CENTER FOR BRAIN-INSPIRED COMPUTING

ENABLING AUTONOMOUS INTELLIGENCE

KAUSHIK ROY

ANAND RAGHUNATHAN

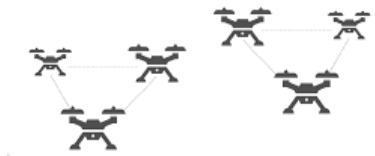
PURDUE UNIVERSITY

SRC JUMP

C-BRIC

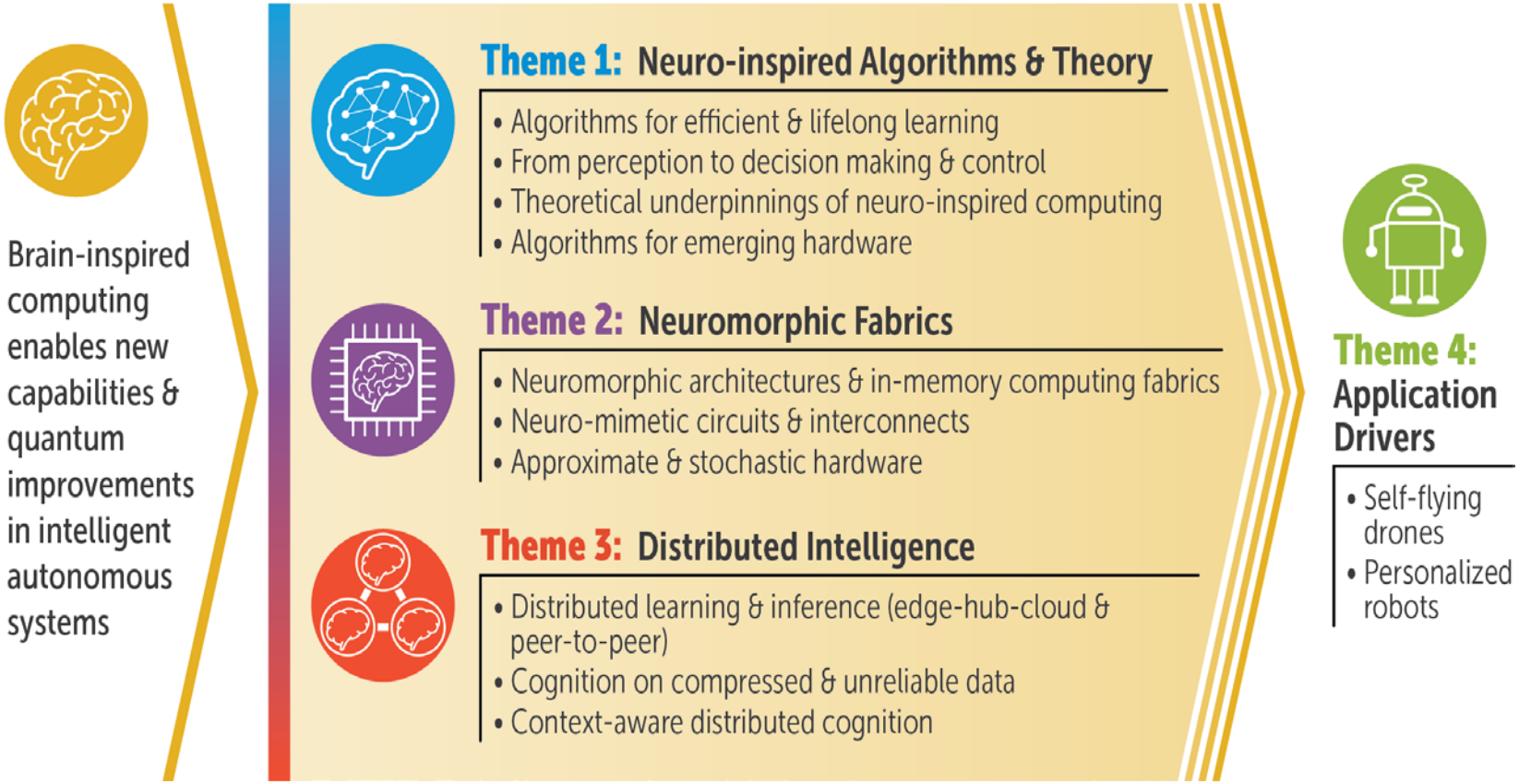
C-BRIC Vision

- Enable next generation of intelligent autonomous systems
  - Narrow the orders-of-magnitude computing efficiency gap between current computing systems and the brain
  - Drive improvements in the robustness of cognitive computing systems
  - Explore distributed intelligence across edge/hub/cloud and peer-to-peer networks
  - Demonstrate the impact of these advances in end-to-end systems such as autonomous drones and personal robotics



C-BRIC

C-BRIC Organization



C-BRIC

Theme 1: Neuro-inspired Algorithms and Theory

State-of-the-Art: Deep Neural Nets

- Largely supervised learning
- Static (one-time) learning
- Training requires global updates (Backpropagation / SGD)
- Perception (speech, images, text)
- Unknown generalization behavior
- Manually designed network topologies

