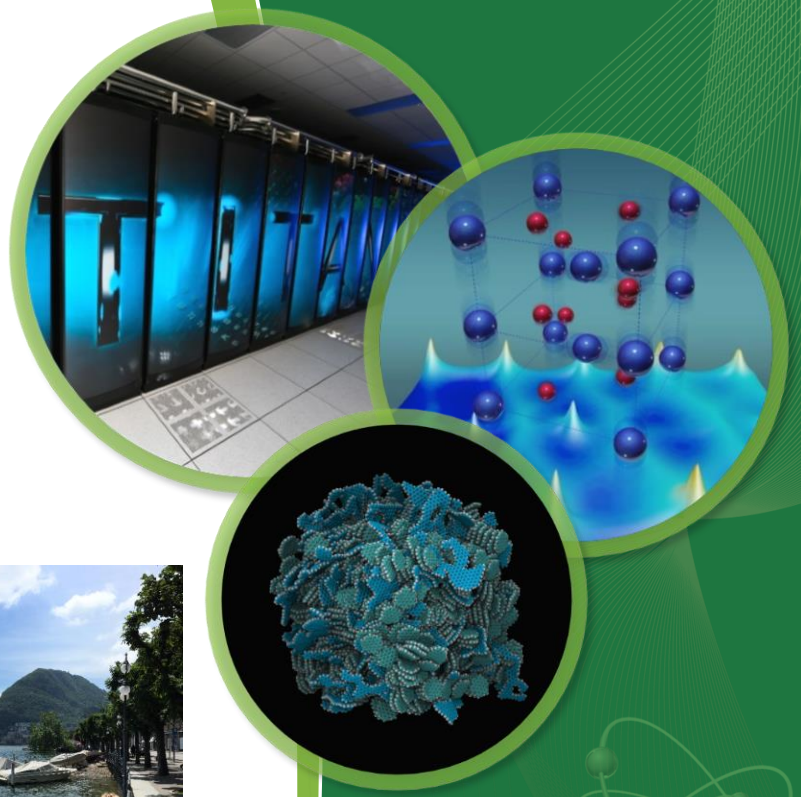
Understanding Portability of a High-Level Programming Model on Diverse HPC Architectures

Jeffrey S. Vetter

Seyong Lee, Jeremy Meredith,

and many collaborators



Presented to
Accelerated Data and Computing Workshop

13 June 2016
Lugano

http://ft.ornl.gov vetter@computer.org

Overview

- Both architectures and applications are growing more complex
  - Trends dictate that this will get worse, not better
  - This complexity creates irregularity in computation, communication, and data movement
- Programming models must provide performance portability in this environment
- Examples
  - Heterogeneous computing
  - Memory hierarchies

OAK RIDGE National Laboratory

Current ASCR Computing At a Glance

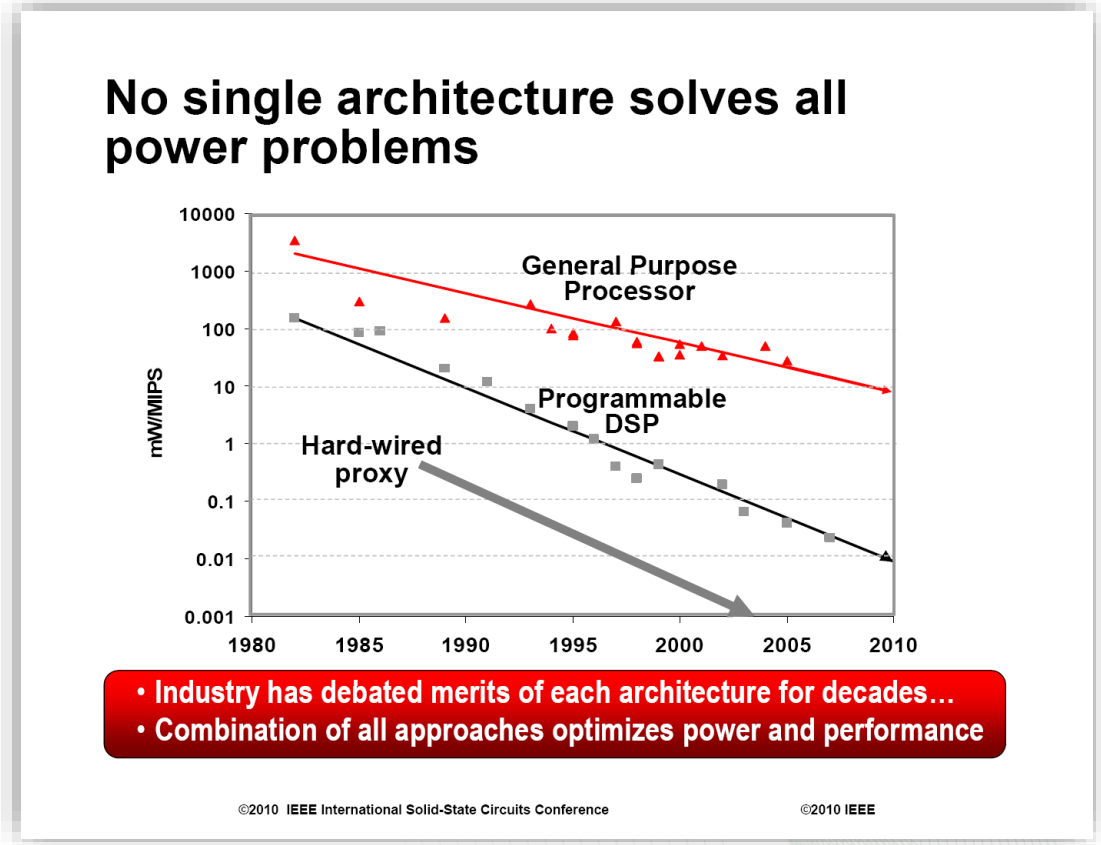
Complexity α T

System attributes	NERSC Now	OLCF Now	ALCF Now	NERSC Upgrade	OLCF Upgrade	ALCF Upgrades	
Planned Installation	**Edison**	**TITAN**	**MIRA**	**Cori 2016**	**Summit 2017-2018**	**Theta 2016**	**Aurora 2018-2019**
System peak (PF)	2.6	27	10	> 30	150	>8.5	180
Peak Power (MW)	2	9	4.8	< 3.7	10	1.7	13
Total system memory	357 TB	710TB	768TB	~1 PB DDR4 + High Bandwidth Memory (HBM)+1.5PB persistent memory	> 1.74 PB DDR4 + HBM + 2.8 PB persistent memory	>480 TB DDR4 + High Bandwidth Memory (HBM)	> 7 PB High Bandwidth On-Package Memory Local Memory and Persistent Memory
Node performance (TF)	0.460	1.452	0.204	> 3	> 40	> 3	> 17 times Mira
Node processors	Intel Ivy Bridge	AMD Opteron Nvidia Kepler	64-bit PowerPC A2	Intel Knights Landing many core CPUs Intel Haswell CPU in data partition	Multiple IBM Power9 CPUs & multiple Nvidia Voltas GPUS	Intel Knights Landing Xeon Phi many core CPUs	Knights Hill Xeon Phi many core CPUs
System size (nodes)	5,600 nodes	18,688 nodes	49,152	9,300 nodes 1,900 nodes in data partition	~3,500 nodes	>2,500 nodes	>50,000 nodes
System Interconnect	Aries	Gemini	5D Torus	Aries	Dual Rail EDR-IB	Aries	2nd Generation Intel Omni-Path Architecture
File System	7.6 PB 168 GB/s, Lustre®	32 PB 1 TB/s, Lustre®	26 PB 300 GB/s GPFS™	28 PB 744 GB/s Lustre®	120 PB 1 TB/s GPFS™	10PB, 210 GB/s Lustre initial	150 PB 1 TB/s Lustre®

OAK RIDGE National Laboratory

Architectural specialization will continue and accelerate

- Accelerators and SoCs already dominate multiple markets
- Vendors, lacking Moore's Law, will need to continue to offer new products (to stay in business)
  - Grant that advantage of better CMOS process stalls
  - Use the same transistors differently to enhance performance
- Architectural design will become extremely important, critical
  - Address new parameters for benefits/curse of Moore's Law

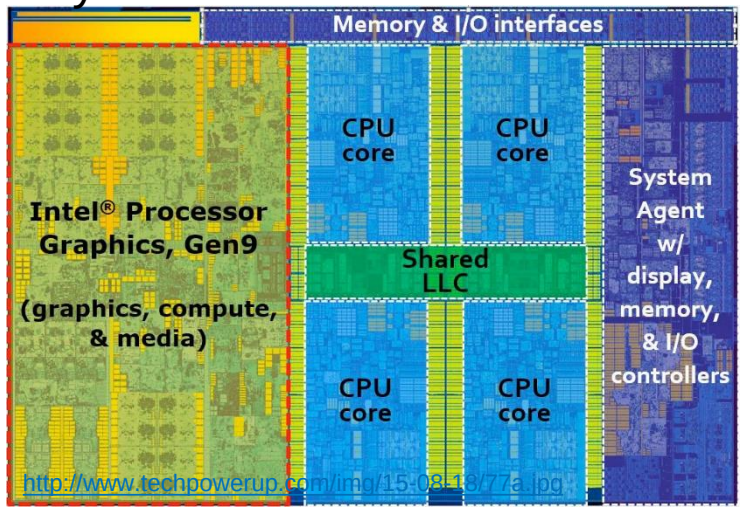


Delagi, ISSCC 2010

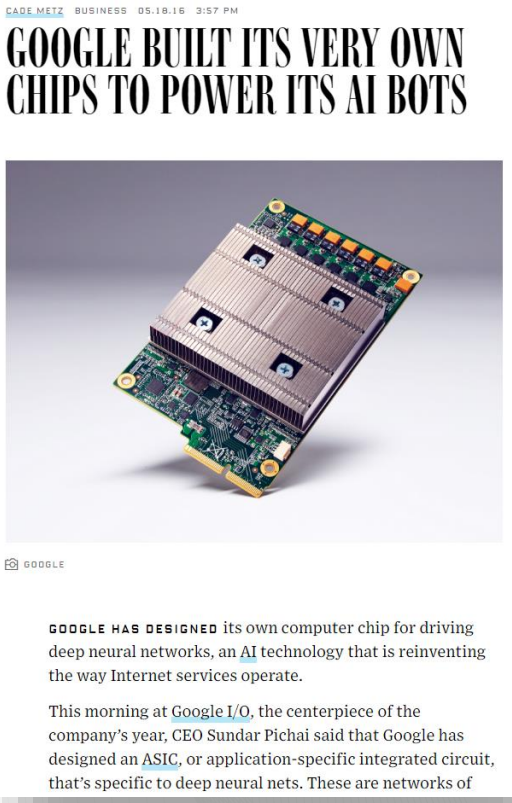
Specialization is here to stay: Core, Processor Architectures

- Cores
  - CPU
  - GPUs (discrete, integrated)
  - FPGAs
  - Special purpose engines
    - RNGs
    - AES, video engines
    - Transactional memory
    - Virtualization support
- SIMD/short vector
- SMT, threading models
- DVFS (incl Turboboost)
- etc

Skylake



http://www.techpowerup.com/img/15-08-18/77a.jpg



CADE METZ BUSINESS 05.18.16 3:57 PM

GOOGLE BUILT ITS VERY OWN CHIPS TO POWER ITS AI BOTS

GOOGLE

GOOGLE HAS DESIGNED its own computer chip for driving deep neural networks, an AI technology that is reinventing the way Internet services operate.

