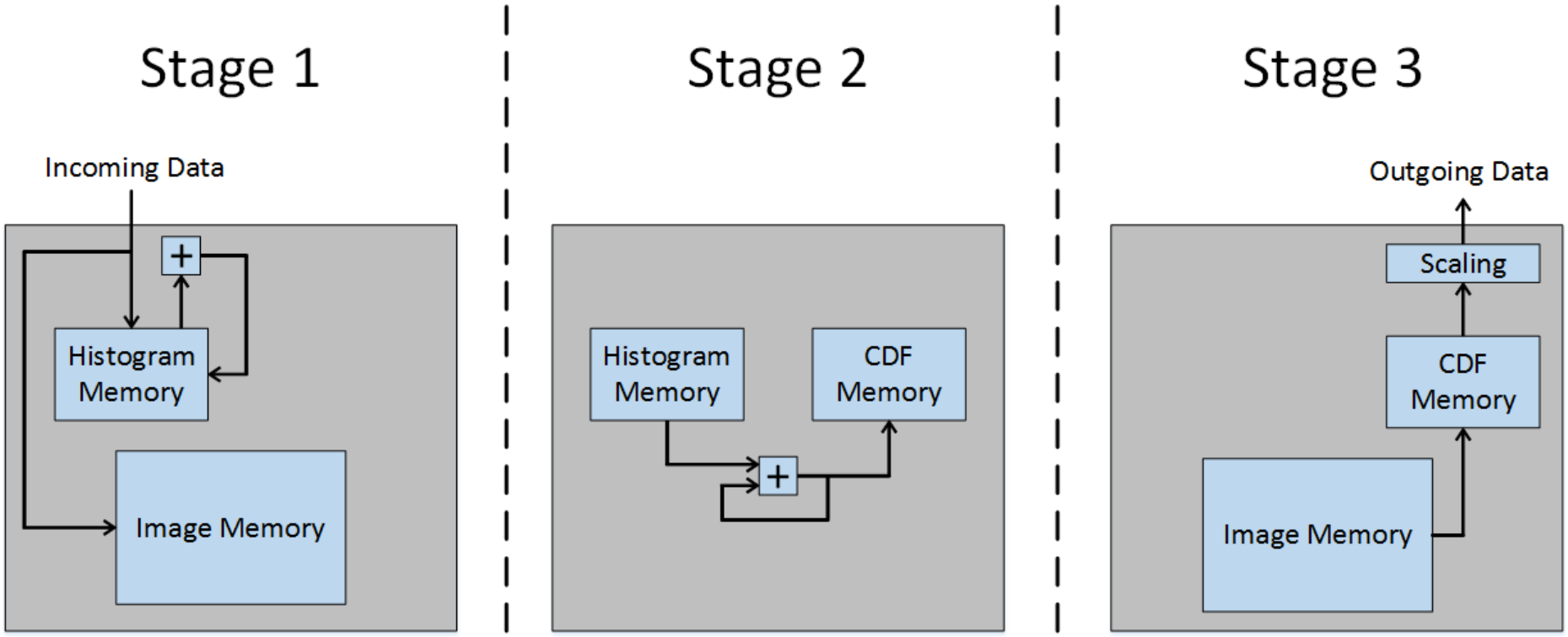
Histogram Equalization

- Common kernel for image enhancement
- Focus:
  - High throughput
  - Low area
  - Low cost
- Energy efficiency

RAW Image

After Histogram Equalization

Histogram Equalization Processing



Histogram Equalization Stages

1. PDF/Histogram creation
   – Main operation: Addition

2. CDF creation
   – Main operation: Addition

3. Pixel scaling
   – Main operation: Addition, division

Optimizations Overview

- Architecture and circuit-level optimizations
  - **Metric**: Frames per second (throughput)
  - Pipelining and data forwarding

- Memory access optimization
  - **Metric**: GOPS/W (energy efficiency)
  - Memory activation scheduling
  - DRAM power-down mode

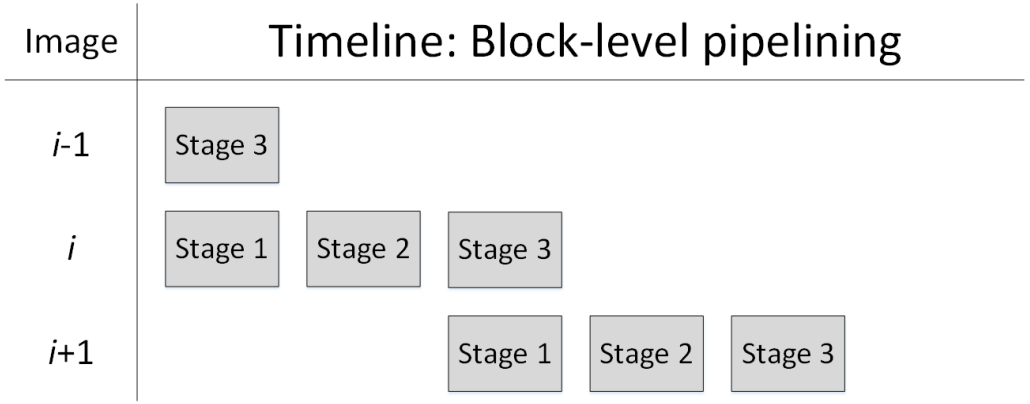
Pipelining and Data Forwarding

- **Goal**: Optimizing for higher throughput

- Issues:
  - Non-overlapping stages
  - Multi-cycle per pixel
  - Data hazards

Pipelining and Data Forwarding

- Block-level pipelining (Stage 1 and 3)
- Circuit-level pipelining
  - Within a stage, between components
- Data forwarding unit
  - Avoid RAW hazards



Pipelining and Data Forwarding

- Results:
  - One pixel processed each cycle
  - If $M \times N$ image, L histogram bins:
    - Completion time for one image without pipelining