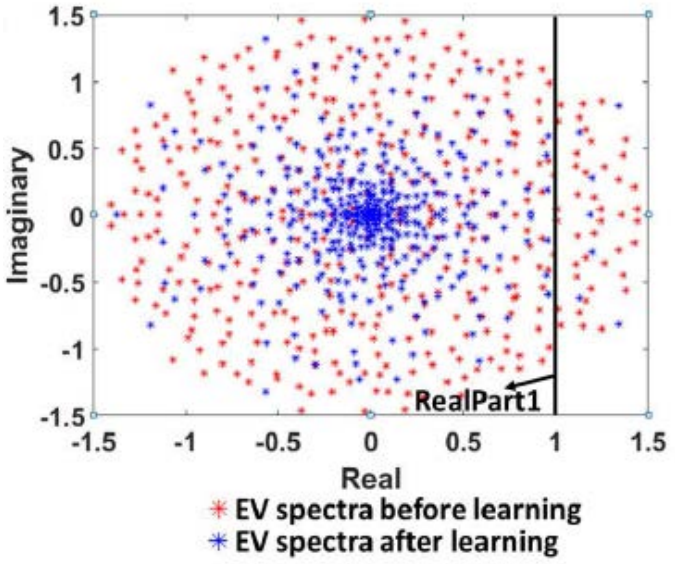
C-BRIC Theme 1

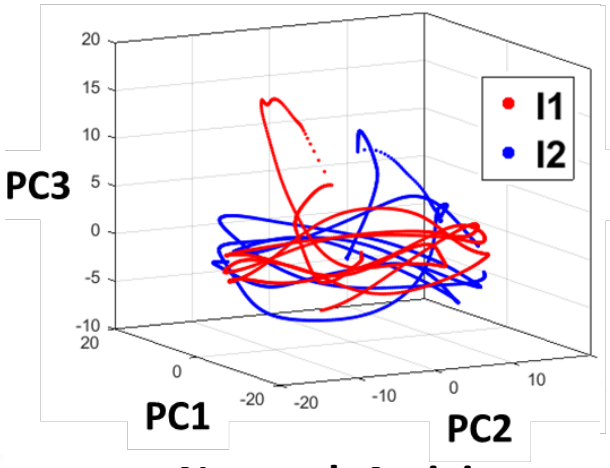
Future Neuro-Inspired Algorithms

- **Computationally efficient algorithms**
- **Theory of neural computing from DNN to emerging models**
- **Learning with less data**
- **Incremental and lifelong learning**
- **Algorithms that leverage stochastic and approximate computation**
- **Learning and inference on emerging computing fabrics**

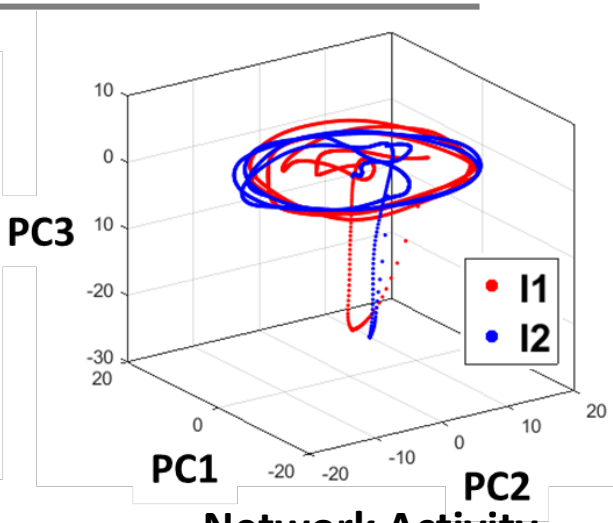
C-BRIC

Theoretical Understanding of Learning





Network Activity before Learning



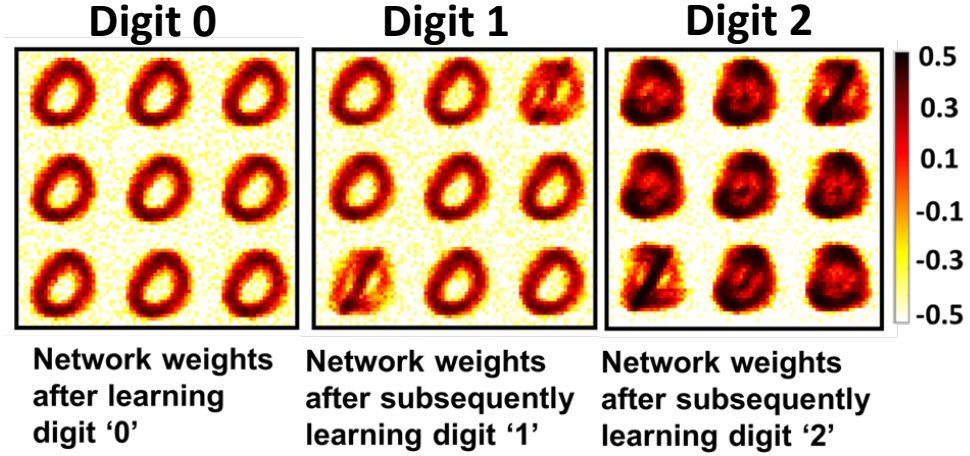
Network Activity after Learning

- Shrinking of the EigenValue spectral circle represents the stabilizing effect of the learning mechanism
- Understanding network behavior from Random Matrix theory and Principal Component Analysis
- Quantification of stabilizing hyperparameters from network activity

