

ARM high performance computing

“Disruptive Technologies”

ARM

David Brash

Architecture Program Director

Architecture & Technology Group

IDC HPC User Forum – Oxford, UK

29th September 2016

Technology disruption

(Rob Aitken: Director of Technology & Fellow, ARM Research)

- Disruptive means significantly changing the roadmap, not extending or tweaking it
 - Disruptive innovation creates new classes of customer
 - Disruptive technology may also eliminate some old customers/markets
- Scale of disruptiveness
 - Context dependent: FinFETs were a major disruption to foundries, much less so to ARM
- Other scaling factors
 - **Likelihood:** Relates to timescale (2, 10, 25 years ...)
 - **Importance:** Does this matter?
To whom and why?
 - **Uncertainty:** Confidence in the predictions and assumptions behind them?

This talk: disruptors – by example

- The world's growing appetite for technology and its underlying evolution
- The ARM business model and development of the ARM architecture
- Key technology changes and the resulting effect(s)
 - Non-Volatile Main Memory (NVMM)
 - blurring the line between memory and storage
 - promising cost effective increases in memory density & close proximity to cores

ARM vision – the architecture for the digital world

Embedded intelligence driving change:

- Efficiency improvements
- The IoT revolution
 - upload/download dynamics
 - device count, ...

Network loads require a wide range of solutions:

- Latency
- Bandwidth
- Hierarchical processing (Edge or *FOG* computing)

Cloud computing driving infrastructure diversity:

- Capacity
- Usage
- Security
- All classes of server
 - through HPC



The scale of device deployment and needs such as computer vision/machine learning across the ecosystem will create new opportunities and change.

Why ARM in HPC?

Energy efficiency

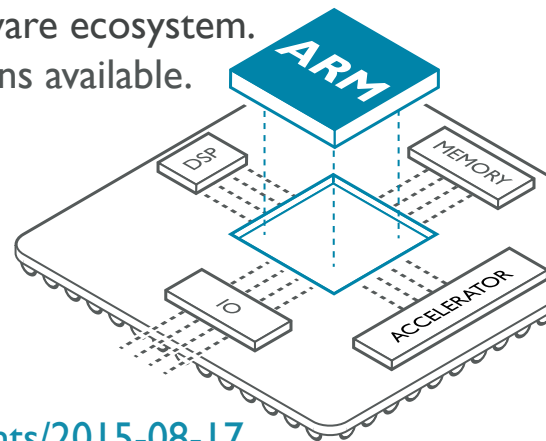
- Energy efficiency in our DNA:
 - a first-class design constraint at ARM
 - our IP goal is to maintain a power-performance efficiency advantage
- The CPU core is only one element:
 - efficiency in memory & on-chip interconnect spaces needed too.

Choice

- Independent silicon providers within a shared software ecosystem.
- Wide range of price/performance/power/size options available.

Customisation

- ARM's partners can build SoCs very quickly.
- Getting interesting in the US:
 - SoC for HPC: <http://www.socforhpc.org>
 - DARPA CRAFT: <http://www.darpa.mil/news-events/2015-08-17>



HPC leadership: International Exascale programs

- **United States**
 - **Data Movement Dominates**
 - **FastForward II**
 - **PathForward (proposals)**



- **Japan**
 - **ARM based Fujitsu Post-K supercomputer will be deployed to RIKEN**
- **China**
 - **Multiple licensees pursuing HPC oriented systems**

- **European Union**
 - **Montblanc 1, 2, & 3**
 - **ExaNode**
 - **Hartree Centre Deployment**
 - **H2020-FET-HPC (proposals)**

ARMv8-A vector processing - introduction

- Advanced SIMD introduced in ARMv7-A (*aka* ARM NEON instructions)
- AArch64 Advanced SIMD was an evolution
 - Gained full IEEE double-precision float and 64-bit integer vector ops
 - Vector register file grew from 16×128b to 32×128b
- New markets for ARMv8-A are demanding more radical changes
 - ✓ Gather load & Scatter store
 - ✓ Per-lane predication
 - ✓ Longer vectors
- But what is the preferred vector length?



ARMv8-A Scalable Vector Extension (SVE)

- There is no preferred vector length
 - Vector Length (VL) is hardware choice, from 128 to 2048 bits, in increments of 128
 - *Vector Length Agnostic (VLA)* programming adjusts dynamically to the available VL
 - No need to recompile, or to rewrite hand-coded SVE assembler or C intrinsics
- SVE is not an extension of Advanced SIMD
 - A separate architectural extension with a new set of A64 instruction encodings
 - Focus is HPC scientific workloads, not media/image processing
- Amdahl says you need high vector utilisation to achieve significant speedups
 - Compilers often unable to vectorize due to intra-vector data & control dependencies
 - SVE begins to address some of the traditional barriers to auto-vectorization

ARM Performance Libraries

Enable the wide variety of ARM cores available today without adding complexity to the software ecosystem.

