IEEE Rebooting Computing Initiative & International Roadmap of Devices and Systems

Tom Conte, 2015 IEEE Computer Society President,
Co-Chair, IEEE Rebooting Computing Initiative,
Schools of CS & ECE, Georgia Institute of Technology

What the problem is

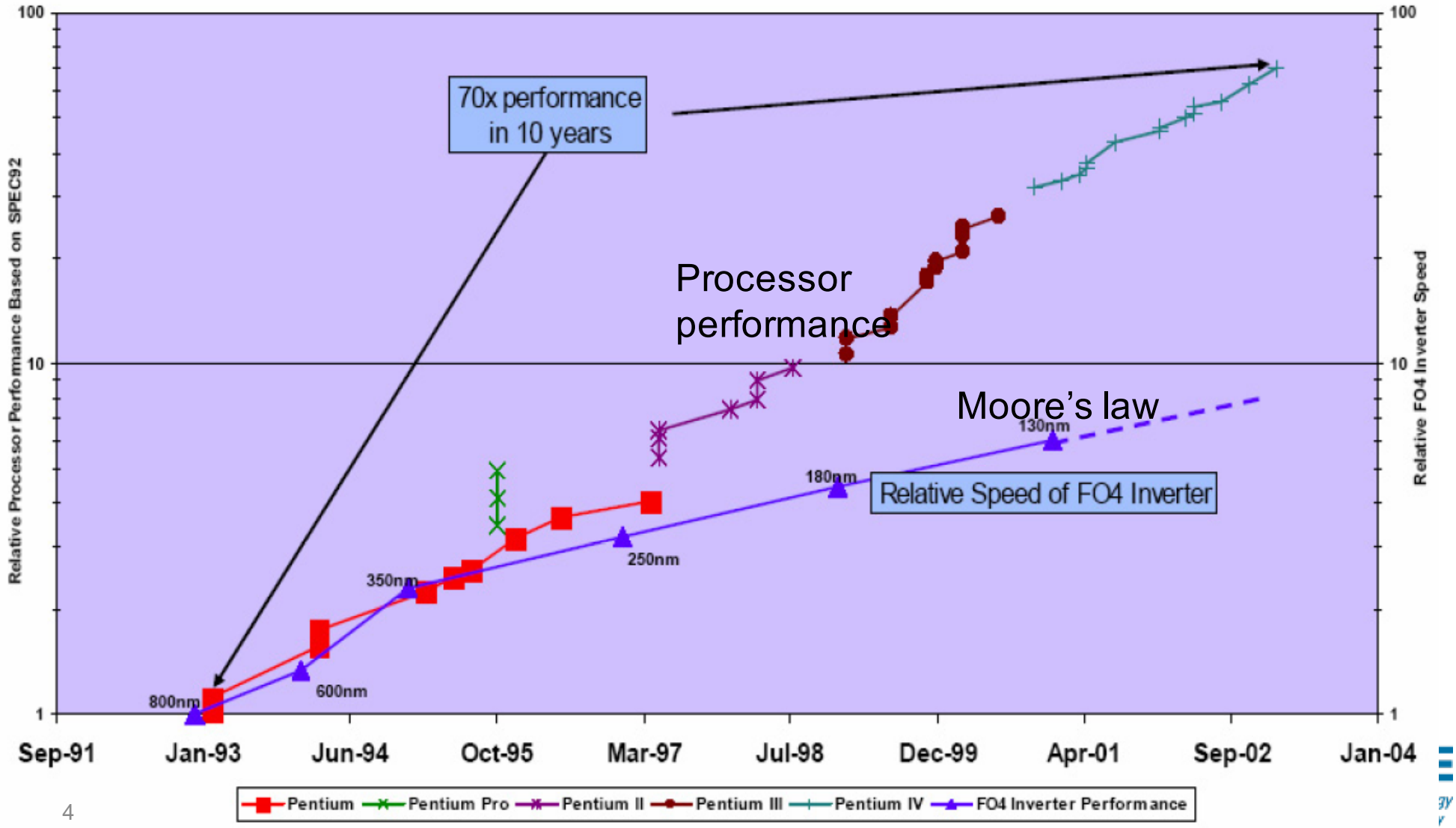
- Transistors are getting smaller *but not faster*
  - From a microarchitect's perspective:

    10nm isn't any better than 14nm, which was only marginally better than 22nm
- Moore's Law for 2D *really* ends in 2021
- Single thread exponential performance scaling ended in 2005
  - *Multicore didn't continue scaling*

How we got here

- 1945: Von Neumann's report describing computer arch.
- 1965: Software industry begins (IBM 360), Moore #1
- 1975: Moore's Law update; Dennard's geo. scaling rule
- 1985: "Killer micros": HPC, general-purpose hitch a ride on Moore's law
- 1995: Slowdown in CMOS logic: superscalar era *begins*

Superscalar to the rescue



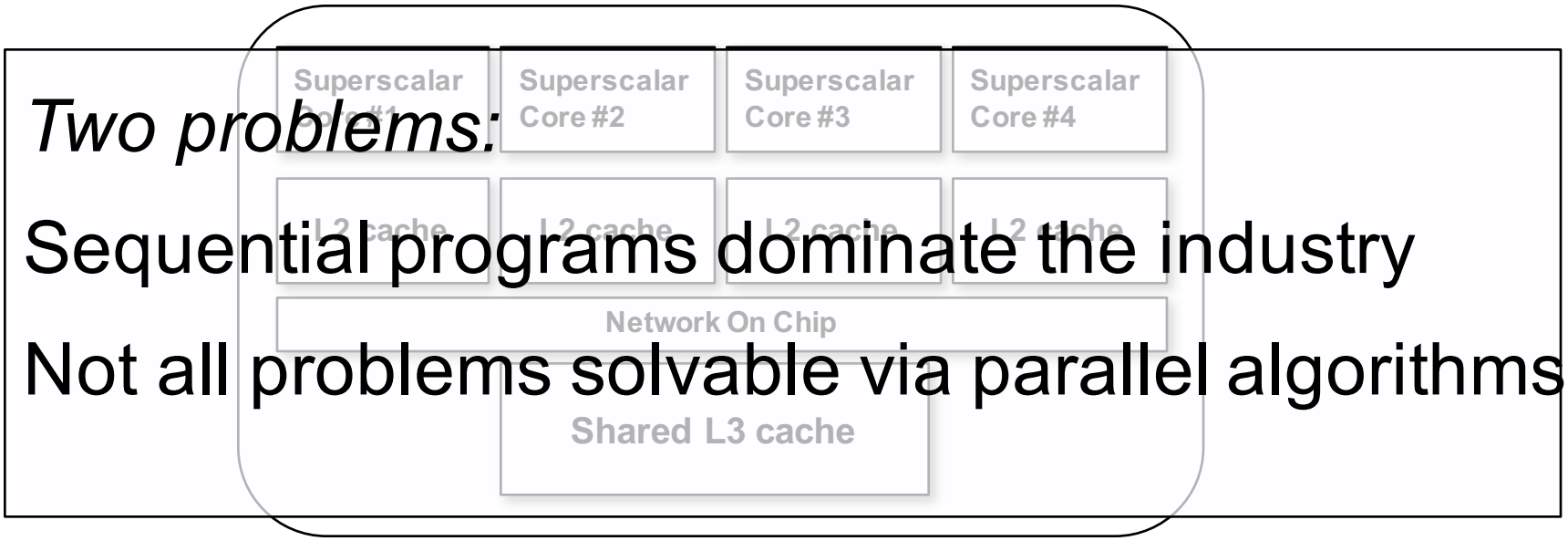
Source: Sanjay Patel, UIUC (used with permission)

How we got here, part 2

- 1945: Von Neumann's report describing computer arch.
- 1965: Software industry begins (IBM 360)
- 1975: Moore's Law; Dennard's geometric scaling rule
- 1985: "Killer micros": HPC, general-purpose hitch a ride on Moore's law
- 1995: Slowdown in CMOS logic: superscalar era *begins*
- 2005: The Power Wall: Single thread exponential scaling *ends* (Intel Prescott) …

What did we do next?

