Oak Ridge National Laboratory

Preparing for Extreme Heterogeneity in High Performance Computing

Jeffrey S. Vetter

With many contributions from FTG Group and Colleagues

18th Annual Workshop on Charm++ and its Applications
October 20, 2020
University of Illinois at Urbana-Champaign (Virtual)

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Future Technologies Group

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

Highlights

Recent trends in computing paint an ambiguous future for architectures

- Contemporary systems provide evidence that power constraints are driving architectures to change rapidly
- Multiple architectural dimensions are being (dramatically) redesigned: Processors, node design, memory systems, I/O

- **Entering an era of Extreme Heterogeneity**
- **Complexity is our main challenge**

Applications and software systems are all reaching a state of crisis

- Applications will not be functionally or performance portable across architectures
- Programming and operating systems need major redesign to address these architectural changes
- Procurements, acceptance testing, and operations of today's new platforms depend on performance prediction and benchmarking.

- **This is a crisis!**

Programming systems must provide performance portability (beyond functional portability)!!

- Strive for 'Write once, run anywhere'
- Descriptive models of parallelism and data movement
- Introspective runtime systems
- Layered, modular, open source approaches required

- **Examples**
  - **ECP investments in LLVM**
    - **FORTRAN with GPU offloading**
  - **Programming FPGAs**
    - **Without Verilog**
  - **Memory systems are changing too**
    - **Language support for NVM**

Time for a short poll…

History (circa 2010)

Q: Think back 10 years. How many of you would have predicted that many of our top HPC systems would be heterogeneous (GPU-based) architectures?

Yes

No

Revisionists ☺

Future (circa 2030)

Q: Think forward 10 years. How many of you predict that our top 100 HPC systems will have the following architectural features?

Assume general purpose multicore CPU

- GPU
- FPGA/Reconfigurable processor
- Neuromorphic processor
- Deep learning processor
- Quantum processor
- RISC-V processor
- Some new unknown processor
- All/some of the above in one SoC

OAK RIDGE
National Laboratory

Implications for Science Applications Teams

Q: Now, imagine you are building a new application with an expected ~3M LOC and 20 team members over the next 10 years. What on-node programming model/system do you use to future-proof your app?

Assume C and C++ (?)

- Fortran XX
- Metaprogramming, DSEL, etc (e.g., AMP, Kokkos, RAJA, SYCL)
- CUDA, cu***, HIP, OpenCL
- Directives: OpenMP, OpenACC
- Python, Julia, Rust, R, Matlab, etc
- Domain Specific Language (e.g., Claw, Hallide, PySL) or Domain Specific Framework (e.g., PetSc, AMReX)
- Legion, Charm++, HPX, etc
- Some new unknown programming approach
- Some combination of the above

Motivating Trends

News & Analysis

Foundries' Sales Show Hard Times Continuing

Peter Clarke
5/23/2016 09:33 P
2 comments

Like 6 4

LONDON--Taiwan
semiconductor se
market slow down

Both companies a
those they achieve
an annual basis in
TSMC and UMC a

partial
ales f
yet ov
signi
vere a
ators
e of s

eetasia.com

SEMICONDUCTOR ENGINEERING

Home > Manufacturing, Design & Test > Uncertainty Grows For 5nm, 3nm

MANUFACTURING, DESIGN & TEST

Uncertainty Grows For 5nm, 3nm

797 74

Nanosheets and nanowire FETs under development, but costs are skyrocketing. New packaging options could provide an alternative.

DECEMBER 19TH, 2016 - BY: MARK LAPEDUS

As several chipmakers ramp up their 10nm finFET processes, with 7nm just around the corner, R&D has begun for 5nm and beyond. In fact, some are already moving full speed ahead in the arena.

TSMC recently announced plans to build a new fab in Taiwan

GlobalFoundries Forfeit 7nm Manufacturing - EE Times Asia

6-7 minu

SAN
the bl

Globa
than
subsi

Samsung to Invest $115 Billion in Foundry & Chip Businesses by 2030

t makers of
ing developers

Intel's 10nm Is Broken, Delayed Until 2019

37 COMMENTS

by Paul Alcorn April 26, 2018 at 6:30 PM

DESIGNLINES | WIRELESS AND NETWORKING DESIGNLINE

GlobalFoundries Selling ASIC Business to Marvell

By Dylan McGrath, 05.20.19 1

Sh

re
its
revi
business to Marvell
ree-phase

Another Step Toward the End of Moore's Law

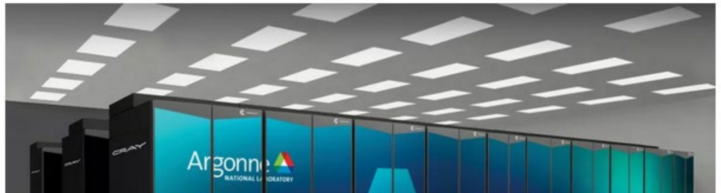
Samsung and TSMC move to

US Government's Aurora Supercomputer Delayed Due to Intel's 7nm Setback

By Anton Shilov 18 days ago

Intel-powered exascale supercomputer will come online later than expected.

Comments (7)



Number of Foundries with a Cutting Edge Logic Fab

SilTerra										
X-FAB										
Dongbu HiTek										
ADI	ADI									
Atmel	Atmel									
Rohm	Rohm									
Sanyo	Sanyo									
Mitsubishi	Mitsubishi									
ON	ON									
Hitachi	Hitachi									
Cypress	Cypress	Cypress								
Sony	Sony	Sony								
Infineon	Infineon	Infineon								
Sharp	Sharp	Sharp								
Freescale	Freescale	Freescale								
Renesas (NEC)	Renesas	Renesas	Renesas	Renesas						
SMIC	SMIC	SMIC	SMIC	SMIC						
Toshiba	Toshiba	Toshiba	Toshiba	Toshiba						
Fujitsu	Fujitsu	Fujitsu	Fujitsu	Fujitsu						
			TI	TI						
			asonic	Panasonic	Panasonic					
			TM	STM	STM					
			MC	UMC	UMC					
			BM	IBM	IBM	IBM				
			oundries	GF	GF	GF	GF			
			nsung	Samsung	Samsung	Samsung	Samsung	Samsung	Samsung	
			SMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	
			ntel	Intel	Intel	Intel	Intel	Intel	Intel	Future
			nm	45 nm/40 nm	32 nm/28 nm	22 nm/20 nm	16 nm/14 nm	10 nm	7 nm	5 nm

13 Dec 2019 | 20:20 GMT

TSMC's 5-Nanometer Process on Track for First Half of 2020

Devices are 15 percent faster, 30 percent more energy efficient

By Samuel K. Moore

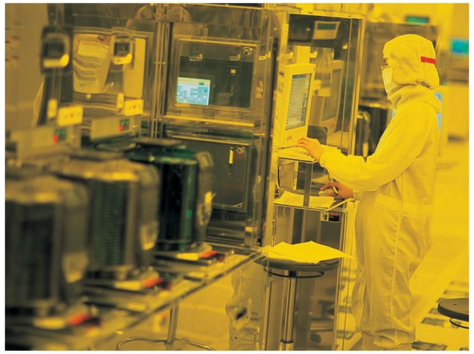
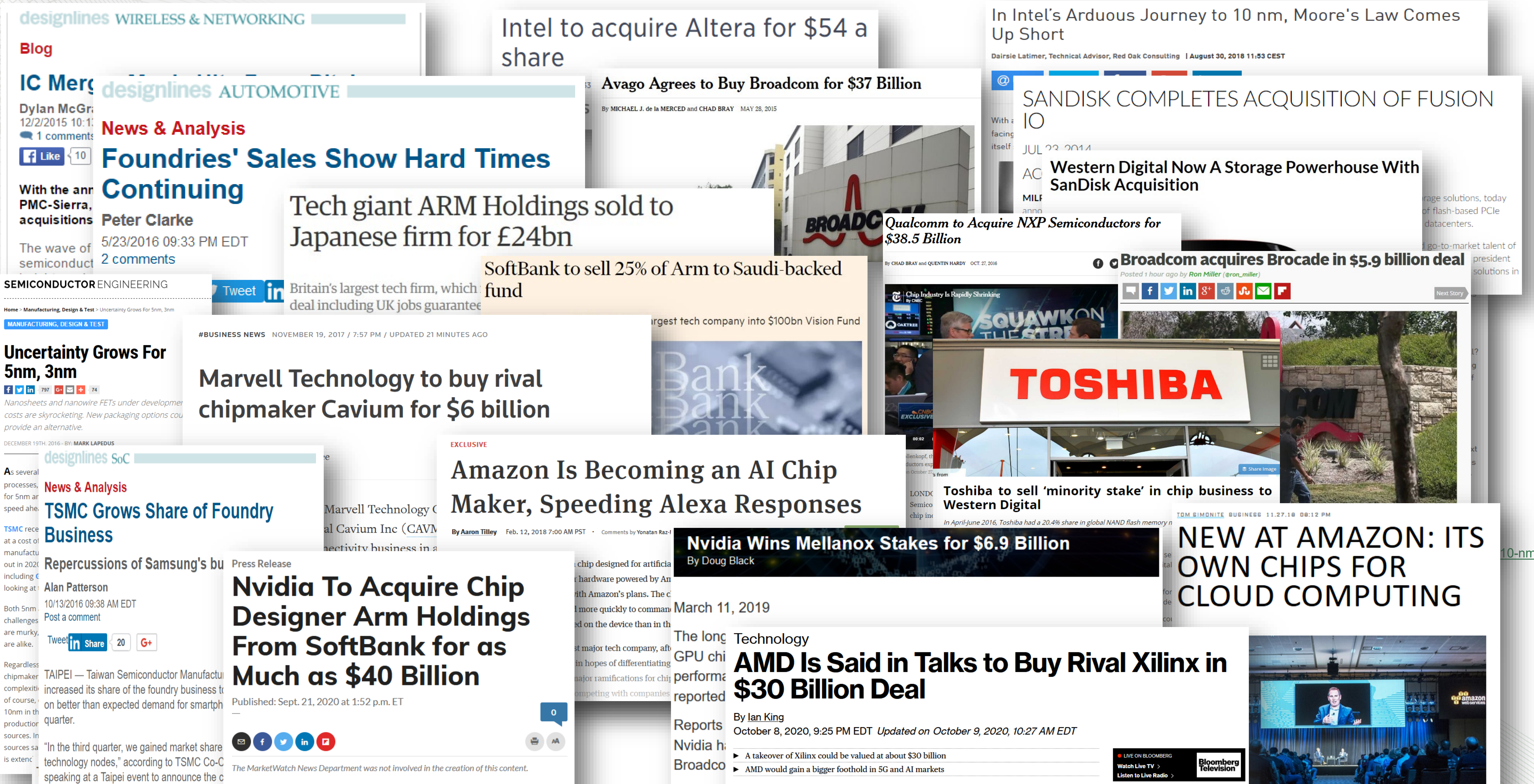


Photo: Taiwan Semiconductor Manufacturing Co.

The performance enhancement achieved by TSMC's new 5-nanometer process is partly due to the inclusion of a "high-mobility channel." How is it created? TSMC wouldn't reveal.

Business climate reflects this uncertainty, cost, complexity, consolidation



designlines WIRELESS & NETWORKING

Blog

IC Merg

Dylan McGra
12/2/2015 10:1
1 comments
Like 10

With the ann
PMC-Sierra,
acquisitions

The wave of
semiconduct

designlines AUTOMOTIVE

News & Analysis

Foundries' Sales Show Hard Times Continuing

Peter Clarke
5/23/2016 09:33 PM EDT
2 comments

Tweet

Intel to acquire Altera for $54 a share

Avago Agrees to Buy Broadcom for $37 Billion

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

In Intel's Arduous Journey to 10 nm, Moore's Law Comes Up Short

Dairsie Latimer, Technical Advisor, Red Oak Consulting | August 30, 2018 11:53 CEST

With a
facing
itself

SANDISK COMPLETES ACQUISITION OF FUSION IO

JUL 23, 2014

AC

MILP
anno

Western Digital Now A Storage Powerhouse With SanDisk Acquisition

rage solutions, today
of flash-based PCIe
datacenters.

go-to-market talent of
president
solutions in

Tech giant ARM Holdings sold to Japanese firm for £24bn

Britain's largest tech firm, which
deal including UK jobs guarantee

SoftBank to sell 25% of Arm to Saudi-backed fund

argest tech company into $100bn Vision Fund

Qualcomm to Acquire NXP Semiconductors for $38.5 Billion

By CHAD BRAY and QUENTIN HARDY OCT. 27, 2016

Chip Industry Is Rapidly Shrinking
By CNBC

llenkopf, th
ductors exp
n October 27

LONDO
Semico
chip inc

Broadcom acquires Brocade in $5.9 billion deal

Posted 1 hour ago by Ron Miller (@ron_miller)

Next Story

SEMICONDUCTOR ENGINEERING

Home > Manufacturing, Design & Test > Uncertainty Grows For 5nm, 3nm

MANUFACTURING, DESIGN & TEST

Uncertainty Grows For 5nm, 3nm

797 74

Nanosheets and nanowire FETs under developmen
costs are skyrocketing. New packaging options cou
provide an alternative.

DECEMBER 19TH, 2016 - BY: MARK LAPEDUS

As several
processes,
for 5nm ar
speed ahe

TSMC rece
at a cost of
manufactu
out in 2020
including G
looking at t

Both 5nm
challenges
are murky,
are alike.

Regardless
chipmaker
complexiti
of course,
10nm in th
production
sources. In
sources sa
is extend

#BUSINESS NEWS NOVEMBER 19, 2017 / 7:57 PM / UPDATED 21 MINUTES AGO

Marvell Technology to buy rival chipmaker Cavium for $6 billion

Marvell Technology C
al Cavium Inc (CAVM
nectivity business in a

designlines SoC

News & Analysis

TSMC Grows Share of Foundry Business

Repercussions of Samsung's bu

Alan Patterson
10/13/2016 09:38 AM EDT
Post a comment

Tweet Share 20

TAIPEI — Taiwan Semiconductor Manufactu
increased its share of the foundry business t
on better than expected demand for smartph
quarter.

"In the third quarter, we gained market share
technology nodes," according to TSMC Co-C
speaking at a Taipei event to announce the c

EXCLUSIVE

Amazon Is Becoming an AI Chip Maker, Speeding Alexa Responses

By Aaron Tilley Feb. 12, 2018 7:00 AM PST · Comments by Yonatan Raz-F

chip designed for artificia
hardware powered by Am
ith Amazon's plans. The c
more quickly to comman
ed on the device than in th

st major tech company, aft
in hopes of differentiating
najor ramifications for chip
ompeting with companies

Press Release

Nvidia To Acquire Chip Designer Arm Holdings From SoftBank for as Much as $40 Billion

Published: Sept. 21, 2020 at 1:52 p.m. ET

0

The MarketWatch News Department was not involved in the creation of this content.

