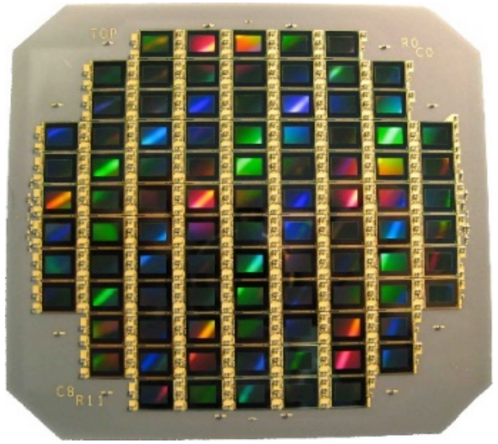
PERFECT Case Studies Demonstrating Order of Magnitude Reduction in Power Consumption

David K. Wittenberg, Edin Kadric, Andre DeHon, Jonathan Edwards, Jeffrey Smith, and Silviu Chiricescu

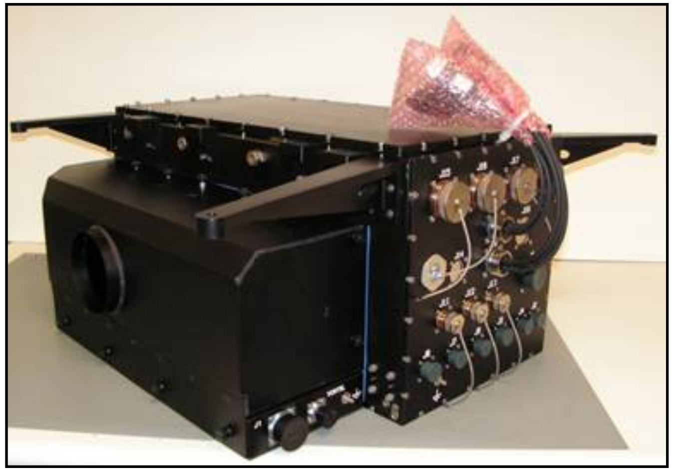
The Problem – Wide Area Motion Imaging (WAMI)

- Want real-time data
- High resolution - 368 Cell phone cameras
- 1.8 Gpixels @ 10 Hz
- Airborne
- Limited bandwidth to ground
- Limited power

ARGUS-IS System Components



ARGUS-IS system components. Left: sensor in 25” gimbal; Right: CFPA with 92 5 MPixel FPAs.



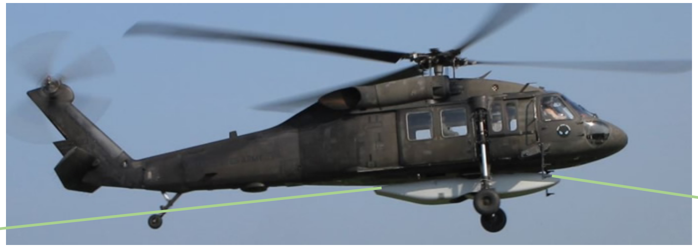
ARGUS-IS ruggedized airborne processor

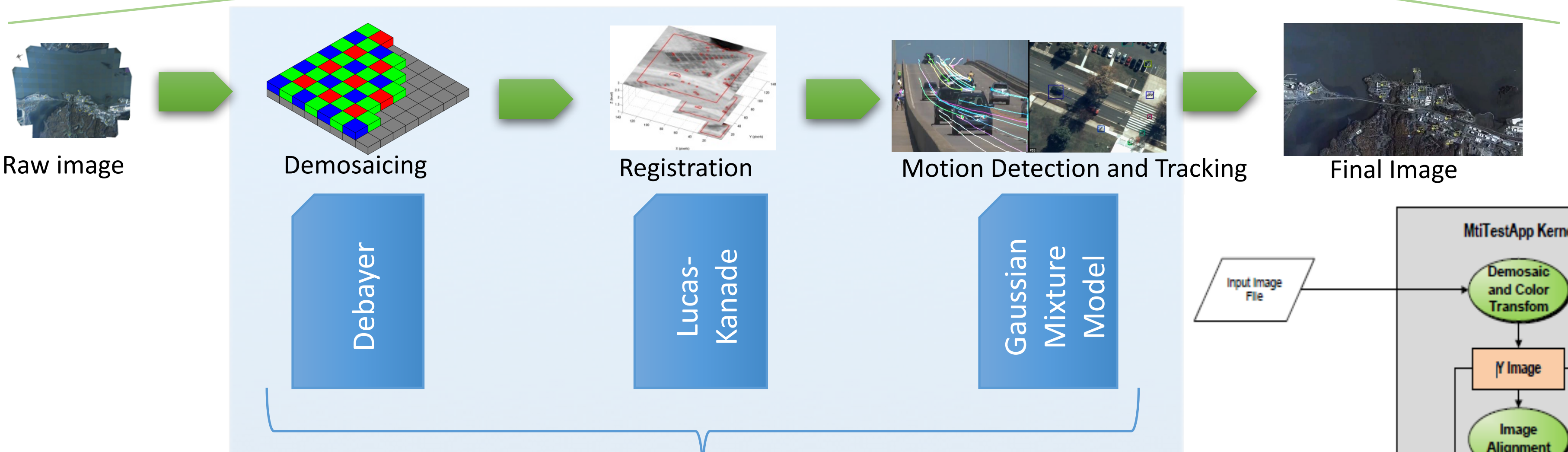
Our Solution

- Lower Voltage
- Light-Weight Checks
- Explore Parallelism
- Continuous Hierarchy Memory

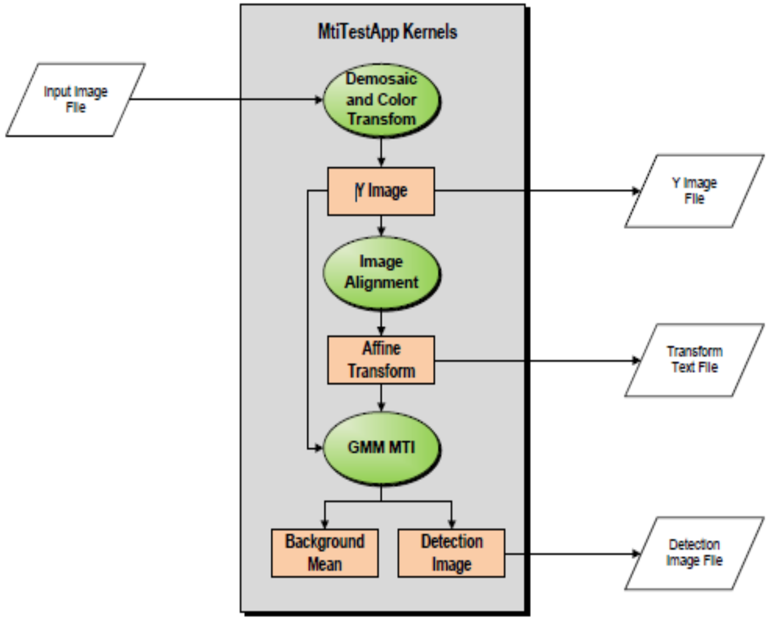
ARGUS-IS context drives PERFECT energy improvements

