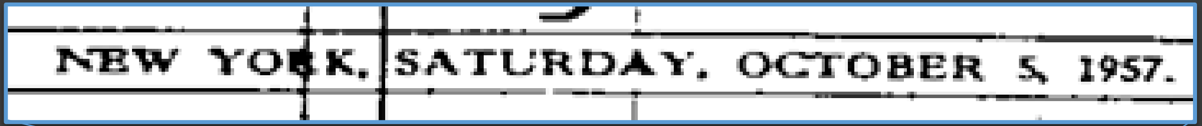
Page 3 Investments Workshop

Part of the

Electronics Resurgence Initiative

July 18, 2017



NEW YORK, SATURDAY, OCTOBER 5, 1957.

"All the News That's Fit to Print"

The New York Times.

LATE CITY EDITION

VOL. CVII..No. 36,414. NEW YORK, SATURDAY, OCTOBER 5, 1957. FIVE CENTS

SOVIET FIRES EARTH SATELLITE INTO SPACE; IT IS CIRCLING THE GLOBE AT 18,000 M. P. H.; SPHERE TRACKED IN 4 CROSSINGS OVER U. S.

—Dallas News Staff Photo.

SIGNALS FROM THE SATELLITE

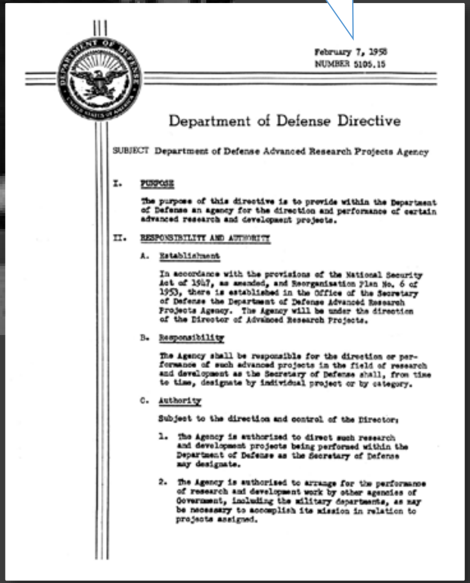
Ham operator Roy Welch of Dallas, seated, plays a tape-recorded signal from the Russian space satellite for fellow hams at the State Fair of Texas. Welch recorded the signals on a receiver at his home.

Nytimes.com

"The purpose of this directive is to provide within the Department of Defense an agency for the direction and performance of certain advanced research and development projects."

February 7, 1958
NUMBER 5105.15

Gettyimages.com



February 7, 1958
NUMBER 5105.15

Department of Defense Directive

SUBJECT Department of Defense Advanced Research Projects Agency

I. PURPOSE

The purpose of this directive is to provide within the Department of Defense an agency for the direction and performance of certain advanced research and development projects.

II. RESPONSIBILITY AND AUTHORITY

A. Establishment

In accordance with the provisions of the National Security Act of 1947, as amended, and Reorganization Plan No. 6 of 1953, there is established in the Office of the Secretary of Defense the Department of Defense Advanced Research Projects Agency. The Agency will be under the direction of the Director of Advanced Research Projects.

B. Responsibility

The Agency shall be responsible for the direction or performance of such advanced projects in the field of research and development as the Secretary of Defense shall, from time to time, designate by individual project or by category.

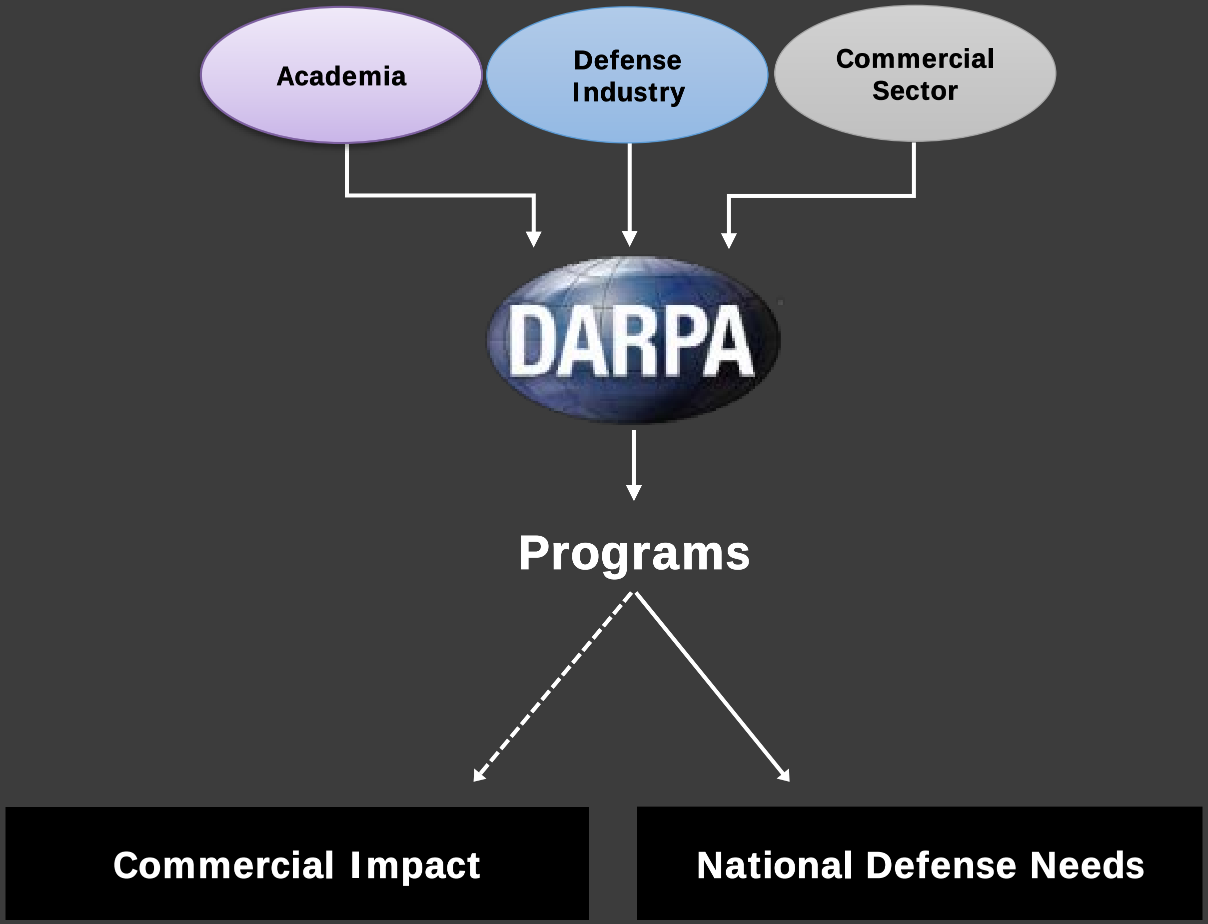
C. Authority

Subject to the direction and control of the Director:

1. The Agency is authorized to direct such research and development projects being performed within the Department of Defense as the Secretary of Defense may designate.

2. The Agency is authorized to arrange for the performance of research and development work by other agencies of Government, including the military departments, as may be necessary to accomplish its mission in relation to projects assigned.

How do we operate?



Program managers from the community...

on a temporary 3 to 5 year assignment...

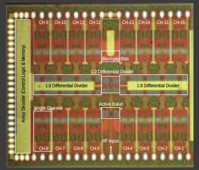
executing ~$3 billion in the hands of ~90 PM's through ~250 programs...

to eliminate technical surprise.

1970 … 2000 2005 2010 2015 Today 2020

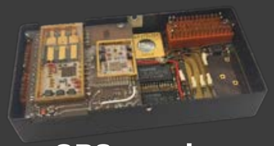
ARPANET

TEAM



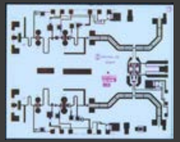
mmW arrays

MGR



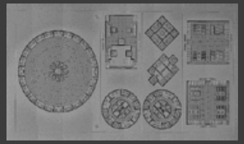
GPS receiver

MIMIC



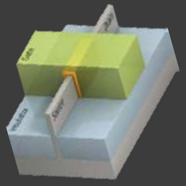
GaAS

MEMS



Inertial sensors

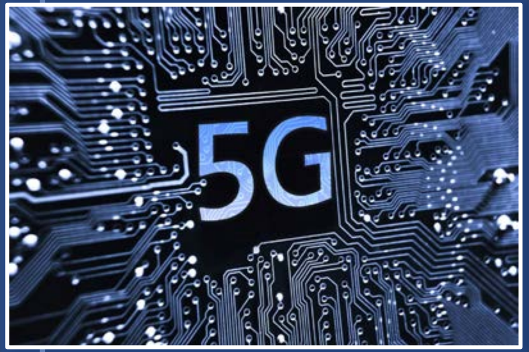
AME



FinFET

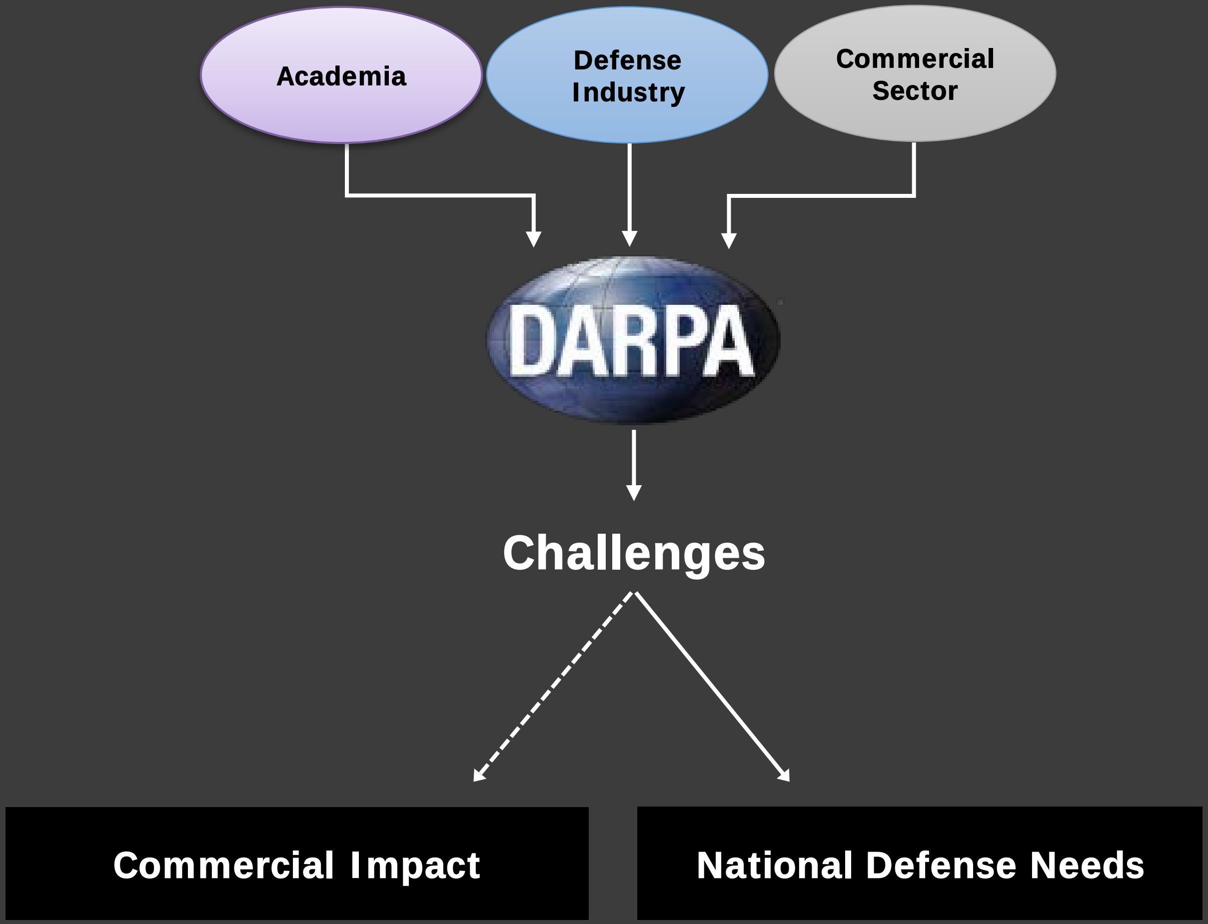
PAL

Siri



SiGe – Silicon Germanium
mmW – Millimeter wave
FinFET - Fin-Shaped Field Effect Transistor
AME – Advanced Microelectronics
MEMS – Micro Electrical Mechanical Systems
MGR – Miniature GPS Receiver
MIMIC – Microwave/Millimeter-Wave Monolithic Integrated Circuits
RF CMOS – Radio Frequency Complimentary Metal Oxide Semiconductor
PAL – Personal Assistant that Learns
TEAM – Technology for Efficient, Agile Microsystems

How do we operate?



Program managers from the community...

on a temporary 3 to 5 year assignment...

executing ~$3 billion in the hands of ~90 PM's through ~250 programs...

to eliminate technical surprise.

DARPA has evolved to using challenges

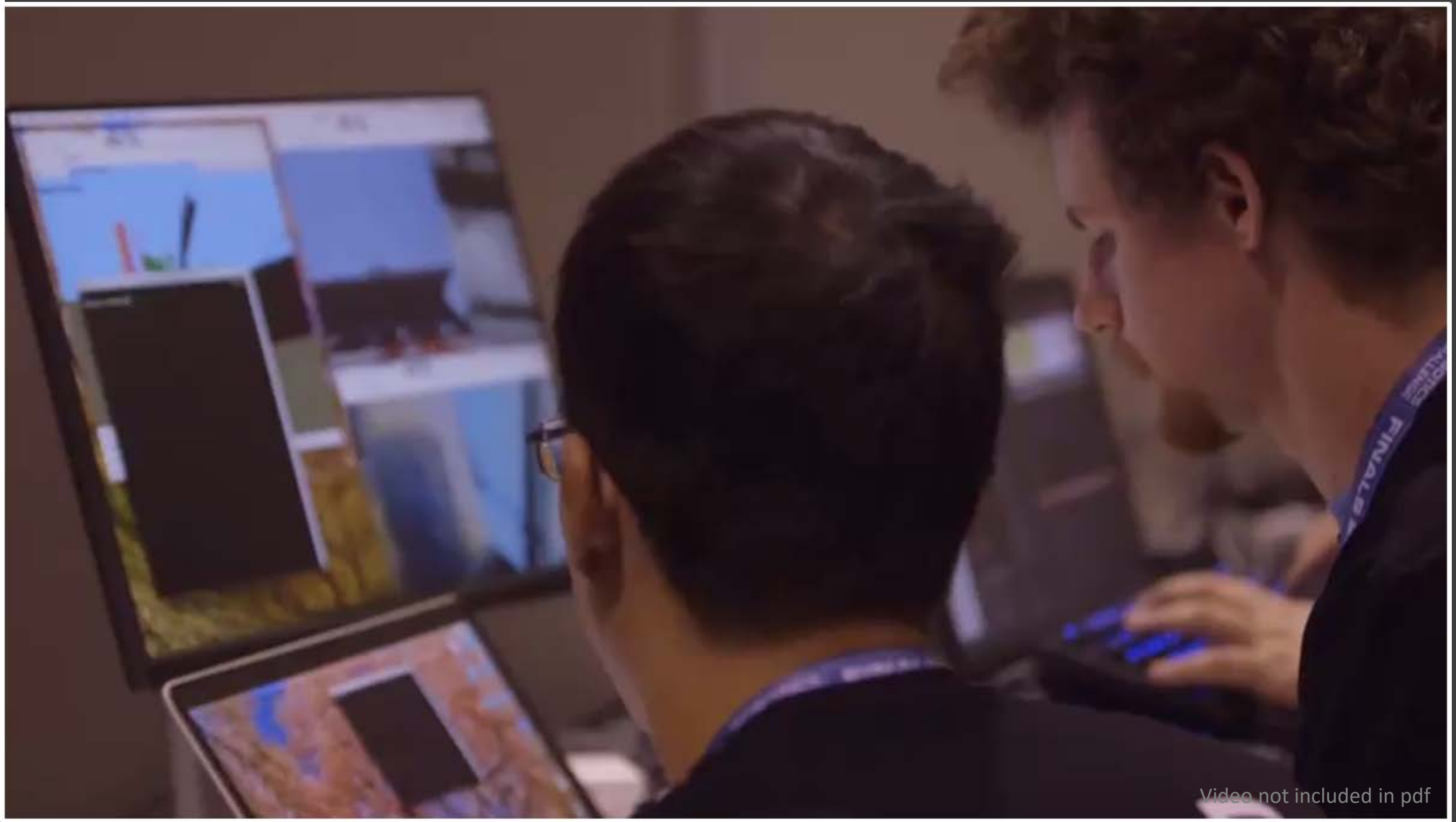
2005 — 2015 — Today — 2020

Grand Challenge (2005-2007)

2014 2015 2016 2017

2015

Robotics Challenge



2014 — 2015 — 2016 — 2017

2016

Cyber Grand Challenge

2014 — 2015 — 2016 — 2017

DARPA Spectrum Challenge

Video not included in pdf

Spectrum Collaboration Challenge

2017

2005 2015 Today 2020

Grand Challenge (2005-2007)

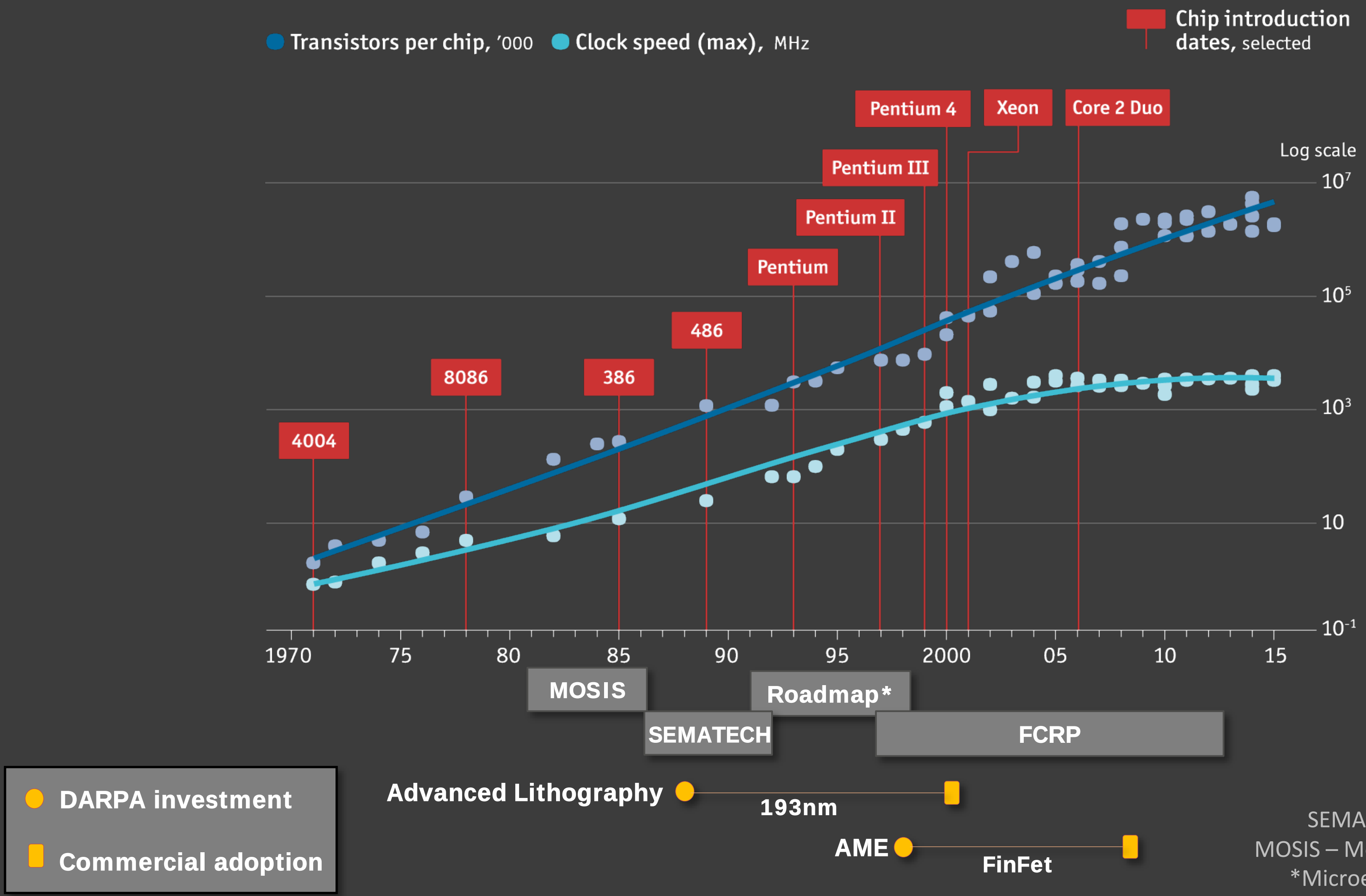
Exploring the capabilities of learning / autonomy and their societal impact

Robotics Challenge (2012-2015)

Cyber Grand Challenge (2016)

Spectrum Collaboration Challenge (2017-2018)

The miracle of Moore's Law has taken us incredibly far...



Sources: Intel; press reports; Bob Colwell; Linley Group; IB Consulting; *The Economist*

AME – Advanced Microelectronics
FCRP – Focus Center Research Program
FinFET – Fin-Shaped Field Effect Transistor
SEMATECH – Semiconductor Manufacturing Technology
MOSIS – Metal Oxide Semiconductor Implementation Service
*Microelectronics Manufacturing Science and Technology (MMST)

Page 2 set us on a 50 year journey


170 THE FUTURE OF COMPUTING PERFORMANCE

Cramming More Components onto Integrated Circuits

GORDON E. MOORE, LIFE FELLOW, IEEE

APPENDIX C

171

Electronics, April 19, 1965: Cramming More Components onto Integrated Circuits; Gordon Moore

P.1

P.2

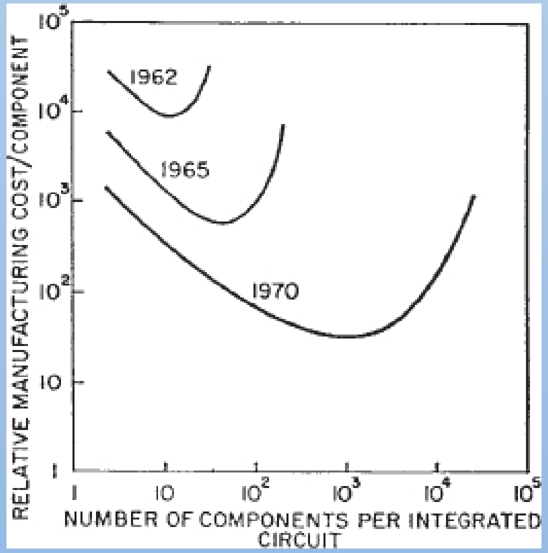


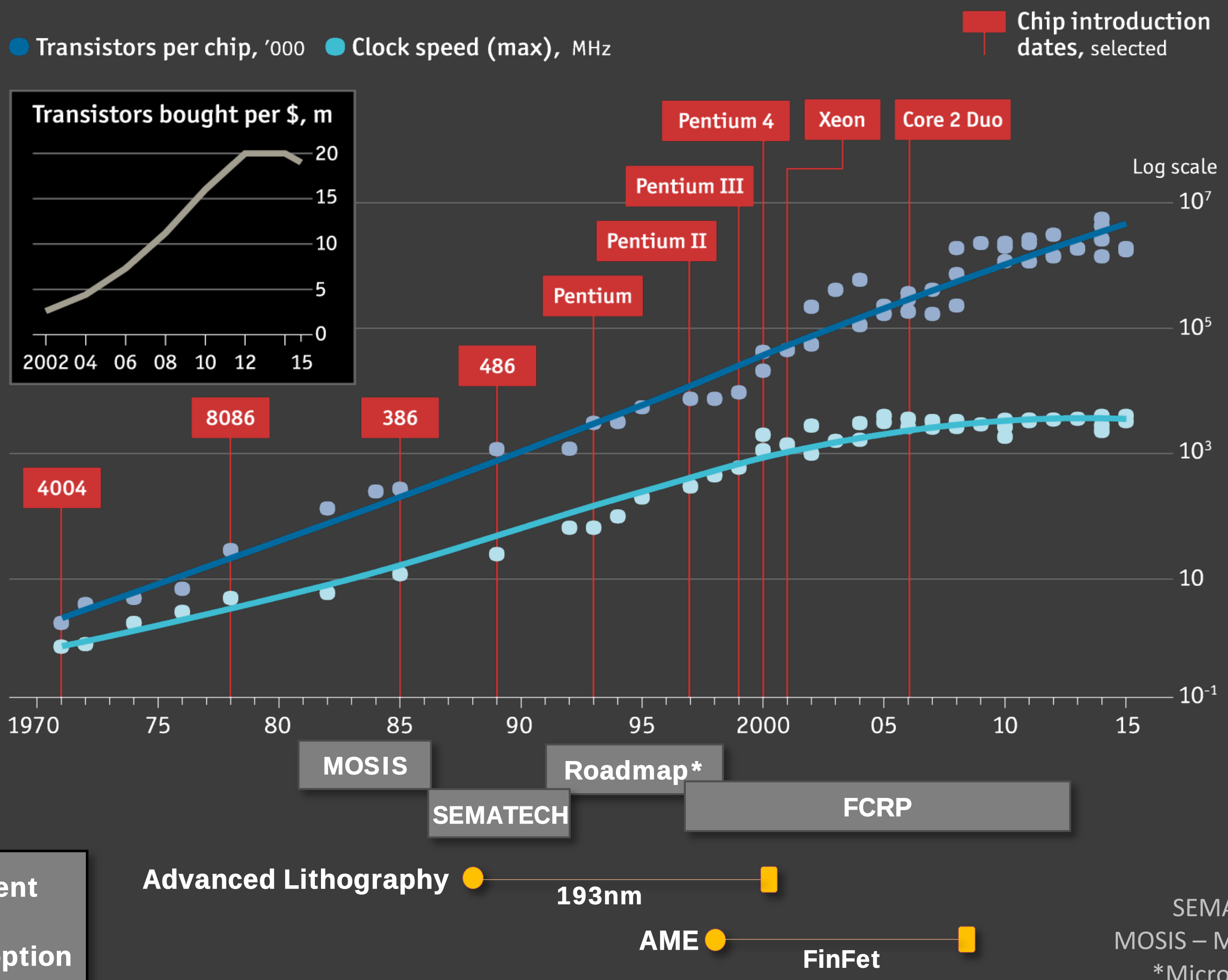
Fig.1

"...The complexity for minimum component costs has increased at a rate of roughly a factor of two per year (see graph)..."

