Computational Self-Awareness: A Paradigm for Adaptive, Resilient Computing Platforms

Nikil D. Dutt

Departments of CS, CogSci, and EECS

University of California, Irvine

dutt@uci.edu

http://www.ics.uci.edu/~dutt

https://duttgroup.ics.uci.edu

Work done jointly with DRG

Research Partially Supported by the National Science Foundation

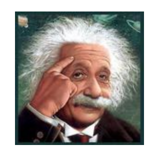
What is Self-Awareness?

Awareness & Adaptation in Biology

[David Gallo: Underwater astonishments, TED]

Awareness vs. Self-Awareness in Biology

Species	Neurons	Synapses
Nematode	302	10^3
Fruit Fly	100,000	10^7
Honeybee	960,000	10^9
Mouse	75,000,000	10^{11}
Cat	1,000,000,000	10^{13}
Human	85,000,000,000	10^{15}



Source – http://en.wikipedia.org/wiki/List_of_animals_by_number_of_neurons

Self-awareness in Sci-Fi

2001, A Space Odyssey (MGM 1968)

- Recognition of mission goals
- Situational Awareness
  - Sense if "something" is wrong or abnormal
- Compensate for sensory impairments
- Deductive reasoning
  - Can see several steps ahead
- Draw conclusions

What is Self-Awareness?

Self-awareness

From Wikipedia, the free encyclopedia

Not to be confused with Self-concept, Self-consciousness, Self-perception, or Self image.

This article has multiple issues. Please help improve it or discuss these issues on the **talk page**. [hide]

- This article **may require cleanup to meet Wikipedia's quality standards**. *(March 2009)*
- This article **needs attention from an expert on the subject**. *(May 2009)*

Self-awareness is the capacity for introspection and the ability to recognize oneself as an individual separate from the environment and other individuals.

The mirror test is a simple measure of self-awareness.

Contents [hide]

Computational Self-Awa
A Hierarchical Vie

self-managing, self-governing, self-maintenance, self-control, self-evaluating, self organizing

Fig. 1. Hierarchy of the self-* properties.

Global Properties

in accordance
to biological self-adaptation

self-monitoring, self-situated

[SALEHIE 2009, TAAS]

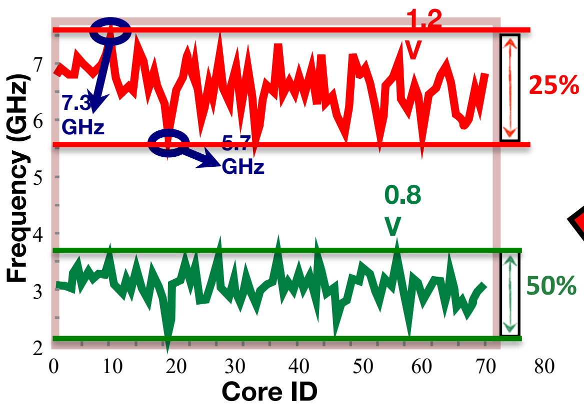
Why On-Chip Self-Awareness (1)?

[Source: NSF Variability Expedition Project]

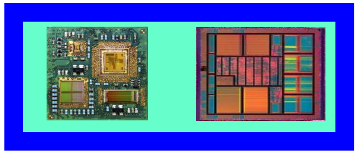
Why On-Chip Self-Awareness (1)?

Variability-induced challenges

Manufacturing



Frequency variation across 80-core processor within a single die in Intel's 65nm node at 1.2V and 0.8V [Dighe10]

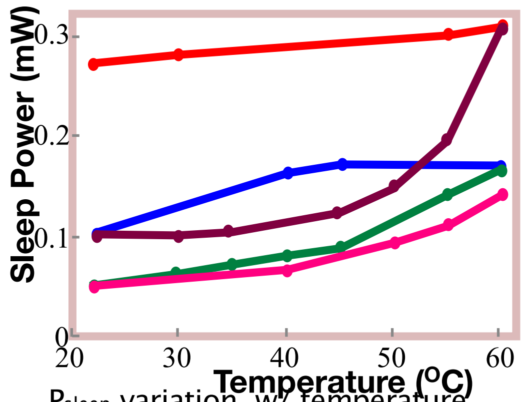


[Source: NSF Variability Expedition Project]

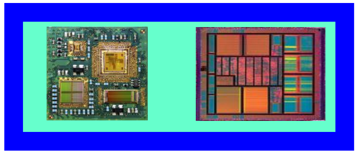
Why On-Chip Self-Awareness (1)?

Variability-induced challenges

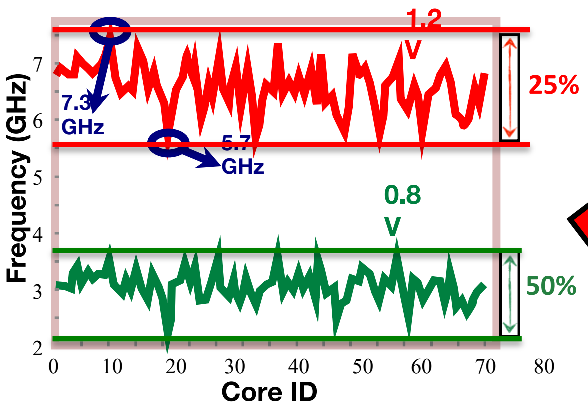
Environment



P_{sleep} variation w/ temperature across 5 ARM Cortex M3 processor instances



Manufacturing

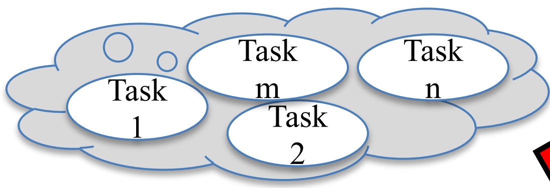


Frequency variation across 80-core processor within a single die in Intel's 65nm node at 1.2V and 0.8V [Dighe10]

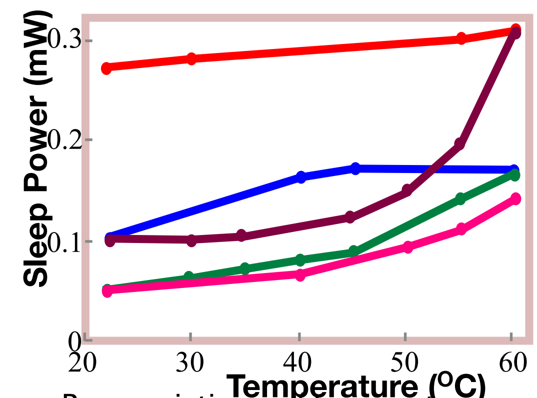
Why On-Chip Self-Awareness (1)?

Variability-induced challenges

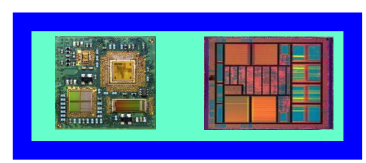
Applications: *varying compute, memory, communication*



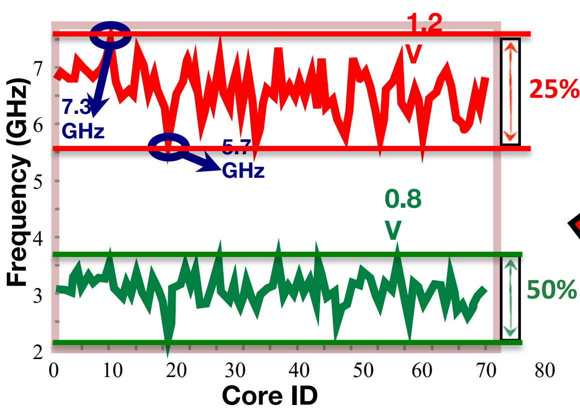
Environment



P_{sleep} variation w/ temperature across 5 ARM Cortex M3 processor instances



Manufacturing

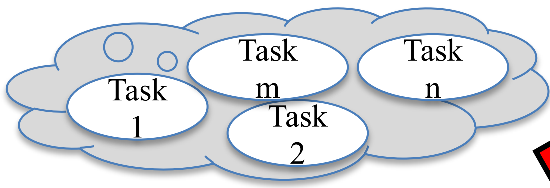


Frequency variation across 80-core processor within a single die in Intel's 65nm node at 1.2V and 0.8V [Dighe10]

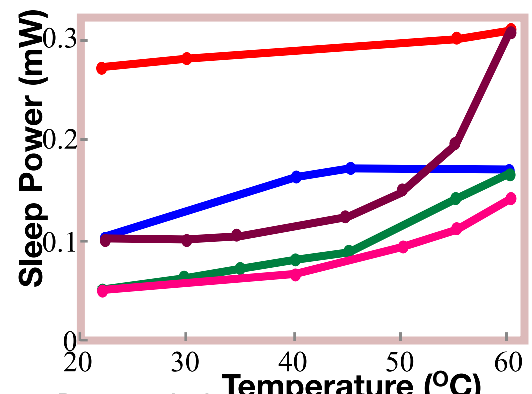
Why On-Chip Self-Awareness (1)?

Variability-induced challenges

Applications: *varying compute, memory, communication*

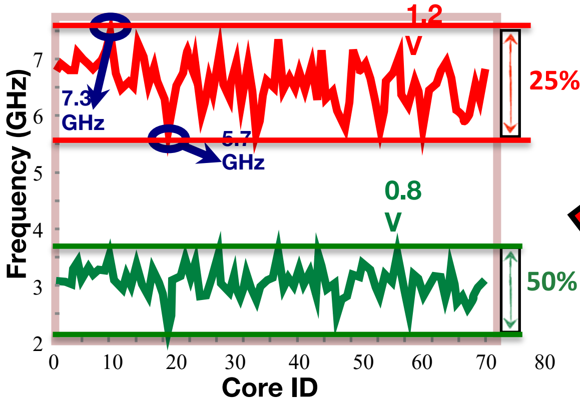


Environment

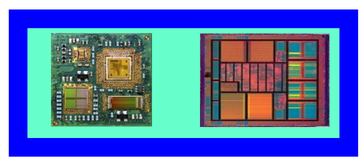


P_{sleep} variation w/ temperature across 5 ARM Cortex M3 processor instances

Manufacturing



Frequency variation across 80-core processor within a single die in Intel's 65nm node at 1.2V and 0.8V [Dighe10]

