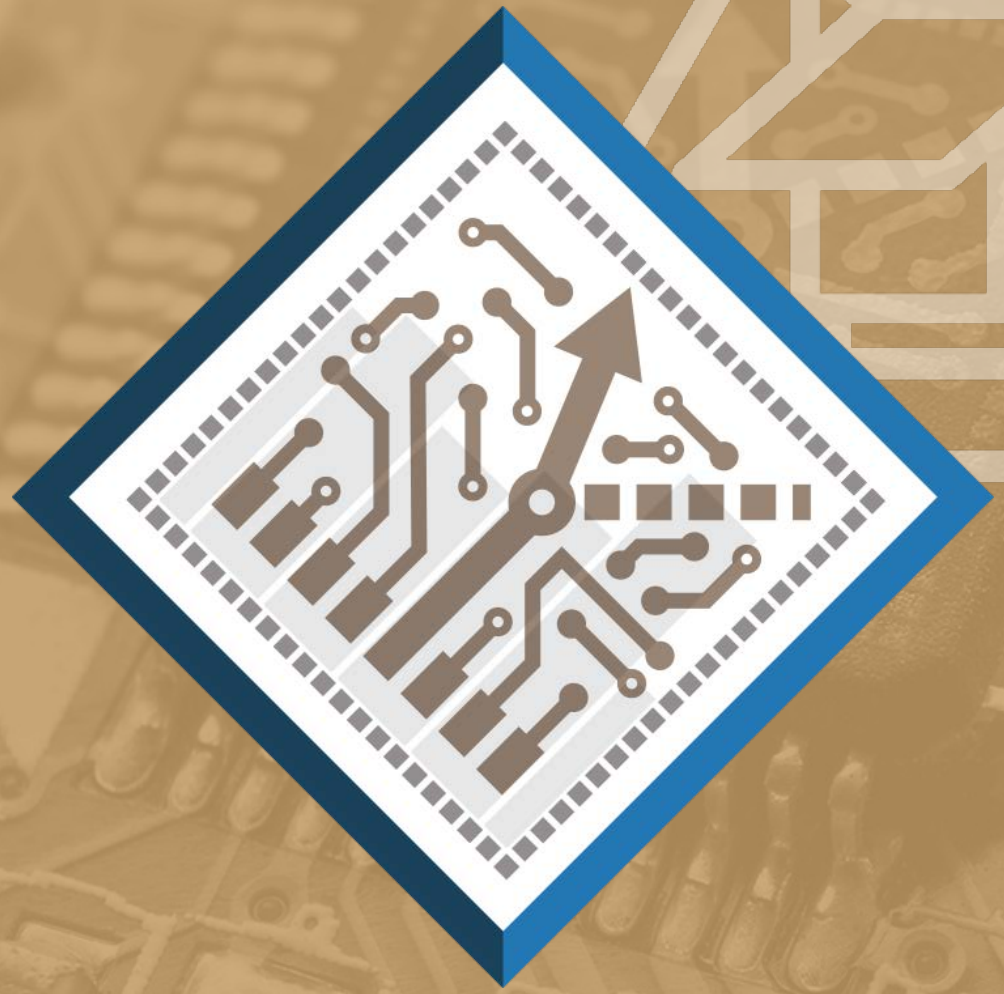
ANDREAS OLOFSSON

PROGRAM MANAGER, DARPA MTO

SANKAR BASU

PROGRAM DIRECTOR, NSF

Approved for Public Release; Distribution Unlimited

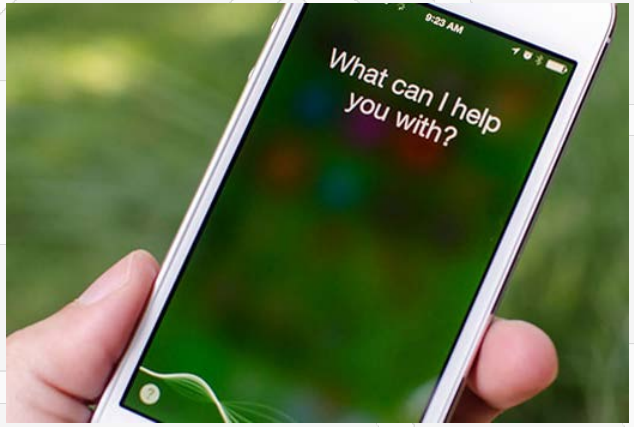


REAL TIME MACHINE LEARNING – DARPA AND NSF COLLABORATION

Approved for Public Release; Distribution Unlimited

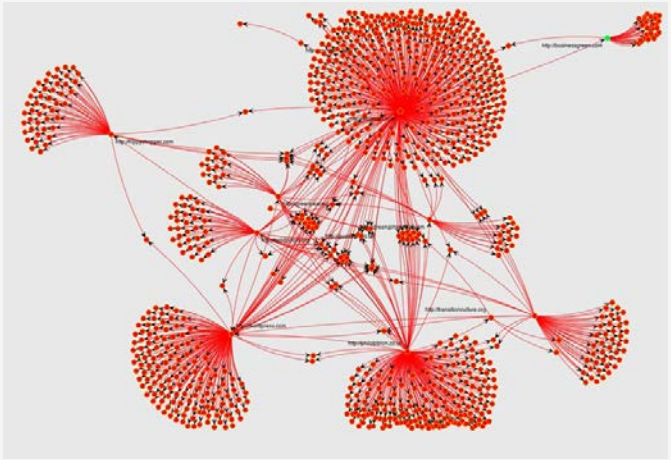
IMPACT OF MACHINE LEARNING

Autonomous vehicles



