arm Research

Arm Research Overview

Stuart Biles

© 2019 Arm Limited

History of Arm

Joint venture between Acorn Computers and Apple

1990

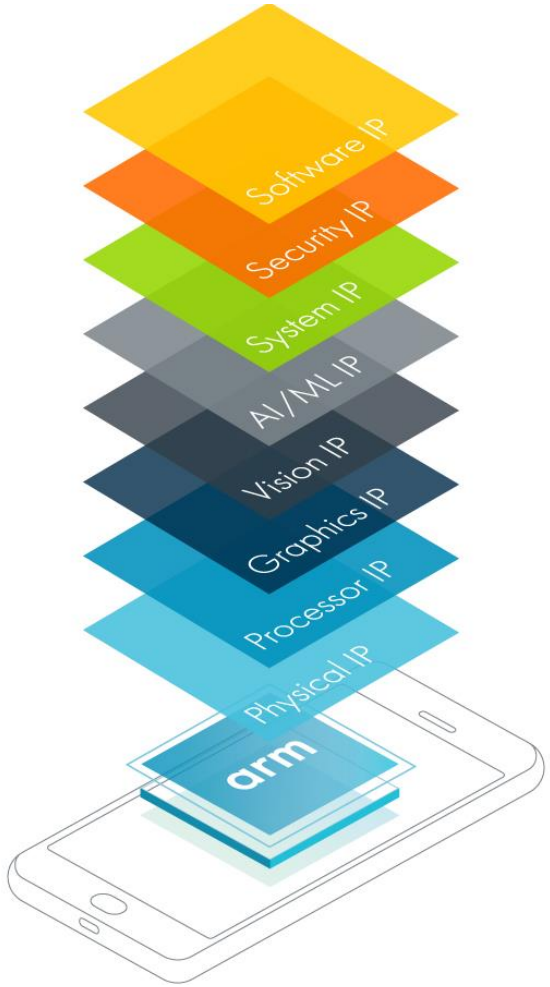
Designed into first mobile phones and then smartphones

1993 onwards

Now all electronic devices and systems can use smart Arm technology

Today

Arm's current business

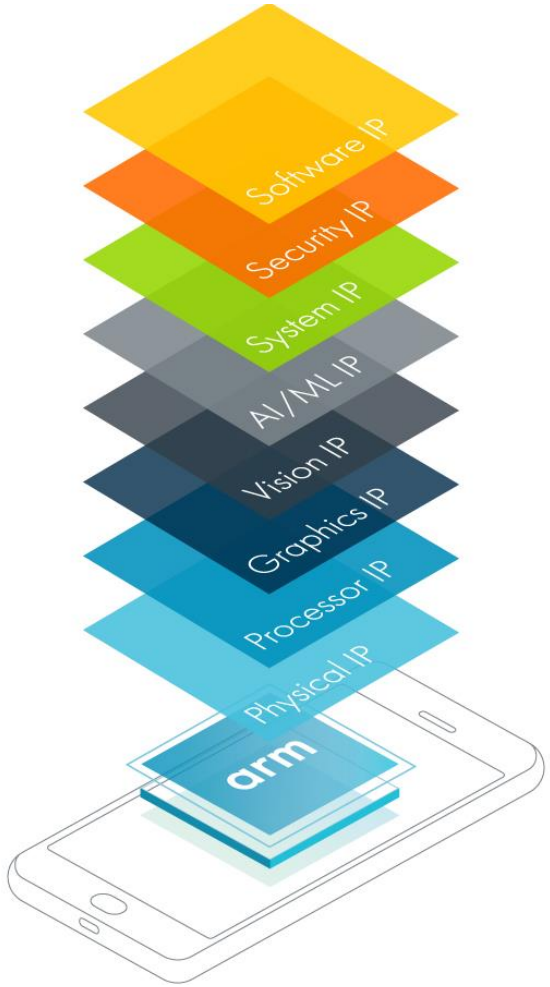


Arm primary business is the development of **intellectual property** (IP) blocks which are used in silicon chips

Our partners combine Arm IP with their own IP to create complete chip designs

We earn **license fees** when we deliver Arm IP to our partners and **royalties** when our partners ship chips that contain Arm IP

Accelerating investment to increase share gains



Generating profits and cash to be reinvested

Investing in new processor technology

- Machine learning processors
- Computer vision
- Augmented reality
- Platform security

Investing in new IoT software and services

- Device Management
- Connectivity as a Service
- Data Management as a Service
- Pelion platform

Introducing Arm Research

Research Collaboration	Research Enablement
Research Programs	Arm Education

arm Research

Education
Collaboration &
Enablement

© 2019 Arm Limited

Arm Education

Helping to close the education and skills gap in Computer Engineering and STEM

Arm University Program

- Free access to academically rigorous education kits using state-of-the-art Arm ecosystem technologies - support Electrical, Electronic and Computer Engineering, and Computer Science curricula
- Email university@arm.com

Arm Education Media

- Enabling learning through rich multimedia online courses and digital-first books allowing students, hobbyists, and engineers to learn, up-skill and become work ready
- Email edumedia@arm.com

Arm School/K-12 Program

- Empowering learners and communities of practice with evidence-based research and curriculum-linked educational materials and platforms - Enable a lifetime positive engagement in STEM
- Email school@arm.com

Arm Research Collaboration Team

arm Research

Help connect people and ideas

Lead outreach and inbound activities

Director US/North America
Raffaella Montelli, PhD.