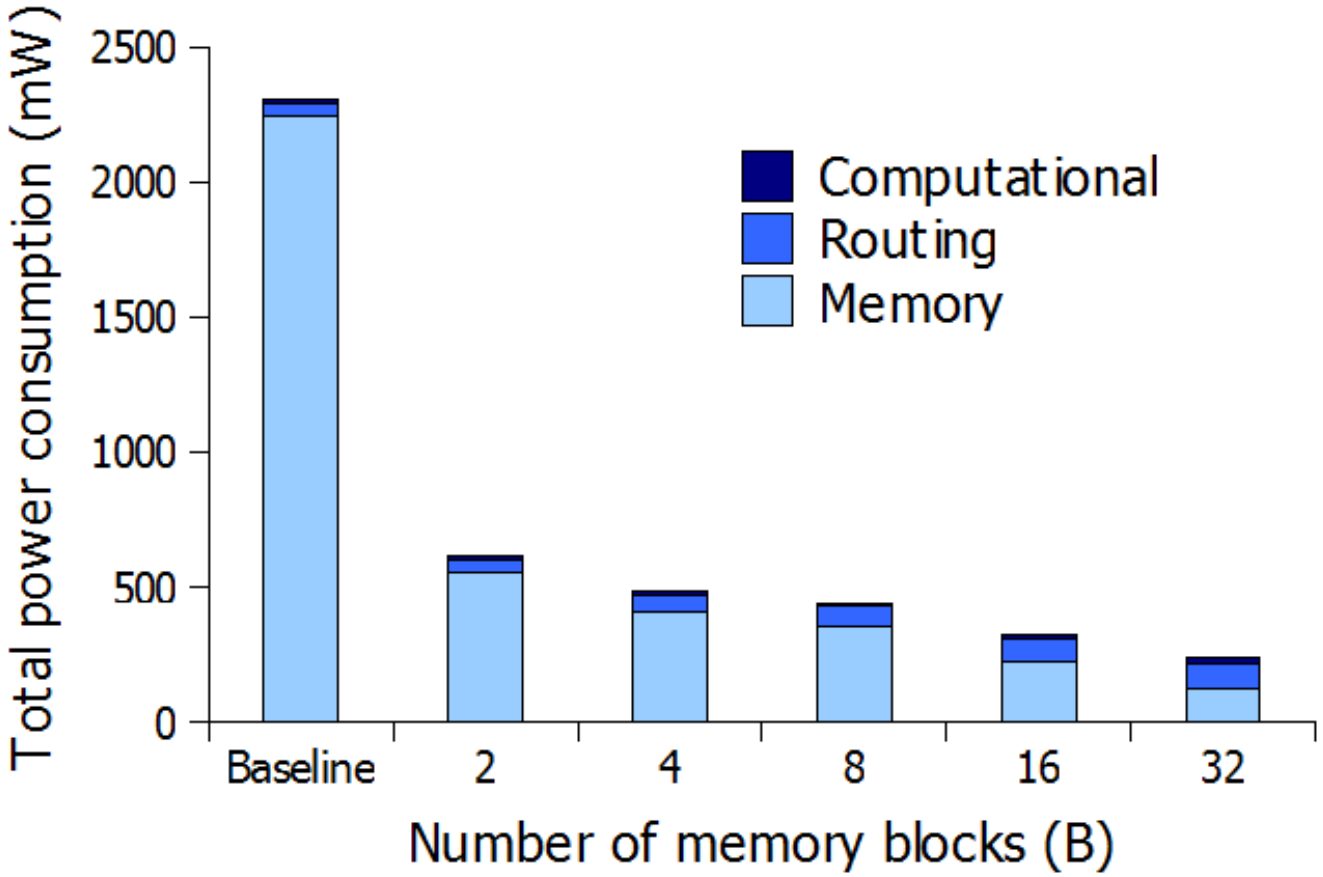
$$2MN + L$$

    - Completion time for one image with pipelining

$$MN + L$$

- For large image sizes, $MN >> L$
  - Almost 2× cycle speedup

Power Profile



- Memory power is dominant

Memory Activation Scheduling

- **Goal**: Optimizing for lower power consumption

- Baseline:
  - All used on-chip memory is active
  - Memory regarded as 1 block
  - No enable control

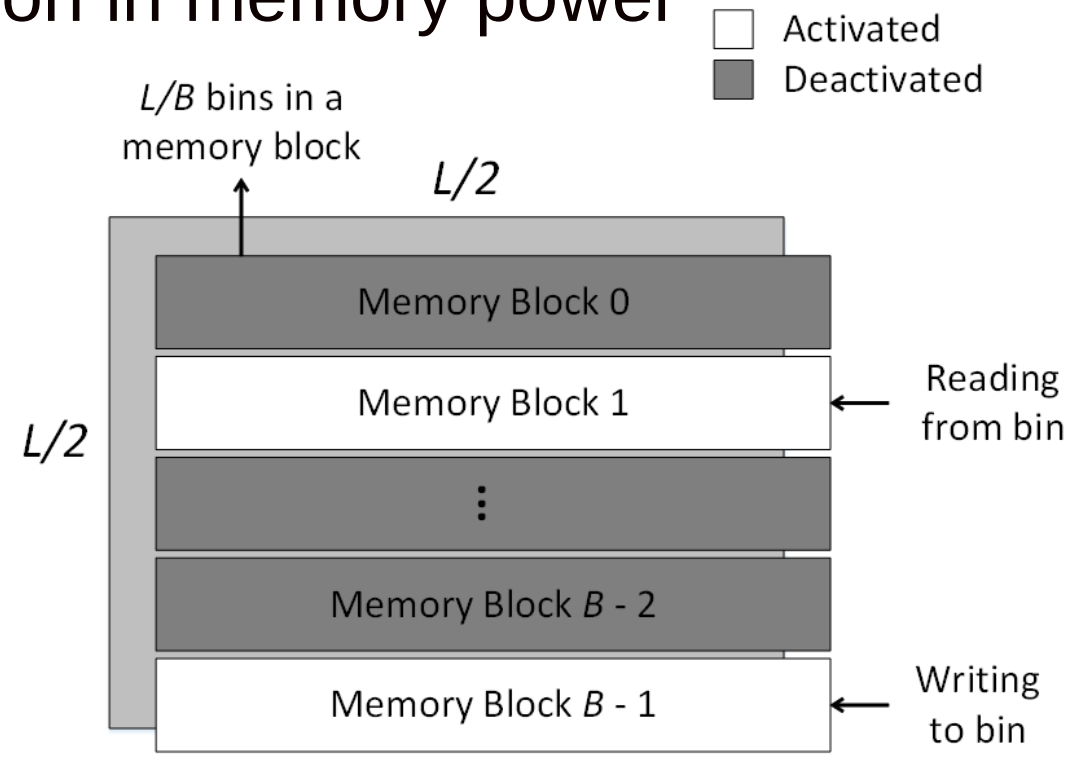
Memory Activation Scheduling

- Assumptions:
  - There are a total of *B* memory blocks with enable control on each block
  - *M*×*N* image size, *L* histogram bins