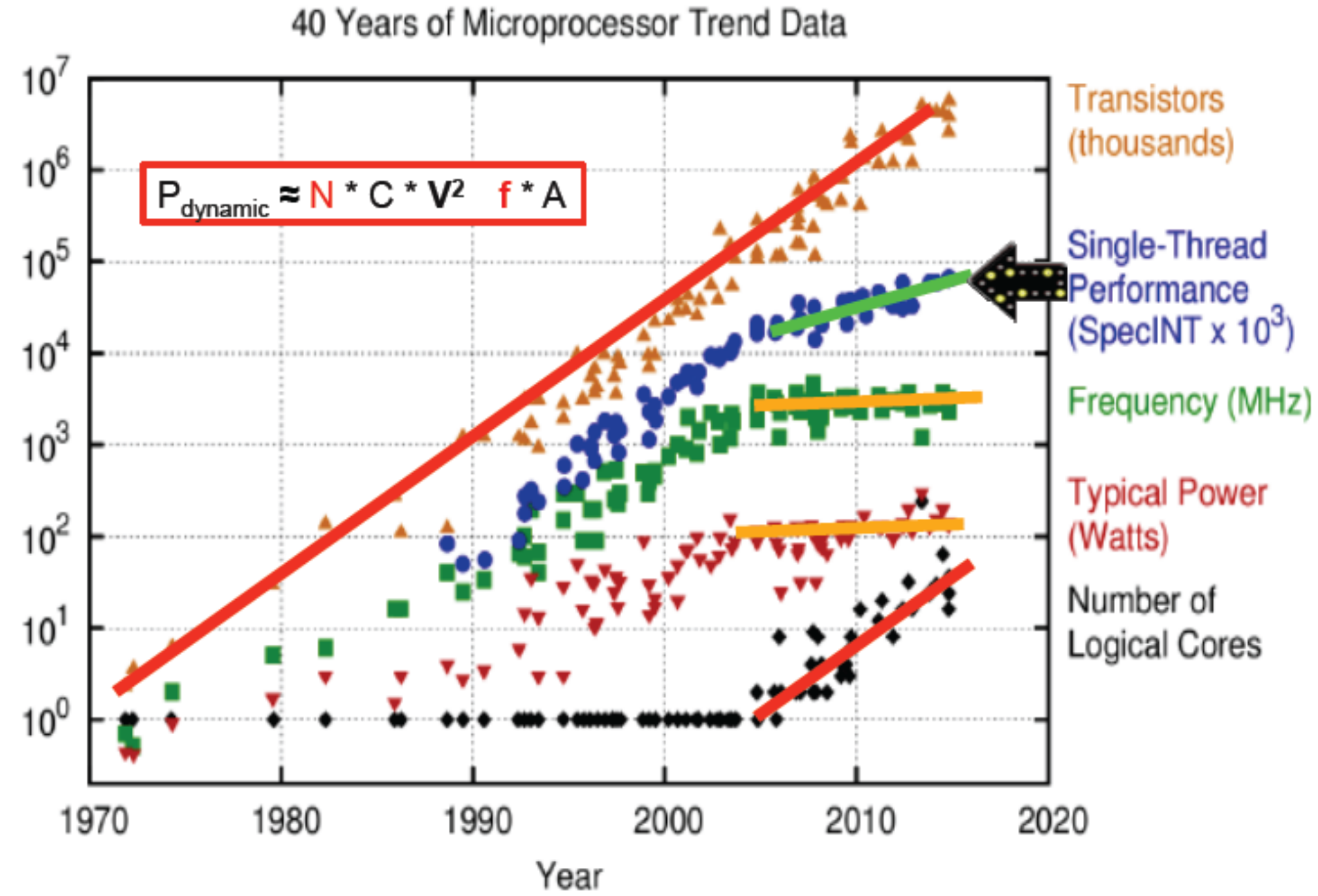
- Kept on getting more transistors, but couldn't build a bigger superscalar: ***The Multicore era began***



Two problems:

Sequential programs dominate the industry

Not all problems solvable via parallel algorithms


40 Years of Microprocessor Trend Data
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
$P_{dynamic} \approx N * C * V^2 \quad f * A$
Transistors (thousands)
Single-Thread Performance (SpecINT x 10^3)
Frequency (MHz)
Typical Power (Watts)
Number of Logical Cores
1970
1980
1990
2000
2010
2020
Year

How we got here, part 3

- 1945: Von Neumann's report describing computer arch.
- 1965: Software industry begins (IBM 360)
- 1975: Moore's Law; Dennard's geometric scaling rule
- 1985: "Killer micros": HPC, general-purpose hitch a ride on Moore's law
- 1995: Slowdown in CMOS logic: superscalar era *begins*
- 2005: The Power Wall: Single thread exponential scaling *ends* (Intel Prescott)
- 2011-2016: Realizing the problem: IEEE Rebooting Computing Initiative

IEEE rebooting COMPUTING

Goal: *Rethink Everything*: Turing & Von Neumann to now

Why IEEE? Encompasses the whole computing stack

SSCS IEEE SOLID-STATE CIRCUITS SOCIETY Where ICs are in IEEE

IEEE computer society

CAS Circuits & Systems Society

IEEE NANO Nanotechnology Council

IEEE CSC Council on Superconductivity

EDS IEEE Electron Devices Society

ReliabilitySociety

CEDA Council on Electronic Design Automation

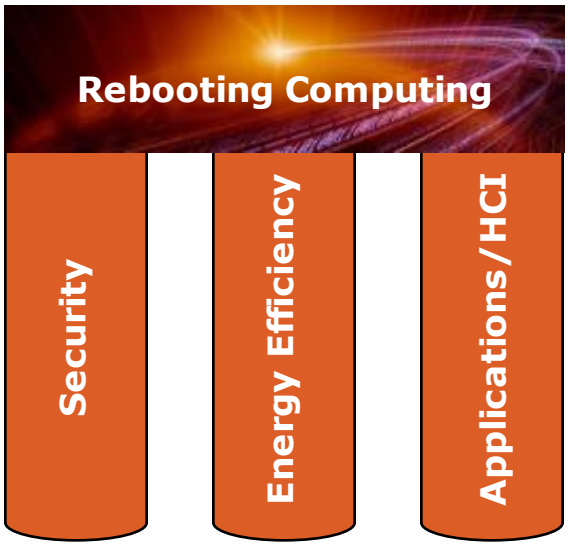
IEEE MAGNETICS 50 YEARS 1964 2014

IEEE

Advancing Technology for Humanity

IEEE Rebooting Computing

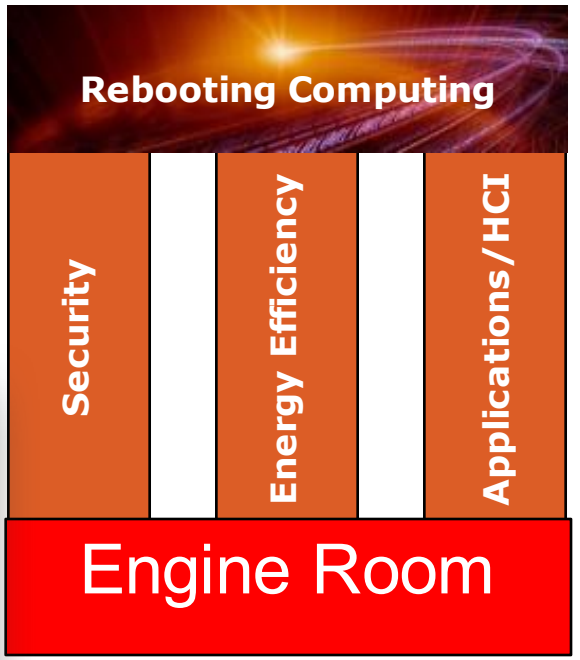
- **Summit 1:** 2013 Dec. 12-13 (summary online)
  - Invitation only
  - Three Pillars:
  - **Energy Efficiency**
  - **Security**
  - **Applications/HCI**



IEEE
Advancing Technology for Humanity

IEEE Rebooting Computing

- **Summit 2**: 2014 May 14-16
  - **Engines of Computation**
    - Adiabatic/Reversible Computing
    - Approximate Computing
    - Neuromorphic Computing
    - Augmentation of CMOS