Facilitate execution of agreements

Director UK/Europe
Andrea Kells, PhD.

Manage collaborations and partnerships

Help identify funding opportunities

Why work with Arm Research?

- **Proven IP, tools, platforms and expertise** to help strengthen the impact of your research
- **Vibrant partner ecosystem** to help you successfully develop and tape out your research products
- **Talent development opportunities** to help your students more easily enter the semiconductor industry
- **Credible industry support** to improve the success of your research proposals
- **Dedicated team** to provide you with what you need to meet your research goals

The Arm Research Ecosystem

Ecosystem Success: Arm High Performance Computing

Mission
Enable the world's first Arm supercomputer(s)

Strategy
Enablement + Co-Design + Partnership

Building Blocks

Enablement

- **Address gaps** in computational capability and data movement within Architecture
- **Seed the software ecosystem** with open source support for Armv8 and SVE libraries, tools, and optimized workloads
- Provide **world class tools** for compilation, analysis, and debug at large scale.

Co-Design

- Work with key end-customers in DoE, DoD, RIKEN, and EU to **design balanced architecture**, uArchitecture and SoCs based on real-world workloads, not benchmarks.
- Develop **simulation and modeling tools** to support co-design development with end-customers, partners, and academia.

Partnership

- Work with Architecture partners to quickly bring **optimized solutions to market**.
- Work with Architecture and uArchitecture design teams to **steer future designs** to be more relevant for HPC, HPDA, and ML
- Work with key ISVs to enable mid-market

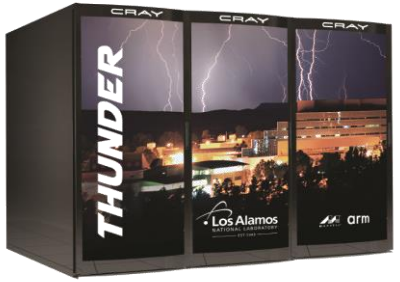
High Performance Computing Research

Dibona (MontBlanc)

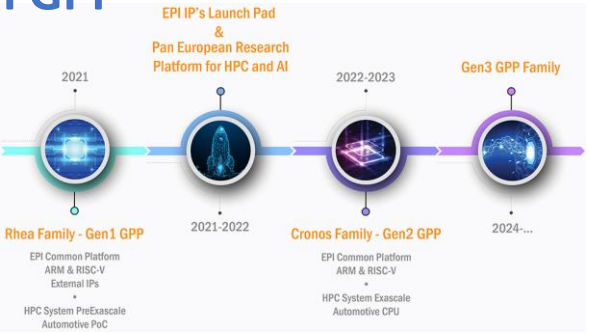
Astra (SANDIA)

Fugaku (RIKEN CCS)

Thunder (LANL)



EPI GPP



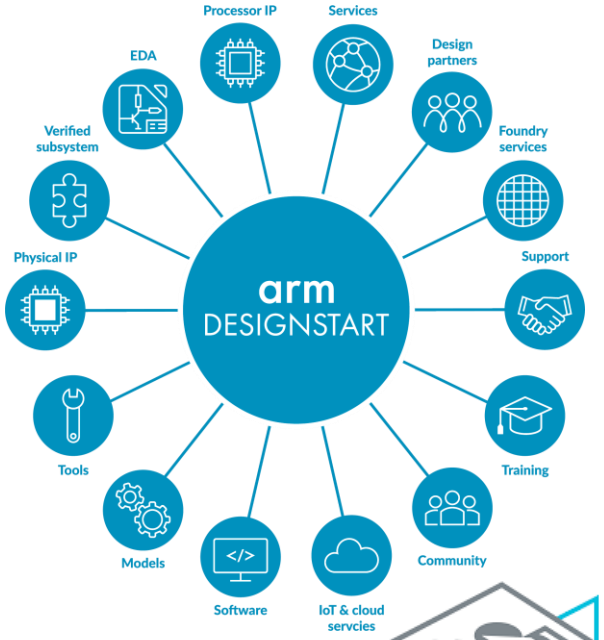
Isambard (U. Bristol)

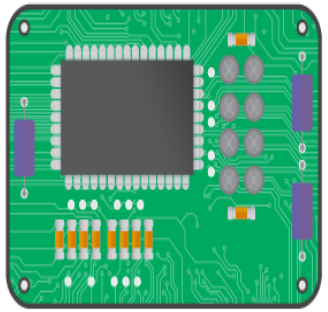
Arm Research Enablement Activities

- Arm has always provided our technology for research purposes under various schemes

- Research Enablement team focus on
  - Broadening the technology portfolio available to researchers
  - Enhancing your experience working with Arm
  - Building an academic, collaborative ecosystem focused on Arm SoCs

- Go to "Your Research Enabled by Arm" workshop to find out details about our existing and new enablement initiatives
  - Tuesday, 11:00- 17:30, Tannehill





Arm Flexible Access for Research

Wider range of IP

- Access to a wide range of Arm IP for research purposes
- Complete RTL for cores and subsystems

Quicker and easier process

- Sped-up approval process from several weeks or months to **two days**

Simplified legal framework

- Reduced friction with **simple** click-through End-User License Agreement (EULA)