Toshiba to sell 'minority stake' in chip business to Western Digital

In April-June 2016, Toshiba had a 20.4% share in global NAND flash memory m

Nvidia Wins Mellanox Stakes for $6.9 Billion
By Doug Black

March 11, 2019

The long
GPU chi
performa
reported

Reports
Nvidia ha
Broadco

Technology

AMD Is Said in Talks to Buy Rival Xilinx in $30 Billion Deal

By Ian King
October 8, 2020, 9:25 PM EDT *Updated on October 9, 2020, 10:27 AM EDT*

- A takeover of Xilinx could be valued at about $30 billion
- AMD would gain a bigger foothold in 5G and AI markets

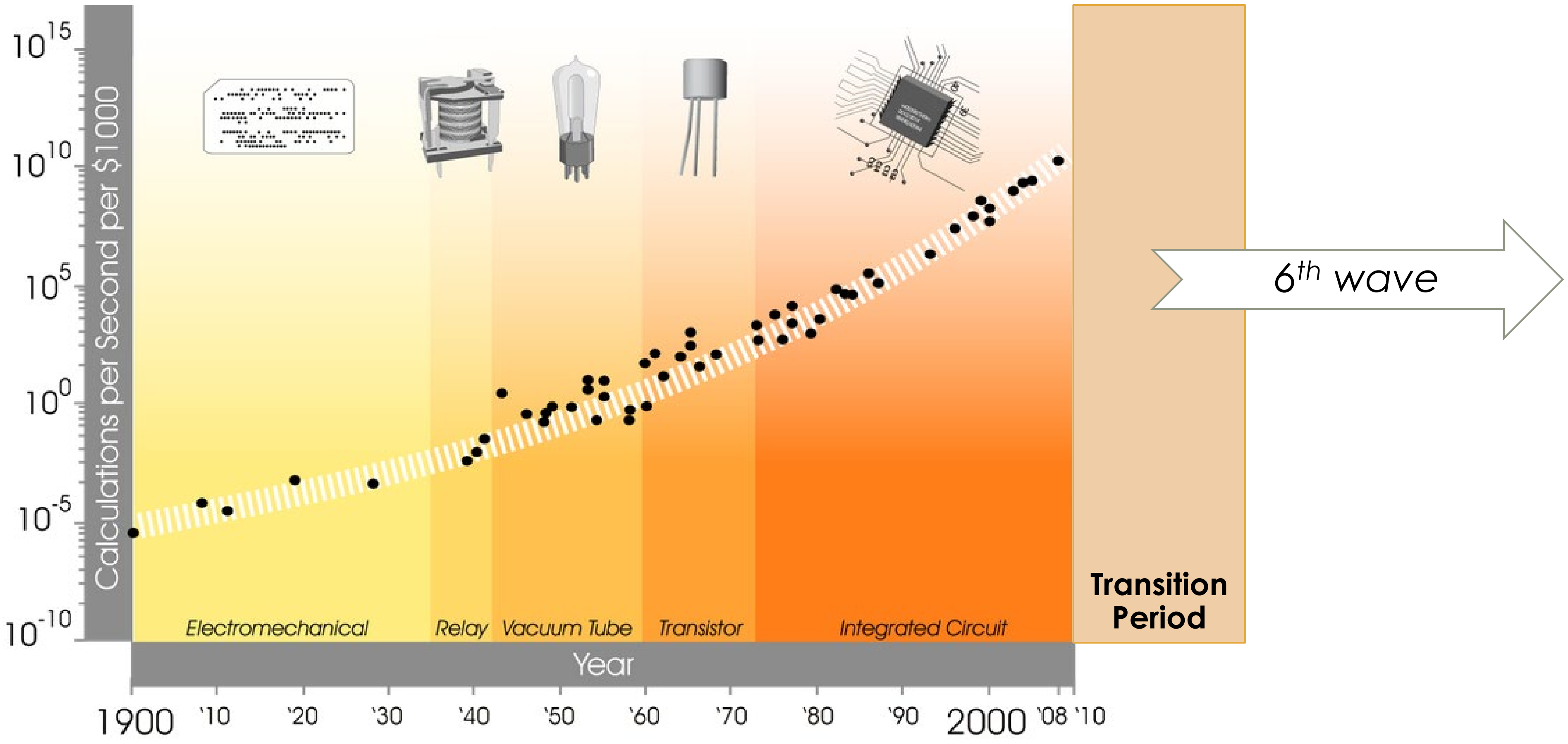
LIVE ON BLOOMBERG
Watch Live TV >
Listen to Live Radio >
Bloomberg Television

TOM SIMONITE BUSINESS 11.27.18 08:12 PM

NEW AT AMAZON: ITS OWN CHIPS FOR CLOUD COMPUTING

10-nm

Sixth Wave of Computing



http://www.kurzweilai.net/exponential-growth-of-computing

Predictions for Transition Period

Optimize Software and Expose New Hierarchical Parallelism

- Redesign software to boost performance on upcoming architectures
- Exploit new levels of parallelism and efficient data movement

Architectural Specialization and Integration

- Use CMOS more effectively for specific workloads
- Integrate components to boost performance and eliminate inefficiencies
- Workload specific memory+storage system design

Emerging Technologies

- Investigate new computational paradigms
  - Quantum
  - Neuromorphic
  - Advanced Digital
  - Emerging Memory Devices

Predictions for Transition Period

Optimize Software and Expose New Hierarchical Parallelism

- Redesign software to boost performance on upcoming architectures
- Exploit new levels of parallelism and efficient data movement

Architectural Specialization and Integration

- Use CMOS more effectively for specific workloads
- Integrate components to boost performance and eliminate inefficiencies
- Workload specific memory+storage system design

Emerging Technologies

- Investigate new computational paradigms
  - Quantum
  - Neuromorphic
  - Advanced Digital
  - Emerging Memory Devices

Predictions for Transition Period

Optimize Software and Expose New Hierarchical Parallelism

- Redesign software to boost performance on upcoming architectures
- Exploit new levels of parallelism and efficient data movement

Architectural Specialization and Integration

- Use CMOS more effectively for specific workloads
- Integrate components to boost performance and eliminate inefficiencies
- Workload specific memory+storage system design

Emerging Technologies

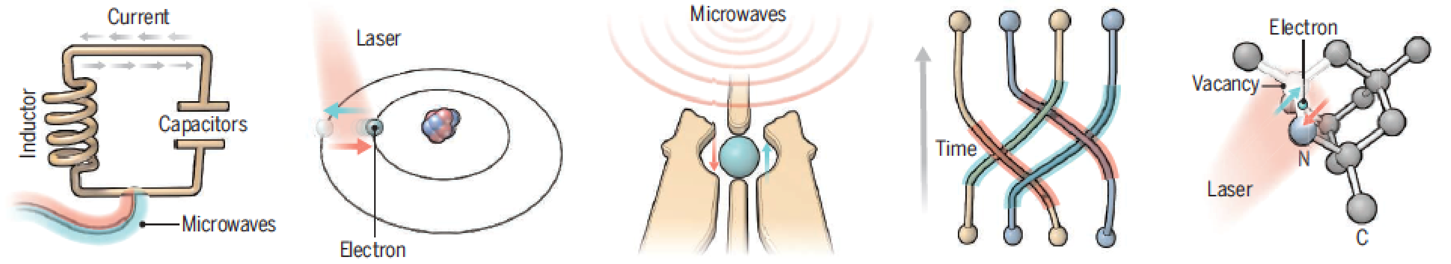
- Investigate new computational paradigms
  - Quantum
  - Neuromorphic
  - Advanced Digital
  - Emerging Memory Devices

Quantum computing: Qubit design and fabrication have made recent progress but still face challenges

Science 354, 1091 (2016) – 2 December

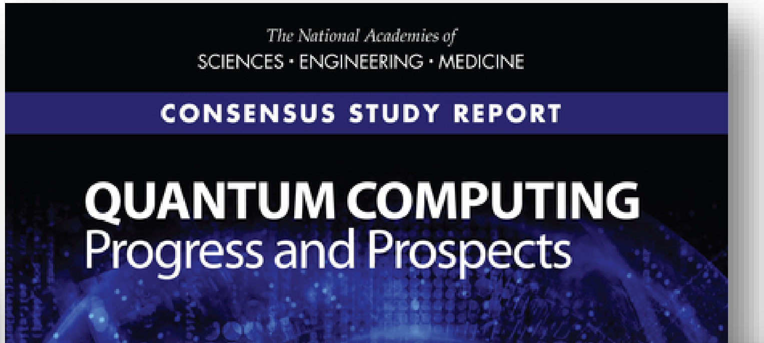
A bit of the action

In the race to build a quantum computer, companies are pursuing many types of quantum bits, or qubits, each with its own strengths and weaknesses.



Superconducting loops	Trapped ions	Silicon quantum dots	Topological qubits	Diamond vacancies
A resistance-free current oscillates back and forth around a circuit loop. An injected microwave signal excites the current into superposition states.	Electrically charged atoms, or ions, have quantum energies that depend on the location of electrons. Tuned lasers cool and trap the ions, and put them in superposition states.	These "artificial atoms" are made by adding an electron to a small piece of pure silicon. Microwaves control the electron's quantum state.	Quasiparticles can be seen in the behavior of electrons channeled through semiconductor structures. Their braided paths can encode quantum information.	A nitrogen atom and a vacancy add an electron to a diamond lattice. Its quantum spin state, along with those of nearby carbon nuclei, can be controlled with light.
Longevity (seconds) 0.00005	>1000	0.03	N/A	10
Logic success rate 99.4%	99.9%	~99%	N/A	99.2%
Number entangled 9	14	2	N/A	6
Company support Google, IBM, Quantum Circuits	ionQ	Intel	Microsoft, Bell Labs	Quantum Diamond Technologies
Pros Fast working. Build on existing semiconductor industry.	Very stable. Highest achieved gate fidelities.	Stable. Build on existing semiconductor industry.	Greatly reduce errors.	Can operate at room temperature.
Cons Collapse easily and must be kept cold.	Slow operation. Many lasers are needed.	Only a few entangled. Must be kept cold.	Existence not yet confirmed.	Difficult to entangle.

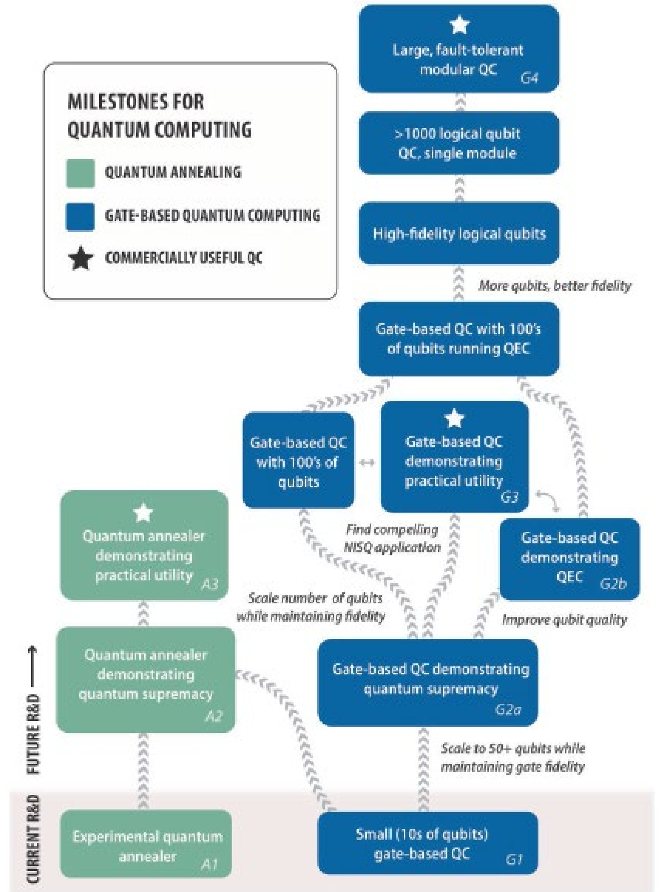
Note: Longevity is the record coherence time for a single qubit superposition state, logic success rate is the highest reported gate fidelity for logic operations on two qubits, and number entangled is the maximum number of qubits entangled and capable of performing two-qubit operations.



The National Academies of SCIENCES · ENGINEERING · MEDICINE

CONSENSUS STUDY REPORT

QUANTUM COMPUTING
Progress and Prospects



MILESTONES FOR QUANTUM COMPUTING
🟩 QUANTUM ANNEALING
🟦 GATE-BASED QUANTUM COMPUTING
★ COMMERCIALLY USEFUL QC

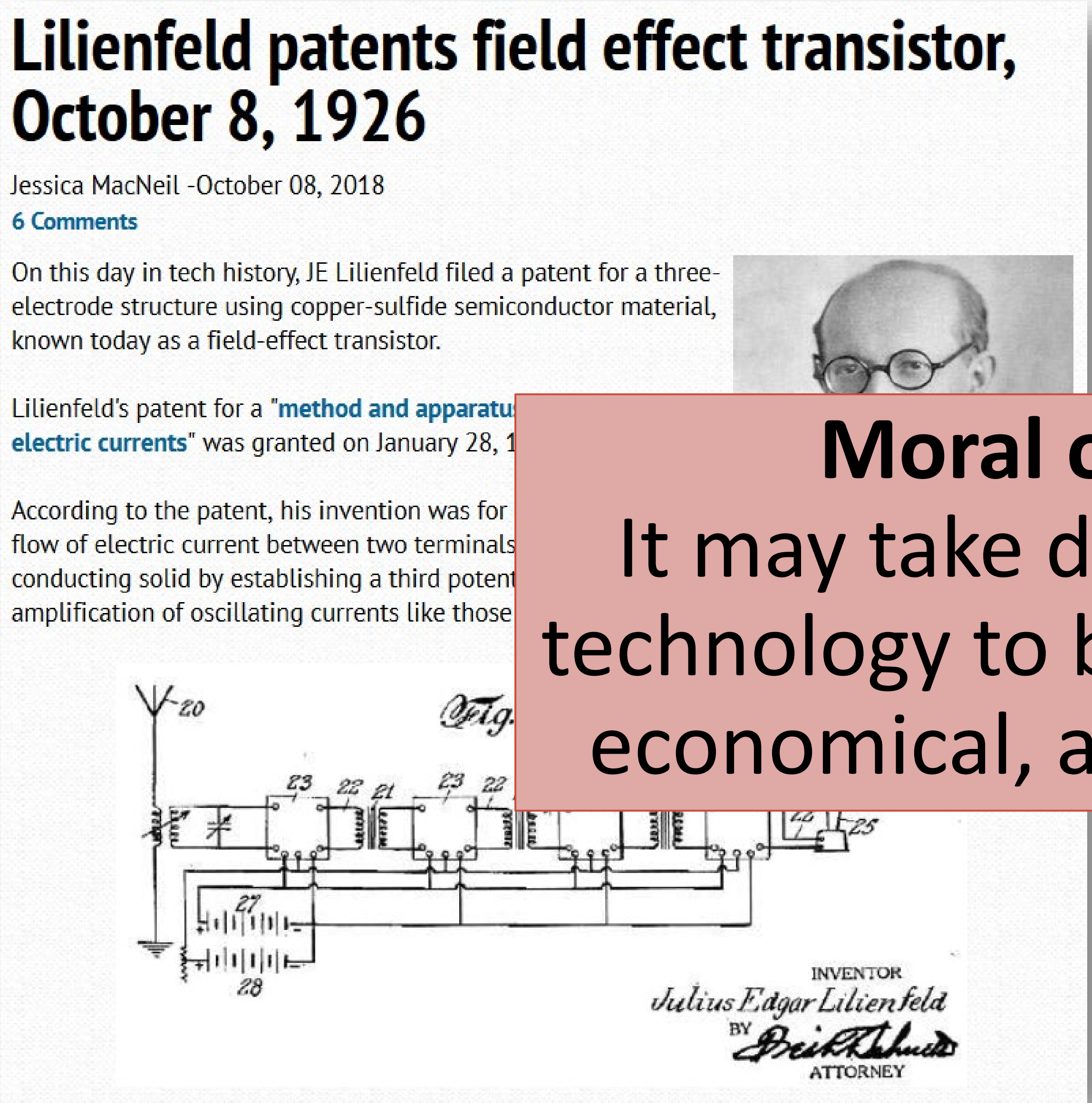
★ Large, fault-tolerant modular QC G4
>1000 logical qubit QC, single module
High-fidelity logical qubits
More qubits, better fidelity
Gate-based QC with 100's of qubits running QEC
Gate-based QC with 100's of qubits
★ Gate-based QC demonstrating practical utility G3
★ Quantum annealer demonstrating practical utility A3
Find compelling NISQ application
Gate-based QC demonstrating QEC G2b
Scale number of qubits while maintaining fidelity
Improve qubit quality
Quantum annealer demonstrating quantum supremacy A2
Gate-based QC demonstrating quantum supremacy G2a
Scale to 50+ qubits while maintaining gate fidelity
FUTURE R&D
CURRENT R&D
Experimental quantum annealer A1
Small (10s of qubits) gate-based QC G1

FIGURE 7.4 An illustration of potential milestones of progress in quantum computing. The arrangement of milestones corresponds to the order in which the committee thinks they are likely to be achieved; however, it is possible that some will not be achieved, or that they will not be achieved in the order indicated.

National Laboratory

http://nap.edu/25196

Fun Question: when was the field effect transistor patented?