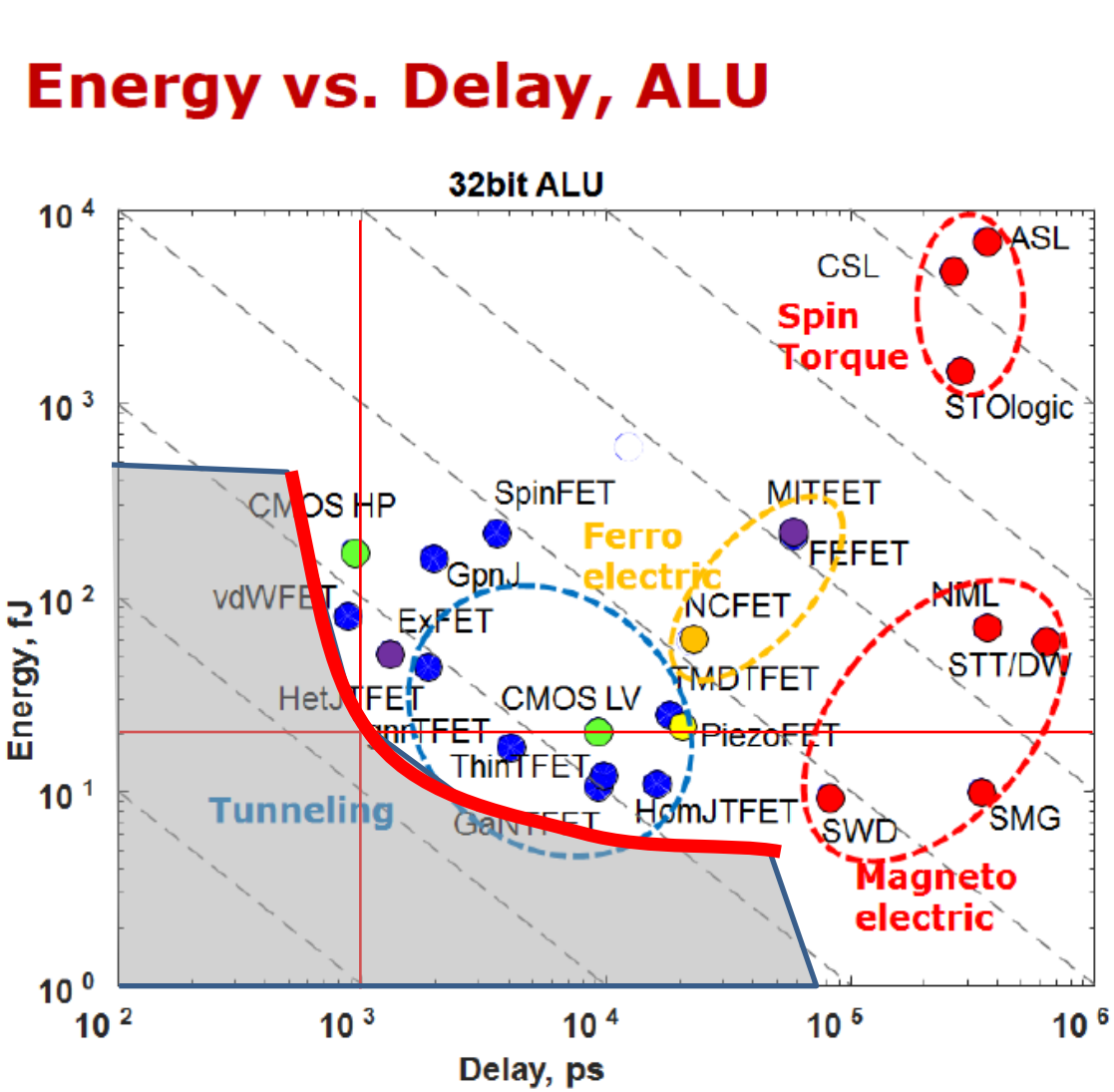


IEEE
Advancing Technology for Humanity

RCI Summit 2: Ways to compute

- Many alternatives
  - New switch (SEE NEXT SLIDE)
  - 3D Integration
  - Adiabatic/Reversible logic
  - Unreliable switch
  - Approximate, Stochastic
  - Cryogenic
  - Neuromorphic
  - Quantum
  - Analog neuro memetic ("sensible machines")
  - …
- *not all are general-purpose drop ins*
  - *(nor do they need to be)*

A better switch?

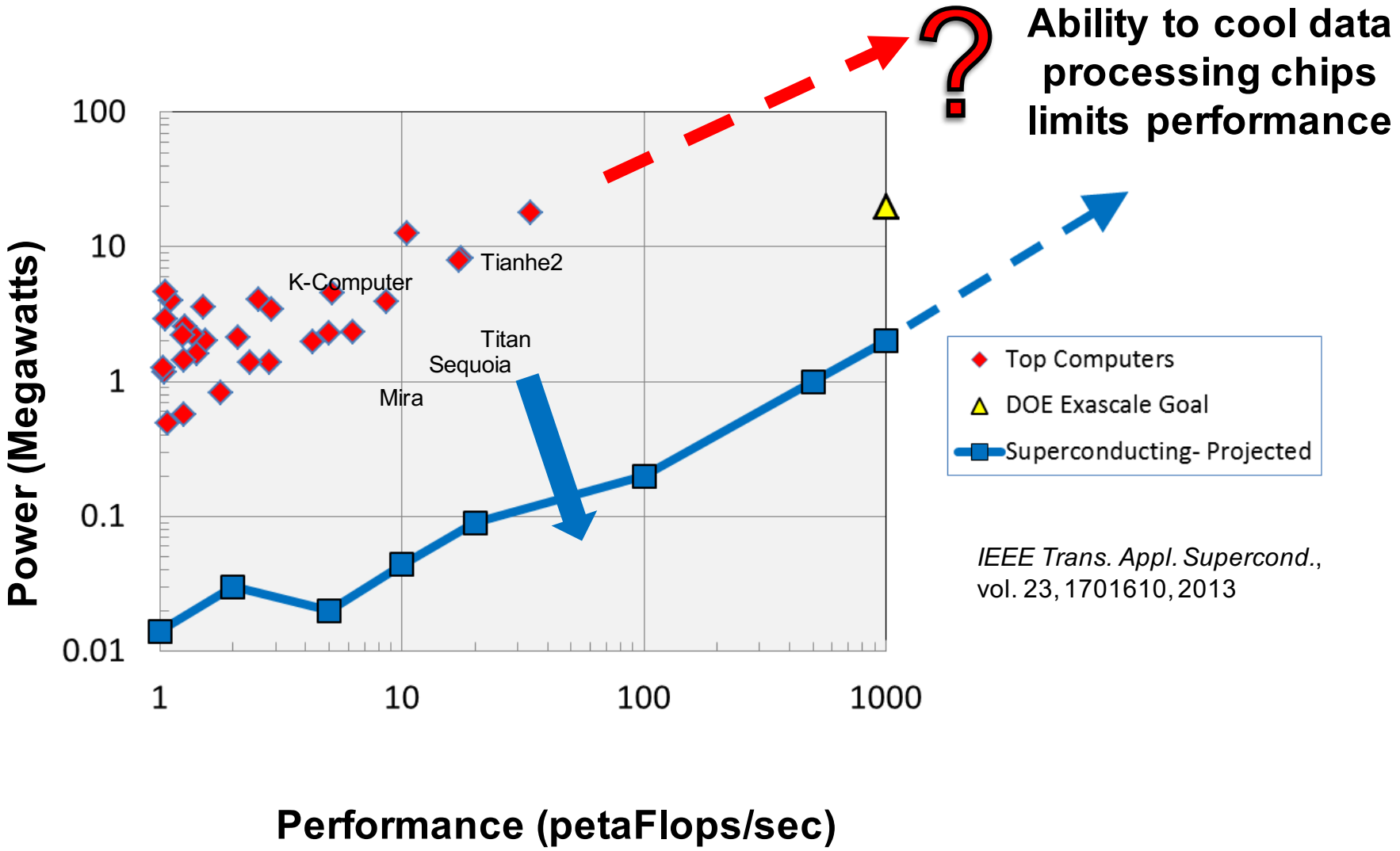


CMOS ref
Electronic
Spintronic
Ferroelectric
Orbitronic
Straintronic

NONE OF THE NEW DEVICES WILL RESTART SCALING

Courtesy Dimitri Nikonov and Ian Young

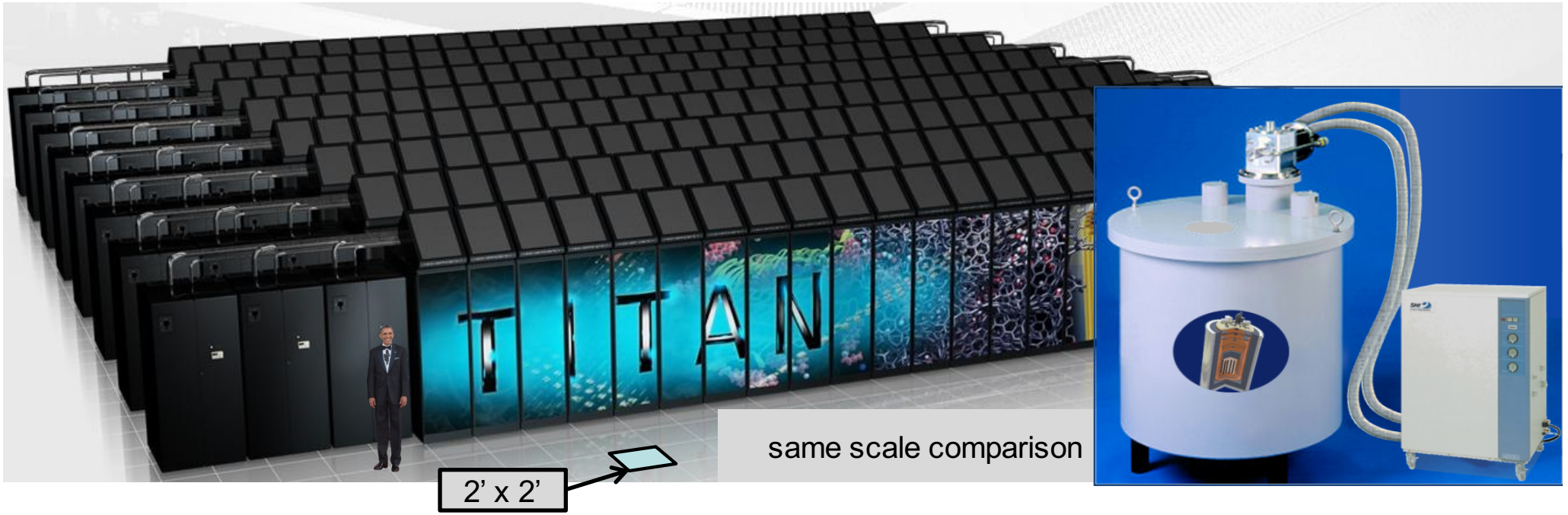
Superconducting computing looks promising!