To be available to academic researchers in early 2020

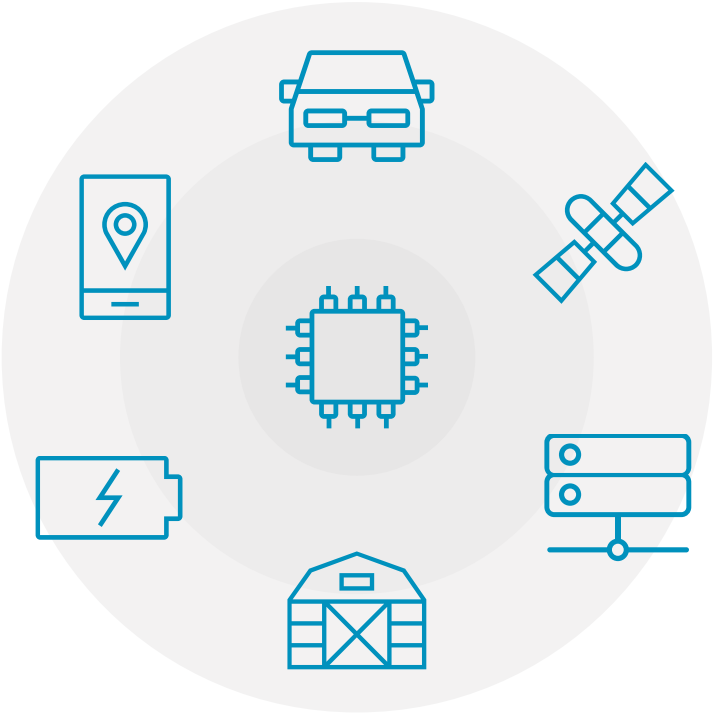
arm Research

Research Programs

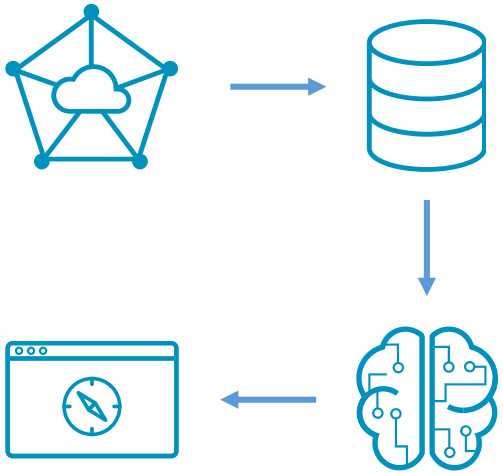
© 2019 Arm Limited