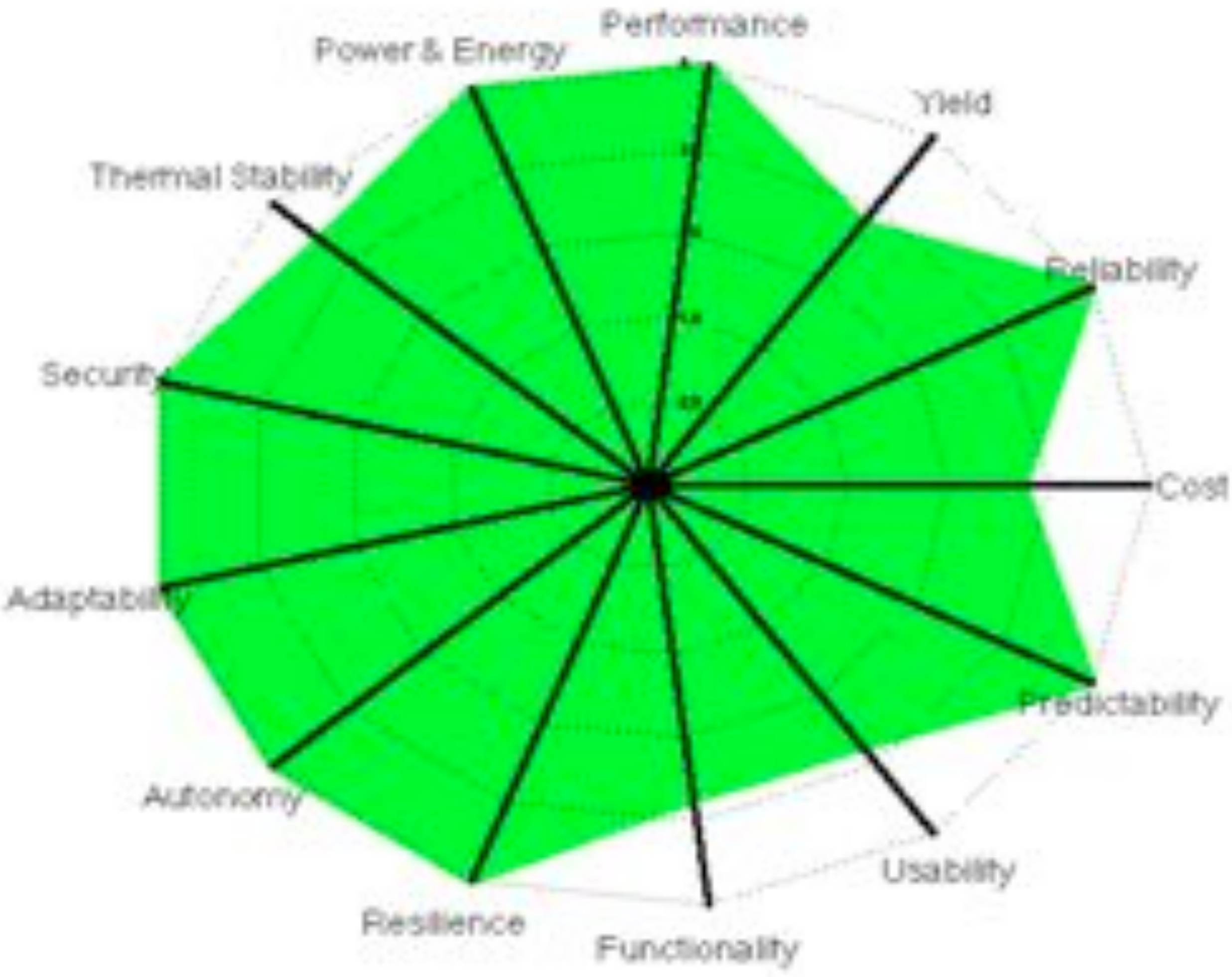


Triple Whammy!

Why On-Chip Self-Awareness (2) ?

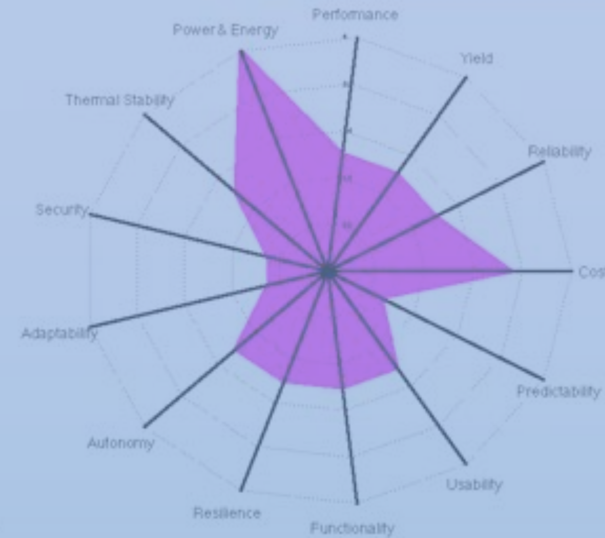
- Chips must adapt to Dynamic Performance, Power, Resilience, Security,….
  - See Radar chart (Kiviat graph) examples
- Provide Guarantees
- Exploit trade-offs in several dimensions

Ideal Radar Chart?

