A Low-Power 0.7-V H.264 720p Video Decoder

D. Finchelstein, V. Sze,
M.E. Sinangil, Y. Koken,
A.P. Chandrakasan

A-SSCC 2008

Outline

- Motivation for low-power video decoders
- Low-power techniques
  - pipelining and parallelism
  - independent voltage/clock domains
  - efficient memory accessing
- ASIC results
- Comparison with state of the art
- Summary

Motivation

- High demand for video capture and playback on mobile devices
- H.264 state of the art video coding standard
- Goal: Ultra Low Power H.264 decoder in 65nm
  - 1280x720 @ 30fps

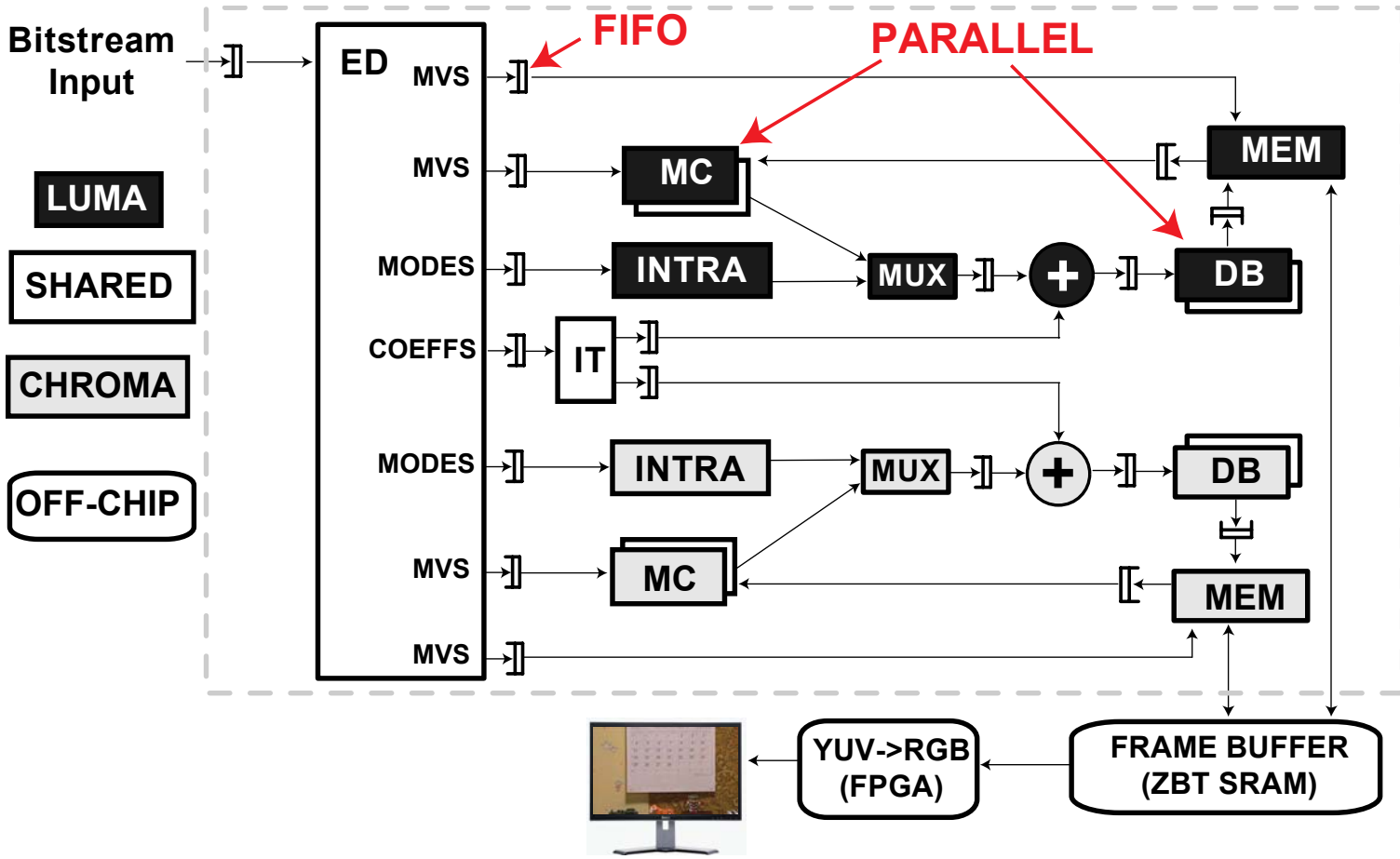
iPhone

DVC

Digital Camera

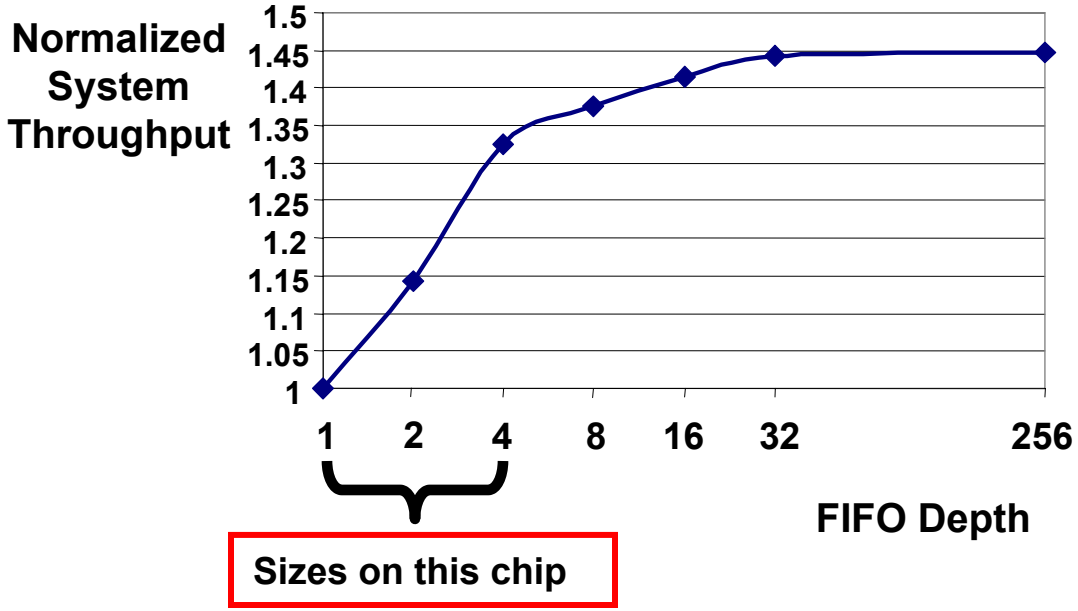