Ever-expanding roles of compute and data drive Arm’s opportunity

Arm everywhere

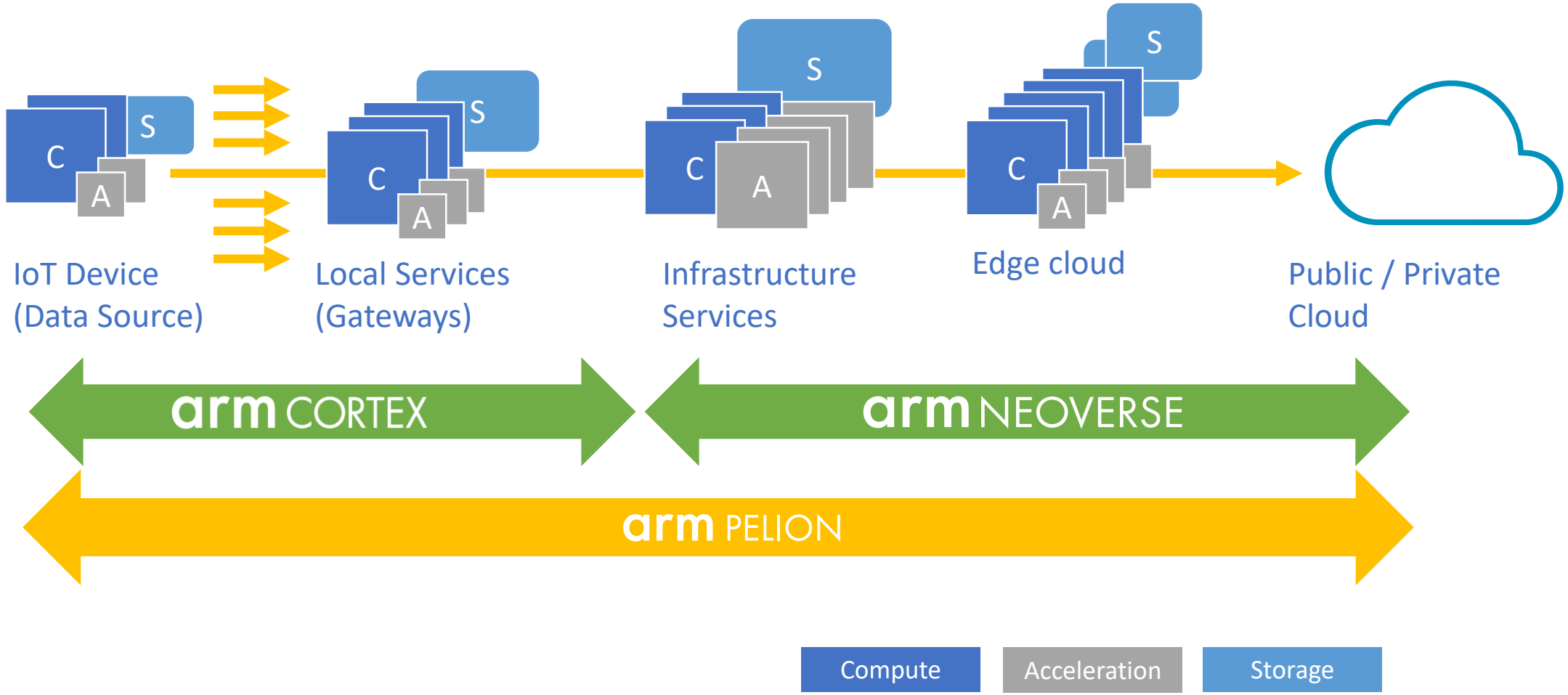


Value from data



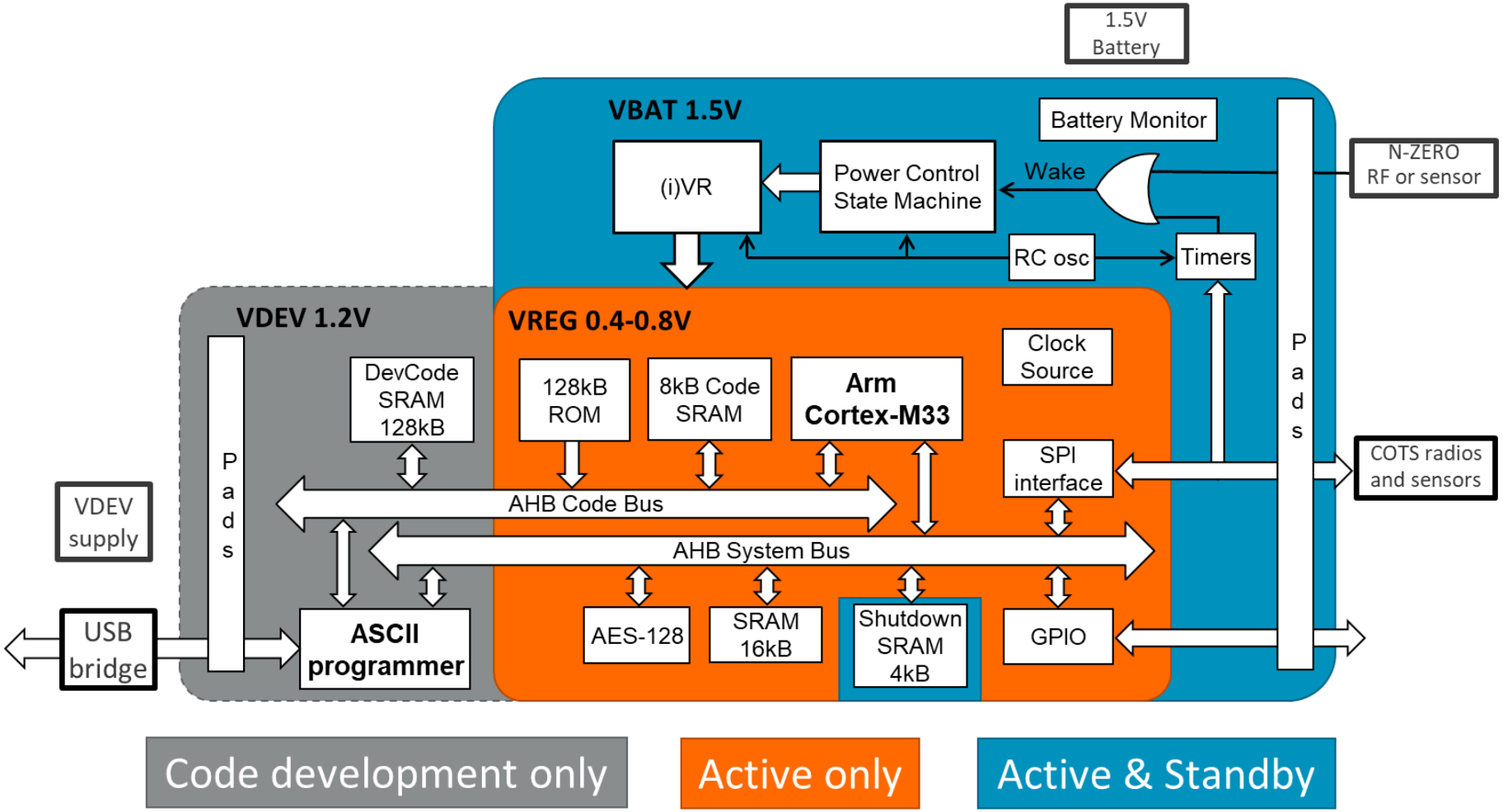
Arm is uniquely positioned in “device-to-cloud”