Lilienfeld patents field effect transistor, October 8, 1926

Jessica MacNeil -October 08, 2018

6 Comments

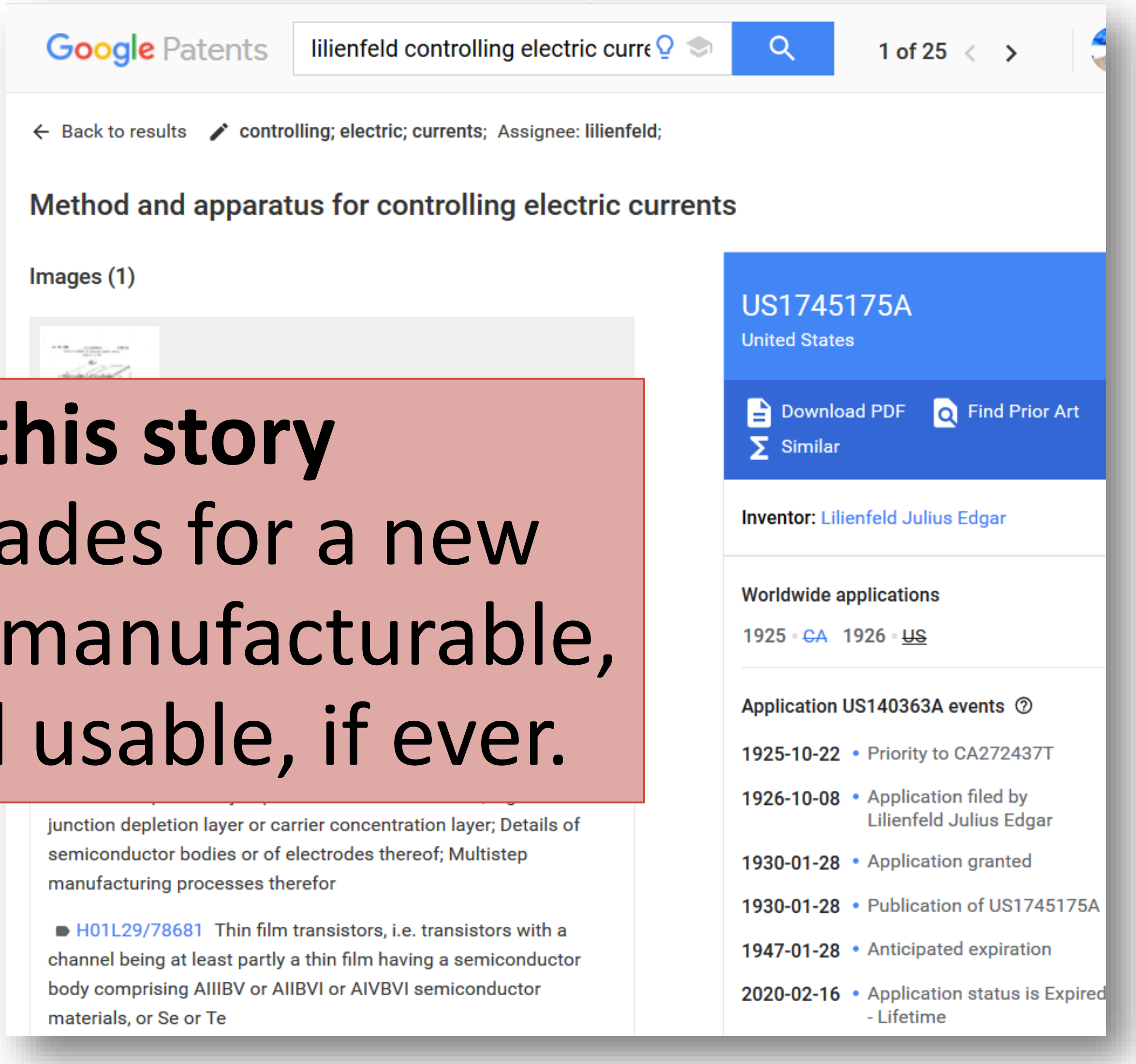
On this day in tech history, JE Lilienfeld filed a patent for a three-electrode structure using copper-sulfide semiconductor material, known today as a field-effect transistor.

Lilienfeld's patent for a "method and apparatu electric currents" was granted on January 28, 1

According to the patent, his invention was for flow of electric current between two terminals conducting solid by establishing a third potent amplification of oscillating currents like those

Moral of this story

It may take decades for a new technology to be manufacturable, economical, and usable, if ever.



controlling; electric; currents; Assignee: lilienfeld;

Method and apparatus for controlling electric currents

Images (1)

junction depletion layer or carrier concentration layer; Details of semiconductor bodies or of electrodes thereof; Multistep manufacturing processes therefor

H01L29/78681 Thin film transistors, i.e. transistors with a channel being at least partly a thin film having a semiconductor body comprising AIIIBV or AIIBVI or AIVBVI semiconductor materials, or Se or Te

US1745175A
United States

Inventor: Lilienfeld Julius Edgar

Worldwide applications

1925 • CA 1926 • US

Application US140363A events

1925-10-22	Priority to CA272437T
1926-10-08	Application filed by Lilienfeld Julius Edgar
1930-01-28	Application granted
1930-01-28	Publication of US1745175A
1947-01-28	Anticipated expiration
2020-02-16	Application status is Expired - Lifetime

https://www.edn.com/electronics-blogs/edn-moments/4422371/Lilienfeld-patents-field-effect-transistor--October-8--1926

Predictions for Transition Period

Optimize Software and Expose New Hierarchical Parallelism

- Redesign software to boost performance on upcoming architectures
- Exploit new levels of parallelism and efficient data movement

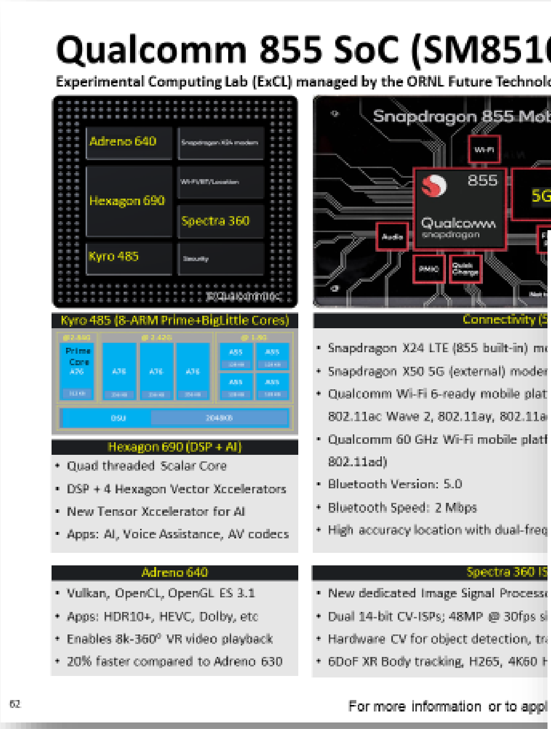
Architectural Specialization and Integration

- Use CMOS more effectively for specific workloads
- Integrate components to boost performance and eliminate inefficiencies
- Workload specific memory+storage system design

Emerging Technologies

- Investigate new computational paradigms
  - Quantum
  - Neuromorphic
  - Advanced Digital
  - Emerging Memory Devices

Various Markets already Experiencing these Architectural Trends



https://excl.ornl.gov/

Qualcomm 855 SoC (SM8510P) Snapdragon™

Experimental Computing Lab (ExCL) managed by the ORNL Future Technologies Group

Kyro 485 (8-ARM Prime+BigLittle Cores)

Hexagon 690 (DSP + AI)

- Quad threaded Scalar Core
- DSP + 4 Hexagon Vector Xccelerators
- New Tensor Xccelerator for AI
- Apps: AI, Voice Assistance, AV codecs

Adreno 640

- Vulkan, OpenCL, OpenGL ES 3.1
- Apps: HDR10+, HEVC, Dolby, etc
- Enables 8k-360° VR video playback
- 20% faster compared to Adreno 630

Connectivity (S

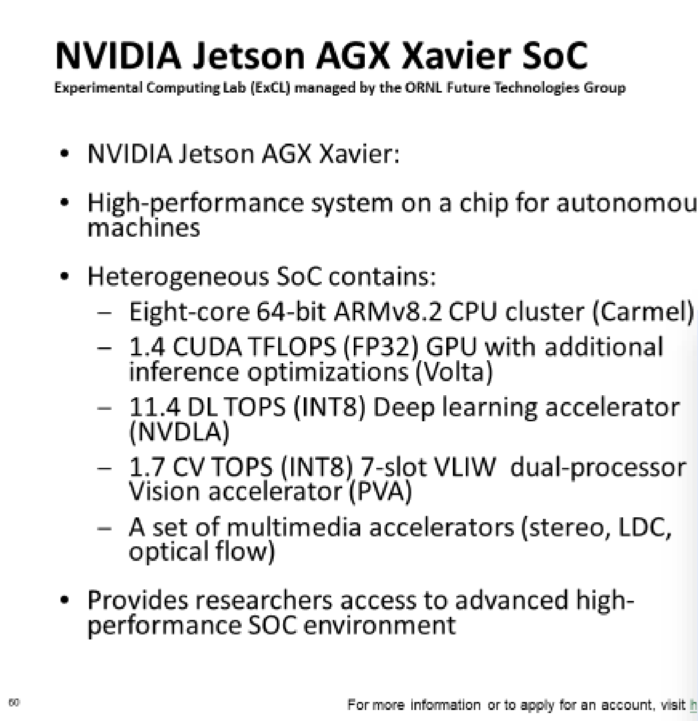
- Snapdragon X24 LTE (855 built-in) m
- Snapdragon X50 5G (external) mode
- Qualcomm Wi-Fi 6-ready mobile plat 802.11ac Wave 2, 802.11ay, 802.11a
- Qualcomm 60 GHz Wi-Fi mobile platf 802.11ad)
- Bluetooth Version: 5.0
- Bluetooth Speed: 2 Mbps
- High accuracy location with dual-freq

Spectra 360 IS

- New dedicated Image Signal Process
- Dual 14-bit CV-ISPs; 48MP @ 30fps s
- Hardware CV for object detection, tr
- 6DoF XR Body tracking, H265, 4K60 H

62

For more information or to appl



https://excl.ornl.gov/

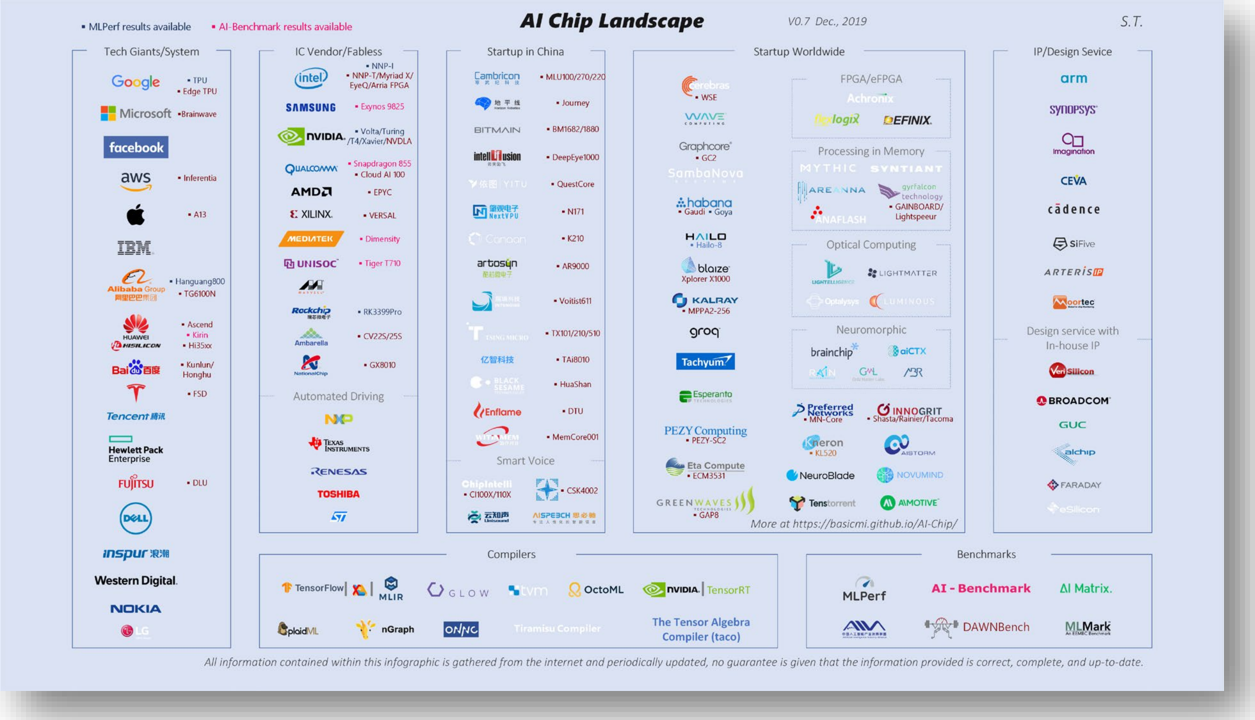
NVIDIA Jetson AGX Xavier SoC

Experimental Computing Lab (ExCL) managed by the ORNL Future Technologies Group

- NVIDIA Jetson AGX Xavier:
- High-performance system on a chip for autonomous machines
- Heterogeneous SoC contains:
  - Eight-core 64-bit ARMv8.2 CPU cluster (Carmel)
  - 1.4 CUDA TFLOPS (FP32) GPU with additional inference optimizations (Volta)
  - 11.4 DL TOPS (INT8) Deep learning accelerator (NVDLA)
  - 1.7 CV TOPS (INT8) 7-slot VLIW dual-processor Vision accelerator (PVA)
  - A set of multimedia accelerators (stereo, LDC, optical flow)
- Provides researchers access to advanced high-performance SOC environment

60

For more information or to apply for an account, visit h



AI Chip Landscape

V0.7 Dec., 2019

S.T.

- MLPerf results available
- AI-Benchmark results available

Tech Giants/System

Google • TPU • Edge TPU
Microsoft •Brainwave
facebook
aws • Inferentia
• A13
IBM
Alibaba Group 阿里巴巴集团 • Hanguang800 • TG6100N
HUAWEI HISILICON • Ascend • Kirin • Hi35xx
Bai du 百度 • Kunlun/ Honghu
• FSD
Tencent 腾讯
Hewlett Pack Enterprise
FUJITSU • DLU
DELL
inspur 浪潮
Western Digital
NOKIA
LG

IC Vendor/Fabless

intel • NNP-I • NNP-T/Myriad X/ EyeQ/Arria FPGA
SAMSUNG • Exynos 9825
NVIDIA • Volta/Turing /T4/Xavier/NVDLA
QUALCOMM • Snapdragon 855 • Cloud AI 100
AMD • EPYC
XILINX • VERSAL
MEDIATEK • Dimensity
UNISOC • Tiger T710
MARVELL
Rockchip 瑞芯微电子 • RK3399Pro
Ambarella • CV22S/25S
NationalChip • GX8010

Automated Driving

NXP
Texas Instruments
RENESAS
TOSHIBA
ST

Startup in China

Cambricon 寒武纪科技 • MLU100/270/220
地平线 Horizon Robotics • Journey
BITMAIN • BM1682/1880
intellifusion 云天励飞 • DeepEye1000
依图 YITU • QuestCore
肇观电子 NextVPU • N171
Canaan • K210
artosyn 酷芯微电子 • AR9000
INTENGINE • Voitist611
TSING MICRO • TX101/210/510
亿智科技 • TAi8010
BLACK SESAME • HuaShan
Enflame • DTU
WITINMEM • MemCore001

Smart Voice

ChipIntelli • CI100X/110X
• CSK4002
云知声 Unisound
AISPEECH 思必驰

Startup Worldwide

cerebras • WSE
WAVE COMPUTING
Graphcore • GC2
SambaNova
habana • Gaudi • Goya
HAILO • Hailo-8
blaize Xplorer X1000
KALRAY • MPPA2-256
groq
Tachyum
Esperanto
PEZY Computing • PEZY-SC2
Eta Compute • ECM3531
GREENWAVES • GAP8

FPGA/eFPGA

Achronix
flexlogix
EFINIX

Processing in Memory

MYTHIC
SYNTIANT
AREANNA
gyrfalcon technology • GAINBOARD/ Lightspeeur
ANAFLASH

Optical Computing

LIGHTELLIGENCE
LIGHTMATTER
Optalysys
LUMINOUS

Neuromorphic

brainchip
aiCTX
RAIN
GML
ABR

Preferred Networks • MN-Core
INNOGRIT • Shasta/Rainier/Tacoma
Kneron • KL520
AISTORM
NeuroBlade
NOVUMIND
Tenstorrent
AIMOTIVE

More at https://basicmi.github.io/AI-Chip/

IP/Design Sevice

arm
synopsys
Imagination
CEVA
cadence
SiFive
ARTERIS IP
Moortec

Design service with In-house IP

VeriSilicon
BROADCOM
GUC
alchip
FARADAY
eSilicon

Compilers

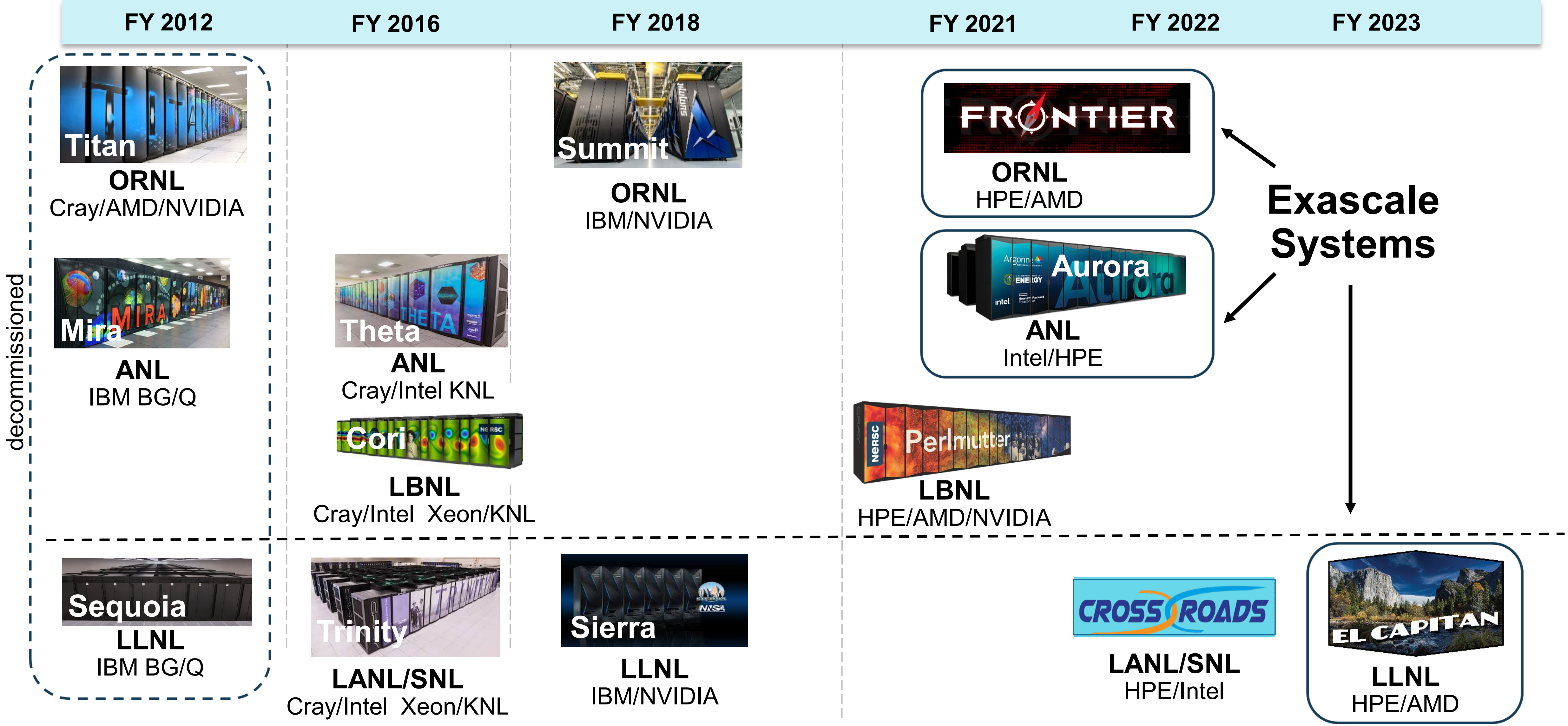
TensorFlow | XLA | MLIR
GLOW
tvm
OctoML
NVIDIA | TensorRT
plaidML
nGraph
ONNC
Tiramisu Compiler
The Tensor Algebra Compiler (taco)

Benchmarks

MLPerf
AI - Benchmark
AI Matrix.
AIIA
DAWNBench
MLMark

All information contained within this infographic is gathered from the internet and periodically updated, no guarantee is given that the information provided is correct, complete, and up-to-date.