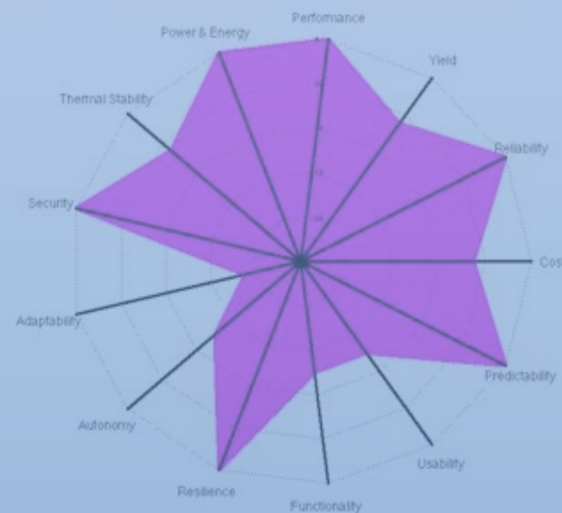


Reality

Performance Driven

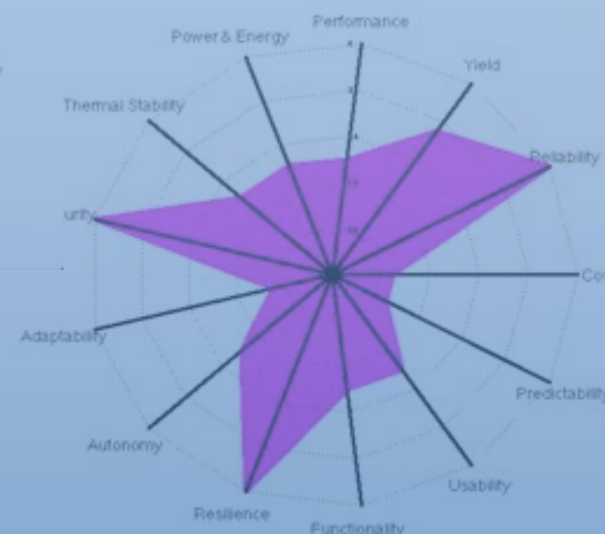


Energy/Power Driven



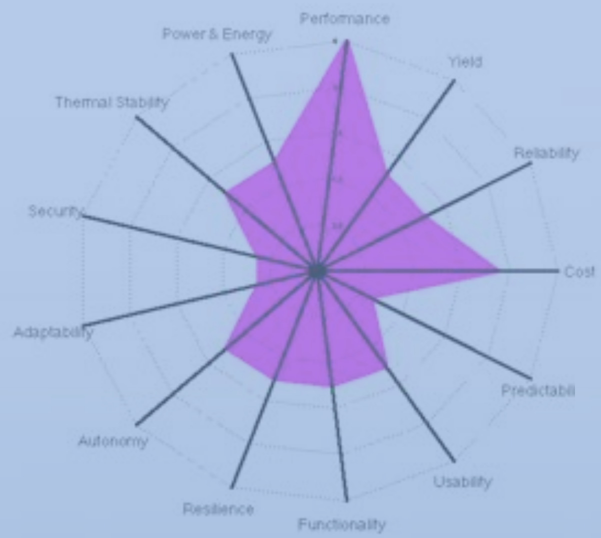
QoS Combination

Reliability Driven

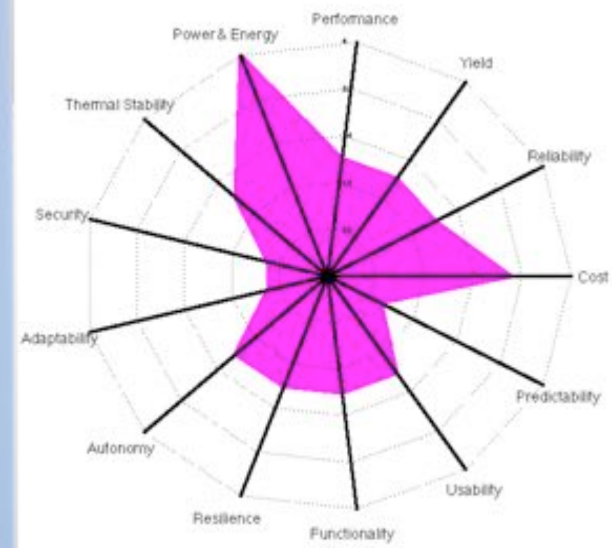


Security Driven

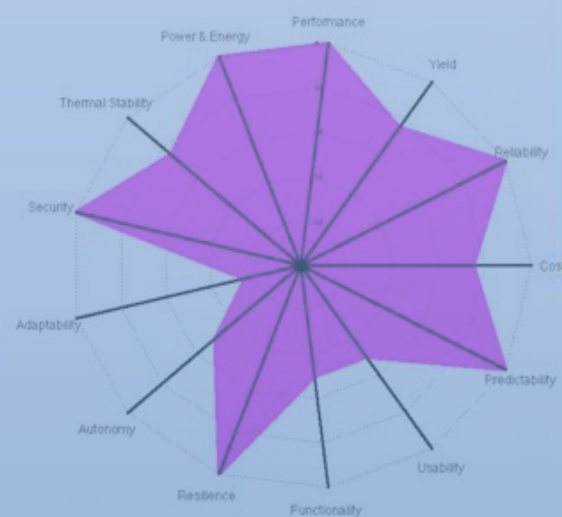
Reality



Performance Driven

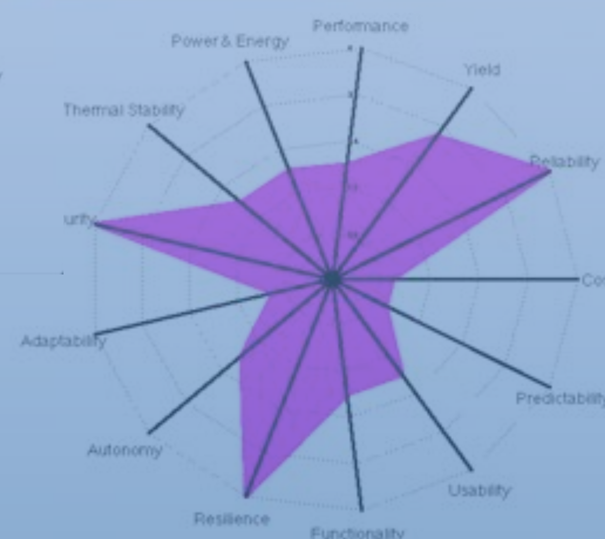


Energy/Power Driven



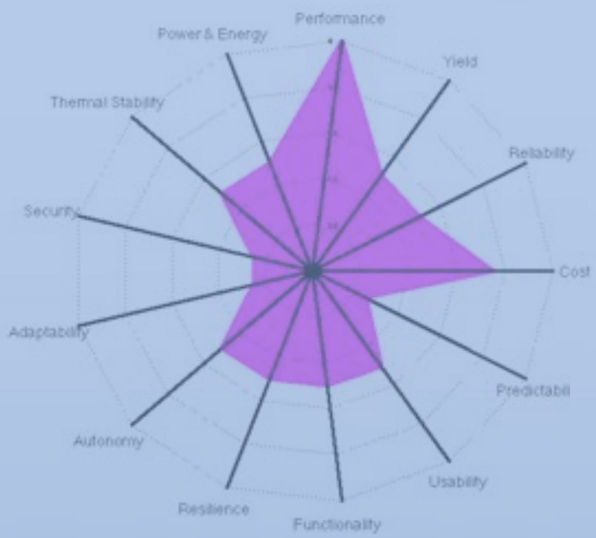
QoS Combination

Reliability Driven

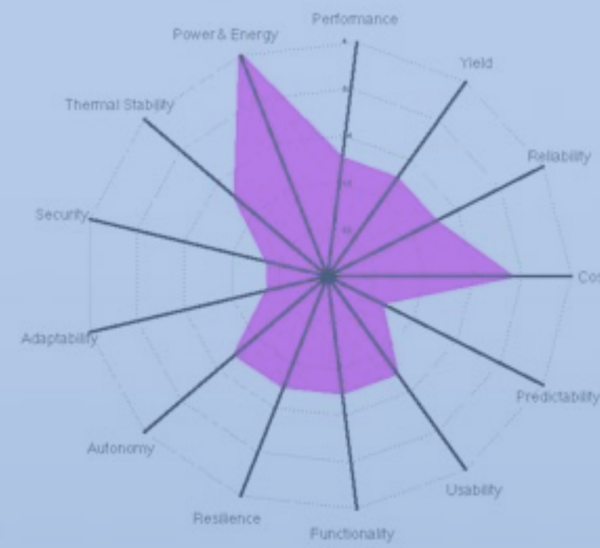


Security Driven

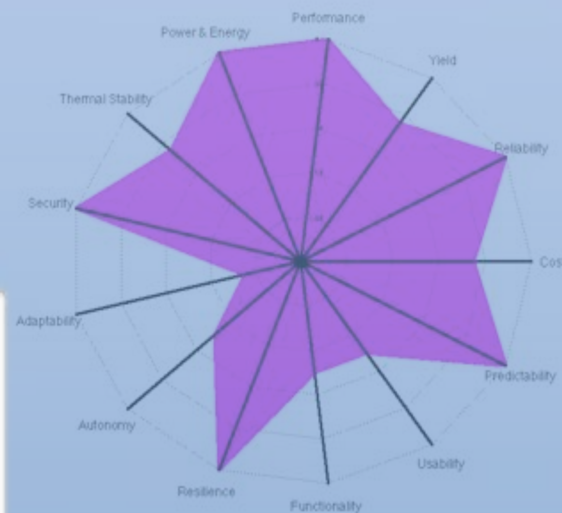
Reality



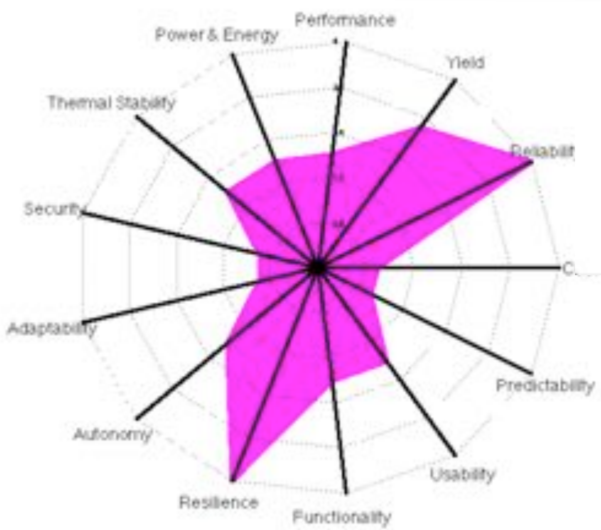
Performance Driven



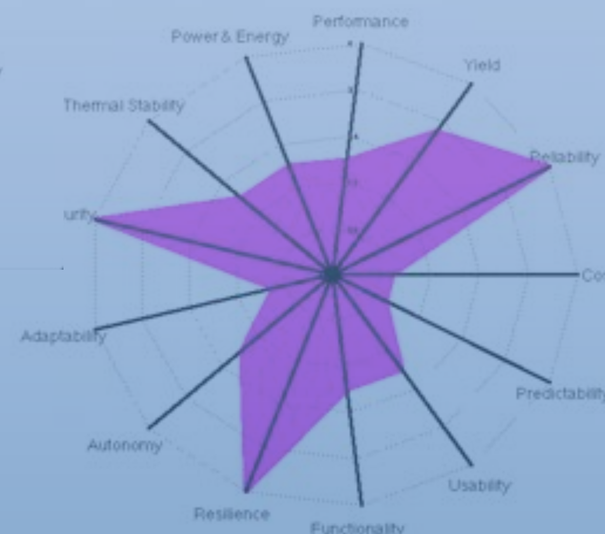
Energy/Power Driven



QoS Combination

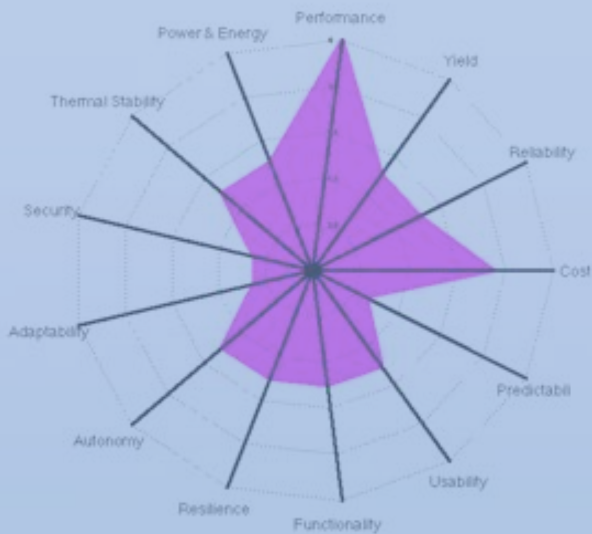


Reliability Driven

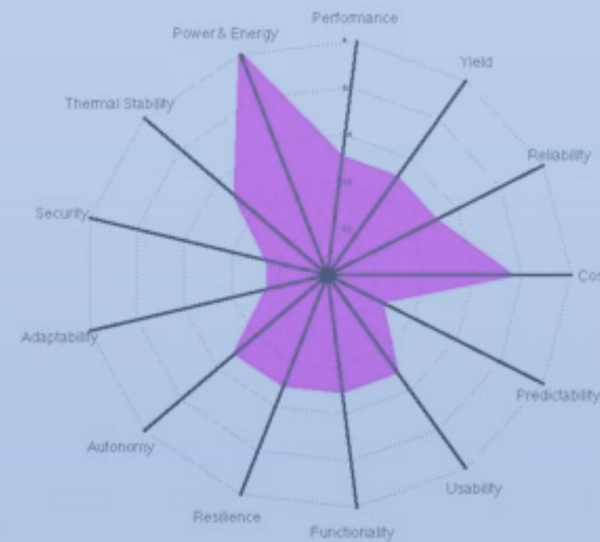


Security Driven

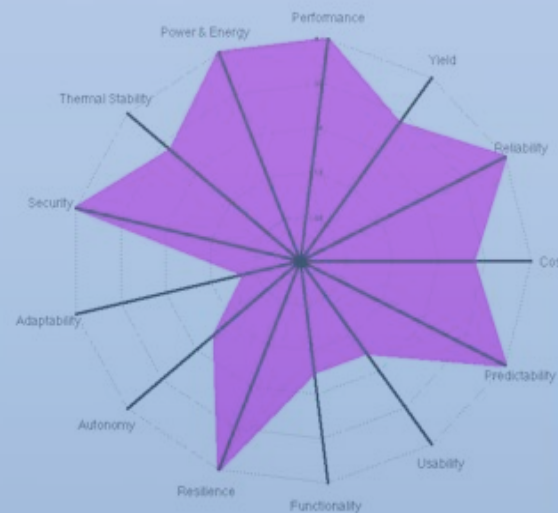
Reality



Performance Driven

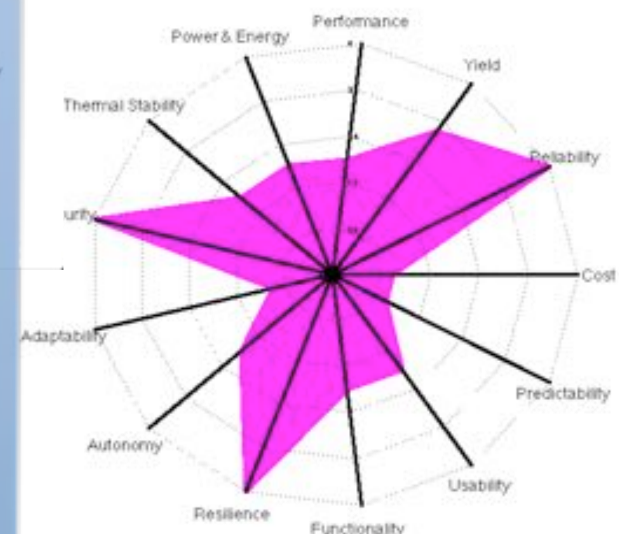


Energy/Power Driven



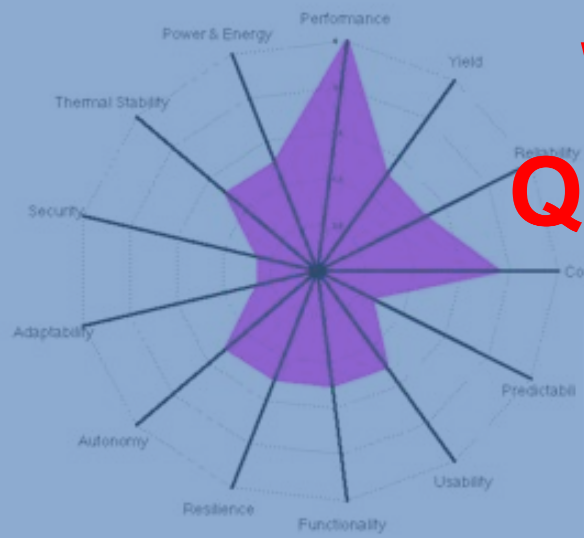
QoS Combination

Reliability Driven

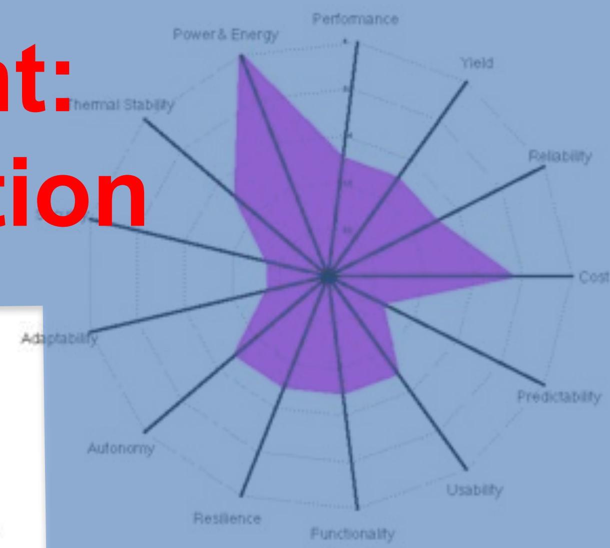


Security Driven

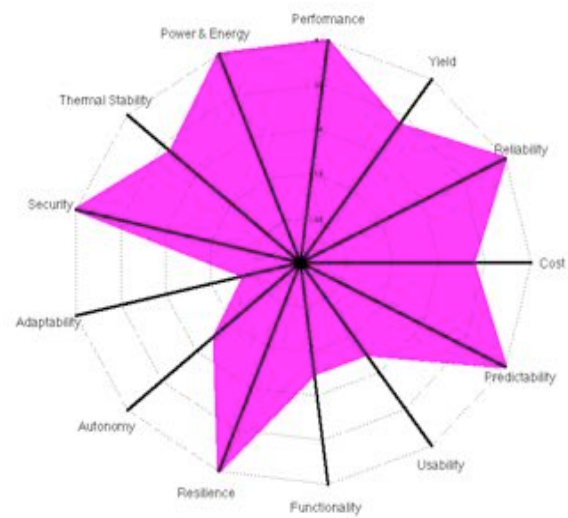
What we want: QoS Combination



Performance Driven



Energy/Power Driven

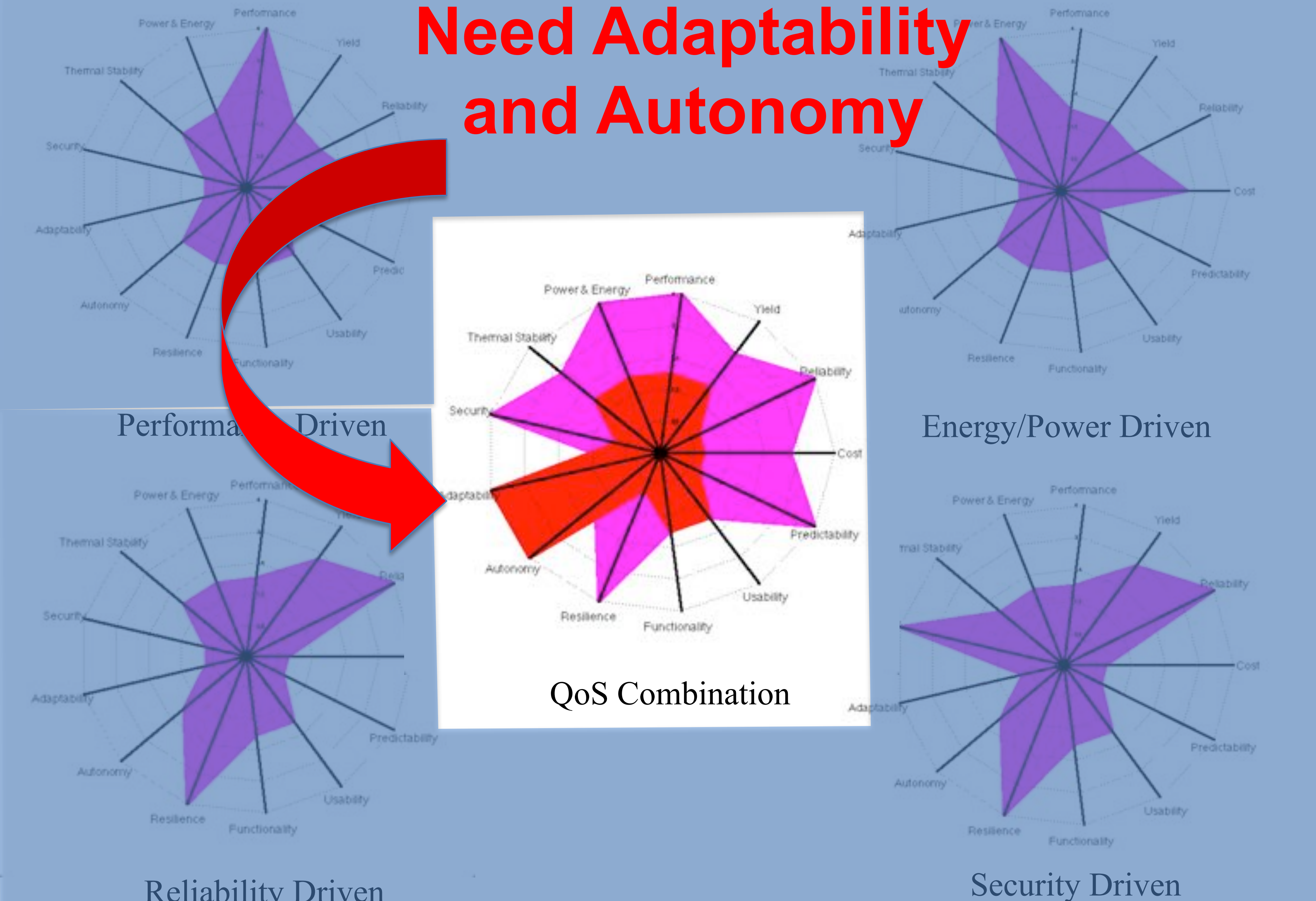


QoS Combination

Reliability Driven

Security Driven

Need Adaptability and Autonomy



Outline

- Computational Self-Awareness
- Exemplar: Cyber-Physical Systems-on-Chip (CPSoC)
- Managing Self-Awareness and Autonomy
- Wrap-up

Outline

- Computational Self-Awareness
- Exemplar: Cyber-Physical Systems-on-Chip (CPSoC)
- Managing Self-Awareness and Autonomy
- Wrap-up

What’s a Cyber-Physical System-on-Chip (CPSoC)?

- Cyber-Physical System
  - Cyber = Embedded Systems and IT
    - HW, SW, Networking/Communication
  - Physical = Environment
    - Sensing, Actuation, Control
- System-on-Chip (SoC):
  - Many-core computing platform: processors, memories, interconnects
- Cyber-Physical System-on-Chip (CPSoC):
  - Sensor-actuator rich adaptive SoC platform
    - Self-awareness
    - Predictive modeling and learning

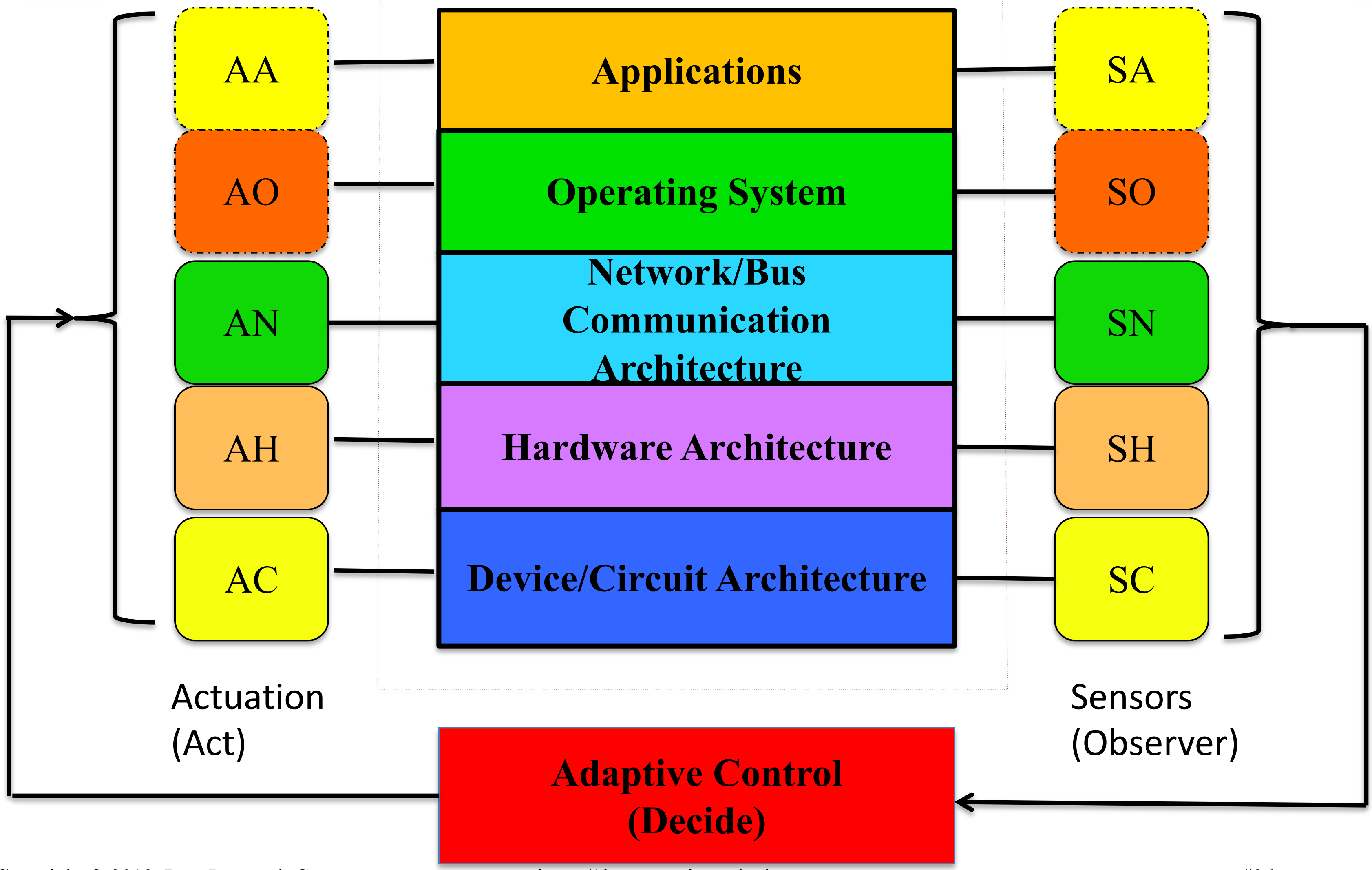
Cyber-Physical System-on-Chip (CPSoC)

- ***Cross-Layer Virtual and Physical Sensing & Actuation***