Edge computing – necessary to handle the massive upstream data



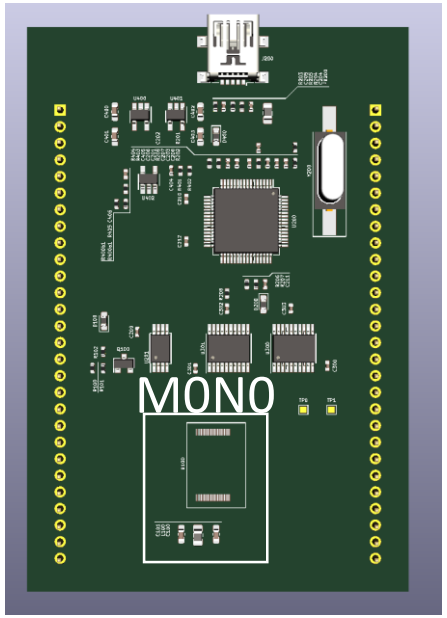
M0N0 Full System Diagram

Evolved to Cortex-M33 with ROM, RAM, peripherals, power management and clocking



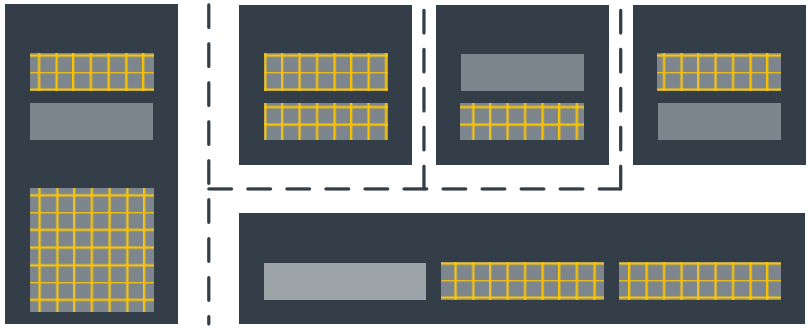
M0N0 is a World Leading Ultra Low Power MCU system

- At least 6x lower shutdown power than 16/32b COTS MCUs
- At least 40% better active energy efficiency than COTS MCUs
- Great target for TinyML applications
  - Cortex-M33 with 4x8b SIMD
- Always-on keyword spotting for 200 days on coincell (158mAh 1.5V)
  - M0N0 chip only - excludes microphone/ADC
- Potential for reduced PCB footprint
  - 6mm^2 chip plus only three passives



SW DevBoard

SSITH – Capabilities for Cortex-M



- Evaluating CHERI technology on Cortex-M
  - Fine grained protection within compartments
- Pointers → Capabilities
  - Permissions bits
    - Data vs code pointer
    - Read only vs. read/write pointer
  - Base and bounds information
- Hidden 1-bit tag makes Capabilities unforgeable

~70%

Of security vulnerabilities due to memory safety

- Protects against a wide range of vulnerabilities
  - E.g. Buffer overflows, return orientated programming, and confused deputy attacks
  - Addresses core memory safety issue, even on legacy languages (e.g. C/C++)
- Low software porting effort
  - Especially for well written code (compiler does most/all of the work)

TinyML

What is it?

- ”Swimming in sensors, drowning in data”
- Model design and optimization for highly constrained hardware platforms
- Can we get 10X+ reduction in ops or memory with minimal accuracy loss?

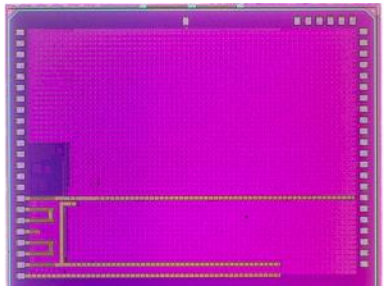
Near term results

- Hybrid neural and non-neural techniques
- New training approaches for binary/ternary networks
- Compression techniques for recurrent neural networks (RNNs) that operate on time-series data

BBC Micro:Bit (Arm Cortex M0, 16KB RAM)

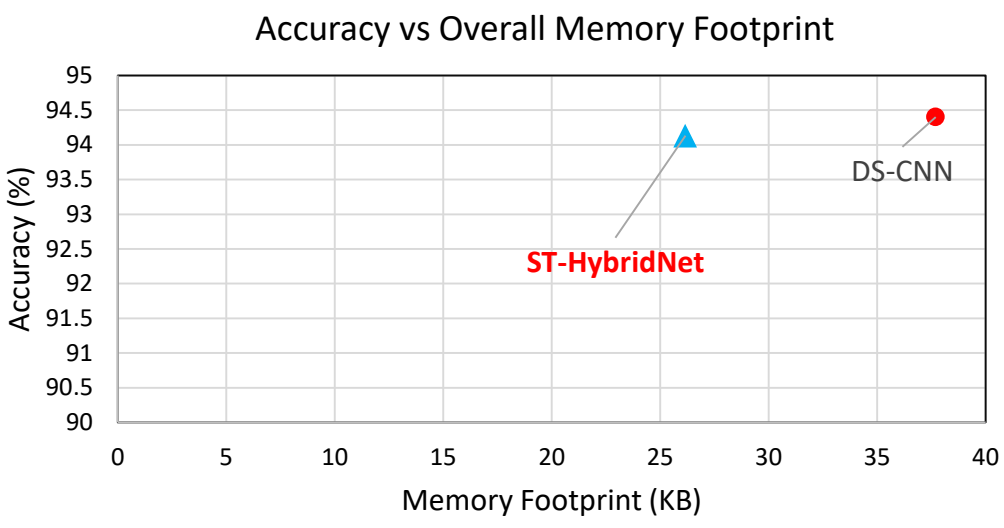
LPCXpresso 1125 (Arm Cortex M0, 8KB SRAM)