IEEE Trans. Appl. Supercond., vol. 23, 1701610, 2013

IEEE rebooting COMPUTING

Superconducting System: smaller, lower power, same performance
System Comparison (~20 PFLOP/s) 8MW→80kW
Scale-up: 1 ExaFLOP/s at only 4 MW



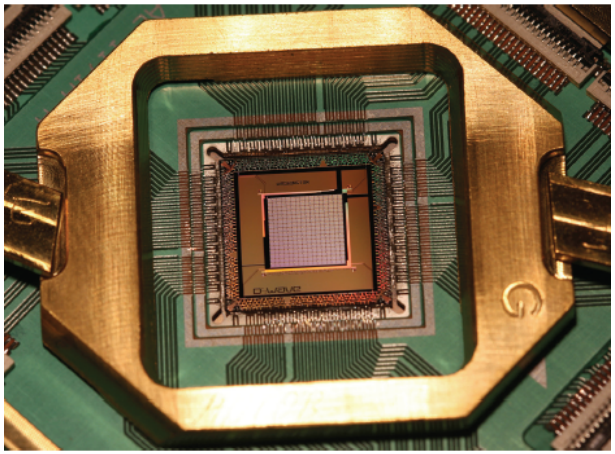
	Supercomputer Titan at ORNL - #2 of Top500	Superconducting Supercomputer	
Performance	**17.6** PFLOP/s (#2 in world*)	**20** PFLOP/s	~1x
Memory	**710** TB (0.04 B/FLOPS)	**5** PB (0.25 B/FLOPS)	7x
Power	**8,200** kW avg. (not included: cooling, storage memory)	**80** kW total power (includes cooling)	0.01x
Space	**4,350** ft^2 (404 m^2, not including cooling)	**~200** ft^2 (includes cooling)	0.05x
Cooling	additional power, space and infrastructure required	All cooling shown	

Courtesy of M. Manheimer, IARPA Cryogenic Computing Complexity (C3) Program

IEEE rebooting COMPUTING

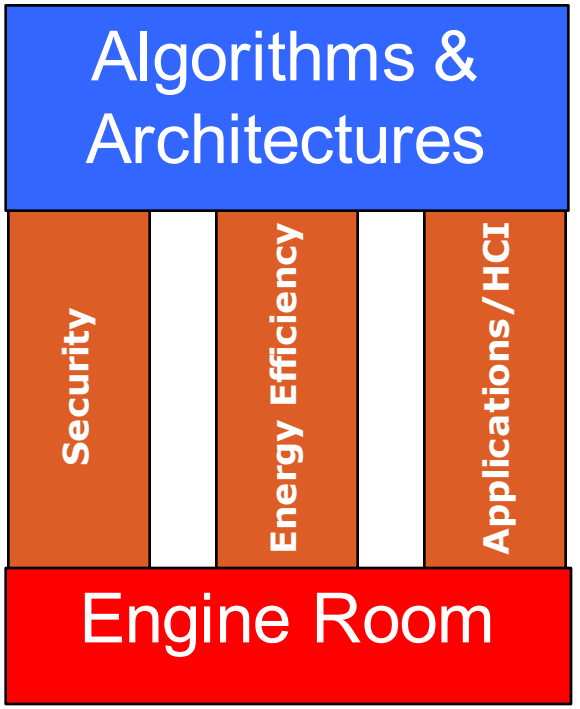
D-Wave Quantum Annealing Processor

- Integrated circuit of 1000 flux qubits
  - Chip with 128,000 JJs, mostly classical superconducting circuits for programming and readout of qubits
- Complete turnkey system
  - "Dilution Refrigerator" with T = 0.02K
  - Digital I/O for interfacing with room-temperature computer
- Designed to solve optimization problems
  - Also machine learning and protein folding.



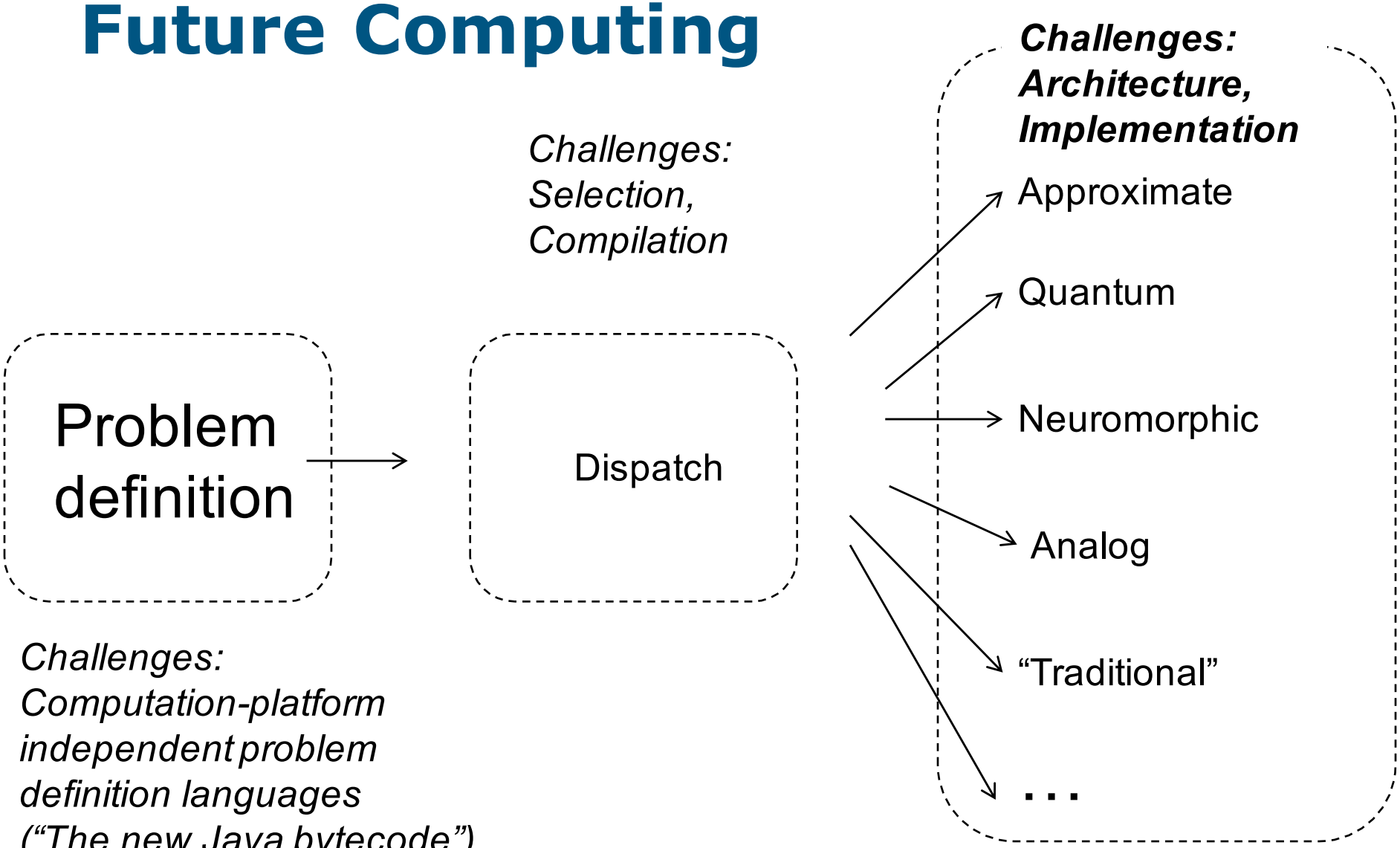
IEEE Rebooting Computing

- **Summit 3: 2014 Oct. 23-24**
  - **Algorithms and Architectures**
    - Security
    - HCI and Applications
    - Randomness & Approximate Computing
    - ITRS joins forces with RCI