Measure energy on the ARGUS-IS WAMI pipeline components

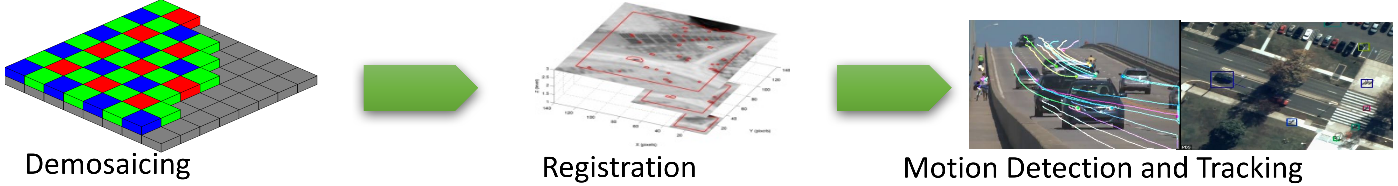




Measure energy on the PRACTICE LEARN WAMI implementation



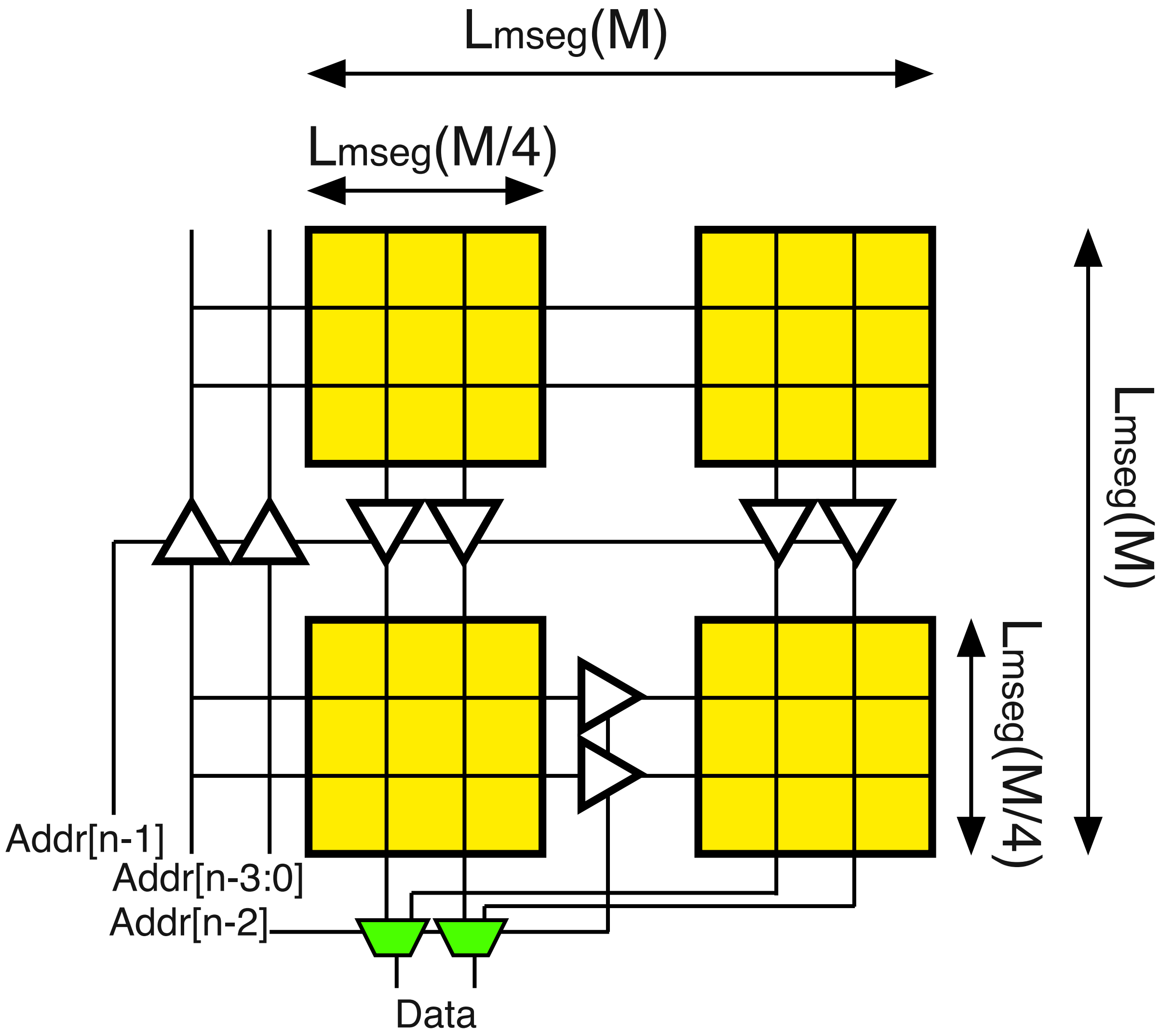
Parallel Optimized Design

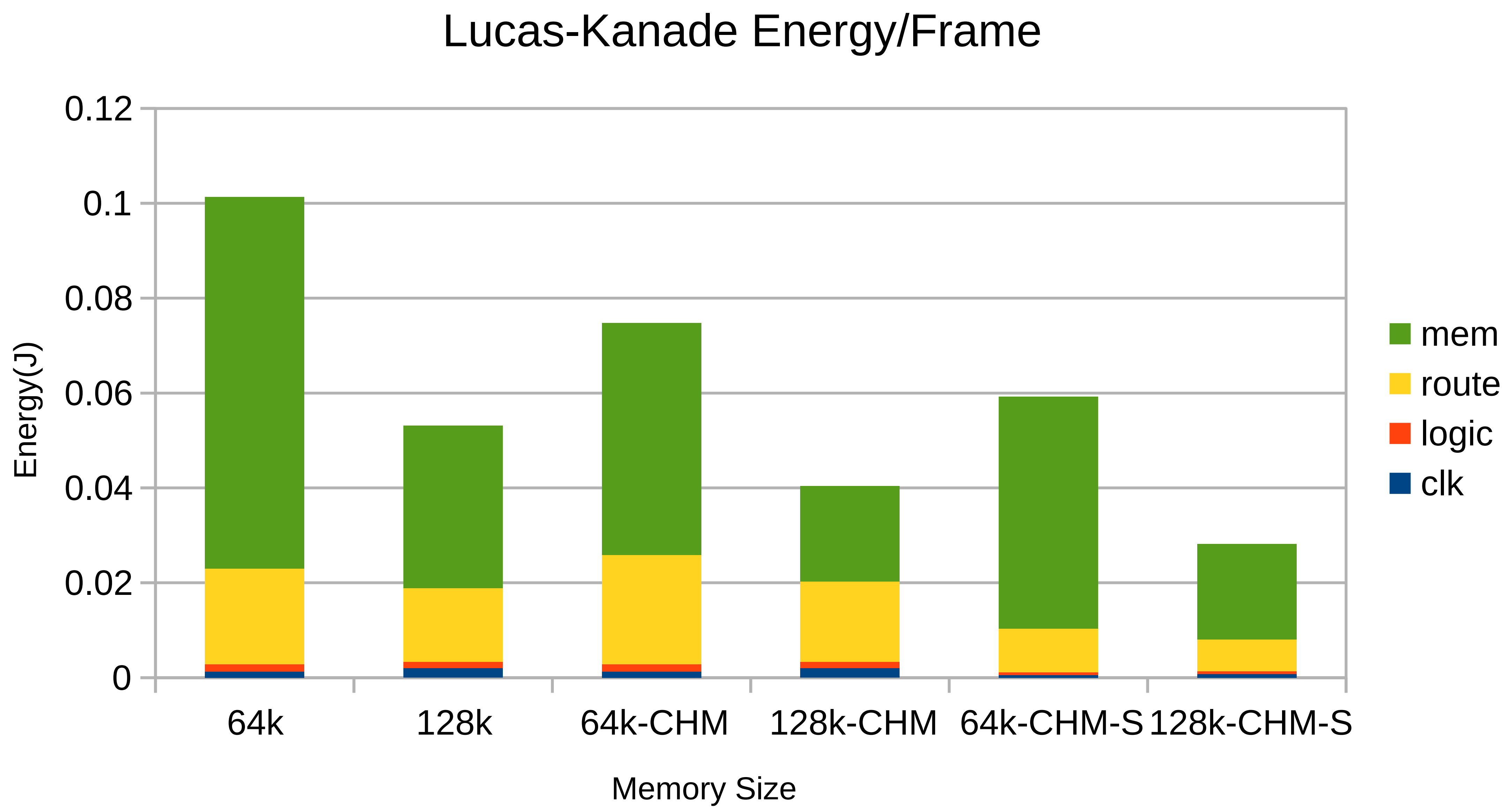


512×512 Image

		Debayer	LK	GMM	WAMI
	PEs	1	4	8	(4,4,4)
	Ops/Cycle	20	220	680	360
	Area (cm^2)	0.0046	0.33	0.2	0.55
0.7V Vdd	Throughput (fps)	540	150	1600	150
	Efficiency (Gops/W)	740	280	910	340
0.5V Vdd	Throughput (fps)	110	30	320	50
	Efficiency (Gops/W)	1500	560	1800	680