Assistive AI

Smart AR



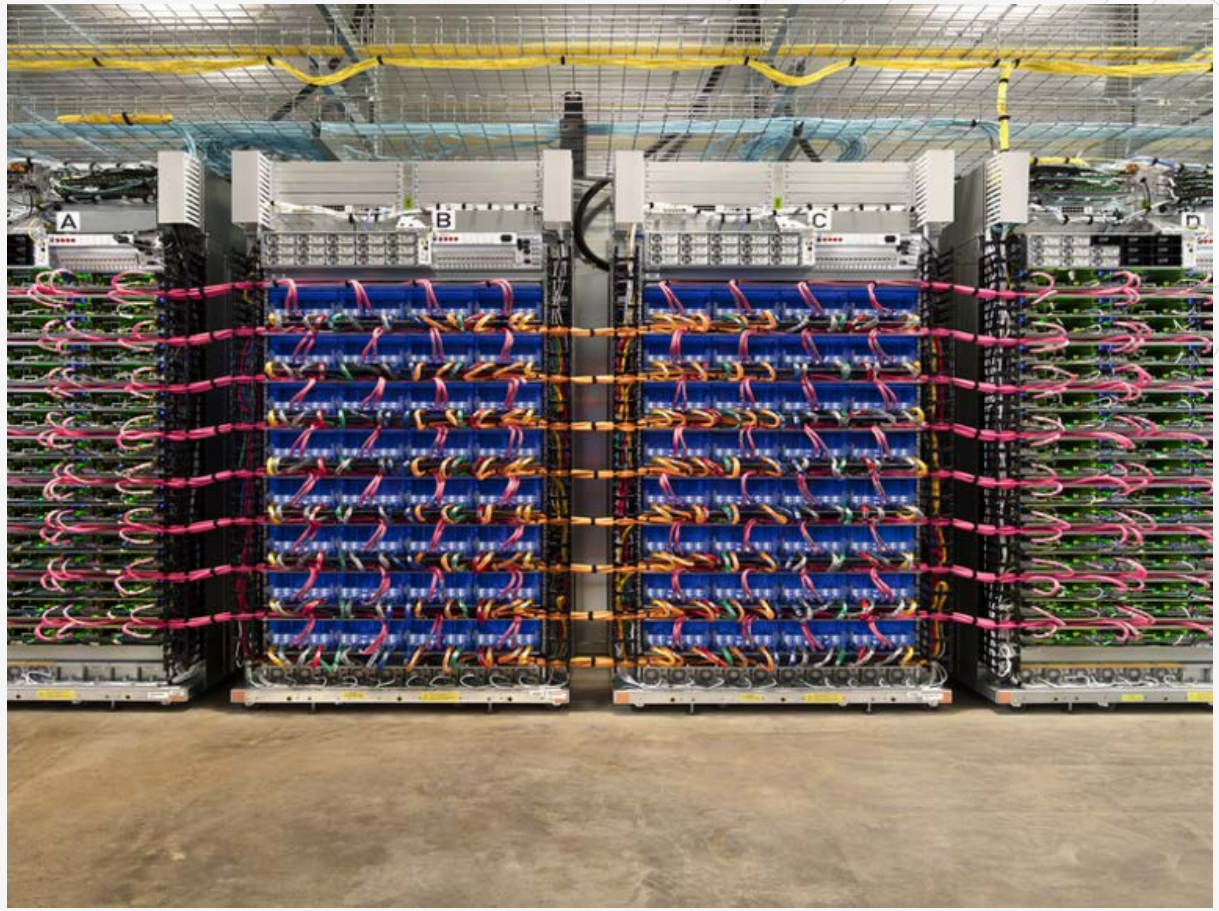
Cybersecurity advances

Facial Recognition

Photos: Wikimedia, US Marine Corps, Flickr, NYT

Approved for Public Release; Distribution Unlimited

CURRENT MACHINE LEARNING LIMITATIONS



>>kW

- Large amounts of data stored
- Processing/training in the cloud
- Edge ML limited to smaller models or reduced sampling/frame rates
- Models are limited/unreliable