DOE HPC Roadmap to Exascale Systems



FY 2012

ORNL
Cray/AMD/NVIDIA

ANL
IBM BG/Q

LLNL
IBM BG/Q

decommissioned

FY 2016

ANL
Cray/Intel KNL

LBNL
Cray/Intel Xeon/KNL

LANL/SNL
Cray/Intel Xeon/KNL

FY 2018

ORNL
IBM/NVIDIA

LLNL
IBM/NVIDIA

FY 2021

ORNL
HPE/AMD

ANL
Intel/HPE

LBNL
HPE/AMD/NVIDIA

FY 2022

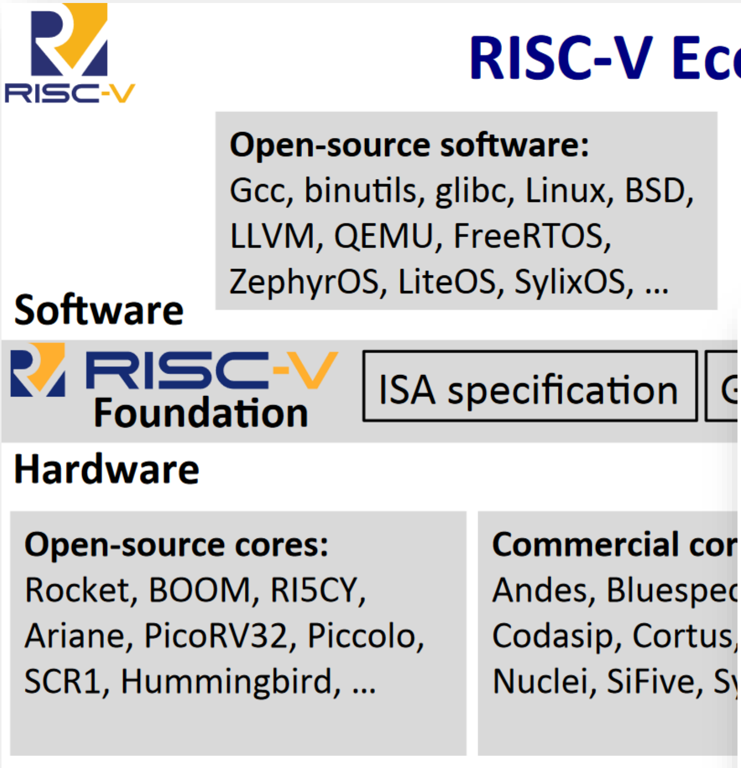
LANL/SNL
HPE/Intel

FY 2023

LLNL
HPE/AMD

Exascale Systems

Future -> Open Source Hardware Enables a Rapid Design of Specialized Chips and Effectively Mass Customization



RISC-V Ecosystem

Software

Open-source software:
Gcc, binutils, glibc, Linux, BSD, LLVM, QEMU, FreeRTOS, ZephyrOS, LiteOS, SylixOS, …

Commercial software:
Lauterbach, Segger, Micrium, ExpressLogic, …

RISC-V Foundation

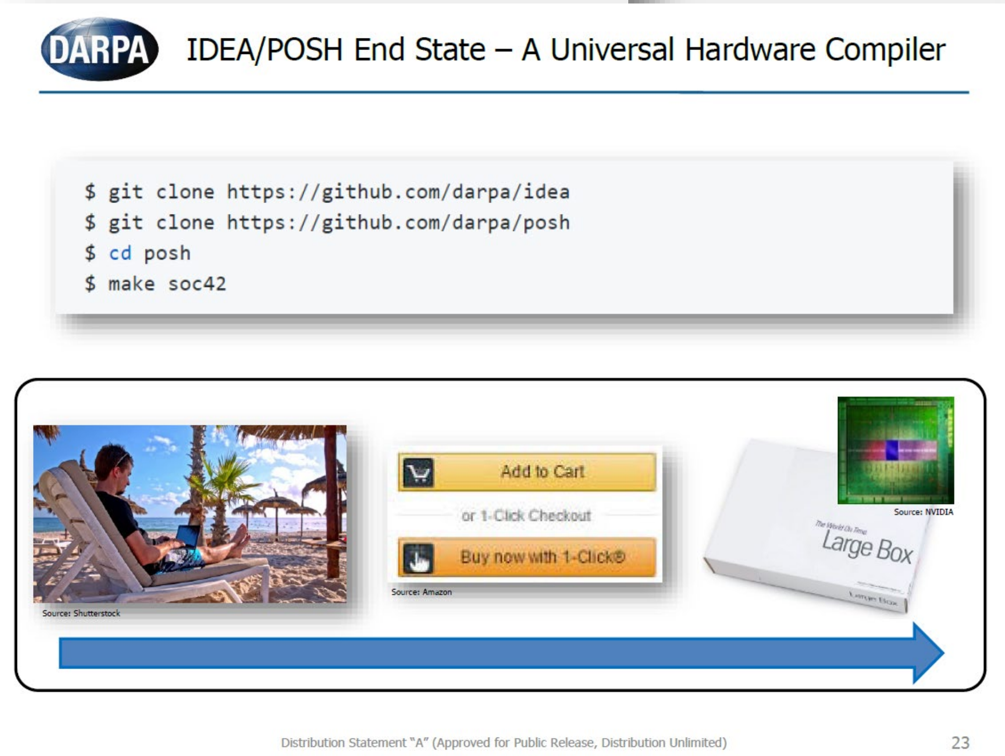
ISA specification

G

Hardware

Open-source cores:
Rocket, BOOM, RI5CY, Ariane, PicoRV32, Piccolo, SCR1, Hummingbird, …

Commercial cor
Andes, Bluespec
Codasip, Cortus,
Nuclei, SiFive, Sy

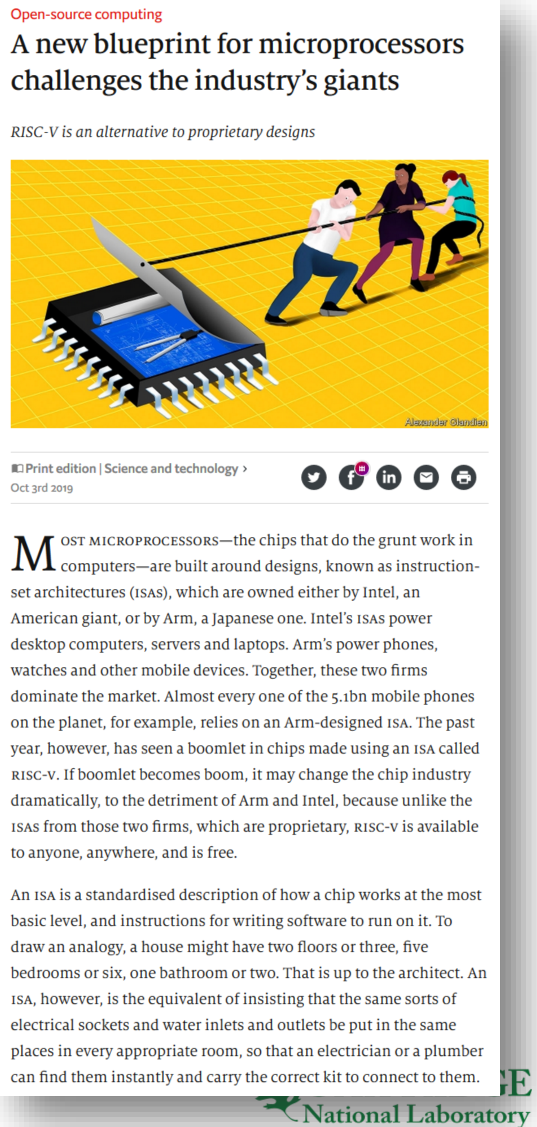


DARPA IDEA/POSH End State – A Universal Hardware Compiler

```
$ git clone https://github.com/darpa/idea
$ git clone https://github.com/darpa/posh
$ cd posh
$ make soc42
```

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited) 23

A. Olofsson, 2018



Open-source computing

A new blueprint for microprocessors challenges the industry's giants

RISC-V is an alternative to proprietary designs

Print edition | Science and technology ›
Oct 3rd 2019

MOST MICROPROCESSORS—the chips that do the grunt work in computers—are built around designs, known as instruction-set architectures (ISAs), which are owned either by Intel, an American giant, or by Arm, a Japanese one. Intel's ISAs power desktop computers, servers and laptops. Arm's power phones, watches and other mobile devices. Together, these two firms dominate the market. Almost every one of the 5.1bn mobile phones on the planet, for example, relies on an Arm-designed ISA. The past year, however, has seen a boomlet in chips made using an ISA called RISC-V. If boomlet becomes boom, it may change the chip industry dramatically, to the detriment of Arm and Intel, because unlike the ISAs from those two firms, which are proprietary, RISC-V is available to anyone, anywhere, and is free.

An ISA is a standardised description of how a chip works at the most basic level, and instructions for writing software to run on it. To draw an analogy, a house might have two floors or three, five bedrooms or six, one bathroom or two. That is up to the architect. An ISA, however, is the equivalent of insisting that the same sorts of electrical sockets and water inlets and outlets be put in the same places in every appropriate room, so that an electrician or a plumber can find them instantly and carry the correct kit to connect to them.

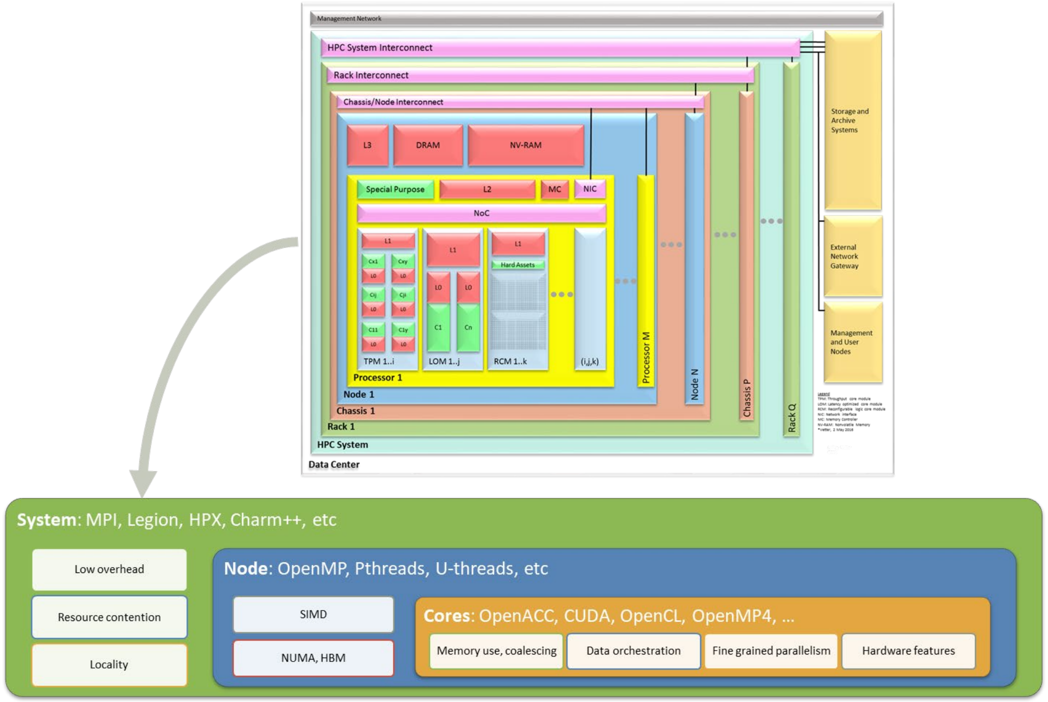
Take away message →
During this Sixth Wave transition, Complexity is our major challenge!

Architecture

- How do we design future systems so that they are better than current systems on important applications?
- Simulation and modeling are more difficult
- Entirely possible that the new system will be slower than the old system!
- Expect 'disaster' procurements

Programmability

- How do we design applications with some level of performance portability?
- Software lasts much longer than transient hardware platforms
- Proper abstractions for flexibility and efficiency
- Adapt or die



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

Final Report on Workshop on Extreme Heterogeneity : https://doi.org/10.2172/1473756

NITRD Software in the Era of Extreme Heterogeneity (Sep 2020) https://www.nitrd.gov/nitrdgroups/index.php?title=Software-Extreme-Heterogeneity

Implication for Applications and Software → Explosion of (incomplete) Programming Systems →

Programming Models

- OpenMP
- OpenMP Offload
- OpenACC
- SYCL
- DPC++
- HIP
- CUDA
- OpenCL
- Kokkos
- Raja
- *Many others…*

Implementations (SYCL Example)

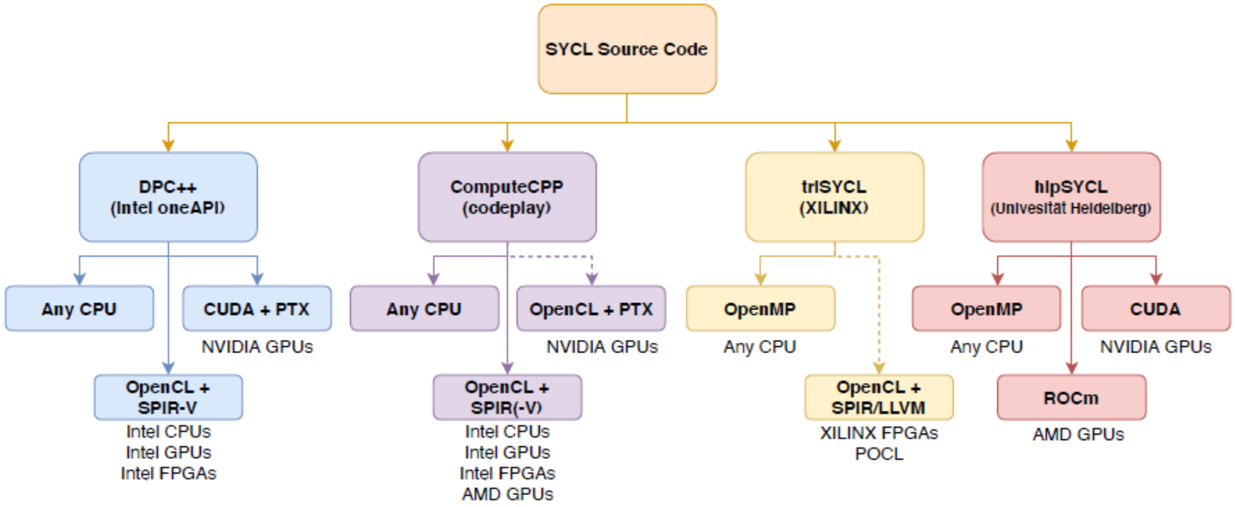


Fig. 1: Current SYCL implementations and their corresponding API backends.