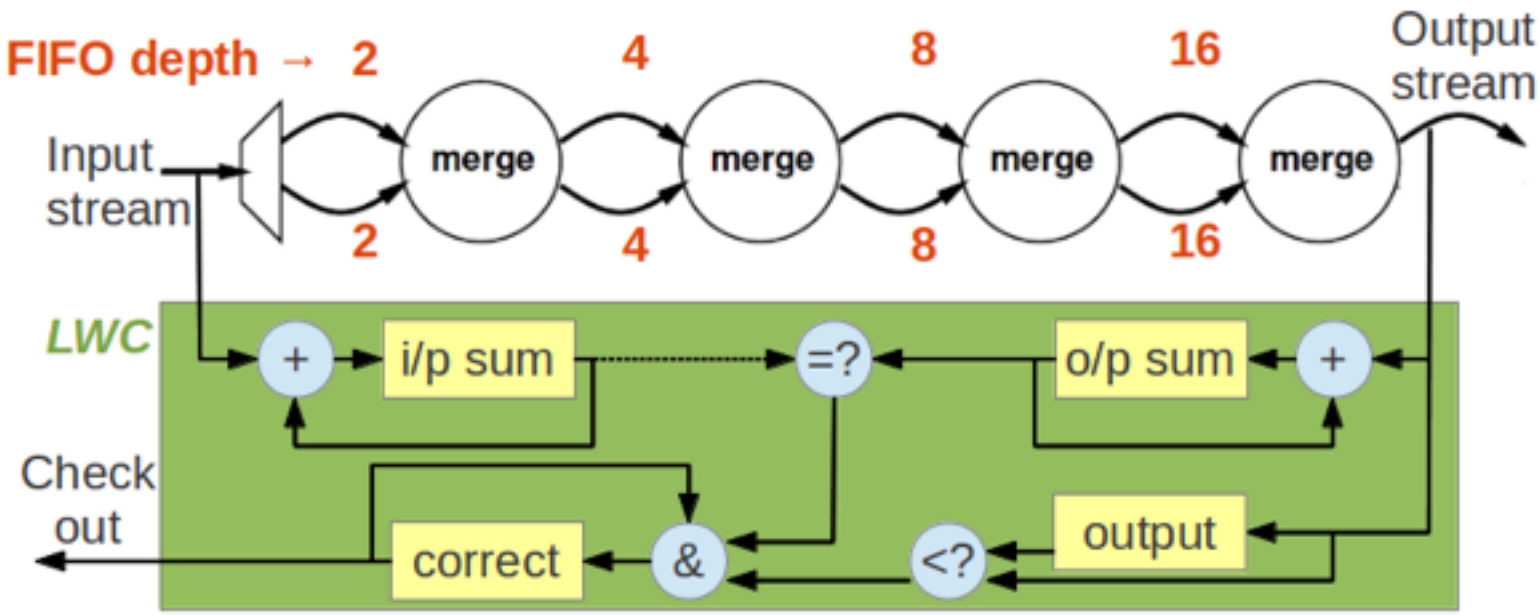
Continuous Hierarchy Memory



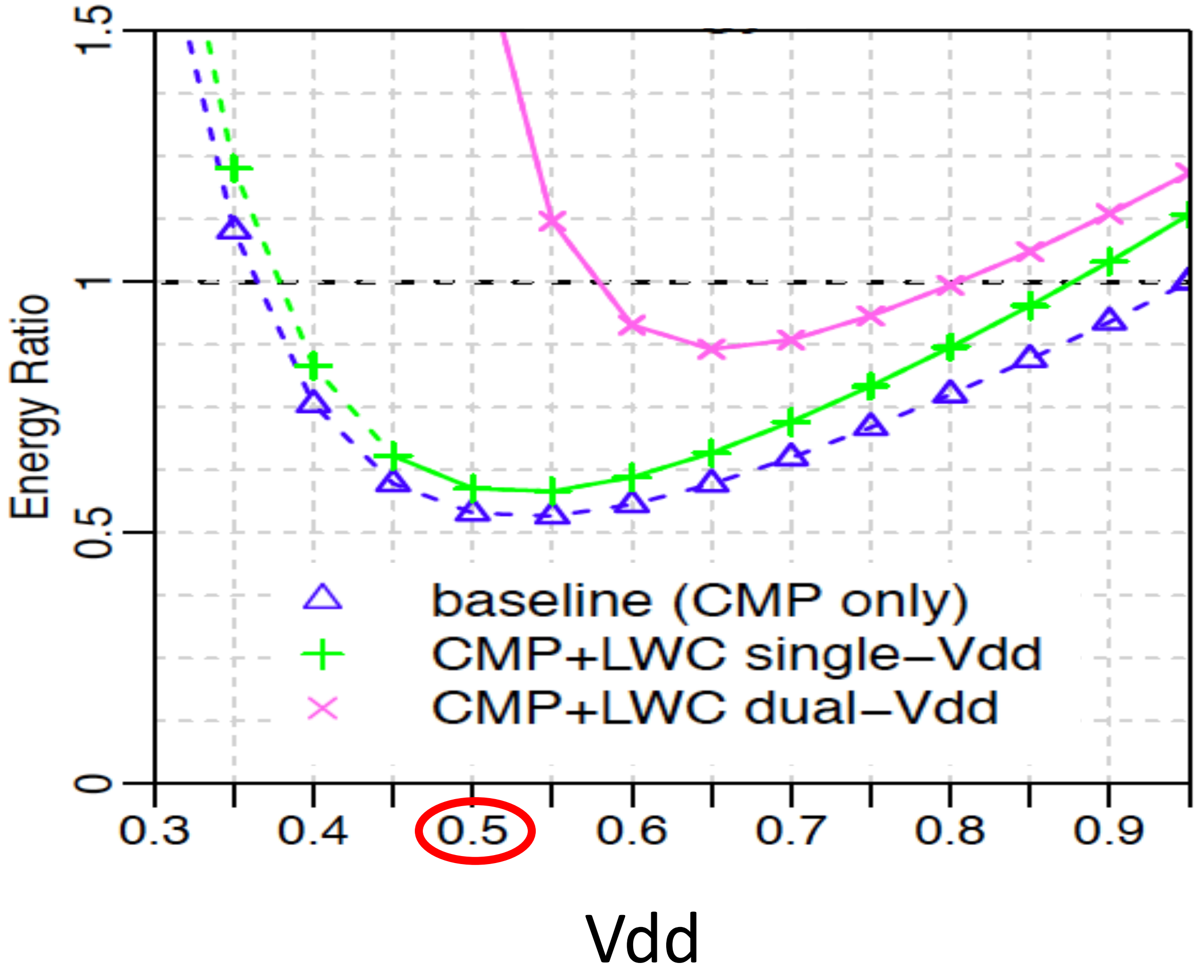

Lucas-Kanade Energy/Frame
Energy(J)
0.12
0.1
0.08
0.06
0.04
0.02
0
64k
128k
64k-CHM
128k-CHM
64k-CHM-S
128k-CHM-S
Memory Size
mem
route
logic
clk

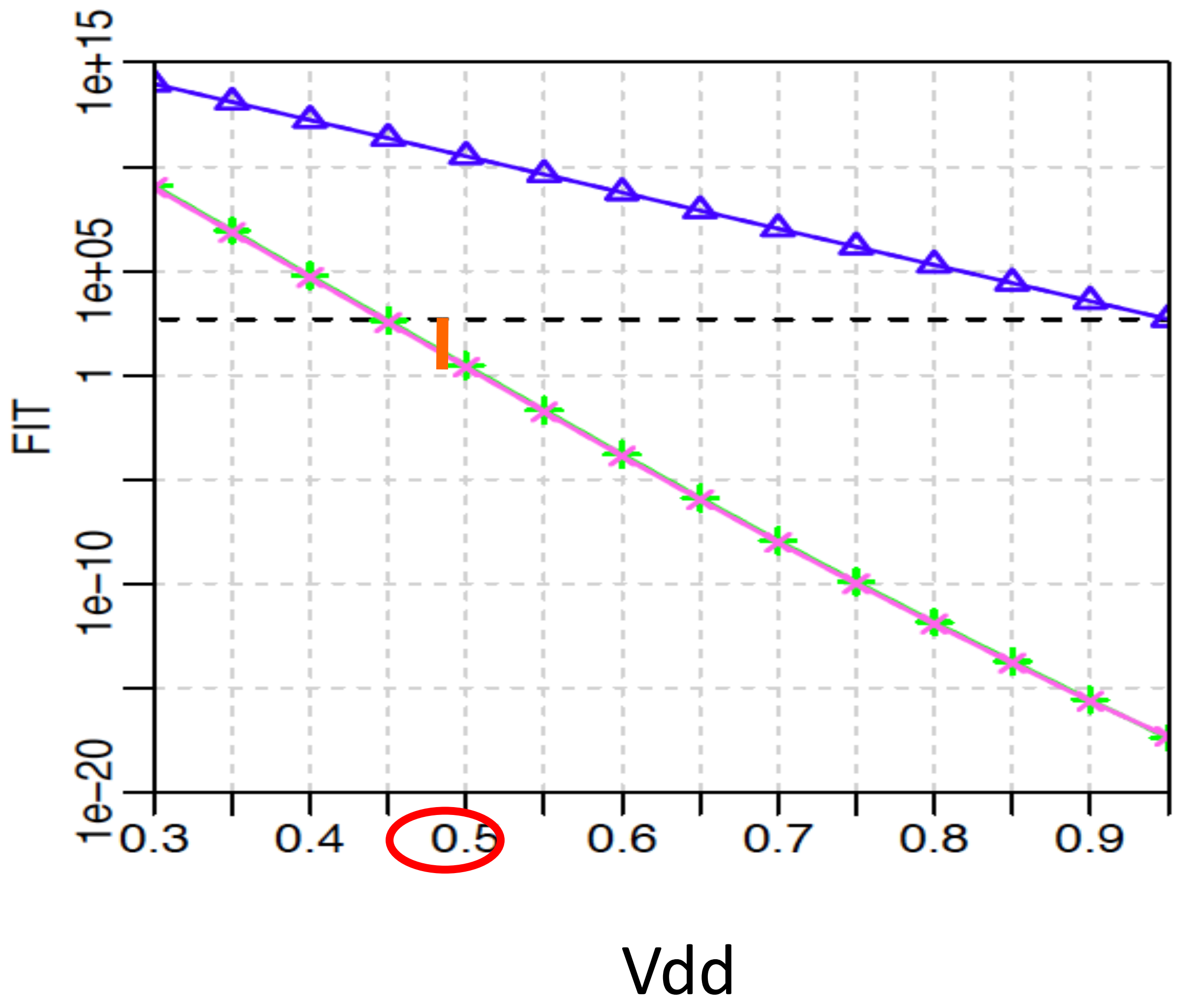
Light Weight Checks

- Example: Sorting vs. checking that it is sorted
- Not needed in many imaging algorithms, as they iterate until errors are small enough



Differential vs. Single Voltage (22nm example)





PERFECT Results and Comparison with ARGUS

- **PERFECT Simulated Output for WAMI Pipeline**
  - Scaled 512x512 pixel 5 Mpixel and further scaled to achieve 3000 frame run
    - 7 nm process technology assumed as best-case
  - Varying Vdd and Mem-size led to optimized GOPS / W
    - Vdd = 0.7, Mem size = 64k
- **Key Results:**
  - 237X GOPS/W increase even with memory intensive pipeline
  - Performed 4X more operations due to differences in implementation

Comparison of ARGUS-IS GPU and PERFECT output

Technology	Power (W)	Time (s)	Energy (J)	GOps	GOPS / W
ARGUS-IS GPU	77.4	6.5	507.2	940.6	1.9
PERFECT	0.5	17.1	0.9	4096.6	439.9
Ratio	0.01	2.6	0.002	4.4	236.9

TAILWIND: PERFECT gives 6-15X Improvement

	TAILWIND SYSTEM		PERFECT SYSTEM			
	Image Registration (SLAM)	Motion Detection (RFD)	Image Registration (LK)		Motion Detection (GMM)	
			STD	CHM	STD	CHM
Energy/Frame (mJ)	752	237	175	129	19	14
Time/Frame (ms)	253	83	444		37	