… but nothing lasts forever

“The total cost of making a particular system function must be minimized”

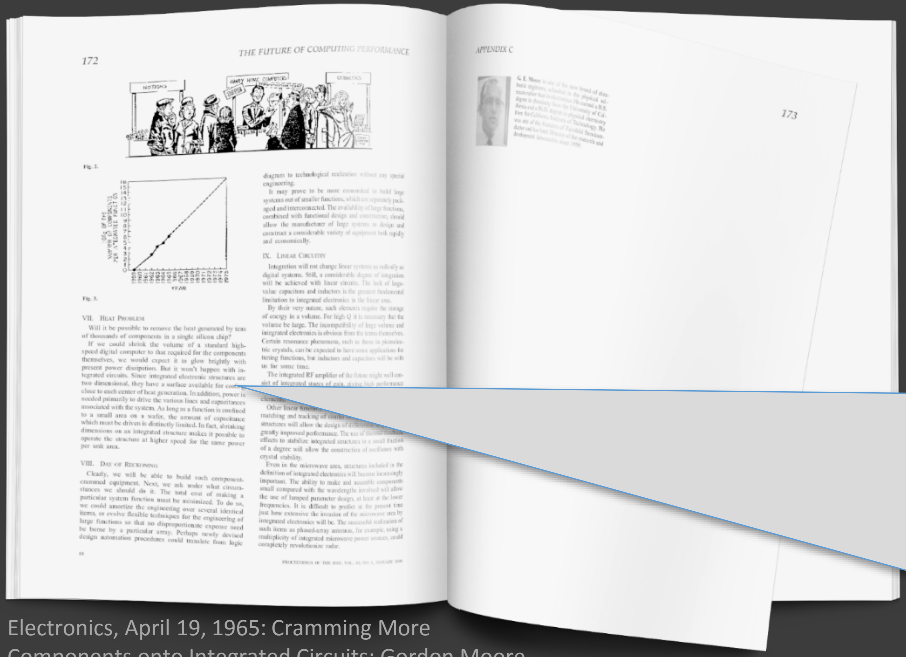
- Gordon Moore



AME – Advanced Microelectronics
FCRP – Focus Center Research Program
FinFET – Fin-Shaped Field Effect Transistor
SEMATECH – Semiconductor Manufacturing Technology
MOSIS – Metal Oxide Semiconductor Implementation Service
*Microelectronics Manufacturing Science and Technology (MMST)

Sources: Intel; press reports; Bob Colwell; Linley Group; IB Consulting; *The Economist*

We need to turn the page



Electronics, April 19, 1965: Cramming More Components onto Integrated Circuits; Gordon Moore

P.3

VIII. Day of Reckoning

Clearly, we will be able to build such component-crammed equipment. Next, we ask under what circumstances we should do it. **The total cost of making a particular system function must be minimized.** To do so, we could amortize the engineering over several identical items, or **evolve flexible techniques for the engineering of large functions** so that no disproportionate expense need be borne by a particular array. Perhaps **newly devised design automation procedures could translate from logic diagram to technological realization** without any special engineering.

It may prove to be **more economical to build large systems out of smaller functions, which are separately packaged** and interconnected. The availability of large functions, combined with functional design and construction, should allow the manufacturer of large systems to design and construct a considerable variety of equipment both rapidly and economically.

Architecture

Maximizing specialized functions

Design

Quickly enabling specialization

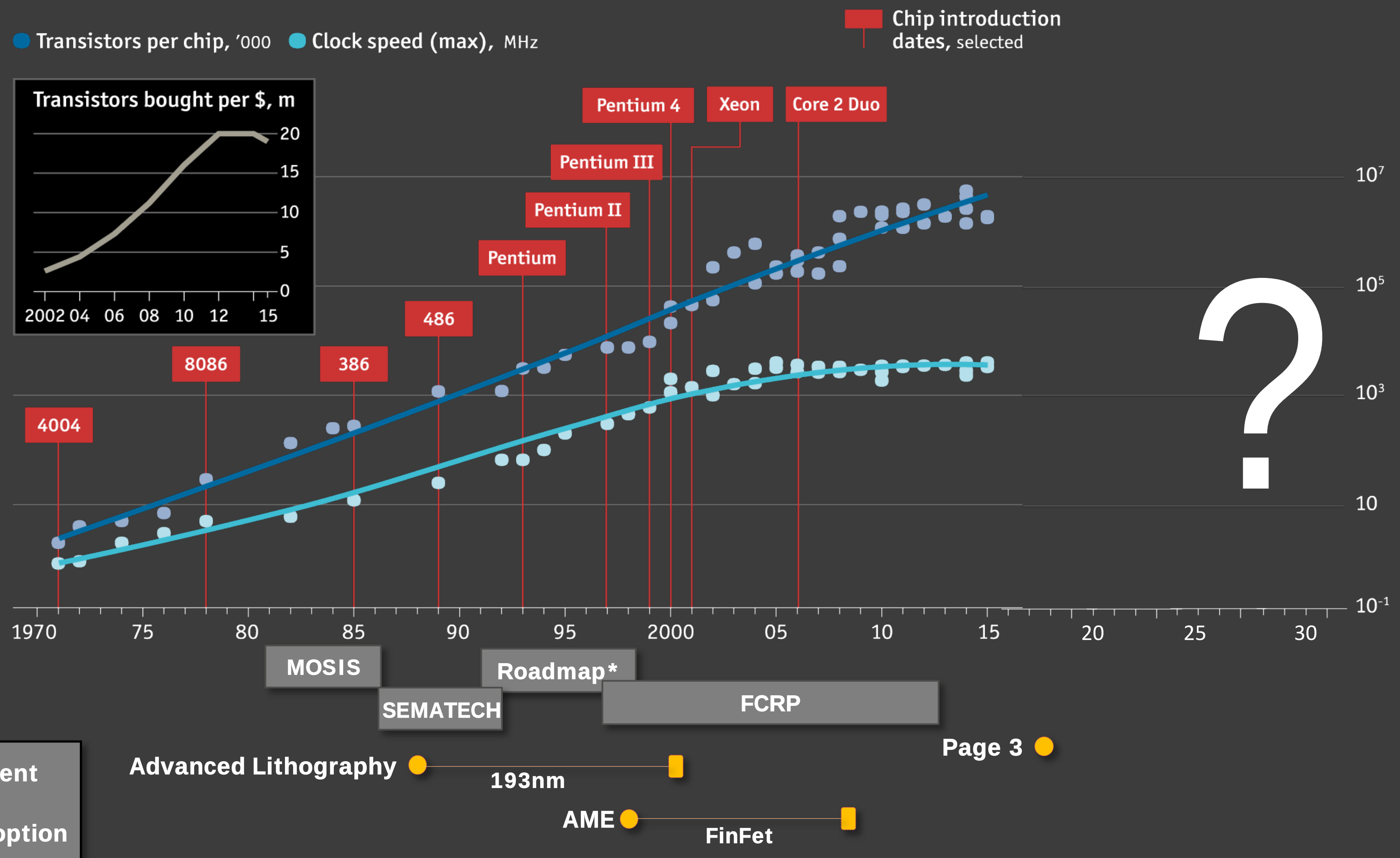
Materials & Integration

Adding separately packaged novel materials and using integration to provide specialized computing

What are the national capabilities we should be investing in next?

"The total cost of making a particular system function must be minimized"

- Gordon Moore

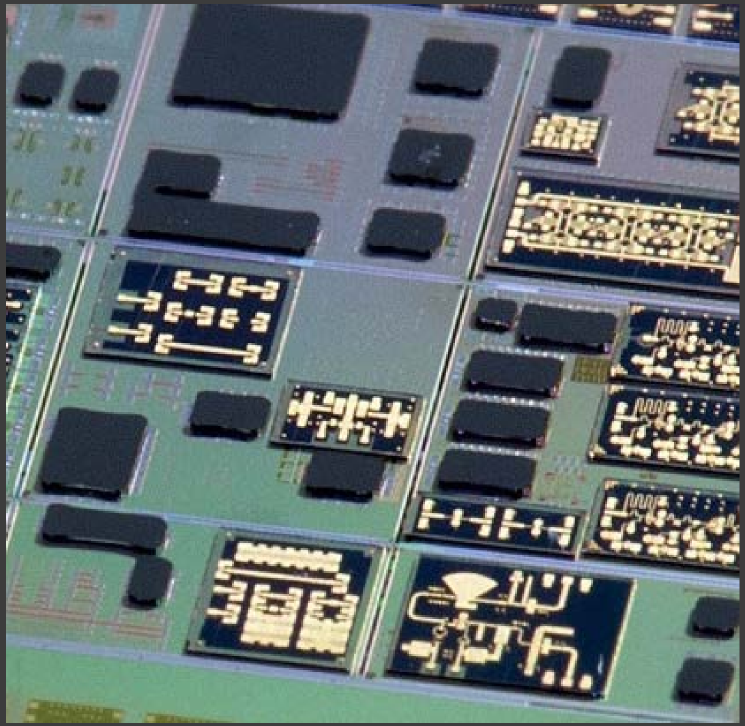


DARPA investment

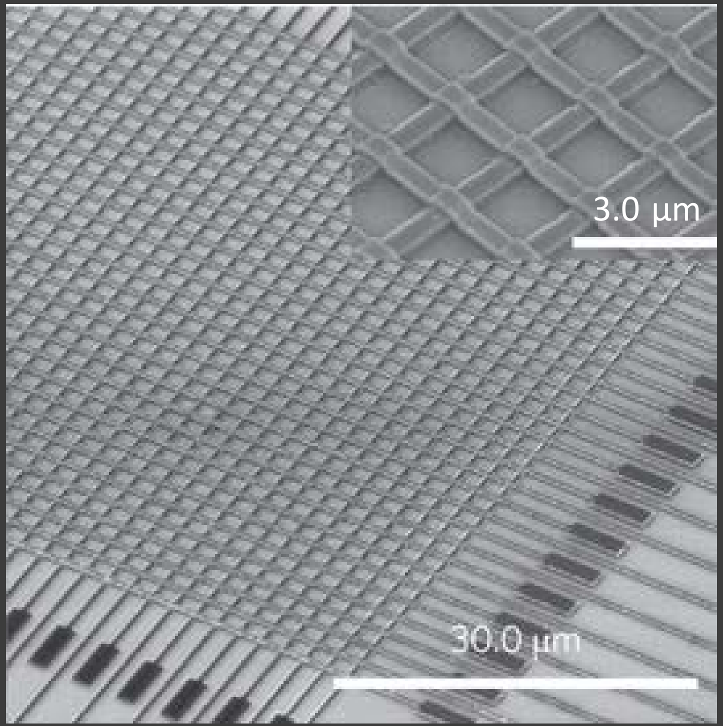
Commercial adoption

Sources: Intel; press reports; Bob Colwell; Linley Group; IB Consulting; *The Economist*

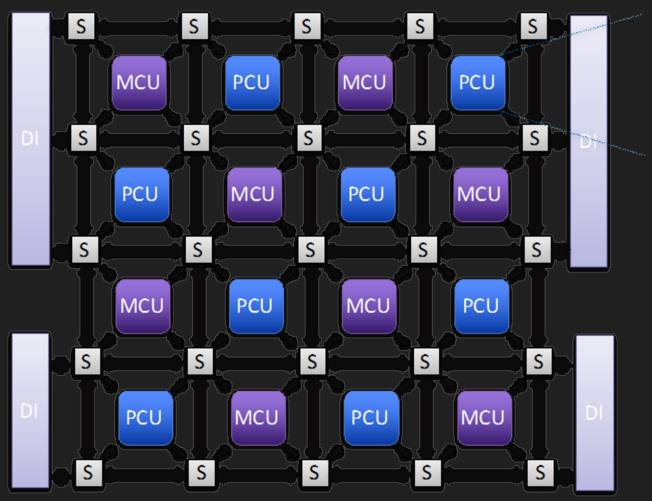
Pseudolithic Integration



Specialized Hardware Blocks



Software Hardware Co-design

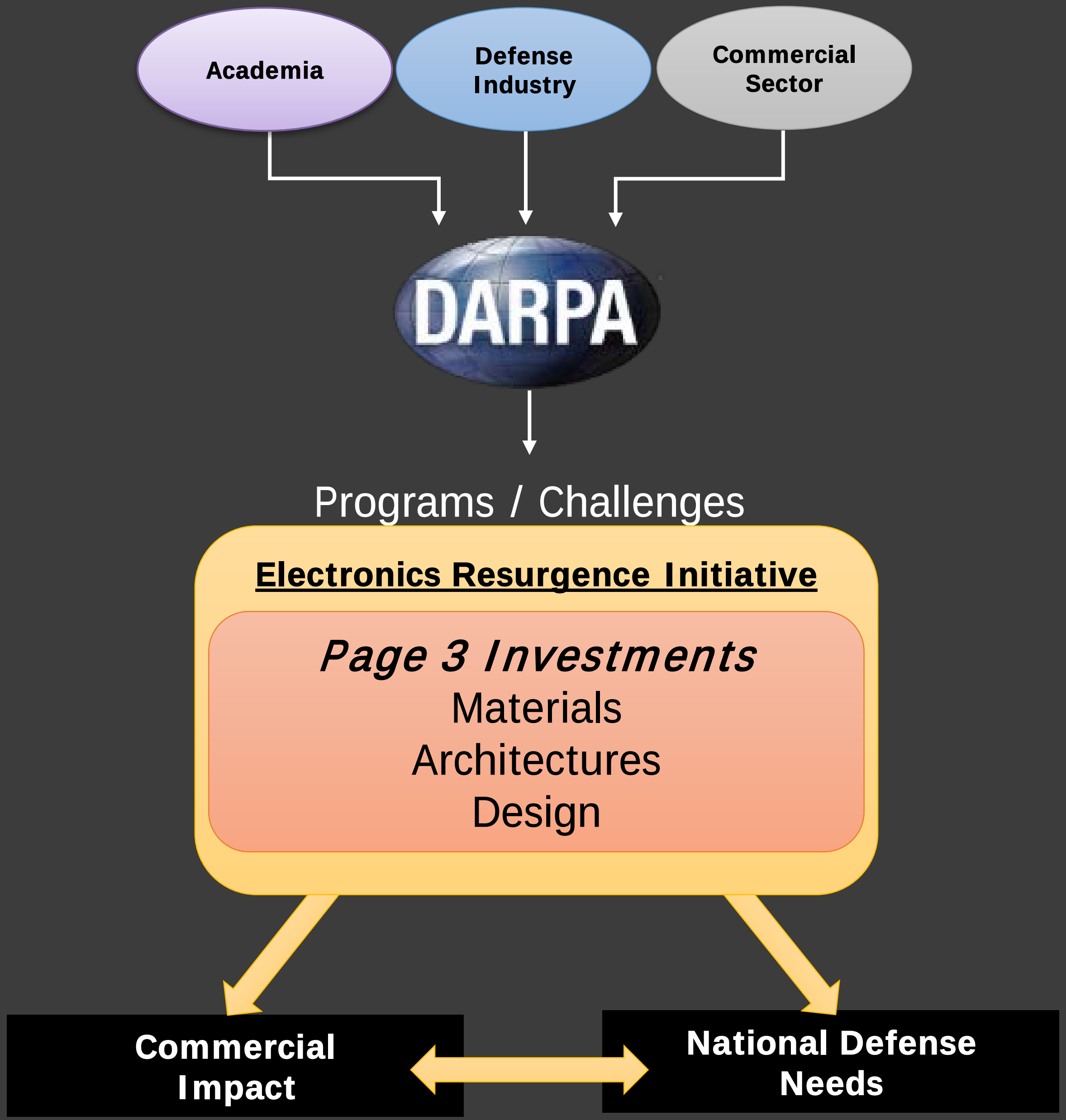


Compiler-directed Hardware Reconfiguration

Sparse + Dense

Where are we heading?

Sowing the seeds for a revolution in processing



What is the initiative?

Program managers hired directly from the electronics community...

Aligning incentives as we both stare at an uncertain future

Co-developing electronics to manage the coming inflection to support both a national electronics base and national defense

MTO ELECTRONICS RESURGENCE INITIATIVE TIMELINE

Launch, Learn, & Organize

6/21: Industry Discussion
7/11: Defense Base Summit

Completed

Summer of Listening

7/18: 2-day workshop on Materials, Architectures, Designs

Completed

Open Competition

9/12: Proposals Requested (Expected)

Fall 2017

Complete Contracting

4/20: Start Work

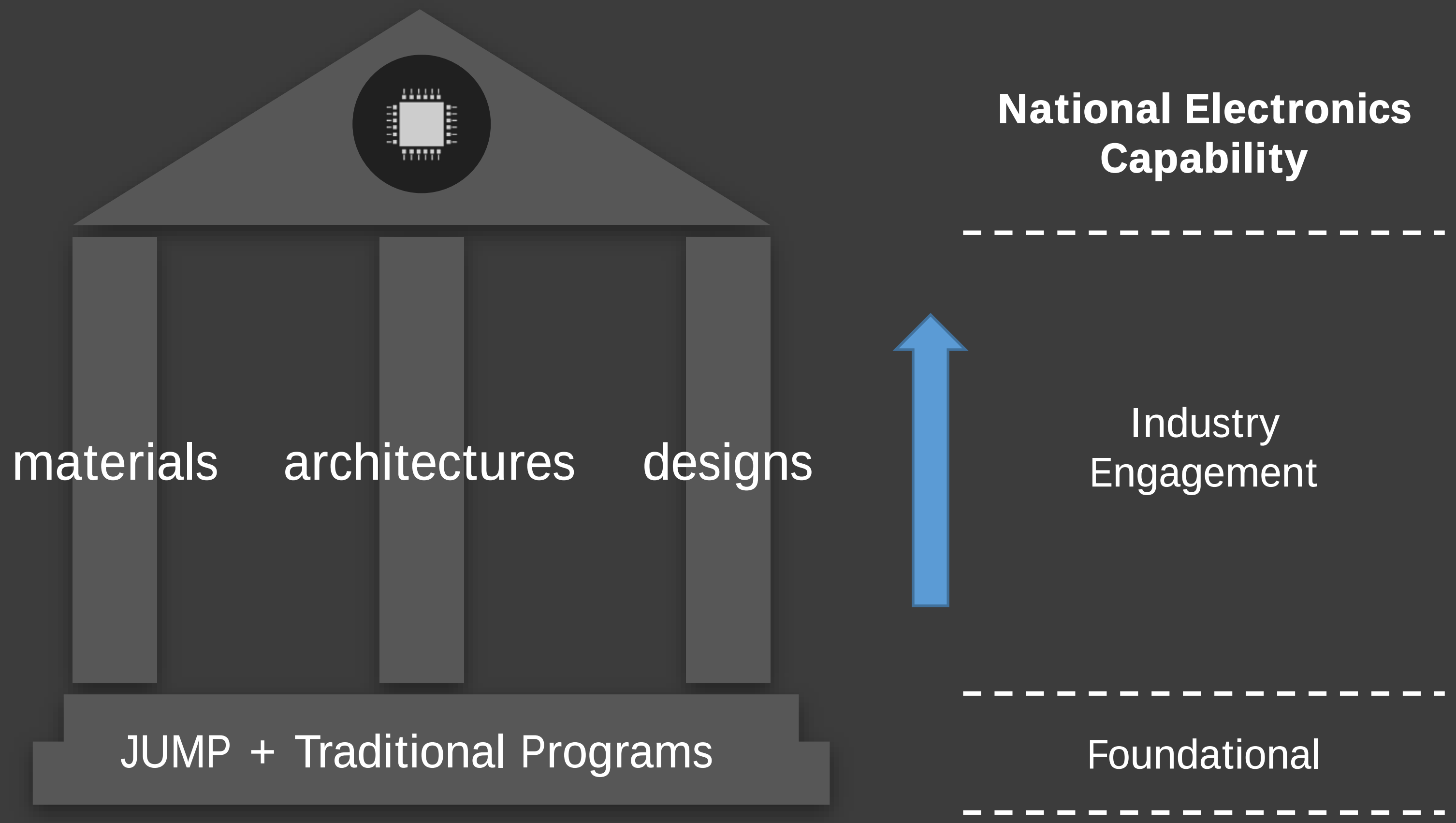
Spring 2018

May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Jan | Apr

Today

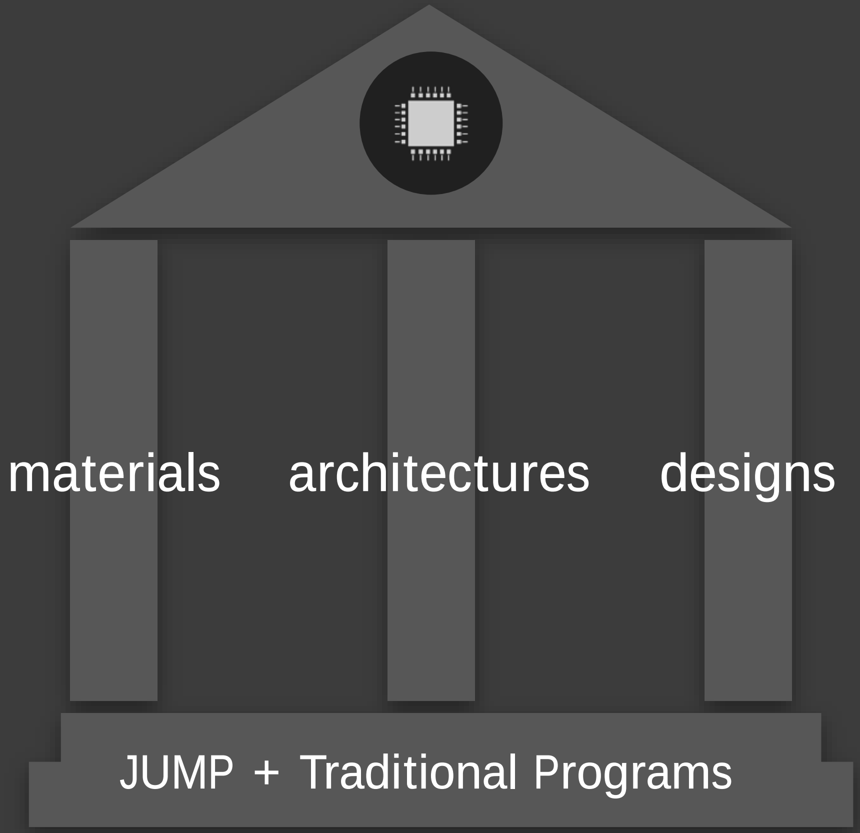
$75 Million Additional in FY18 Budget

Press Release Announcing Initiative


National Electronics Capability
materials
architectures
designs
Industry Engagement
JUMP + Traditional Programs
Foundational

2025 - 2030

NATIONAL ELECTRONICS CAPABILITY



$216 MILLION TOTAL (FY18)

$75 million
Of New Funding
(FY18)

$141 million in Current Efforts (FY18)

Traditional Programs Currently Funded

- **JUMP** – Joint University Microelectronics Program
- **CHIPS** – Common Heterogeneous Integration and IP Reuse Strategies
- **HIVE** – Hierarchical Identify Verify Exploit
- **L2M** – Lifelong Learning Machines
- **N-ZERO -** Near-Zero Power Radio Frequency Receivers
- **CRAFT** – Circuit Realization at Faster Time Scales
- **SSITH** – System Security Integrated Through Hardware and firmware