  - **Sensor fusion and Actuation**
    - Combine hardware and software sensors
- ***Self-Awareness and Adaptation***
  - Combines *Simple* and *Self-Aware* adaptions
  - A reflective (**Observe-Decide-Adapt**) architecture to achieve closed loop system control
- ***Predictive Modeling & Learning***
  - Dynamic characterization of platform variability across multiple levels of the system stack.

Cyber-Physical System-on-Chip (CPSoC)

- ***Cross-Layer Virtual and Physical Sensing & Actuation***
  - **Sensor fusion and Actuation**
    - Combine hardware and software sensors
- ***Self-Awareness and Adaptation***
  - Combines ***Simple*** and ***Self-Aware*** adaptions
  - A reflexive (**Observe-Decide-Adapt**) architecture to achieve closed loop system control
- ***Predictive Modeling & Learning***
  - Dynamic characterization of platform variability across multiple levels of the system stack.

Cross-Layer Physical/Virtual Sensing & Actuation



Examples of Virtual Sensors and Actuators Across Layers of CPSoC

Layers	**Virtual/Physical Sensors**	**Virtual/Physical Actuators**
Application	Execution Time, Workload Power, Energy,	Loop perforation Algorithmic Choice
Operating System	System Utilization Peripheral States	Task Allocation, Scheduling, Migration, Duty Cycling
Network/Bus Communication	Bandwidth; Packet/Flit status; Channel Status, Congestion, Latency	Adaptive Routing Dynamic Bandwidth Allocation Ch. no and direction
Hardware Architecture	Cache misses, Miss rate; access rate; IPC, Throughput, ILP/MLP, Core asymmetry	Cache Sizing; Reconfiguration, Resource Provision Static/Dynamic Redundancy
Circuit/Device	Circuit Delay, Aging, leakage Temperature, oxide breakdown	DVFS, DFS, DVS ABB, Clock and Power-gating

Cyber-Physical System-on-Chip (CPSoC)

- ***Cross-Layer Virtual and Physical Sensing & Actuation***
  - **Sensor fusion and Actuation**
    - Combine hardware and software sensors

- ***Self-Awareness and Adaptation***
  - Combines ***Simple*** and ***Self-Aware*** adaptions
  - A reflective (**Observe-Decide-Adapt**) architecture to achieve closed loop system control

- ***Predictive Modeling & Learning***
  - Dynamic characterization of platform, application, and environment variability across multiple system levels

What are Self-Aware Chips?