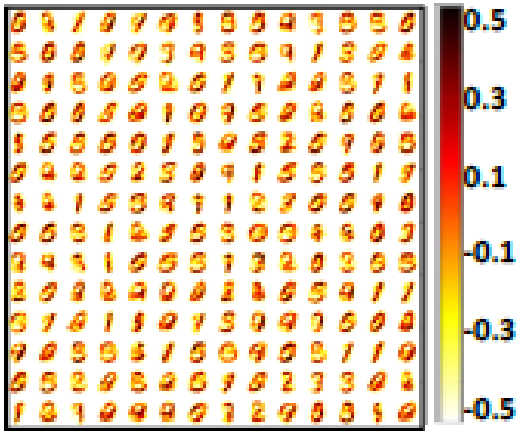
C-BRIC

Learning to Forget with Adaptive Decay

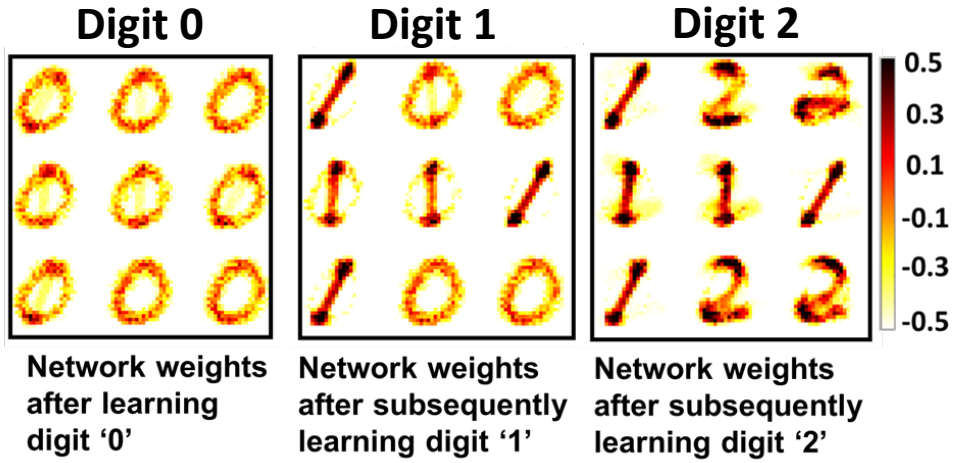
STDP Learning



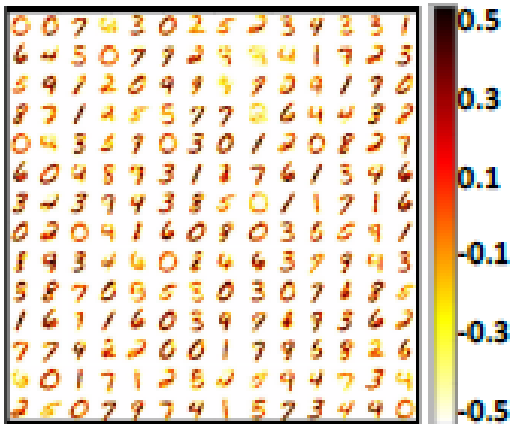
Digits
0, 1, 2, ..9



Adaptive Plasticity Learning



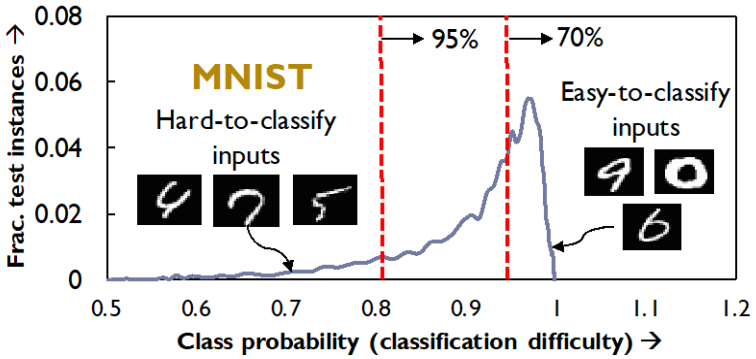
Digits
0, 1, 2, ..9



C-BRIC

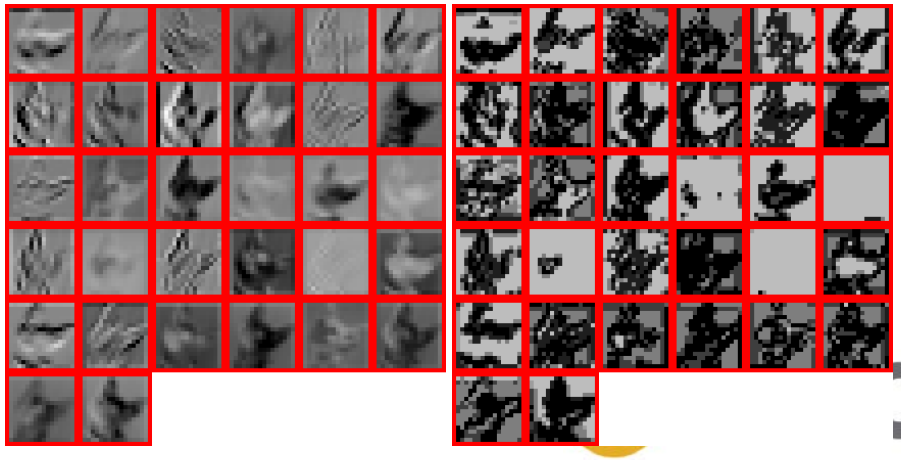
Dynamic, Variable-Effort Deep Networks

- Deep nets are fixed-effort and static
- Inputs differ greatly in their difficulty
- Mechanisms to dynamically modulate computational effort of neural nets



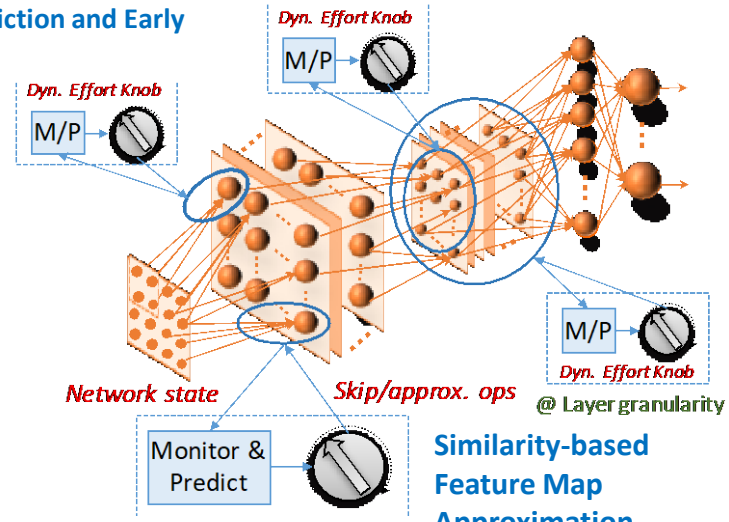
Input (CIFAR-10)

Feature maps (C1 layer) Computational effort



Significance-driven Feature Evaluation

Saturation Prediction and Early Termination



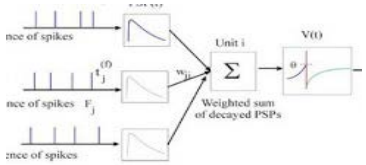
Similarity-based Feature Map Approximation

Stochastic Neurons and Synapses

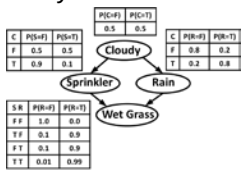
Explore applications that exploit probabilistic inference