- Memory activation scheduling
  - Scheduler determines block enabling
  - Disable unused memory blocks

Memory Activation Scheduling

- Results:
  - Significant reduction in memory power
  - Maintain same throughput
- Some image sizes cannot fit fully on-chip



DRAM Power-Down Mode

- **Goal**: Optimizing for lower power consumption
- Store large images into off-chip DDR3 memory
  - *b* buffers required between DRAM and processor
- Baseline:
  - Image fully stored on DRAM → Read by processor → Written back to DRAM before outputted
  - **Result**: 2 DRAMs required to satisfy BW requirement

DRAM Power-Down Mode

- Careful data layout minimizes row activations

- DRAM power-down mode
  - Mode with lowest power consumption
  - Cannot read or write in this mode
  - Power-down mode can be used up to 75% of total run time

DRAM Power-Down Mode

- Results:
  - DRAM used as temporary image storage
    - **Result**: 1 DRAM required to satisfy BW
  - Power-down mode can be used up to 75% of total run time
  - Up to 2× reduction of average DRAM power consumption

Experimental Setup

- FPGA Virtex 7 XC7VX980T with -2L speed grade

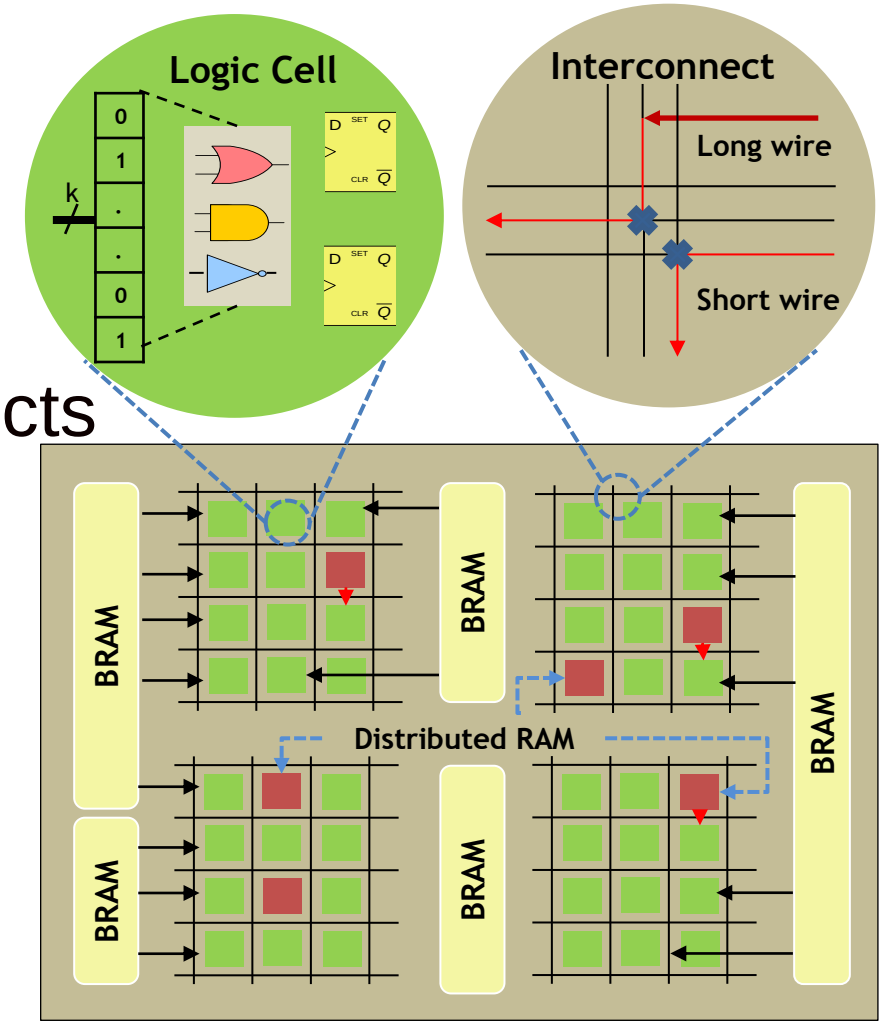
- Tools:
  - Vivado 2013.4
  - Vivado Power Analysis Tool
  - Micron DDR3 SDRAM System-Power Calculator

Experimental Setup

- Image details:
  - Image sizes from 240×128 to 3840×2160
  - 16 bits per pixel
- VCD file input
  - Input to Vivado Power Analysis Tool
- At least 30 fps

FPGA Platform

- Field Programmable Gate Arrays
  - Programmable logic cells
  - Programmable interconnects
  - Programmable on-chip storage
    - BRAM & LUT based Distributed RAM
    - Both width and depth



FPGA Platform

- Huge on-chip bandwidth
  - All BRAM and Distributed RAM can be accessed in parallel
    - 1k x 36 bits at 200 MHz = 7.2Tb/s