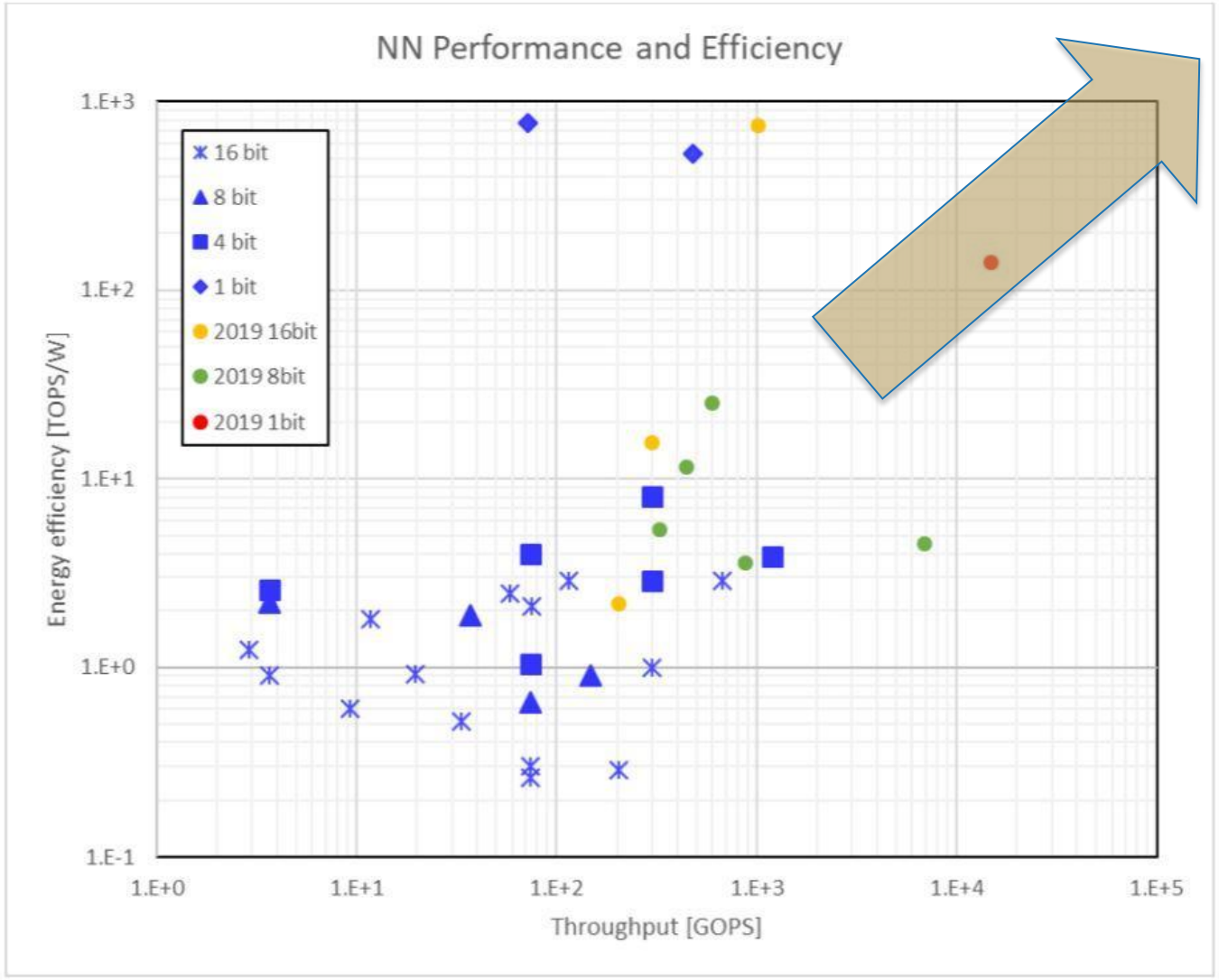
?

<<1W

Photos: Wikimedia, Stanford.Edu, Pixabay, OpenAI

Approved for Public Release; Distribution Unlimited

WHAT IS THE UPPER BOUND ON EFFICIENCY?