Applications Enabled by Probabilistic Inference

Spiking Networks

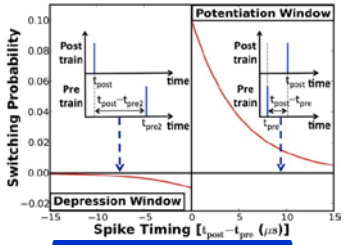


Bayesian Networks

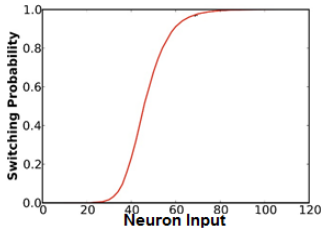


Adapt traditional deterministic neural and synaptic computing to stochastic models

Stochastic Neural and Synaptic Computing



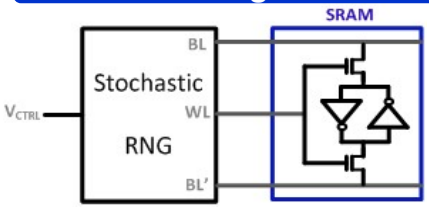
Synaptic Learning



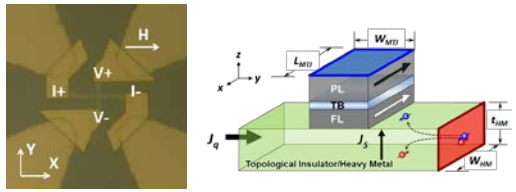
Neural Inference

Underlying hardware fabrics – CMOS & post-CMOS technologies

Enabling Technologies: CMOS and post-CMOS



CMOS Technology



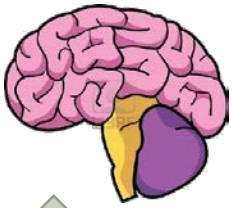
Post-CMOS Technology

Algorithm-Hardware-Application Co-Design

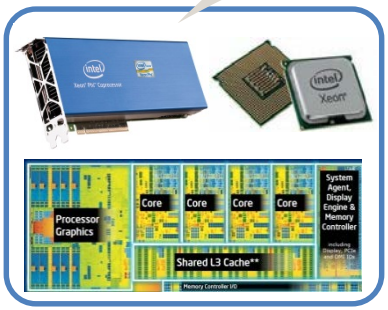
C-BRIC

Theme 2: Neuromorphic Fabrics

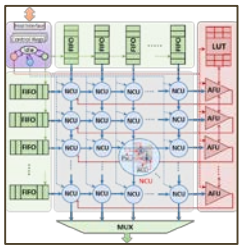
- CMOS and Post-CMOS neuro-mimetic devices and interconnects
- Compute-near-memory / Compute-in-memory
- Approximate and stochastic neuronal and synaptic hardware
- Architectures that embody computing principles from the brain (sparse, irregular, event-driven, massively parallel)
- Programming and evaluation frameworks