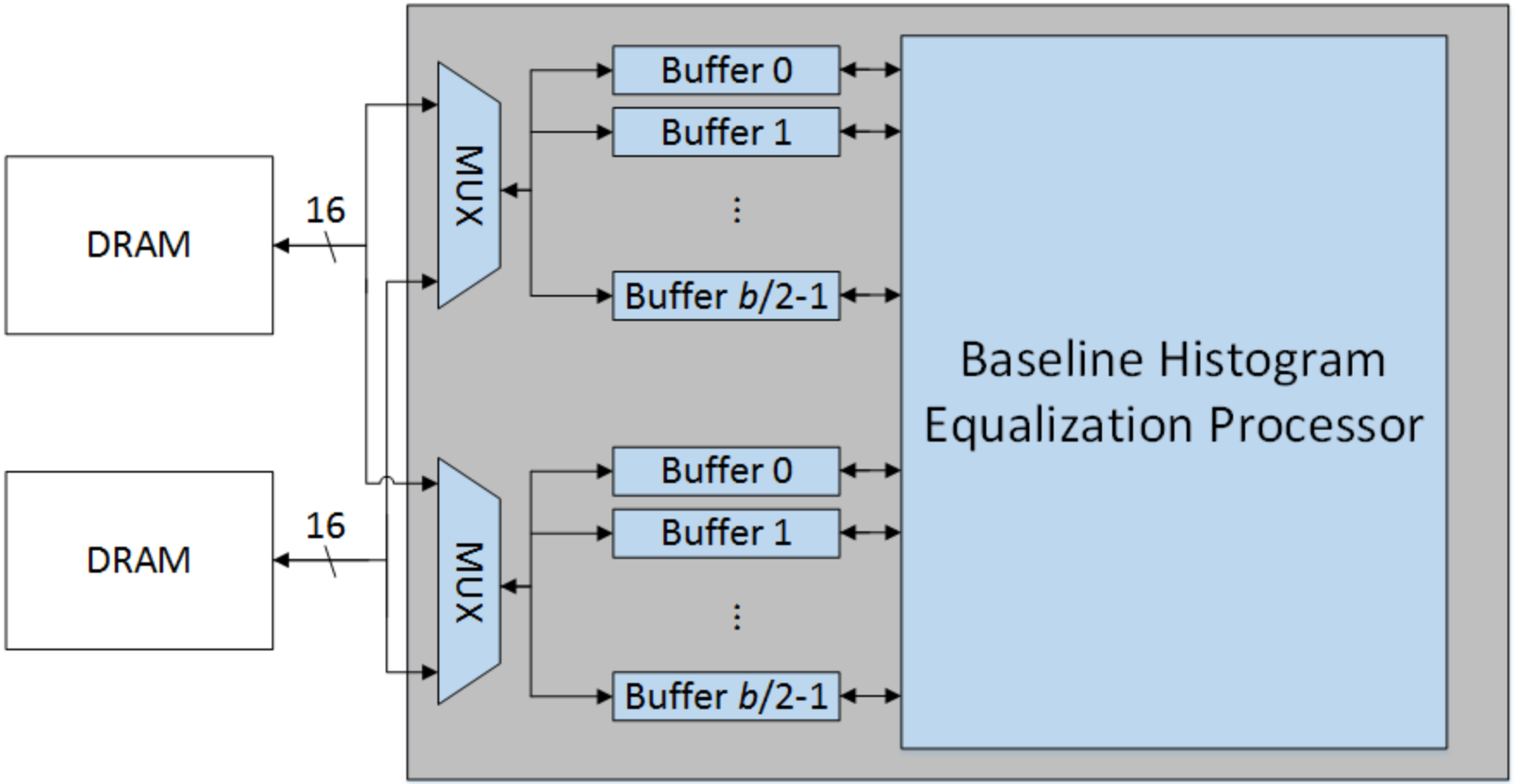
- Other Embedded features for specific functions
  - General purpose processors (PowerPC, ARM), Multipliers (DSP blocks)

Baseline Architecture

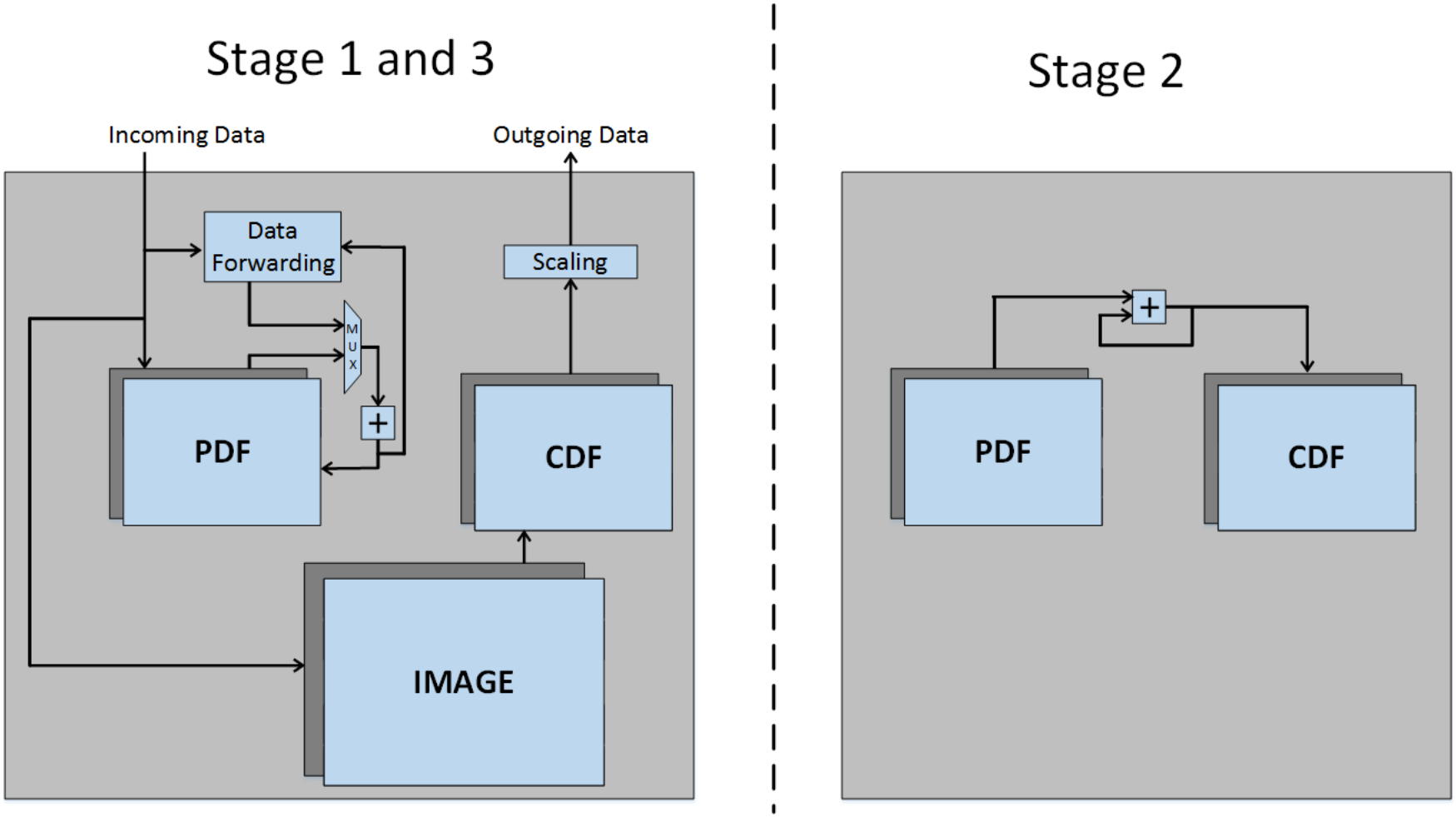
- Throughput optimizations
  - Block- and circuit-level pipelining, data forwarding

