INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS

2016 EDITION

SYSTEM AND ARCHITECTURE WHITE PAPER

PRELIMINARY REPORT

SYSTEM AND ARCHITECTURE

1. CHARTER

Systems and Architectures FT is to establish a top down system driven roadmapping framework for key market drivers of the semiconductor industry. SA uses the cross matrix map of application(s) to market drivers provided by the AB FT. For each market driver, SA identifies and roadmaps one or more architectures vs. key metrics.

2. MISSION

The health of the Information Technologies (IT) sector requires computer systems that can run applications important to society with improving function, speed, or cost over time. These systems have traditionally been based on the von Neumann computer architecture, but other architectures are expected to offer greater benefit in the future.

3. SCOPE

The group takes responsibility for roadmapping systems and architectures that are in production and others that are at the research stage. Production architectures at the time of the group's founding comprised smartphones, [micro] servers, and Internet of Things (IoT) gadgets. The previous architectures are expected to be augmented by some additional architectures now in production, such as FPGAs and GPUs. Research systems and architectures include neural network systems, specialized architectures for novel devices such as spintronic devices, probabilistic devices, superconducting devices, and quantum computers.

The group's scope will include hardware and some software. An application's resource requirements are usually most evident from an analysis of the application's software under scaling. However, this group is not intending to roadmap the development of algorithms or software methods like computer languages.

Aside from picking what is to be roadmapped, the roadmap for system architecture will need to cover the right set of properties and metrics. In the near term, the scope should be issues pertinent to manufacturing. For example, bus width, processor throughput, and amount of memory per system. For the long term, the scope shifts to potential under scaling. To use image recognition as an example, long-term issues would be minimum energy and time per recognition task as a function of image recognition accuracy.

4. CROSS TEAM INTERACTIONS

This group expects to get a list of applications from the Applications Benchmarking group. The group will roadmap architectures built with components in the roadmaps of Beyond CMOS, Heterogeneous MicroSystems, More Moore, and Outside System Connectivity.

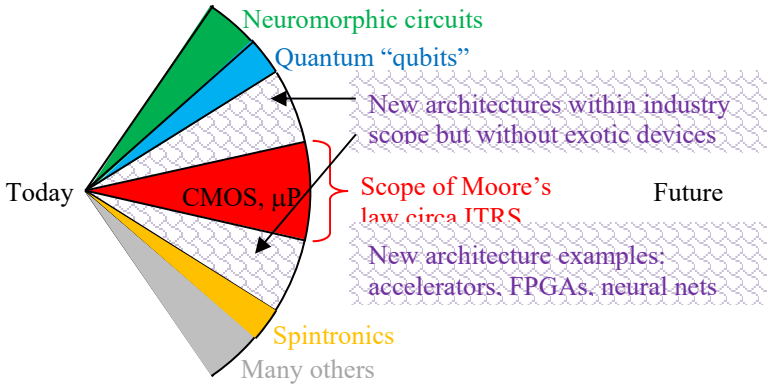
5. STAKEHOLDERS

SA roadmaps should provide an objective, informative resources for computer manufacturers, users, researchers, and research funding organizations.

6. TECHNOLOGY STATUS, NEW REQUIREMENTS AND POTENTIAL SOLUTIONS

The section below has been created as a proposal for discussion of a future consensus direction.

The figure below proposes the form of the top-level roadmap. The previous SI group charter was tied to CMOS and the μP, although subdivided into smartphone, server, and IoT. This segment is shown as the red "narrow" slice of the pie moving left-to-right or forward in time.



Many organizations are proposing remedies to "Moore's law is ending" based on new device physics. Representative new devices include neuromorphic circuits, quantum "qubits," spintronics, and many others. These new devices represent a significant broadening of the search space over the previous focus on CMOS and the μP and are shown with the widest deviation from the current path. The organizations include IRDS's Emerging Research Devices now Beyond CMOS focus area and US Government research funding agencies like DARPA, OSTP, IARPA, etc.

While architectures appropriate for all the device physics in the previous two paragraphs are within the scope of systems architecture, the hatched region is another area not covered by any other IRDS interest group. Other IRDS documents [Paolo's viewgraphs – need to cite] take the position that Moore's law hasn't ended but has merely shifted into a new realm that will require advances in computer architecture to be fully exploited. As an example of this architectural shift, just about every desktop and smartphone manufactured today includes a GPU, essentially forming a counter example to the widely cited claim that nearly all computers are based on the von Neumann architecture. The hatched area could be defined as new non-von Neumann architectures suitable for implementation in CMOS, perhaps with 3D memory. These are often called accelerators

.