This morning at Google I/O, the centerpiece of the company's year, CEO Sundar Pichai said that Google has designed an ASIC, or application-specific integrated circuit, that's specific to deep neural nets. These are networks of

http://www.wired.com/2016/05/google-tpu-custom-chips/

Pascal

http://cdn.wccftech.com/wp-content/uploads/2014/03/NVIDIA-Pascal-GPU-Chip-Module.jpg

D.E. Shaw, M.M. Deneroff, R.O. Dror *et al.*, *"Anton, a special-purpose machine for molecular dynamics*

OAK RIDGE National Laboratory

Recent news

Nvidia and IBM create GPU interconnect for faster supercomputing

"NVLink" shares up to 80GB of data per second between CPUs and GPUs.

by Jon Brodkin - Mar 25 2014, 2:45pm EST

NETWORKING SUPERCOMPUTING 34

It Begins: AMD Announces Its First ARM Based Server SoC, 64-bit/8-core Opteron A1100

by Anand Lal Shimpi on January 28, 2014 6:35 PM EST

Posted in CPUs | IT Computing | Enterprise | enterprise CPUs | AMD | Opteron | Opter

"SEATTLE" 64-BIT ARM SERVER PROCESSOR

FIRST 28NM ARM SERVER CPU TO SAMPLE IN MARCH

Industry's only 64-bit ARM Server SoC from a proven server processor supplier

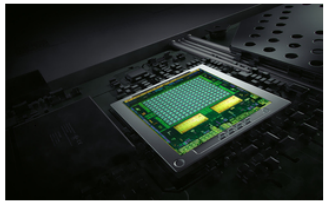
- The most server experience of any ARM licensee

Nvidia Jetson TK1 mini supercomputer is up for pr

Will ship on 15 May

By Lee Bell

Fri May 02 2014, 11:38



NVIDIA'S JETSON TK1 mi development kit is now up f priced at $192.

Despite Nvidia having anno that it is "now shipping", the that is powered by a Tegra actually ship until 15 May.

Claiming to be "the world's supercomputer", the Jetson embedded systems to aid th computers attempting to simulate human recognition of physical objects, suc self-driving cars.

Speaking at the GPU Technology Conference (GTC) in March, Nvidia co-foun Hsun Huang described it as capable of running anything the Geforce GTX Tita can run, but at a slower pace.

With a total performance of 326 GFLOPS, the Jetson TK1 should be more powerful than the Raspberry Pi board, which delivers just 24 GFLOPS, but will retail for much more, costing $192 in the US - a number that matches the number of cores in the Tegra K1 processor that Nvidia launched at CES in Las Vegas in January.

"The Jetson TK1 also comes with this new SDK called Vision Works. Stacked onto CUDA, it comes with a whole bunch of primitives whether it's recognising corners or detecting edges, or it could be classifying objects.

Parameters are loaded into this Vision Works primitives syste recognises objects," Huang said on stage during the Jetson T

A-SERIES REDEFINES COMPUTE

Kaveri

4 "Steamroller" CPU Cores | Multimedia | AMD TrueAudio technology | UVD | VCE | Connectivity | PCIe Gen 3 | PCIe Gen 2 | Display | DisplayPort 1.2 | HDMI 1.4a | DVI

MarketWatch

PRESS RELEASE

Altera and IBM Unveil FPGA-accelerated POWER Systems with Coherent Shared Memory

By

Published: Nov 17, 2014 8:00 a.m. ET

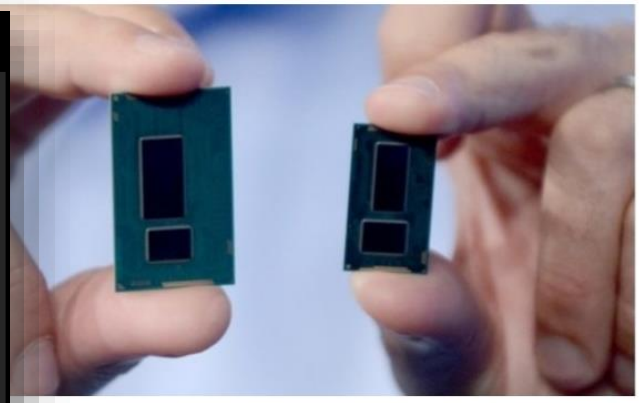
POWER8 Systems that Leverage Reprogrammable FPGA Accelerators Gain Significant Improvements in System Performance, Efficiency and Flexibility

NEW ORLEANS, Nov. 17, 2014 /PRNewswire/ -- **SuperComputing 2014** -- Altera Corporation ALTR, +0.00% and IBM IBM, +0.00% today unveiled the industry's first FPGA-based acceleration platform that coherently connects an FPGA to a POWER8 CPU leveraging IBM's Coherent Accelerator Processor Interface (CAPI). The reconfigurable hardware accelerator features shared virtual memory between the FPGA and processor which significantly improves system performance, efficiency and flexibility in high-performance computing (HPC) and data center applications. Altera and IBM are presenting several POWER8 systems that are coherently accelerated using FPGAs at SuperComputing 2014.

Working together through the OpenPOWER Foundation, Altera and IBM are

Intel's 14nm Broadwell GPU takes shape, indicates major improvements over Haswell

By Sebastian Anthony on November 5, 2013 at 10:21 am | 16 Comments



This Article

Ahead of its 2014 launch, Intel has started open-sourcing the Linux driver for Broadwell's GPU. Broadwell Intel's while the CPU cted to l's GPU looks ris) GPU odified U, and the

Avago Agrees to Buy Broadcom for $37 Billion

By MICHAEL J. de la MERCED and CHAD BRAY MAY 28, 2015

Intel to acquire Altera for $54 a share

Monday, 1 Jun 2015 | 8:33 AM ET

REUTERS

Intel Mates FPGA With Future Xeon Server Chip

June 18, 2014 by Timothy Prickett Morgan

intel inside

Intel is taking field programmable gate arrays seriously as a means of accelerating applications and has crafted a hybrid chip that marries an FPGA to a Xeon E5 processor and puts them in the same processor socket.

OAK RIDGE National Laboratory

Current ASCR Computing At a Glance

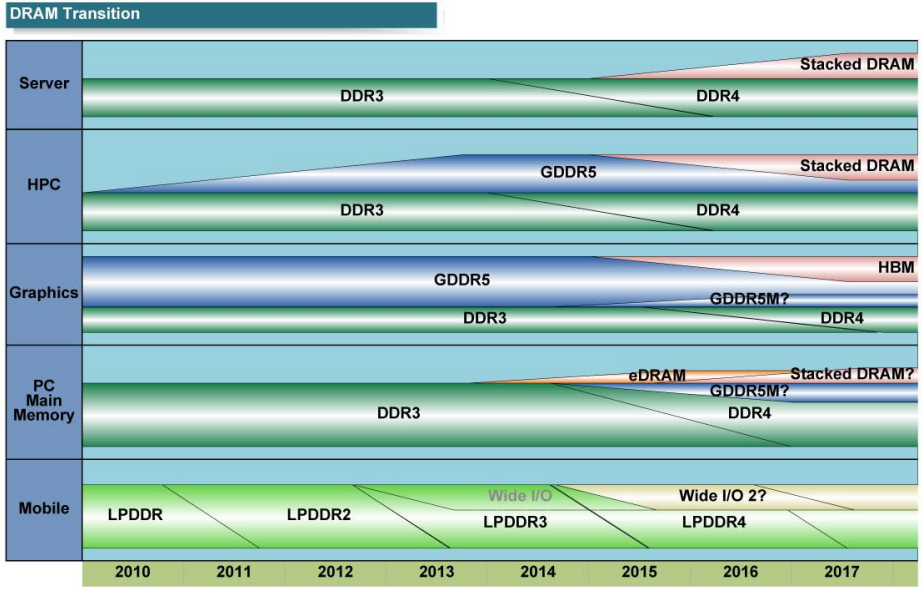
Complexity α T

System attributes	NERSC Now	OLCF Now	ALCF Now	NERSC Upgrade	OLCF Upgrade	ALCF Upgrades	
Planned Installation	**Edison**	**TITAN**	**MIRA**	**Cori 2016**	**Summit 2017-2018**	**Theta 2016**	**Aurora 2018-2019**
System peak (PF)	2.6	27	10	> 30	150	>8.5	180
Peak Power (MW)	2	9	4.8	< 3.7	10	1.7	13
Total system memory	357 TB	710TB	768TB	~1 PB DDR4 + High Bandwidth Memory (HBM)+1.5PB persistent memory	> 1.74 PB DDR4 + HBM + 2.8 PB persistent memory	>480 TB DDR4 + High Bandwidth Memory (HBM)	> 7 PB High Bandwidth On-Package Memory Local Memory and Persistent Memory
Node performance (TF)	0.460	1.452	0.204	> 3	> 40	> 3	> 17 times Mira
Node processors	Intel Ivy Bridge	AMD Opteron Nvidia Kepler	64-bit PowerPC A2	Intel Knights Landing many core CPUs Intel Haswell CPU in data partition	Multiple IBM Power9 CPUs & multiple Nvidia Voltas GPUS	Intel Knights Landing Xeon Phi many core CPUs	Knights Hill Xeon Phi many core CPUs
System size (nodes)	5,600 nodes	18,688 nodes	49,152	9,300 nodes 1,900 nodes in data partition	~3,500 nodes	>2,500 nodes	>50,000 nodes
System Interconnect	Aries	Gemini	5D Torus	Aries	Dual Rail EDR-IB	Aries	2nd Generation Intel Omni-Path Architecture
File System	7.6 PB 168 GB/s, Lustre®	32 PB 1 TB/s, Lustre®	26 PB 300 GB/s GPFS™	28 PB 744 GB/s Lustre®	120 PB 1 TB/s GPFS™	10PB, 210 GB/s Lustre initial	150 PB 1 TB/s Lustre®

OAK RIDGE National Laboratory

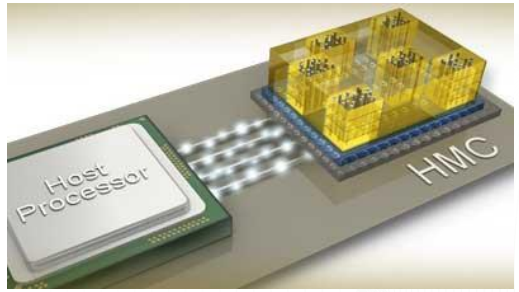
Memory Systems are Diversifying

- HMC, HBM/2/3, LPDDR4, GDDR5X, WIDEIO2, etc
- 2.5D, 3D Stacking
- New devices (ReRAM, PCRAM, STT-MRAM, Xpoint)
- Configuration diversity
  - Fused, shared memory
  - Scratchpads
  - Write through, write back, etc
  - Consistency and coherence protocols
  - Virtual v. Physical, paging strategies



http://gigglehd.com/zbxe/files/attach/images/1404665/988/406/011/788d3ba1967e2db3817d259d2e83c88e_1.jpg

Copyright (c) 2014 Hiroshige Goto All rights reserved.



https://www.micron.com/~/media/track-2-images/content-images/content_image_hmc.jpg?la=en

	SRAM	DRAM	eDRAM	2D NAND Flash	3D NAND Flash	PCRAM	STTRAM	2D ReRAM	3D ReRAM
Data Retention	N	N	N	Y	Y	Y	Y	Y	Y
Cell Size (F^2)	50-200	4-6	19-26	2-5	<1	4-10	8-40	4	<1
Minimum F demonstrated (nm)	14	25	22	16	64	20	28	27	24
Read Time (ns)	< 1	30	5	10^4	10^4	10-50	3-10	10-50	10-50
Write Time (ns)	< 1	50	5	10^5	10^5	100-300	3-10	10-50	10-50
Number of Rewrites	10^{16}	10^{16}	10^{16}	10^4-10^5	10^4-10^5	10^8-10^{10}	10^{15}	10^8-10^{12}	10^8-10^{12}
Read Power	Low	Low	Low	High	High	Low	Medium	Medium	Medium
Write Power	Low	Low	Low	High	High	High	Medium	Medium	Medium
Power (other than R/W)	Leakage	Refresh	Refresh	None	None	None	None	Sneak	Sneak
Maturity									