Joint University Microelectronics Program (JUMP)

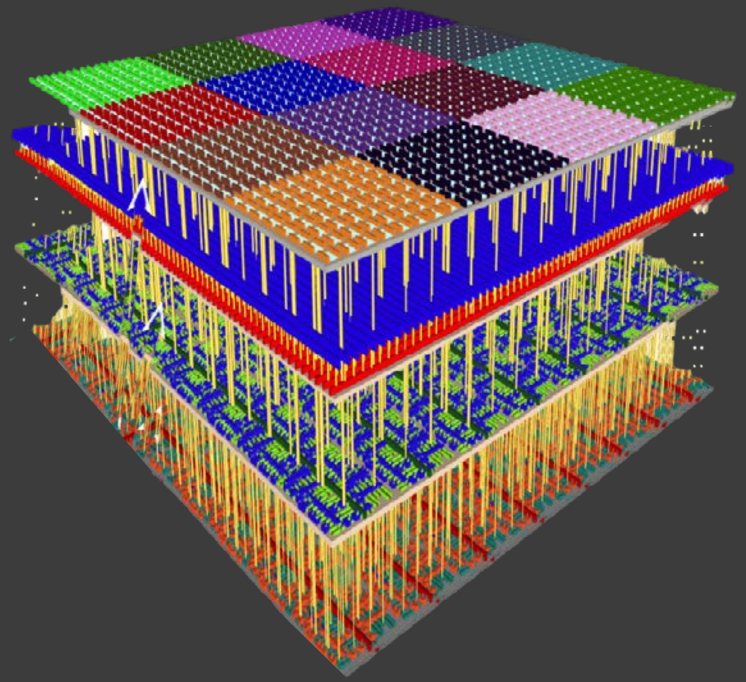
DARPA 40%

Industry 60%

Joint University Microelectronics Program (JUMP)

- RF to THz
- Distributed Computing
- Cognitive Computing
- Intelligent Memory/Storage
- Advanced IC Architectures
- Devices/ Materials

World Class Idea Generation

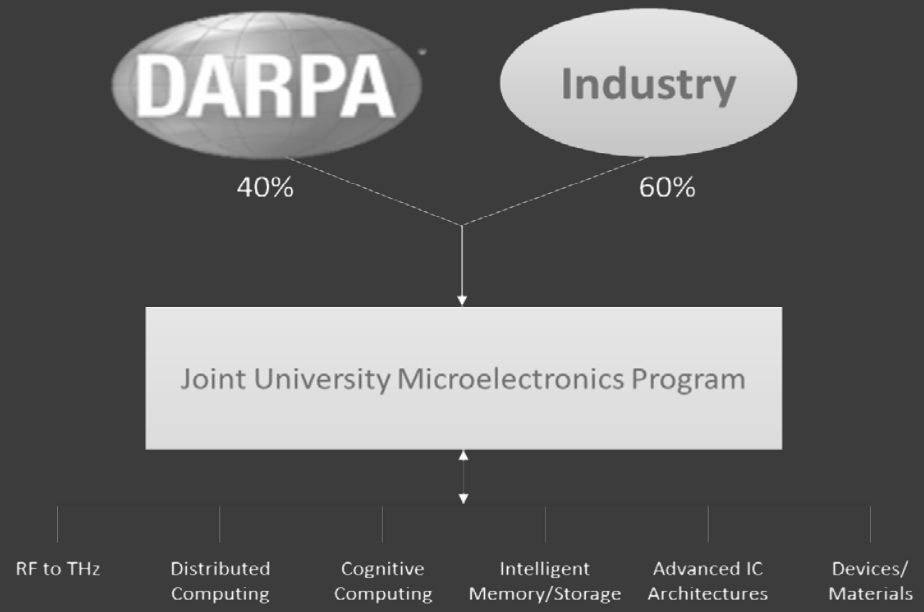


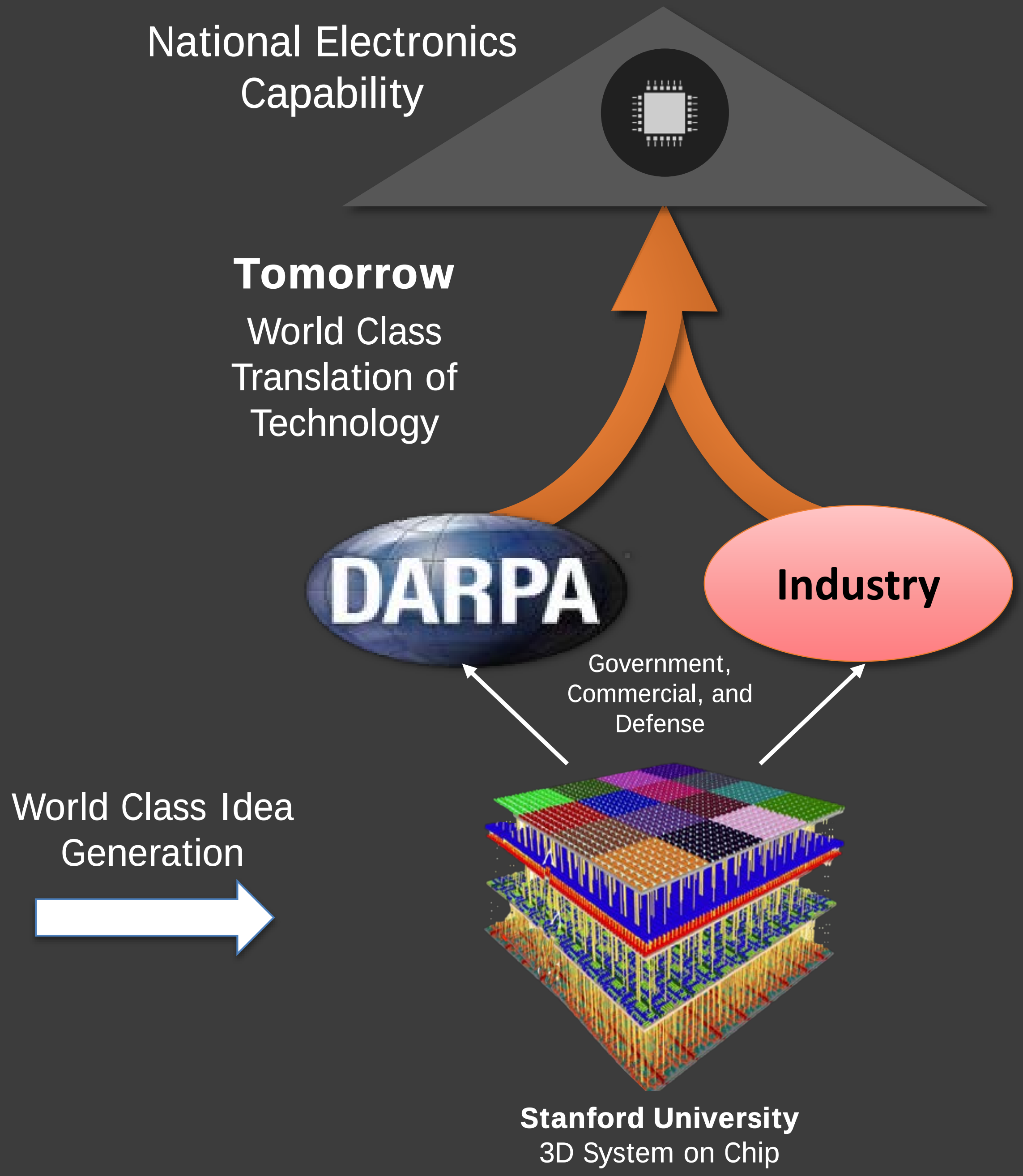
Stanford University
3D System on Chip

Linton Salmon

DARPA Program Manager

The intersection of industry, academics, and government





DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

MTO Electronics Timeline

2016 | 2017 | 2018 | 2019

11/2015
N-ZERO
Kickoff

4/2016
CRAFT
Kickoff

6/2016
CHIPS
Approved

8/22/2016
JUMP
Approved

1/2017
L2M
Approved

4/2017
HIVE
Kickoff

4/2017
SSITH
BAA Released

Today

JUMP
University Driven

+

Page 3 Investments
Industry Driven

+

Traditional Programs

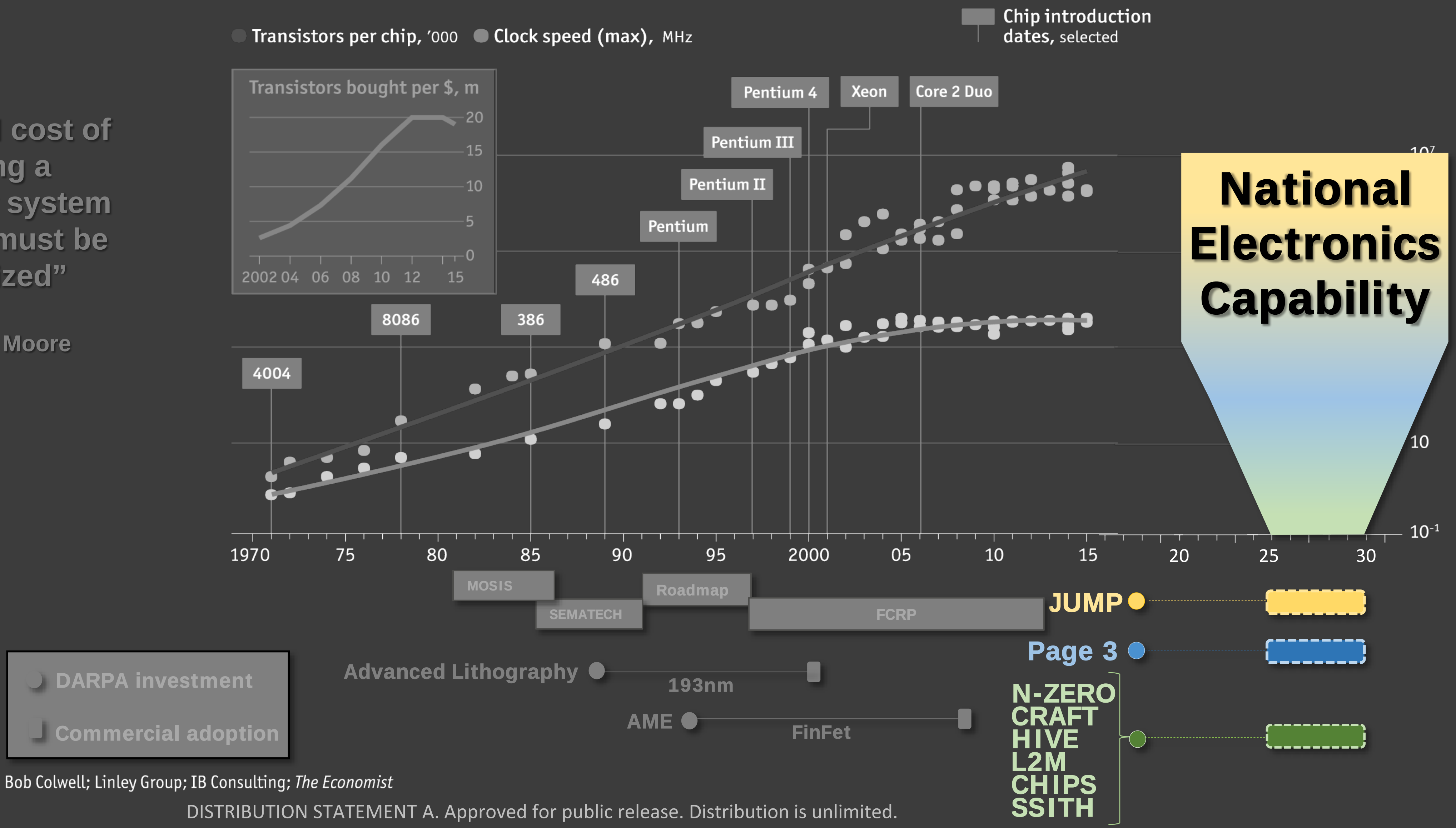
- **N-ZERO**
- **CRAFT**
- **L2M**
- **HIVE**
- **CHIPS**
- **SSITH**

L2M – Lifelong Learning Machines
BAA – Broad Agency Announcement
HIVE – Hierarchical Identify Verify Exploit
CRAFT – Circuit Realization at Faster Timescales
JUMP – Joint University Microelectronics Program N-
ZERO – Near Zero Power RF and Sensor Operations
SSITH – System Security Integrated Through Hardware and Firmware
CHIPS – Common Heterogeneous Integration and IP Reuse Strategies

The goal of the Electronics Resurgence investment today is to reach a national capability between 2025 and 2030

"The total cost of making a particular system function must be minimized"

- Gordon Moore



Sources: Intel; press reports; Bob Colwell; Linley Group; IB Consulting; *The Economist*

So how do you get involved?

Timeline and structure

MTO ELECTRONICS PAGE 3 INVESTMENTS TIMELINE

Launch, Learn, & Organize

6/21: Industry Discussion
7/11: Defense Base Summit

Completed

Summer of Listening

7/18: 2-day workshop on Materials, Architectures, Designs

Completed

Open Competition

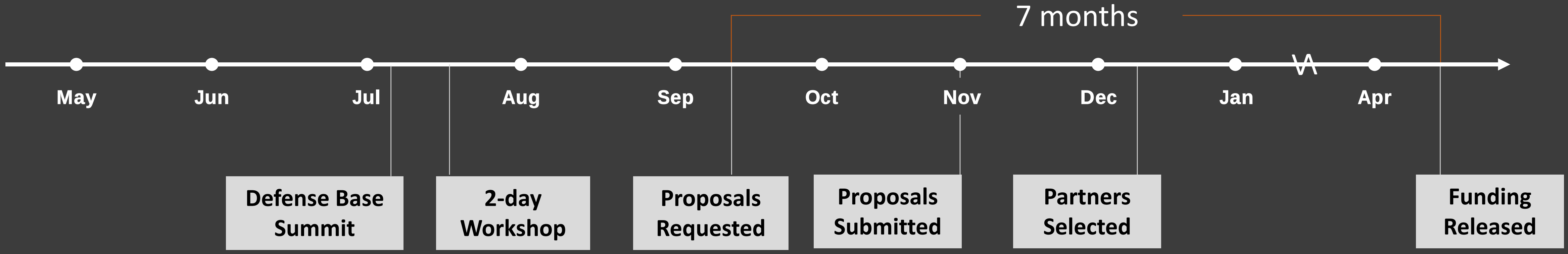
9/12: Proposals Requested (Expected)

Fall 2017

Complete Contracting

4/20: Start Work

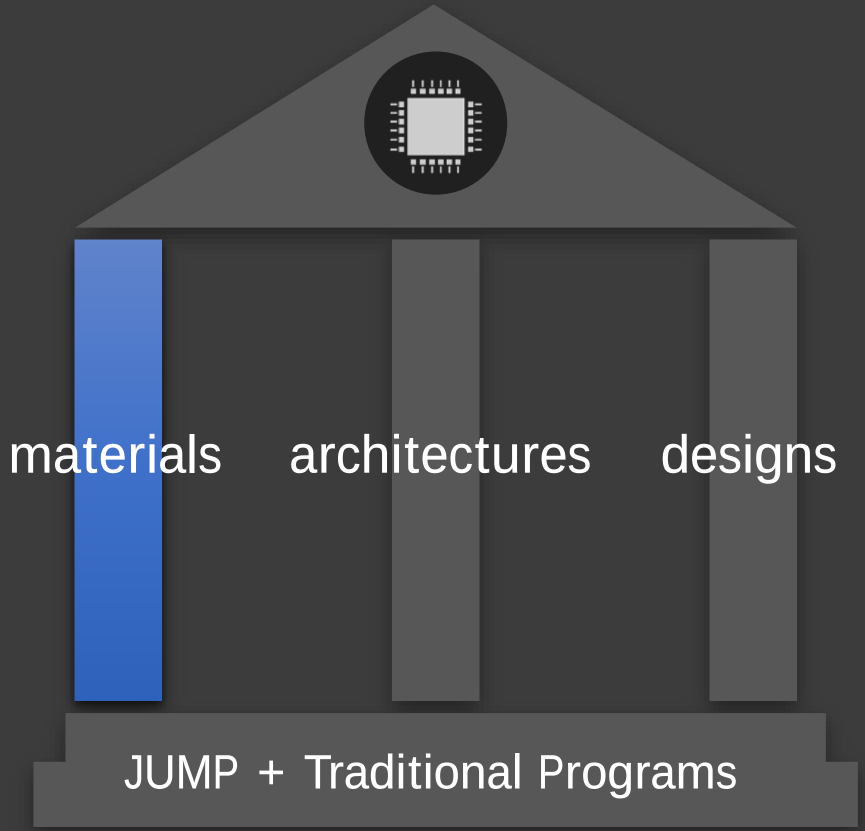
Spring 2018



Dan Green

Materials

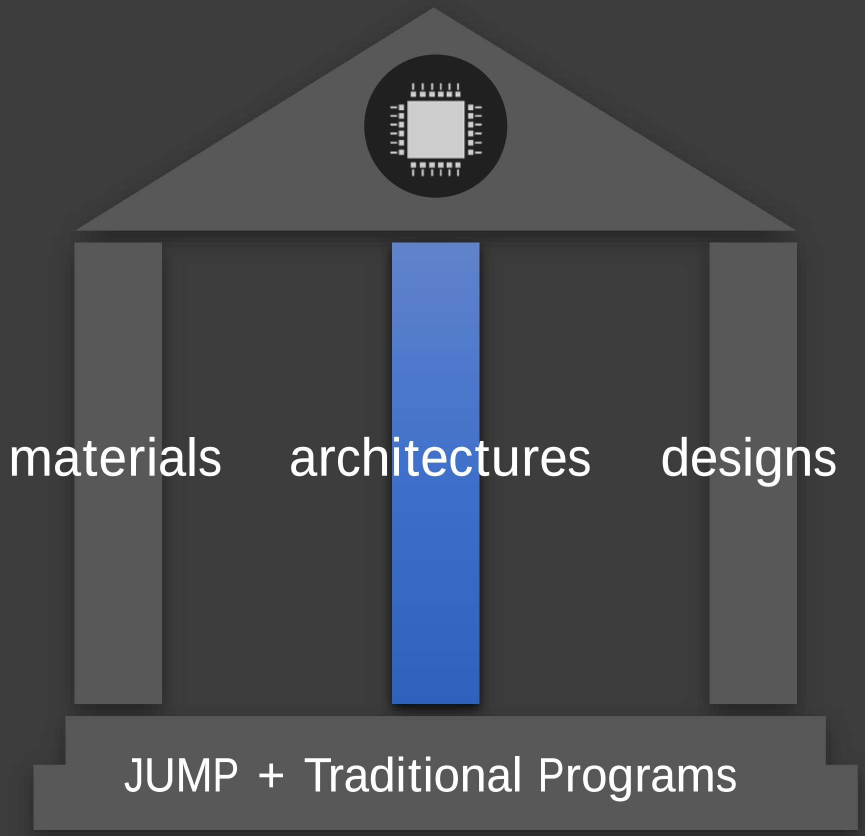
Steering the science of materials to commercial product lines



Tom Rondeau

Architectures

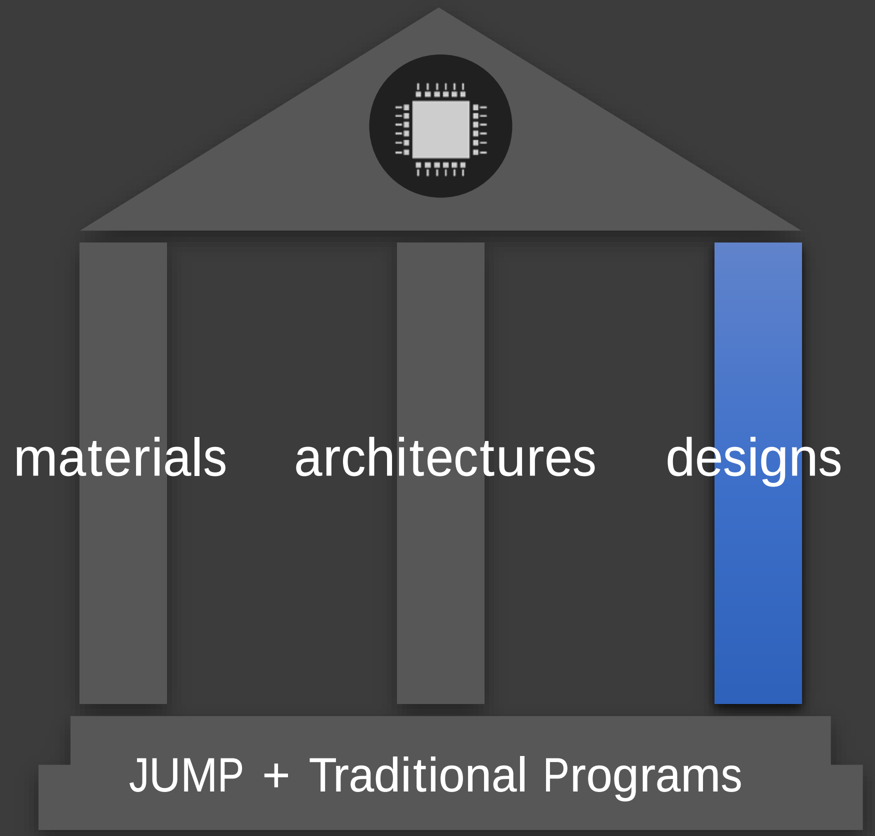
The intersection of connectivity and computation



Andreas Olofsson

Designs

From Kickstarter to Supercomputer



ENSURING LONG-TERM U.S. LEADERSHIP IN SEMICONDUCTORS

ELECTRONICS RESURGENCE INITIATIVE WORKSHOP

FAIRMONT SAN JOSE, 170 S MARKET ST, SAN JOSE, CA 95113

JULY 18 – JULY 19, 2017

CRAIG MUNDIE


REPORT TO THE PRESIDENT
Ensuring Long-Term U.S. Leadership in Semiconductors

Executive Office of the President
President's Council of Advisors on Science and Technology

January 2017

PCAST WORKING GROUP

Co-Chairs

John Holdren* Director, OSTP	Assistant to the President for Science and Technology &
Paul Otellini	Former President and CEO Intel

Industry Working Group Members

Richard Beyer	Former Chairman and CEO Freescale Semiconducto
Ajit Manocha	Former CEO Global Foundries
Wes Bush	Chairman, CEO, and President Northrop Grumman
Jami Miscik	Co-CEO and Vice Chairman Kissinger Associates
Diana Farrell	President and CEO JP Morgan Chase Institute
Craig Mundie*	President Mundie & Associates
John Hennessy	President Emeritus Stanford University
Mike Splinter	Former CEO and Chairman Applied Materials
Paul Jacobs	Executive Chairman Qualcomm
Laura Tyson	Distinguished Professor - Graduate School UC Berkeley

CHALLENGES AND OPPORTUNITIES

- TECHNOLOGICAL BARRIERS TO LOWER-POWER AND SCALING
- RAPIDLY SHIFTING GLOBAL MARKETS
- STRATEGIC POLICY AND FINANCIAL INVESTMENTS OUTSIDE USA
- MARKET ACCESS CONSTRAINTS
- UNEVEN INTELLECTUAL PROPERTY ENFORCEMENT
- FAB CAPACITY IN USA NOW LESS THAN 13%
- DESIGN COMPLEXITY AND DEGREE OF SPECIALIZATION INCREASING

WE'VE SEEN THIS MOVIE BEFORE…

- IN THE 1980'S JAPAN WAS OVERTAKING THE U.S. IN MEMORY CIRCUITS
- BUT, THE MARKET WAS SHIFTING, DRIVEN BY MICROPROCESSOR ADVANCES
- THE USA POLICY AND INDUSTRY FOCUS WAS ON SPEED IMPROVEMENT AND TECHNOLOGY FUNDAMENTALS, AND THE JAPANESE FELL BEHIND
- KOREA, MORE RECENTLY, HAS MADE BIG INVESTMENTS
- CHINA IS INVESTING STRATEGICALLY
- A SUCCESSFUL U.S. STRATEGY TODAY MUST BE DIFFERENT…

WIN THE RACE BY RUNNING FASTER!