IEEE
Advancing Technology for Humanity

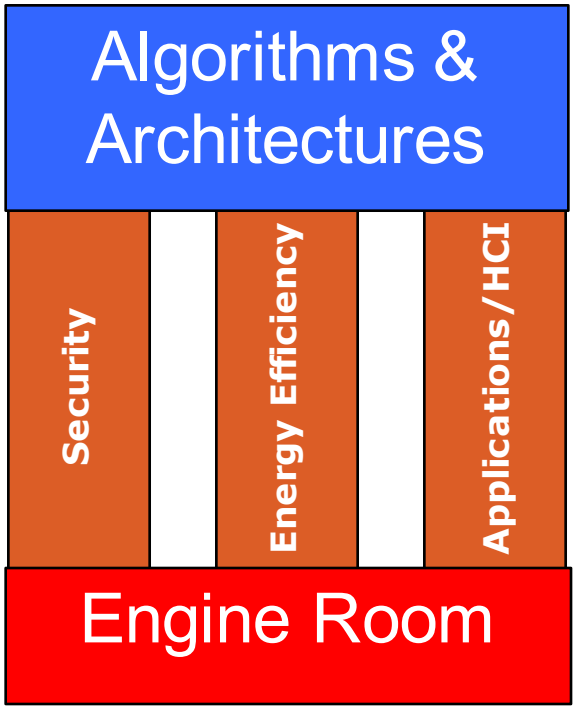
Future Computing



IEEE
Advancing Technology for Humanity

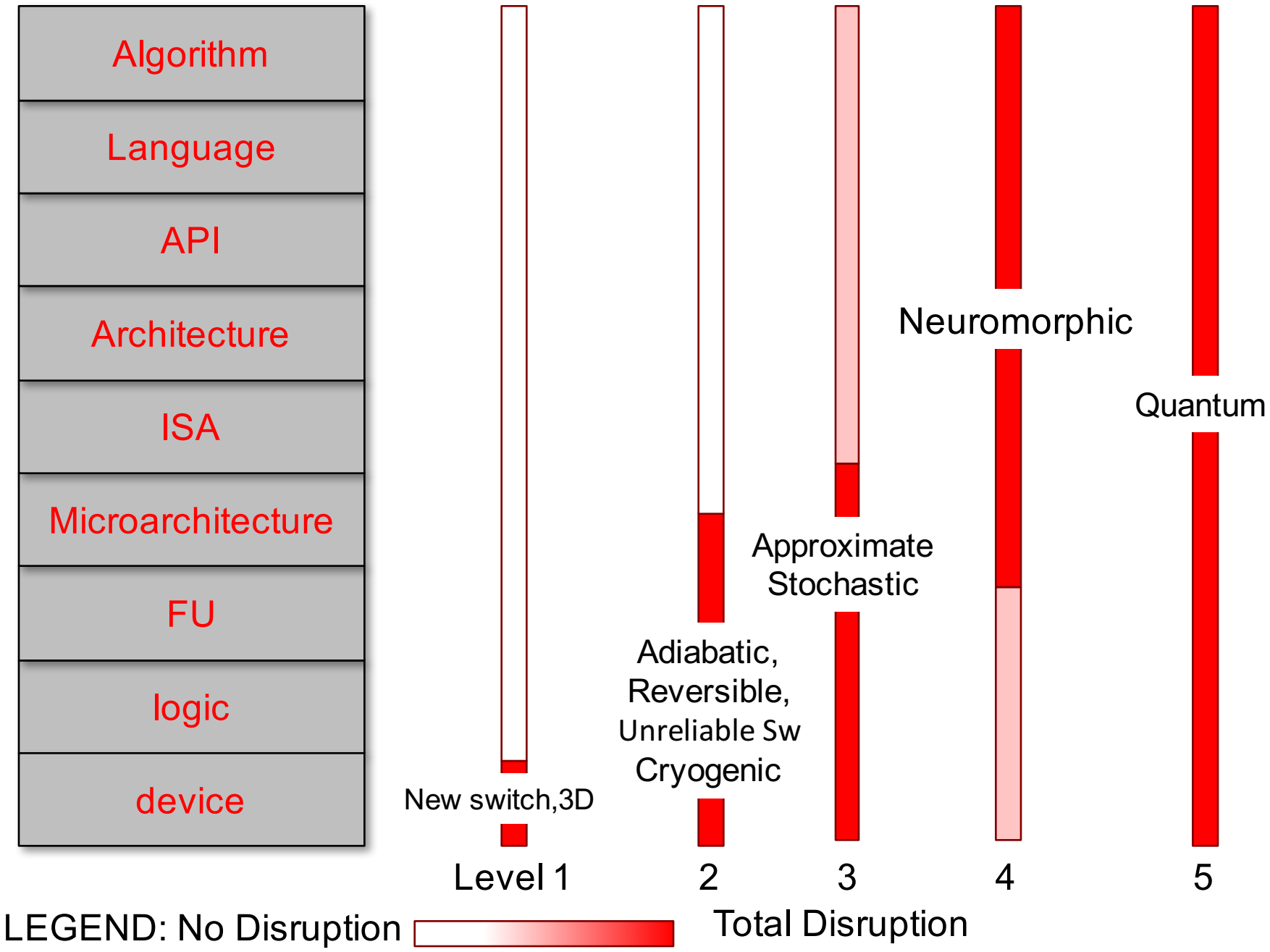
IEEE Rebooting Computing

- **Summit 4: 2015 Dec. 10-11**
  - **Coordinating efforts**
  - **Industry (HP, Intel, NVIDIA)**
  - **US: DOE, DARPA, IARPA, NSF**
  - **How to roadmap the future**



IEEE
Advancing Technology for Humanity

Differing Levels of Disruption in Computing Stack



Briefing Room

Your Weekly Address

Speeches & Remarks

Press Briefings

Statements & Releases

White House Schedule

Presidential Actions

Executive Orders

Presidential Memoranda

Proclamations

Legislation

Nominations & Appointments

Disclosures

The White House
Office of the Press Secretary

For Immediate Release July 29, 2015

Executive Order -- Creating a National Strategic Computing Initiative

EXECUTIVE ORDER

- - - - - - -

CREATING A NATIONAL STRATEGIC COMPUTING INITIATIVE

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing (HPC) research, development, and

PRELIMINARY IEEE RC-ITRS REPORT
SEPTEMBER 2015

On The Foundation Of The New Computing Industry Beyond 2020

Thomas M. Conte and Paolo A. Gargini
IEEE Fellows

EXECUTIVE SUMMARY

FEDERAL REGISTER
The Daily Journal of the United States Government

Nanotechnology-Inspired Grand Challenges for the Next Decade

A Notice by the Science and Technology Policy Office on 06/17/2015

ACTION Notice Of Request For Information.

SUMMARY The purpose of this Request for Information (RFI) is to seek suggestions for *Nanotechnology-Inspired Grand Challenges for the Next Decade:* Ambitious but achievable goals that harness nanoscience, nanotechnology, and innovation to solve important national or global problems and have the potential to capture the public's imagination. This RFI is intended to gather information from external stakeholders about potential grand challenges that will help guide the science and technology priorities of Federal agencies, catalyze new research activities, foster the commercialization of nanotechnologies, and inspire different sectors to invest in achieving the goals. Input is sought from nanotechnology stakeholders including researchers in academia and industry, non-governmental organizations, scientific and professional societies, and all other interested members of the public.


IEEE
rebooting
COMPUTING
Search IEEE Rebooting Computing
Search
Join the Rebooting Computing Technical Community
IEEE Rebooting Computing Responds to White House Nanotechnology Grand Challenge: "Sensible Machines" That are Smaller, Faster, and Lower Power
Rebooting Computing Initiative, October 2015

OSTP RFI

stan.williams@hp.com
epdeben@sandia.gov

OSTP Nanotechnology-Inspired Grand Challenge: Sensible Machines (extended version 2.5)

R. Stanley Williams
Hewlett-Packard Laboratories

Erik P. DeBenedictis
Sandia National Laboratories

October 20, 2015

Committee (also reviewers):	Reviewers and team members
Thomas M. Conte, IEEE and Georgia Tech Paolo A. Gargini, ITRS 2.0 David J. Mountain, IEEE and LPS Elie K. Track, IEEE and nVizix	IEEE Rebooting Computing: Arvind Kumar, IBM, Mark Stalzer, Caltech ITRS 2.0: Mustafa Badaroglu, Qualcomm, Geoff W. Burr, IBM, An Chen, Globalfoundaries, Shamik Das, MITRE, Andrew B. Kahng, UCSD, Matt Marinella, Sandia Sandia: Sapan Agarwal, John B. Aidun, Frances S. Chance, Michael P. Frank, Conrad D. James, Fred Rothganger, John S. Wagner SRC: Ralph Cavin, Victor Zhirnov

A Nanotechnology-Inspired Grand Challenge for Future Computing

OCTOBER 20, 2015 AT 6:00 AM ET BY LLOYD WHITMAN, RANDY BRYANT, AND TOM KALIL

Summary: Today, the White House is announcing a grand challenge to develop transformational computing capabilities by combining innovations in multiple scientific disciplines.

In June, the Office of Science and Technology Policy issued a Request for Information seeking suggestions for *Nanotechnology-Inspired Grand Challenges for the Next Decade.* After considering over 100 responses, OSTP is excited to announce the following grand challenge that addresses three Administration priorities—the National Nanotechnology Initiative, the National Strategic Computing Initiative (NSCI), and the BRAIN initiative:

Create a new type of computer that can proactively interpret and learn from data, solve unfamiliar problems using what it has learned, and operate with the energy efficiency of the human brain.