J.S. Vetter and S. Mittal, "Opportunities for Nonvolatile Memory Systems in Extreme-Scale High Performance Computing," CiSE, *17(2):73-82, 2015.*

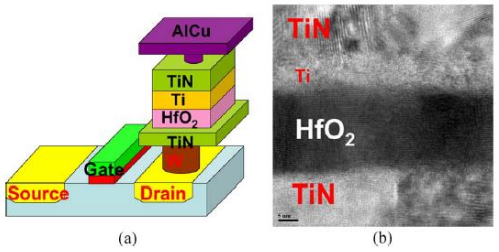
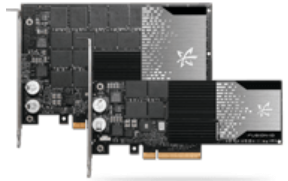


Fig. 4. *(a) A typical 1T1R structure of RRAM with HfO_x; (b) HR-TEM image of the TiN/Ti/HfO_x/TiN stacked layer; the thickness of the HfO_2 is 20 nm.*

H.S.P. Wong, H.Y. Lee, S. Yu *et al.*, *"Metal-oxide RRAM," Proceedings of the IEEE, 100(6):1951-70, 2012.*

NVRAM Technology Continues to Improve – Driven by Market Forces



designlines MEMORY

Blog

First Look at Samsung's 48L 3D V-NAND Flash

Kevin Gibb, Product Line Manager, TechInsights
4/6/2016 04:40 PM EDT
9 comments post a comment

NO RATINGS
LOGIN TO RATE

Like 16 Tweet Share 61 G+1 2

The highly anticipated Samsung's 48 layer V-NAND 3D flash memory is out in the market, and we at TechInsights have the first look.

Samsung had announced its 256 Gb 3-bit multi-level cell K9AFGY8S0M 3D V-NAND as early as August 2015, stating that it would be used ir
on the market in
them in their 2 T
Figure 1.

designlines WIRELESS & NETWORKING

Slideshow

Facebook Likes Intel's 3D XPoint

Google joins open hardware effort

Rick Merritt
3/10/2016 07:56 AM EST
7 comments

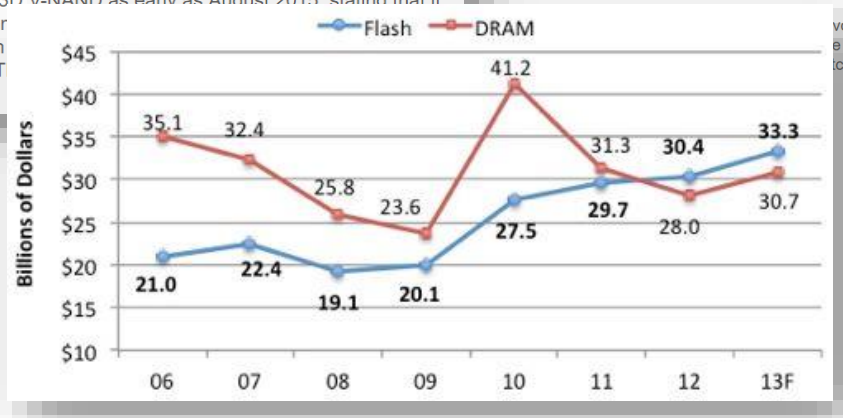
NO RATINGS
LOGIN TO RATE

Like 115 Tweet Share 46 G+1 3

SAN JOSE, Calif.—Facebook said it hopes to use Intel's emerg
3D XPoint memories in its data centers. Meanwhile Google join
its archrival's open hardware efforts to drive standards ranging
high-power compute racks to giant form factors for disk drives.

The two moves were likely the highest impact announcements
the annual event of the Facebook-led Open Compute Project
(OCP) here. Among other news, Intel showed a new 16-core X
SoC with dual 10G Ethernet controllers and a prototype chip
merging Xeon with an Arria FPGA in a single package.

volatile memories be
e endorsement," said
tcher Insight64

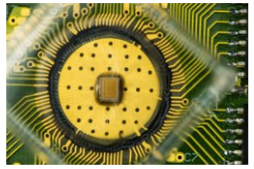


http://www.eetasia.com/STATIC/ARTICLE_IMAGES/201212/EEOL_2012DEC28_STOR_MFG_NT_01.jpg

May 18, 2016

IBM Puts 3D XPoint on Notice with 3 Bits/Cell PCM Breakthrough

Tiffany Trader



IBM scientists have broken new ground in the development of a pha
change memory technology (PCM) that puts a target on competing
XPoint technology from Intel and Micron. IBM successfully stored 3
per cell in a 64k-cell array that had been pre-cycled 1 million times
exposed to temperatures up to 75∘C. A paper describing the advan
was presented this week at the IEEE International Memory Worksh
Paris.

Phase-change memory is an up-and-coming non-volatile memory technology — a storage-class memory that bridges the divide betw
expensive performant, volatile memory (namely DRAM), and slower persistent storage (flash or hard disk drives
According to IBM, having the ability to reliably fit 3 bits per cell is what will make this technology price-competitive
flash.

With memory demands riding the tide of big data, phase change memory has a lot to recommend it but to be a
success, the economics must work, say the authors, and being able to store multiple bits per memory cell is es
for keeping costs under control.

Using a combination of electrical sensing techniques and signal processing technologies, the researchers have
shown for the first time the the viability of Triple-Level-Cell (TLC) storage in phase-change memory cells. The
researchers addressed challenges related to multi-bit PCM including drift, variability, temperature sensitivity and
endurance cycling with two innovative enabling technologies:

*(a) an advanced, nonresistance cell-state metric that exhibits robustness to drift and PCM noise, and (b) an ada
level-detection and modulation-coding framework that enables further resilience to drift, noise and temperature
variation effects.*

Original URL: http://www.theregister.co.uk/2013/11/01/hp_memristor_2018/

HP 100TB Memristor drives by 2018 – if you're lucky, admits tech titan

Universal memory slow in coming

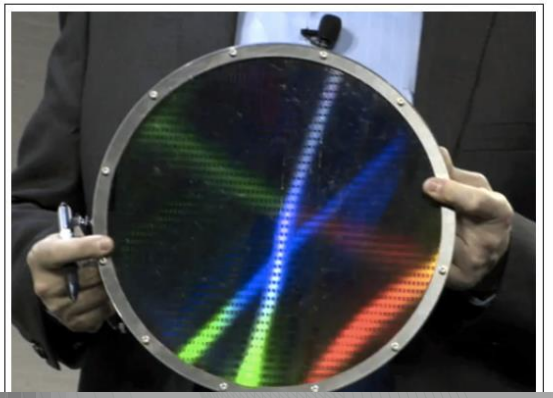
By Chris Mellor

Posted in Storage, 1st November 2013 02:28 GMT

Blocks and Files HP has warned *El Reg* not to get its hopes up too high after the tech titan's CTO Martin Fink suggested StoreServ arrays could be packed with 100TB Memristor drives come 2018.

In five years, according to Fink, DRAM and NAND scaling will hit a wall, limiting the maximum capacity of the technologies: process shrinks will come to a shuddering halt when the memories' reliability drops off a cliff as a side effect of reducing the size of electronics on the silicon dies.

The HP answer to this scaling wall is Memristor, its flavour of resistive RAM technology that is supposed to have DRAM-like speed and better-than-NAND storage density. Fink claimed at an HP Discover event in Las Vegas that Memristor devices will be ready by the time flash NAND hits its limit in five years. He also showed off a Memristor wafer, adding that it could have a 1.5PB capacity by the end of the decade.



Forbes / Tech

JUL 28, 2015 @ 2:46 PM 7,391 VIEWS

Intel And Micron Jointly Announce Game-Changing 3D XPoint Memory Technology

OAK RIDGE National Laboratory

Comparison of Emerging Memory Technologies

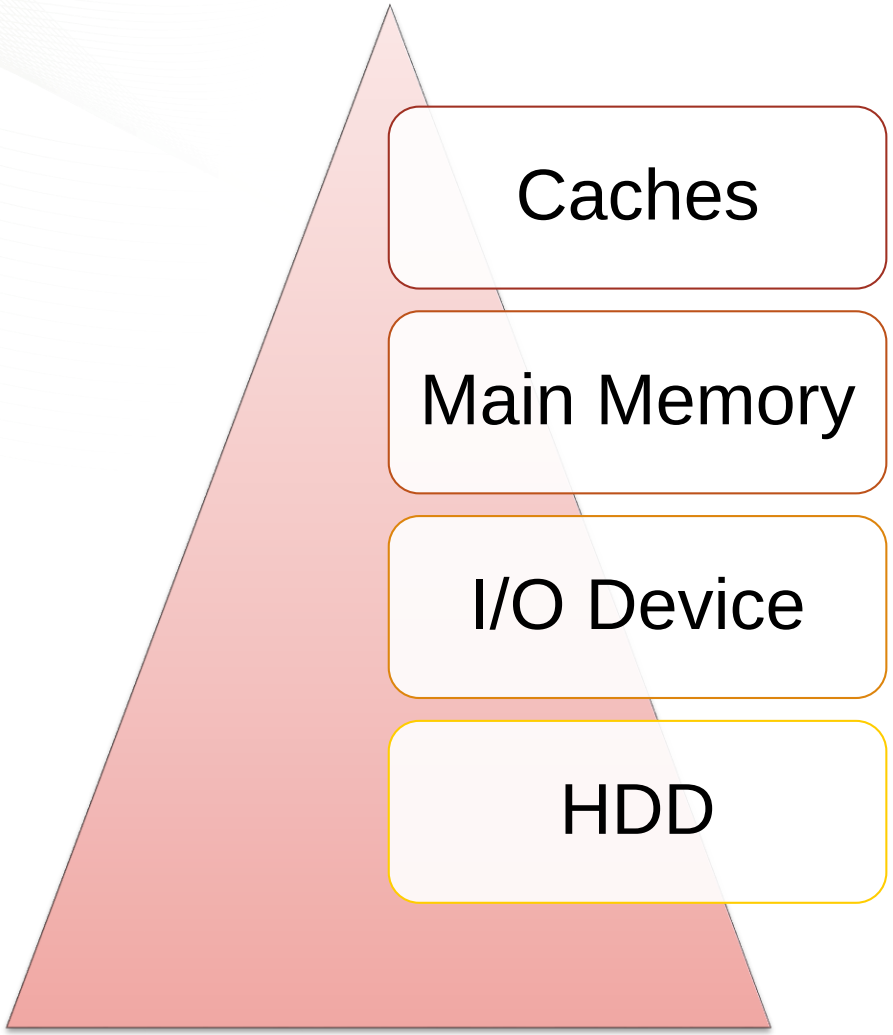
	SRAM	DRAM	eDRAM	2D NAND Flash	3D NAND Flash	PCRAM	STTRAM	2D ReRAM	3D ReRAM
Data Retention	N	N	N	Y	Y	Y	Y	Y	Y
Cell Size (F^2)	50-200	4-6	19-26	2-5	<1	4-10	8-40	4	<1
Minimum F demonstrated (nm)	14	25	22	16	64	20	28	27	24
Read Time (ns)	< 1	30	5	10^4	10^4	10-50	3-10	10-50	10-50
Write Time (ns)	< 1	50	5	10^5	10^5	100-300	3-10	10-50	10-50
Number of Rewrites	10^{16}	10^{16}	10^{16}	10^4-10^5	10^4-10^5	10^8-10^{10}	10^{15}	10^8-10^{12}	10^8-10^{12}
Read Power	Low	Low	Low	High	High	Low	Medium	Medium	Medium
Write Power	Low	Low	Low	High	High	High	Medium	Medium	Medium
Power (other than R/W)	Leakage	Refresh	Refresh	None	None	None	None	Sneak	Sneak
Maturity									