Inconsistent: feature support, performance, tools ecosystem

Oak Ridge National Laboratory

Software Strategies for Extreme Heterogeneity

Strategies for Programming Systems in this Era of Rapidly Designed, Diverse Architectures

Goals

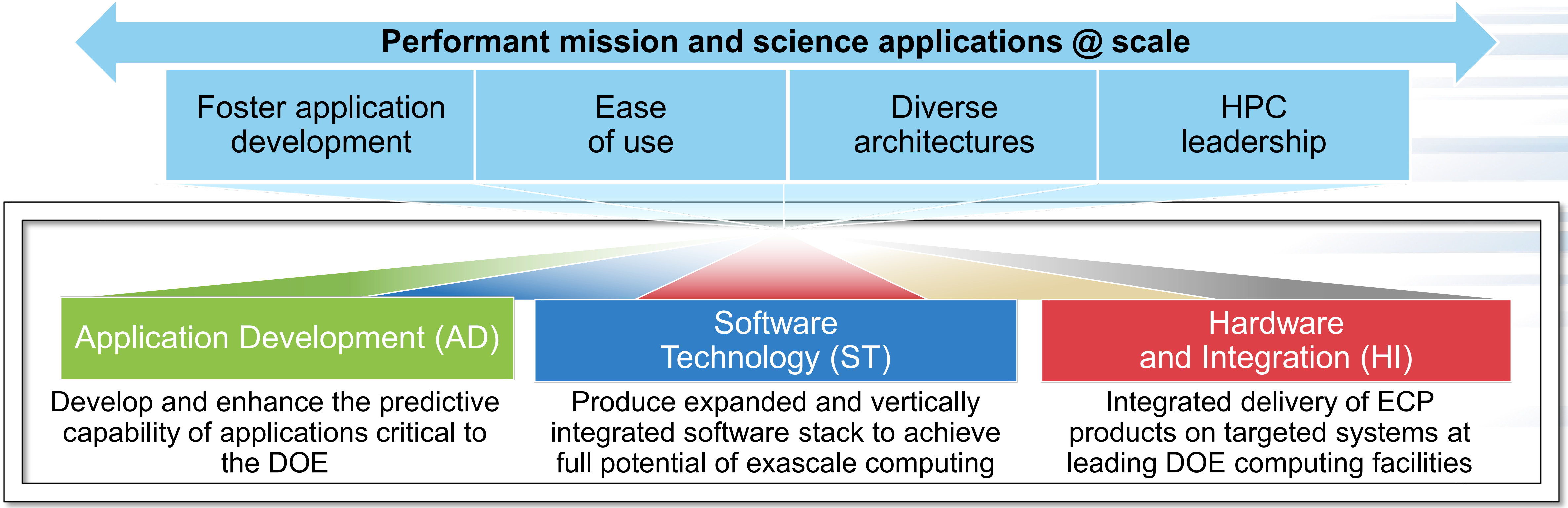
- Strive for ‘Write one, run anywhere’
- Descriptive models of parallelism and data movement that enable effective code generation
- Introspective runtime systems
- Layered, modular, open source approaches required
  - One organization can’t do it all

Examples

- Contributing to LLVM
  - FORTRAN with GPU offloading
- Programming FPGAs
  - Without Verilog
- Memory systems are changing too
  - Language support for NVM

Contributing to LLVM

The three technical areas in ECP have the necessary components to meet national goals



Performant mission and science applications @ scale

Foster application development

Ease of use

Diverse architectures

HPC leadership

Application Development (AD)

Develop and enhance the predictive capability of applications critical to the DOE

Software Technology (ST)

Produce expanded and vertically integrated software stack to achieve full potential of exascale computing

Hardware and Integration (HI)

Integrated delivery of ECP products on targeted systems at leading DOE computing facilities

25 applications ranging from national security, to energy, earth systems, economic security, materials, and data

80+ unique software products spanning programming models and run times, math libraries, data and visualization

6 vendors supported by PathForward focused on memory, node, connectivity advancements; deployment to facilities

ECP is Improving the LLVM Compiler Ecosystem

LLVM
COMPILER INFRASTRUCTURE

LLVM

- Very popular open source compiler infrastructure
- Easily extensible
- Widely used and contributed to in industry
- Permissive license
- Used for heterogeneous computing

+SOLLVE

- Enhancing the implementation of OpenMP in LLVM
- Unified memory
- OMP Optimizations
- Prototype OMP features for LLVM
- OMP test suite
- Tracking OMP implementation quality

+PROTEAS-TUNE

- Core optimization improvements to LLVM
- OpenACC capability for LLVM
    - Clacc
    - Flacc
- Autotuning for OpenACC and OpenMP in LLVM
- Integration with Tau performance tools

+FLANG

- Developing an open-source, production Fortran frontend
- Upstream to LLVM public release
- Support for OpenMP and OpenACC
- Recently approved by LLVM

+HPCToolkit

- Improvements to OpenMP profiling interface OMPT
- OMPT specification improvements
- Refine HPCT for OMPT improvements

+ATDM

- Enhancing LLVM to optimize template expansion for FlexCSI, Kokkos, RAJA, etc.
- Flang testing and evaluation

Vendors

- Increasing dependence on LLVM
- Collaborations with many vendors using LLVM
    - AMD
    - ARM
    - Cray
    - HPE
    - IBM
    - Intel
    - NVIDIA

Active involvement with broad LLVM community: LLVM Dev, EuroLLVM

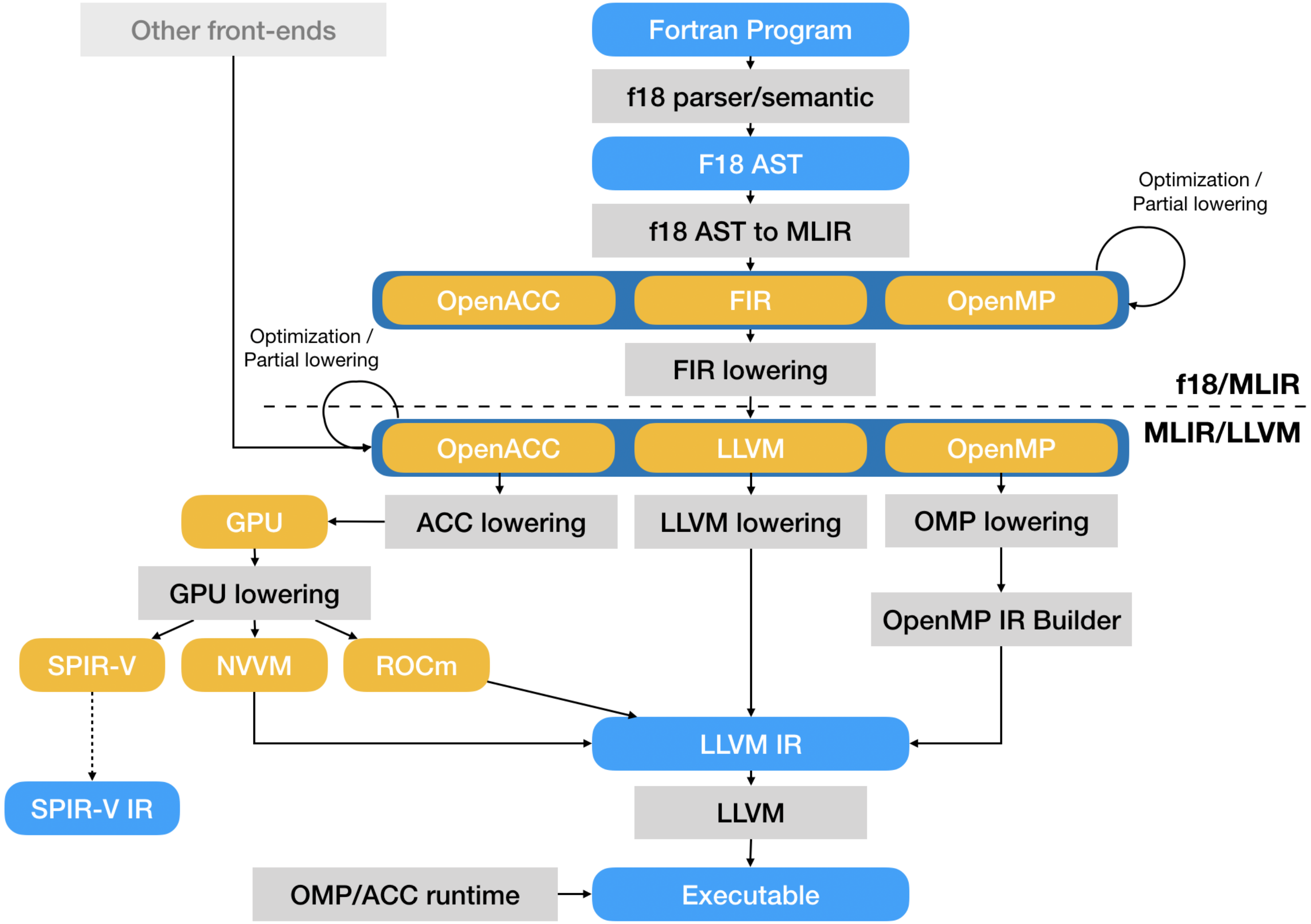
https://github.com/llvm-doe-org/llvm-project

EXASCALE COMPUTING PROJECT

Leveraging LLVM Ecosystem to Meet a Critical ECP (community) need : FORTRAN

ECP Projects: Flang, SOLLVE, PROTEAS-TUNE
Many other contributors: NNSA, NVIDIA, ARM, Google, …

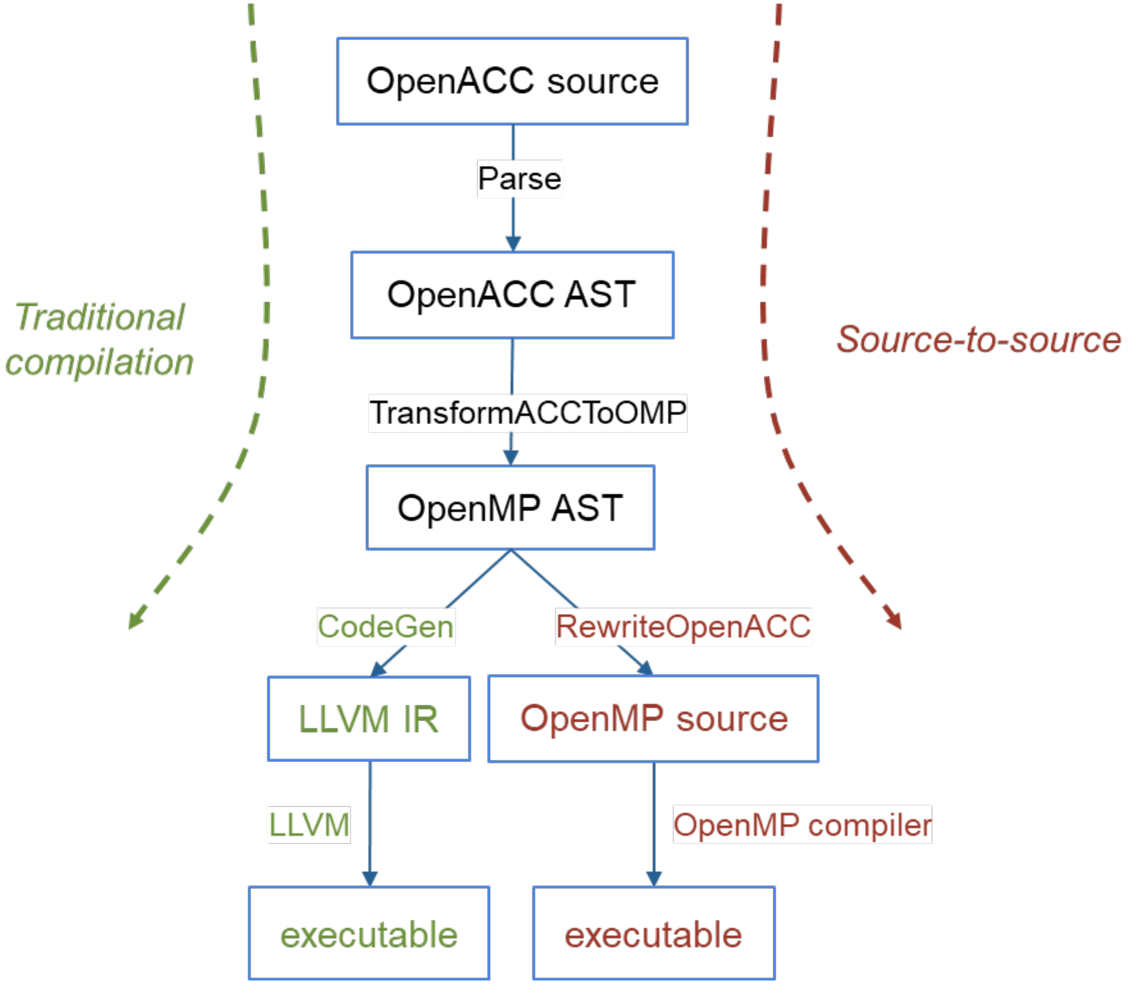
- Fortran support continues to be an ongoing requirement
- Flang project started in NNSA funding NVIDIA/PGI to open source compiler front-end into LLVM ecosystem
- SOLLVE is improving OpenMP dialect, implementation, and core optimizations
- PROTEAS-TUNE is creating OpenACC dialect and improving MLIR
- ECP projects are contributing many changes upstream to LLVM core, MLIR, etc
- Many others are contributing: backends for processors, optimizations in toolchain, …
  - Google contributed MLIR



PROTEAS-TUNE: Clacc – OpenACC in Clang/LLVM

2.3.2.10	PROTEAS-TUNE
POC	Joel Denny, ORNL

- Develop production-quality, standard-conforming traditional OpenACC compiler and runtime support by extending Clang and LLVM
  - Build on existing OpenMP infrastructure
- Enable research and development of source-level OpenACC tools
  - Design compiler to leverage Clang/LLVM ecosystem extensibility
  - E.g., Pretty printers, analyzers, lint tools, and debugger and editor extensions
- Actively contribute improvements to the OpenACC specification
- Actively contribute upstream all Clang and LLVM improvements that are mutually beneficial
  - Many contributions are already in LLVM
- Open-source with multiple collaborators (vendors, universities)



Traditional compilation

OpenACC source

Parse

OpenACC AST

TransformACCToOMP

OpenMP AST

CodeGen

RewriteOpenACC

LLVM IR

OpenMP source

LLVM

OpenMP compiler

executable

executable

Source-to-source

Clacc: Translating OpenACC to OpenMP in Clang, Joel E. Denny, Seyong Lee, and Jeffrey S. Vetter, 2018 IEEE/ACM 5th Workshop on the LLVM Compiler Infrastructure in HPC (LLVM-HPC), Dallas, TX, USA, (2018).

Programming FPGAs with OpenACC

Challenges in FPGA Computing

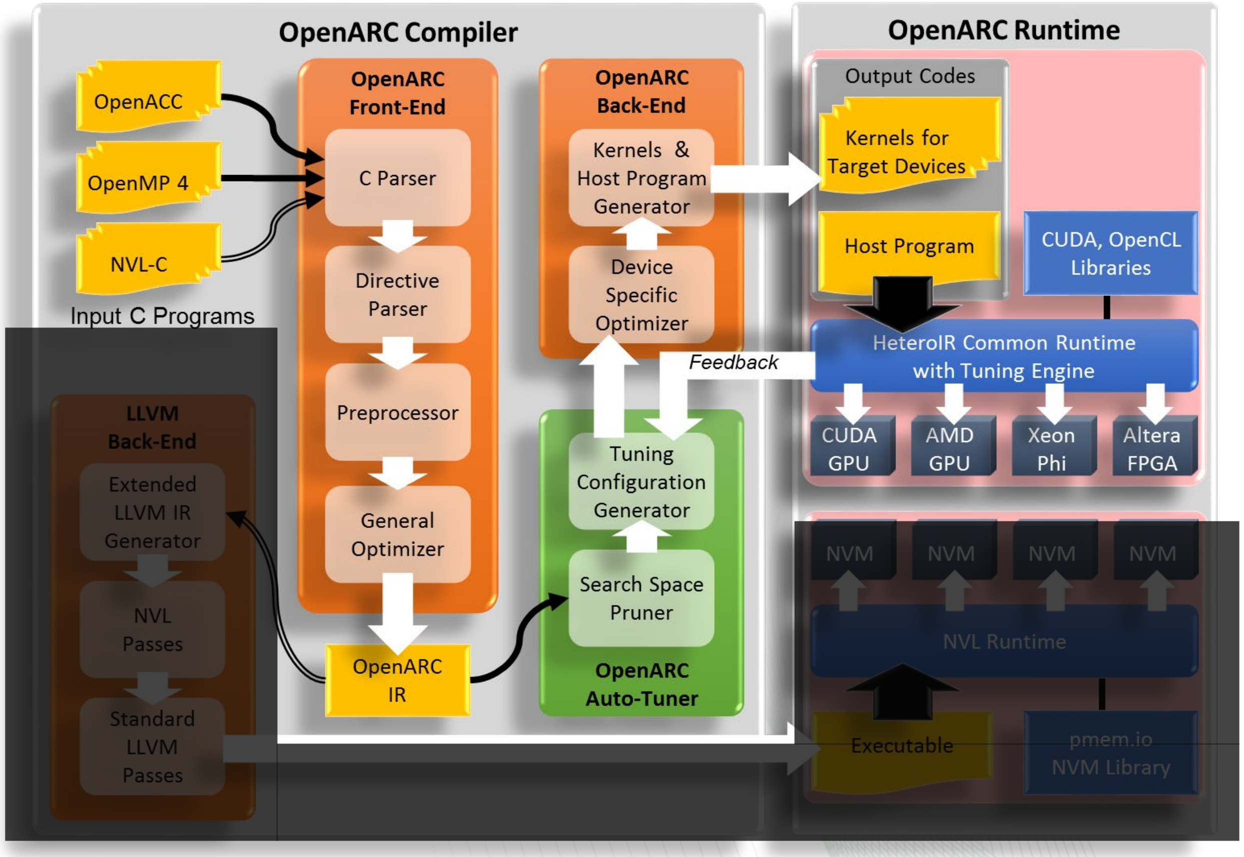
- Programmability and Portability Issues
  - Best performance for FPGAs requires writing Hardware Description Languages (HDLs) such as VHDL and Verilog; too complex and low-level
    - HDL requires substantial knowledge on hardware (digital circuits).
    - Programmers must think in terms of a state machine.
    - HDL programming is a kind of digital circuit design.
  - High-Level Synthesis (HLS) to provide better FPGA programmability
    - SRC platforms, Handel-C, Impulse C-to-FPGA compiler, Xilinx Vivado (AutoPilot), FCUDA, etc.
    - None of these use a portable, open standard.