- Self-Aware chips
  - ***Construct model*** of behaviors and environment using sensor data
  - Achieve ***self-awareness*** through on-chip sensors and monitors
    - Experience phenomena
    - Aware of state and behavior
  - The ability to ***introspect***
  - ***Adapt behavior*** based on model of external and internal environment

Reflex vs Reflect

Reflexive, Reactive

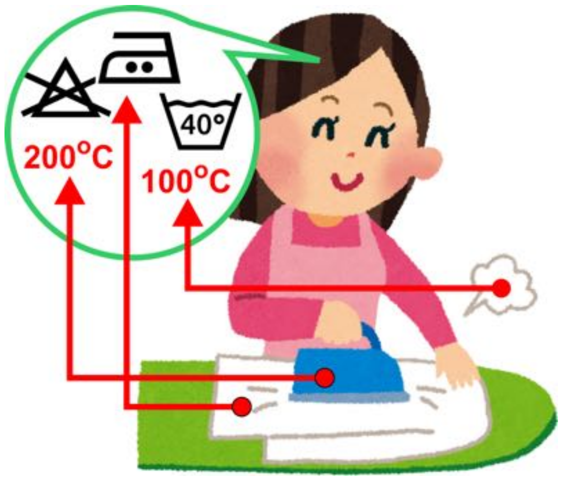
- **Actions driven solely on external feedback**
  - **E.g., our autonomic nervous systems**

Reflex vs Reflect

Reflexive, Reactive

- Actions driven solely on external feedback
  - E.g., our autonomic nervous systems

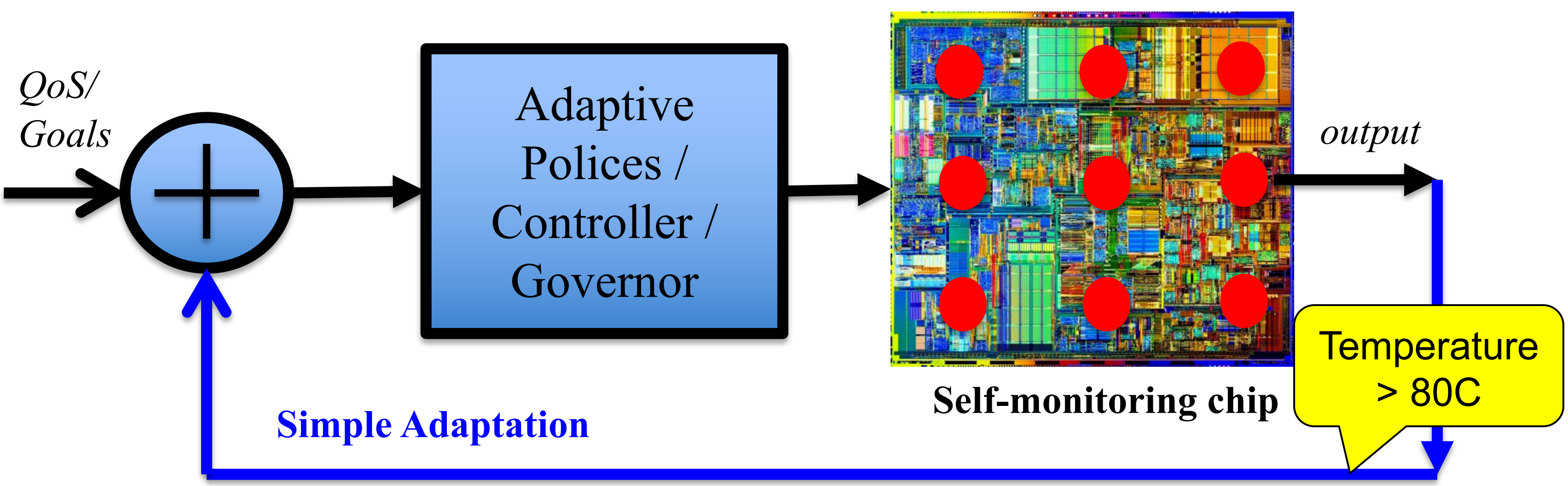
Reflection, Introspection



- Consider past and future outcomes
  - E.g., planning, strategies, policies, …

Towards Self-Aware Chips: What we do now

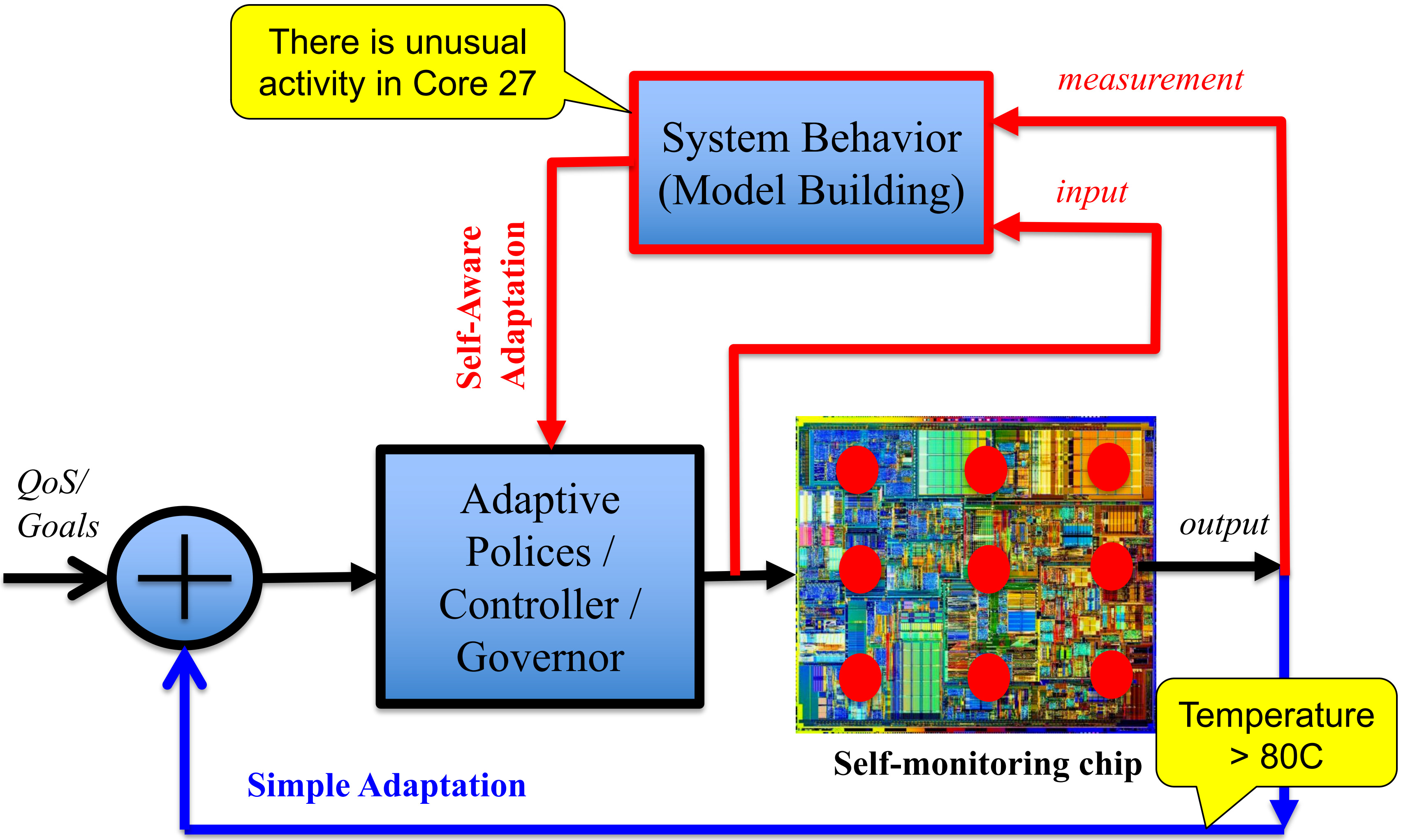
Reflexive, Reactive



[Sarma14, CODES+ISSS14]

Towards Self-aware chips

Beyond simple reactive models

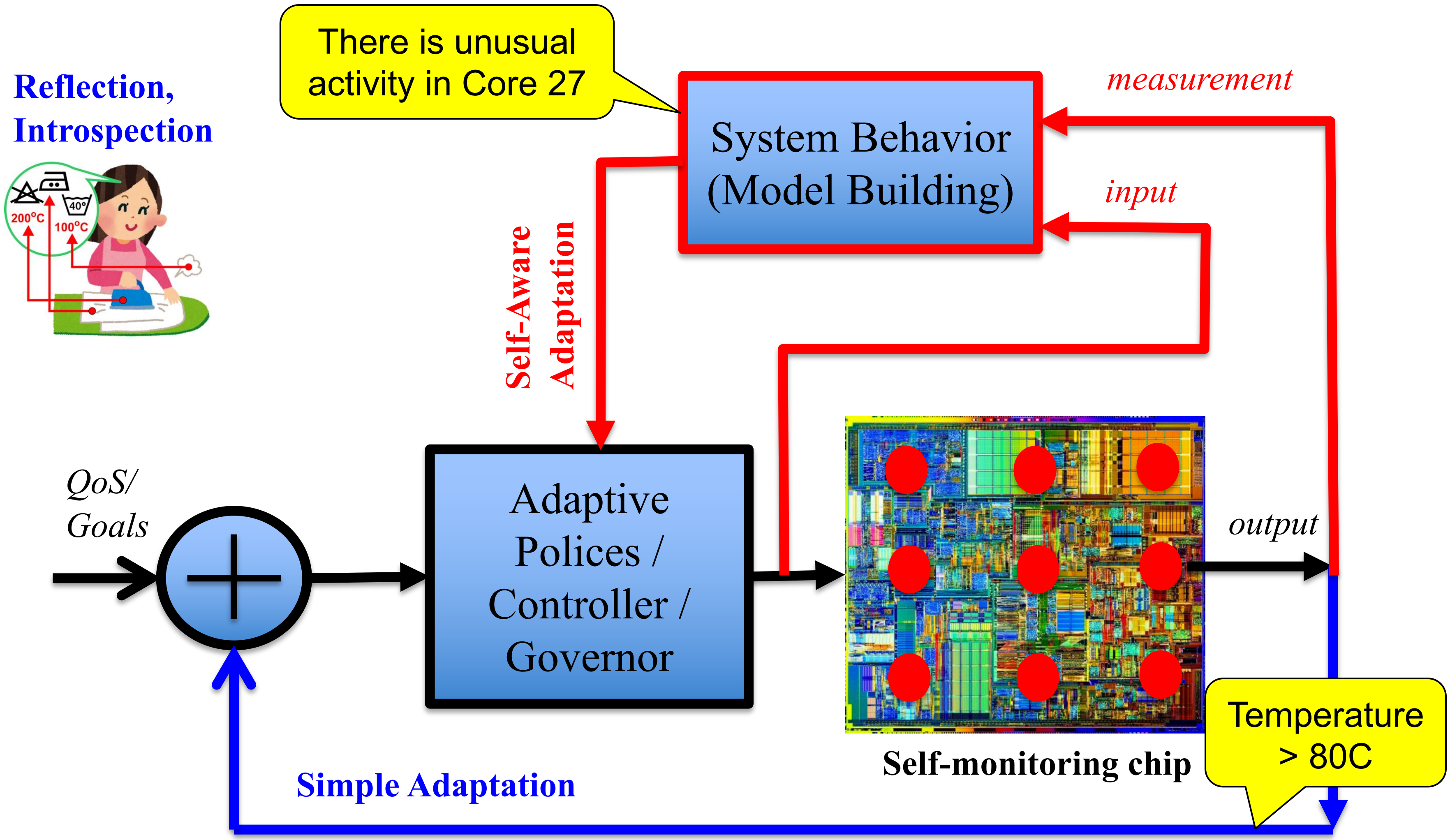


Self-monitoring and **Self-modeling**

[Sarma14, CODES+ISSS14]

Towards Self-aware chips

Beyond simple reactive models

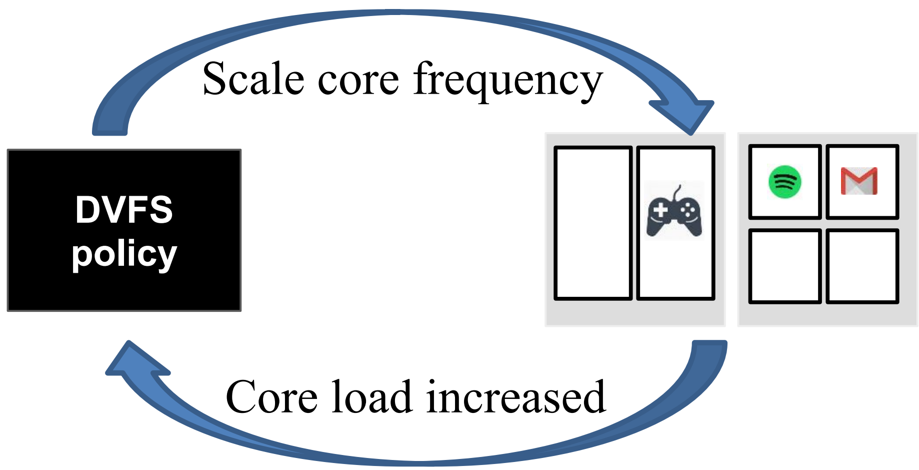


Self-monitoring and **Self-modeling**

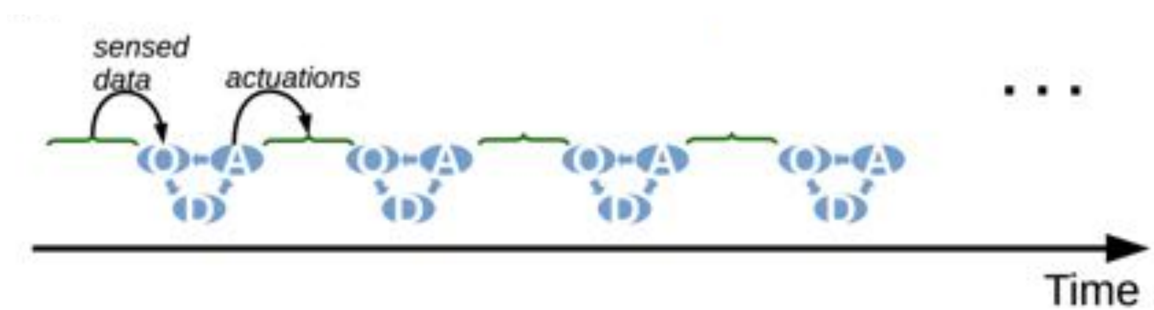
[Sarma14, CODES+ISSS14]