PSP

H.264 Decoder Architecture



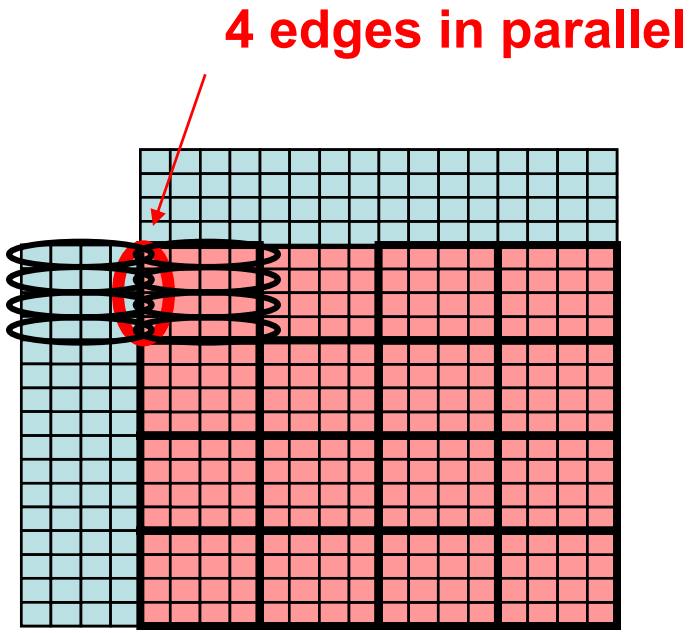
Pipelined, highly parallel architecture to reduce voltage (and power)

Pipeline FIFO Sizing



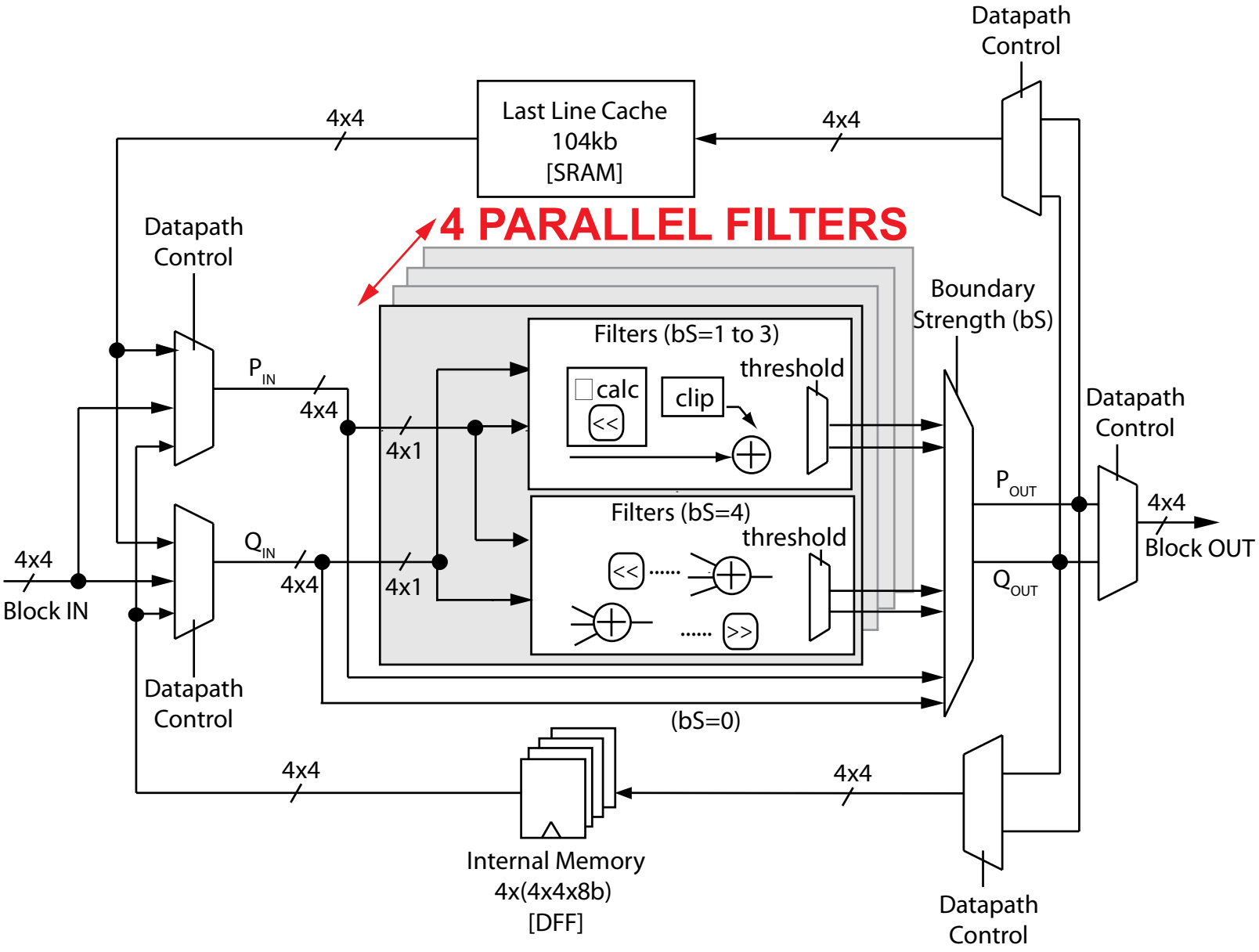
- Pipeline stages have variable latencies
  - ex: ED latency is 0-33 cycles per 4x4 block
- Larger FIFOs help average out workload
  - increase performance by up to 45%
  - FIFOs of depths 1-4 chosen to reduce area

Deblocking Filter Parallelism

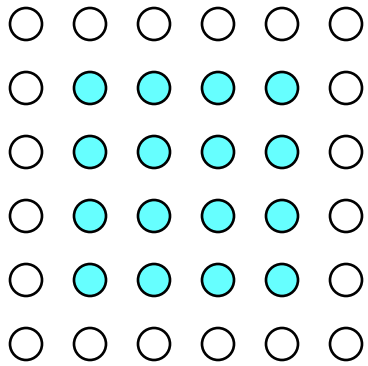


- Process entire 4x4 edge (4 filters) in parallel
- Filter luma and chroma in parallel
- 192 cycles reduced to ~ 46 cycles per 16x16 block