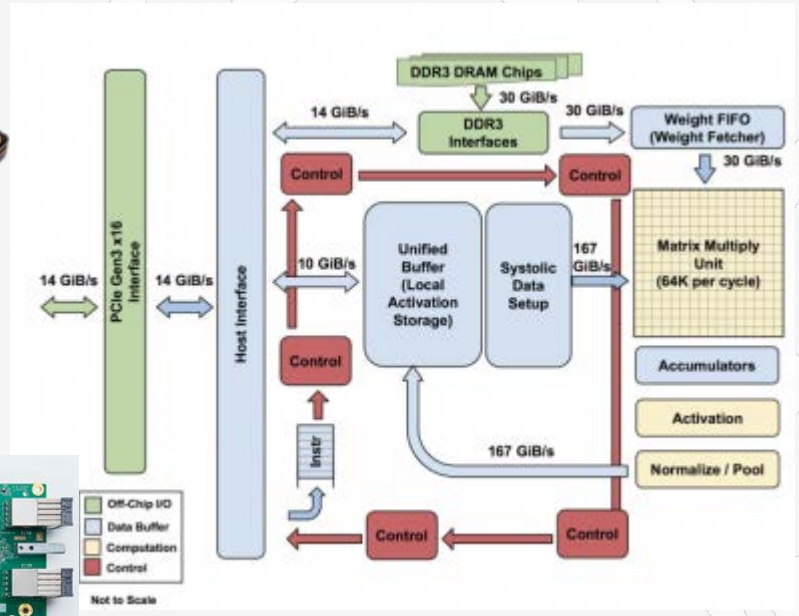


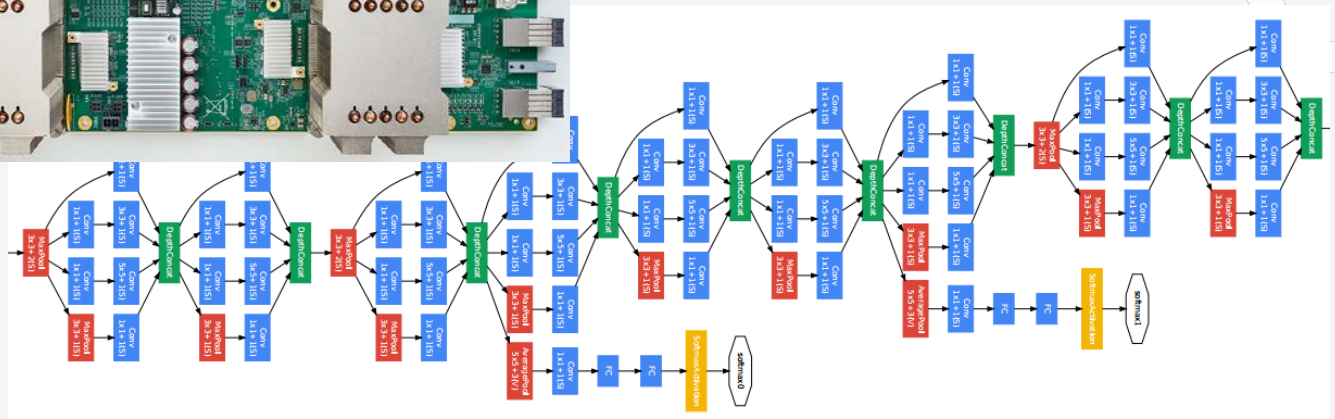
- Extreme expense of HW development means extremely sparse dataset of implementations
- How can we know where to go without an baseline and gradient?

Photos: isscc.org

Approved for Public Release; Distribution Unlimited

WHAT IS THE LOWER BOUND ON LATENCY?



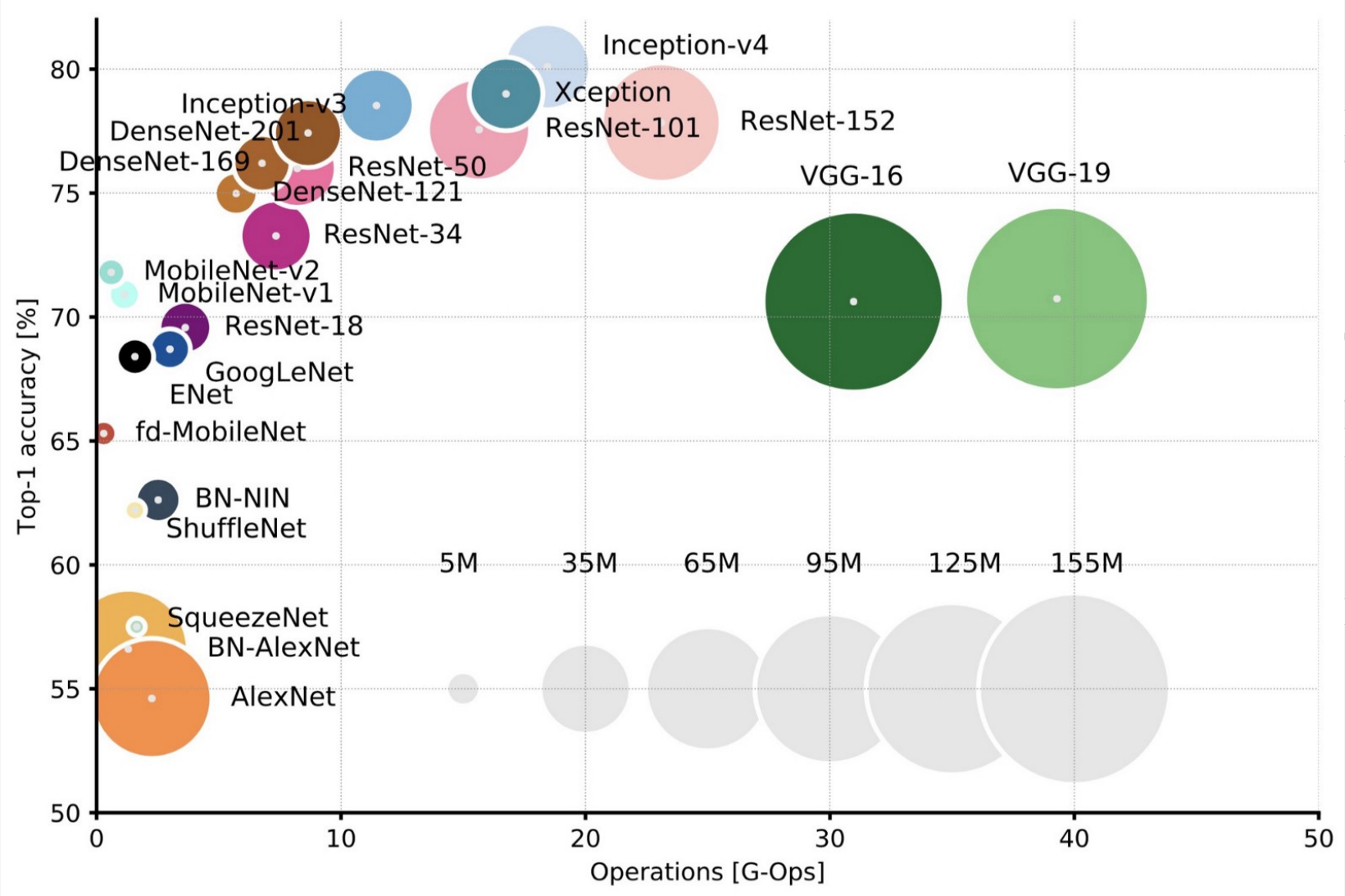


- Current HW optimized for throughput and programmability
- Extreme expense of HW development means latency of ASICs is unexplored
- How low can we go?

Photos: Google, Nvidia

Approved for Public Release; Distribution Unlimited

ACCURACY VS LATENCY VS POWER TRADEOFFS?