Intel/Micron Xpoint?

http://ft.ornl.gov/trac/blackcomb

OAK RIDGE National Laboratory

As NVM improves, it is working its way toward the processor core



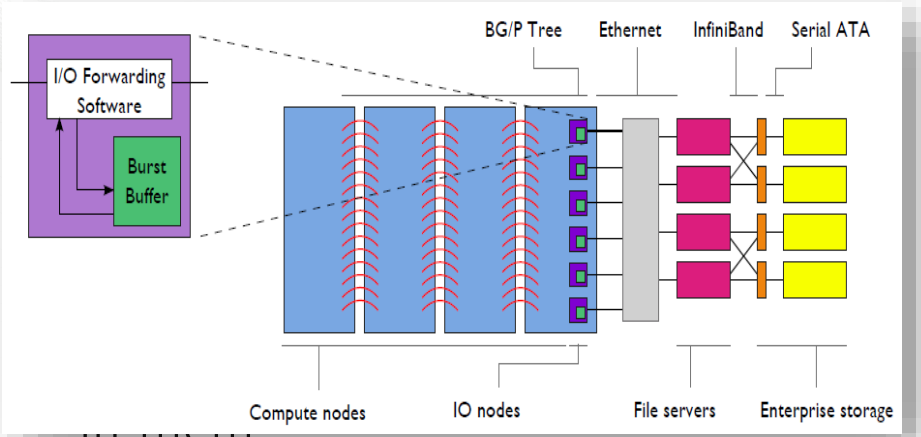
- Newer technologies improve
  - density,
  - power usage,
  - durability
  - r/w performance
- In scalable systems, a variety of architectures exist
  - NVM in the SAN
  - NVM nodes in system
  - NVM in each node

OAK RIDGE National Laboratory

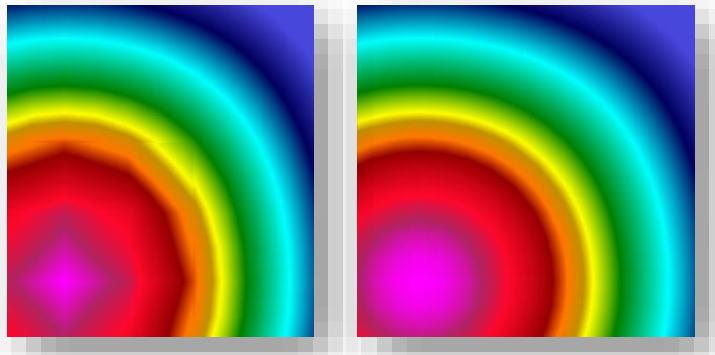
Opportunities for NVM in Emerging Systems

- Burst Buffers, C/R

[Liu, et al., MSST 2012]



- In situ visualization



http://ft.ornl.gov/eavl

tables

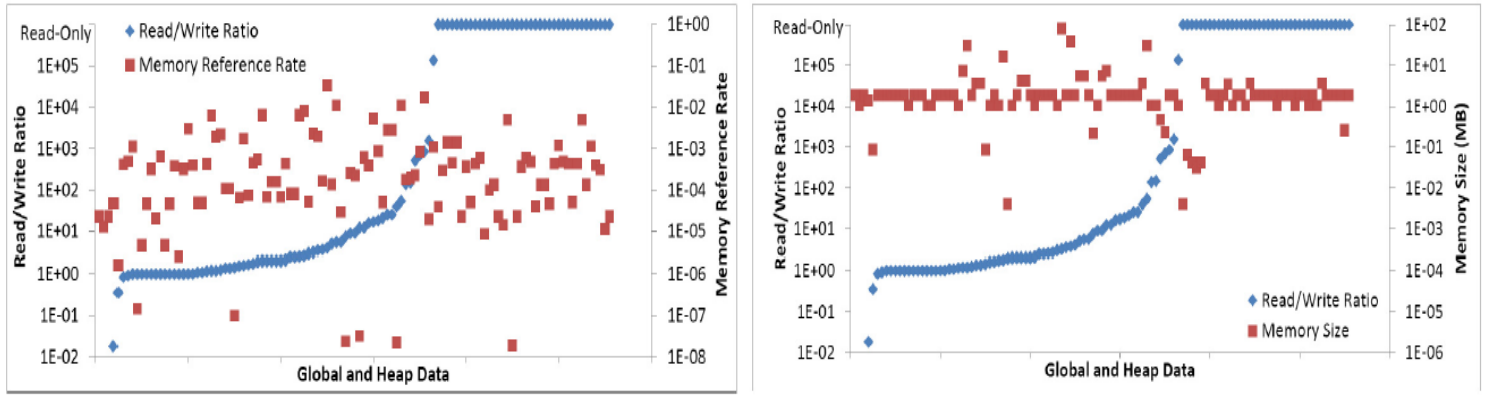
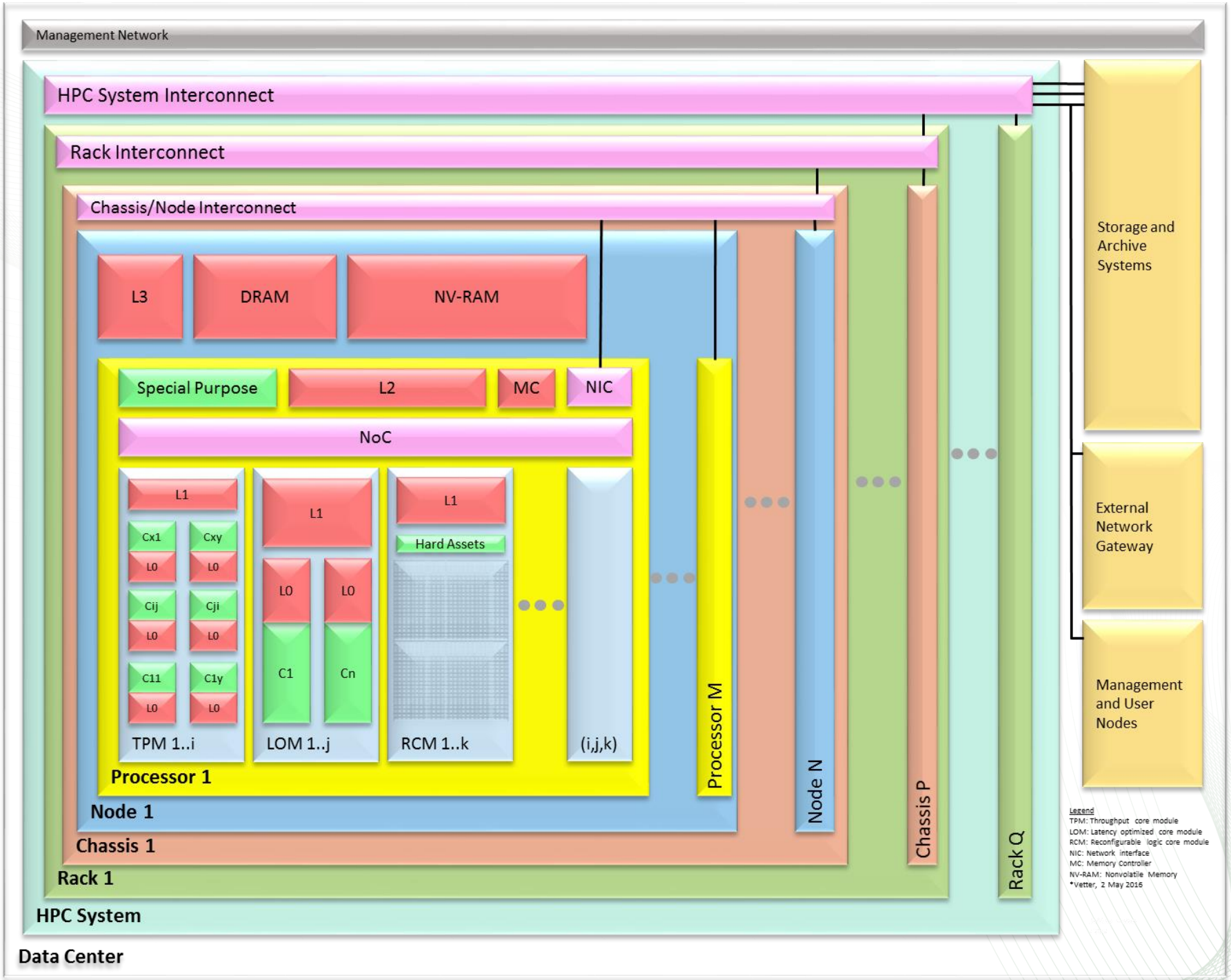


Figure 3: Read/write ratios, memory reference rates and memory object sizes for memory objects in Nek5000

J.S. Vetter and S. Mittal, "Opportunities for Nonvolatile Memory Systems in Extreme-Scale High-Performance Computing," *Computing in Science & Engineering, 17(2):73-82, 2015, 10.1109/MCSE.2015.4.*

OAK RIDGE National Laboratory



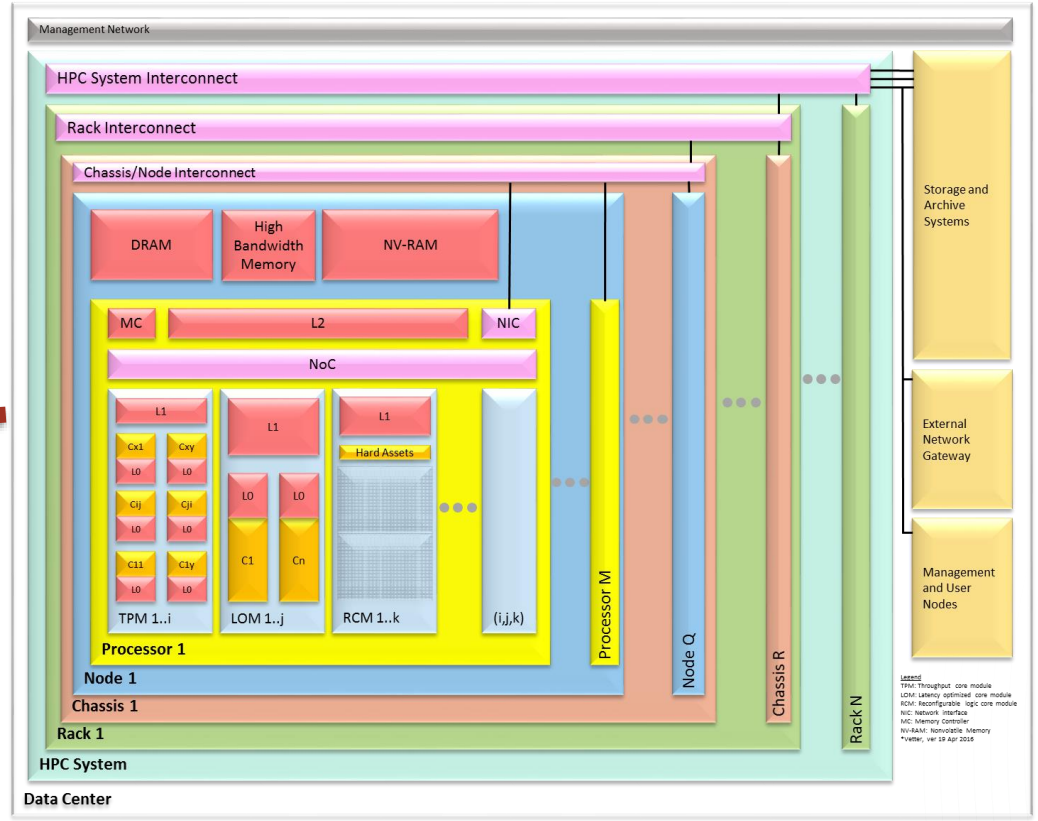
OAK RIDGE National Laboratory

Complexity is the next major challenge!