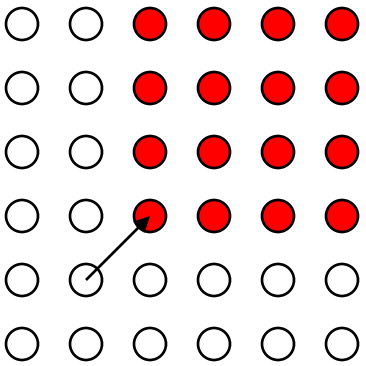
Deblocking Filter Architecture



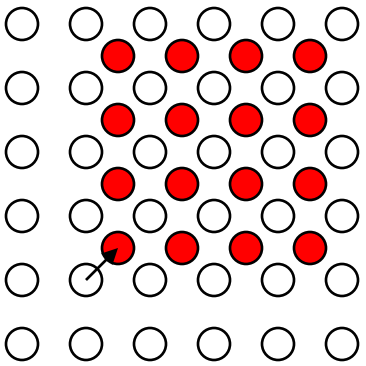
Motion Compensation (MC)



4x4 block in current frame



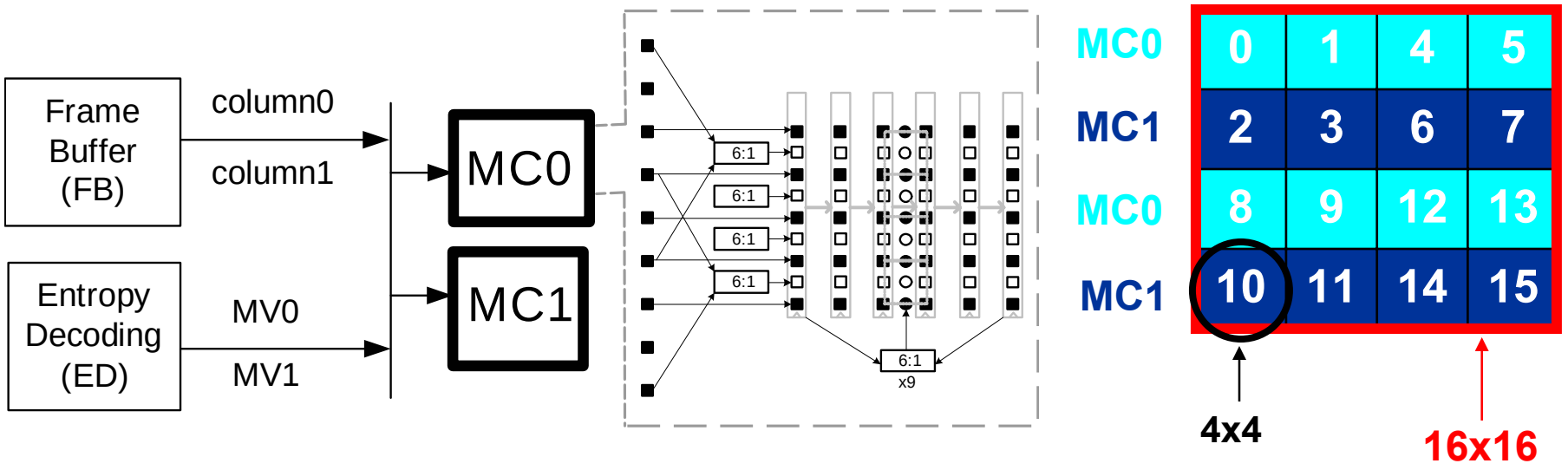
Reference block Vector (1, -1)



Reference block Vector (0.5, -0.5)

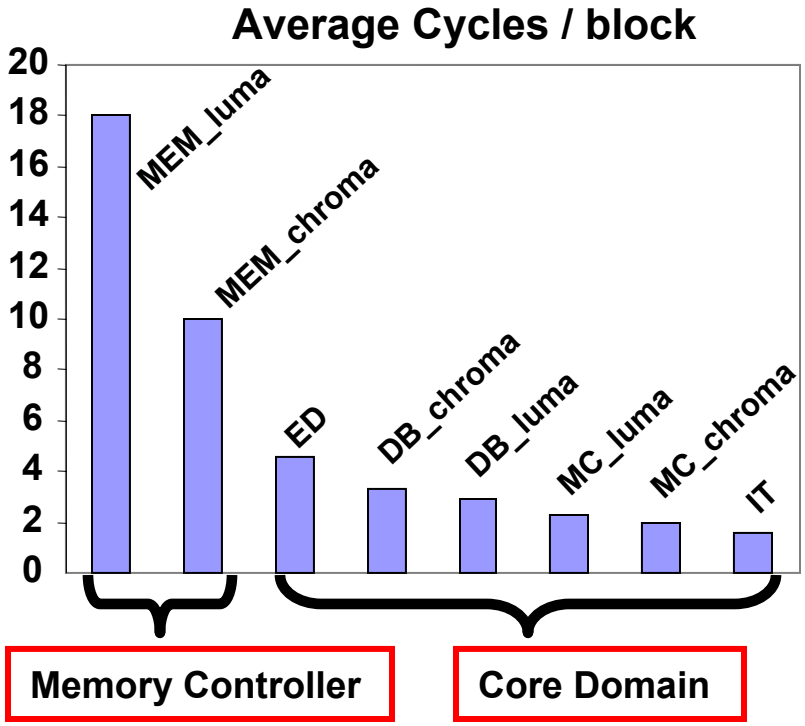
- Use two interpolators in parallel
- Interpolate luma and chroma in parallel
- 176 reduced to ~ 72 cycles per 16x16 block