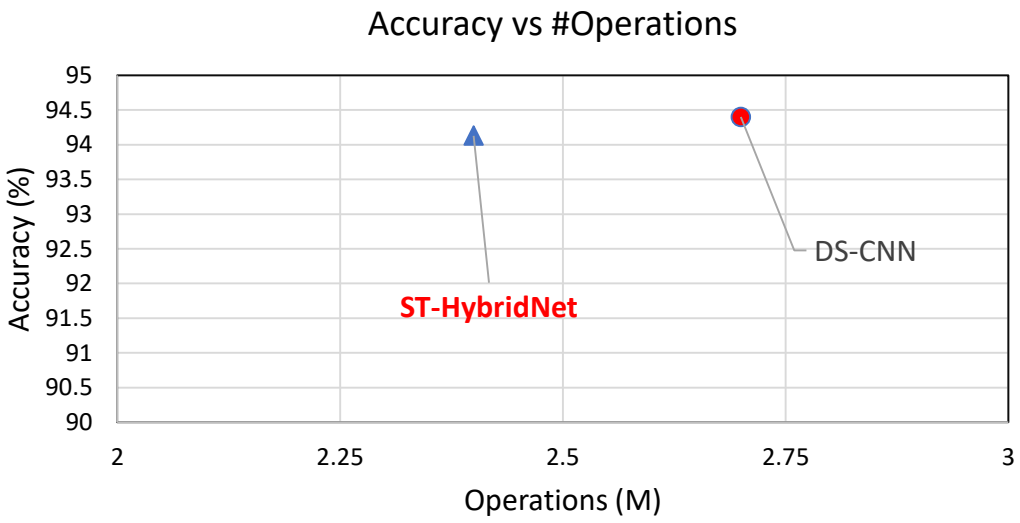
M0N0 (Arm Cortex M33, 16KB SRAM)



TinyML: HybridNet

- "DS-CNN" is a highly optimized network for the key word spotting (KWS) task
  - How do we optimize it further at iso-accuracy?
- Ternarize weight values using Strassen's algorithm
  - Overall memory footprint reduced by 30%
- Selectively use decision trees to reduce compute
  - Total number of operations reduced by 12%
- Less than 0.3% loss in accuracy for these savings





Published in SysML'19 - https://arxiv.org/abs/1903.01531

AutoBot

What is it?

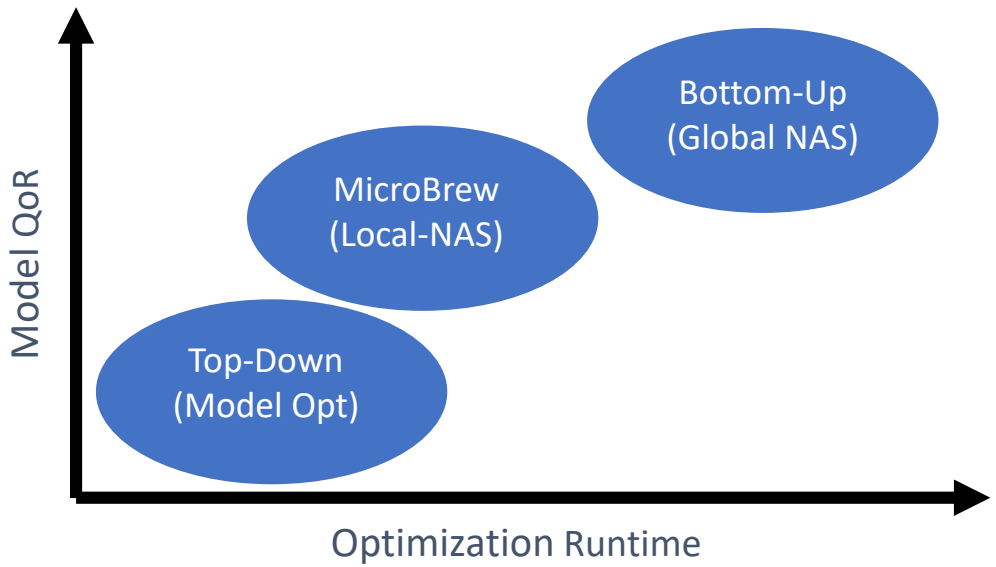
- Automate Neural Architecture Search (NAS) on Arm
- Incorporate information about Arm hardware into the optimization flow
- Reduce search runtime

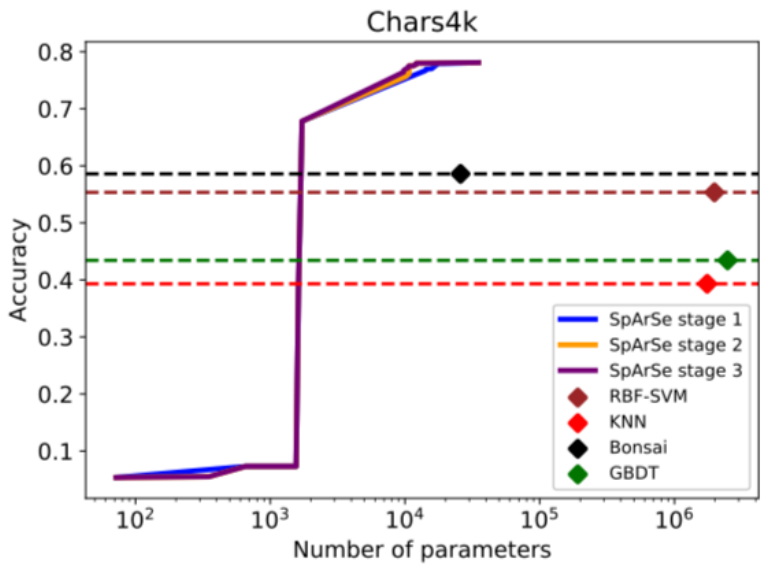
Near term goal: Top-Down (Optimization)