6.1. Technology Status and Update

The table below is a top-level summary of the previous System Integration group status, summarizing key issues for the smartphone, [micro] server, and IoT drivers.

Market drivers	Technology focus	Scaling focus	Specific technologies
Mobile Driver	Form factor Power, thermal Performance Features	Cost	Multi-Vt 3D integration Unified logic/RF Advanced DRAM (HBM, HMC) Integrated multi-standard comm. Circuit Sensor/MEMS/logic integration
Microserver Driver	Latency Bandwidth Off-MPU bandwidth	IO device Leakage 2.5D/3D integration Memory BW/latency	DRAM, eDRAM, MRAM, RRAM Silicon photonics (incl. silicon compatible laser source) High-radix networks Distributed compression/encryption engines Novel memory devices 64-bit ARM core Modularized 3D stacks
Datacenter Driver	Latency Bandwidth Off-MPU bandwidth Power density	Performance Reliability Memory BW/latency	
IoT Driver	Form factor Power, thermal Energy harvesting	Specialty devices at baseline tech SiP miniaturization	Vt control eFLASH, HV, MRAM, RRAM 3D integration On-chip passive components Configurable/fine-grain regulation New computation paradigm (near-threshold, asynchronous, stochastic, approximate)

6.2. New Requirements & Difficult Challenges

The shift to the new IRDS structure expands the scope of the previous System Integration group to include Systems Architecture. Some new requirements will originate with the Applications Benchmarking group, which is currently based on the table below.

*Initial AB focus team's application areas are used across market drivers that the systems and architectures (SA) team identified.**										
Application area	***Medical diagnosis***	***Bioinfomatics***	***Medical device***	***IoT edge devices***	***Cloud***	***Big data***	***Robotics***	***CPSs***	***Smartphones***	***Automotive***
Big data analytics	G	G			X	G				
Feature recognition	G	X	P	X	X	X	G, P	P	P	P
Discrete event simulation					X					
Physical system simulation	X				X					
Optimization	X	X	P		X		G, P	P		
Graphics rendering	X				X				P	P
Media processing	X	G	P	X	X		X	P	G	P
Cryptographic codec	X	X	G,P	G, P	X	X	X	G. P	G, P	G, P

Definitions of X, G, P TODO.

The SA group's plan will be to identify the best implementation approach for each application area in the first column, possibly selecting an existing architecture like von Neumann or non-von Neumann or possibly selecting a research approach such as neuromorphic or quantum. This plan will be limited by resources, so it seems likely that the group will produce detailed analyses for only a few top options.

6.3 POTENTIAL SOLUTIONS

A tentative list of systems and architectures appears below:

Systems	
Smartphone	Mobile, battery operated handheld device with various sensors. Road mapping evolving requirements and expected performance.
Server [micro server?]	Stationary computer like a desktop, Web server, and often based on ideas originating from supercomputers. Roadmapping evolving requirements and expected performance.
IoT	Small embedded devices generally standalone (no user interface). Roadmapping evolving requirements and expected performance.
Quantum computer?	Irrespective of commercial interest or non-interest, a quantum computer represents a new computer system since it would use distinct hardware, software, and algorithms.
Architectures in production	
CPU	Projections of architectural evolution and performance expectations. Includes impact of improved transistors like TFETs, 3D stacked/layered memory, and photonic interconnect.
GPU	As traditionally defined (non-traditional designs covered elsewhere). Projections of architectural evolution and performance expectations. Includes impact of 3D stacked/layered memory and photonic interconnect.
FPGA	As traditionally defined.
New architectures not requiring new devices	
Architectures accommodating or exploiting probabilistic effects	Approximate, stochastic, Redundant Residue Number Systems (RRNS)
New non-von Neumann architectures	Example: Knight's [location] (Intel)
New neural network architectures	TrueNorth (IBM), Knight's Mill (Intel), others
Accelerators	A generalization of some of the above: Perhaps defined as computationally powerful elements but not necessarily general purpose. Example: Accelerators for cryptography, image processing, mathematics, etc. All the previous possibly combined with memory to be "processor-in-memory."
New computing approaches with both new devices and new architectures	
Neuromorphic	Neural networks implemented by new circuitry (Memristors, PCM). Roadmaps for functional diversity and energy efficiency limits (including in comparison to digital approaches)
Quantum	Computers based on qubits. Goes with quantum computer driver. Probably requires superconducting electronics (below).
Superconducting electronics	Josephson junction-based computer designs
Spintronics and other physical effects	Spin logic per Nikonov and Young study; coupled oscillators; photon interference effects

7. SUMMARY

TBD