- Motion Detection – PERFECT kernel consumes ≈ ***15X*** less energy than the TAILWIND kernel
- Image Registration – PERFECT kernel consumes ≈ ***6X*** less energy than the TAILWIND kernel

Summary

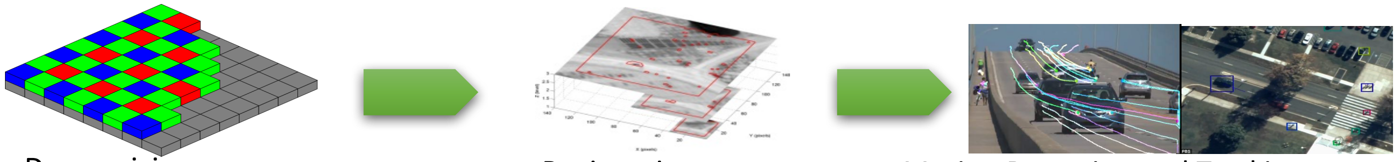
1. Exceeded PERFECT energy efficiency goals by up to 10x (440-680 Gops/W) on common set of Wide Area Motion Imagery (WAMI) kernels
2. The projected Argus power consumption reduction enables:
   - Airborne, real-time 3D situational awareness to the warfighter along with:
     - Miniaturization of 3D mapping technology currently housed in multiple server racks
     - Increased the flight mission time and track detection accuracy with improved data compression in manned flights
   - Autonomous threat detection, tracking and multi-sensor fusion
3. Designed a novel FPGA based architecture that combines ultra low power operation with high reliability
   - Optimize the substrate and design mapping to minimize communication energy

Acronyms Used

ARGUS	Autonomous Real-time Ground-Ubiquitous Surveillance-Imaging System
CHM	Continuous Hierarchy Memory
FIT	Failures In Time
FPGA	Field Programmable Gate Array
GMM	Gaussian Mixture Model
LEARN	Low-Energy Architecture of Reconfigurable Nodes
L-K	Lucas-Kanade
LWC	Light Weight Checks
PERFECT	Power Efficiency Revolution for Embedded Computing Technology
RANSAC	RANdom SAmple Consensus
RFD	Robust Frame Differences
SLAM	Simultaneous Location And Mapping
TAILWIND	Tactical Aircraft to Increase Long Wave Infrared Nighttime Detection
UAV	Unmanned Aerial Vehicle
WAMI	Wide Area Motion Imagery

BACKUP

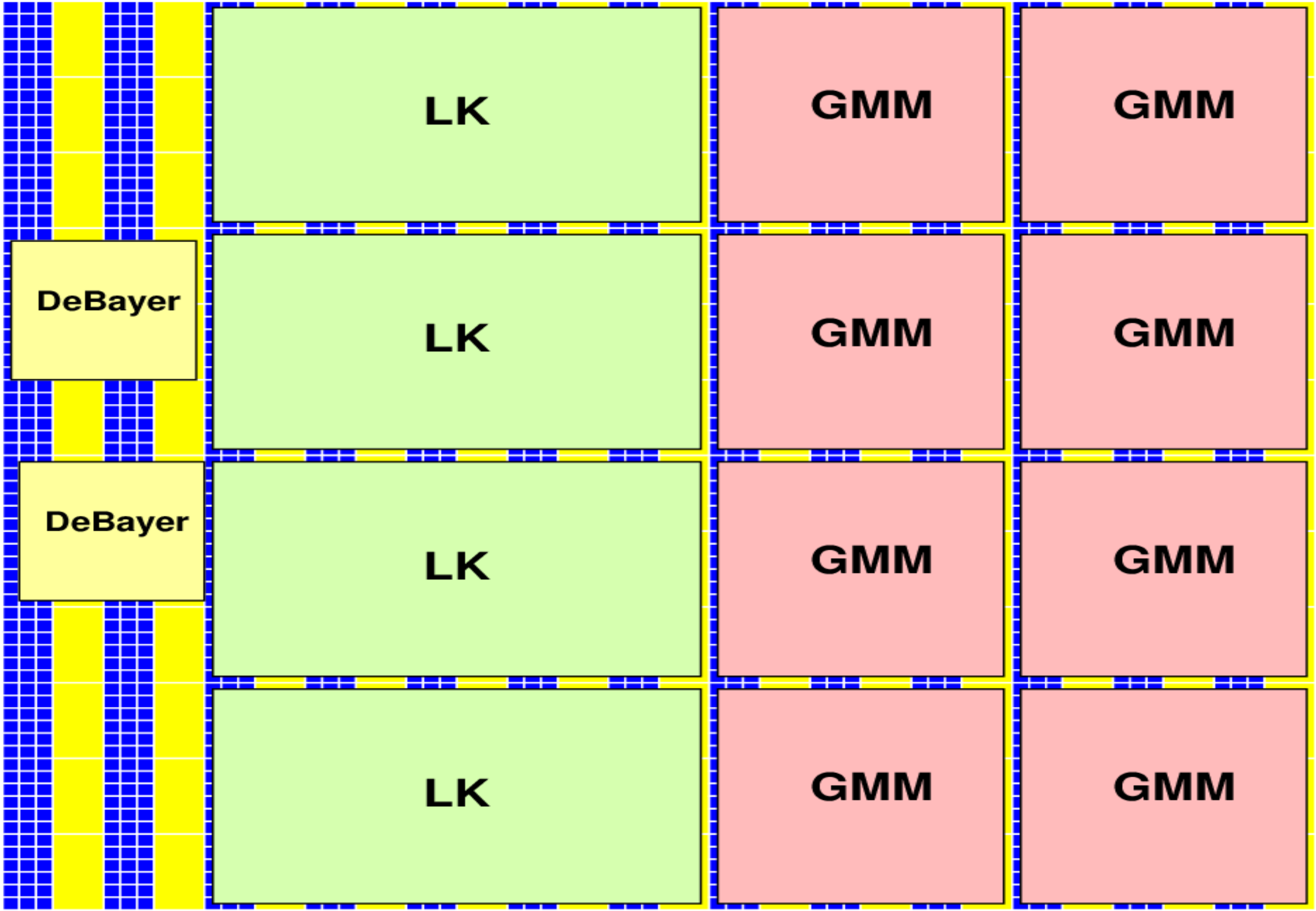
Minimal Streaming Design



Fully functional pipeline

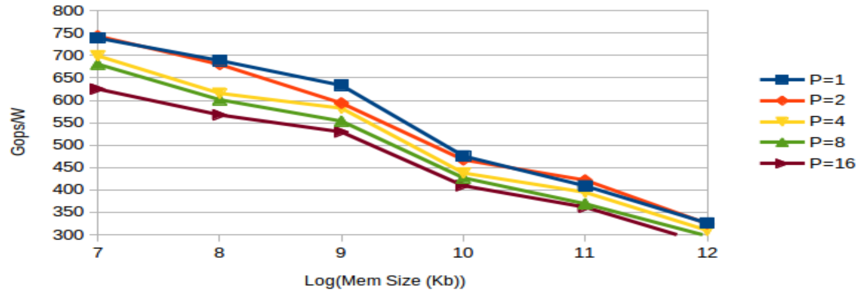
512×512 Image

		Debayer	LK	GMM	WAMI
	PEs	1	1	1	(1,1,1)
	Ops/Cycle	20	55	85	90
	Area (cm^2)	0.0046	0.13	0.034	0.17
0.7V V_{dd}	Throughput (fps)	540	58	430	57
	Efficiency (Gops/W)	740	250	650	300
0.5V V_{dd}	Throughput (fps)	110	12	85	11
	Efficiency (Gops/W)	1500	500	1300	600