Today: “Reactive” Resource Management

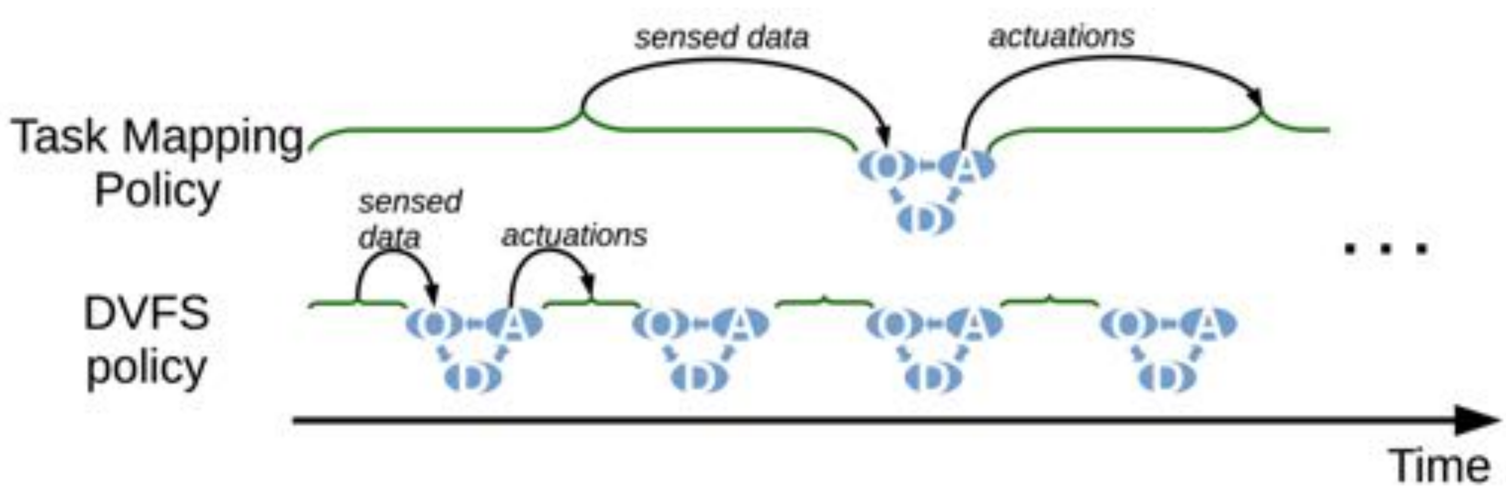
- Dynamic Voltage/Frequency scaling (DVFS)



- Observe-Decide-Adapt approaches

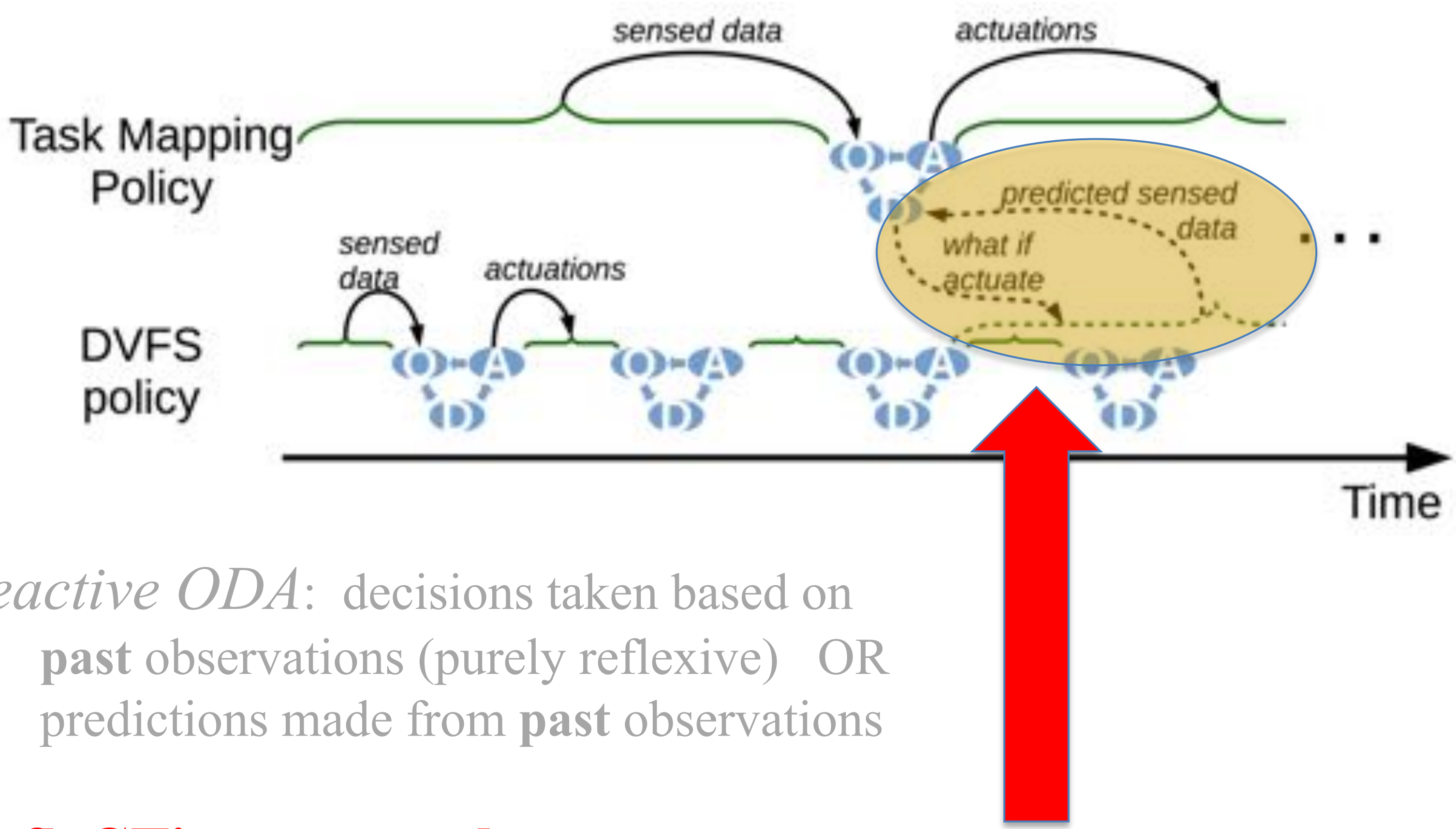


Reactive vs RefleCTive Resource Management



- *Reactive ODA*: decisions taken based on
  - **past** observations (purely reflexive) OR
  - predictions made from **past** observations

Reactive vs RefleCTive Resource Management



- *Reactive ODA*: decisions taken based on
  - **past** observations (purely reflexive) OR
  - predictions made from **past** observations

- **RefleCTive approach**: considers **future** events that could happen in the next iteration of the ODA loop

Adaptive Resource Management

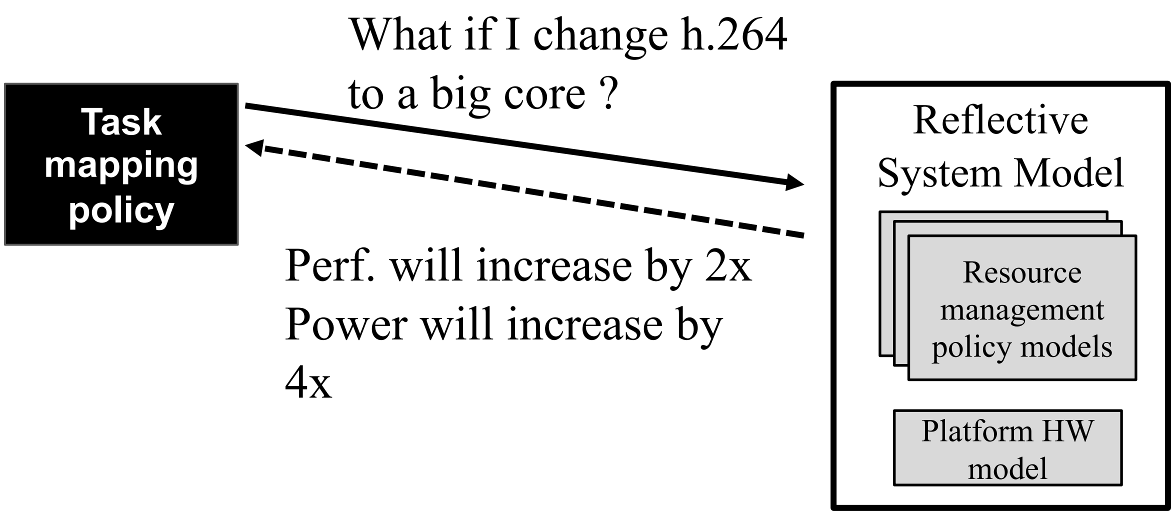
- Use concept of **reflection**
  - **Reflection**: *change your actions based on both external feedback and* ***introspection*** *(i.e., self-assessment)*

Adaptive Resource Management

- Use concept of **reflection**
  - Reflection: *change your actions based on both external feedback and* ***introspection*** *(i.e., self-assessment)*
- **Reflective resource management combines:**
  - Current system state assessed from sensing information (e.g., readings from performance counters, power sensors, etc.)
  - Models to predict the behavior of other system components before performing an action

MARS: Our coordination approach

- Coordination though **reflective** resource management
  - **MARS: M**iddleware for **A**daptive **R**eflective **S**ystems



Do we have room for reflection ?