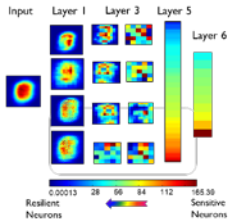
Multicores/GPUs



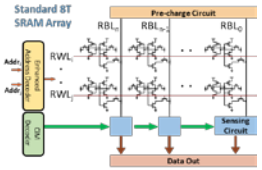
Accelerators



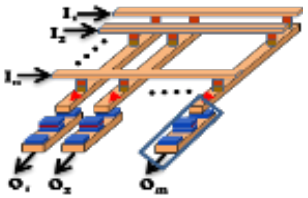
Approximate & Stochastic Hardware



In-memory computing



Post-CMOS Devices

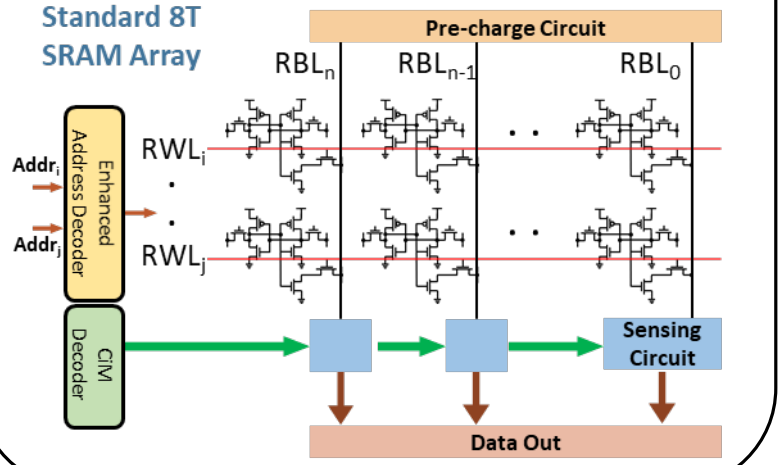


~10^4 Energy Gap

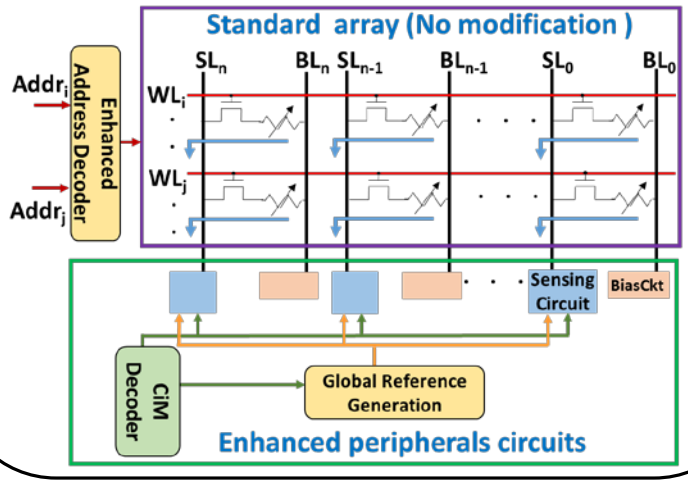
C-BRIC

Compute-in-Memory

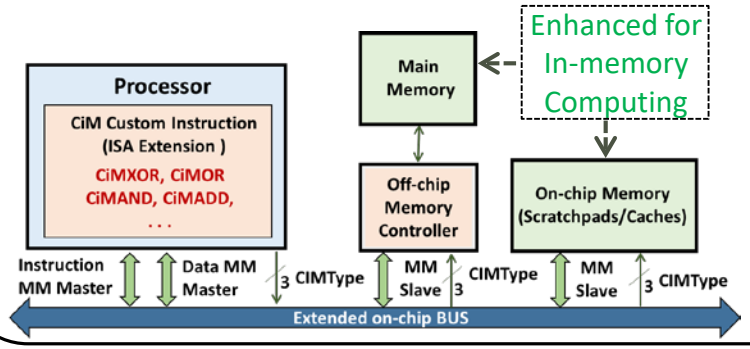
Compute-in-memory with CMOS



Compute-in-memory with Post-CMOS



System for In-memory Computing



Load R_{ADDR1} R_{DEST0}
Load R_{ADDR2} R_{DEST1}
ADD R_{DEST0} R_{DEST1} R_{OUT}

Conventional Computing

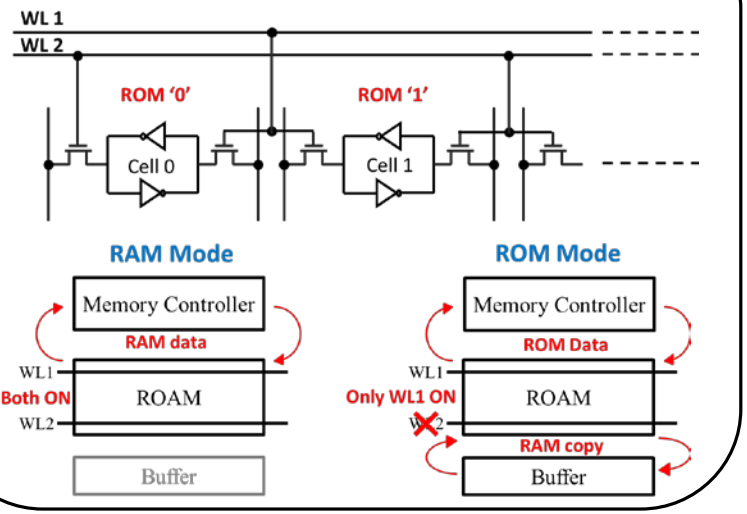
CiMADD R_{ADDR1} R_{ADDR2} R_{OUT}

In-memory Computing

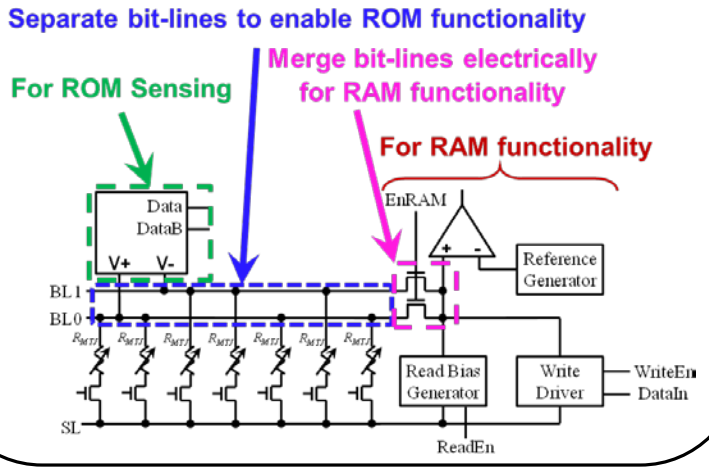
C-BRIC

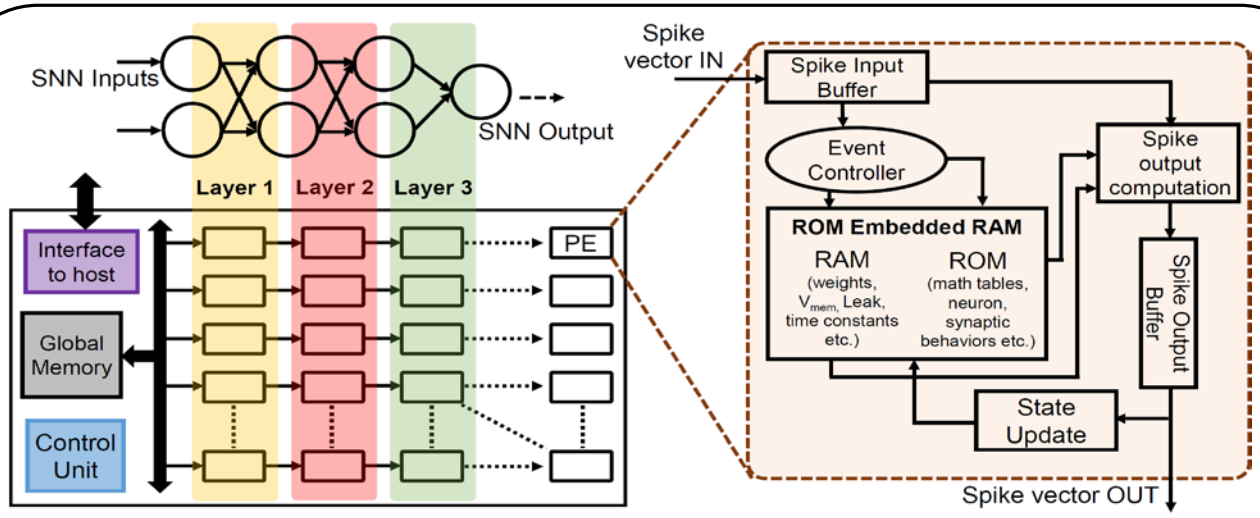
Compute-in-Memory: Embedding ROM In RAM

ROM Embedded CMOS memories



ROM Embedded Post-CMOS memories

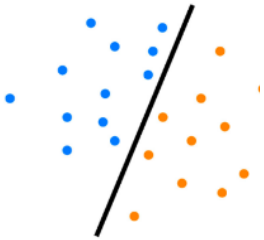




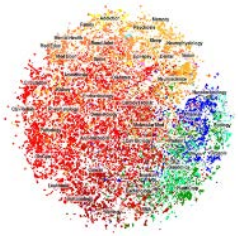
Distributed architecture for in-memory SNN computations with ROM-Embedded RAMs