- "Exciting" times in computer architecture
  - Heterogeneous cores
  - Multimode memory systems
  - Fused memory systems
  - I/O architectures
  - Error correction
  - Changing system balance
- Uncertainty, Ambiguity
  - How do we design future systems so that they are faster than current systems on mission applications?
    - Entirely possible that the new system will be slower than the old system!
  - How do we provide some level of performance portability for applications teams?
  - How do we understand reliability and performance problems?
- Managing complexity is our main challenge!

OAK RIDGE National Laboratory

…Yields Complex Programming Models



This approach is not scalable, affordable, robust, elegant, etc.

System: MPI, Legion, HPX, Charm++, etc

- Low overhead
- Resource contention
- Locality

Node: OpenMP, Pthreads, U-threads, etc

- SIMD
- NUMA, HBM

Cores: OpenACC, CUDA, OpenCL, OpenMP4, …

- Memory use, coalescing
- Data orchestration
- Fine grained parallelism
- Hardware features

OAK RIDGE National Laboratory

Performance Portability : what is it?

- "I know it when I see it"
- Performance portability is not a new topic
  - Kuck, 1996
- For two decades, expectations were set by '(Curse of) Moore's Law' with exception for MPI
  - Recompile and relink
- As discussed, becoming difficult to hide complexity for even functional portability

- "write once, run anywhere efficiently"
- Efficiently use resource of interest

		Application				
		A	B	C	D	
Architecture	W					?
	X					?
	Y					?
	Z					?
		?	?	?	?	

D.J. Kuck, *High performance computing: challenges for future systems. New York: Oxford University Press, 1996.*

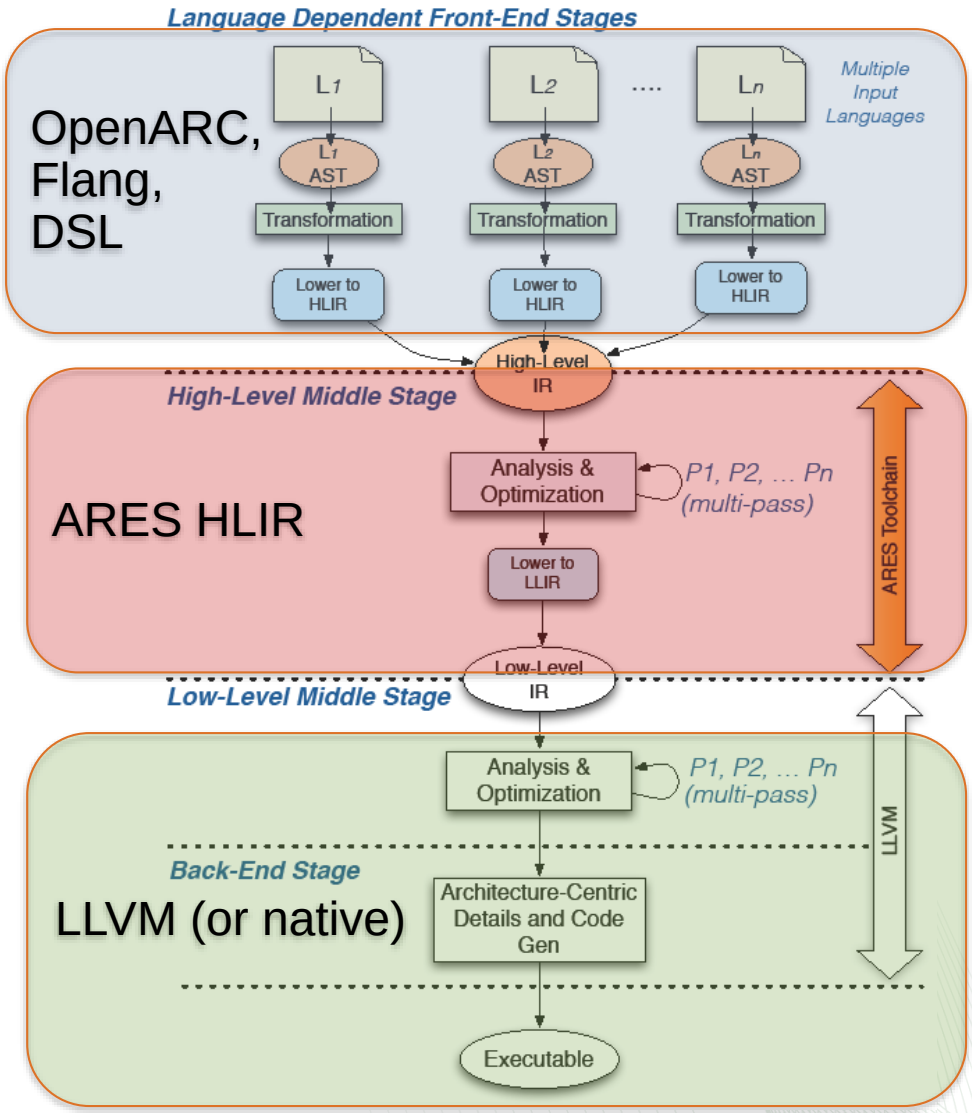
OAK RIDGE National Laboratory

Programming Heterogeneous Systems

OAK RIDGE
National Laboratory

HLIR Approach

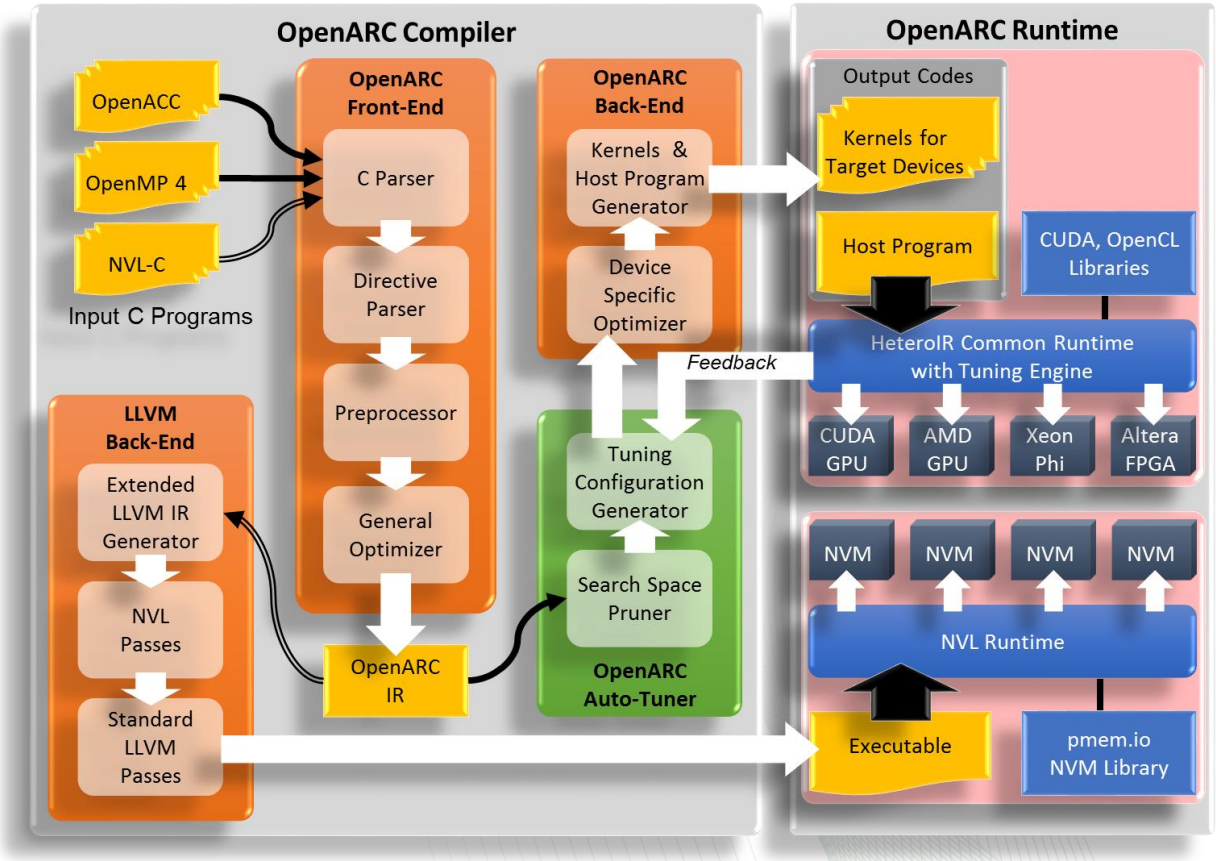
- Vertically integrated toolchain for programming systems
  - Not trying to build a complete toolchain, but rather leverage other software
- Define an open-source, extensible, universal High-Level Intermediate Representation (HLIR) leveraging the widely adopted LLVM infrastructure
- HLIR Analysis and optimization passes can be applied to any Frontend
- HLIR enables higher level analysis and transformation than low level IRs
- Lowered to LLVM or native support (e.g., CUDA)



OAK RIDGE National Laboratory

OpenARC: Open Accelerator Research Compiler

- Open-Sourced, High-Level Intermediate Representation (HIR)-Based, Extensible Compiler Framework.
  - Perform source-to-source translation from OpenACC C to target accelerator models.
    - Support full features of OpenACC V1.0 (+ array reductions and function calls)
    - Support both CUDA and OpenCL as target accelerator models
  - Provide common runtime APIs for various back-ends
  - Can be used as a research framework for various study on directive-based accelerator computing.
    - Built on top of Cetus compiler framework, equipped with various advanced analysis/transformation passes and built-in tuning tools.
    - OpenARC's IR provides an AST-like syntactic view of the source program, easy to understand, access, and transform the input program.