Standard, Portable Programming Models for Heterogeneous Computing

- OpenCL
  - Open standard portable across diverse heterogeneous platforms (e.g., CPUs, GPUs, DSPs, Xeon Phis, FPGAs, etc.)
  - Much higher than HDL, but still complex for typical programmers.
- Directive-based accelerator programming models
  - OpenACC, OpenMP4, etc.
  - Provide higher abstraction than OpenCL.
  - Most of existing OpenACC/OpenMP4 compilers target only specific architectures; none supports FPGAs.

Directive-based Strategy with 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.



S. Lee and J.S. Vetter, "OpenARC: Open Accelerator Research Compiler for Directive-Based, Efficient Heterogeneous Computing," in *ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC). Vancouver: ACM, 2014*

FPGAs|Approach

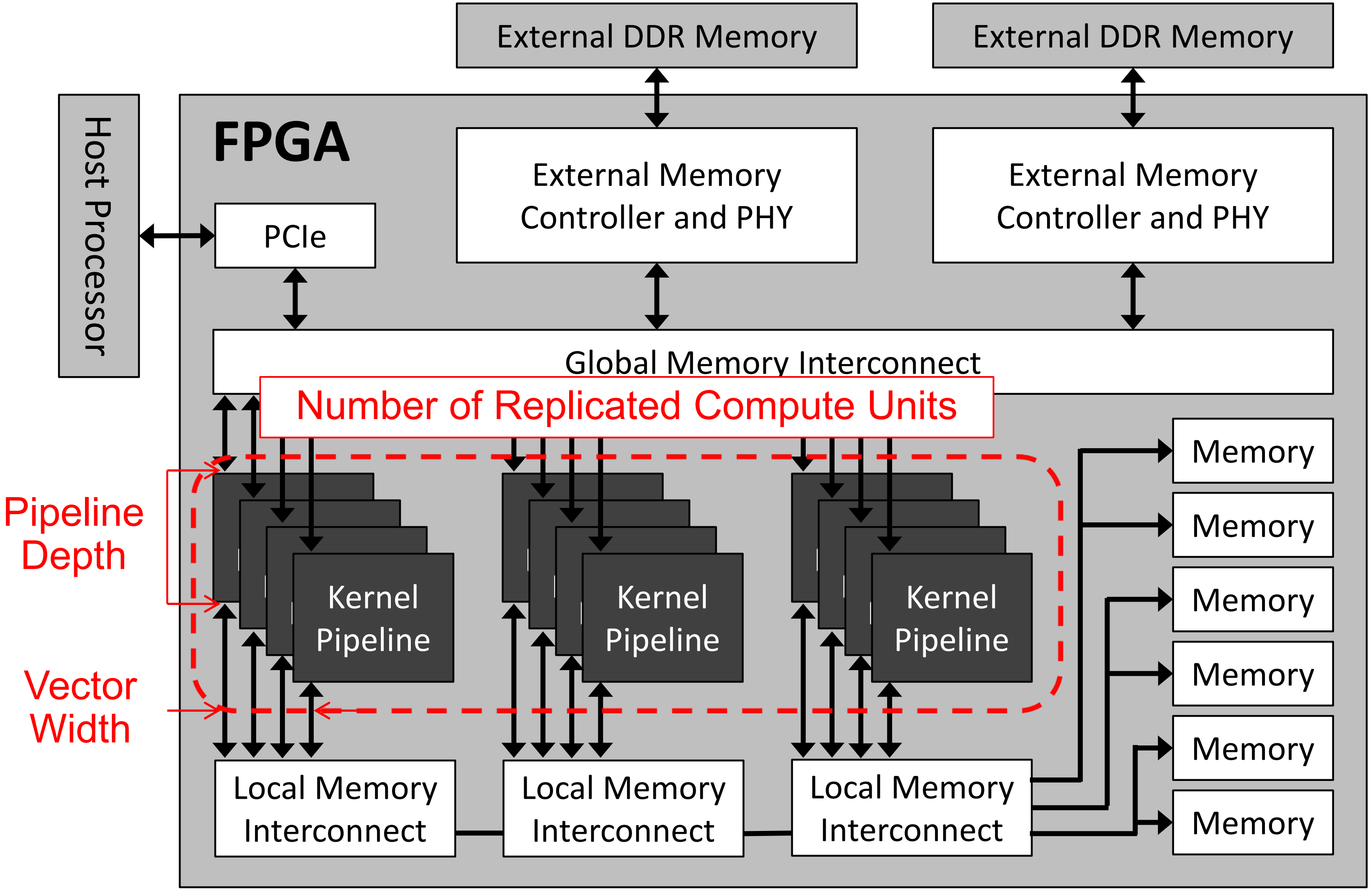
- Design and implement an OpenACC-to-FPGA translation framework, which is the first work to use a standard and portable directive-based, high-level programming system for FPGAs.
- Propose FPGA-specific optimizations and novel pragma extensions to improve performance.
- Evaluate the functional and performance portability of the framework across diverse architectures (Altera FPGA, NVIDIA GPU, AMD GPU, and Intel Xeon Phi).

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, 10.1109/IPDPS.2016.28.

Baseline Translation of OpenACC-to-FPGA

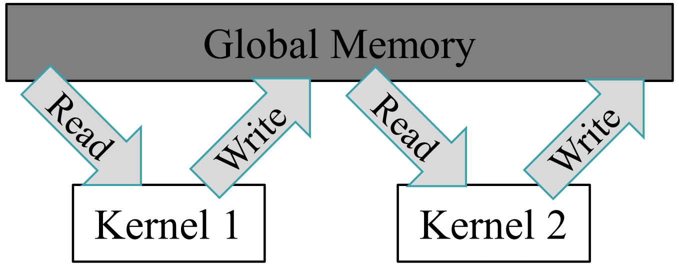
- Use OpenCL as the output model and the Altera Offline Compiler (AOC) as its backend compiler.
- Translates the input OpenACC program into a host code containing HeteroIR constructs and device-specific kernel codes.
  - Use the same HeteroIR runtime system of the existing OpenCL backends, except for the device initialization.
  - Reuse most of compiler passes for kernel generation.

FPGA OpenCL Architecture

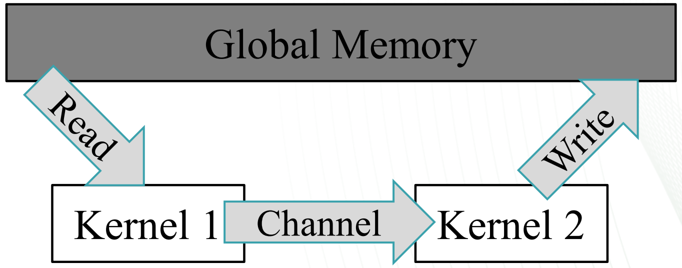

External DDR Memory
External DDR Memory
Host Processor
FPGA
PCIe
External Memory Controller and PHY
External Memory Controller and PHY
Global Memory Interconnect
Number of Replicated Compute Units
Pipeline Depth
Vector Width
Kernel Pipeline
Kernel Pipeline
Kernel Pipeline
Memory
Memory
Memory
Memory
Memory
Memory
Local Memory Interconnect
Local Memory Interconnect
Local Memory Interconnect

Kernel-Pipelining Transformation Optimization

- Kernel execution model in OpenACC
  - Device kernels can communicate with each other only through the device global memory.
  - Synchronizations between kernels are at the granularity of a kernel execution.
- Altera OpenCL channels
  - Allows passing data between kernels and synchronizing kernels with high efficiency and low latency



Kernel communications through global memory in OpenACC

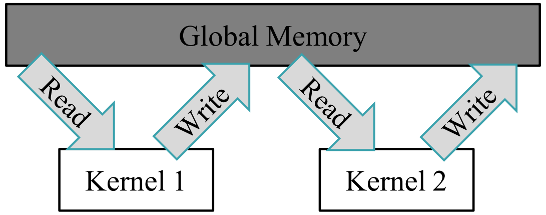


Kernel communications with Altera channels

Kernel-Pipelining Transformation Optimization (2)

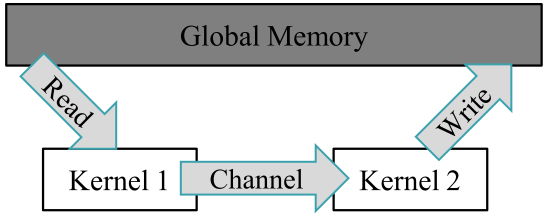
(a) Input OpenACC code

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



(b) Altera OpenCL code with channels

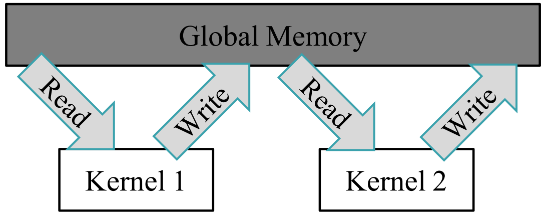
```
channel float pipe_b;
__kernel void kernel1(__global float* a) {
    int i = get_global_id(0);
    write_channel_altera(pipe_b, a[i]*a[i]);
}
__kernel void kernel2(__global float* c) {
    int i = get_global_id(0);
    c[i] = read_channel_altera(pipe_b);
}
```



Kernel-Pipelining Transformation Optimization (3)

(a) Input OpenACC code

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

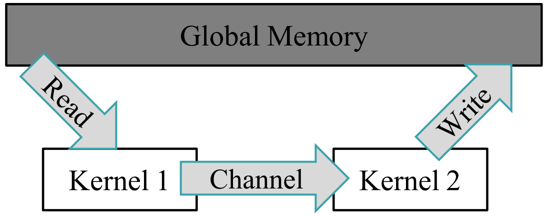


Kernel-pipelining transformation

Valid under specific conditions

(c) Modified OpenACC code for kernel-pipelining

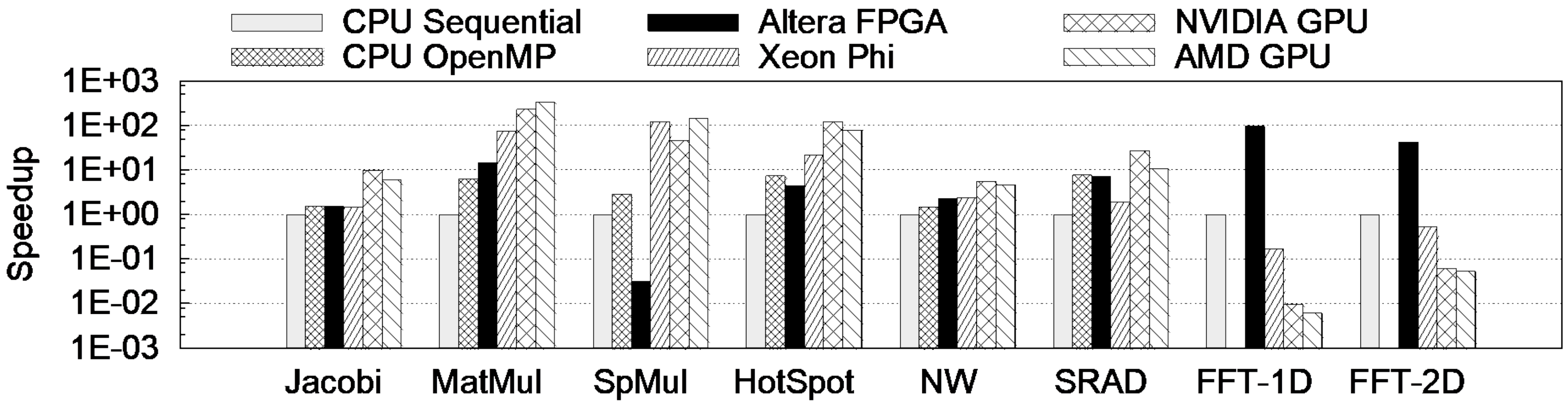
```
#pragma acc data copyin (a) pipe (b) copyout (c)
{
    #pragma acc kernels loop gang worker pipeout (b) present (a)
    For(i=0; i<N; i++) { b[i] = a[i]*a[i]; }
    #pragma acc kernels loop gang worker pipein (b) present (c)
    For(i=0; i<N; i++) {c[i] = b[i];}
}
```



FPGA-specific Optimizations

- Single work-item
- Collapse
- Reduction
- Sliding window
- (Branch-variant code motion)
- (Custom unrolling)

Overall Performance of OpenARC FPGA Evaluation



FPGAs prefer applications with deep execution pipelines (e.g., FFT-1D and FFT-2D), performing much higher than other accelerators.

For traditional HPC applications with abundant parallel floating-point operations, it seems to be difficult for FPGAs to beat the performance of other accelerators, even though FPGAs can be much more power-efficient.

- Tested FPGA does not contain dedicated, embedded floating-point cores, while others have fully-optimized floating-point computation units.

Current and upcoming high-end FPGAs are equipped with hardened floating-point operators, whose performance will be comparable to other accelerators, while remaining power-efficient.