- Systems actuations happen at different timescales

Task mapping

Which core executes a task?
(100's – 10's ms)

DVFS

What's the next frequency?
(10's ms – 10's us)

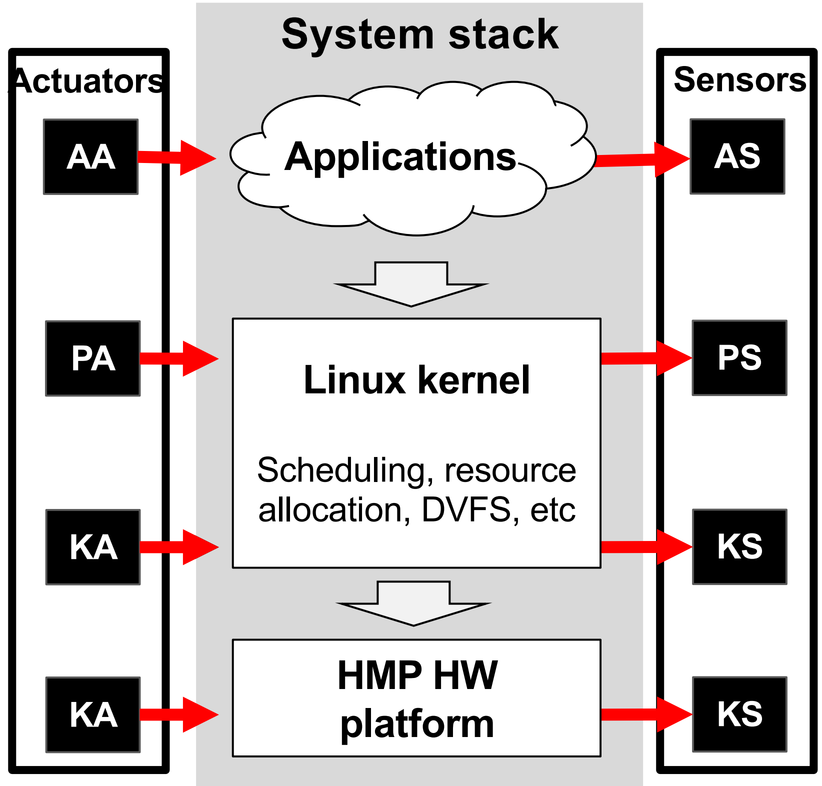
Scheduling

Should I preempt a task ?
(<1 us)

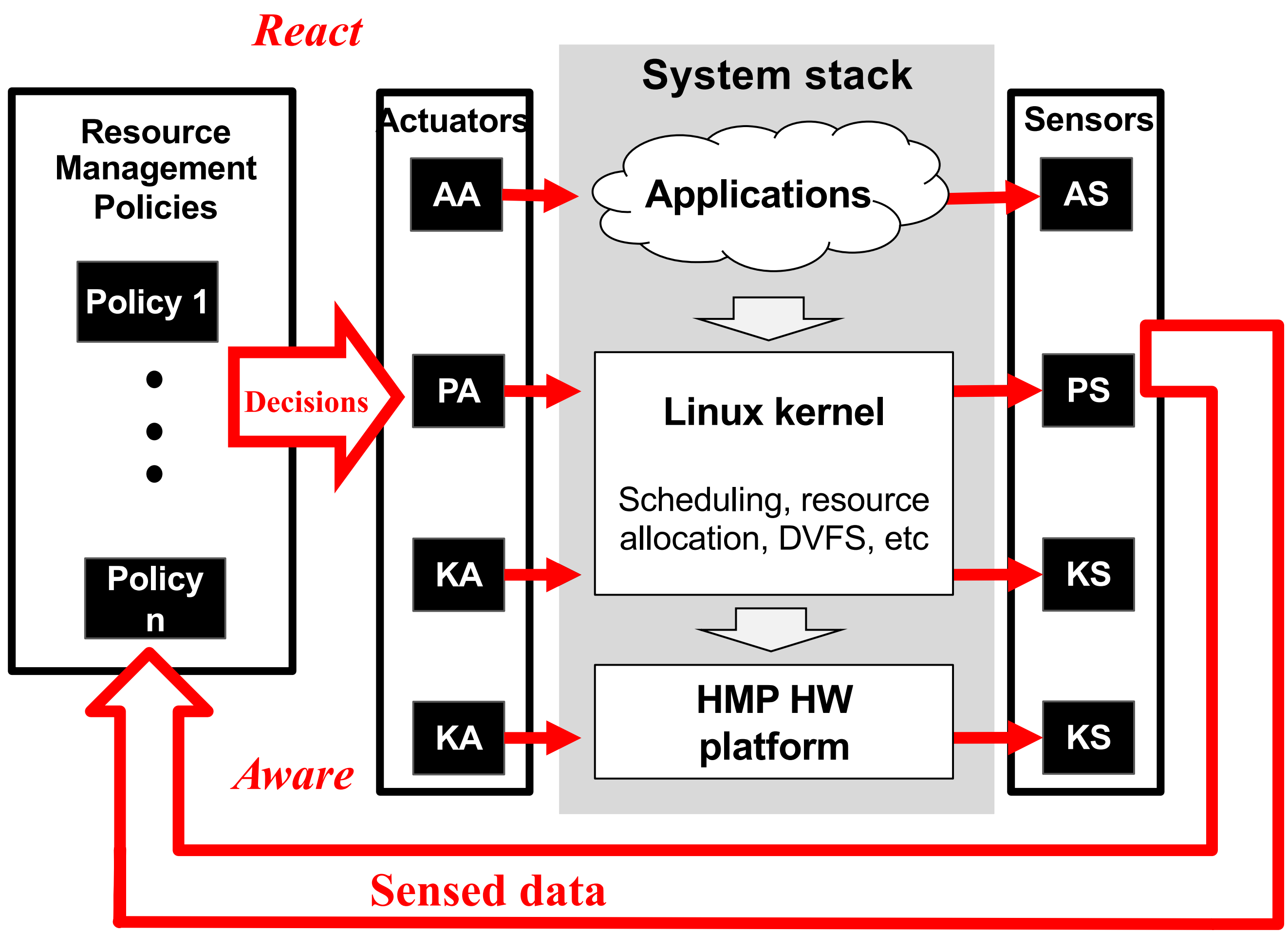
Coarser grained actuation ⟷ ***Finer grained actuation***

- Some actuations happen quickly with little room for reasoning
- Other actuations can occur on larger timescales
  - **Task mapping, Wear-leveling (for aging)….**

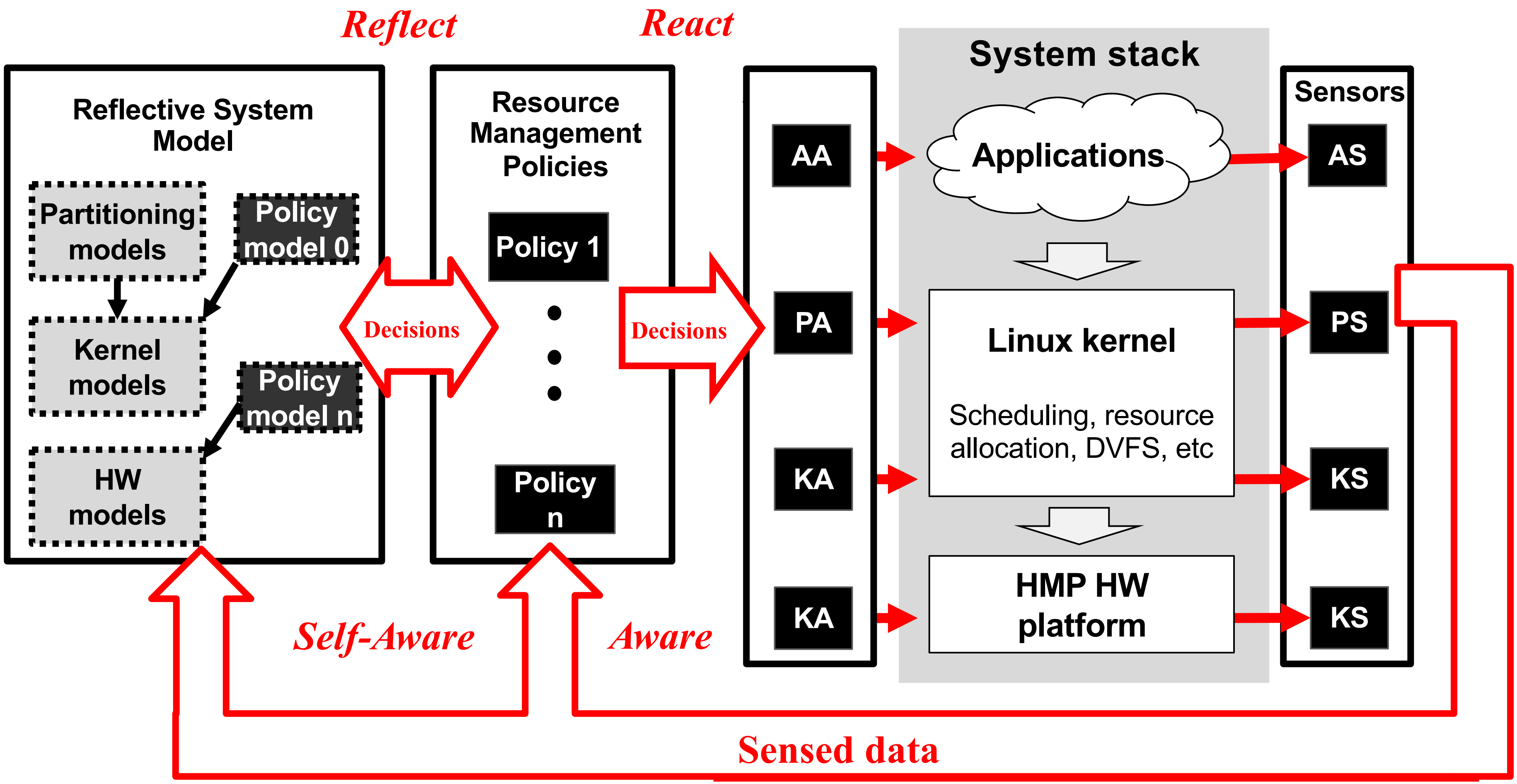
MARS middleware for reflective resource management



MARS middleware for reflective resource management



MARS middleware for reflective resource management



MARS: Middleware for Adaptive Reflective Computer Systems

- Framework and tools for developing reflective resource/power management policies
  - Use models to predict system behavior
  - Enable easy adaptation to runtime changes
  - Case studies show promise

MARS framework is open source!

https://duttgroup.ics.uci.edu/projects/self-aware-adaptive-computing/mars/

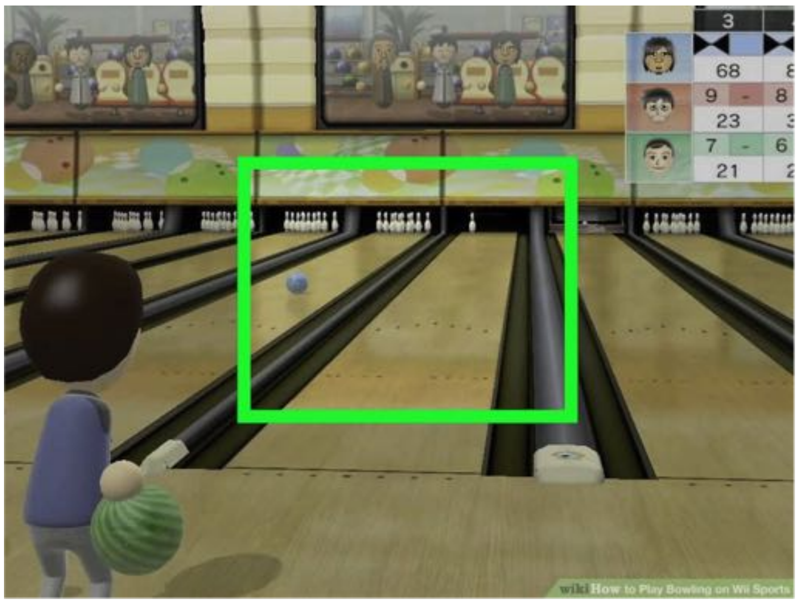
https://github.com/duttresearchgroup/MARS

Outline

- Computational Self-Awareness
- Exemplar:Cyber-Physical Systems-on-Chip (CPSoC)
- Managing Self-Awareness and Autonomy
- Wrap-up