OAK RIDGE National Laboratory

HLIR-based, Directive-agnostic Framework provides portability

HLIR Concepts Encapsulating Existing Directive-based Accelerator Programming Models (OpenMP and OpenACC)

Features		OpenACC 2.0	OpenMP 4.0	HLIR Concepts
Data Parallelism	Coarse-grain parallelism	Yes	Yes	Coarse-parallel
	Fine-grain parallelism	Yes	Yes	Fine-parallel
	Vectorization	Yes	Yes	Vector
Task Parallelism		No	Yes	Task
Memory Management	Structured data region	Yes	Yes	Data-region
	Unstructured data region	Yes	No	Enter/exit data
	Memory transfer	Yes (rich)	Yes	Copyin, copyout,…
	Flush	No	Yes	Flush
Execution control	Synchronization	Yes (limited)	Yes	Barrier
	Asynchronous operation	Yes	Yes (indirect)	Async
	Critical section	No	Yes	Critical section
	Atomic operation	Yes	Yes	Atomic
Execution Configuration	Device type	Yes	No	Device_type
	Device number	Yes	Yes	Device_number

OAK RIDGE National Laboratory

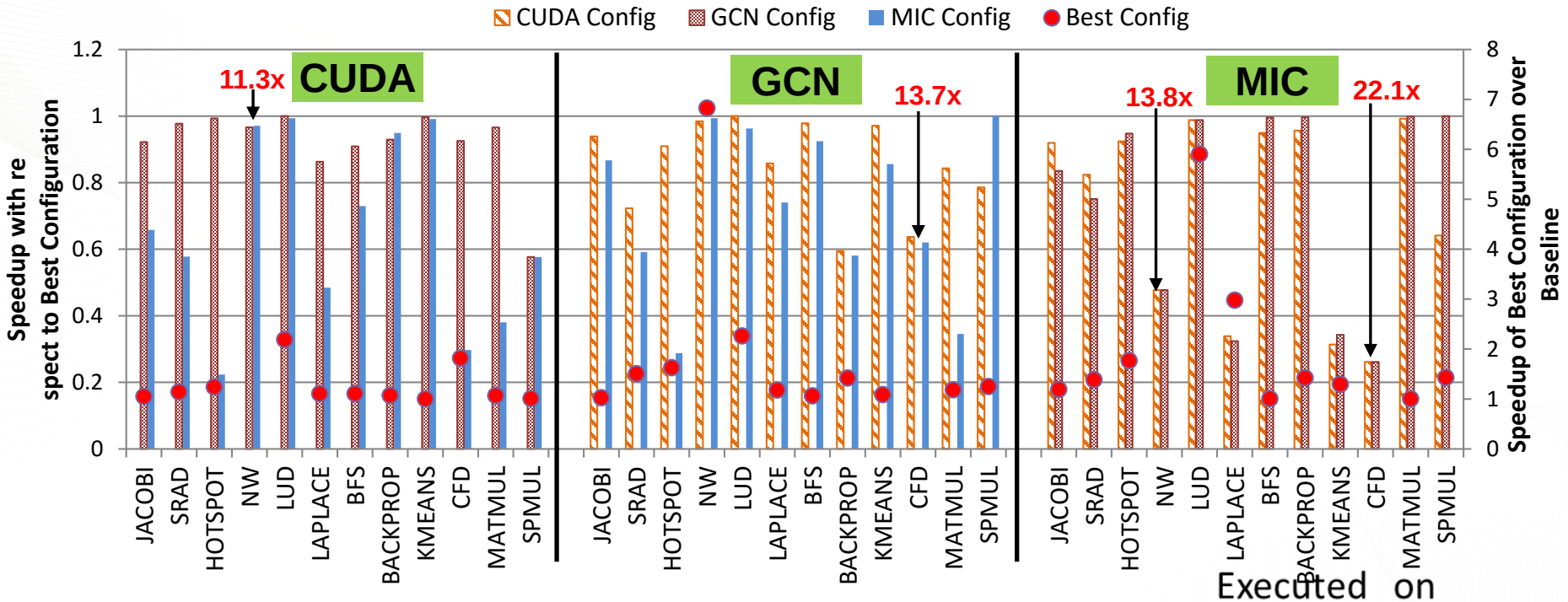
Understanding Performance Portability of High-level Programming Models for Heterogeneous Systems

- Problem
  - Directive-based, high-level accelerator programming models such as OpenACC provide code portability.
    - How does it fare on performance portability?
    - And what architectural features/compiler optimizations affect the performance portability? And how much?
- Solution
  - Proposed a high-level, architecture-independent intermediate language (HeteroIR) to map high-level programming models (e.g., OpenACC) to diverse heterogeneous devices while maintaining portability.
  - Using HeteroIR, port and measure the performance portability of various OpenACC applications on diverse architectures.
- Results
  - Using HeteroIR, OpenARC ported 12 OpenACC applications to diverse architectures (NVIDIA CUDA, AMD GCN, and Intel MIC), and measured the performance portability achieved across all applications.
  - HeteroIR abstracts out the common architecture functionalities, which makes it easy for OpenARC (and other compilers) to support diverse heterogeneous architectures.
  - HeteroIR, combined with rich OpenARC directives and built-in tuning tools, allows OpenARC to be used for various tuning studies on diverse architectures.

Best Program version of	Executed on: CUDA	GCN	MIC
CUDA	100	84	65
GCN	91	100	67
MIC	58	68	100

Amit Sabne, Putt Sakdhnagool, Seyong Lee, and Jeffrey S. Vetter. Understanding Portability of a High-level Programming Model on Contemporary Heterogeneous Architectures, IEEE Micro Volume 35, Issue 4 (DOI: 10.1109/MM.2015.73), 2015.

Overall Performance Portability



- Better perf. portability among GPUs
- Lesser across GPUs and MIC
- Main reasons
  - Parallelism arrangement
  - Compiler optimizations : e.g. device-specific memories, unrolling etc.

Executed on

Best Program version of	CUDA	GCN	MIC
CUDA	100	84	65
GCN	91	100	67
MIC	58	68	100

Performance Portability Matrix

Amit Sabne, Putt Sakdhnagool, Seyong Lee, and Jeffrey S. Vetter. Understanding Portability of a High-level Programming Model on Contemporary Heterogeneous Architectures, IEEE Micro Volume 35, Issue 4 (DOI: 10.1109/MM.2015.73), 2015.

OAK RIDGE National Laboratory

Intelligent selection of optimizations based on target architecture

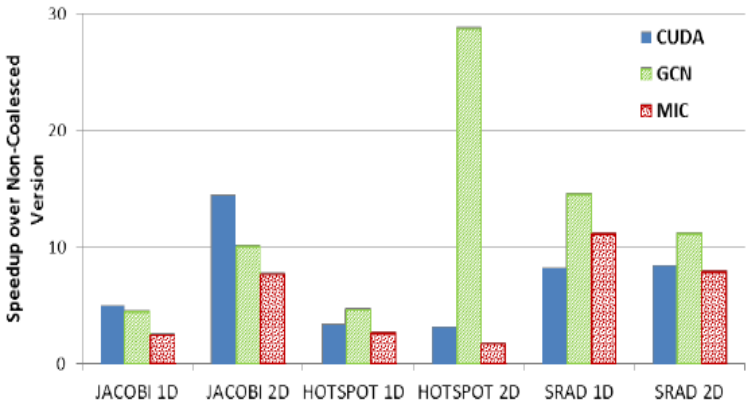


Figure 5: Memory Coalescing Benefits on Different Architectures : MIC is impacted the least by the non-coalesced accesses

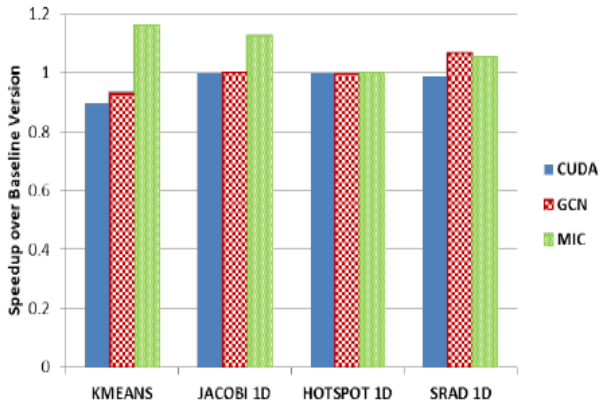


Figure 9: Effects of Loop Unrolling - MIC shows benefits on unrolling

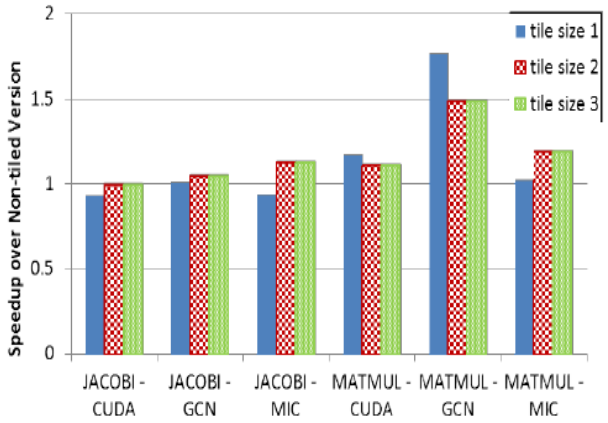


Figure 7: Impact of Tiling Transformation : *MATMUL* shows higher benefits than *JACOBI* owing to more contiguous accesses

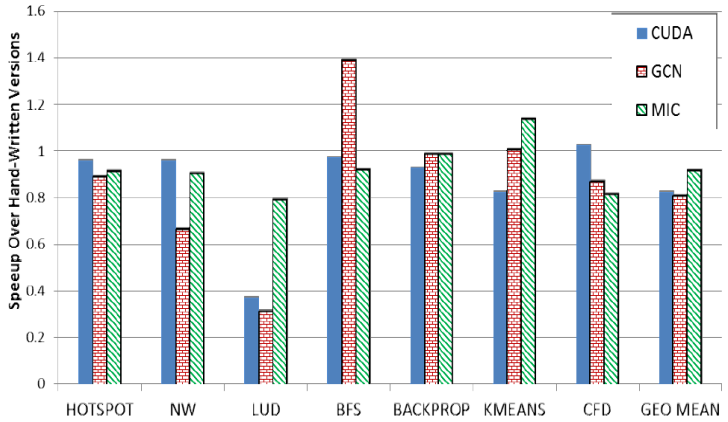
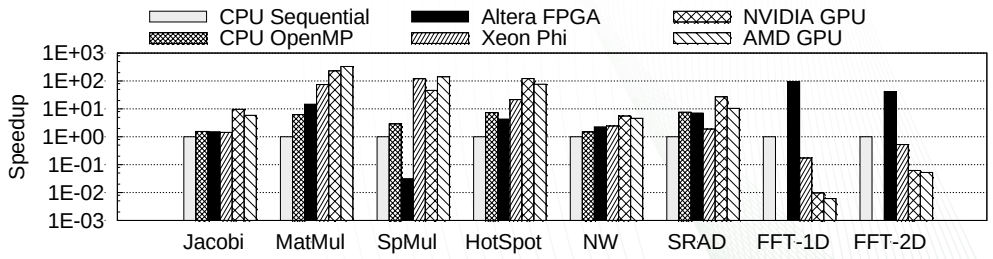


Fig. 11: Comparison of hand-written CUDA/OpenCL programs against auto-tuned OpenARC code versions : Tuned OpenACC programs perform reasonably well against hand-written codes

OAK RIDGE National Laboratory

OpenACC to FPGA: A Framework for Directive-Based High-Performance Reconfigurable Computing

- Problem
  - Reconfigurable computers, such as FPGAs, offer more performance and energy efficiency for specific workloads than other heterogeneous systems, but their programming complexities and low portability have limited their deployment in large scale HPC systems.
- Solution
  - Proposed an OpenACC-to-FPGA translation framework, which performs source-to-source translation of the input OpenACC program into an output OpenCL code, which is further compiled to an FPGA program by the underlying backend Altera OpenCL compiler.
- Recent Results
  - Proposed several FPGA-specific OpenACC compiler optimizations and pragma extensions to achieve higher throughput.
  - Evaluated the framework using eight OpenACC benchmarks, and measured performance variations on diverse architectures (Altera FPGA, NVIDIA/AMD GPUs, and Intel Xeon Phi).
- Impact
  - Proposed translation framework is the first work to use a standard and portable, directive-based, high-level programming system for FPGAs.
  - Preliminary evaluation of eight OpenACC benchmarks on an FPGA and comparison study on other accelerators identified that the unique capabilities of an FPGA offer new performance tuning opportunities different from other accelerators.