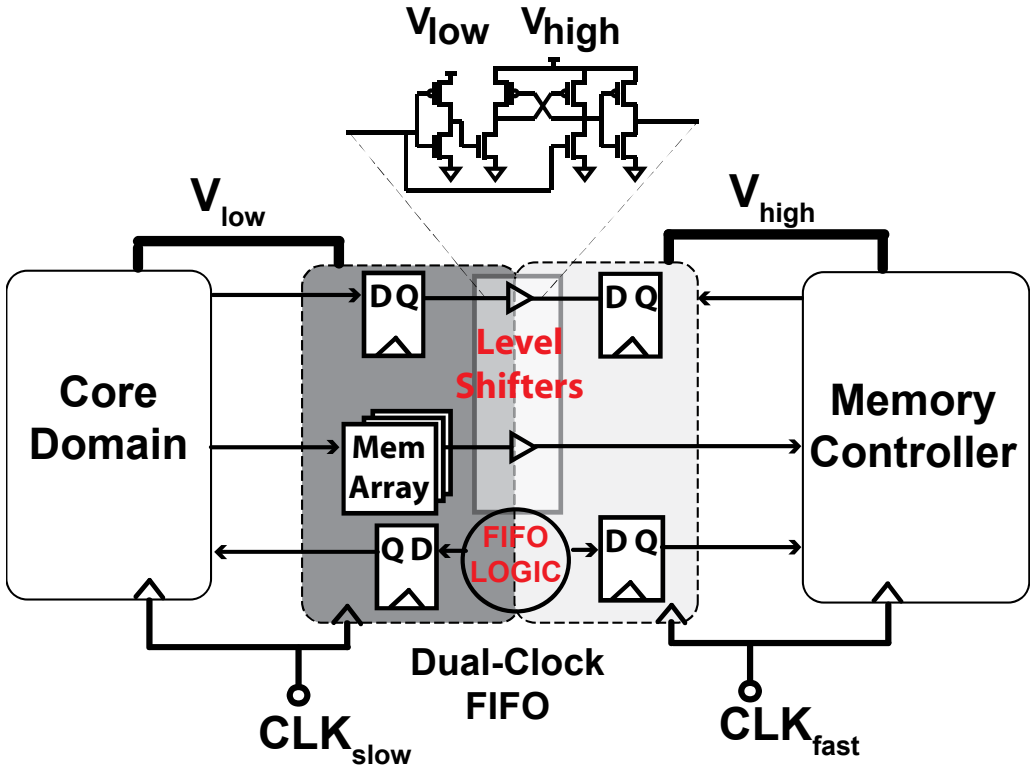
Parallel MC Interpolators



- Interpolators can run in same cycle when…
  - motion vectors are all available
  - memory interface supplies 2 columns per cycle
- Interpolators are synchronized
  - MC0: even 4x4 rows, MC1: odd 4x4 rows
  - shared interpolation data reused

Clock & Voltage Domains

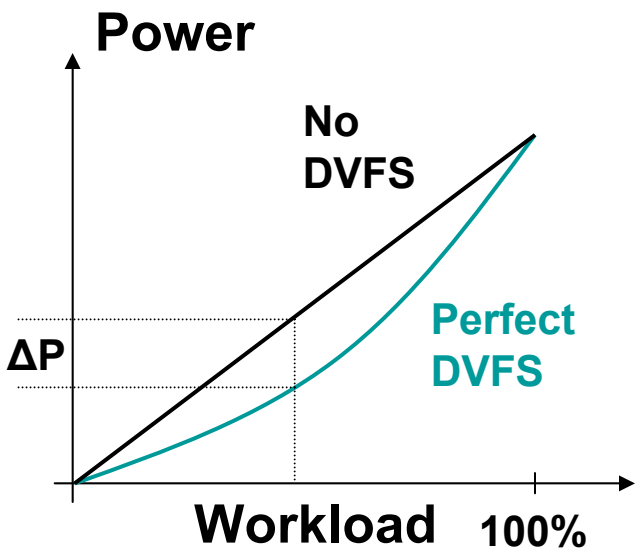




- Decouple voltage / clock domains
  - lower core voltage and frequency
  - 25% power savings vs. single domain
    - 4% further savings if we used 3 domains
  - dual-clock FIFOs and level-shifters link domains

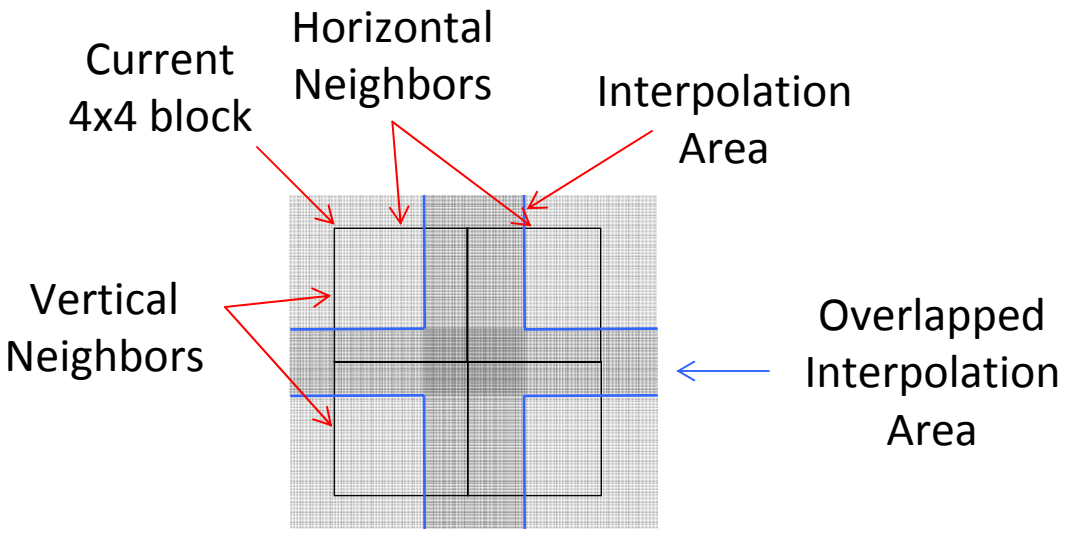
Workload Variation

	Core Domain [MHz @ V]	Memory Controller [MHz @ V]	Relative Power @720p 1 I-frame, 14 P-frames [%]
P-frame	14 @ 0.70	50 @ 0.84	
I-frame	53 @ 0.90	25 @ 0.76	
MAX	53 @ 0.90	50 @ 0.84	100
DVFS	14 @ 0.70 53 @ 0.90	25 @ 0.76 50 @ 0.84	73
FA	17 @ 0.72	48 @ 0.83	73