- No memory optimizations
  - On-chip memory is always active
  - For large images, 2 DRAMs required

Baseline Architecture



Histogram Equalization Processor

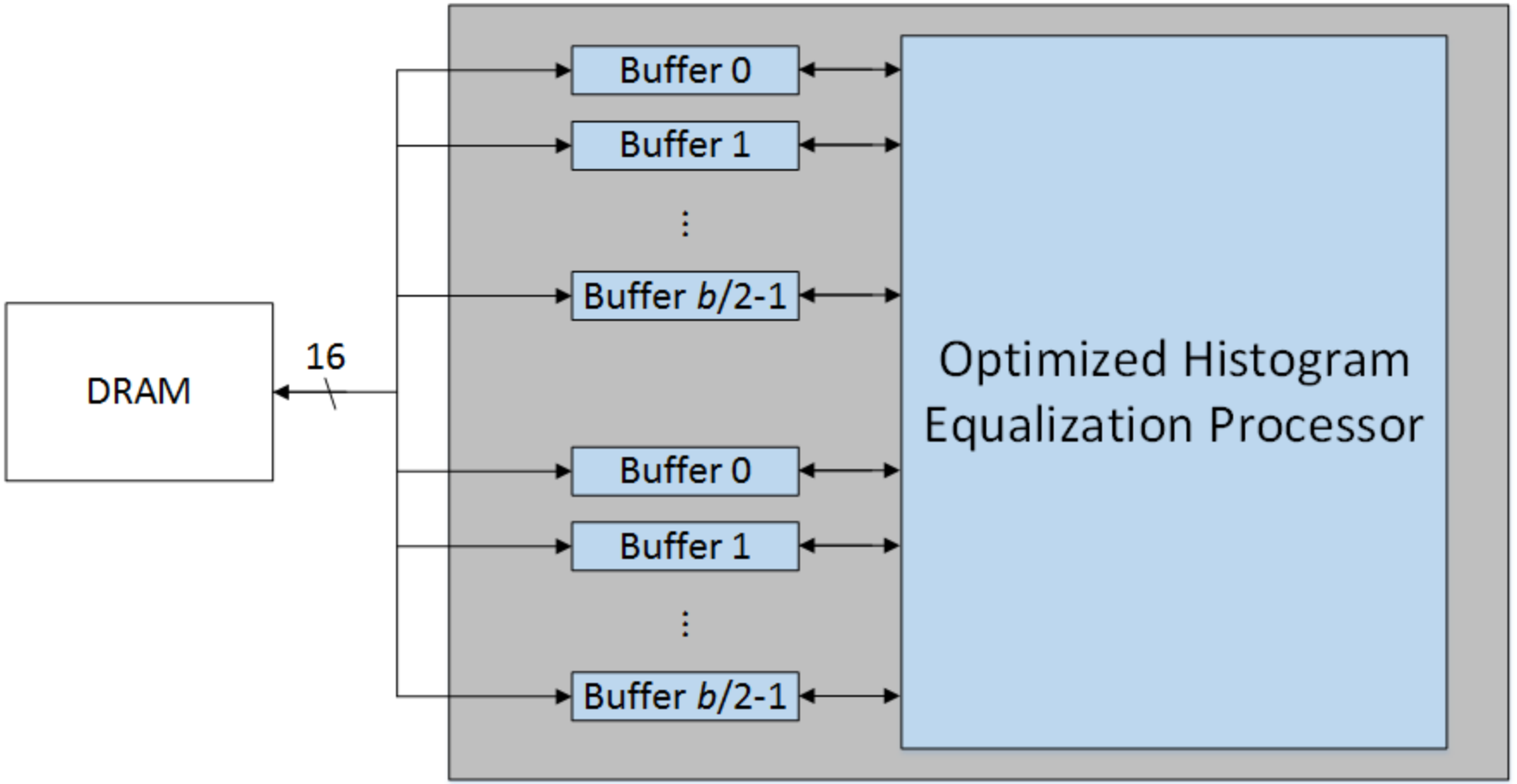


Optimized Architecture

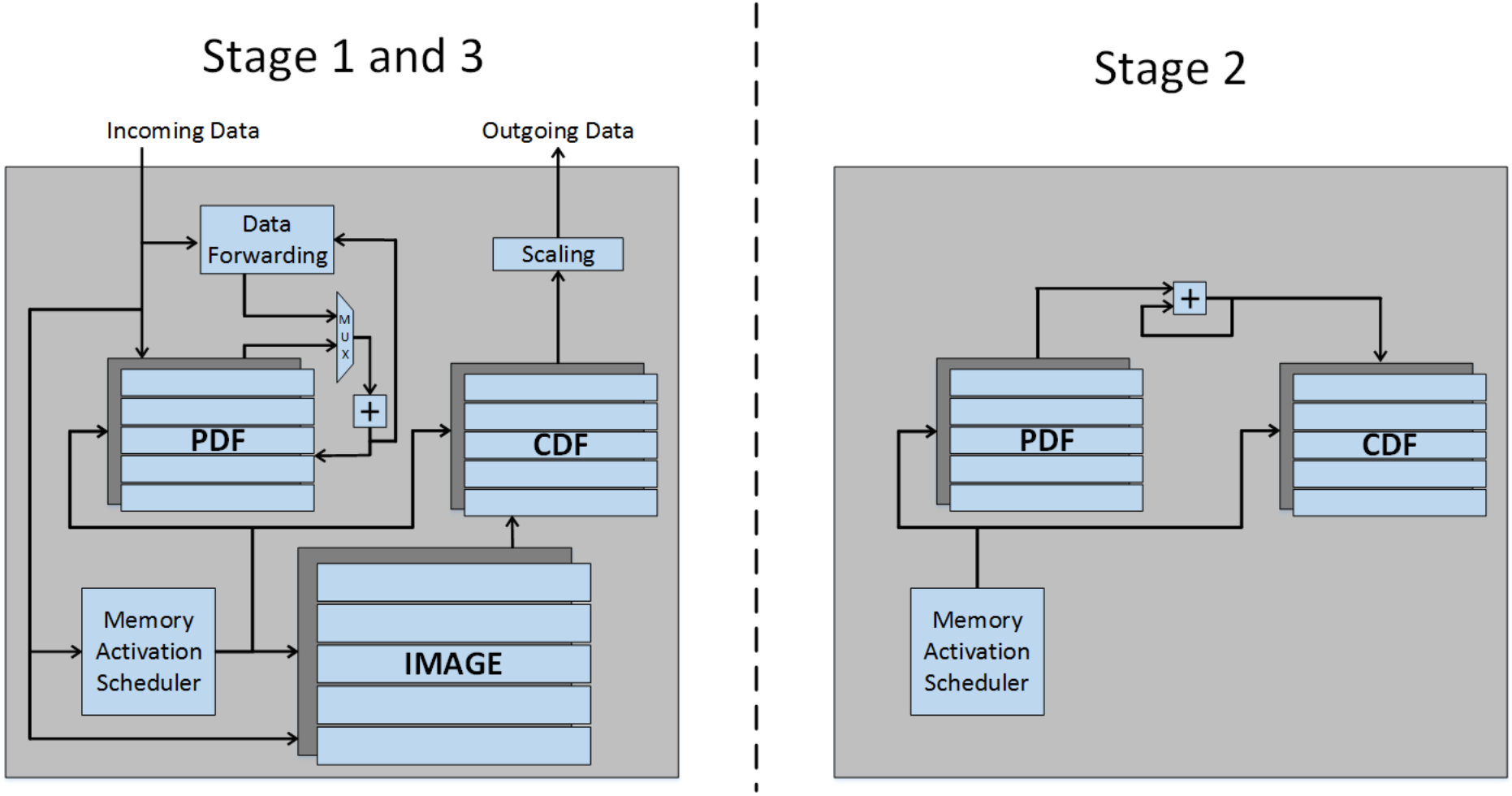
- Throughput optimizations
  - Block- and circuit-level pipelining, data forwarding