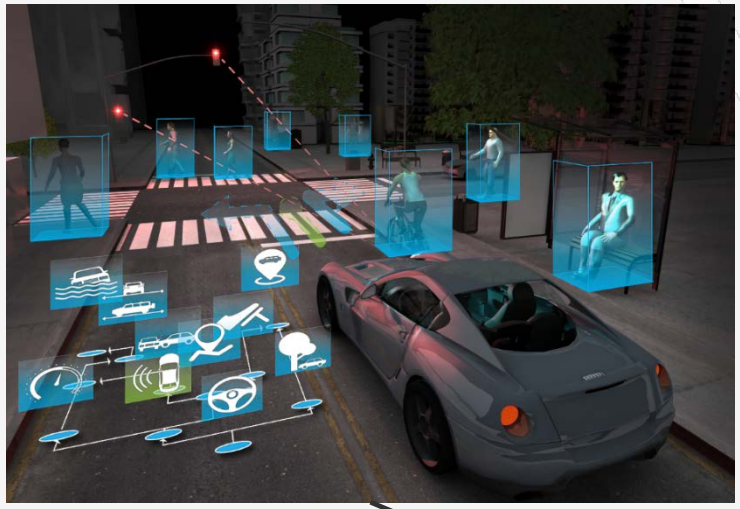


- Overdesign wastes energy
- Can we auto-tune a network for a set of real world constraints?
- Gains of 100X possible

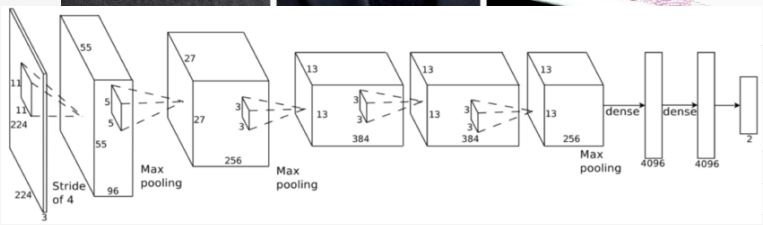
Approved for Public Release; Distribution Unlimited

THE POWER OF BRIDGES

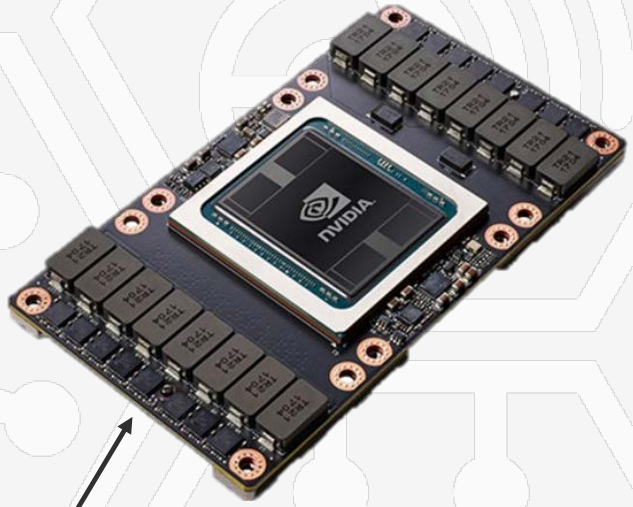
Application Experts

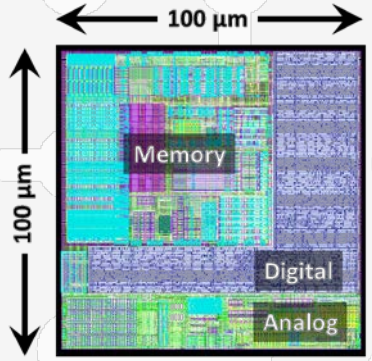


ML Experts



Platform Experts





TensorFlow

PyTorch

RTML (New!)