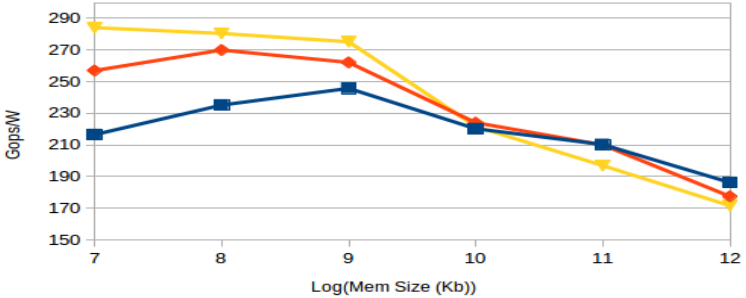

LK
GMM
GMM
DeBayer
LK
GMM
GMM
DeBayer
LK
GMM
GMM
LK
GMM
GMM

Optimizing Parallelism

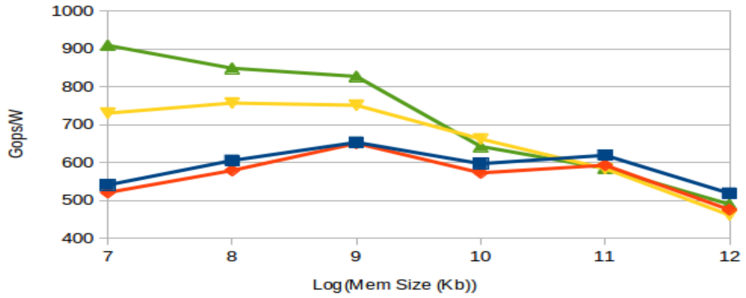
DeBayer



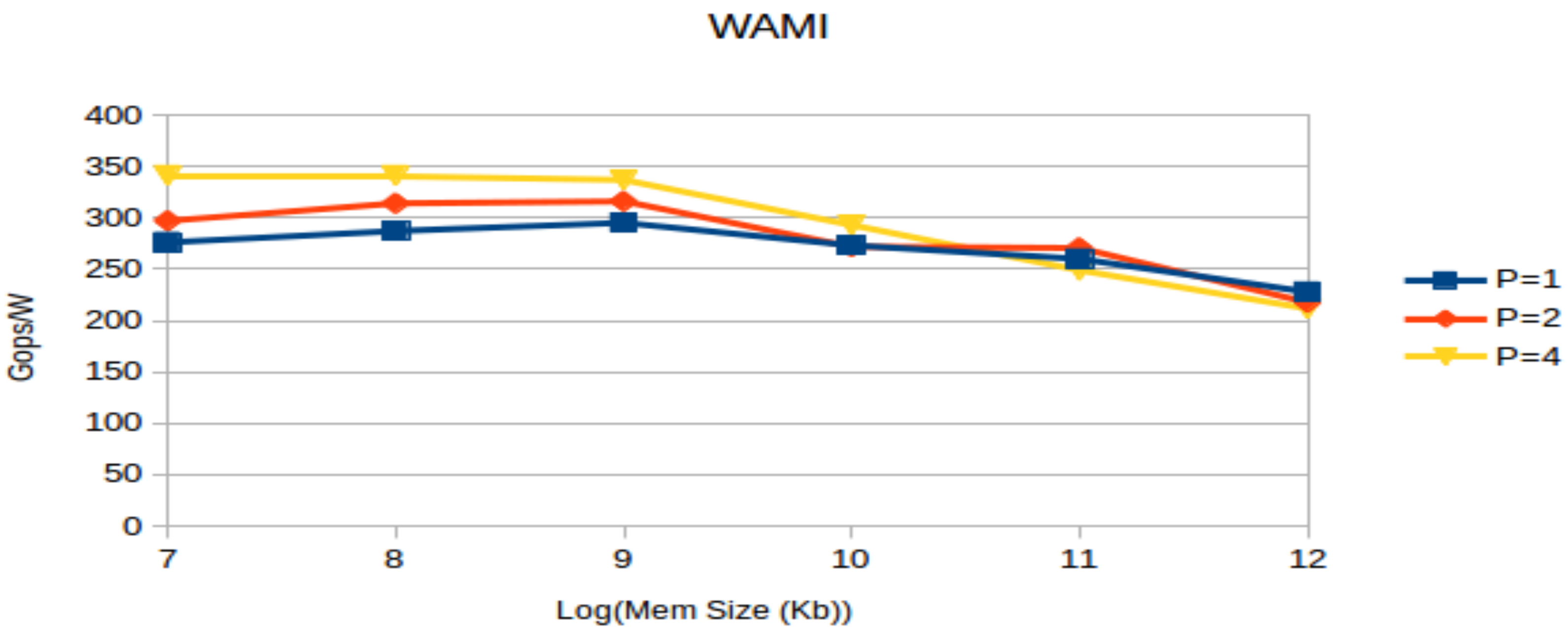
LK

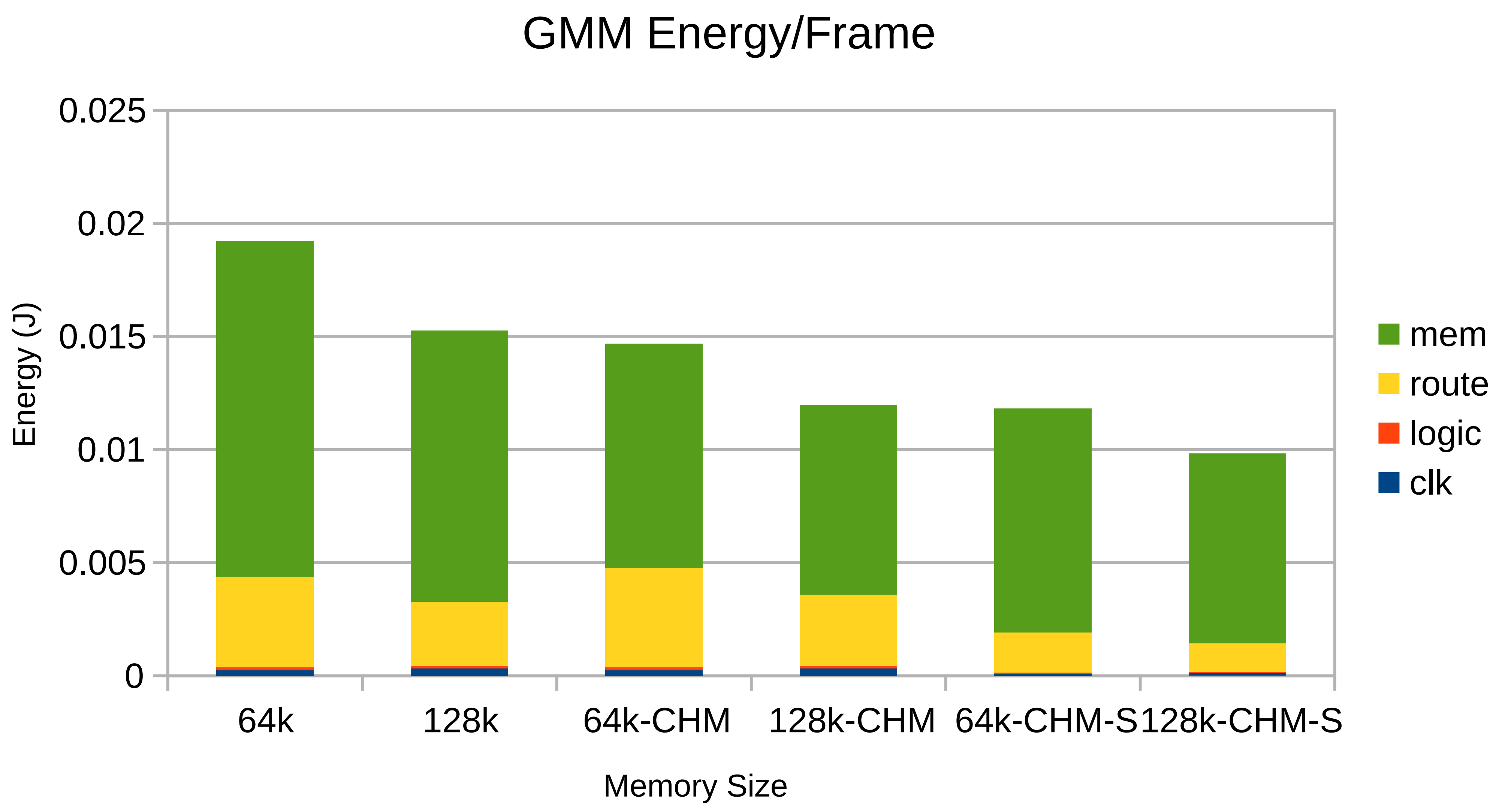


GMM



512×512 Image

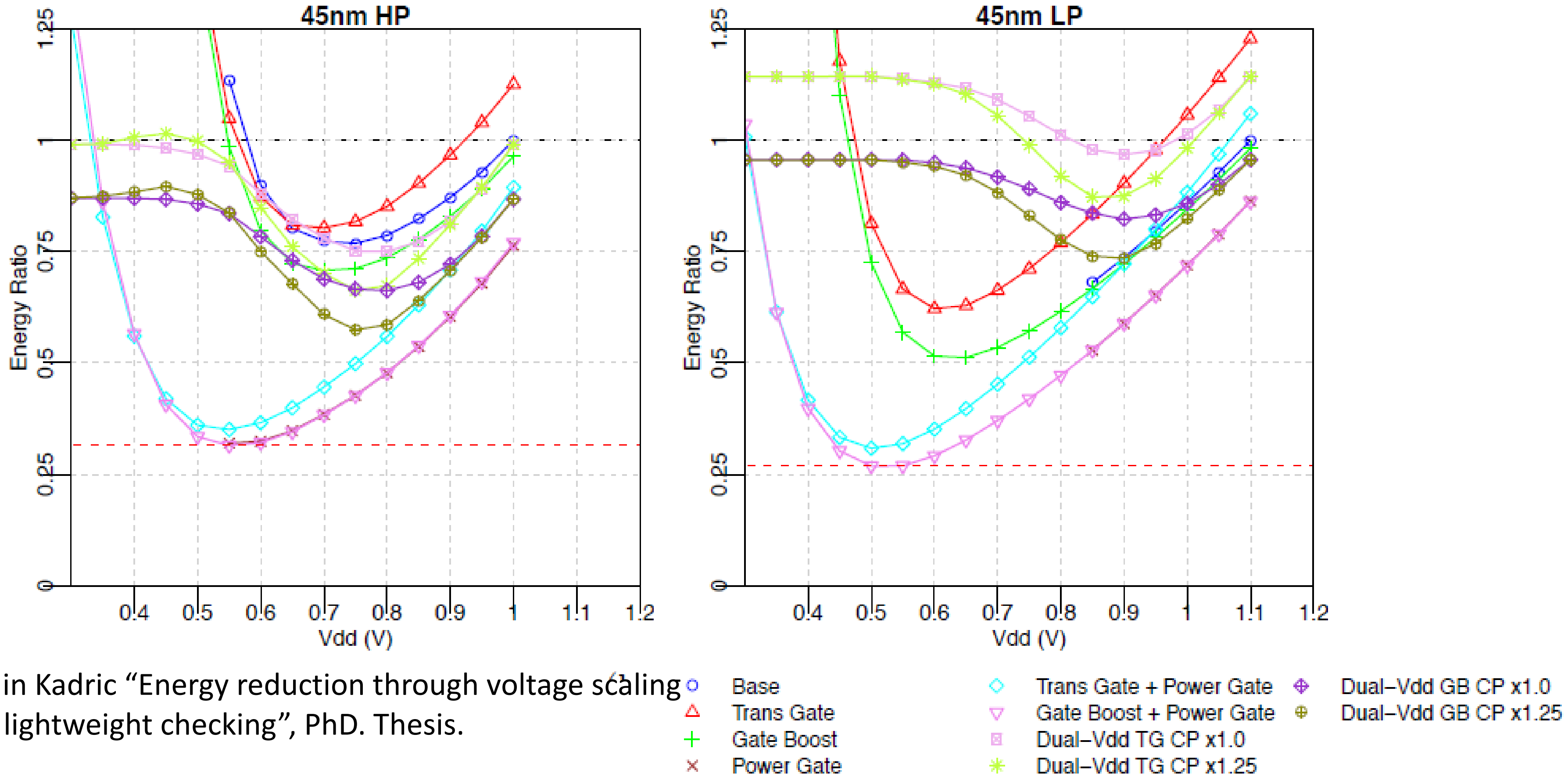




GMM Energy/Frame
Energy (J)
0.025
0.02
0.015
0.01
0.005
0
64k
128k
64k-CHM
128k-CHM
64k-CHM-S
128k-CHM-S
Memory Size
mem
route
logic
clk

Energy reduction techniques*

- Use low-power technology
- Voltage scaling
- Use T-gates or gate-boosting (GB) instead of pass-T w/ level restorer
- Use dual-Vdd
- Power gating