1. Input a trained model
2. Optimize for Arm IP – reduce latency/energy at iso-accuracy

Long term goal: Bottom-Up (Design)

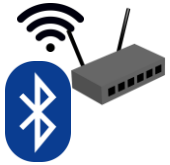
1. Input a dataset
2. Create a from-scratch model optimized for Arm IP





Submitted to NeurIPS'19 - https://arxiv.org/abs/1905.12107

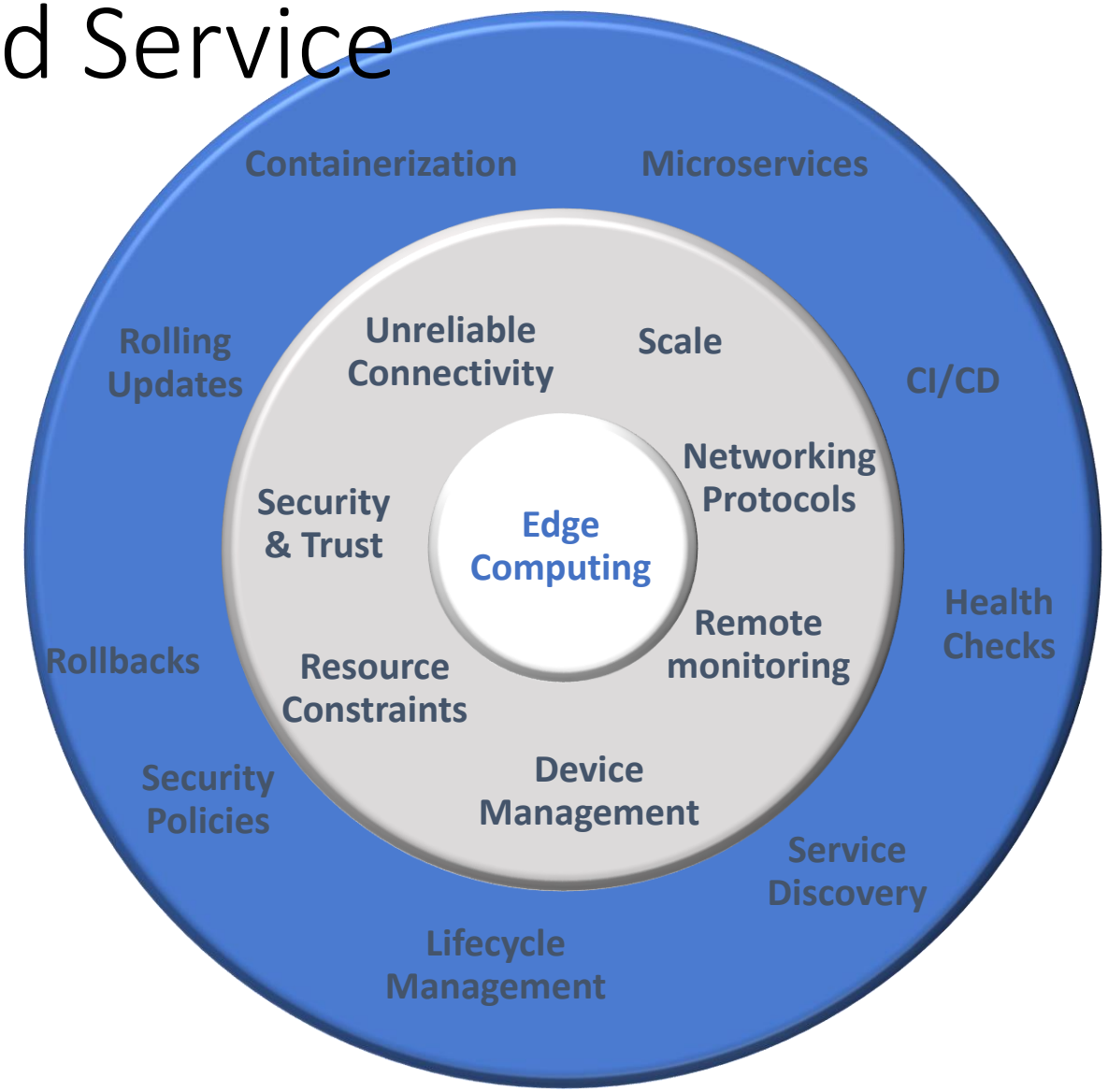
Edge Computing Cloud Service



Designed and optimized for IoT and edge

Brings server style modern application management to edge computing

Based on Kubernetes and compatible with open source ecosystem, APIs, and tool chains



© 2019 Arm Limited

Research Focus Areas

Computer Systems

- Computer system architecture
- Application and domain specialization
- System and circuit implementation

Emerging Technologies & Materials

- Post-Moore materials
- Novel memory technologies
- New frontiers of compute

Security

- Threat intelligence
- Isolation and control mechanisms
- Specification, testing and proof

Machine Learning

- Best in class ML
- Big data analytics
- Applied ML

Distributed & Information Systems

- Device commissioning and management
- Edge computing and services
- Data services

Disruptive Technology

- Big paradigm shifts
- Fundamental enabling technologies
- Tracking and analysing

Careers at Arm

Join the team!