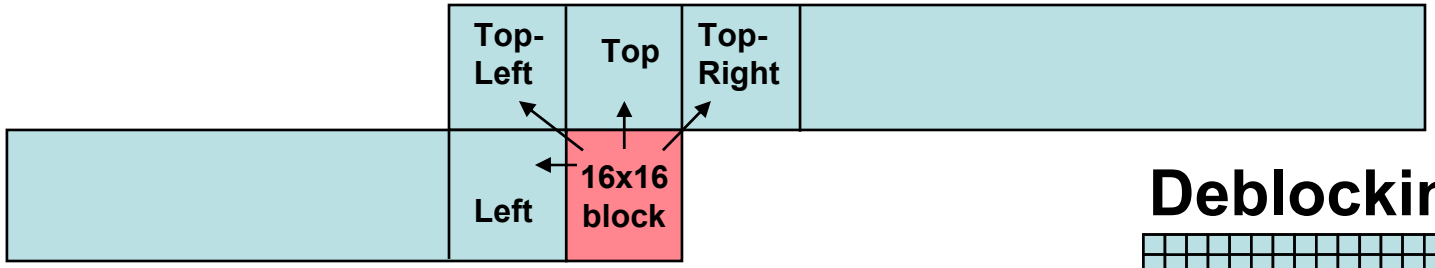
- INTER-INTRA workload variation
  - MAX: maximum frequency on each domain
  - DVFS: 1 frame every 33ms
  - Frame Averaging (FA): 15 frames every 15 * 33ms
    - switches less often than DVFS, but needs output buffer

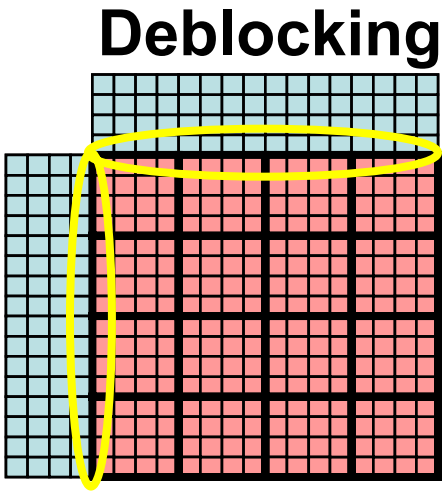
MC Data Overlap

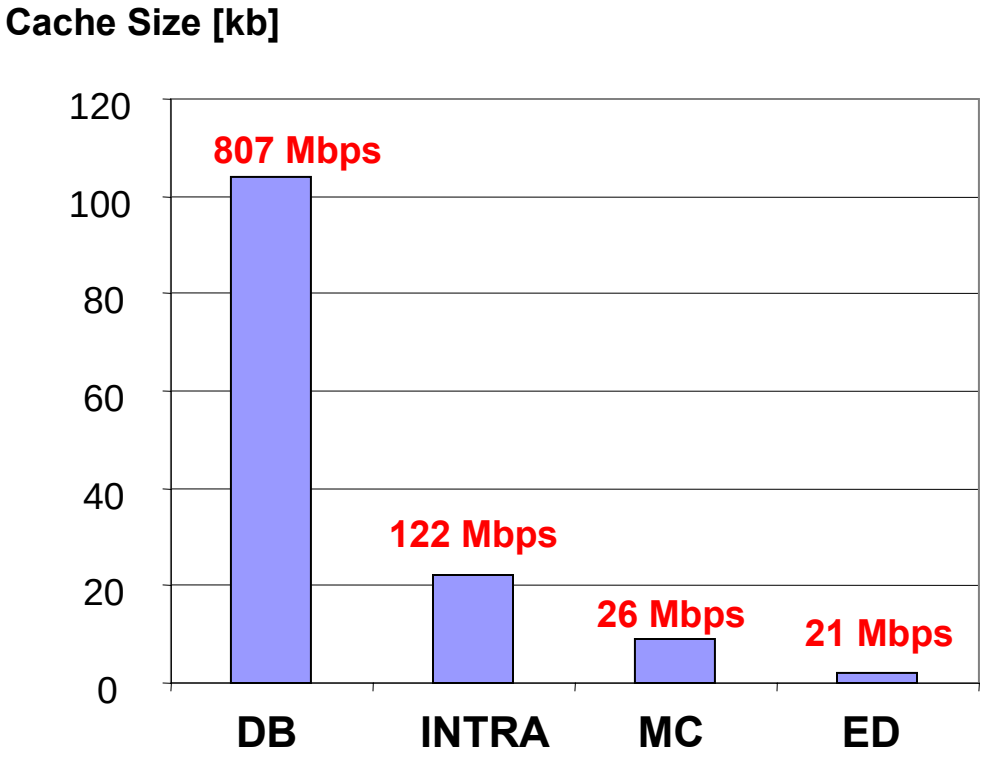


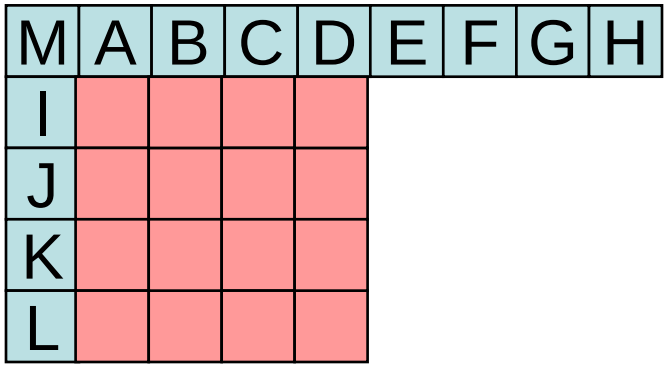
- Neighboring 4x4 with same MVs
- Overlap area shared horizontally and vertically
- Reduced MC read bandwidth