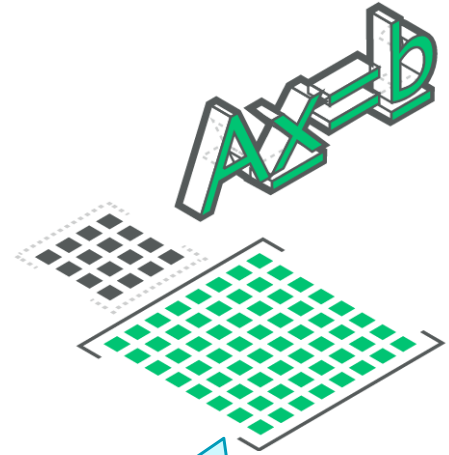
- Commercially supported **64-bit ARMv8-A vendor math libraries** for scientific computing.
- **Built and validated** using technology from the Numerical Algorithms Group (NAG).
- **ARM silicon partners provide us with tuned kernels.**

Capabilities:

- BLAS
- LAPACK
- FFT

Tuned for:

- ARM Cortex-A57,
- Applied Micro X-Gene®
- Cavium® ThunderX



Find out more from the ARM tools group at SC16:

- Salt Lake City, Utah
- 13-18 November 2016
- <http://sc16.supercomputing.org/>

NVMM

- Memory (size and bandwidth) critical to system performance
 - SRAM too expensive (\$'s and area) for large arrays => used for core cache(s)
 - DRAM provides local memory arrays (GB's) => DIMMs
 - NVRAM traditionally too slow => Flash cards, SSDs (persistent storage)
- Persistent memory based on emerging memory technologies
 - phase change (PCM); spin-torque transfer (STTM); memristor
- Impacts architecture and supporting software across the memory hierarchy
 - Point of Unification // unification of I\$ and D\$ caches
 - Point of Coherency // provides access guarantees to all agents in a shareable domain
 - Point of Persistency // provides guarantees for power management too

Developer website : developer.arm.com/hpc

ARM have launched an HPC-specific microsite

This is home to our HPC ecosystem offering:

- technical reference material
- how-to guides
- latest news and updates from partners
- downloads of HPC libraries
- third-party software recommendations
- web forum for community discussion and help



Participate and help drive the community

Thank you

ARM

The trademarks featured in this presentation are registered and/or unregistered trademarks of ARM Limited (or its subsidiaries) in the EU and/or elsewhere. All rights reserved. All other marks featured may be trademarks of their respective owners.

Copyright © 2016 ARM Limited

©ARM 2016

Acknowledgement: Material presented derived from various groups across ARM

Background

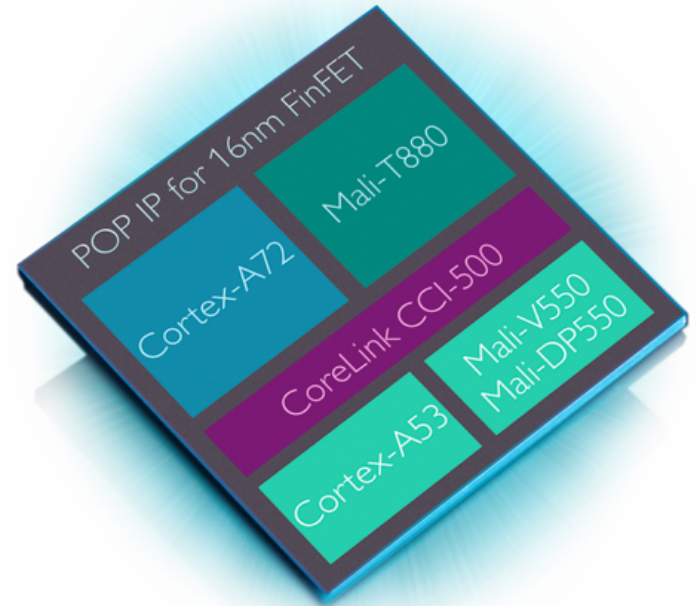
An introduction to ARM

ARM is the world's leading semiconductor intellectual property supplier.

We license to over 440 partners, are present in 95% of smart phones, 80% of digital cameras, 35% of all electronic devices, and more than 80 billion ARM chips have been shipped since 1990.

Our CPU business model:

- License technology to partners, who use it to create their own system-on-chip (SoC) products.
- We may license an ISA (e.g. “ARMv8-A”) or a specific microarchitectural **implementation** (e.g. “Cortex-A72”)



...and our IP extends beyond the CPU

ARM & OpenHPC

Participation:

- Silver member of OpenHPC
- ARM is on OpenHPC TSC in order to drive ARM architecture build support
- OpenHPC defines a clear target baseline of codes, many of which have already ported and are available on ARM. See developer.arm.com/hpc for details.

Status:

- 131 out of 166 packages building (79%)
- Closing on fixing remaining relevant packages and moving towards continuous build integration on multiple ARM platforms

Functional Areas	Components
Base OS	RHEL/CentOS 7.1, SLES 12
Administrative Tools	Conman, Ganglia, Lmod, LosF, ORCM, Nagios, pdsh, prun
Provisioning	Warewulf
Resource Mgmt.	SLURM, Munge, Altair PBS Pro*
I/O Services	Lustre client (community version)
Numerical/Scientific Libraries	Boost, GSL, FFTW, Metis, PETSc, Trilinos, Hypre, SuperLU, Mumps
I/O Libraries	HDF5 (pHDF5), NetCDF (including C++ and Fortran interfaces), Adios
Compiler Families	GNU (gcc, g++, gfortran)
MPI Families	OpenMPI, MVAPICH2
Development Tools	Autotools (autoconf, automake, libtool), Valgrind, R, SciPy/NumPy
Performance Tools	PAPI, Intel IMB, mpiP, pdtoolkit TAU

International standards



OPEN
Compute Project