Emerging Memory Systems

Memory Hierarchy is Specializing too

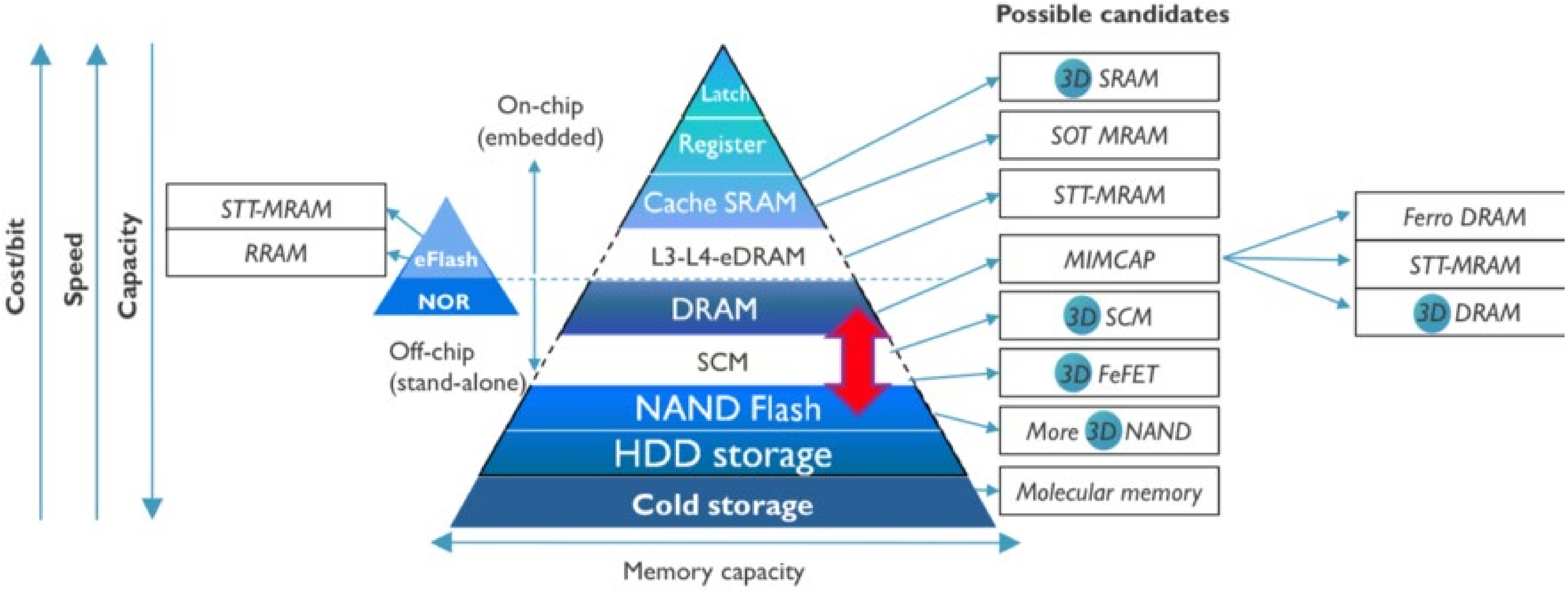
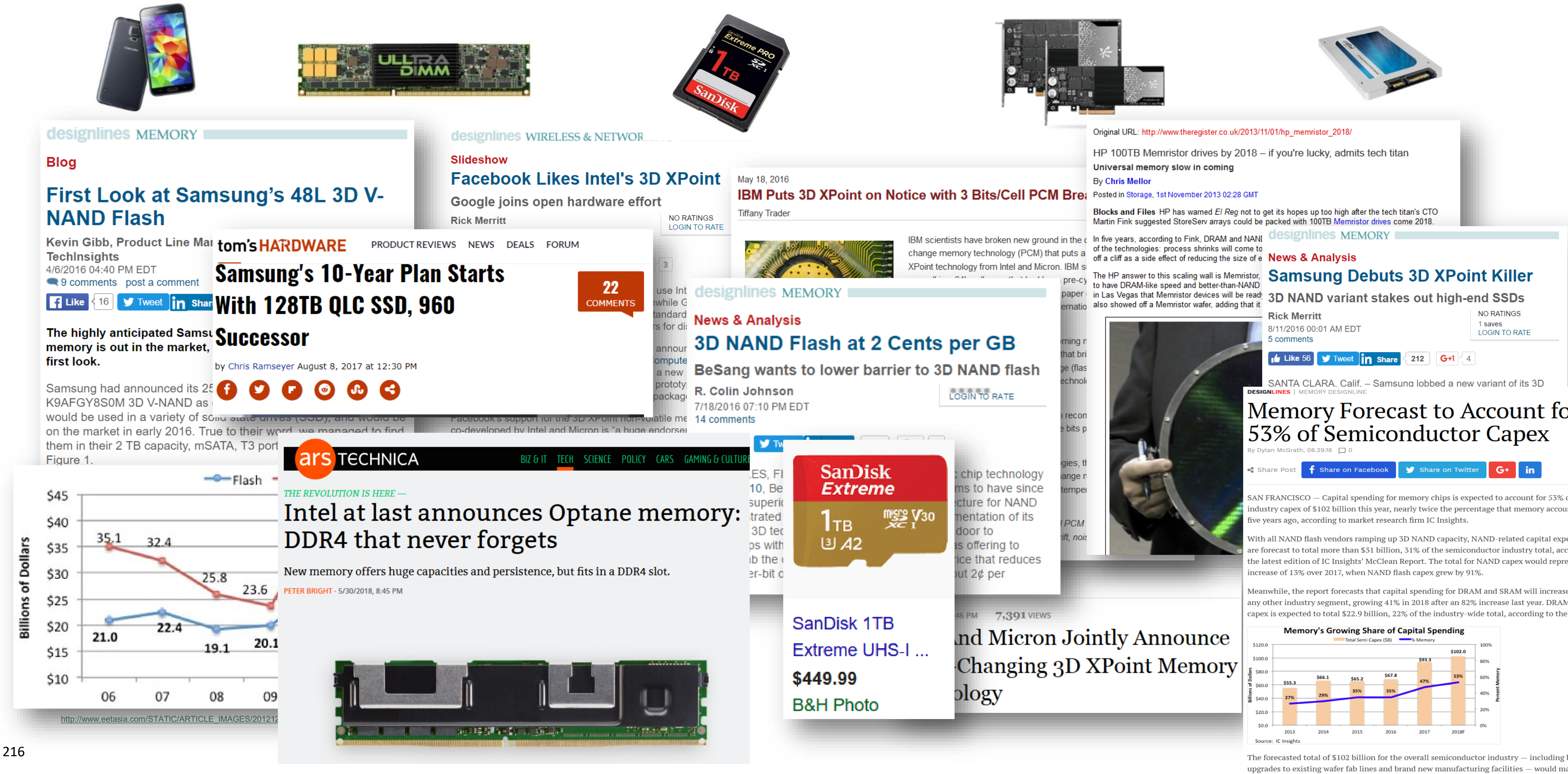


Image Source: IMEC

NVRAM Technology Continues to Improve – Driven by Broad Market Forces



designlines MEMORY

Blog

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

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

Like 16 Tweet Shar

The highly anticipated Samsu
memory is out in the market,
first look.

Samsung had announced its 25
K9AFGY8S0M 3D V-NAND as
would be used in a variety of so
on the market in early 2016. True to their word, we managed to find
them in their 2 TB capacity, mSATA, T3 port
Figure 1.

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

designlines WIRELESS & NETWOR

Slideshow

Facebook Likes Intel's 3D XPoint

Google joins open hardware effort

Rick Merritt

NO RATINGS
LOGIN TO RATE

3

use Int
while G
tandard
rs for di
announ
Compute
a new
prototy
packag

Facebook's support for the 3D XPoint non-volatile me
co-developed by Intel and Micron is "a huge endorser

tom's HARDWARE PRODUCT REVIEWS NEWS DEALS FORUM

Samsung's 10-Year Plan Starts With 128TB QLC SSD, 960 Successor

22 COMMENTS

by Chris Ramseyer August 8, 2017 at 12:30 PM

ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

THE REVOLUTION IS HERE —

Intel at last announces Optane memory: DDR4 that never forgets

New memory offers huge capacities and persistence, but fits in a DDR4 slot.

PETER BRIGHT - 5/30/2018, 8:45 PM

May 18, 2016

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

Tiffany Trader

IBM scientists have broken new ground in the c
change memory technology (PCM) that puts a
XPoint technology from Intel and Micron. IBM s
pre-cy
paper
ternatio
oming
that bri
ge (flas
echnol
recon
e bits p
gies, th
ange
temper
PCM
ift, nois

designlines MEMORY

News & Analysis

3D NAND Flash at 2 Cents per GB

BeSang wants to lower barrier to 3D NAND flash

R. Colin Johnson
7/18/2016 07:10 PM EDT
14 comments

LOGIN TO RATE

Tw

ES, FI
10, Be
superi
trated
3D tec
ps with
b the
er-bit o

chip technology
ms to have since
cture for NAND
mentation of its
door to
as offering to
ice that reduces
ut 2¢ per

SanDisk 1TB Extreme UHS-I ...
$449.99
B&H Photo

46 PM 7,391 VIEWS

nd Micron Jointly Announce Changing 3D XPoint Memory ology

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
of the technologies: process shrinks will come to
off a cliff as a side effect of reducing the size of e

The HP answer to this scaling wall is Memristor,
to have DRAM-like speed and better-than-NAND
in Las Vegas that Memristor devices will be ready
also showed off a Memristor wafer, adding that it

designlines MEMORY

News & Analysis

Samsung Debuts 3D XPoint Killer

3D NAND variant stakes out high-end SSDs

Rick Merritt
8/11/2016 00:01 AM EDT
5 comments

NO RATINGS
1 saves
LOGIN TO RATE

Like 56 Tweet Share 212 G+1 4

SANTA CLARA. Calif. – Samsung lobbed a new variant of its 3D

DESIGNLINES | MEMORY DESIGNLINE

Memory Forecast to Account fo 53% of Semiconductor Capex

By Dylan McGrath, 08.29.18 0

Share Post Share on Facebook Share on Twitter G+ in

SAN FRANCISCO — Capital spending for memory chips is expected to account for 53% c industry capex of $102 billion this year, nearly twice the percentage that memory accou five years ago, according to market research firm IC Insights.

With all NAND flash vendors ramping up 3D NAND capacity, NAND-related capital expe are forecast to total more than $31 billion, 31% of the semiconductor industry total, acc the latest edition of IC Insights' McClean Report. The total for NAND capex would repre increase of 13% over 2017, when NAND flash capex grew by 91%.

Meanwhile, the report forecasts that capital spending for DRAM and SRAM will increase any other industry segment, growing 41% in 2018 after an 82% increase last year. DRAM capex is expected to total $22.9 billion, 22% of the industry-wide total, according to the

Memory's Growing Share of Capital Spending

Source: IC Insights

The forecasted total of $102 billion for the overall semiconductor industry — including l upgrades to existing wafer fab lines and brand new manufacturing facilities — would ma

Language support for NVM:
NVL-C - extending C to support NVM

Design Goals: Familiar programming interface

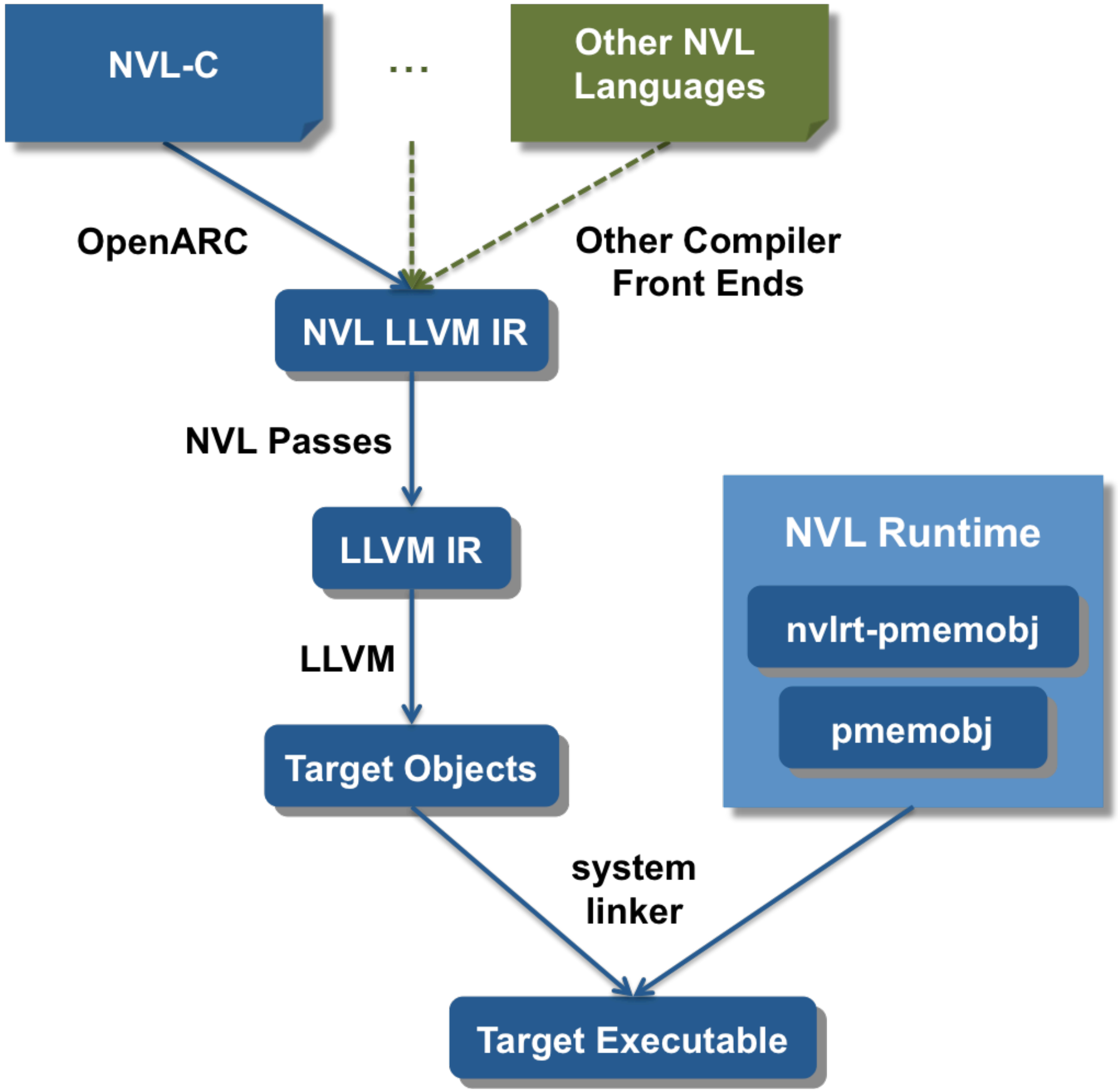
```
#include <nvl.h>
struct list {
  int value;
  nvl struct list *next;
};
void add(int k, nvl struct list *after) {
  nvl struct list *node
    = nvl_alloc_nv(heap, 1, struct list);
  node->value = k;
  node->next  = after->next;
  after->next = node;
}
```

- Small set of C language extensions:
  - Header file
  - Type qualifiers
  - Library API
  - Pragmas
- Existing memory interfaces remain:
  - NVL-C is a superset of C
  - Unqualified types as specified by C
  - Local/global variables stored in volatile memory (DRAM or registers)
  - Use existing C standard libraries for HDD

Design Goals: Avoiding persistent data corruption

- New categories of pointer bugs:
  - Caused by multiple memory types:
    - E.g., pointer from NVM to volatile memory will become dangling pointer
  - Prevented at compile time or run time
- Automatic reference counting:
  - No need to manually free
  - Avoids leaks and dangling pointers
- Transactions:
  - Avoids persistent data corruption across software and hardware failures
- High performance:
  - Performance penalty from memory management, pointer safety, and transactions
  - Compiler-based optimizations
  - Programmer-specified hints

Design Goals: Modular implementation



- Core is common compiler middle-end
- Multiple complier front ends for multiple high-level languages:
  - For now, just OpenARC for NVL-C
- Multiple runtime implementations:
  - For now, just Intel’s pmem (pmemobj)

NVL-C: Programming Model

- 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);
}
```

Denny, J.E., Lee, S., and Vetter, J.S.: 'NVL-C: Static Analysis Techniques for Efficient, Correct Programming of Non-Volatile Main Memory Systems'. Proc. Proceedings of the 25th ACM International Symposium on High-Performance Parallel and Distributed Computing, Kyoto, Japan2016 pp. Pages

Programming Model: NVM Pointers

```
#include <nvl.h>
struct list {
  int value;
  nvl struct list *next;
};
void add(int k, nvl struct list *after) {
  struct list *node
    = malloc(sizeof(struct list));
  node->value = k;
  node->next  = after->next;
  after->next = node;
}
```

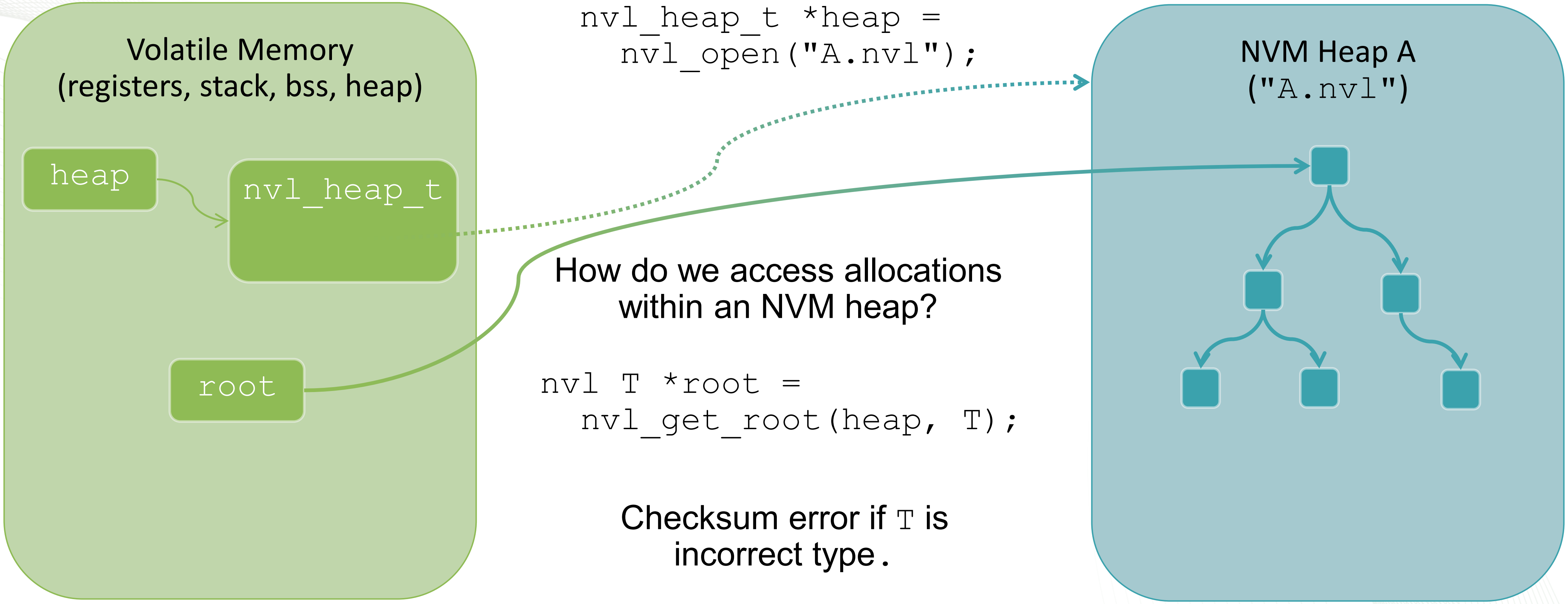
compile-time error
explicit cast won't help

- nvl type qualifier:
  - Indicates NVM storage
  - On target type, declares NVM pointer
  - No NVM-stored local or global variable
- Stricter type safety for NVM pointers:
  - Does not affect other C types
  - Avoids persistent data corruption
  - Facilitates compiler analysis
  - Needed for automatic reference counting
  - E.g., pointer conversions involving NVM pointers are strictly prohibited

Programming Model: Bare NVM Pointers

- NVM pointers are wide:
  - Facilitates: automatic reference counting, pointer constraints, transactions
  - NVM pointers must be decoded for target architecture’s load/store
  - Bare NVM pointer = virtual address with all NVL-C metadata stripped away
- NVM pointer hoisting is important for performance:
  - Conversion to bare NVM pointer is many instructions longer than load/store
  - In tight loop, the performance penalty is severe
  - If conversion is loop-invariant, it can be hoisted
  - Currently, we implement per application with an informal NVL-C extension
  - Future work: eliminate extension and automate in compiler

Programming Model: Accessing NVM