- Memory optimizations
  - Memory activation scheduling
  - For large images, 1 DRAM required
  - DRAM power-down mode for large images

Optimized Architecture



Histogram Equalization Processor

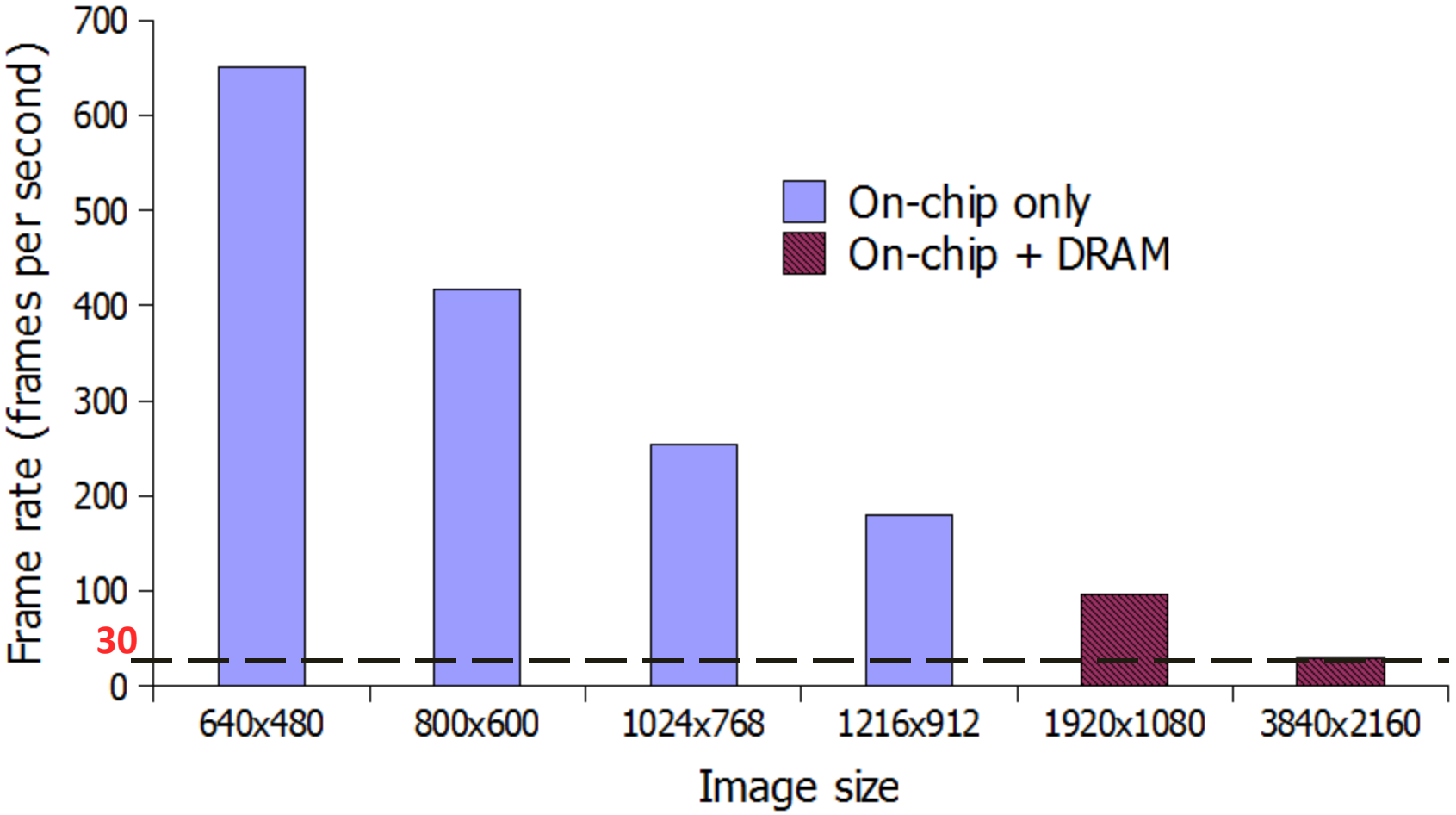


Energy Efficiency

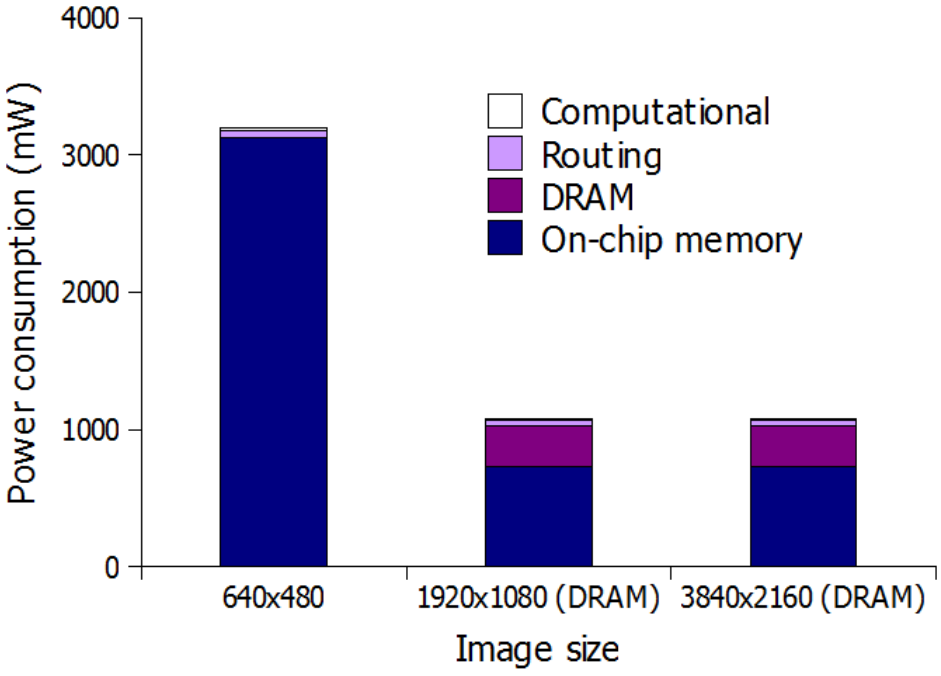
- Energy efficiency
  = # of operations/energy consumed by design
  = GigaOperations/sec/Watt (GOPS/W)

- Operations included for histogram equalization:
  - Required operations to do histogram equalization
  - Ignores extraneous operations specific to an implementation