Last-line On-chip Caches



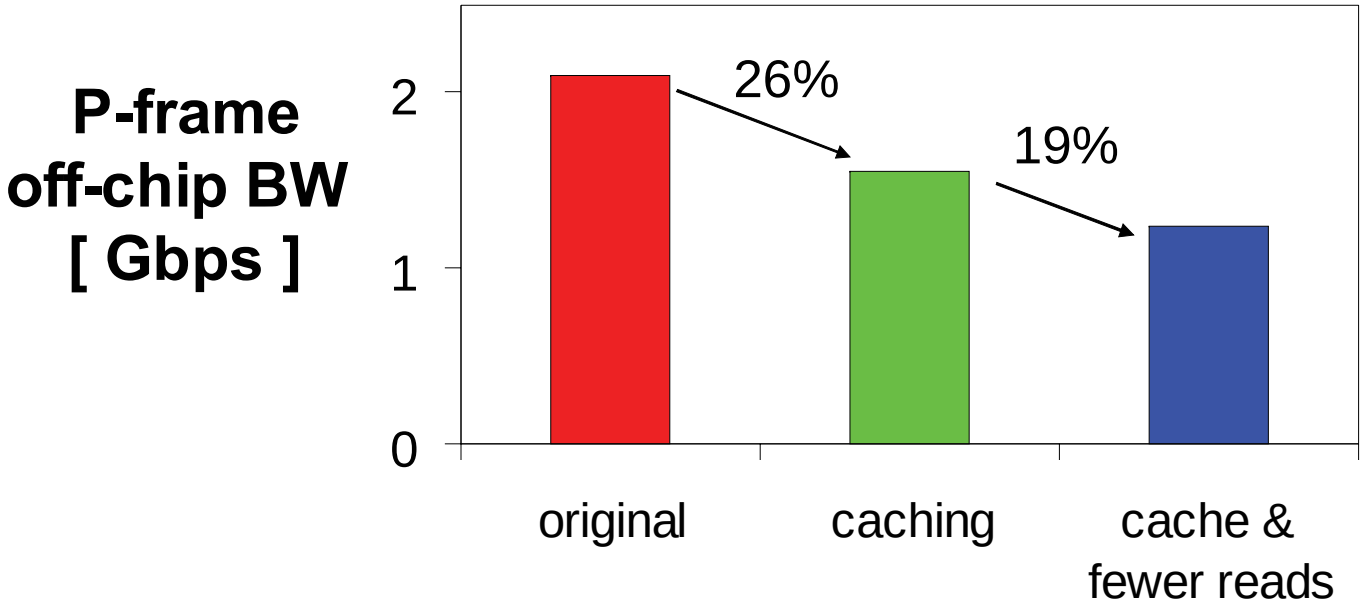






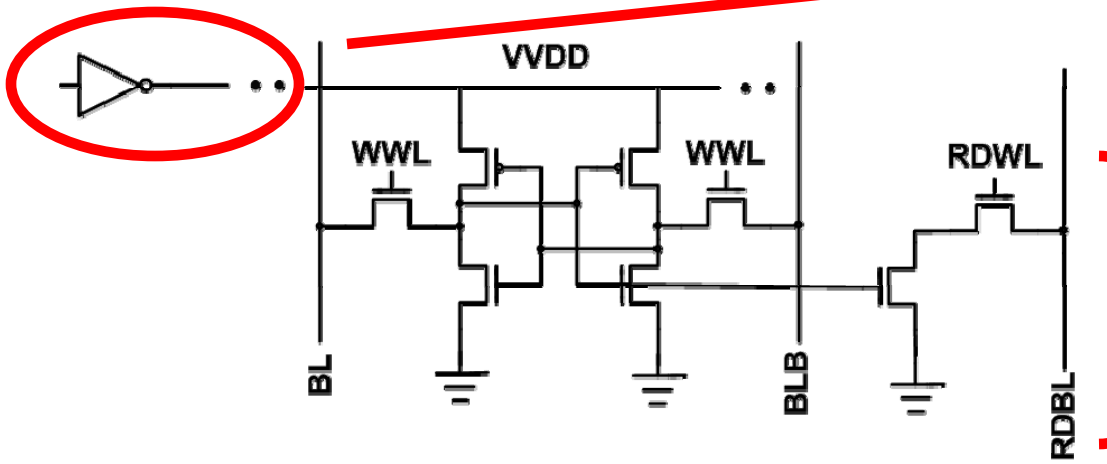
Intra prediction

Off-chip Bandwidth



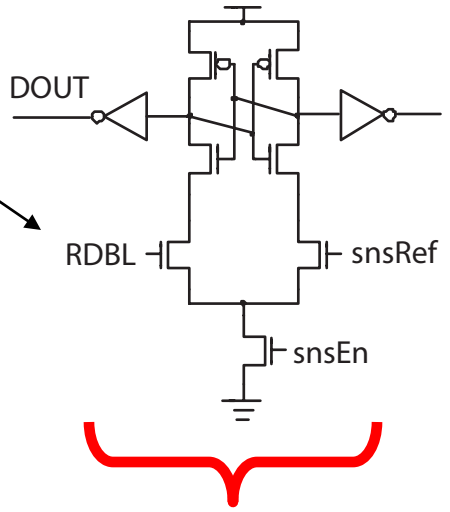
- Frame buffer off-chip (1.4 MByte per frame)
- P-frames more common than I-frames
  - P-frame off-chip BW larger due to MC
- 40% (0.9 Gbps) total reduction
  - last-line caches
  - avoided redundant reads in MC

Voltage Scalable SRAM



Write Assist to improve writability at low voltages

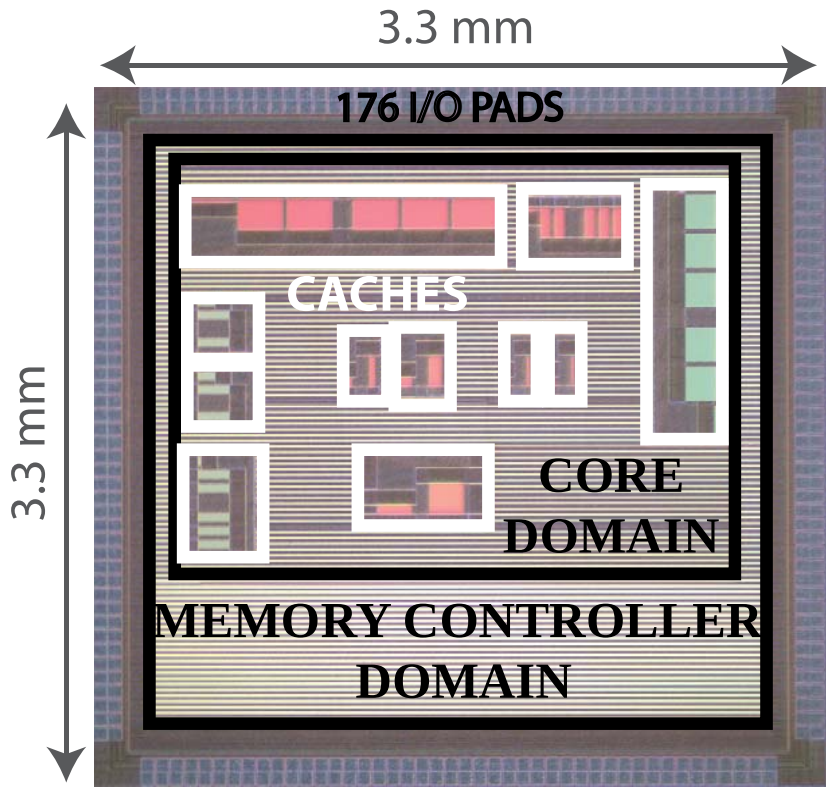
Extra 2 Tx ensures read stability at low voltages