S. Lee, J. Kim, and J.S. Vetter, "OpenACC to FPGA: A Framework for Directive-based High-Performance Reconfigurable Computing," Proc. IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2016.

OAK RIDGE National Laboratory

Reconfigurable Computing Tests Performance Portability in a New Dimension

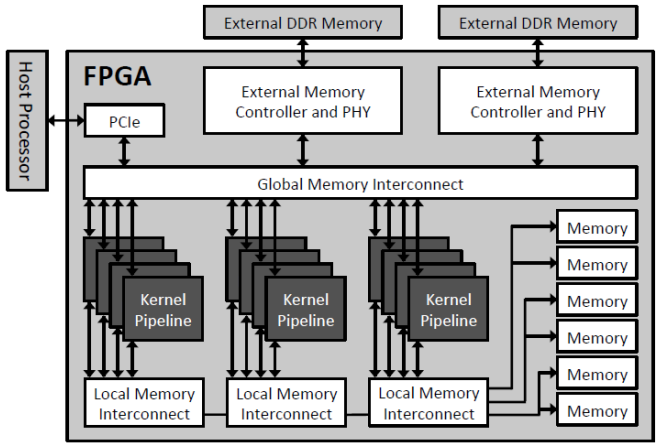
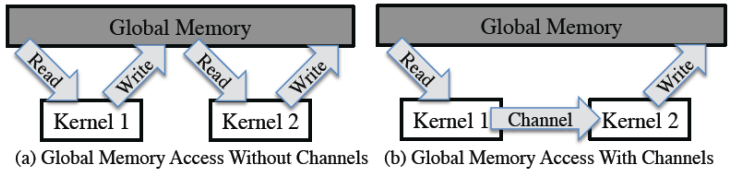


Figure 2: FPGA OpenCL Architecture



(a) Global Memory Access Without Channels (b) Global Memory Access With Channels

Figure 3: Difference in Global Memory Access Pattern as a Result of Channels Implementation

Listing 4: Altera OpenCL (AOCL) Channel Example

```
#pragma acc data copyout(a[0:N]) create(b[0:N]) \\
copyin(c[0:N])
  {
#pragma acc kernels loop gang worker present(b, c)
    for (i=0; i<N; i++) b[i] = c[i]*c[i];
#pragma acc kernels loop gang worker present(a, b)
    for (i=0; i<N; i++) a[i] = b[i];
  }
        (a) Input OpenACC code

#pragma acc data copyout(a[0:N]) pipe(b[0:N]) \\
copyin(c[0:N])
  {
#pragma acc kernels loop gang worker pipeout(b) present(c)
    for (i=0; i<N; i++) b[i] = c[i]*c[i];
#pragma acc kernels loop gang worker pipein(b) present(a)
    for (i=0; i<N; i++) a[i] = b[i];
  }
        (b) Modified OpenACC code for kernel-pipelining

#pragma OPENCL EXTENSION cl_altera_channels : enable
channel float pipe__b;
__kernel void kernel0(__global float * c)
{
  int i = get_global_id(0);
  write_channel_altera(pipe__b, (c[i]*c[i]));
}
__kernel void kernel1(__global float * a)
{
  int i = get_global_id(0);
  a[i] = read_channel_altera(pipe__b);
}
        (c) Output OpenCL code with channels
```

S. Lee, J. Kim, and J.S. Vetter, "OpenACC to FPGA: A Framework for Directive-based High-Performance Reconfigurable Computing," Proc. IEEE International Parallel & Distributed Processing Symposium (IPDPS), 2016. (to appear)

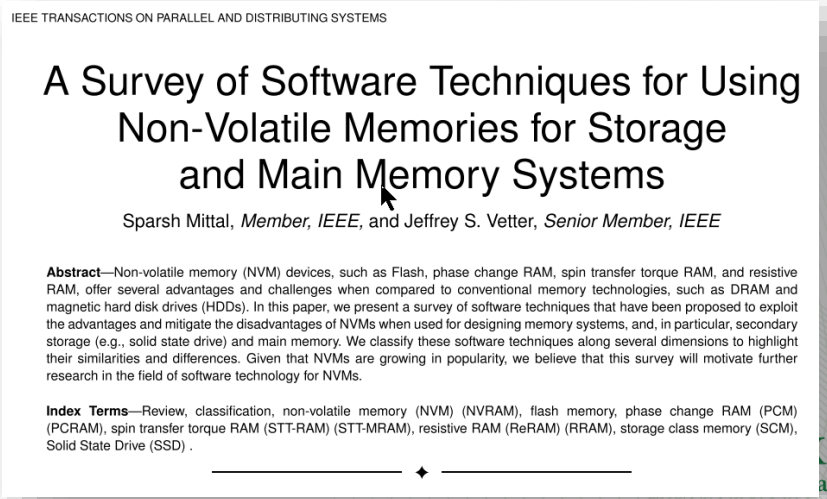
OAK RIDGE National Laboratory

Programming NVM Systems

NVM Programming Systems : Goals

- Architectures will vary dramatically
  - How should we design the node?
  - Portable across various NVM architectures
  - MPI and OpenMP do not solve this problem.
- Two modes of operation
  - Drop in replacement for DRAM
  - Exploit persistence
- Active area of research
  - http://j.mp/nvm-sw-survey
- Performance for HPC scenarios
  - Allow user or compiler/runtime/os to exploit NVM
  - Asymmetric R/W
  - Remote/Local
- Assume lower power costs under normal usage
- Security

- Correctness and durability
  - A crash or erroneous program could corrupt the NVM data structures
  - Programming system needs to provide support for this model
  - Enhanced ECC for NVM devices
- ACID
  - Atomicity: A transaction is “all or nothing”
  - Consistency: Takes data from one consistent state to another
  - Isolation: Concurrent transactions appears to be one after another
  - Durability: Changes to data will remain across system boots



IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTING SYSTEMS

A Survey of Software Techniques for Using Non-Volatile Memories for Storage and Main Memory Systems

Sparsh Mittal, *Member, IEEE,* and Jeffrey S. Vetter, *Senior Member, IEEE*

Abstract—Non-volatile memory (NVM) devices, such as Flash, phase change RAM, spin transfer torque RAM, and resistive RAM, offer several advantages and challenges when compared to conventional memory technologies, such as DRAM and magnetic hard disk drives (HDDs). In this paper, we present a survey of software techniques that have been proposed to exploit the advantages and mitigate the disadvantages of NVMs when used for designing memory systems, and, in particular, secondary storage (e.g., solid state drive) and main memory. We classify these software techniques along several dimensions to highlight their similarities and differences. Given that NVMs are growing in popularity, we believe that this survey will motivate further research in the field of software technology for NVMs.

Index Terms—Review, classification, non-volatile memory (NVM) (NVRAM), flash memory, phase change RAM (PCM) (PCRAM), spin transfer torque RAM (STT-RAM) (STT-MRAM), resistive RAM (ReRAM) (RRAM), storage class memory (SCM), Solid State Drive (SSD) .

NVL-C: Portable Programming for NVMM

- Minimal, familiar, programming interface:
  - Minimal C language extensions.
  - App can still use DRAM.
- Pointer safety:
  - Persistence creates new categories of pointer bugs.
  - Best to enforce pointer safety constraints at compile time rather than run time.
- Transactions:
  - Prevent corruption of persistent memory in case of application or system failure.
- Language extensions enable:
  - Compile-time safety constraints.
  - NVM-related compiler analyses and optimizations.
- LLVM-based:
  - Core of compiler can be reused for other front ends and languages.
  - Can take advantage of LLVM ecosystem.

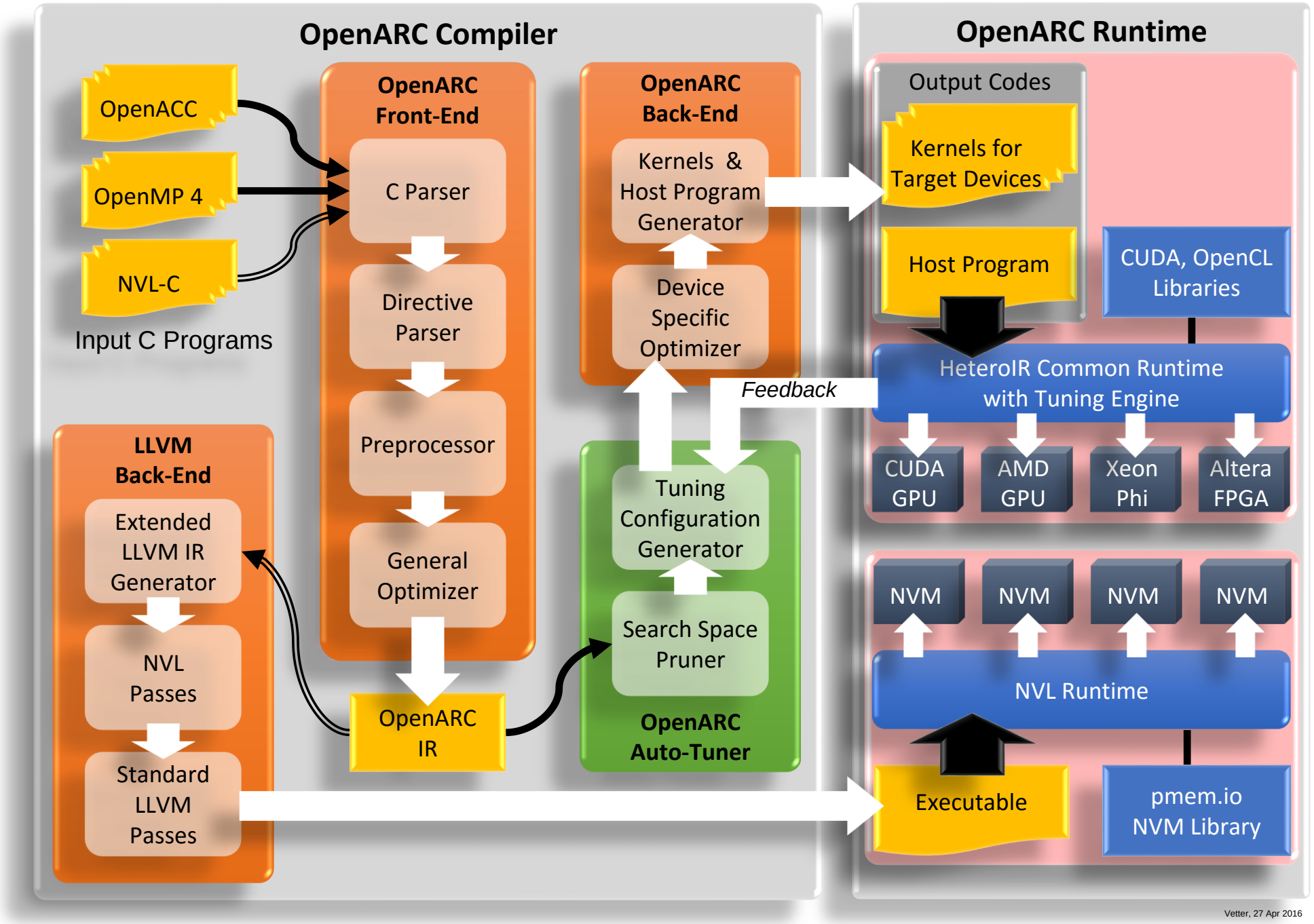
```
#include <nvl.h>
struct list {
  int value;
  nvl struct list *next;
};
void remove(int k) {
  nvl_heap_t *heap
    = nvl_open("foo.nvl");
  nvl struct list *a
    = nvl_get_root(heap, struct list);
  #pragma nvl atomic
  while (a->next != NULL) {
    if (a->next->value == k)
      a->next = a->next->next;
    else
      a = a->next;
  }
  nvl_close(heap);
}
```