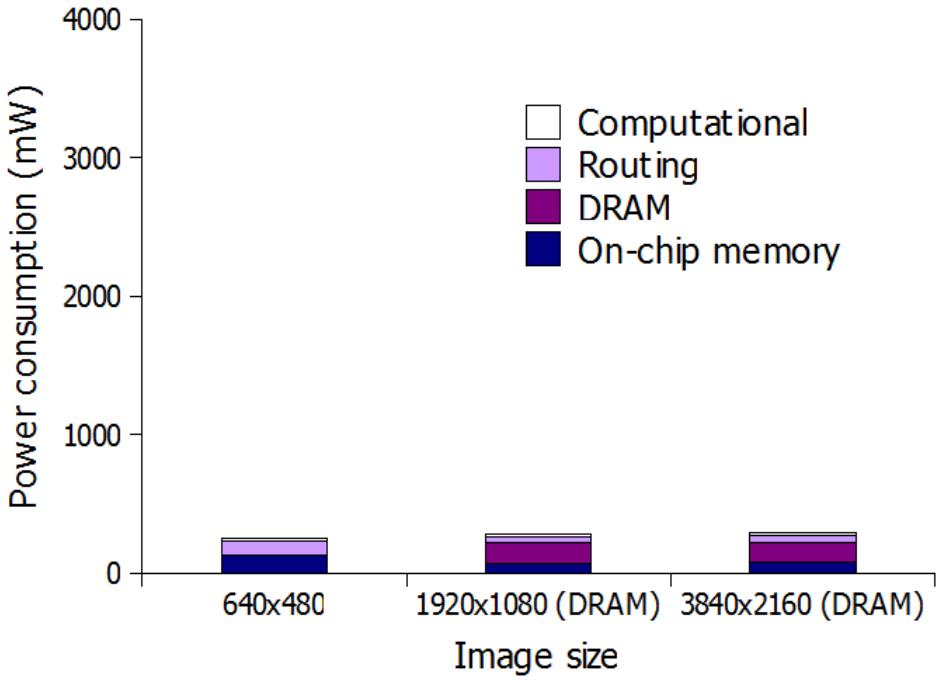
Experimental Results



Power Profile

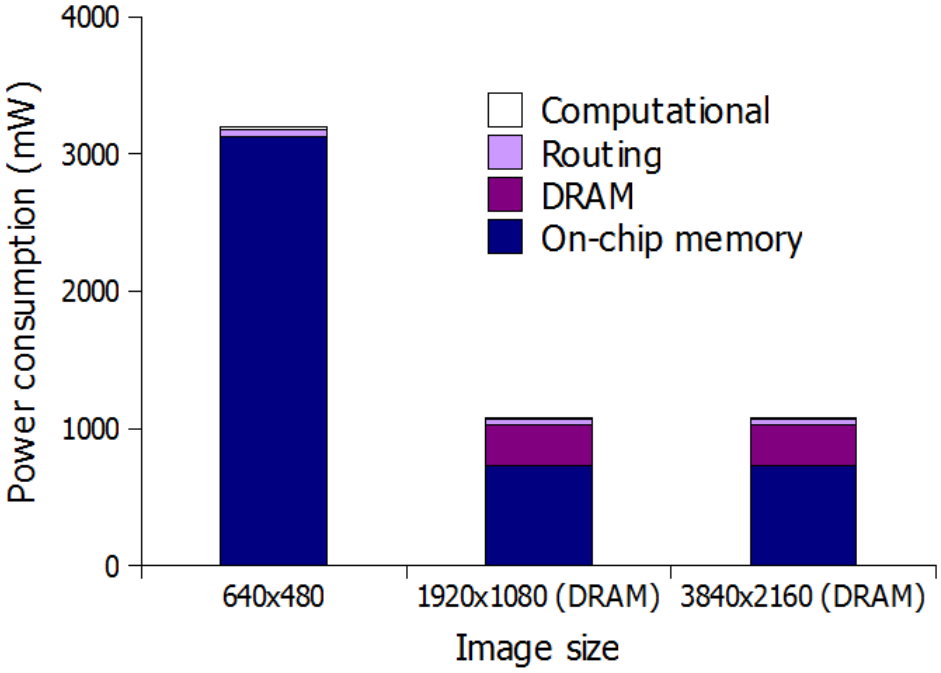


Baseline Results

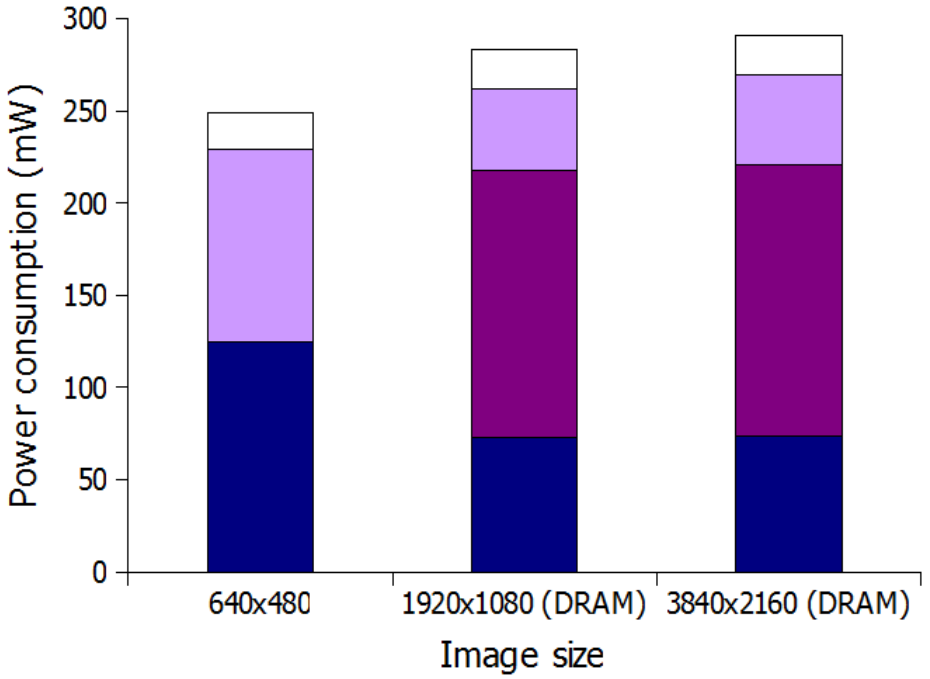


Optimized Results (B = 32)

Power Profile

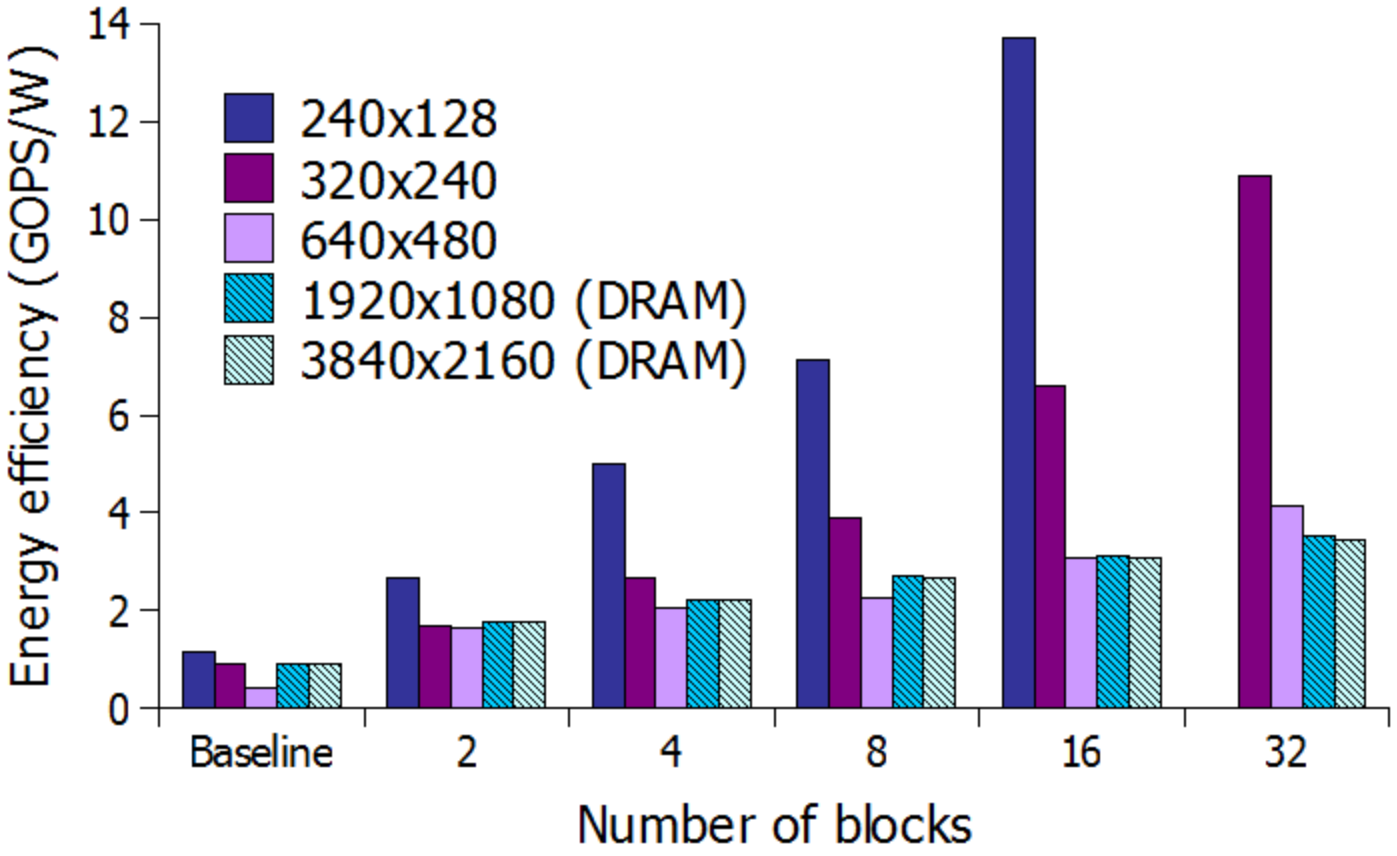


Baseline Results



Optimized Results (B = 32)

Experimental Results



- 3.72× to 12.8× energy efficiency improvement

Conclusion

- Our proposed optimizations achieved:
  - At least 30+ fps
  - Up to 12.8× higher energy efficiency than baseline
- Future work:
  - Develop general optimizations for classes of image processing kernels

Contact Information

- Andrea Sanny – sanny@usc.edu
- Yi-Hua Yang – edward.yang@xilinx.com
- Viktor K. Prasanna – prasanna@usc.edu

- For more information on our work, see:
  http://pgroup.usc.edu/wiki/Parallel_Computing