Pseudo-differential sense amplifier with global snsRef

- Low voltage SRAM needed
  - Typical 6T SRAMs fail at low voltages
  - 8T SRAMs work down to **0.5V**

H.264 Decoder ASIC

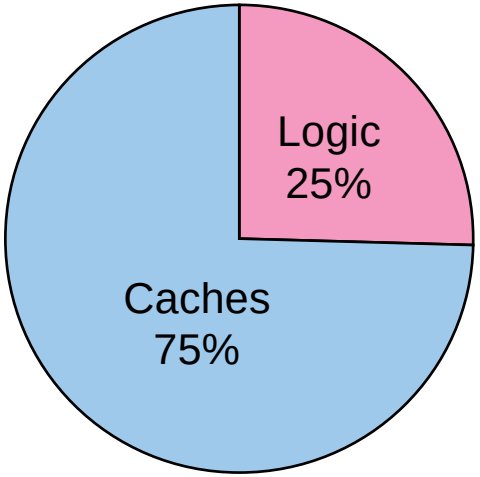


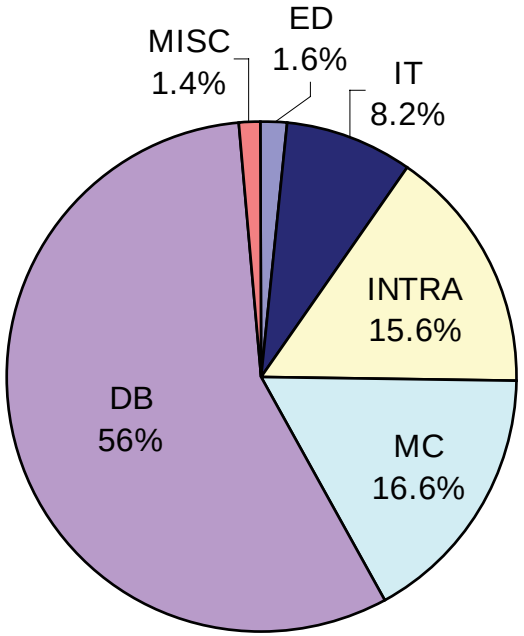
DECODER STATISTICS

Area (w/o pads)	:	2.76 x 2.76 mm^2
Area Utilization	:	31 %
Technology	:	65-nm
I/O Pads	:	176
On-chip SRAM	:	17kB

Area Breakdown

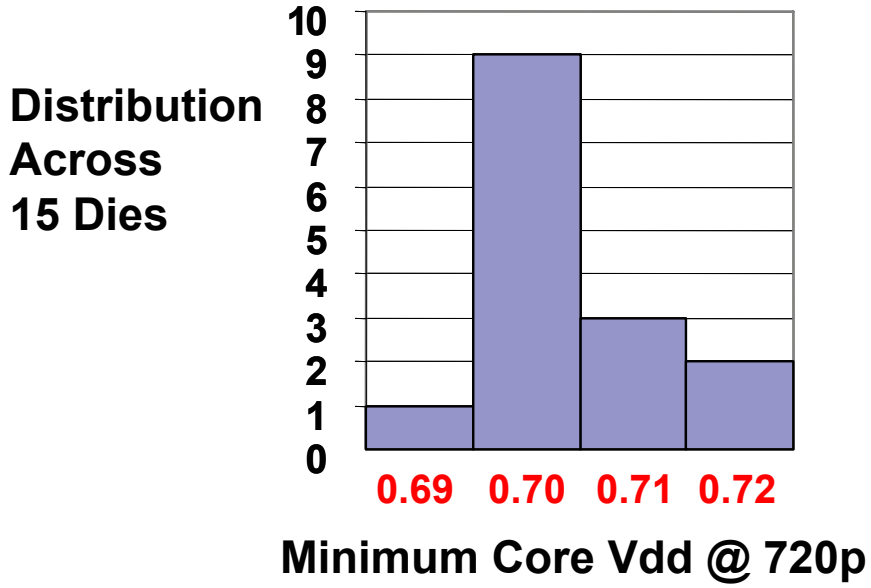
- Cache area 3x larger than logic
- Standard Cells: 134k
- Parallelism Overhead
  - 1.5% of active chip area
  - 4 luma + 2 chroma filters:
    - 1.5% of DB
  - 2 luma + 4 chroma interpolators:
    - 9% of MC