- PICK FOCUS AREAS – MOONSHOTS
- APPLICATIONS-DRIVEN APPROACH
- TEN-YEAR TIME HORIZON
- GOVERNMENT INVESTMENT SHOULD COMPLEMENT NATURAL INDUSTRY INVESTMENT AREAS
- REDUCE DESIGN COSTS WITH RADICAL ADVANCES IN DESIGN TOOLS AND REUSABILITY – GOAL SHOULD BE 10X TO 100X REDUCTIONS IN TIME AND COSTS

BUT IT'S LIKE PLAYING 3D CHESS…

- THE RULES OF THE GAME ARE DETERMINED BY THE APPLICATION DOMAIN
- THE PLANES OF THE GAME INCLUDE:
  - Computing Modalities
  - Computing Architectures
  - Component Technologies

APPLICATION DOMAIN LEADERSHIP & SUPPORT ROLES

STRONG TECH INDUSTRY INTEREST *(GOVERNMENT SUPPORT)*

- Big Data Analytics: Local real-time data analysis and visualization enabled by advances in security, low-power computation, and processor specialization.
- Artificial Intelligence and Machine Learning: Supervised and unsupervised machine learning enabled by new processors, including low-power processers, graphics processing units, and quantum computers.
- Biotechnologies, Human Health Technologies: Medical implants that are capable of ultra-low power processing, communications, and wireless charging.
- Robotics, Autonomous Systems: Speech and image recognition for mobile computing.
- Telepresence, Virtual Reality, Mixed Reality: Local real-time sensory input, such as video and graphics.
- Machine Vision: Imaging-based automatic inspection and analysis for applications such as process control and robot guidance.
- Speech Recognition and Synthesis: Portable systems enabling recognition and artificial production of human speech.
- Nanoscale Systems and Manufacturing: Democratized, small-batch fabrication structures at the nanoscale using a variety of material classes. Nanoscale 3D Printers will provide desktop fab capabilities for rapid prototyping, additive manufacturing, moving beyond silicon and interfacing with soft matter.
- Ultra-High Performance Wireless: Wireless systems with very low latency and extremely reliable communications, for example, between autonomous vehicles.
- Holistic Secure Systems: hardware-based defense in-depth, such as tamper resistant hardware that electronically authenticates software integrity.

WEAKER TECH INDUSTRY INTEREST *(GOVERNMENT LEADERSHIP)*

- Computational Chemistry: Design of novel solutions for catalysis, low-temperature nitrogen fixation, etc.
- Advanced Materials Science and Manufacturing: Simulation of solid state materials, etc.
- Modeling and Simulation: Efficient exascale computing to enable advanced earthquake prediction (CMOS-based high-performance computing capable of 1-10 exaflops), high-fidelity weather modeling (superconducting-based hyperscale computing capable of 10-100 exaflops), and optimization problems (quantum computing).
- Space Technologies: Radiation hardness through circuit design and technologies (e.g., wide-bandgap electronics) rather than special manufacturing processes (e.g., insulating substrates or shielding).

TAKING A FULL-STACK APPROACH

DOMAIN BY DOMAIN…

1. ULTIMATE SOFTWARE APPLICATION
2. APPLICATION PROGRAMMING MODEL
3. PLATFORM SOFTWARE SERVICES
4. PLATFORM PROGRAMMING MODEL
5. OPERATING SYSTEMS SERVICES
6. COMPUTER SYSTEM ARCHITECTURES (PROCESSING, STORAGE, AND INTERCONNECT AT EVERY SCALE)
7. COMPONENT TECHNOLOGIES

COMPUTING MODALITIES

EMBEDDED SYSTEMS: SPECIALIZED SEMICONDUCTORS, RANGING FROM HIGH-VOLUME/LOW-COST FOR APPLICATIONS LIKE INTERNET OF THINGS (IOT) DEVICES TO LOW-VOLUME/HIGH-COST SEMICONDUCTORS FOR ROBOTICS OR DEFENSE SYSTEMS. POWER EFFICIENCY REQUIREMENTS WILL VARY BY APPLICATION (HARVESTING ENERGY FROM THE AMBIENT ENVIRONMENT VERSUS DEDICATED POWER SOURCES, RESPECTIVELY). FLEXIBILITY AND AGILITY IN FABRICATION AND DESIGN WILL BE NEEDED TO MAINTAIN PROFITABILITY.

PERSONAL/PORTABLE SYSTEMS: DESKTOP, MOBILE, AND WEARABLE COMPUTING DEVICES. THESE ARE FREQUENTLY BATTERY-POWERED COMPUTATIONAL DEVICES, WHICH WILL BE OPTIMIZED FOR PERFORMANCE, PRICE, AND POWER EFFICIENCY. GENERAL PURPOSE COMPUTING WILL BE AUGMENTED BY ACCELERATORS, SENSOR ADD-ONS, AND OTHER FUNCTION-AUGMENTING ICT'S.

HYPERSCALE SYSTEMS: SUPERCOMPUTING DEVICES FOR "REMOTE" COMPUTATION THAT WILL BE AGGREGATED TO FORM THE MOST POWERFUL SYSTEMS THAT CAN BE PRODUCED IN EACH ARCHITECTURAL CLASS. THESE SYSTEMS ARE EXPECTED TO SOLVE OTHERWISE INTRACTABLE PROBLEMS; OR, FOR CLASSICAL ARCHITECTURES, TO MAXIMIZE PERFORMANCE WITHIN PRACTICAL POWER CONSTRAINTS. EMERGING ARCHITECTURES PROVIDING NEW CAPABILITIES AND DOMAIN-SPECIFIC OPTIMIZATIONS WILL BECOME INCREASINGLY IMPORTANT AS PERFORMANCE INCREASES LAG AND PRACTICAL POWER LIMITS ARE REACHED IN TRADITIONAL COMPUTING ARCHITECTURES.

COMPUTER SYSTEM ARCHITECTURES

VON NEUMANN: CHANGES IN TECHNOLOGY TO ACCOMMODATE POST-MOORE'S LAW REALITIES, SUCH AS MULTI-CORE CPUS WITH DIFFERENT, COMPLEX MEMORY HIERARCHIES, WILL DEMAND NEW ENGINEERING PARADIGMS ACROSS THE EXISTING RANGE OF TRADITIONAL VON NEUMANN ARCHITECTURES FOR DIGITAL COMPUTATION.

QUANTUM: QUANTUM COMPUTING HAS THE POTENTIAL TO SUBSTANTIALLY ADVANCE OUR COMPUTE CAPABILITIES AND SOLVE CURRENTLY INTRACTABLE PROBLEMS. THERE ARE SEVERAL QUANTUM ARCHITECTURAL APPROACHES WHICH MAY SUPPORT DIFFERENT STRATEGIC DOMAINS, AND ALONG DIFFERENT TIMELINES. THESE APPROACHES, IN ROUGH ORDER OF LIKELY DEPLOYMENT, ARE: ANALOG QUANTUM SIMULATION; ADIABATIC QUANTUM ANNEALING; AND CIRCUIT-BASED QUANTUM COMPUTING.

BIO/NEURO-INSPIRED (NEUROMORPHIC COMPUTING): BIOLOGICALLY-INSPIRED POWER CONSUMPTION AND "TOPOLOGY" OF THE CIRCUITRY (USING THREE DIMENSIONS, MORE LIKE THE BRAIN), ANALOGOUS TO HOW RADIO NETWORKS ARE NOW DESIGNED IN THE POST-SHANNON LIMIT ERA.

ANALOG COMPUTING: ANALOG COMPUTING APPROACHES PREDATE DIGITAL COMPUTING AND IN THEORY CAN SOLVE SOME PROBLEMS THAT ARE INTRACTABLE ON DIGITAL COMPUTERS. IN PRACTICE, DIGITAL COMPUTING TECHNIQUES HAVE OVERTAKEN ANALOG COMPUTING, BUT ADVANCES IN NOISE MINIMIZATION COULD ALLOW SOLUTIONS IN SOME AREAS.

SPECIAL PURPOSE ARCHITECTURES: FIELD-PROGRAMMABLE GATE ARRAYS, GRAPHICS PROCESSING UNITS, AND DEEP LEARNING/MACHINE LEARNING ACCELERATORS, INCLUDING FOR EDGE COMPUTING.

APPROXIMATE COMPUTING: PERFORMING BOUNDED APPROXIMATION INSTEAD OF EXACT CALCULATIONS FOR ERROR-TOLERANT TASKS (SUCH AS MULTIMEDIA PROCESSING, MACHINE LEARNING, AND SIGNAL PROCESSING), SIGNIFICANTLY INCREASING EFFICIENCY AND REDUCING ENERGY CONSUMPTION.

COMPONENT TECHNOLOGY VECTORS AND TIMELINES

1 TO 4 YEARS

- Neuromorphic
- Photonics
- Advanced and Quantum Sensors
- CMOS Sub 7nm and 3D structures
- Magnetic Flash and DRAM Memories
- 3D Wafer Stacking
- 5G wireless technologies

5 TO 7 YEARS

- Magnetic SRAM
- 3D Die-to-Wafer Stacking
- 3D Monolithic Fab
- Advanced non-volatile SRAM
- Carbon Nanotubes
- Phase Change Materials
- Biotech-to-electronic interfaces
- Superconducting Logic, Interconnects and Storage

7 TO 10+ YEARS

- 6G wirelesss technologies
- Quantum Computers
- DNA Storage

WE HAVE MORE THAN ENOUGH TECHNOLOGIES

WE JUST HAVE TO PICK A FEW BIG PROBLEMS TO DRIVE THEM INTO COMMERCIALIZATION

Data, Computation, and Electronics

Wade Shen
DARPA Program Manager

18 July 2017

DARPA

Background

- I2O = Information Innovation Office @ DARPA

- Data analysis and machine learning for national security:
  - Detecting ceasefire violations in Yemen
  - Finding human traffickers from their online ads
  - Machine learning that patches bugs in real-time
  - Tracking targets at the speed of a bullet
  - Machine learning that builds machine learning

- Why do we need better compute capabilities?

Detecting ceasefire violations in Yemen

- Data: Publicly available social media + seismic activity data from WWSSN

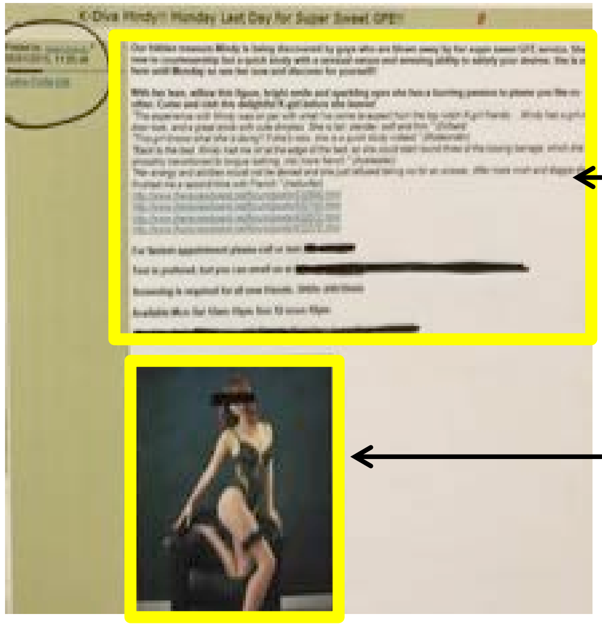
1. Anomaly detection finds events via social media and seismic data

2. Image understanding helps characterize the event

WWSN - World Wide Seismograph Network

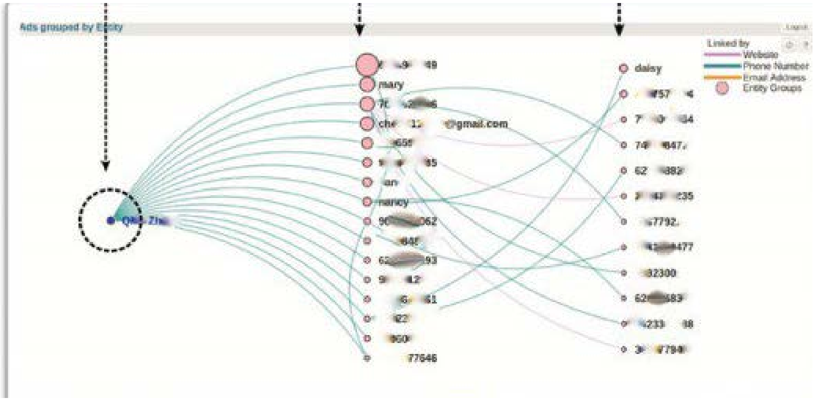
Machine learning for detection of trafficking

Ads and reviews posted online



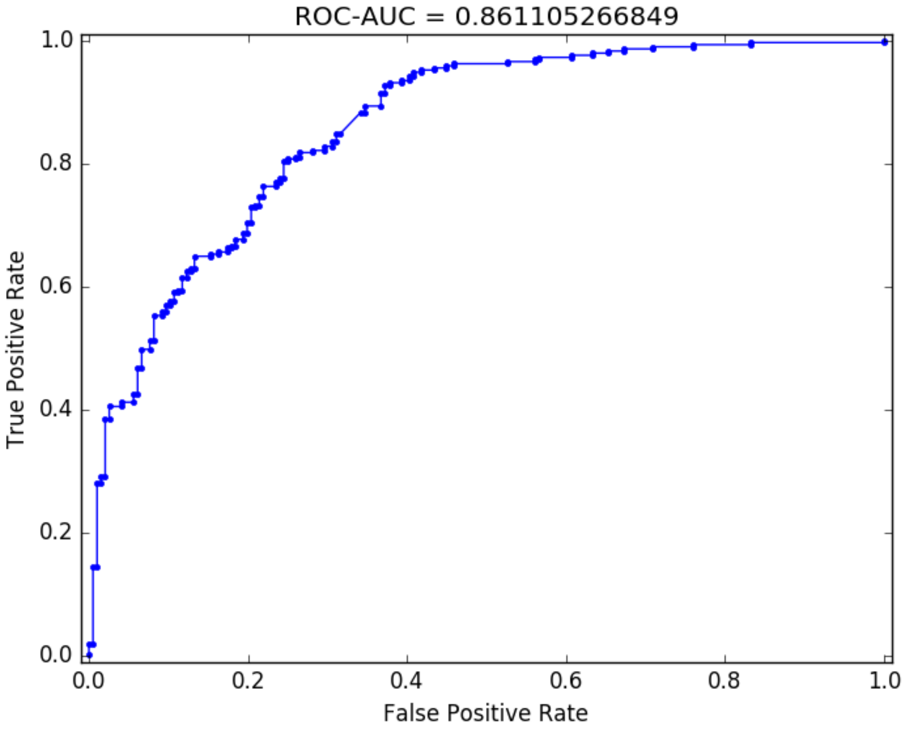
Text helps identify authors and pricing behaviors

Images indicate signs of physical abuse and age



Author networks help discover latent trafficking rings

Predicting trafficking vendors from ad behaviors



400+ arrests, 16+ convictions

Memex Missing Persons Watch (MMPW)

- 1 in 6 missing persons become sex trafficking victims
  [National Center for Missing & Exploited Children (NCMEC)]

- MMPW continuously monitors online prostitution ads for missing persons
  - Compares ad photos vs. missing person photos
  - Alerts when missing person emerges in online advert

NCMEC Photo


MISSING
HELP BRING ME HOME
[redacted]

Missing Since: Dec 16, 2015
Missing From: Greenville, IL
DOB: Feb 14, 2000
Age Now: 17
Race: Black
Hair Color: Brown
Eye Color: Brown
Height: 5'9"
Weight: 180 lbs

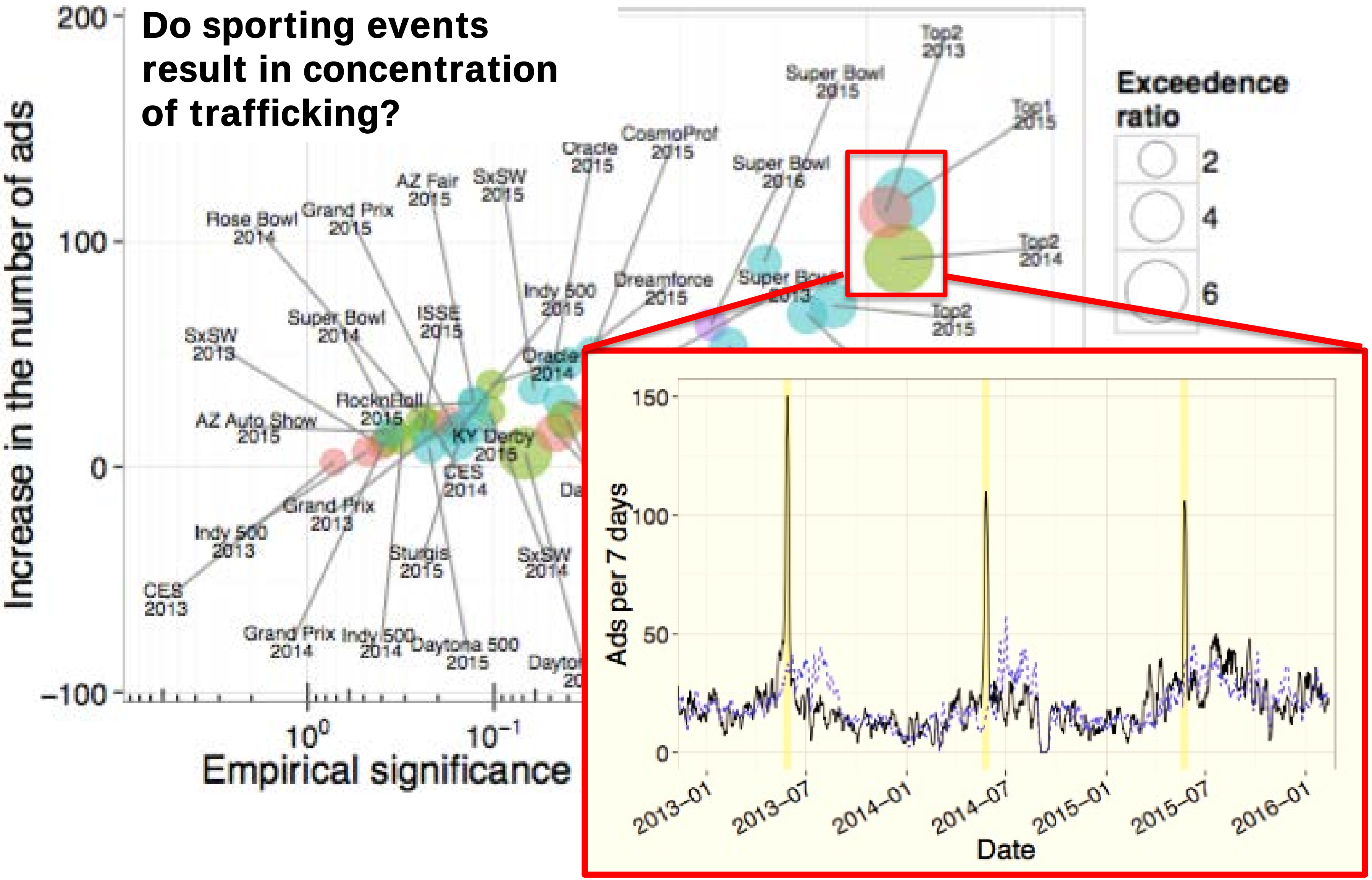
Both photos shown are of [redacted] The child may travel to Belleville, Illi
[redacted] 's nose is pierced. The child has multiple tattoos, including the

MEMEX Ad

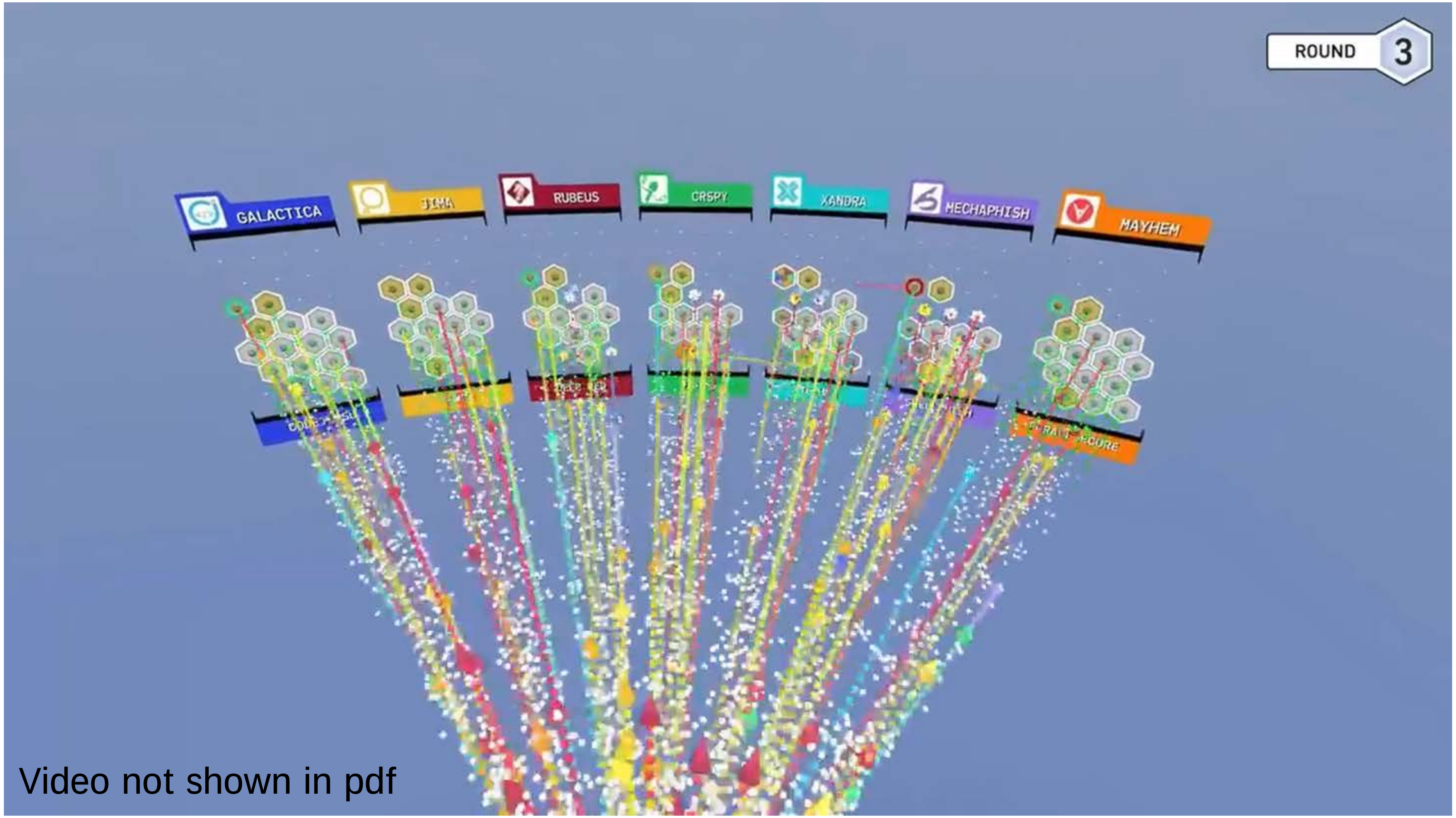
MMPW automatically discovered 4 missing persons searching 17M faces/day

Prevalence estimation for sex trafficking

Do sporting events result in concentration of trafficking?