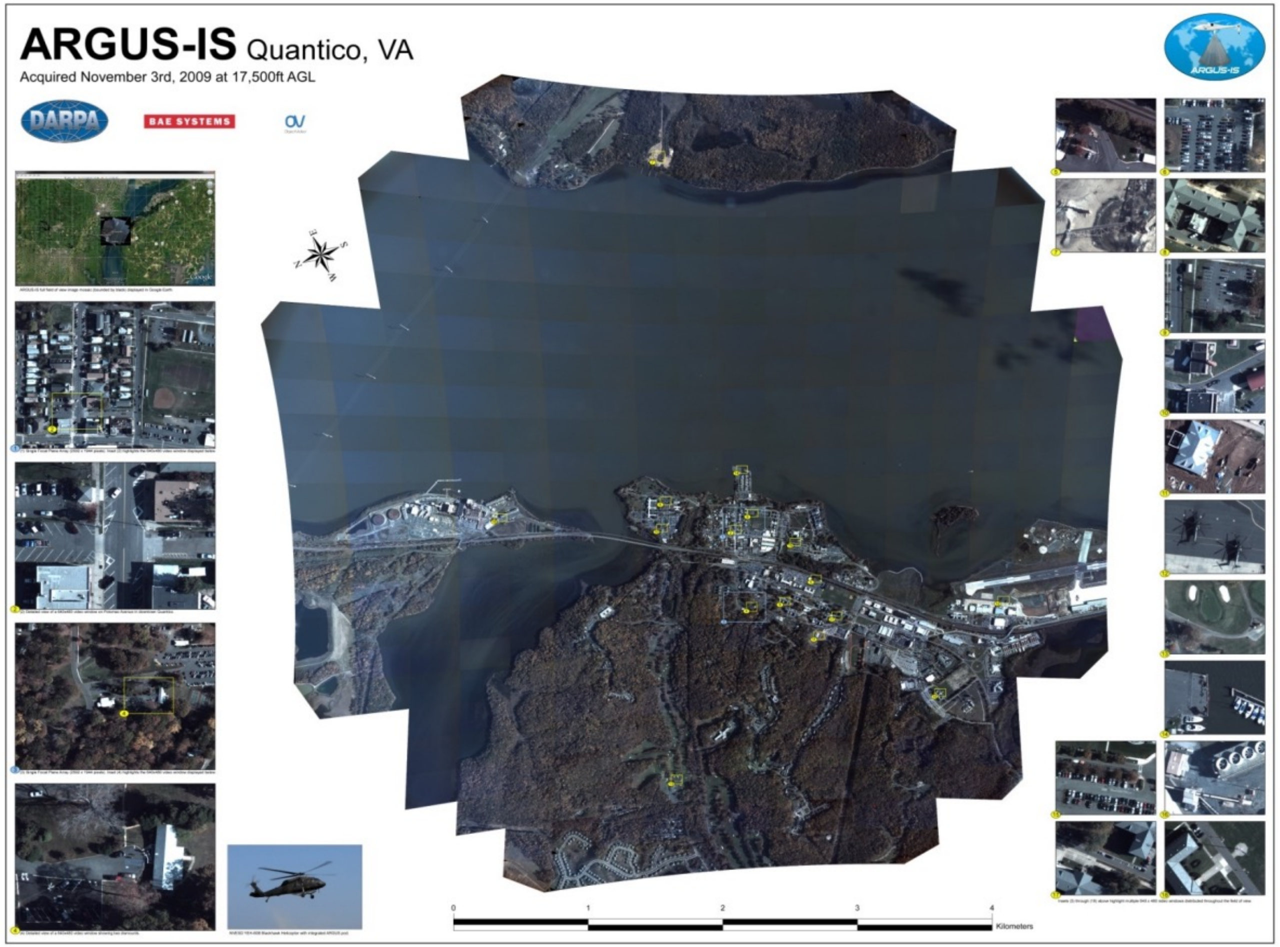
* Edin Kadric "Energy reduction through voltage scaling and lightweight checking", PhD. Thesis.

TAILWIND System Level Metrics

Data Set	System Level Metric	SLAM + RFD	LK + RFD	SLAM + GMM
Quantico Scene 1	P(vehicle detection)	0.9	0.53	0.5
	P(dismount detection)	0.64	0.38	0.46
	False alarm/minute	1.35	103	163
	P(tracking vehicle)	0.3	0.1	0
	P(tracking dismount)	0.33	0.14	0.14
Quantico Scene 2	P(vehicle detection)	0.82	0.89	0.3
	P(dismount detection)	0.52	0.71	0.25
	False alarm/minute	8.9	149	124
	P(tracking vehicle)	0.42	0.25	0
	P(tracking dismount)	0.44	0.28	0.1
Quantico Scene 3	P(vehicle detection)	0.85	0.54	0.24
	P(dismount detection)	0.6	0.42	0.19
	False alarm/minute	4.1	117	126
	P(tracking vehicle)	0.15	0.13	0.03
	P(tracking dismount)	0.21	0.09	0.27

- TAILWIND
  - Fine-tuned parameters for the different kernels
- PERFECT
  - Dismounts are at granularity of a few pixels which is more than the difference between consecutive frames – large false alarm rates
  - GMM has lower levels of accuracy than RFD

ARGUS-IS Public Release Poster



References from PERFECT project

1. Edin Kadrik, David Lakata and André DeHon. “Impact of Memory Architecture on FPGA Energy Consumption”, FPGA’15, Monterey, California, USA, Feb 22-24, 2015, http://ic.ese.upenn.edu/pdf/meme_fpga2015.pdf.
2. Benjamin Gojman, Sirisha Nalmela, Nikil Mehta, Nicholas Howarth, and André DeHon, “GROK-LAB: Generating Real On-chip Knowledge for Intra-cluster Delays using Timing Extraction”, ACM Transactions on Reconfigurable Technology and Systems (TRETS) , Volume 7, Number 4, DOI: 10.1145/2597889, December, 2014.
3. E. Kadric, K. Mahajan, A. DeHon, "Kung Fu Data Energy—Minimizing Communication Energy in FPGA Computations", IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM), May 11–13, 2014
4. E. Kadric, K. Mahajan, A. DeHon, "Energy Reduction through Differential Reliability and Lightweight Checking", FPGA’14, February 26–28, 2014, Monterey, California, USA.
5. Feng Liu, Soumyadeep Ghosh, Nick P. Johnson, David I. August. “CGPA: Coarse-Grained Pipelined Accelerators”, Design Automation Conference (DAC), 2014 51st ACM/EDAC/IEEE, 1-5 June 2014, San Francisco, CA, http://dl.acm.org/citation.cfm?doid=2593069.2593105
6. Andre DeHon, “Fundamental Underpinnings of Reconfigurable Computing Architectures”, accepted for publication in the Proceedings of the IEEE , Volume 103, Issue: 3, DOI: 10.1109/JPROC.2014.2387696, March 2015, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7086421
7. Feng Liu, Heejin Ahn, Stephen R. Beard, Taewook Oh, and David I. August, “DynaSpAM: Dynamic Spatial Architecture Mapping using Out of Order Instruction Schedules”, Proceedings of the 42nd International Symposium on Computer Architecture (ISCA), ISBN 978-1-4503-3402-0/15/06, Portland, OR, USA, June 13-17, 2015.
8. Edin Kadrik, David Lakata and André DeHon, "Impact of Parallelism and Memory Architecture on FPGA Communication Energy", ACM Transactions on Reconfigurable Technology and Systems, Vol. 9, to be published date: 2016. 0/15/06, Portland, OR, USA, June 13-17, 2015.
9. “Demonstration of Applicability of PERFECT/PRACTICE Technology to TAILWIND Processing”, BAE Systems Technical Report for DARPA PERFECT Program
10. “Demonstration of Applicability of PERFECT-PRACTICE Technology to ARGUS-IS Processing”, BAE Systems Technical Report for DARPA PERFECT Program