Internships

Cloud Services, Silicon Technology, Machine Learning, Processor and Memory System Architecture, Security, IoT Services, Emerging Technologies, Large Scale Systems, Digital Subsystems Design

Established

Security

Principal Security Research Engineer, Staff Security Research Engineer, Senior Security Research Engineer

Machine Learning

Senior Machine Learning Research Scientist, Machine Learning Researcher

Other

Data Systems Researcher, Senior Research Engineer, Staff Research Engineer, Staff Research Tools Engineer

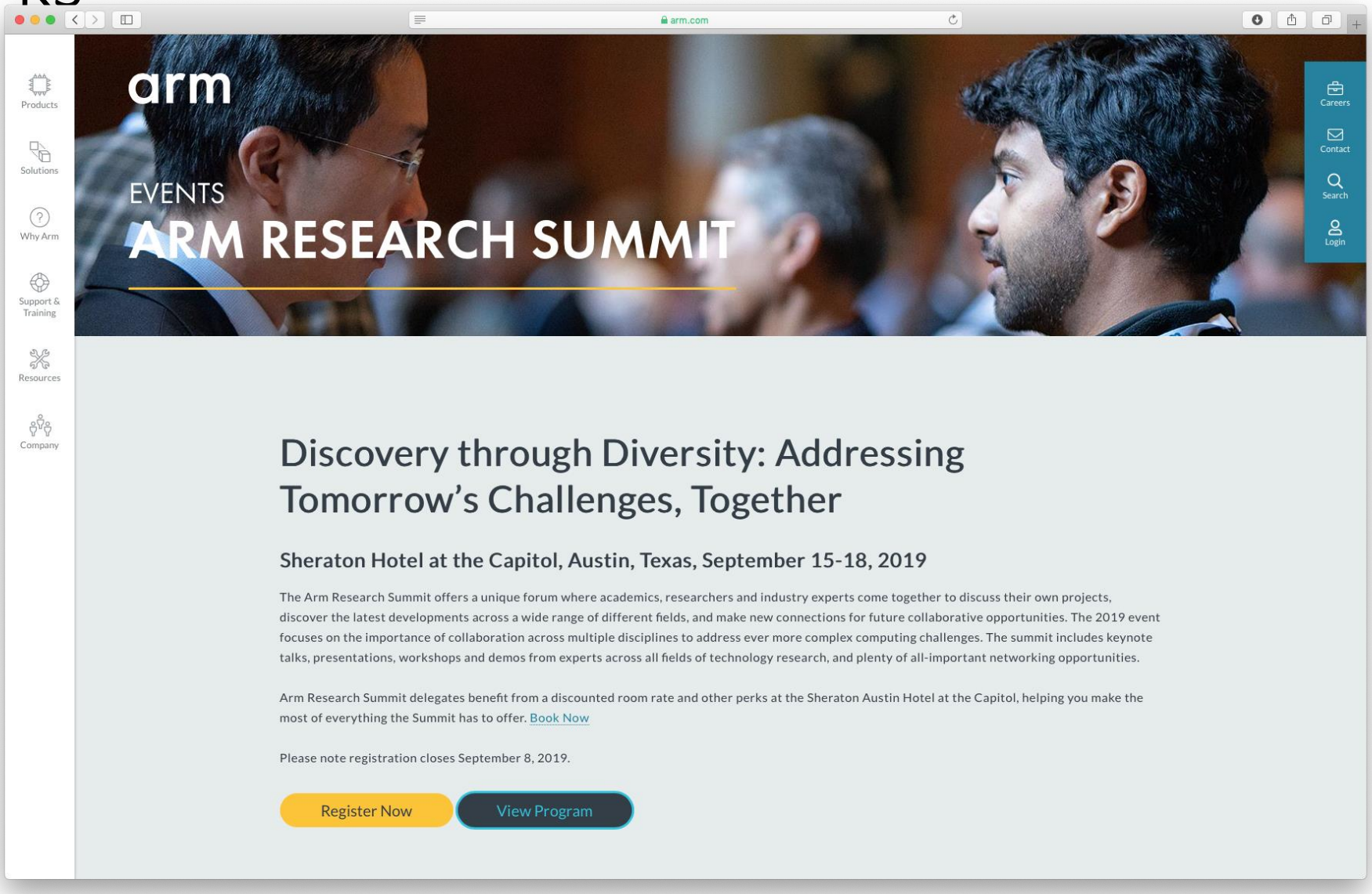
bit.ly/armrshcareers

More Arm Careers

Many more opportunities across Arm beyond Research

arm.com/careers

Closing Remarks



arm Research

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

www.arm.com/company/policies/trademarks

© 2019 Arm Limited