Objective: Speed of innovation

Photos: NVIDIA, Getty Images, Wikipedia

Approved for Public Release; Distribution Unlimited

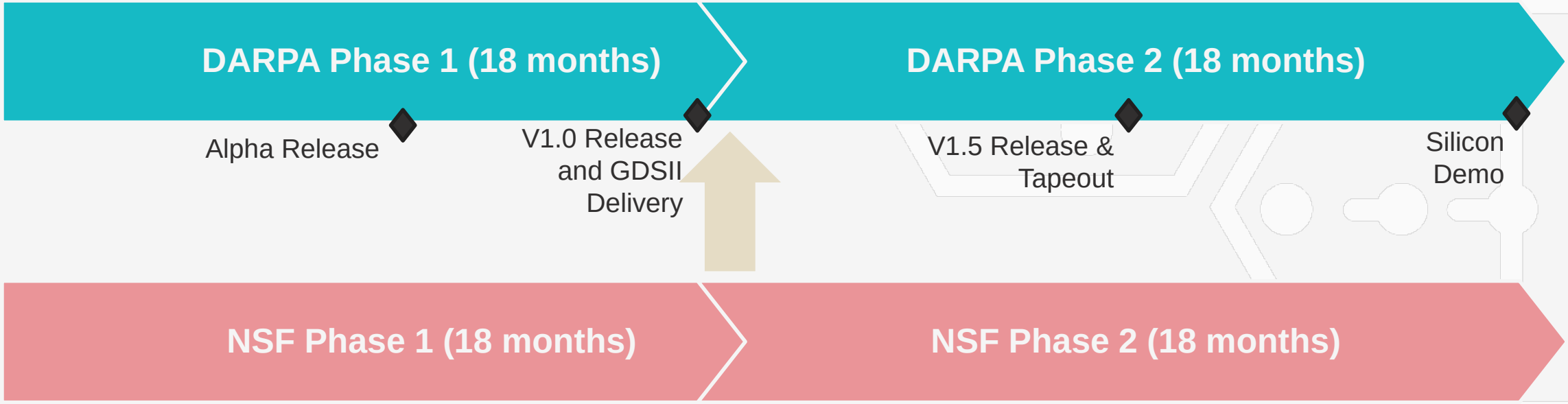
DARPA-NSF REAL-TIME MACHINE LEARNING

NSF

Lay the foundation for next-generation co-design of algorithms and hardware, with the principal focus on developing novel hardware architectures and learning algorithms can be performed in real time.

DARPA

Create a no-human-in-the-loop compiler for automated creation of machine learning ASICs from high level source code.



Approved for Public Release; Distribution Unlimited

NSF LEADERSHIP IN AI

Research Funding

$ NSF invested nearly $450M in AI research (core, applications, systems, infrastructure) in FY 2018

>$100M in “core” AI research

Thought Leadership Across USG

NSTC Select Committee on AI
NSTC Subcommittee on ML & AI
NSTC AI Interagency Working Group (under NITRD): 2016, 2019 *National AI R&D Strategic Plans*
OSTP Assistant Director(s) for AI
International: OECD, G7
Envisioning AI Institutes meeting

Innovative Programmatics

NSF 19-018
Dear Colleague Letter: EArly-concept Grants for Exploratory Research on Artificial Intelligence (AI) and Society - Supported Jointly with the Partnership on AI

NSF Program on Fairness in Artificial Intelligence in Collaboration with Amazon (FAI)

Enabling Access to Cloud Computing Resources for CISE Research and Education (Cloud Access)

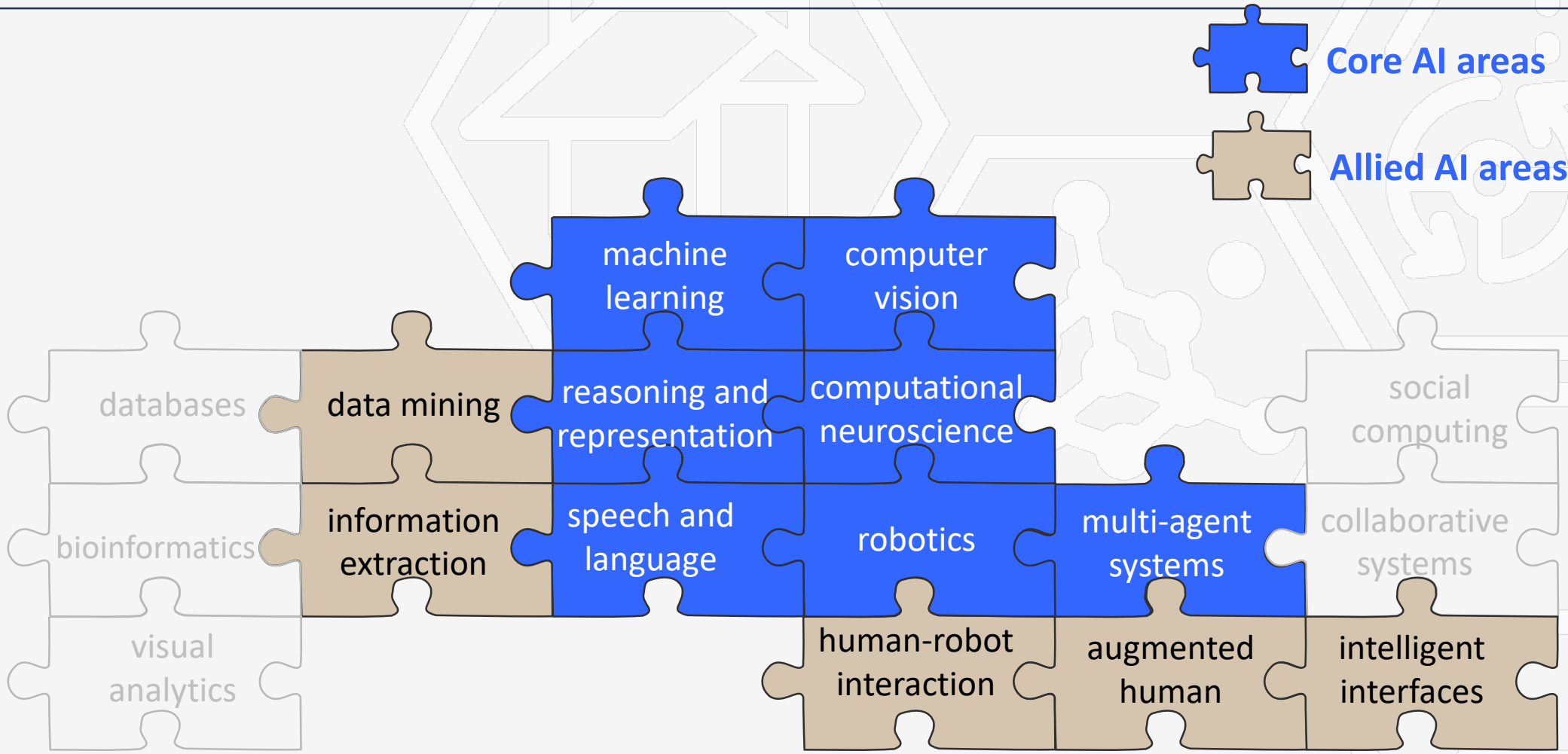
PROGRAM SOLICITATION
NSF 19-510

National Science Foundation

Directorate for Computer & Information Science & Engineering
Division of Computing and Communication Foundations
Division of Information & Intelligent Systems
Division of Computer and Network Systems
Office of Advanced Cyberinfrastructure