Pointer Class	Permitted
NV-to-V	no
V-to-NV	yes
intra-heap NV-to-NV	yes
inter-heap NV-to-NV	no

Table 1: Pointer Classes

J. Denny, S. Lee, and J.S. Vetter, "NVL-C: Static Analysis Techniques for Efficient, Correct Programming of Non-Volatile Main Memory Systems," in *ACM High Performance Distributed Computing (HPDC). Kyoto: ACM, 2016*

OAK RIDGE National Laboratory



Vetter, 27 Apr 2016

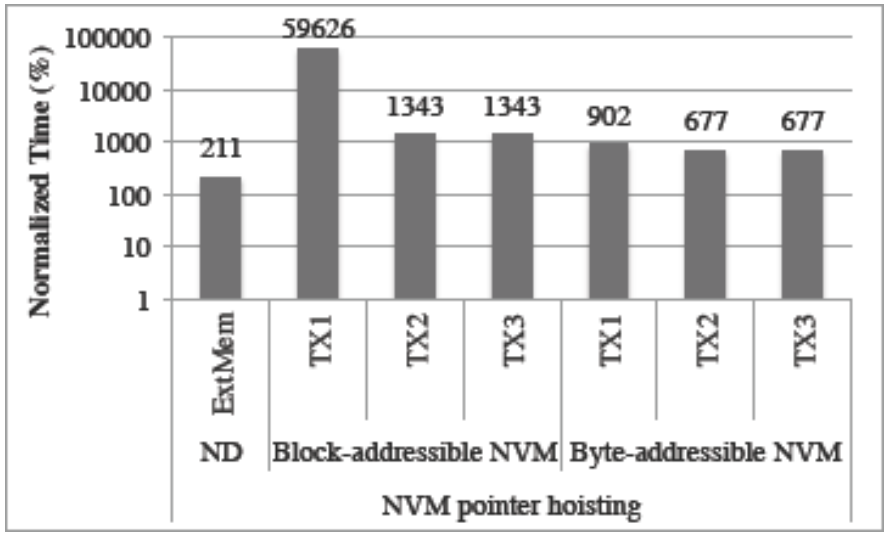
Preliminary Results

- Applications extended with NVL-C
- Compiled with NVL-C
- Executed on Fusion ioScale
- Compared to DRAM
- Various levels of optimization

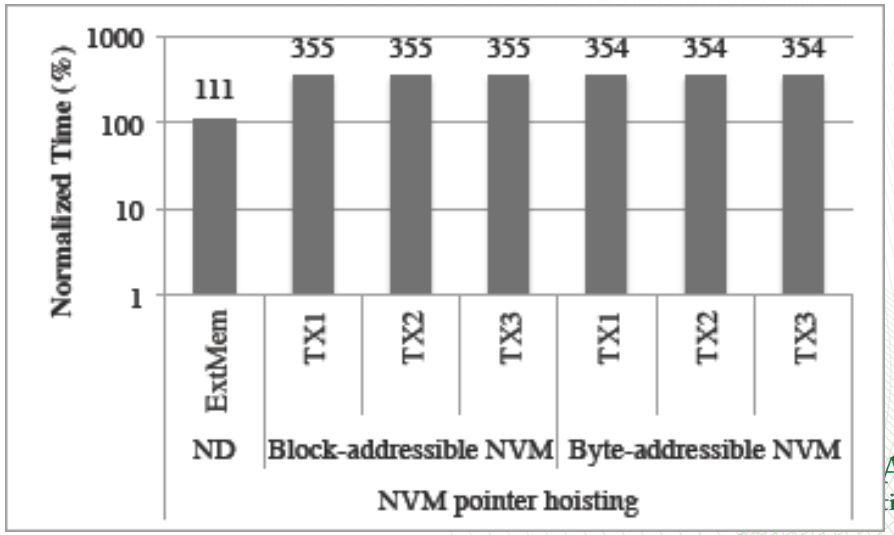
Table 3: Symbols Used in the Result Figures

Symbol	Description
ExtMem or ExM	Use persistent storage as if extended DRAM
No Durability or ND	Skip runtime operations for durability
Base or B	Basic NVL-C version w/o Safety, RefCnt, and transaction (TX0, TX1, ...)
Safety or S	Automatic pointer-safety checking
RefCnt or R	Automatic reference counting
TX0	B+S+R + Enforce only durability of each NVM write
TX1	B+S+R + Enforce ACID properties of each transaction
TX2	TX1 + aggregated transaction using `backup` clauses
TX3	TX2 + skipping unnecessary backup using `clobber` clauses
TX4	TX3 at the granularity of each loop
CLFlush	Flush cache line to memory
MSync	Synchronize memory map with persistent storage

LULESH



XSBENCH



AK RIDGE
tional Laboratory

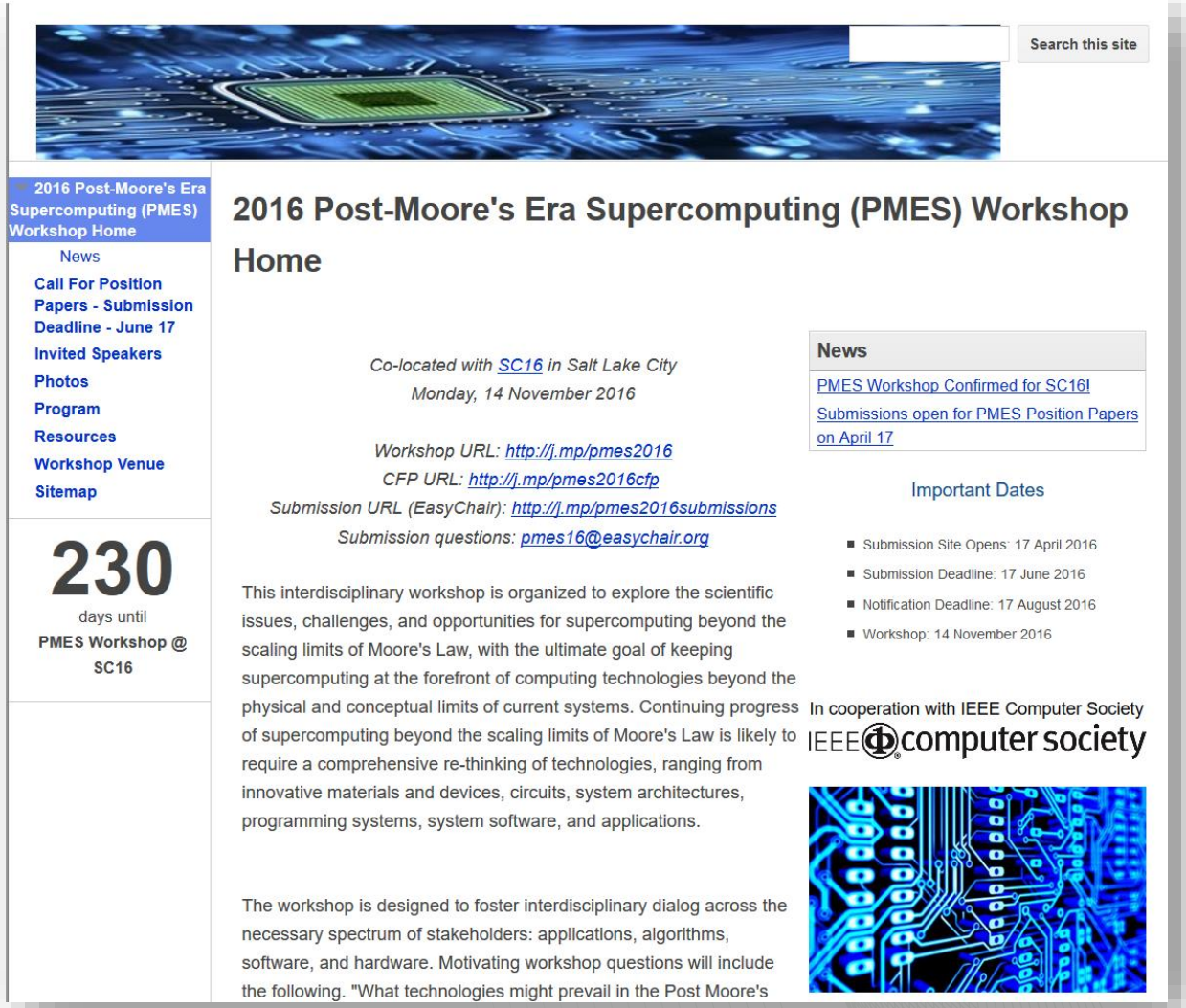
Summary

- Both architectures and applications are growing more complex
  - Trends dictate that this will get worse, not better
  - This complexity creates irregularity in computation, communication, and data movement
- Programming models must provide performance portability in this environment
- Examples
  - Heterogeneous computing
    - OpenARC provides a foundation for ‘normalizing’ hierarchical parallelism
  - Memory hierarchies
    - Experimenting with NVL-C prototype to understand potential

OAK RIDGE National Laboratory

PMES Workshop @ SC16

- https://j.mp/pmes2016
- @SC16
- Position papers due June 17



Search this site

2016 Post-Moore's Era Supercomputing (PMES) Workshop Home

News

Call For Position Papers - Submission Deadline - June 17

Invited Speakers

Photos

Program

Resources

Workshop Venue

Sitemap

230

days until

PMES Workshop @ SC16

2016 Post-Moore's Era Supercomputing (PMES) Workshop Home

Co-located with SC16 in Salt Lake City
Monday, 14 November 2016

Workshop URL: http://j.mp/pmes2016
CFP URL: http://j.mp/pmes2016cfp
Submission URL (EasyChair): http://j.mp/pmes2016submissions
Submission questions: pmes16@easychair.org

This interdisciplinary workshop is organized to explore the scientific issues, challenges, and opportunities for supercomputing beyond the scaling limits of Moore's Law, with the ultimate goal of keeping supercomputing at the forefront of computing technologies beyond the physical and conceptual limits of current systems. Continuing progress of supercomputing beyond the scaling limits of Moore's Law is likely to require a comprehensive re-thinking of technologies, ranging from innovative materials and devices, circuits, system architectures, programming systems, system software, and applications.

The workshop is designed to foster interdisciplinary dialog across the necessary spectrum of stakeholders: applications, algorithms, software, and hardware. Motivating workshop questions will include the following. "What technologies might prevail in the Post Moore's

News

PMES Workshop Confirmed for SC16!

Submissions open for PMES Position Papers on April 17

Important Dates

- Submission Site Opens: 17 April 2016
- Submission Deadline: 17 June 2016
- Notification Deadline: 17 August 2016
- Workshop: 14 November 2016

In cooperation with IEEE Computer Society

IEEE computer society

OAK RIDGE National Laboratory

Acknowledgements

- Contributors and Sponsors
  - Future Technologies Group: http://ft.ornl.gov
  - US Department of Energy Office of Science
    - DOE Vancouver Project: https://ft.ornl.gov/trac/vancouver
    - DOE Blackcomb Project: https://ft.ornl.gov/trac/blackcomb
    - DOE ExMatEx Codesign Center: http://codesign.lanl.gov
    - DOE Cesar Codesign Center: http://cesar.mcs.anl.gov/
    - DOE Exascale Efforts: http://science.energy.gov/ascr/research/computer-science/
  - Scalable Heterogeneous Computing Benchmark team: http://bit.ly/shocmarx
  - US National Science Foundation Keeneland Project: http://keeneland.gatech.edu
  - US DARPA
  - NVIDIA CUDA Center of Excellence

OAK RIDGE National Laboratory

Bonus Material