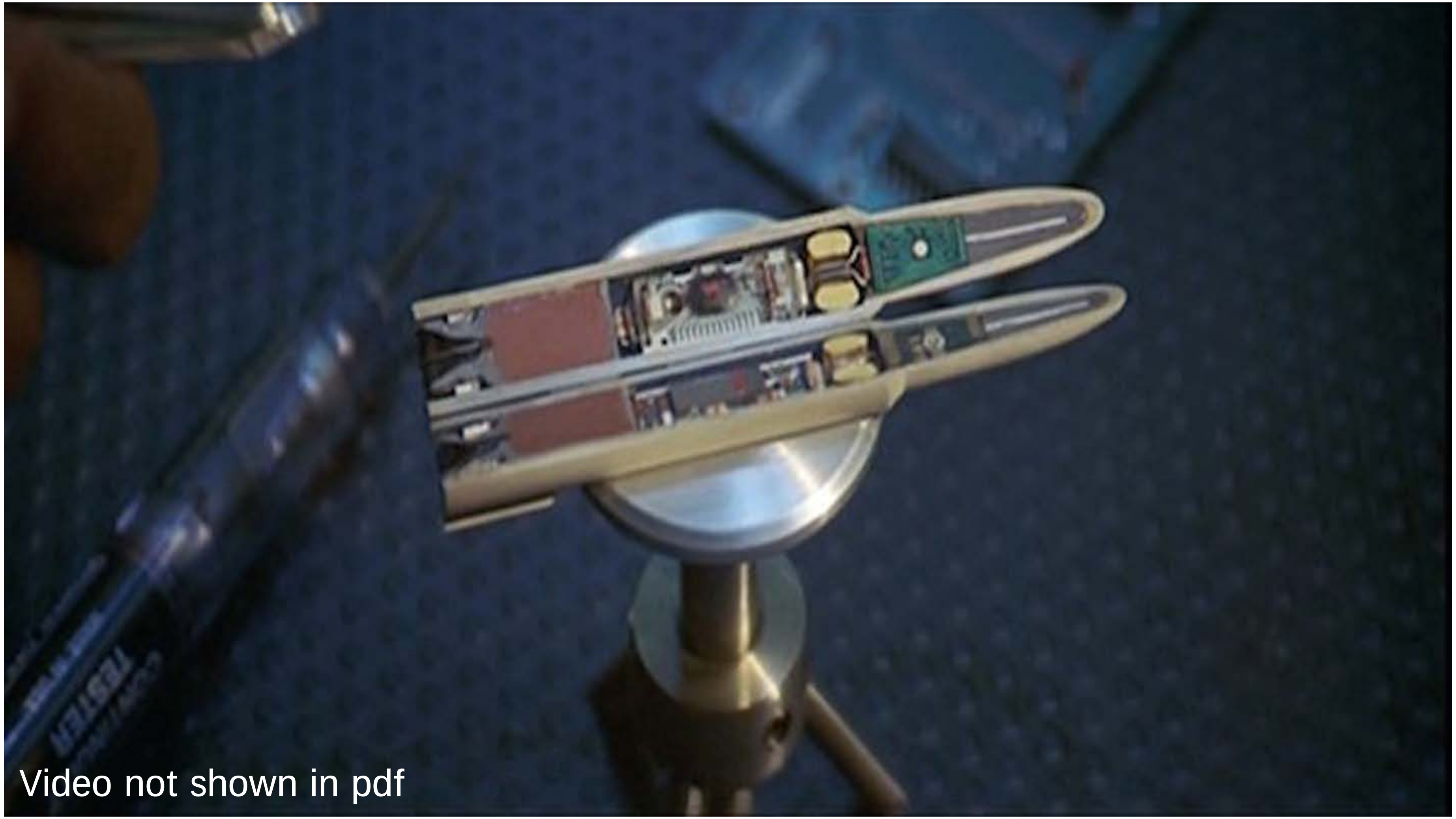
CGC: Finding bugs at machine speed



Video not shown in pdf

Image source: https://www.youtube.com/watch?v=v5ghK6yUJv4

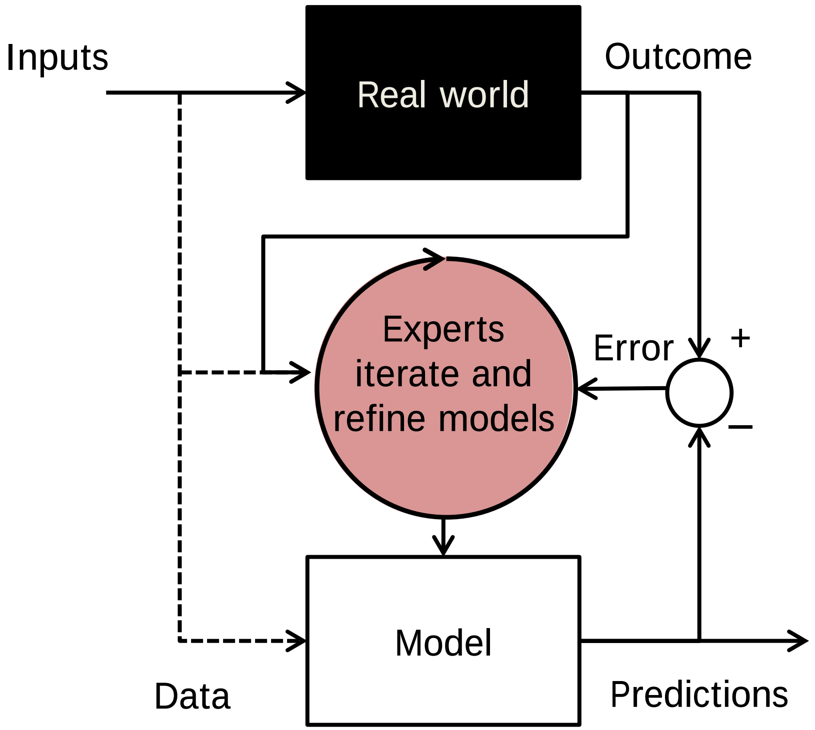
Exacto: Target tracking at the speed of a bullet



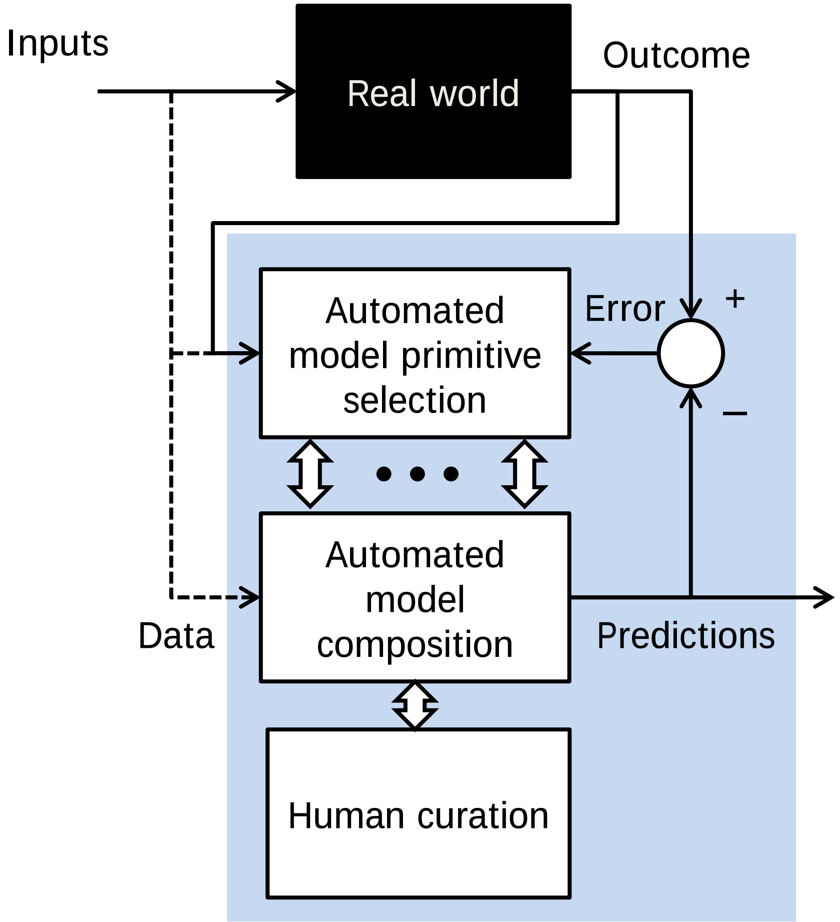
Source youtube: https://www.youtube.com/watch?v=YoOaJclkSZg

D³M: Data-Driven Discovery of Models

Today: Manual



Tomorrow: Automated



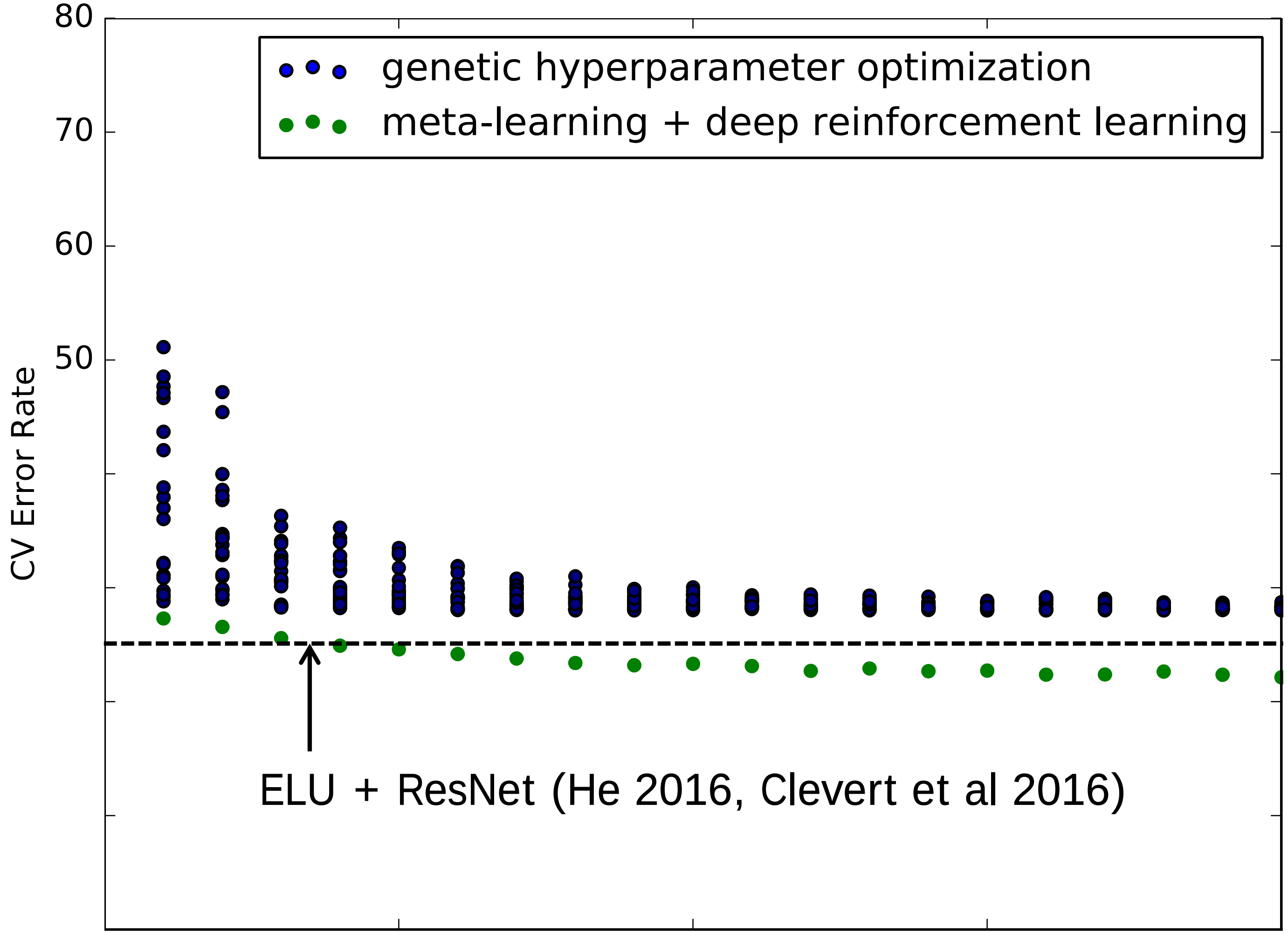
- Model: representation of a real-world system
  - 538 election model
  - NCAR arctic sea ice model
  - N7 IED explosion predictor
- Manual process: 10-1000s of person-years
- Teams of experts required to develop the model

- Automatically select problem-specific model primitives
  - Extend the library of modeling primitives
- Automatically compose complex models from primitives
- Facilitate user interaction with composed models

Machine learning that builds machine learning

Computer vision/object recognition (CIFAR-100)

- Prior state of the art: Google/Microsoft DNN

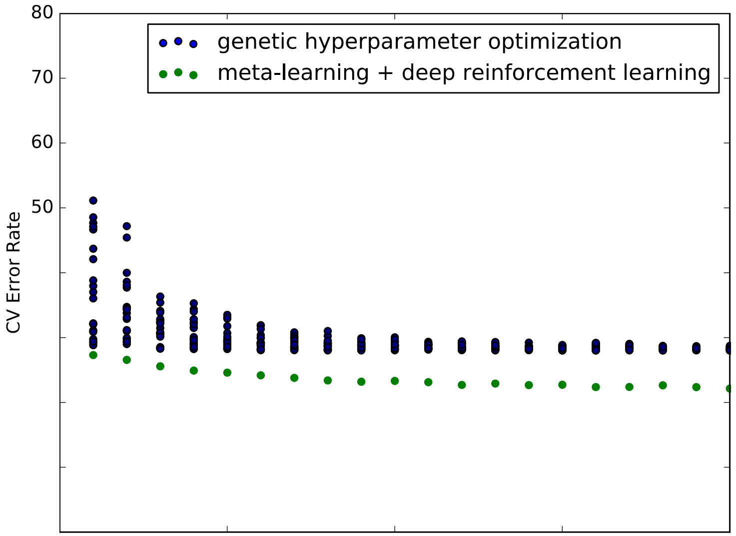


DNN – Deep neural network

The compute problem

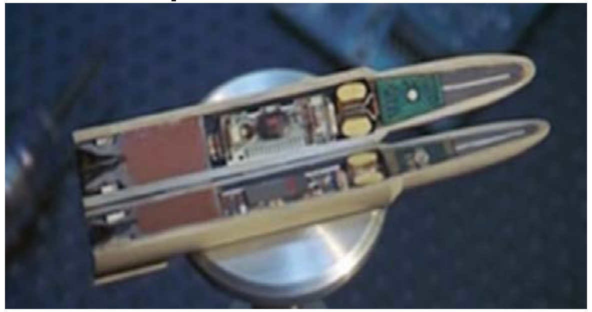
It takes this...

... to protect this



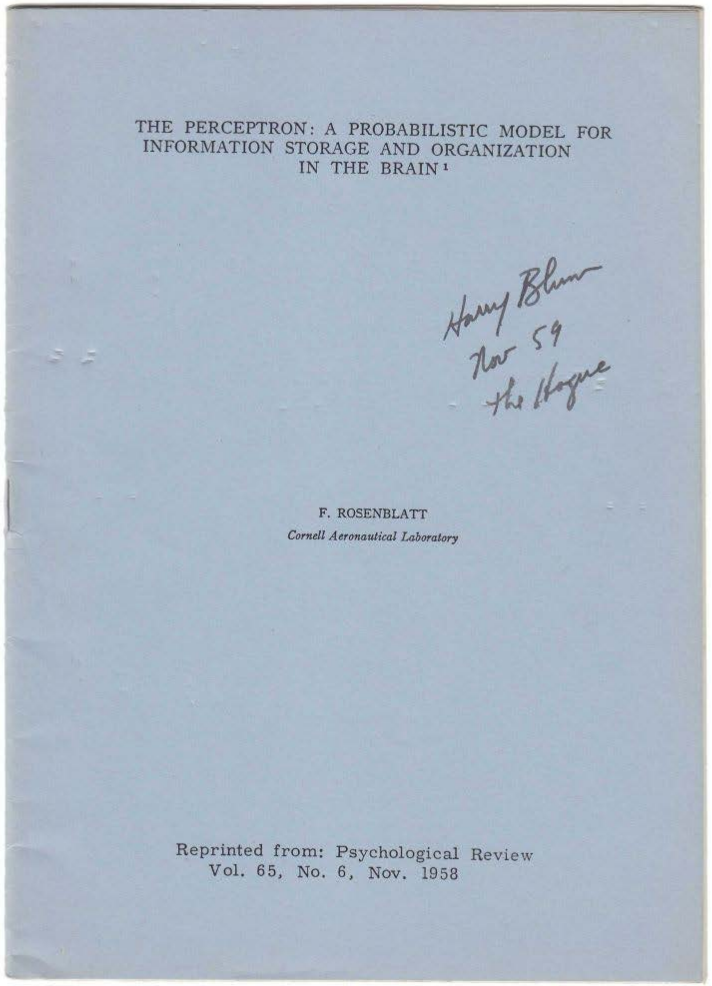
Required 7,000+ compute hours to beat humans

The scope that tracks this...



has 30 minutes of battery life

Case study: deep neural networks

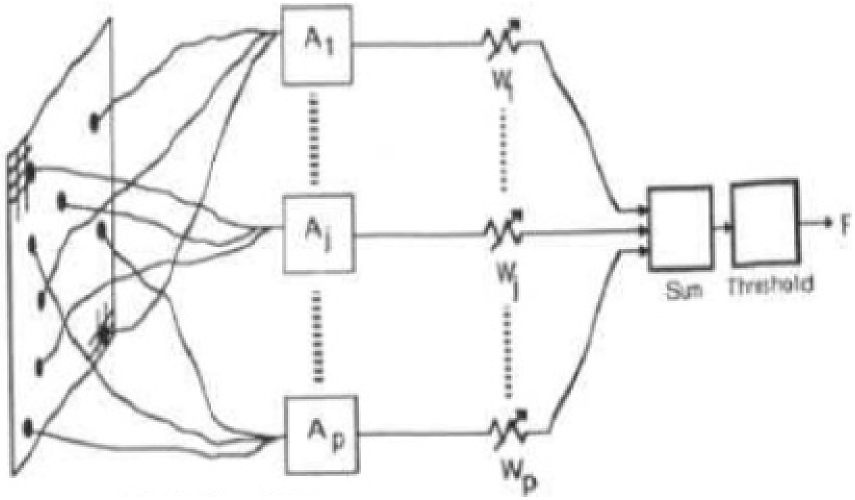


THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN [1]