Approved for Public Release; Distribution Unlimited

NSF CORE AI THRUSTS

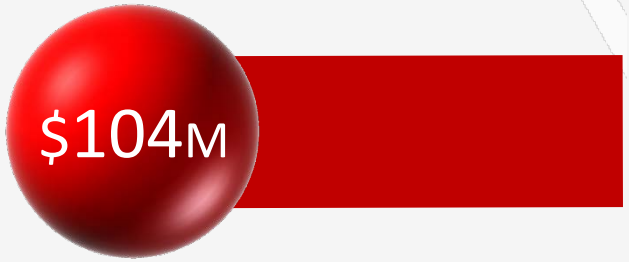


NSF/CISE Division of Information and Intelligent Systems

Approved for Public Release; Distribution Unlimited

NSF INVESTMENT IN CROSS CUTTING AI RESEARCH

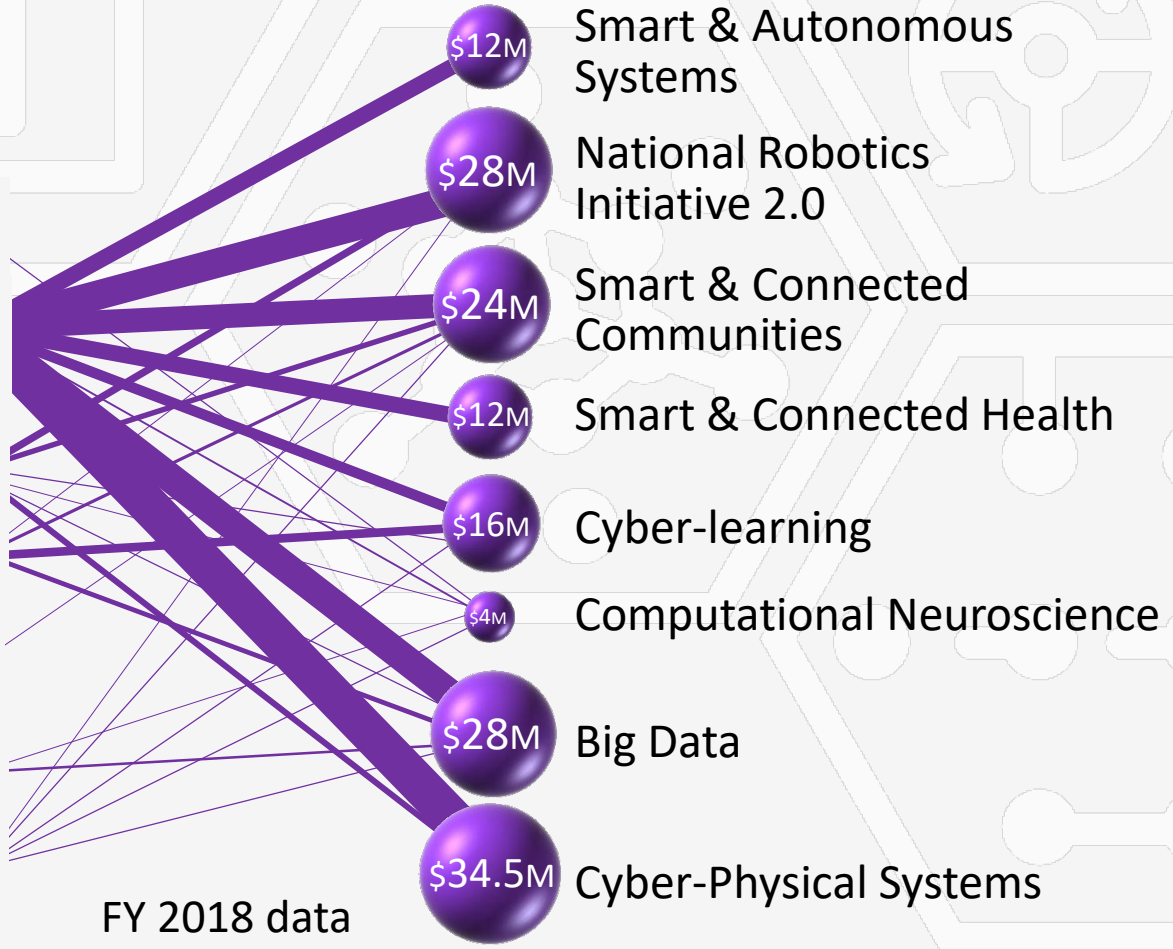
CISE "core" AI



- Robust Intelligence (RI)
- Information Integration and Informatics (III)
- Cyber-Human Systems (CHS)

BIO
CISE
ENG
EHR
GEO
MPS
SBE

"Cross-cutting" AI



FY 2018 data

Approved for Public Release; Distribution Unlimited

NSF INVESTMENT IN CROSS CUTTING AI RESEARCH

CISE "core" AI

- Robust Intelligence (RI)
- Information Integration and Informatics (III)
- Cyber-Human Systems (CHS)