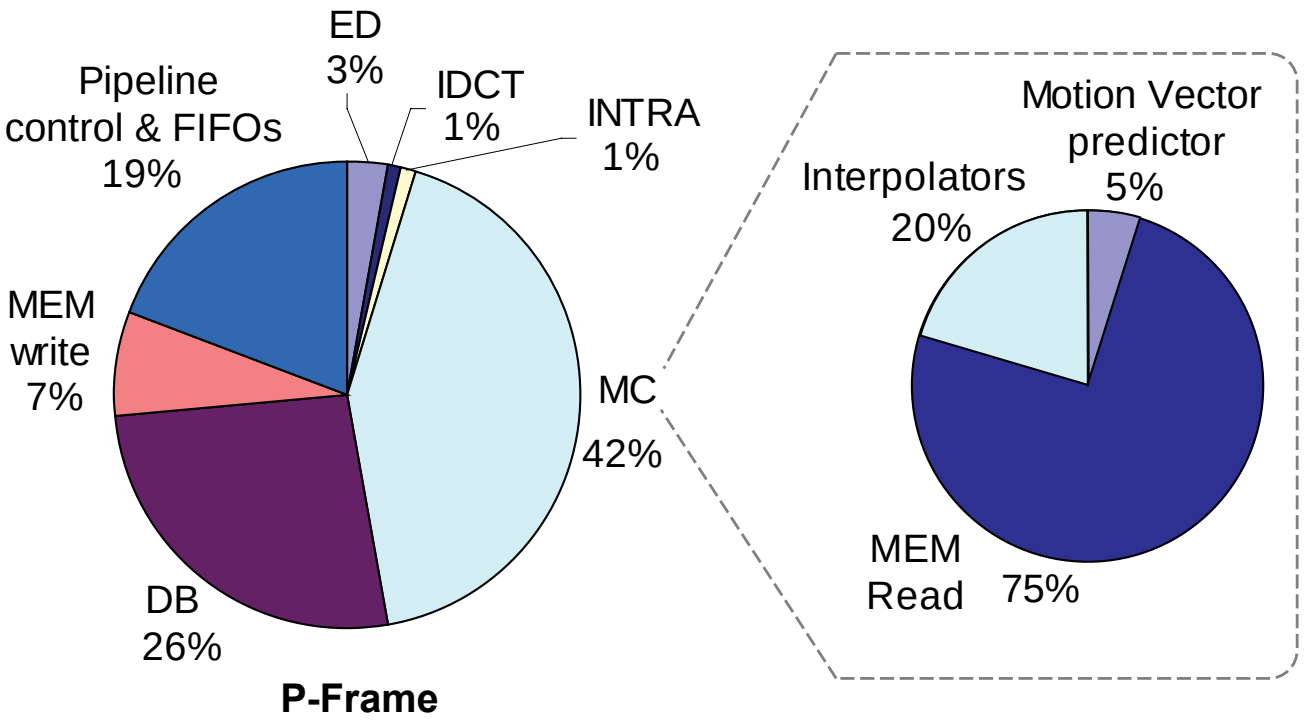


Power Measurements

720p Video	mobcal	shields	parkrun
Input Bitrate [Mbps]	5.4	7.0	26
Core [MHz @ V]	14 @ 0.70	14 @ 0.70	25 @ 0.80
MEM [MHz @ V]	50 @ 0.84	50 @ 0.84	50 @ 0.84
Power [mW]	**1.8**	**1.8**	**3.2**

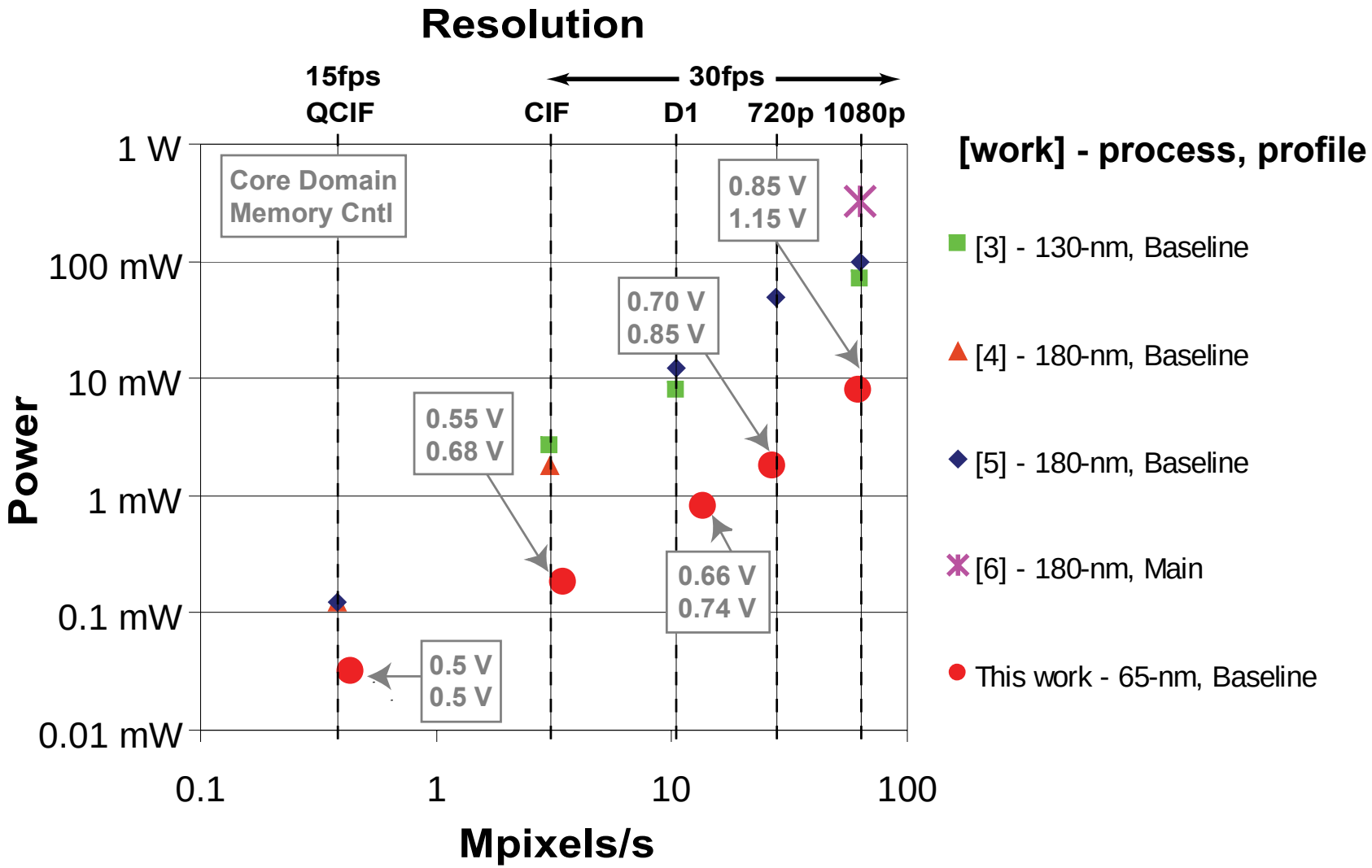


Power Breakdown



- P-frame power dominated by:
  - MC (frame buffer reads)
  - deblocking filter

Survey of Other Decoders



Summary

- Pipeline and parallelism
  - Concurrency allows 14MHz @ 720p
  - Parallelism: luma DB = 4x, luma MC = 2x
- Separate voltage/clock domains
  - 25% P-frame power savings
  - DVFS on each domain for I/P-frame differences
- Efficient memory accesses
  - Low-voltage on-chip caches and data reuse
  - Off-chip BW lowered by 40%

Acknowledgements

- Funding: Nokia, TI, and NSERC
- Chip fabrication: TI
- Valuable feedback:
  - **Nokia**: J. Hicks, G. Raghavan, J. Ankcorn
  - **TI**: M. Budagavi, D. Buss, M. Zhou
  - **MIT**: Arvind, E. Fleming

Video Demo