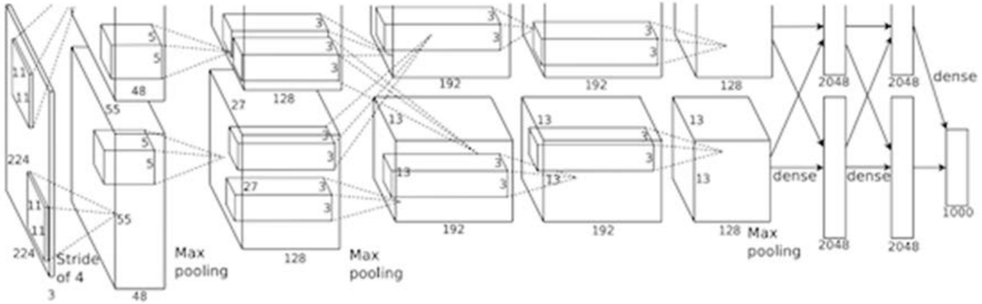
Harry Blum
Nov 59
the Hague

F. ROSENBLATT
Cornell Aeronautical Laboratory

Reprinted from: Psychological Review
Vol. 65, No. 6, Nov. 1958

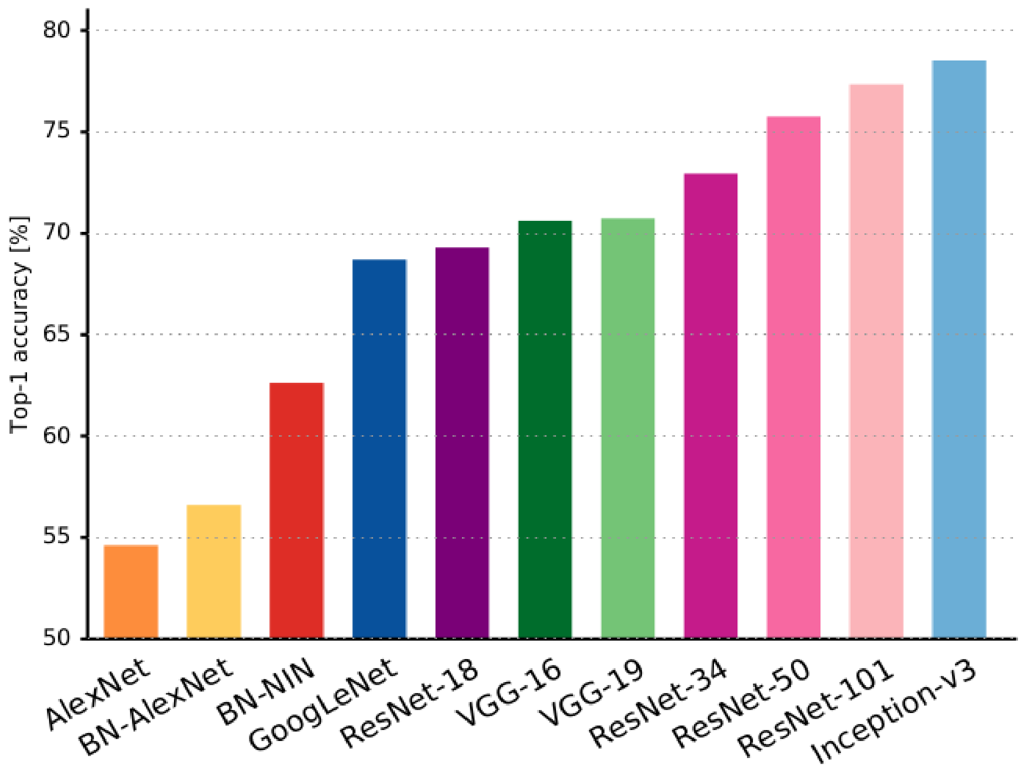


Rosenblatt's digit recognizer, 1958



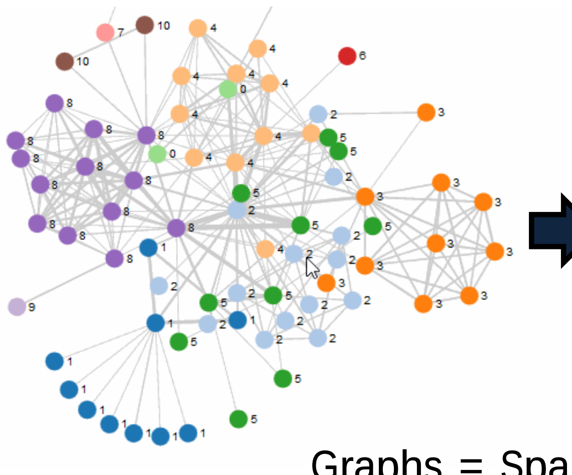
AlexNet, 2012

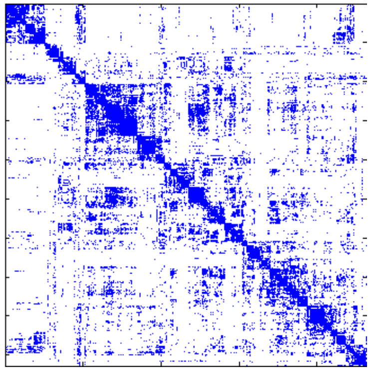
Case study: deep neural networks



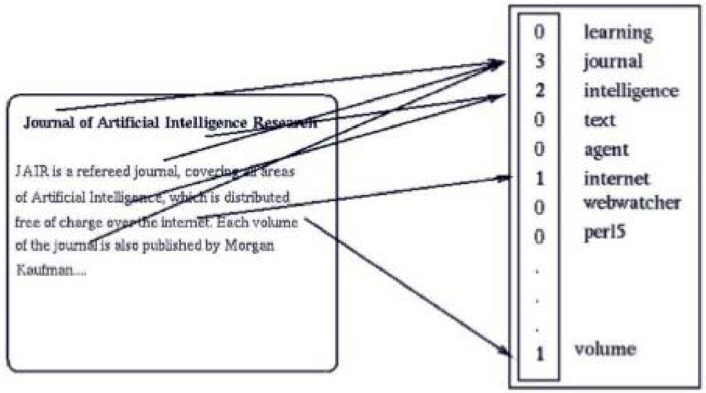
Data are vectors and matrices

Sparse vectors/matrices



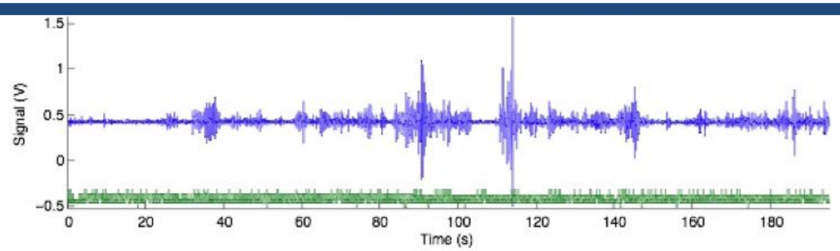


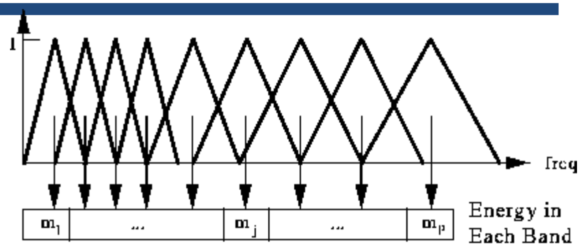
Graphs = Sparse matrices



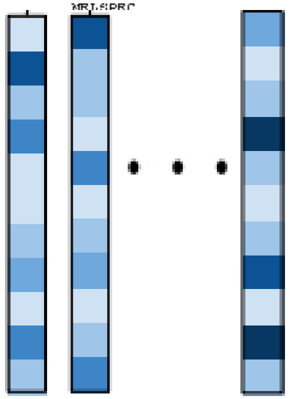
Text and programs -> Sparse vectors

Dense vectors/matrices





Time series = dense cepstrum/spectrum

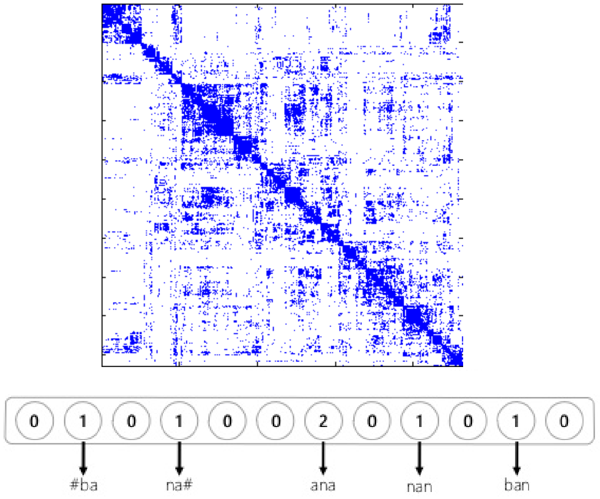


Images = dense vectors

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

Machine learning is projection

Sparse systems



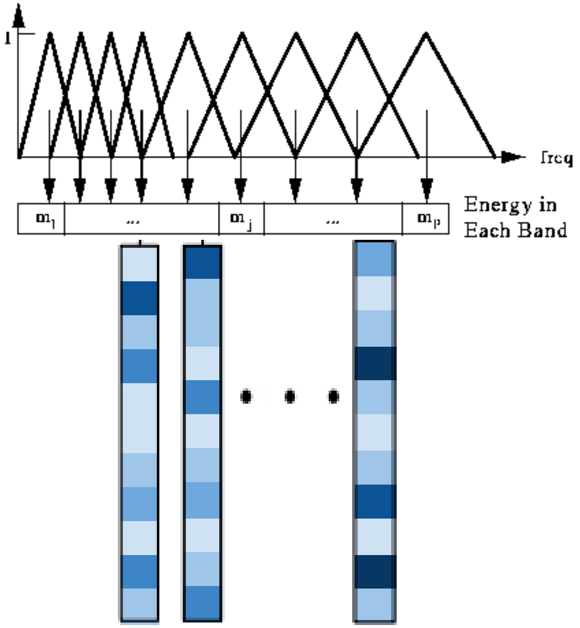
Sparse-to-Sparse

Machine translation
Author ID

Sparse-to-Dense

Graph/word embedding
Vertex classification
PageRank

Dense systems



Dense-to-Dense

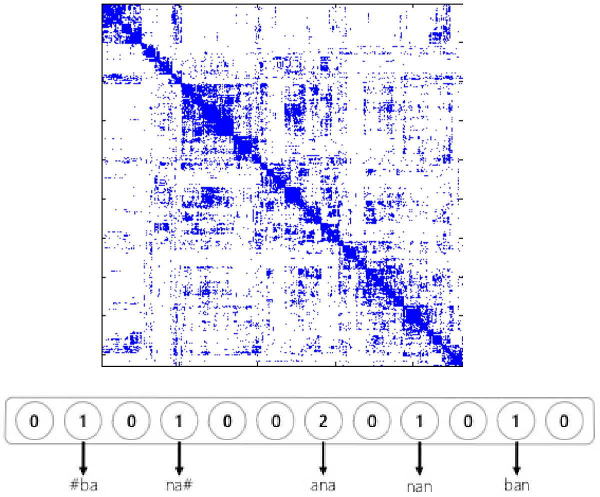
Target tracking
Object detection

Dense-to-Sparse

Image captioning
Speech recognition
Image/audio biometrics

Compute enables machine learning; partially

Sparse systems



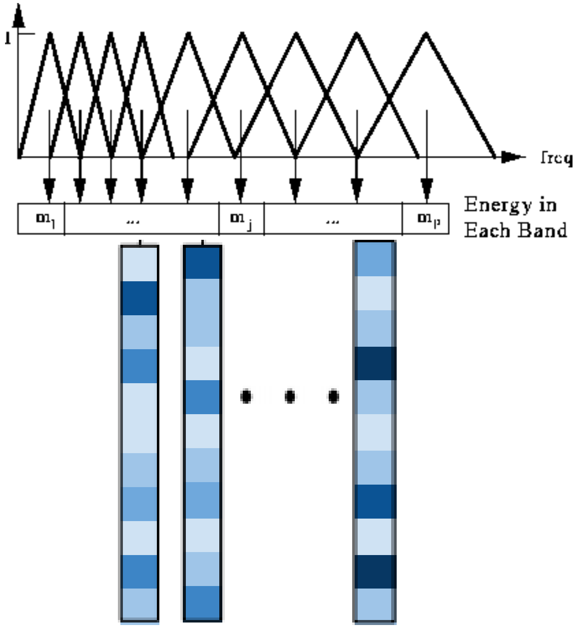
Sparse-to-Sparse
ASICs ~ 10-50x

Machine translation
Author ID

Sparse-to-Dense
GPU ~ 2-10x

Graph/word embedding
Vertex classification
PageRank

Dense systems



Dense-to-Dense
GPU ~ 10-100x
TPU ~ 100-1000x

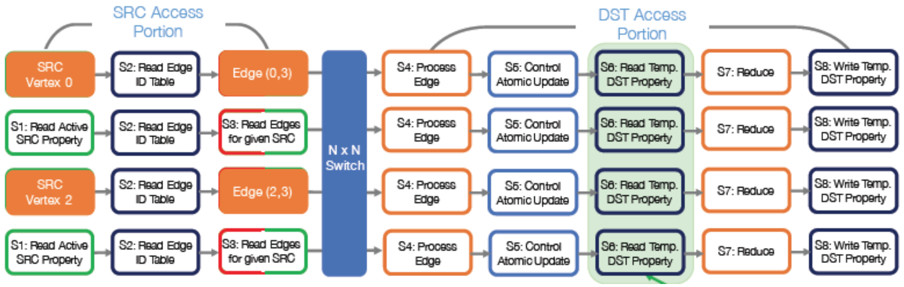
Target tracking
Object detection

Dense-to-Sparse
GPU ~ 2-10x

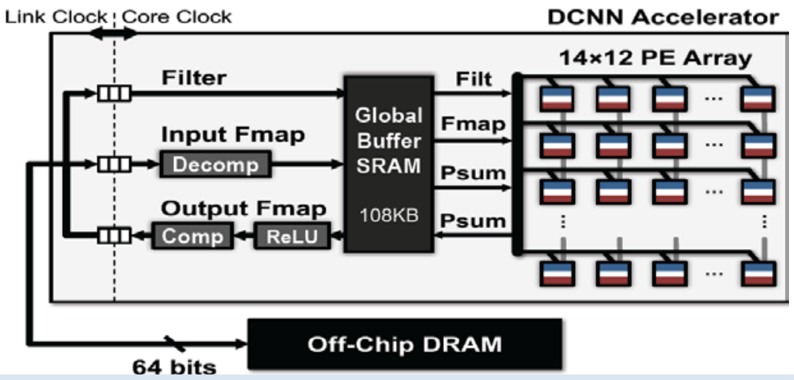
Image captioning
Speech recognition
Image/audio biometrics

Can we have our cake and eat it to?

Hyper-specialization (ASICS)

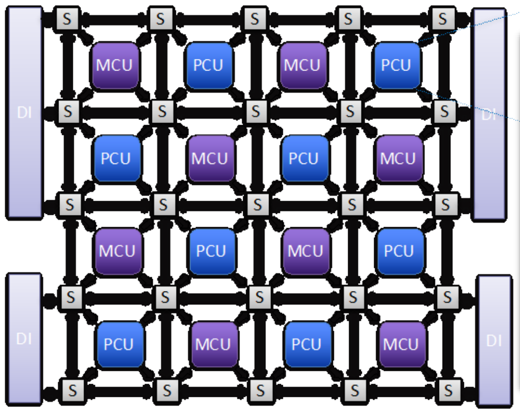


SPARSE - Graphicionado @ 1 GHz in 28nm
- Pipeline for graph analytics data flow
- Multiple pipeline streams with SRC & DST access
- **157K edges/s/mW on BFS**



DENSE - Eyeriss @ 0.6 GHz in 28nm
- Convolution accelerator 168 (MACs) PEs with reconfigurable dataflow
- 182 KB of on-chip SRAM
- **250 images/s/W on AlexNet**

Malleable architectures



SPARSE:
BFS on Twitter
102K edges/s/mW

DENSE:
AlexNet (full app.)
130 images/s/W

Images source – Stanford University

www.darpa.mil

Electronics Resurgence Initiative: Materials and Integration Thrust

Daniel S. Green

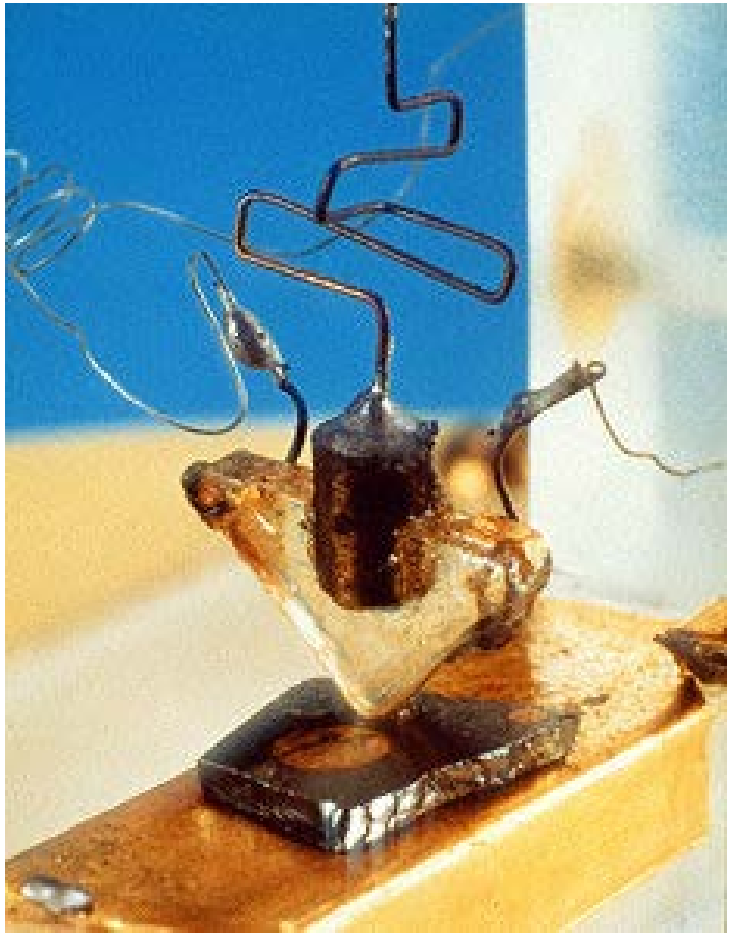
DARPA Program Manager

18 July 2017

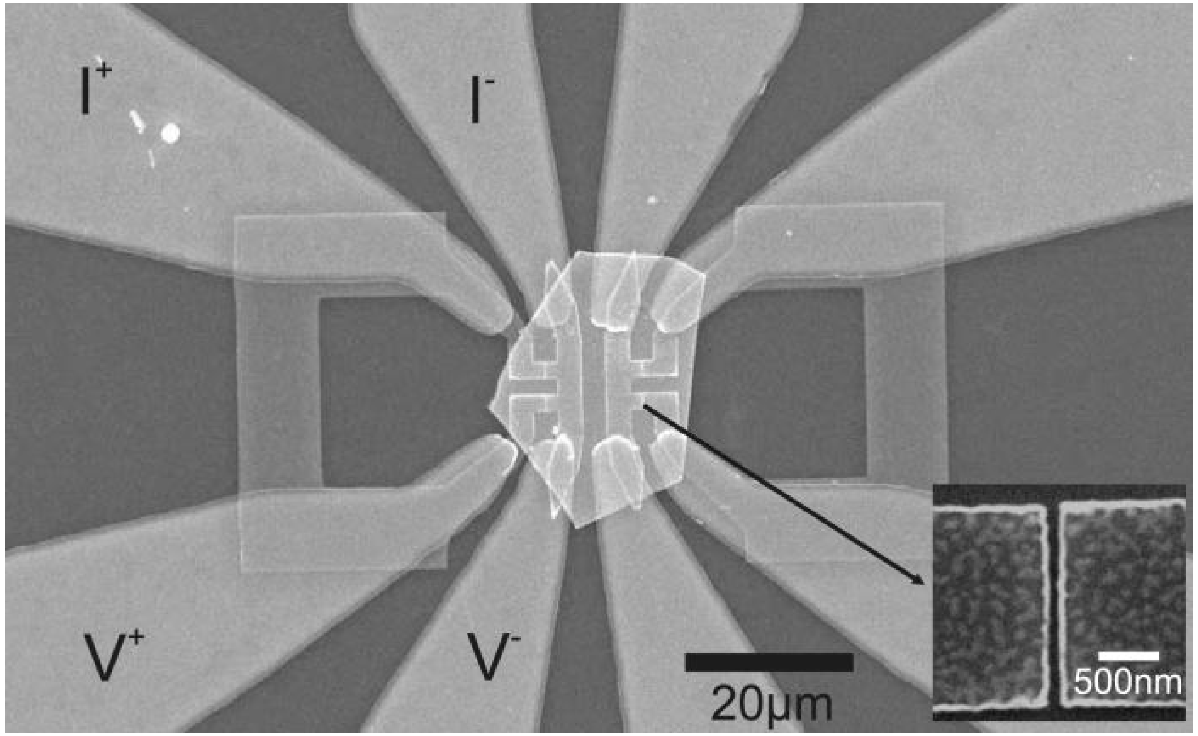
Motivating Materials for Beyond Moore's Law Scaling

A compute problem

Materials have underpinned Moore's Law from the start



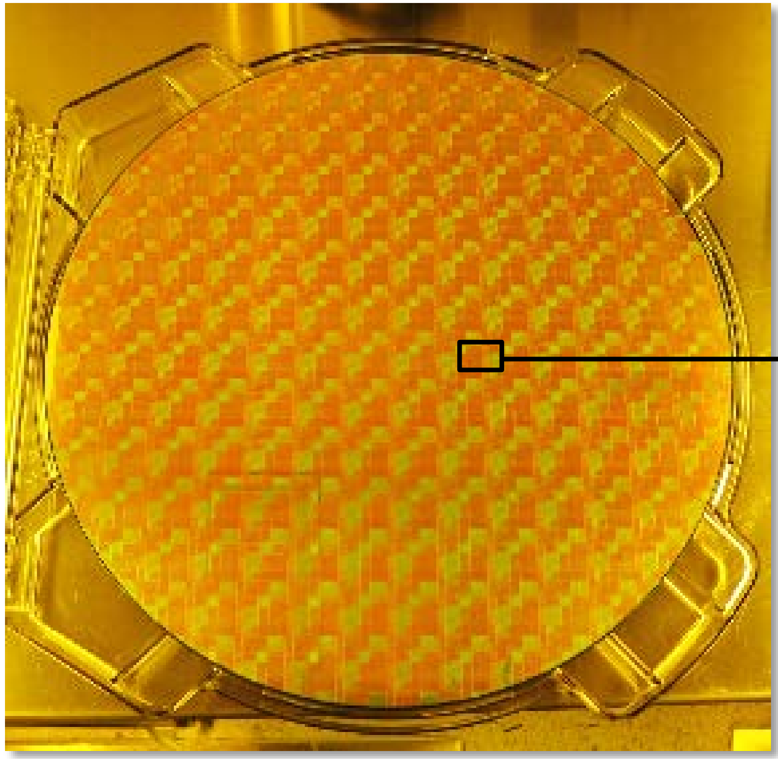
What is a transistor: The World of Modern Electrons; Sam Sattel



Applied Physics: Feb 2012; Experimental realization of superconducting quantum interference devices with topological insulator junctions. M. Veldhorst et. al.

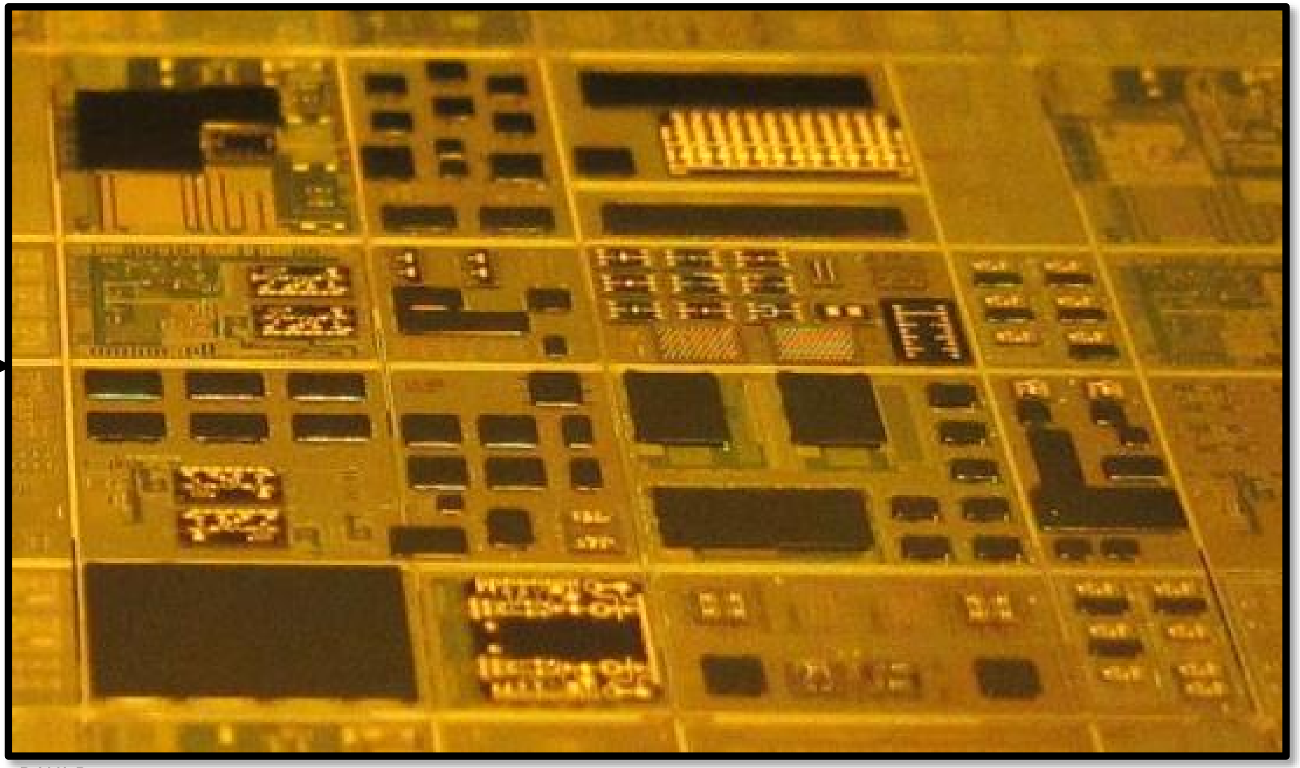
...and continue to present opportunities

At the same time, heterogeneous integration has advanced



DAHI Program

300mm diameter Si CMOS wafer



DAHI Program

Si (45nm), InP (TF5 HBT), GaN (GaN20 HEMT)

...and allowed a faster, flexible mix of materials

The materials thrust aims to advance and combine these pieces together

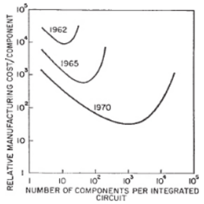
Integration

Panel 1: Accelerating Materials Discovery

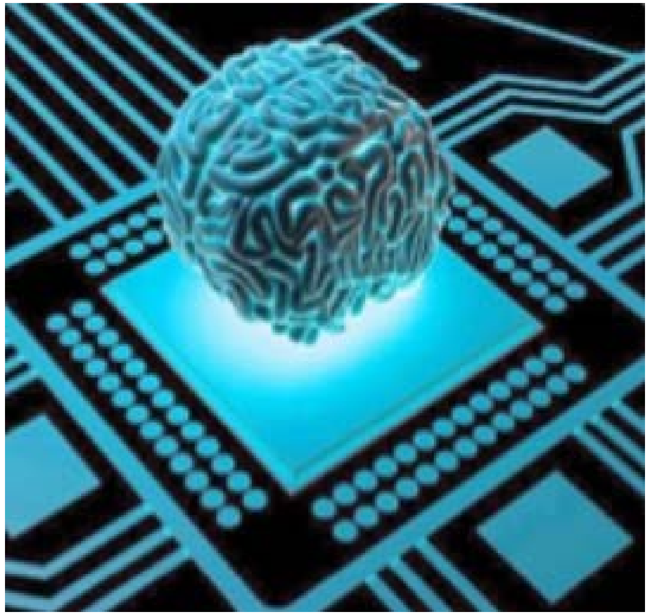
Panel 2: Emerging Materials and Devices

Panel 3: Integrated Processes

Fundamental Science



Beyond Moore's Law



Shutterstock.com

Changing the fundamental compute building blocks allows us to question:

Where and how should we do our thinking?

Accelerating Materials Discovery

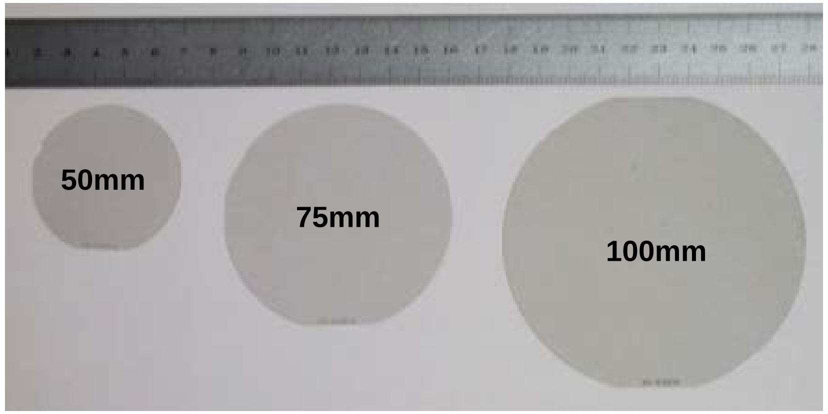
A brief materials story

DARPA Grant: Metalorganic Chemical Vapour dDeposition (MOCVD) Growth Process
~1989

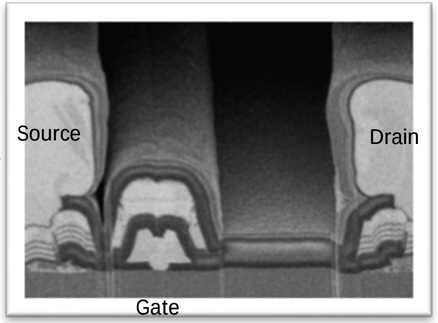
Commercial Material: Gallium Nitride (GaN)
1990s

Blue LED

DARPA Wide Bandgap Semiconductor – Radio Frequency (WBGS-RF) Program
2000s



GaN RF Device

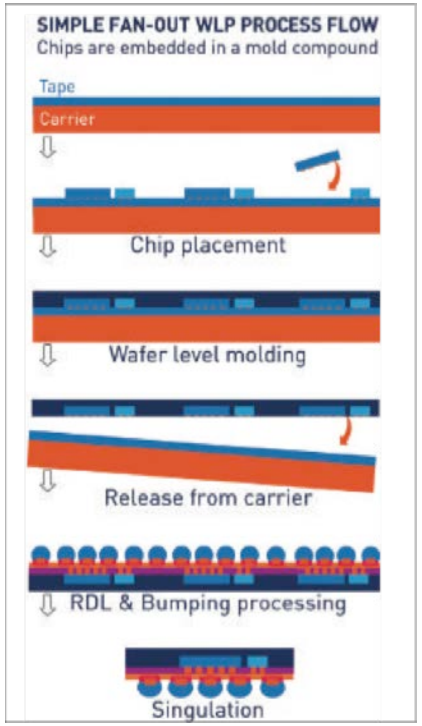


Commercial: Base station technology
2010s

Defense: Radar and Communications
2010s

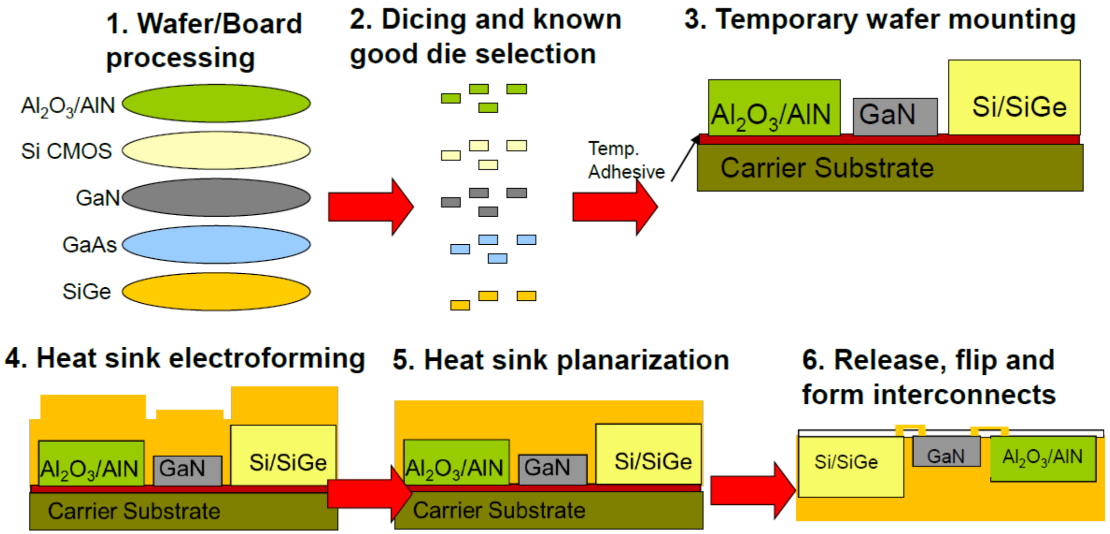
Sources: DARPA, HRL, Solid State Technology

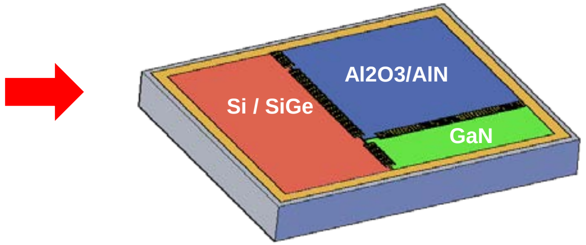
Processes for integration are also possible



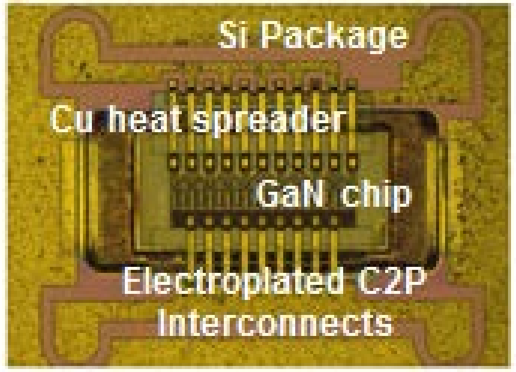
MECA is analogous to wafer-level fan-out packaging technology.