How we got here, part 3

- 1945: Von Neumann's report describing computer arch.
- 1965: Software industry begins (IBM 360)
- 1975: Moore's Law; Dennard's geometric scaling rule
- 1985: "Killer micros": HPC, general-purpose hitch a ride on Moore's law
- 1995: Slowdown in CMOS logic: superscalar era *begins*
- 2005: The Power Wall: Single thread exponential scaling *ends* (Intel Prescott)
- 2016: Int'l Roadmap for Devices & Systems formed
- 2021: Moore's Law *for 2D lithography* ends "for real"

Why's and How's of Roadmapping

Roadmapping is Essential For Success

Moore defined the law, *Roadmapping with buy-in from industry kept it going* because roadmapping:

1. *Tracks progress*
2. *Finds roadblocks*
3. *Identifies and compares potential solutions*
4. *Pre-competitive/standards-like*

Roadmap history

- Gordon Moore published two forecasts on the rate of growth of transistors as a function of time, one in 1965 and one in 1975.
- National Technology Roadmap of Semiconductors in 1998, changed to International Tech. Roadmap of Semi (ITRS), and then ITRS 2.0 - *Produced 20 total roadmaps 1998-2015*

Roadmap history

- In 2013, ITRS2.0 signed an MOU with IEEE Rebooting Computing Initiative
- In 2015, ITRS2.0 declared 'last roadmap' by US SIA leadership
- Industry disagreed: roadmapping should not be ended. Rather, it should be evolved
- IEEE RCI moves roadmapping into IEEE as the International Roadmap of Devices and Systems (IRDS)

IEEE
Advancing Technology for Humanity

IEEE International Roadmap for Devices and Systems

- Formed under IEEE Standards Association, first meeting Leuven Belgium (May, 2016)
- Roadmapping processes from ITRS 2.0
- New "front end" processes to identify:
  - Important application drivers
  - Appropriate computer architectures
  - Systems integration
  - (more on this later)

IEEE
Advancing Technology for Humanity

IRDS FT Focus Teams

ITWG	Deliverable	Chair	Company
Applications Benchmarking **(AB)**	Key App Drivers, App Performance	T. Conte	GA Tech
Beyond CMOS **(BC)**	Emerging Logic Devices, High Perf, Low Pwr, Sensing Applications	A. Chen	IBM
Factory Integration **(FI)**	Semi. Mfg infrastructure R&D for Hdwre, Sftwre, Materials, Metrology 15 Yr. look ahead	J. Moyne	U of Michigan/ Applied Materials
Heterogeneous Integration **(HI)**	IC Package Requirements Package Single and Multiple IC and Components	B. Bottoms	Third Millennium Test Solutions
Outside System Connectivity **(OSC)**	RFIC, IOT, High Perf Wireless CMOS, Bipolar, III-V Photonic Communication, Devices, Components, LAN, Fiber, Vehicle, Aerospace	M. Garner	Garner Nanotechnology Solutions
More Moore **(MM)**	Physical, Electrical, Reliability Logic and Memory IC scaling Requirements	M. Badaroglu	Qualcomm
Systems and Architectures **(SA)**	Computer architectures, IC & Component Performance, Power Requirements 15 yr. outlook	A. Kahng	UCSD

IEEE Advancing Technology for Humanity

IRDS ITWG International Technical Working Groups

ITWG	Deliverable	Chair	Company
Env. Safety & Health **(ESH)**	Identify Risks and Alternative Options. Safe Tech solutions	S. Moffatt	Applied Materials
Emerging Research Materials **(ERM)**	Device, Interconnect, Photonic, Pkg solutions Emerging materials issues and solutions	D. Herr	UNC Greensboro
Interconnect **(I)**	Process, Materials, Equipment for future Logic and Memory Technology	A. Isobayashi	Toshiba
Lithography **(L)**	Assess Litho equipment, materials and and process optionsfor future Semi Mfg.	M. Neisser	Kem Pur
Metrology **(M)**	CMOS Extensions, Beyond CMOS, MFG, Materials Characterization, 15 yr. outlook	G. (Ndubuisi) Oril	NIST
Test **(T)**	Test solutions for Semi devices (planar, 2.5D, 3D, MEMs, SOC, SIP, etc)	D. Armstrong	Advantest
Yield Enhancement **(YE)**	Monitoring, Characterization methods, Fast Yield Learning Contamination Control. Environment, Solutions	I. Thurner	CONVANIT GmbH&co.KG

IEEE Advancing Technology for Humanity

IRDS Press

EE Times

Chip Roadmap Reboots Under New Management

IEEE expands ITRS to include systems

The New York Times

Moore's Law Running Out of Room, Tech Looks for a Successor

Solid State TECHNOLOGY
Insights for Electronics Manufacturing

IEEE introduces new international roadmap for devices and systems to set the course for end-to-end computing

National Public Radio
npr

After Moore's Law: Predicting The Future Beyond Silicon Chips

IEEE Advancing Technology for Humanity

International Roadmap for Devices and Systems

Environmental Analysis

Applications Analysis

Systems & Architectures

Devices Components Integration

MM

FI

HC

HI

BC

OSC

Capabilities

Design

ERM

ESH

FEP

Litho

Interconnects

Metrology

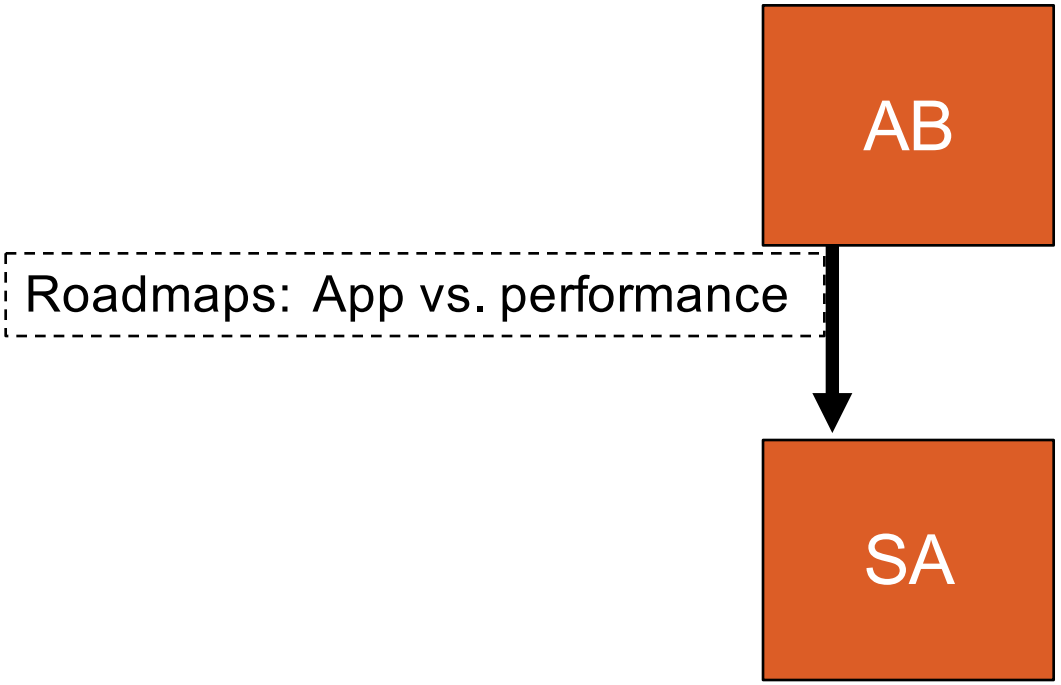
Modeling

Test

Yield

IEEE
Advancing Technology for Humanity

ITRS vs. IRDS: A new "front end"



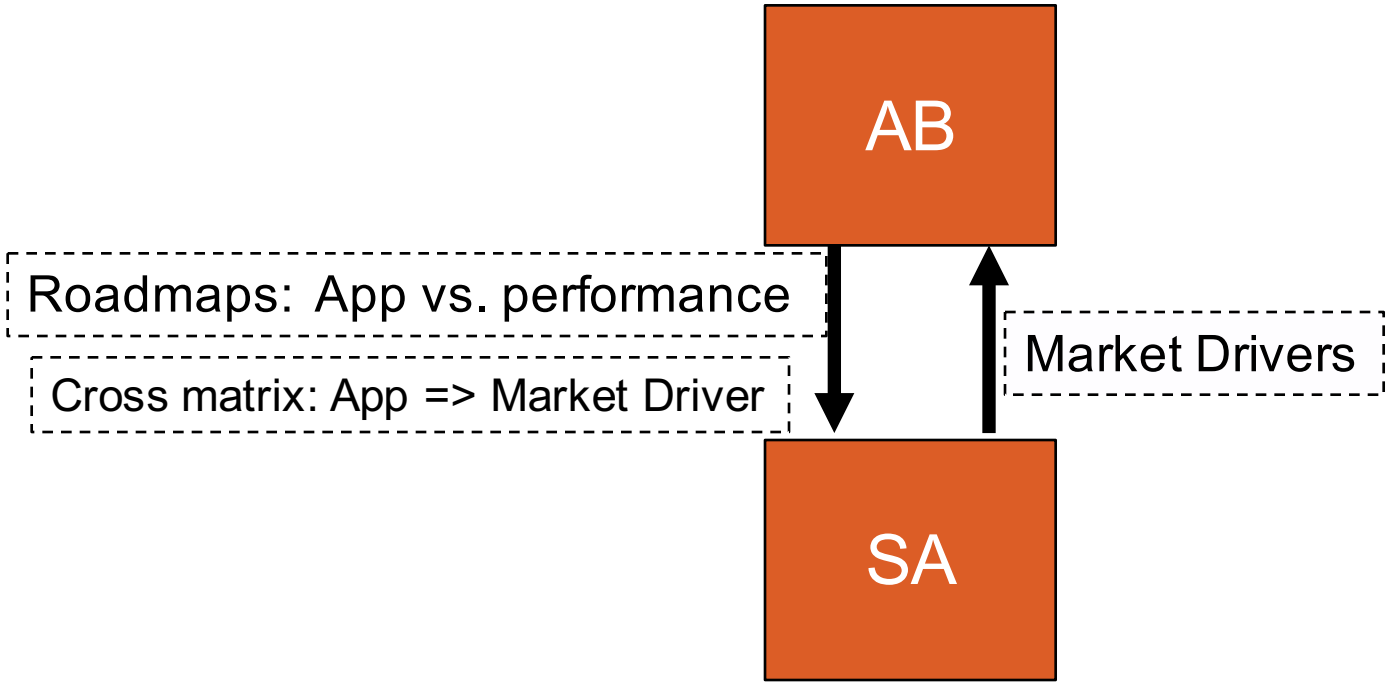
MM FI HC HI BC OSC

IEEE Advancing Technology for Humanity

Application Domains (v 0.2)