C-BRIC

Hardware Demonstration of Autonomous Decision Making via Reinforcement Learning

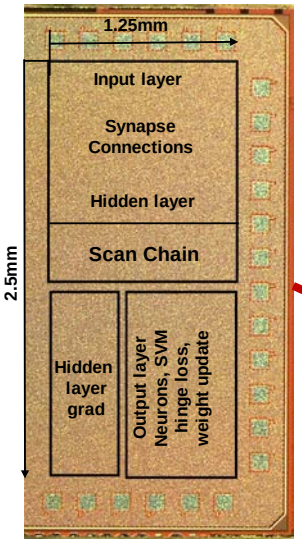


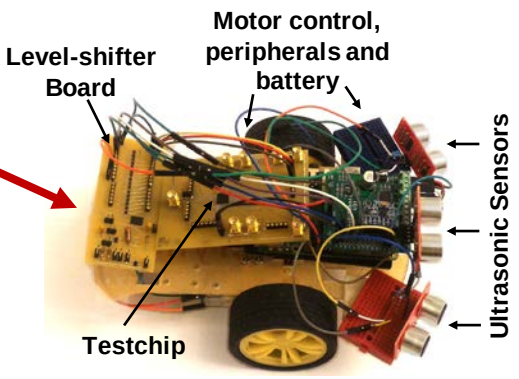
Supervised
Learning
known
patterns

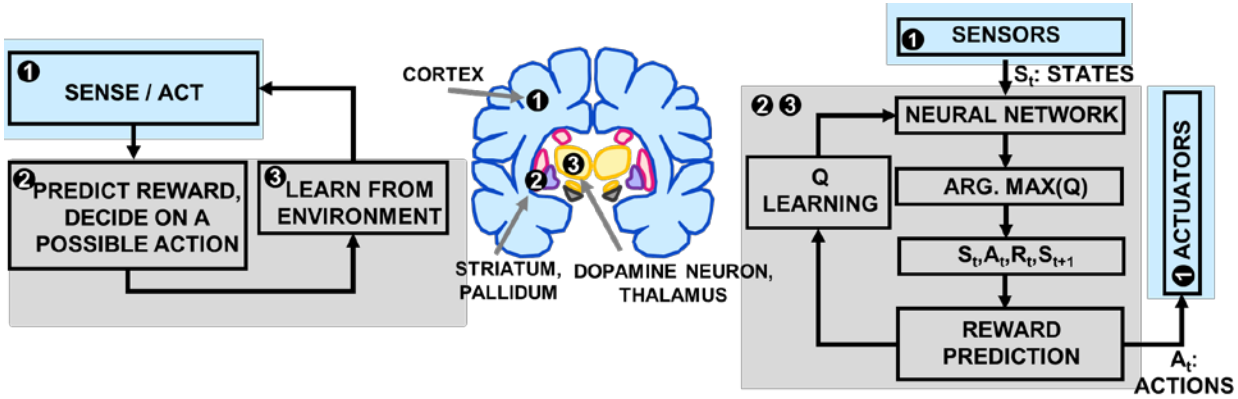


Unsupervised
Learning
unknown
patterns

Reinforcement
Generating data
Learning
patterns







C-BRIC

Theme 3: Distributed Intelligence

State-of-the-Art: Cloud-enabled Intelligence

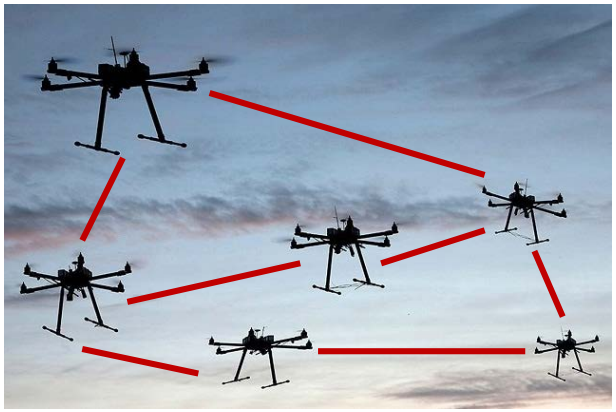
- Centralized training in cloud
- Inference entirely in cloud or entirely on edge device
- Algorithms agnostic to distributed context require high communication

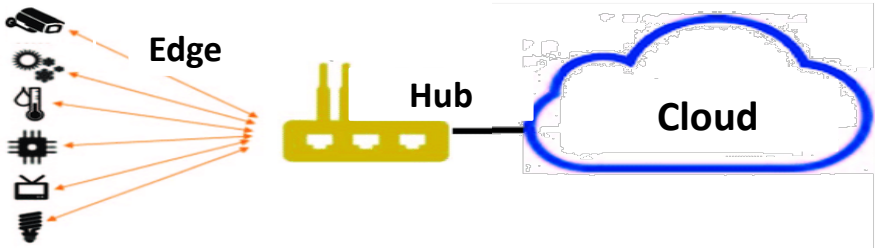
C-BRIC Theme 3

Future Distributed Intelligence

- **Partitioned learning and inference**
  - **Algorithms for hierarchical (edge/hub/cloud) and peer-to-peer networks**
- **Cognition on compressed and unreliable data**
  - **Event-driven sensors, data fusion, learning from incomplete/ unsynchronized/noisy data**
- **In-sensor analytics**
  - **Low-complexity algorithms and hardware to enable in-sensor computing**