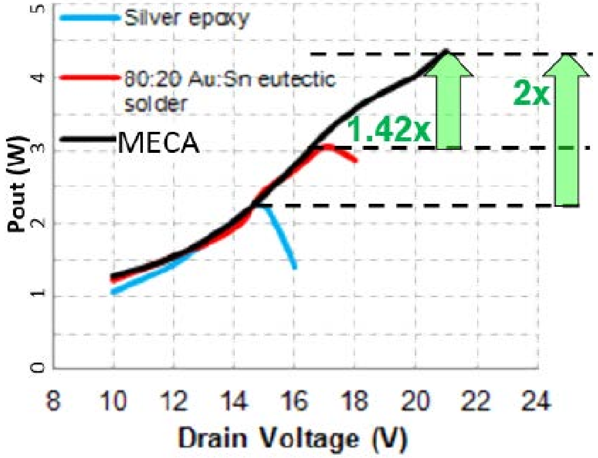
Metal Embedded Chip Assembly (MECA)





MECA-integrated heterogeneous module





Sources: HRL, Solid State Technology

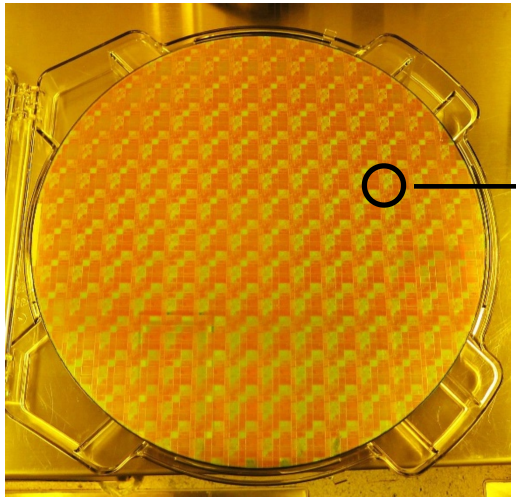
What's different here?

- Focus on enabling Beyond Moore's Law Scaling
  - Not an RF component initiative
  - Not a Moore's Law Scaling Initiative
- Big Question:
  - Can we develop processes to integrate (and identify) new materials quickly?

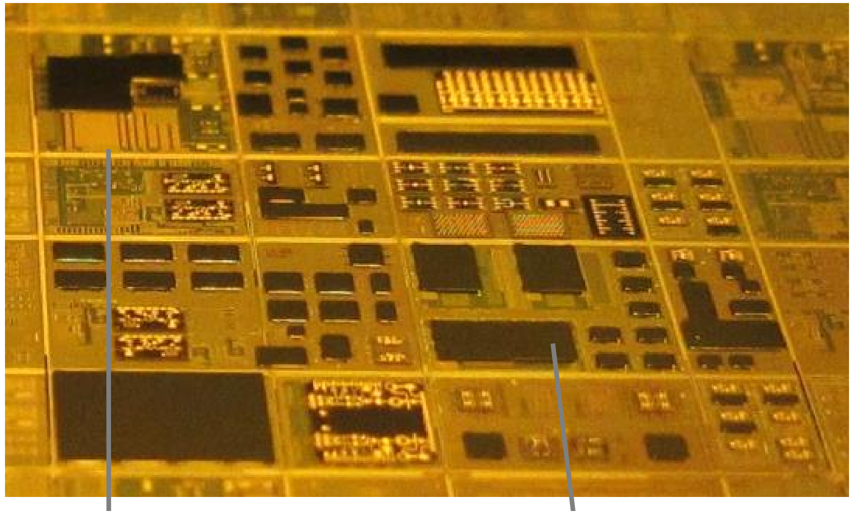
Accelerating Materials Discovery Panel		
	Stephen Bedell	IBM T. J. Watson Research Center
	Joy Watanabe	Intermolecular, Inc.
	Michael Kozicki	Arizona State
	Subu Iyer	UCLA
	Joseph Geddes	Photia Incorporated

Emerging Materials and Devices

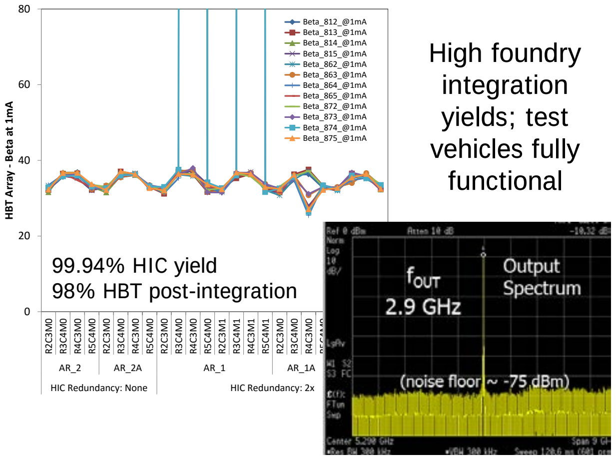
Device opportunity: informed by integration



300mm diameter Si CMOS wafer (45nm node)

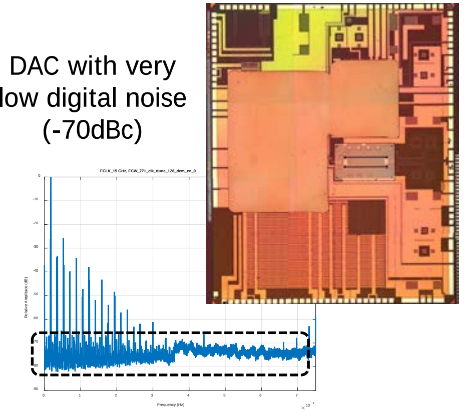


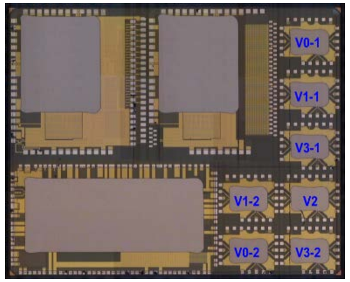
DAHI integration (Dec 2015): Si (45nm), InP (TF5 HBT), GaN (GaN20 HEMT)



High foundry integration yields; test vehicles fully functional

DAC with very low digital noise (-70dBc)

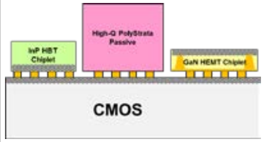
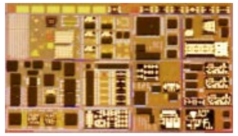
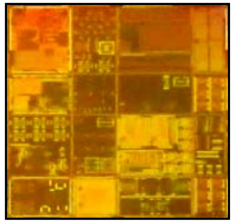
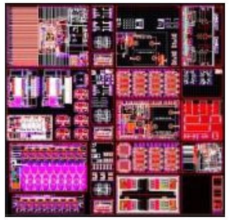
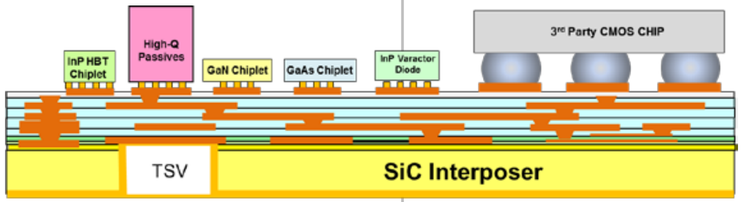




Successful testing identified optimal S/H circuit for ADC (>65dB SFDR @ 2GHz)

Sources: DARPA, Northrop Grumman

DAHI simplicity enables rapid evolution

Technology	MPW0	MPW1	MPW2	MPW3	Future MPWs
CMOS	IBM 65nm	GF 45 nm	GF 45 nm	GF 45 nm	GF 45 nm
InP HBT	TF4 (2 metals)	TF4 (3 metals)	TF4 (4 metals)	TF4 (4 metals)	TF4 (4 metals)
		TF5 (3 metals)	TF5 (4 metals)	TF5 (4 metals)	TF5 (4 metals)
InP Varactor Diode					AD1
GaN HEMT	GaN20	GaN20	GaN20	GaN20	GaN20
	T3 (HRL)	T3 (HRL)	T3 (HRL)	T3 (HRL)	T3 (HRL)
GaAs HEMT				P3K6	P3K6
Passive Components		PolyStrata (Nuvotronics)	PolyStrata (Nuvotronics)	PolyStrata (Nuvotronics)	PolyStrata (Nuvotronics)
Base Substrate	CMOS	CMOS	CMOS	CMOS	CMOS
				SiC Interposer (IWP5)	SiC Interposer (IWP5)
			In test 	In fab 	

Sources: DARPA, Northrop Grumman

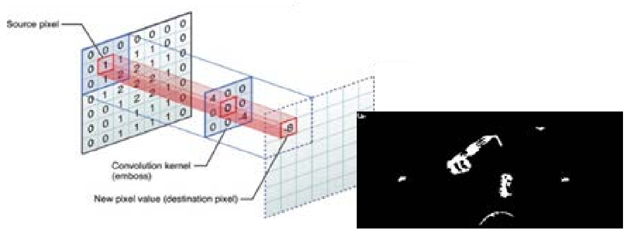
Builds on our initial efforts – with demonstrated UPSIDE

Unconventional Processing of Signals for Intelligent Data Exploitation (UPSIDE) Program

Video surveillance collection and analysis significantly exceed current embedded computing capability

Today: Digital Signal Processing

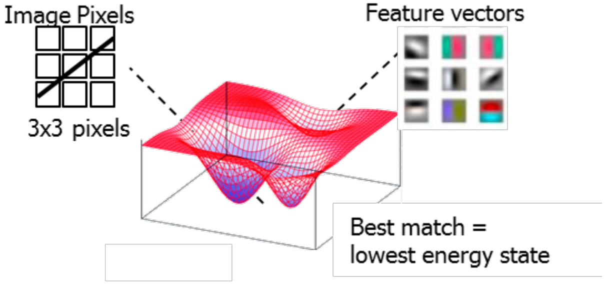
- Current approaches require compute-intensive, exact, sequential operations over all pixels to detect features, objects and tracks.
- Large images require Tera-Ops/sec



Detected Salient Pixels

Unconventional Analog Processing

- UPSIDE replaces compute-intensive exact Boolean operations with probabilistic, best match for significant power efficiency



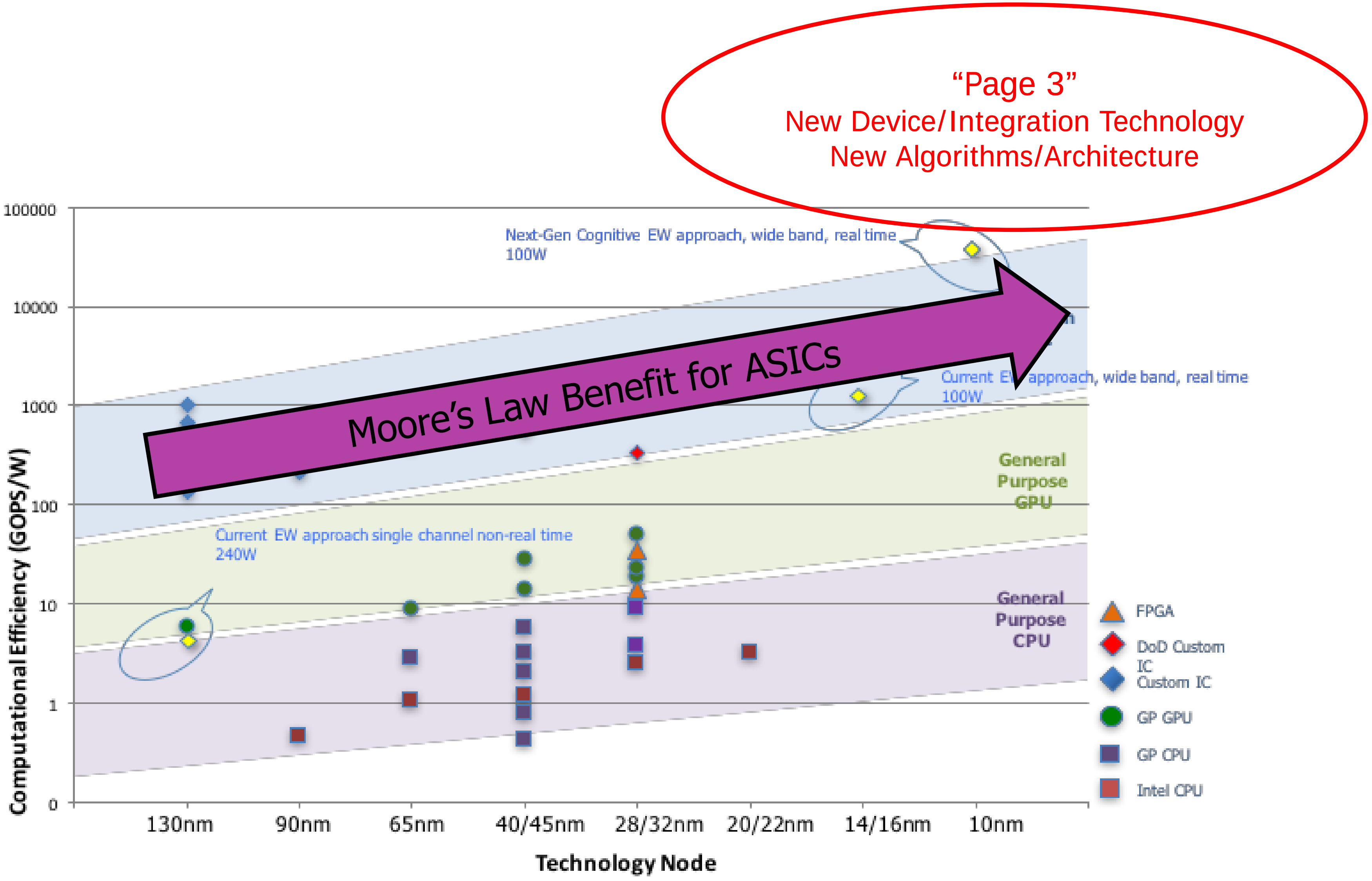
What's different here?

- Focus on enabling Beyond Moore's Law Scaling
  - Not a conventional logic / memory device initiative
- Big Question:
  - What are the NEW materials or devices (and their functions) that should added to the toolbox?

Emerging Materials and Devices Panel		
	Jian-Ping Wang	University of Minnesota
	Sayeef Salahuddin	UC Berkeley / EECS
	Arjit Raychowdhury	Georgia Tech
	Vladimir Stojanovic	University of California, Berkeley
	Noah Sturcken	Ferric, Inc.

Integrated Processes

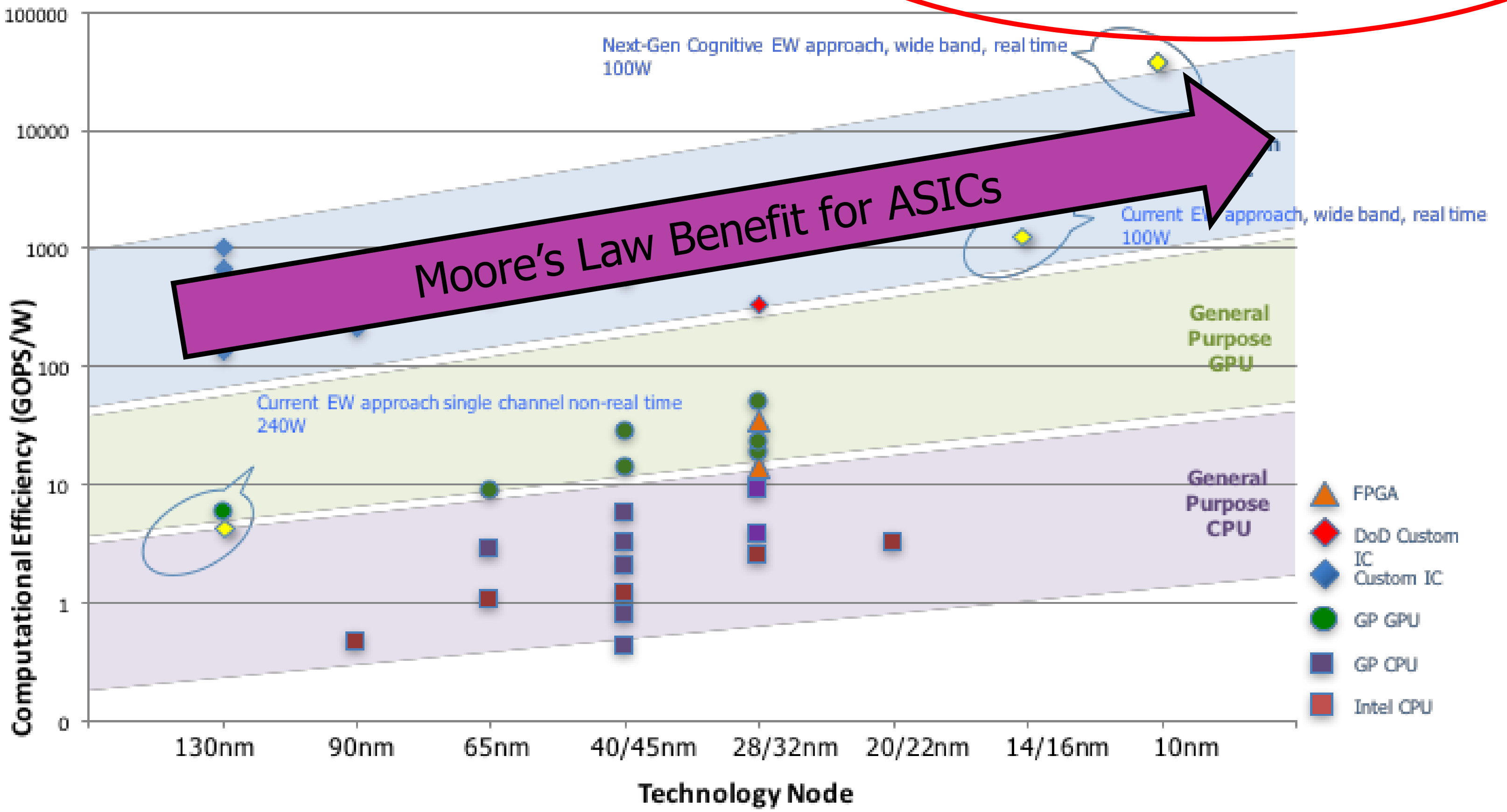
“Page 3” materials and integration



FPGA – Field Programmable Gate Array
GP - General Purpose
GPU – Graphics Processing Unit
CPU – Central Processing Unit
GOPS- Giga Operations per Second

“Page 3” materials and integration

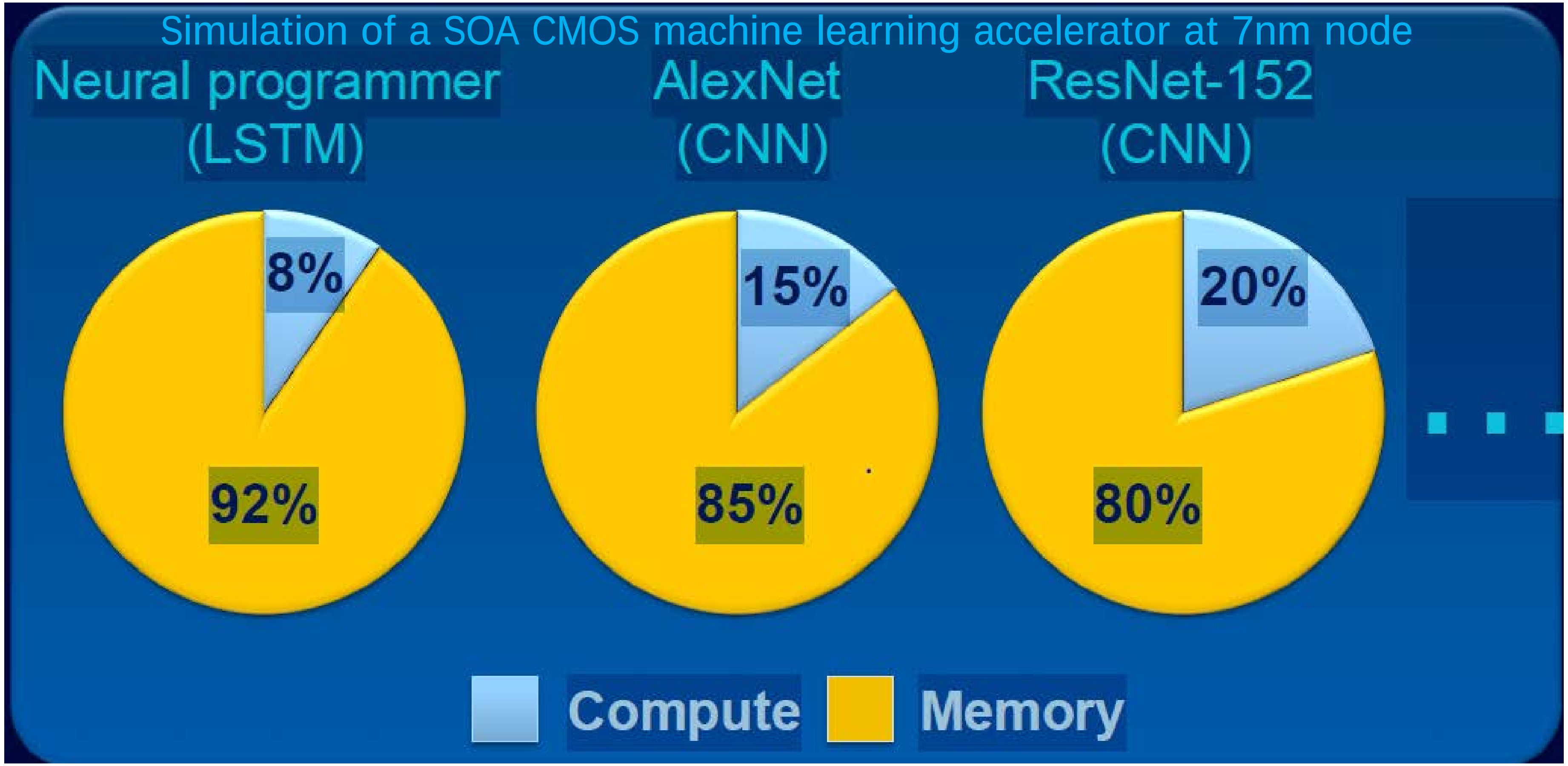
“Page 3”
New Device/Integration Technology
New Algorithms/Architecture



FPGA – Field Programmable Gate Array
GP - General Purpose
GPU – Graphics Processing Unit
CPU – Central Processing Unit
GOPS- Giga Operations per Second

The problem for many applications: SoC performance is driven by data transfer time

- Most of the problem is memory bandwidth and latency
- Even 2D CMOS ML accelerators aren't addressing the memory problem

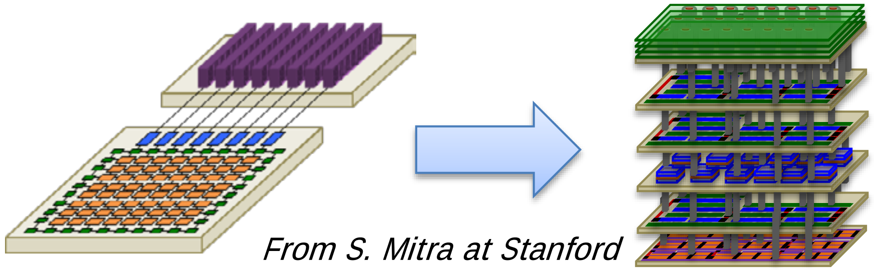


Simulation data from S. Mitra at Stanford

CMOS – Complimentary Metal Oxide Semiconductor
SoC – System on Chip
SOA – State of the Art
LSTM – Long Short Term Memory
CNN – Convolutional Neural Network

Potential solutions

“Bring memory in the compute”
Monolithic 3D SoC



From S. Mitra at Stanford

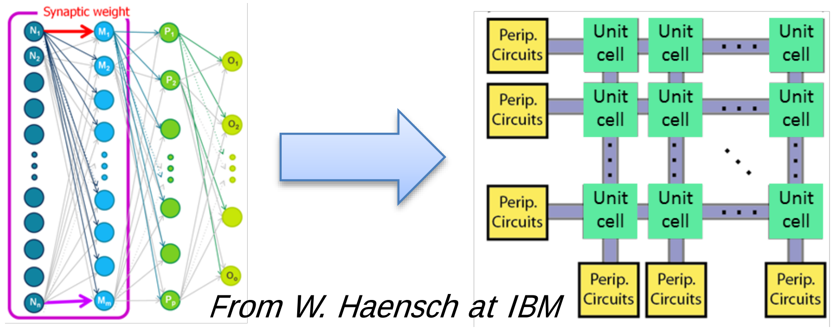
Initial simulations

- Up to 1000X improvement in Energy*time for memory-intensive applications at a common node
- Up to 100X improvement in Energy*time when comparing 3D SoC @ 90nm with 2D at 7nm
- Less cost per area than 2D 14nm fabrication with up to 4GB of on-chip memory storage

Critical needs

- Low temperature logic device fabrication (< 450C)
- Low temperature, dense NVM cell fabrication (< 450C)

“Bring the compute in memory”
DNN Dot Product calculation



From W. Haensch at IBM

Initial simulations

- Initial simulation shows strong improvement to Energy*time for DNN core computation

Critical needs

- Full system simulations
- Optimal memory unit cell

What's different here?