Application area	*Desired metric*	*Description*
Big Data Analytics	Feature/sec	Data mining to identify nodes in a large graph that satisfy a given feature/features
Feature Recognition	Feature/sec	Graphical dynamic moving image (movie) recognition of a class of targets (e.g., face, car) . This can include neromorphic approaches such as CNNs
Discrete Event Simulation	Sim/sec	Large discrete event simulation of a discretized-time system. (e.g., large computer system simulation) Generally used to model engineered systems
Physical system simulation	Sim/sec	Simulation of physical real-world phenomena. Typically finite-element based. Examples include fluid flow, weather prediction, thermo-evolution
Optimization	Solution/sec	Integer NP-hard optimization problems
Graphics rendering	Frame/sec	Large scale, real-time photorealistic rendering driven by physical world models. Examples include RPGs, VR.
Media processing	Frame/sec	Discrete processing, including filtering, compressing, decompressing of streaming media, where the media is unknown (i.e., camera stream based). Includes integrating multiple cameras to feed graphics rendering
Cryptographic codec	Codon/sec	Crypting and decrypting of data at the edge of cryptographic science. Includes asymmetric-key encroption, excludes symmetric-key encryption.

ITRS vs. IRDS: A new "front end"



MM FI HC HI BC OSC

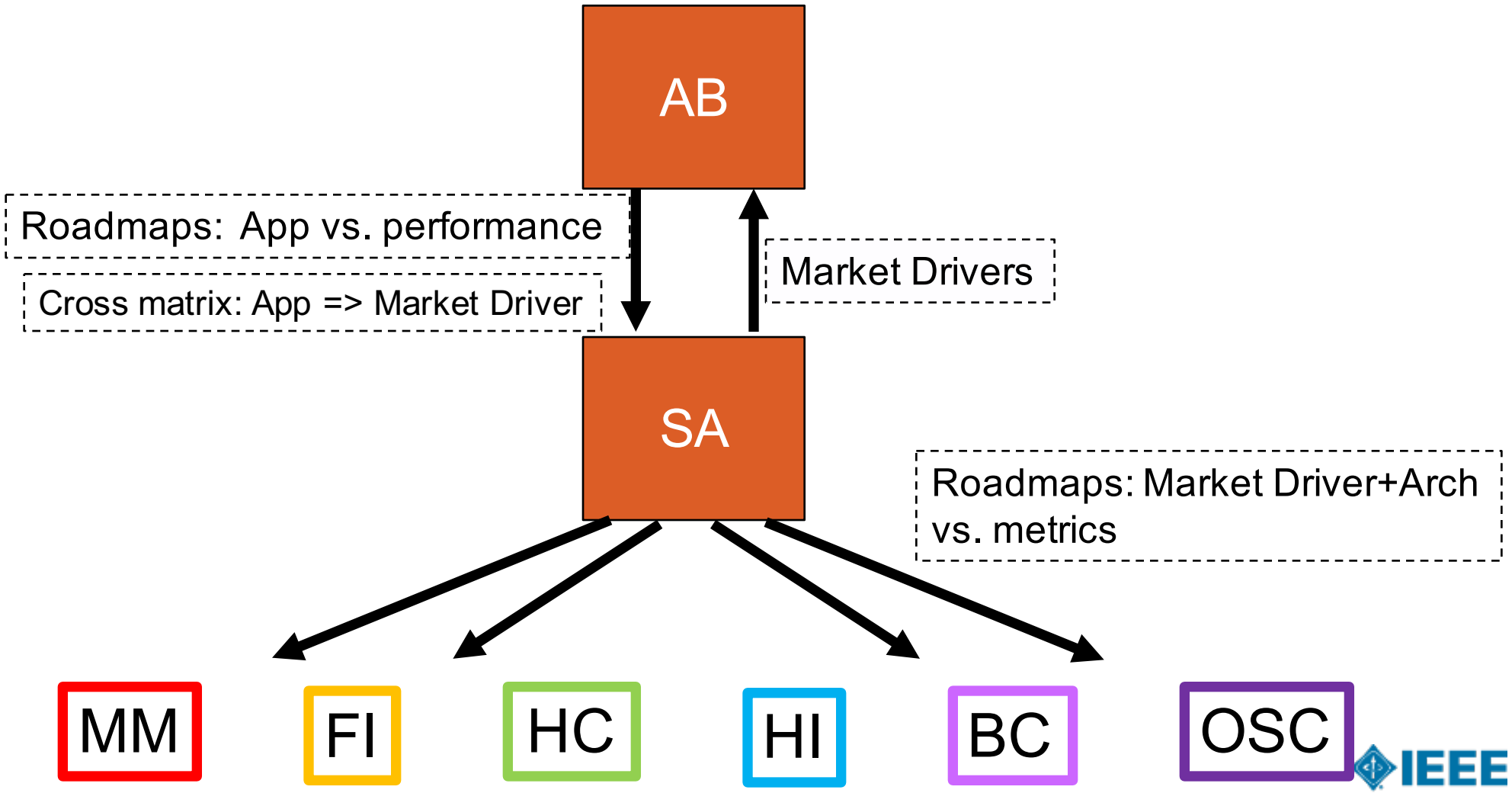
IEEE Advancing Technology for Humanity

Cross matrix (AB interface to SA)

	Medical diagnosis	*Bioinfomatics*	*Medical device*	*IoT (edge)*	*Cloud*	*Big Data*	*Robotics*	*CPS*	*Phone*	*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

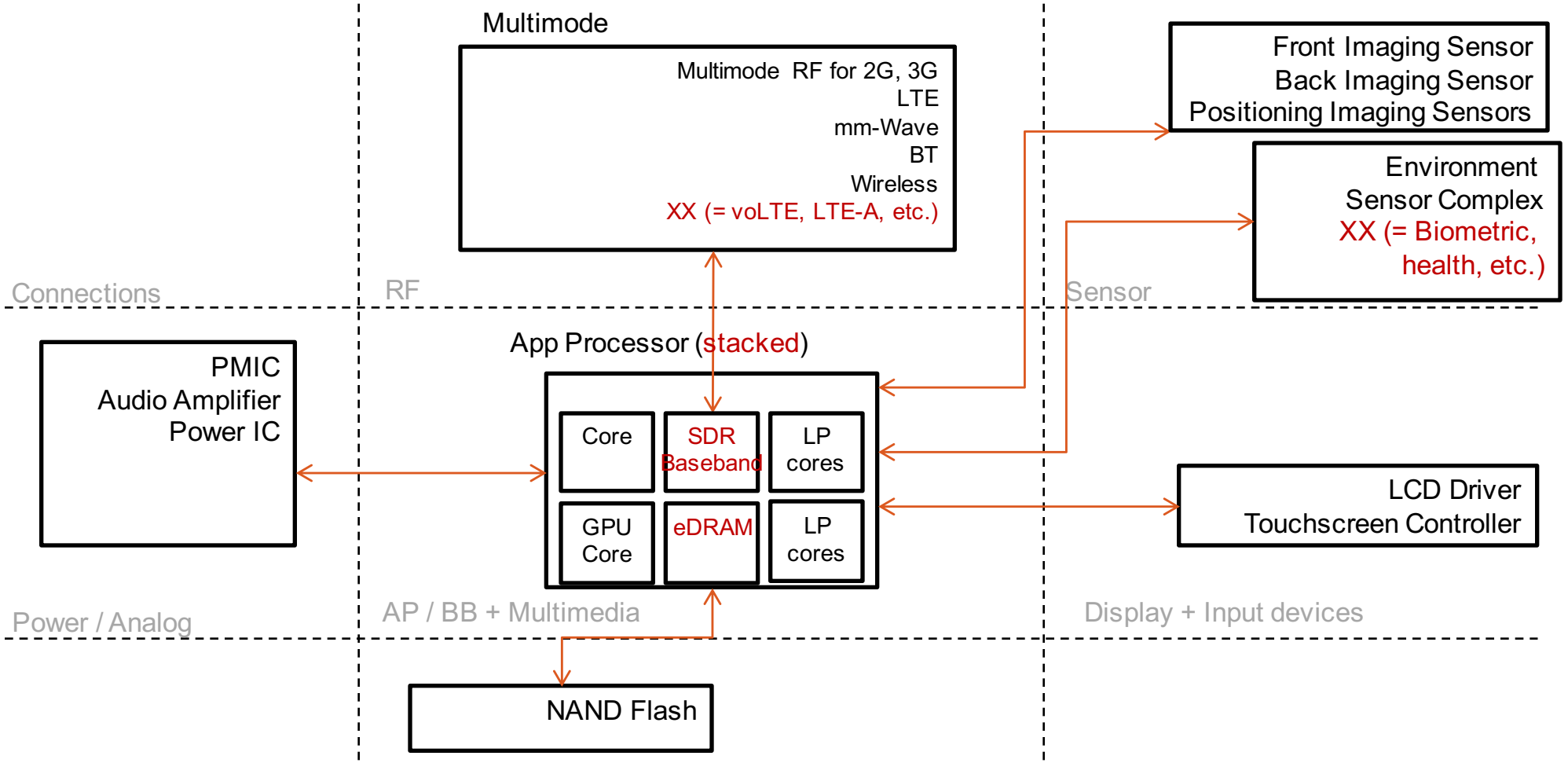
X = important
G = Gating (critical)
P = Power sensitive and important