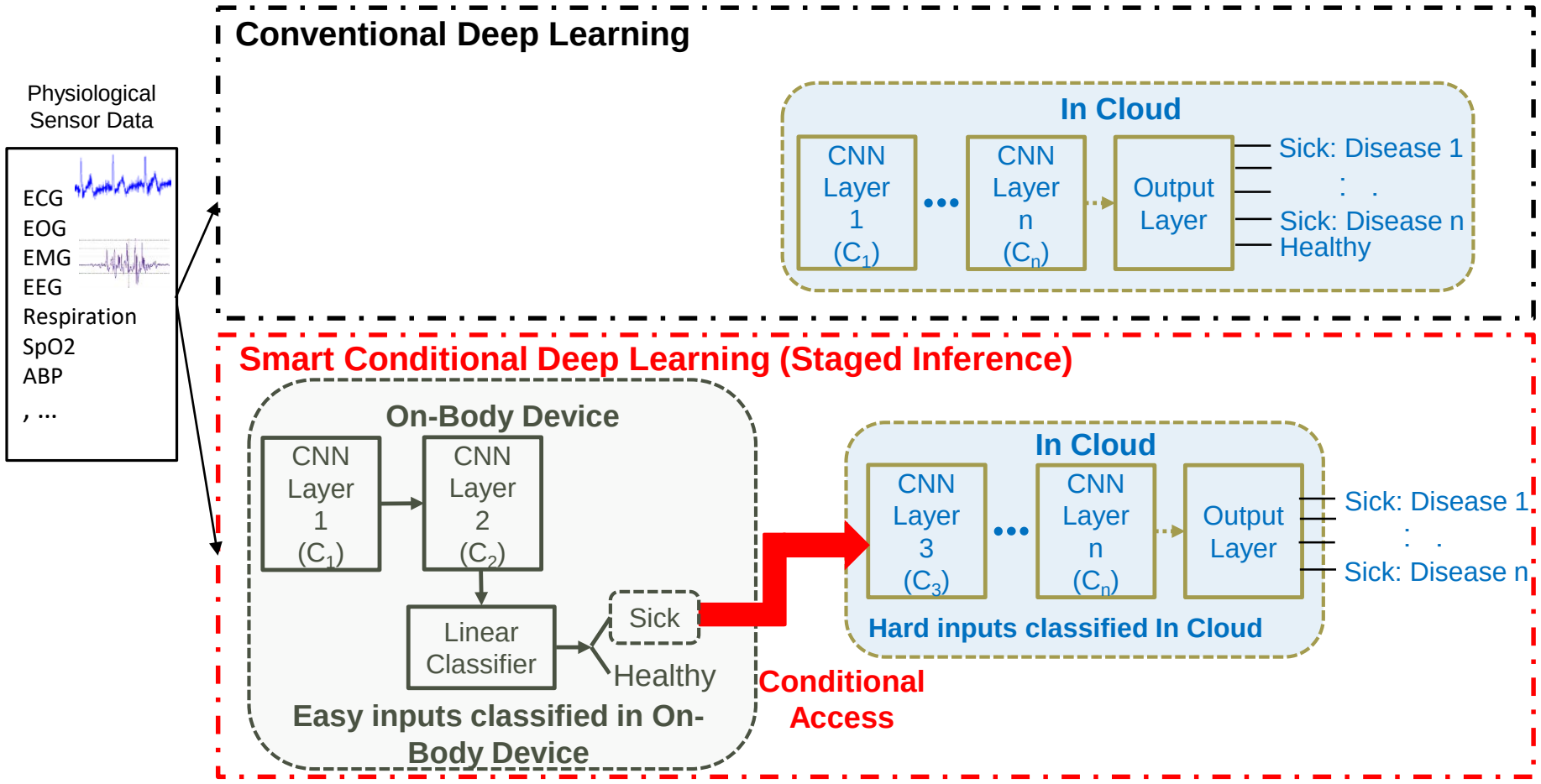
Peer-to-peer



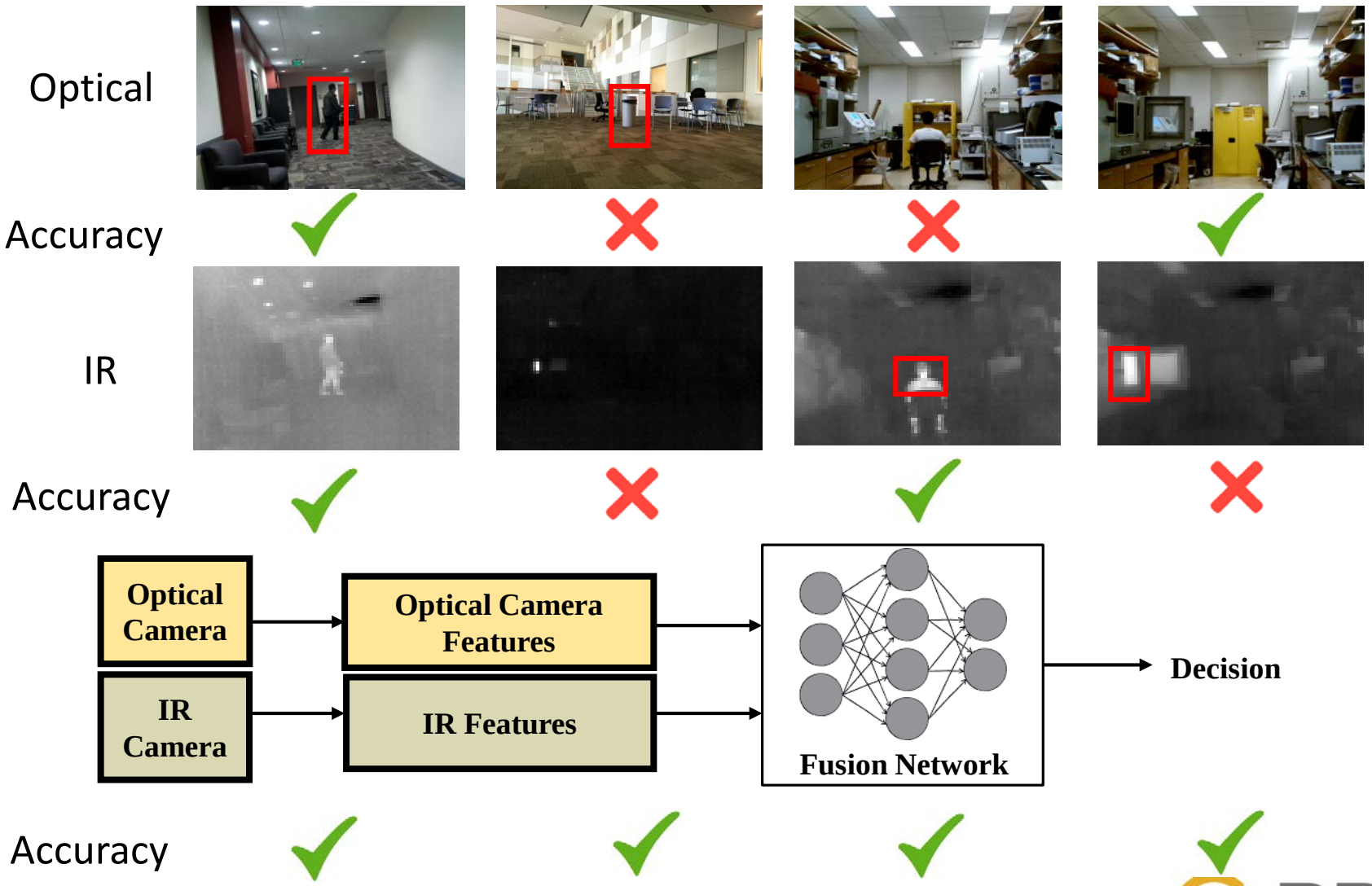


C-BRIC

Staged Conditional Learning/Inference

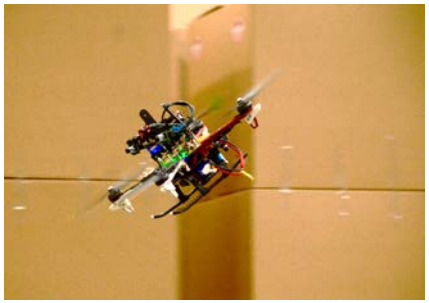


Multi-sensor Cognition in Smart Buildings