"Cross-cutting" AI

Partners

USDA, DOE, DARPA, AFOSR, ONR

NIH (9 Institutes)

ANR, BMBF, BSF, NICT, NIH

Amazon, Google, Microsoft, IBM

DHS, DOT, NASA, NIH, USDA

Approved for Public Release; Distribution Unlimited

NSF-DARPA COLLABORATION FRAMEWORK

- NSF-RTML and DARPA-RTML are two **distinct** programs.
- Solicitation, proposals submission, review criteria are different, and follow guidelines of the respective agencies.
- Funding will not be co-mingled, i.e., no cofunding will be provided by DARPA for NSF projects, or by NSF for DARPA projects.
- The two programs have **closely related** technical goals, and may provide opportunities for collaborations between the awardees through "***partnership supplements***".
- NSF encourages, but will not require, the two DARPA target applications.

Approved for Public Release; Distribution Unlimited

NSF RTML PROGRAM BUDGET

- Anticipated NSF-funding amount: $10,000,000
- Award size:
  - **Small** Awards: up to $500,000 for 3 years;
  - **Large** Awards: up to $1,500,000 for 3 years.
- Estimated number of awards: 8 to 12
- Approximately $1M of program funds reserved for "***partnership supplements***"
- NSF proposals should budget for up to two project personnel to attend **four** joint NSF-DARPA workshops over the duration of the NSF project (36 months), in the Washington DC area.

Approved for Public Release; Distribution Unlimited

NSF RTML SOLICITATION

- Hardware-software-algorithm cross-layer **co-design** is required;
- Learning algorithms in which **all stages of training** (including incremental training, hyperparameter estimation, and deployment) can be performed in **real time** will receive higher priority, recognizing the asymmetry in real-time learning versus real-time inference;
- Proposals submitted to this program should seek to demonstrate radical improvement in the **metrics of performance**, e.g., size, weight, power, latency and efficiency;
- **Approximate algorithms** for efficient implementation, e.g., low-precision gradient computation and sparsely-connected neuronal nodes, are within scope;
- Approaches to **self-assessing** systems are within scope;
- **Distributed ML** algorithms and hardware for real-time performance are within scope;
- Efficient and novel utilization of data and memory paths, e.g., in- or **near-memory** computations, are within scope;
- The program is also interested in **analog/mixed-signal** architectures.

Approved for Public Release; Distribution Unlimited

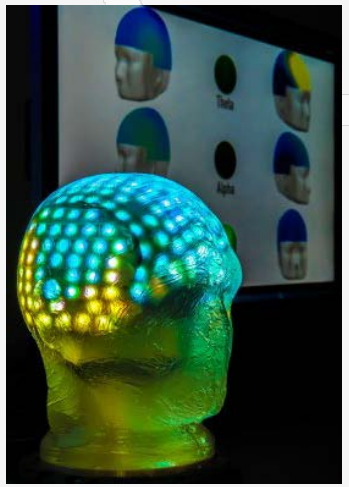
DARPA/NSF RTML PROGRAM END STATE

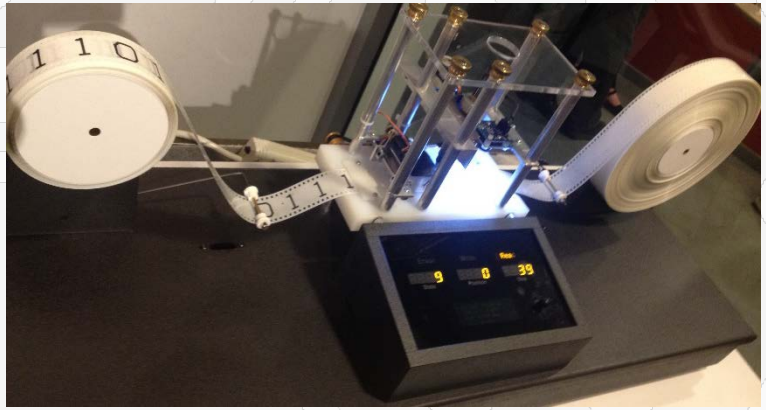
NEARER TERM (DARPA)

- A no human in the loop compiler for rapid design of custom machine learning ASICs
- A baseline of hardware data points to to guide future AI research

LONGER TERM (NSF)

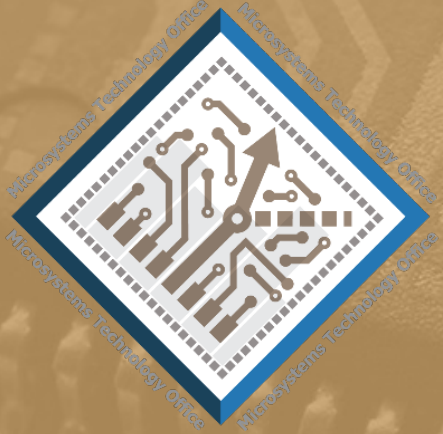
- Breakthroughs in real time adaptability at the edge
- Breakthroughs in processing efficiency, throughput, and latency





Photos: Army.Mil, Wikipedia, Arxiv, MadSciBlog

Approved for Public Release; Distribution Unlimited



ERI ELECTRONICS RESURGENCE INITIATIVE

SUMMIT

2019 | Detroit, MI | **July 15 - 17**

Approved for Public Release; Distribution Unlimited