```
nvl_heap_t *heap =
  nvl_open("A.nvl");
```

How do we access allocations within an NVM heap?

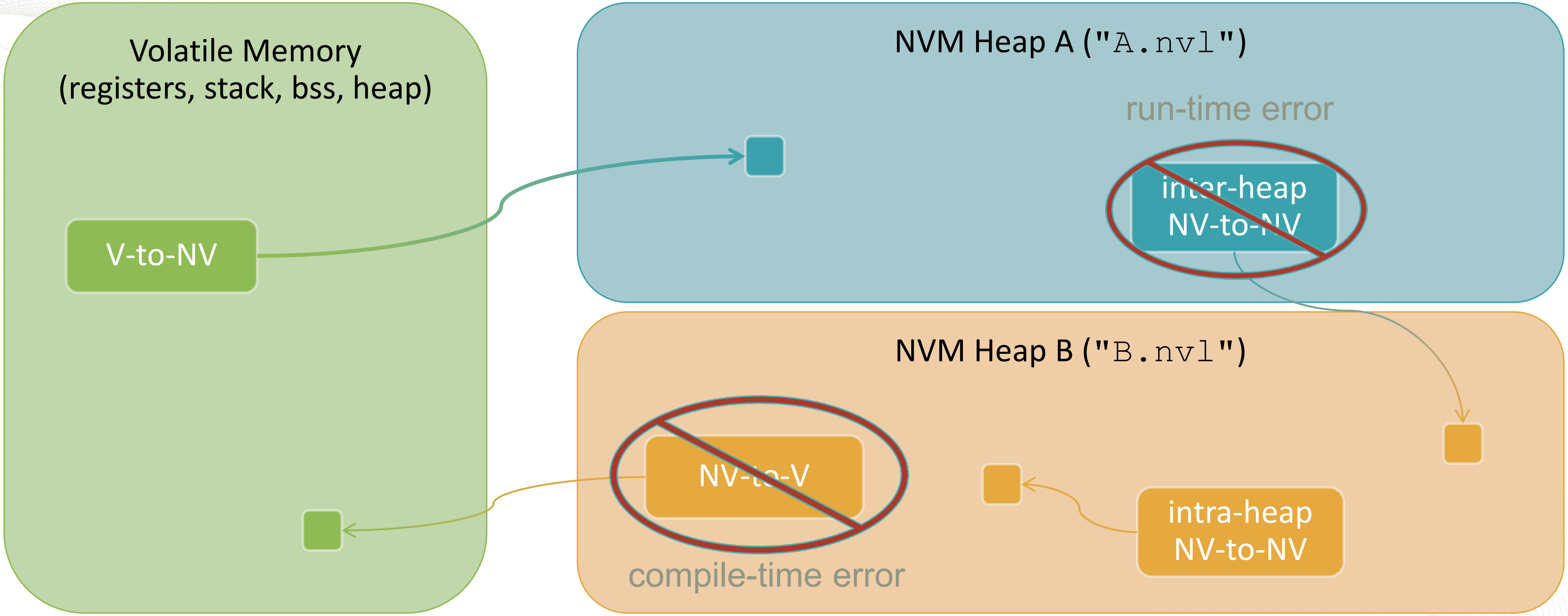
```
nvl T *root =
  nvl_get_root(heap, T);
```

Checksum error if `T` is incorrect type .

Set root with `nvl_set_root`.

Before first `nvl_set_root`, `nvl_get_root` returns null.

Programming Model: Pointer types (like Coburn et al.)



avoids dangling pointers when
memory segments close

Programming Model: Transactions: Purpose

- Ensures data consistency
- Handles unexpected application termination:
  - Hardware failure (e.g., power loss)
  - Application or OS failure (e.g., segmentation fault)
  - NVL-C safety constraint violation (e.g., inter-heap NV-to-NV pointer)
- Does not handle concurrent access to NVM:
  - Future work
  - Concurrency is still possible
  - Programmer must safeguard NVM data from concurrent access

Programming Model: Transactions: Undo logs

```
#include <nvl.h>
void matmul(nvl float a[I][J],
            nvl float b[I][K],
            nvl float c[K][J],
            nvl int *i)
{
  while (*i<I) {
    #pragma nvl atomic heap(heap)
    {
      for (int j=0; j<J; ++j) {
        float sum = 0.0;
        for (int k=0; k<K; ++k)
         sum += b[*i][k] * c[k][j];
        a[*i][j] = sum;
      }
      ++*i;
    }
  }
}
```

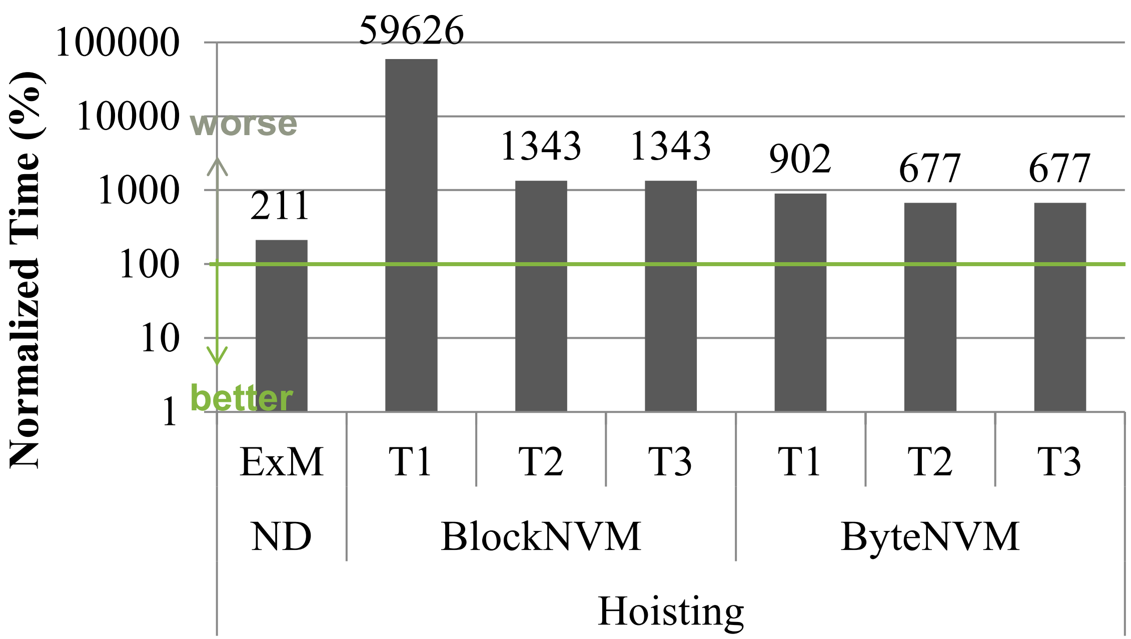
- Before every NVM store, transaction creates undo log to back up old data
- Undo log contains metadata plus old data being overwritten
- Problem: large overhead because an undo log is created for every element of `a` (every iteration of `j` loop)

Programming Model: Transactions: `clobber` clause

```
#include <nvl.h>
void matmul(nvl float a[I][J],
            nvl float b[I][K],
            nvl float c[K][J],
            nvl int *i)
{
  while (*i<I) {
    #pragma nvl atomic heap(heap) \
            clobber(a[*i:1])
    {
      for (int j=0; j<J; ++j) {
        float sum = 0.0;
        for (int k=0; k<K; ++k)
         sum += b[*i][k] * c[k][j];
        a[*i][j] = sum;
      }
      ++*i;
    }
  }
}
```

- **`clobber`** clause suppresses undo logs
- Durability after transaction commit is still guaranteed

Evaluation: LULESH



- ExM = use SSD as extended DRAM
- T1 = BSR + transactions
- T2 = T1 + `backup` clauses
- T3 = T1 + `clobber` clauses
- BlockNVM = `msync` included
- ByteNVM = `msync` suppressed

- **`backup` is important for performance**
- **`clobber` cannot be applied because old data is needed**

NVL-C Summary

- Motivated a new programming model for NVM as persistent memory
- Introduced NVL-C, a new programming system for this purpose
  - First class language construct
  - Transactions
- Described several performance optimizations for NVL-C
- Showed performance results for these optimizations on an SSD
- Working on Optane DIMMs now

Recap

Recent trends in computing paint an ambiguous future for architectures

- Contemporary systems provide evidence that power constraints are driving architectures to change rapidly
- Multiple architectural dimensions are being (dramatically) redesigned: Processors, node design, memory systems, I/O

- **Entering an era of Extreme Heterogeneity**
- **Complexity is our main challenge**

Applications and software systems are all reaching a state of crisis

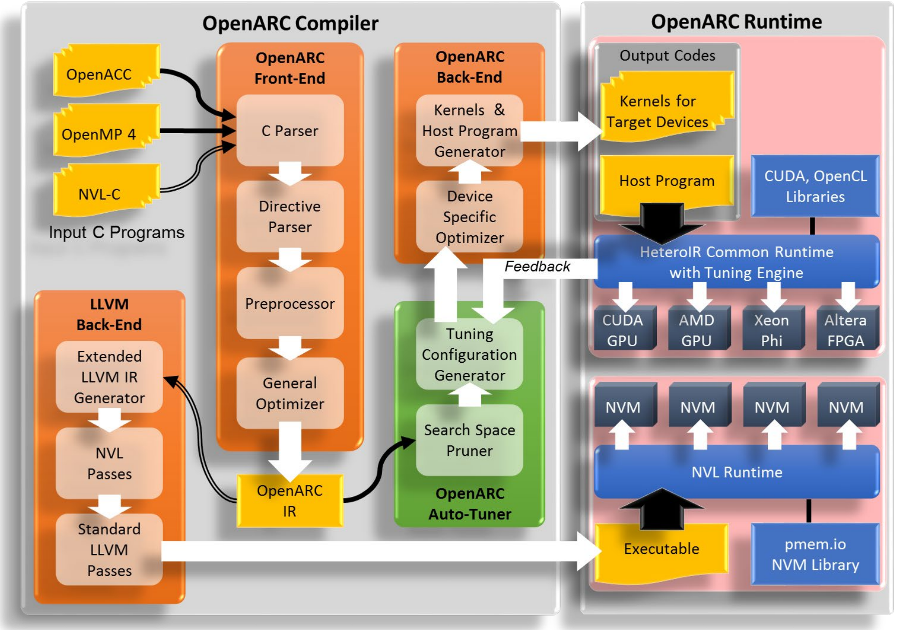
- Applications will not be functionally or performance portable across architectures
- Programming and operating systems need major redesign to address these architectural changes
- Procurements, acceptance testing, and operations of today's new platforms depend on performance prediction and benchmarking.

- **This is a crisis!**

Programming systems must provide performance portability (beyond functional portability)!!

- Strive for 'Write once, run anywhere'
- Descriptive models of parallelism and data movement
- Introspective runtime systems
- Layered, modular, open source approaches required

- **Examples**
  - **ECP investments in LLVM**
    - **FORTRAN with GPU offloading**
  - **Programming FPGAs**
    - **Without Verilog**
  - **Memory systems are changing too**
    - **Language support for NVM**



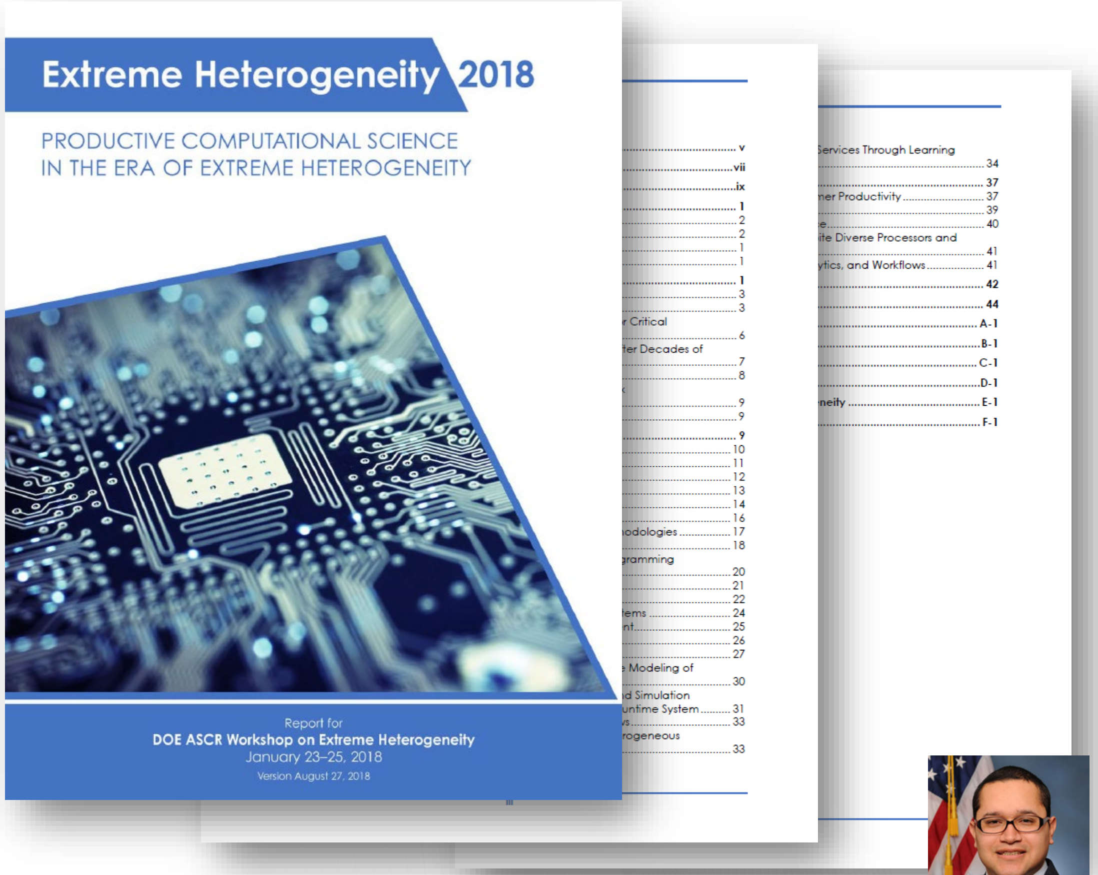
- ~~Visit us~~
  - ~~We host interns and other visitors year round~~
    - ~~Faculty, grad, undergrad, high school, industry~~
- Jobs at ORNL
  - Postdoctoral Research Associate in Computer Science
  - Software Engineer
  - Computer Scientist
  - Visit https://jobs.ornl.gov
- Contact me vetter@ornl.gov

Oak Ridge National Laboratory

Final Report on Workshop on Extreme Heterogeneity

1. Maintaining and improving programmer productivity
   - Flexible, expressive, programming models and languages
   - Intelligent, domain-aware compilers and tools
   - Composition of disparate software components

- Managing resources intelligently
  - Automated methods using introspection and machine learning
  - Optimize for performance, energy efficiency, and availability
- Modeling & predicting performance
  - Evaluate impact of potential system designs and application mappings
  - Model-automated optimization of applications
- Enabling reproducible science despite non-determinism & asynchrony
  - Methods for validation on non-deterministic architectures
  - Detection and mitigation of pervasive faults and errors
- Facilitating Data Management, Analytics, and Workflows
  - Mapping of science workflows to heterogeneous hardware and software services
  - Adapting workflows and services to meet facility-level objectives through learning approaches



Extreme Heterogeneity 2018

PRODUCTIVE COMPUTATIONAL SCIENCE
IN THE ERA OF EXTREME HETEROGENEITY

Report for
DOE ASCR Workshop on Extreme Heterogeneity
January 23–25, 2018
Version August 27, 2018

https://orau.gov/exheterogeneity2018/

https://doi.org/10.2172/1473756