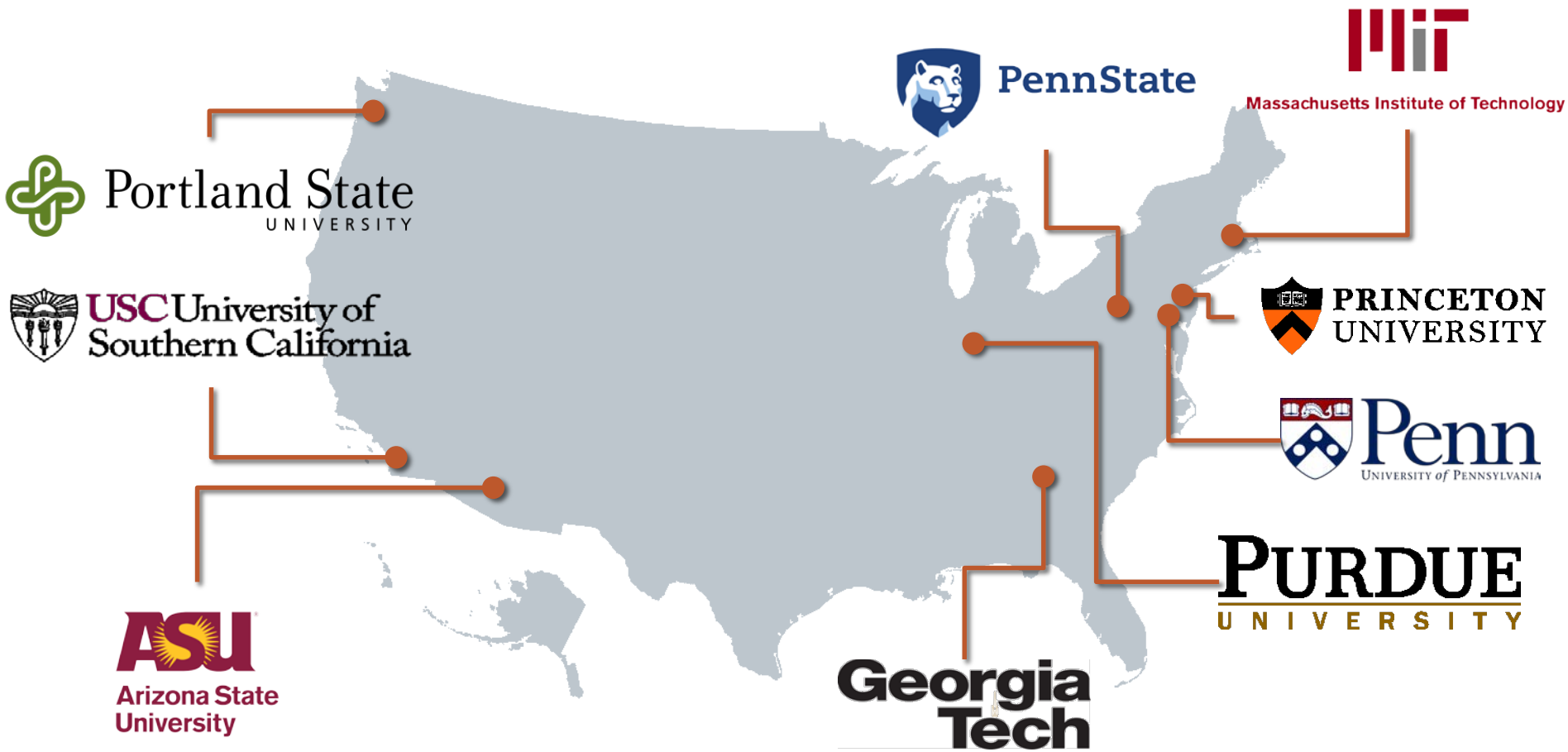
Theme 4: Application Drivers

- Autonomous drones and drone swarms
- Personal robotic assistants
- Technologies from Themes 1-3 enable new capabilities with real-time, autonomous operation




C-BRIC

C-BRIC Universities



C-BRIC