- Focus on enabling Beyond Moore's Law Scaling
  - Not just the 3DIC challenge with conventional architectures
  - Seek to overcome the memory bottleneck
- Big Question:
  - Can we use integrated process to realize new architectures unavailable today?

Integrated Processes Panel		
	Max Shulaker	**MIT**
	Bruce Taol	**Micron**
	Wilfried Haensch	**IBM T. J. Watson Research Center**
	Qiangfei Xia	**UMass Amherst**
	Zvi Or-Bach	**MonolithIC 3D, Inc.**

www.darpa.mil

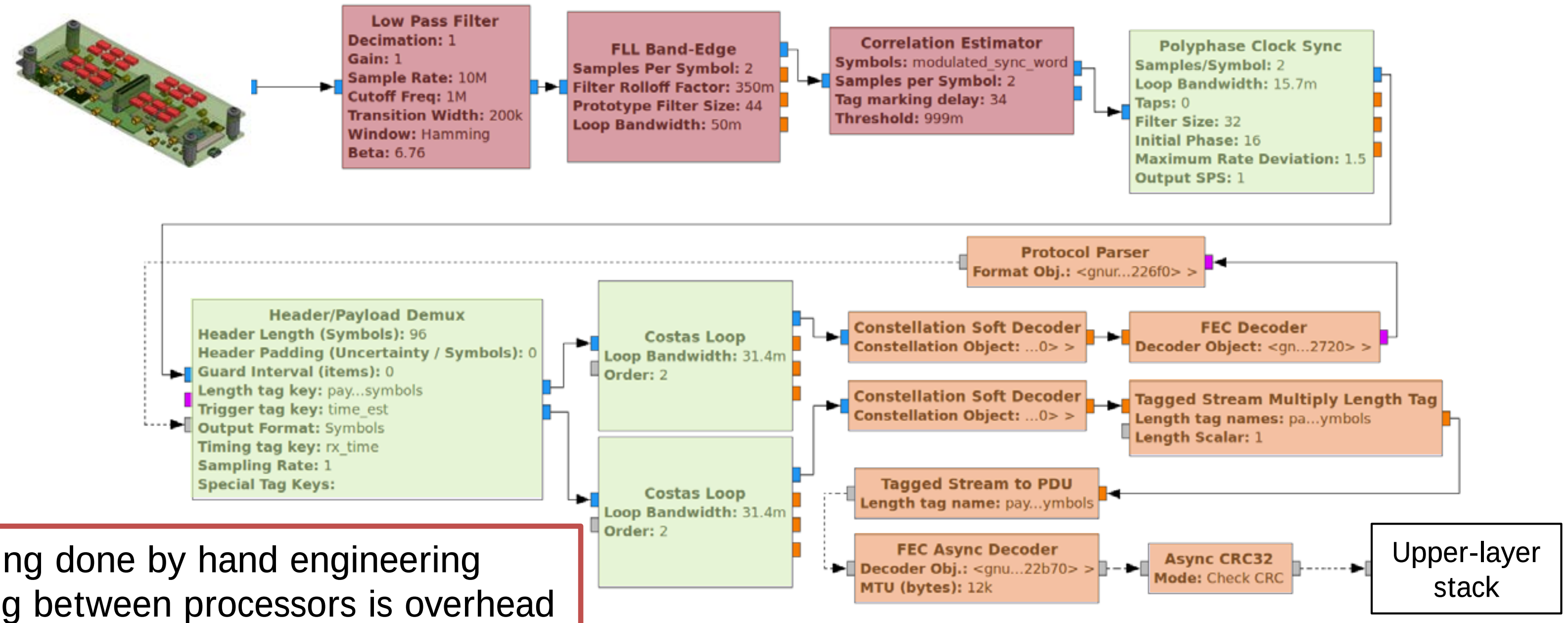
Electronics Resurgence Initiative: Architectures

Tom Rondeau
DARPA Program Manager

18 July 2017

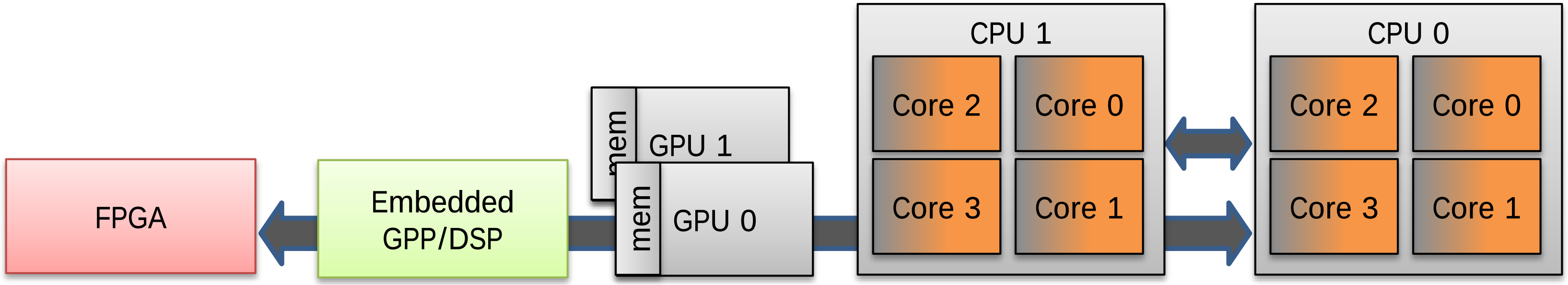
Previous project lead for GNU Radio

Streaming data across multiple processing elements



Mapping done by hand engineering
Moving between processors is overhead

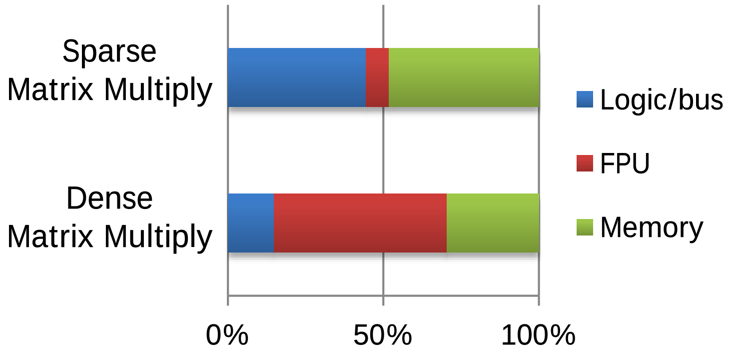
FPGA – Field Programmable Gate Array
GPU – Graphics Processing Unit
CPU – Central Processing Unit
DSP – Digital Signal Processor
GPP – General Purpose Processor



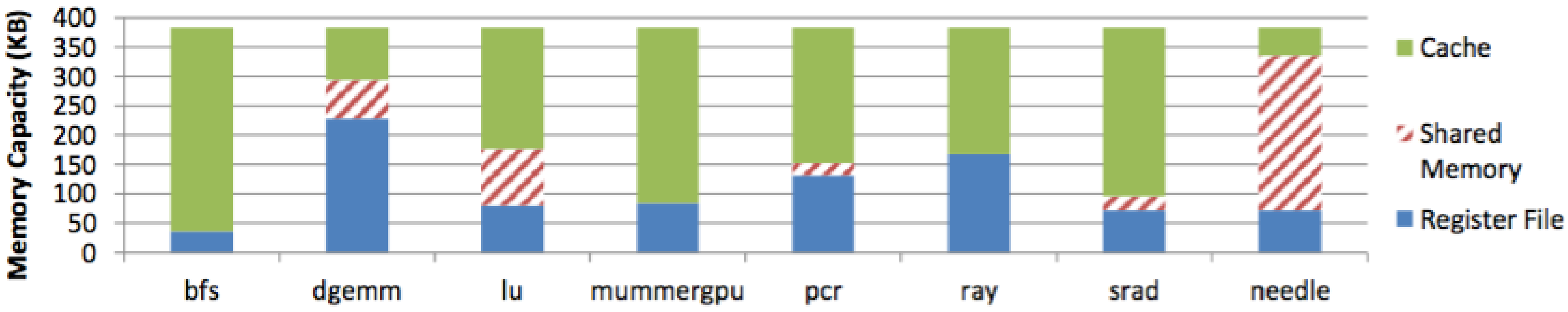
Can we automatically map algorithms to processors?
Can we afford to move data back and forth?

Computing linear algebra is a hard problem

- Processor design trades
  - Math/logic resources
  - Memory (cache vs. register vs. shared)
  - Address computation
  - Data access and flow
- Processor choice depends on:
  - Memory requirements
    (small vs. large) x (random vs. linear)
  - Computation requirements

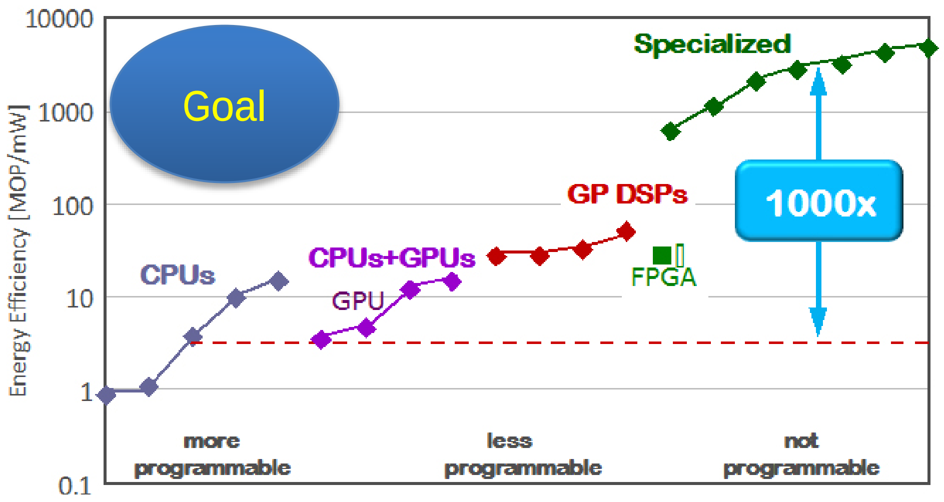


The problem: Can we find optimal hardware configuration across algorithms?



No one hardware solves all problems well

Managing specialization & flexibility

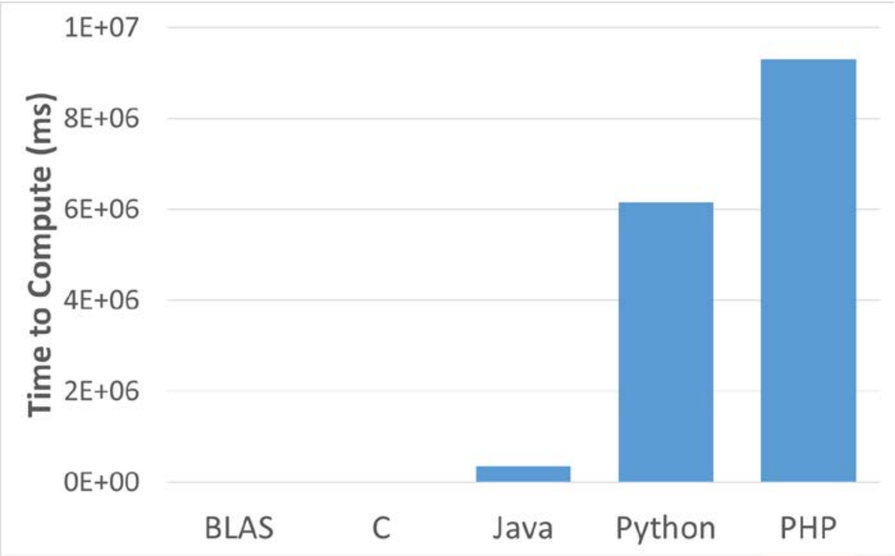


Specialization

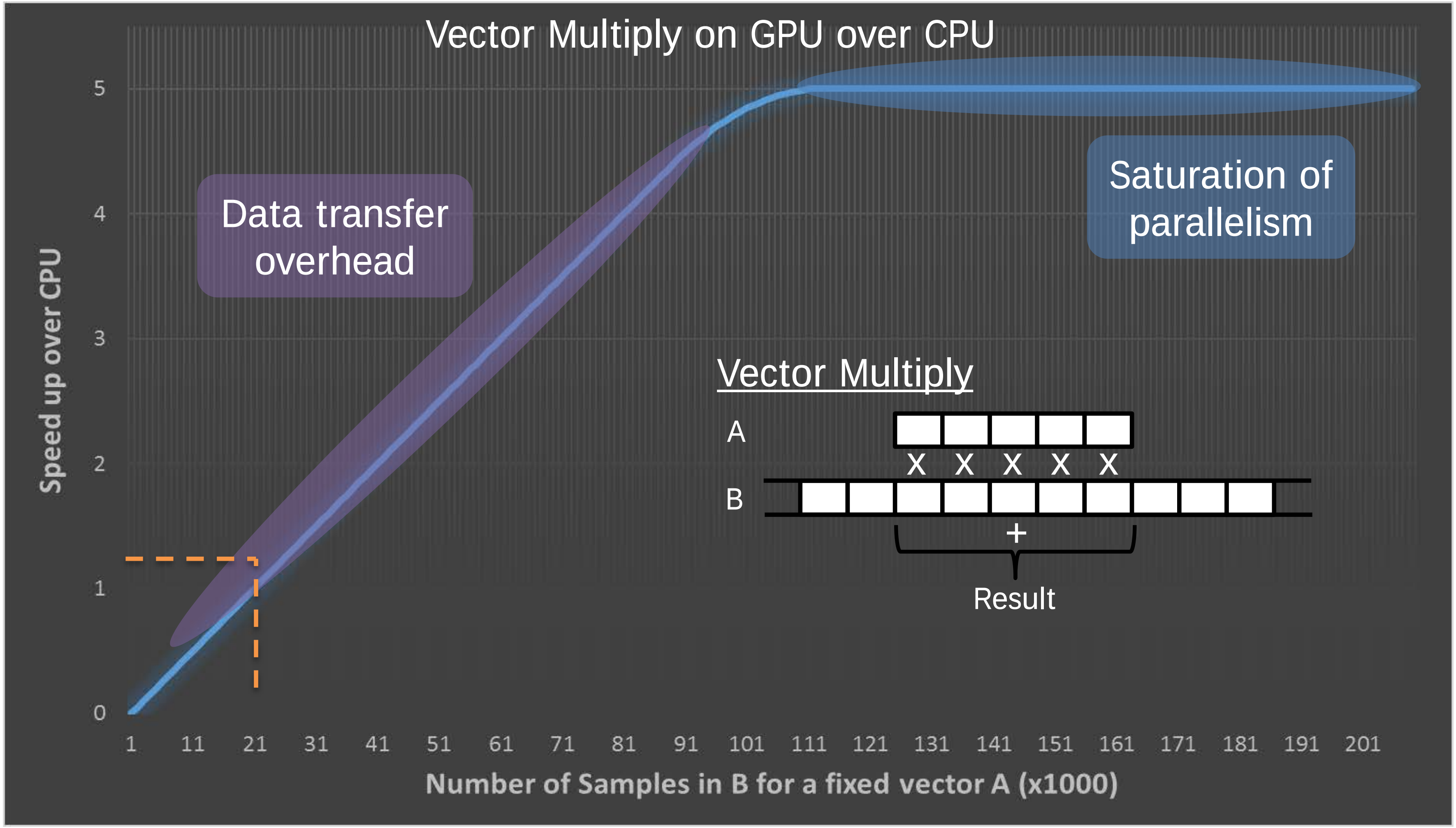
- Performance has come at the cost of usability
- Difficulty in programming and system integration

Flexibility

- Productivity has come at the cost of compute efficiency
- Abstraction tends to ignore the underlying hardware



It’s not just the processor



- GPUs do better at computing convolutions (dense matrix multiplies)
- Cost of data transfer means sometimes the CPU is more efficient
- Resource optimization for multiple applications

System integration requires full-stack programming

Today's model

Single Processor: Significant prior work

- High-level languages, compilers, libraries, tools

System of Processors: Basic tools but significant difficulties

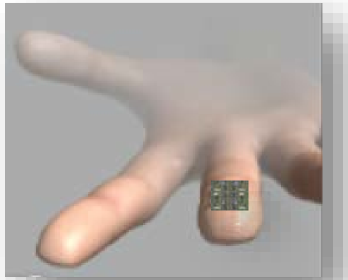
- Middleware, busses/networking, data management

Opportunities

- Full understanding of the processing elements
- Performance monitoring and online updates
- Managing data movement (memory, I/O)
- Better representations of the problems
- Faster time to integration

Building a development ecosystem

Build new compute engines and processors that solve the significant computing needs of today's and tomorrow's applications.



But a chip that can't be used, integrated, and programmed is called sand

This list of processors suggests that solutions exist. So why are we here?

Parallel Processors

Adapteva
Analog Devices-
BlackFin
Altair
Altera
Ambric
AMD-APU
ARM-MP/Neon
ARM-Mali
Asocs
Aspex
AxisSemi
BOPS
Boston Circuits
Brightscale
Calxeda
Cavium
CEVA
Chameleon
Clearspeed
Cognimem

Cognivue
Cognovo
Coherent Logix
CoreSonic
CPUTech
Cradle
Cswitch
DesignArt
ElementCXI
EZChip
Freescale
Greenarrays
HP
IBM-Cell
IBM-Cyclopse
Icera-PowerVR
Imagination-PowerVR
Imec
Inmos-Transputer
Intel-TFLOPS
Intel-Larrabee

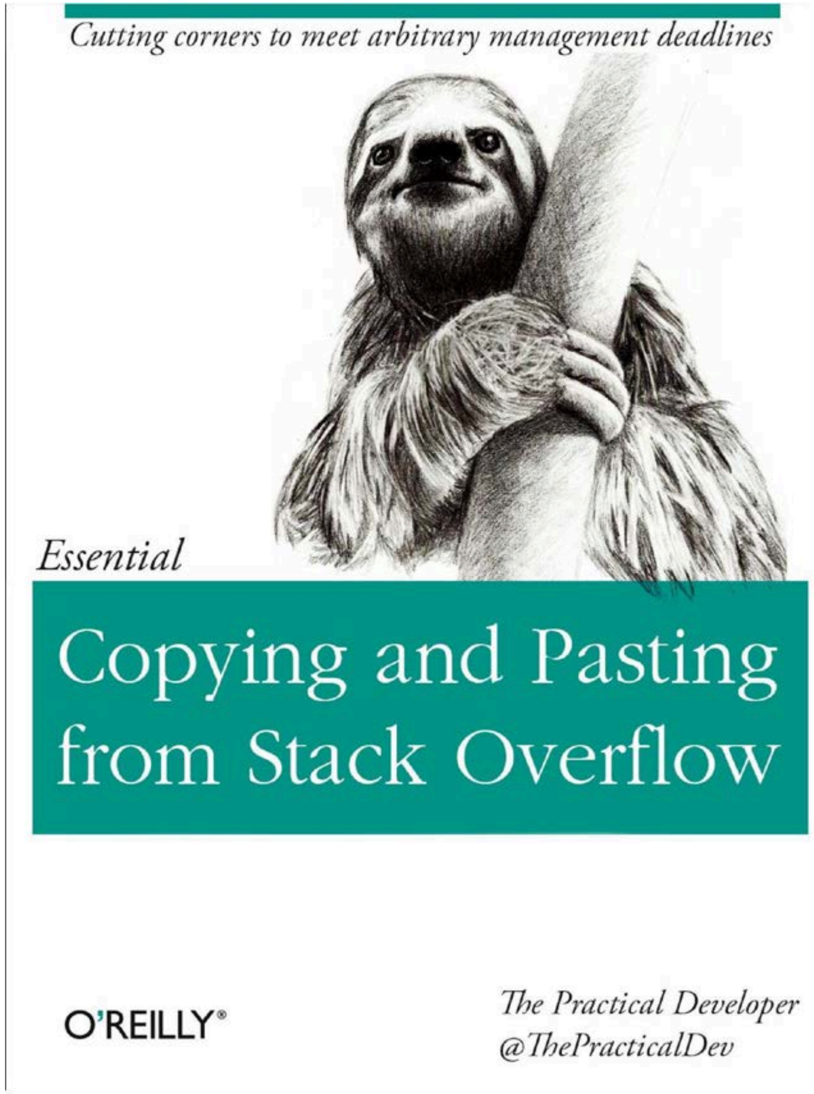
Intel-MIC
Intellasys
Intrinsity
IPFLex
Kalray
Mathstar
MobileEye
ModemArt
Morphics
Morpho
Movidius
NEC
Netlogic
Netronome
Nvidia
Octasic
PACT
Paneve
Picochip
Plurality
Quicksilver

Rapport
Raytheon-Monarch
Recore
Sandbridge
SiByte
SiCortex
Silicon Hive
Silicon Spice
Singular Computing
Sound Design
SpiralGateway
Stream Processors
Stretch
Tabula
Thinking Machines
TI
Tilera
TOPS
Venray
Xelerated
Xilinx

XMOS
Ziilabs

http://www.adapteva.com/andreas-blog/the-siren-song-of-parallel-computing/

Benefits of a rich development ecosystems



https://www.gitbook.com/book/tra38/essential-copying-and-pasting-from-stack-overflow/details

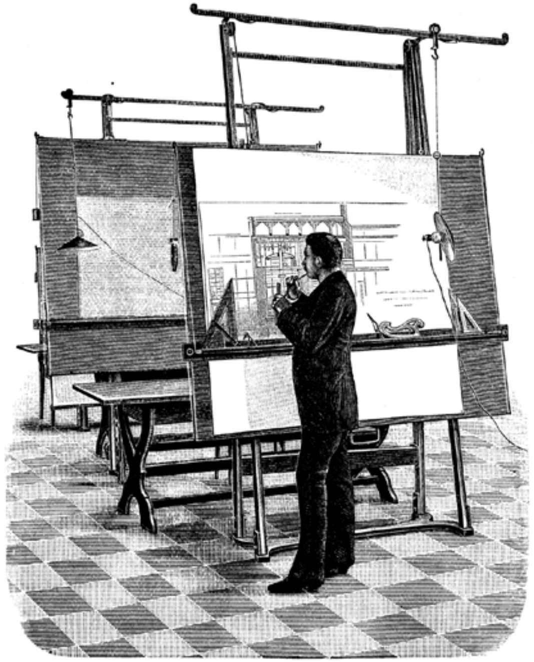
Beyond scaling architectures

<u>Managing specialization & flexibility</u>

- Are flexibility and specialization inherently opposite?
  - Eat your cake and have it, too
- New approaches to processor/SoC designs that change how we specialize?
  - Potential new accelerators and flexible processors that change to meet data needs?

<u>Building a Development Ecosystem</u>

- How do we understand processing needs/capabilities?
  - Cataloged by the math (e.g., dense vs. sparse)?
- Are there better tools to manage the system of processors?
  - Intelligent agents, smart compilers, others?



https://en.wikipedia.org/wiki/Architect

https://commons.wikimedia.org/wiki/File:Chocolate_Fondant.jpg

www.darpa.mil