ITRS vs. IRDS: A new "front end"



IEEE Advancing Technology for Humanity

2015 Smartphone Block Diagram



- Multiple functions integrated into AP, e.g., heterogeneous memory, SDR BB
- Integration of more sensors, antennas
- SDR = Software-Defined Radio

IEEE Advancing Technology for Humanity

SA: Smartphone Metrics Roadmap

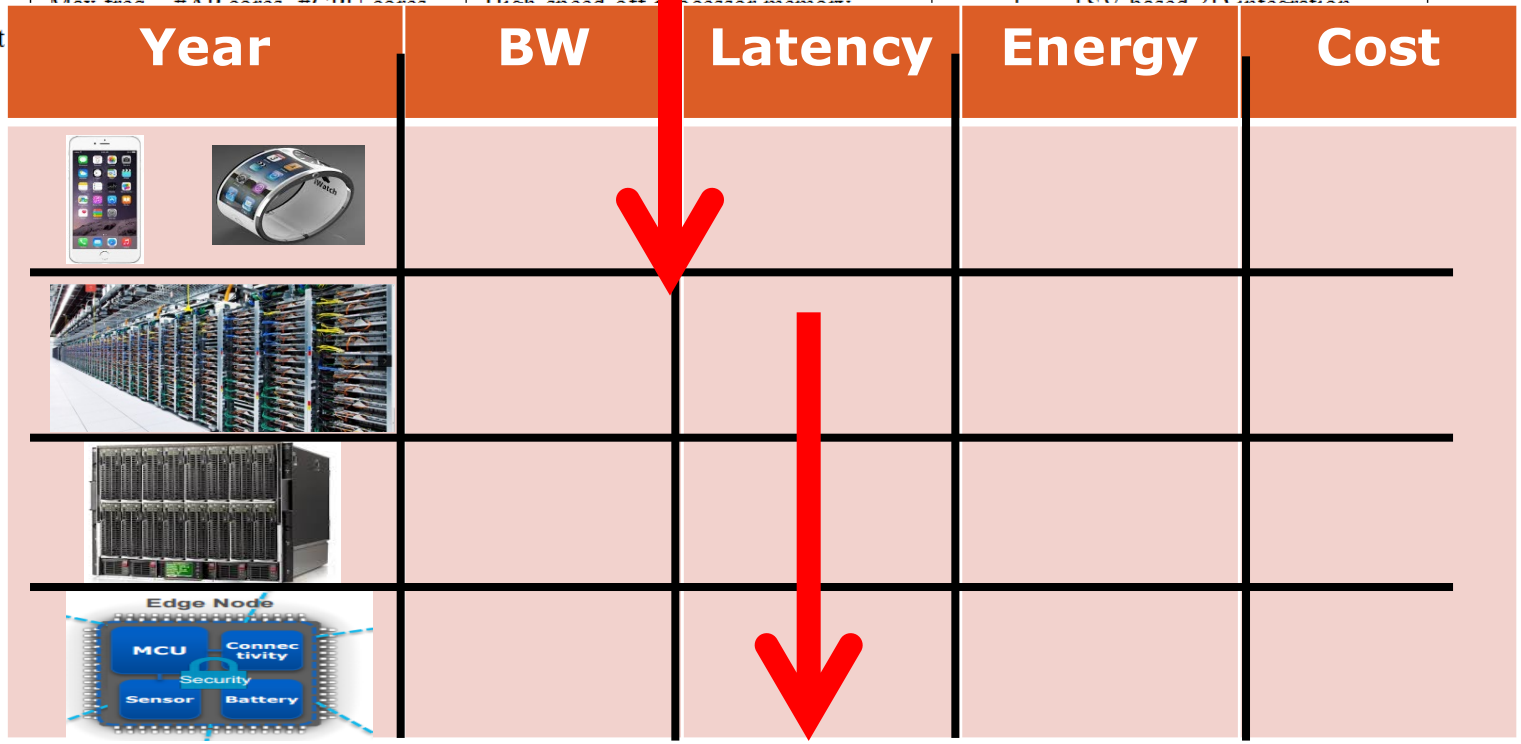
Table SYSINT2

System Metrics	Year	2007	2010	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
Input Metrics	#AP cores	1	2	4	4	4	4	4	6	6	6	8	8	10	10	12	14	14	16	18
	#GPU cores	2	6	10	12	14	18	22	28	34	44	54	68	86	110	138	174	220	280	352
	Max freq. (GHz)	0.6	1.5	2.5	2.6	2.7	2.8	2.9	3	3.2	3.3	3.4	3.6	3.7	3.8	4	4.2	4.3	4.5	4.7
	#MPixels	0.307	0.922	0.922	2.1	2.1	2.1	2.1	3.7	3.7	8.8	8.8	8.8	8.8	33.2	33.2	33.2	33.2	33.2	33.2
	Mem BW (GB/s)	3.2	22.6	22.6	45.3	45.3	45.3	45.3	64.0	64.0	89.6	89.6	89.6	89.6	148.2	148.2	148.2	148.2	148.2	148.2
	#Sensors	4	8	10	12	14	14	16	16	20	20	20	20	21	21	21	21	22	22	22
	#Antennas	6	8	10	11	11	11	13	13	13	13	14	14	15	15	15	15	15	15	15
	#ICs	8	12	9	7	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10	7-10
	Cellular data rate (MB/s)	0.048	1.70	12.50	12.50	12.50	12.50	12.50	21.63	21.63	40.75	40.75	40.75	40.75	40.75	40.75	40.75	40.75	40.75	40.75
	WiFi data rate (Mb/s)	6.75	75	75	867	867	867	867	867	867	867	7000	7000	7000	28000	28000	28000	28000	28000	28000
Output Metrics	Board area (cm^2)	53	53	57	59	62	66	69	73	76	80	84	89	93	98	103	103	103	103	103
	Board power (mW)			3882	4074	4274	4485	4706	4939	5183	5439	5708	5991	6287	6599	6926	7269	7630	8008	8406

IEEE
Advancing Technology for Humanity

Interface from SA to FTs = Metrics

Challenges	Metrics (Description)	Roadblocks	Potential solutions
Form Factor Challenge	#Sensors, #ICs, #Antennas, (#Components ↑) Memory bandwidth (PCB routing complexity ↑, #connectors ↑)	Increasing PCB footprint occupied by connectors and components	1. Package-level integration 2. Through-silicon via (TSV)-based 3D integration 3. Sequential 3D integration[12]
		Integration of heterogeneous components	1. TSV-based 3D integration[13] 2. Unified logic/RF technology
		Die area explosion due to more functionalities	1. Technology scaling 2. Sequential 3D integration
System-Level Power Management			



Technology metrics: M1HP, Ion/Ioff, Vdd, Rsd, #mask layers, CD 3 LWR, EUV mask defectivity, ILD k, …, circuits, design, test, …

IEEE Advancing Technology for Humanity

IRDS 2016 Meeting Schedule

- Kickoff: May 12-13, 2016, Leuven Belgium
  - Establish Working Groups, Governance
  - Regional participation
    - Europe: KSIA, Aeneas
    - Japan: JSAP
    - Korea: KSIA
    - US: IEEE RC
    - Taiwan
- Workshop at SEMICON West : July 12-14, 2016, San Francisco, CA
- Workshop: December 2016 –Taiwan

IRDS Meeting Schedule 2017-2019

- 2017
  - December Meeting: Japan
- 2018
  - December Meeting: Korea
- 2019
  - Spring Meeting, May time frame: Europe, Grenoble

IEEE Rebooting Computing: Take-aways

- Too many solutions: new switch/3D, new circuit tricks, approximate / stochastic, neuromorphic, quantum, …
- Software challenges are vast
- Brace for a heterogeneous world
- *Applications-driven* Roadmapping to direct the industry to produce the right devices

- **Web portal** http://rebootingcomputing.ieee.org

IEEE
Advancing Technology for Humanity