Goals and Autonomy

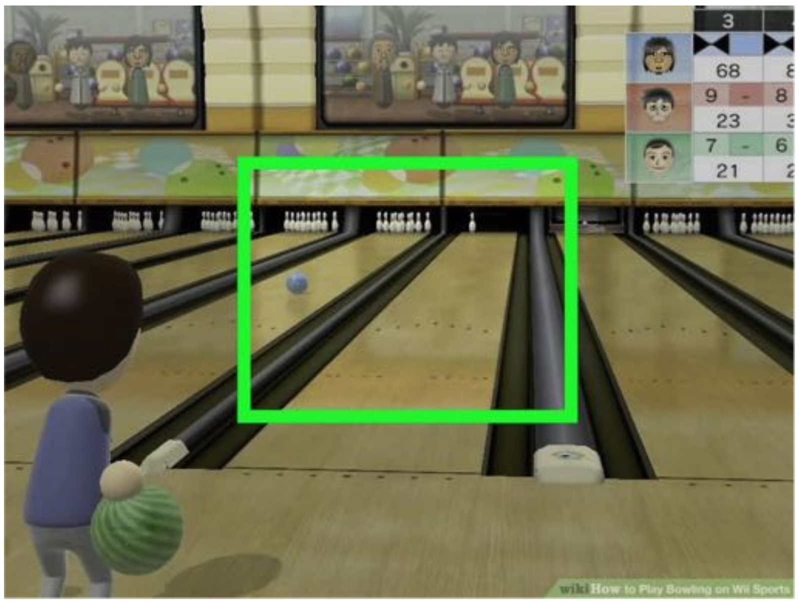
Goal



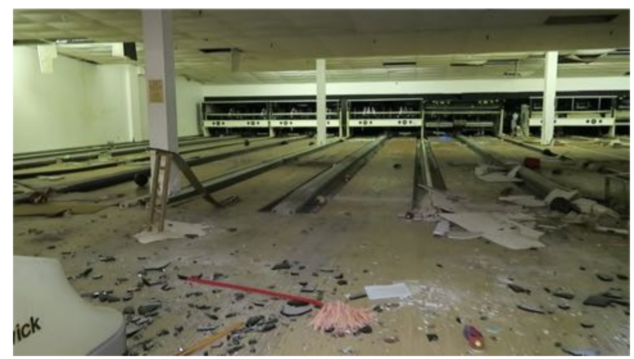
- **Single, straightforward objective**
  - **E.g., hit the pin**

Goals and Autonomy

Goal



Model Imperfection



- Single, straightforward objective
  - E.g., hit the pin

- What happens when we introduce unpredictability?
  - E.g., balls with different sizes, shapes weights; uneven or damaged surfaces

Goals and Autonomy

Supervision

- **Constrain behavior so we are always headed toward the goal**
  - **E.g., bumpers**

Goals and Autonomy

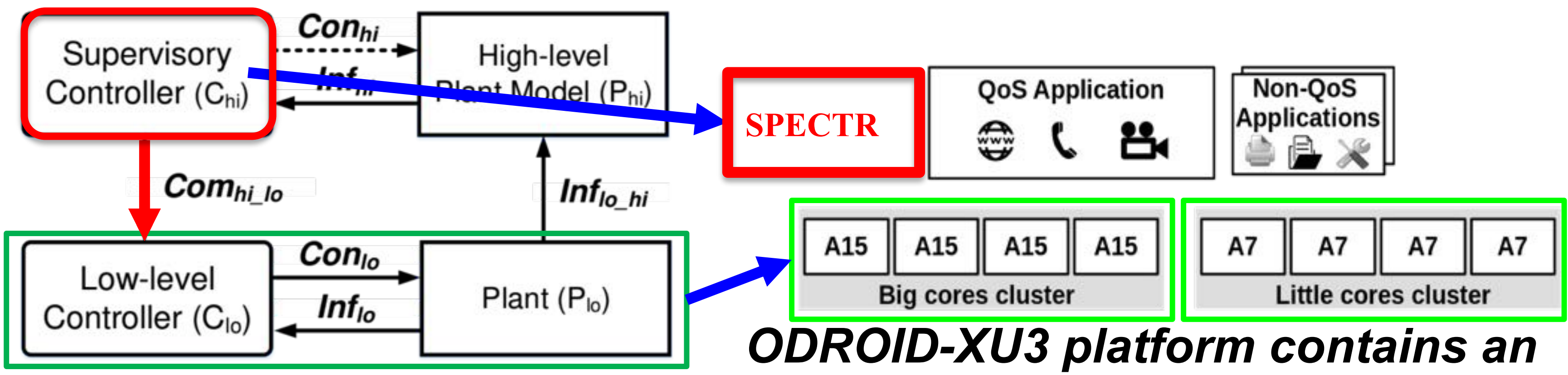
Supervision

- **Constrain behavior so we are always headed toward the goal**
  - **E.g., bumpers**
- **Bonus: what about when we have more complex or multiple goals?**

SPECTR: Our Supervisory Approach

- Autonomy and robustness through supervisory control
- Case Study (ASPLOS 2018 Paper)*

Supervisor bounds behavior of controllers, manages goal



ODROID-XU3 platform contains an Exynos 5422 Octa-core SoC

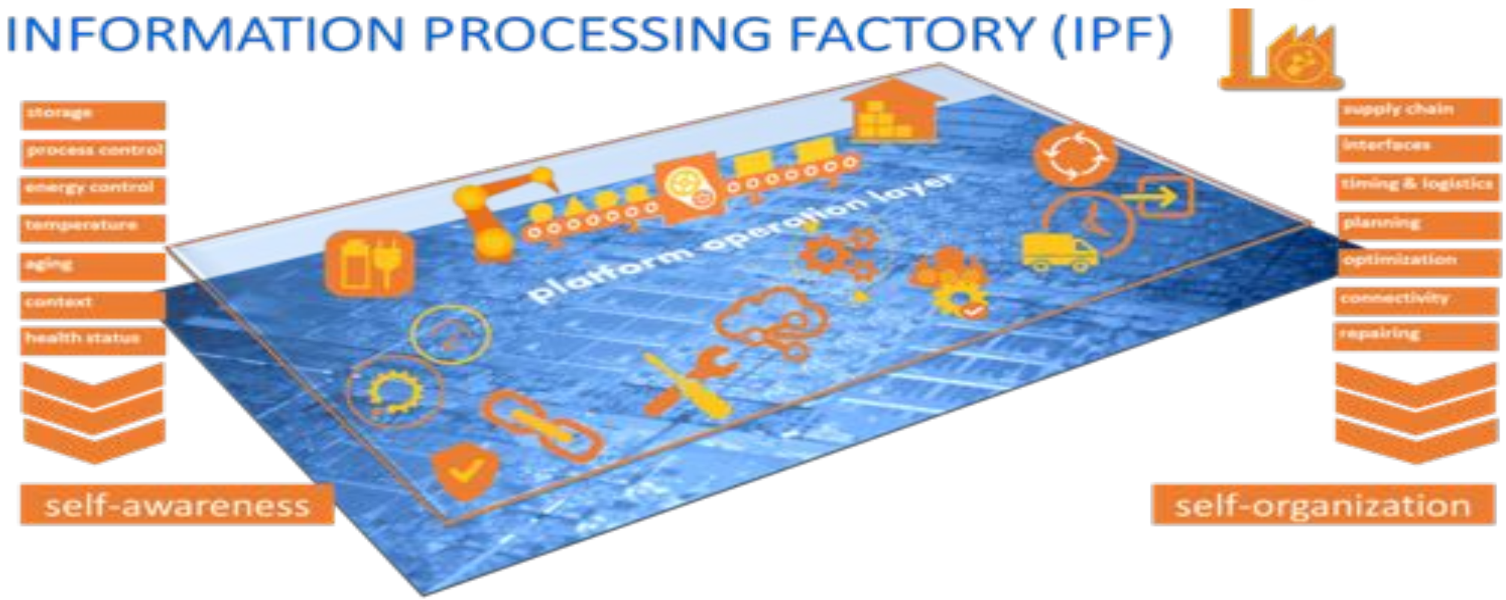
Low-level controllers satisfy objective

*Rahmani, A. M., Donyanavard, B., Mück, T., Moazzemi, K., Jantsch, A., Mutlu, O., & Dutt, N., *SPECTR: Formal Supervisory Control and Coordination for Many-core Systems Resource Management*. ASPLOS '18

Self-* Computing: Related Efforts

- Self-test, Self-diagnosis, Self-Repair for chips and computing platforms (several decades)
- Reflective SW and Autonomic Computing (IBM) (circa 2003)
- ODA-based feedback control strategies for adaptive computing (circa 2011)
  - SEEC (MIT/ Milano)
  - Organic/Invasive Computing Germany)
    - *Software centric/focused adaptation of computing platform resources*
- **Missing pieces:**
  - **Cross-layer sensing/actuation**
  - **Self-modeling and introspection**
  - **Control/coordination of emergent behaviors**

IPF: Self-Aware / Self-Organizing Information Processing Factory

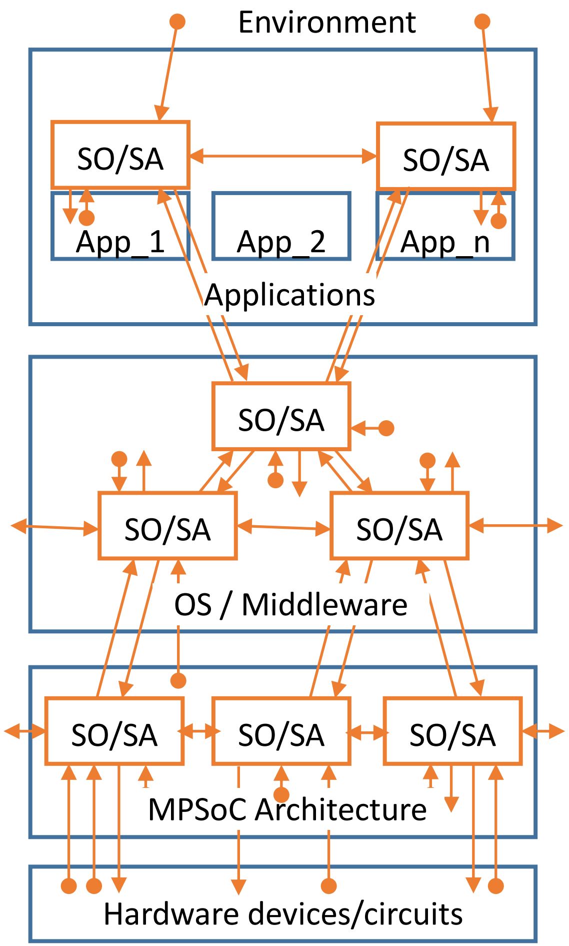


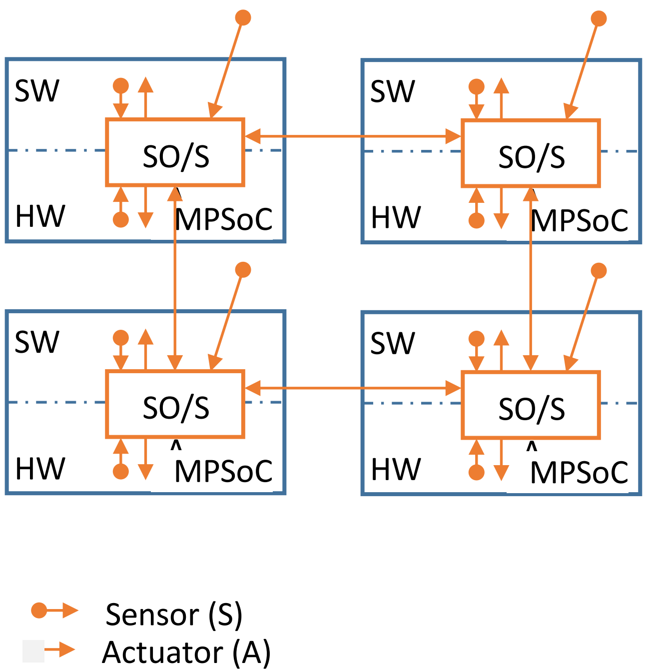
- Scalability of control through higher sub-system autonomy …
- … while operating within a corridor of guaranteed objectives and operation parameters

Objectives of a Self-Aware Information Processing Factory

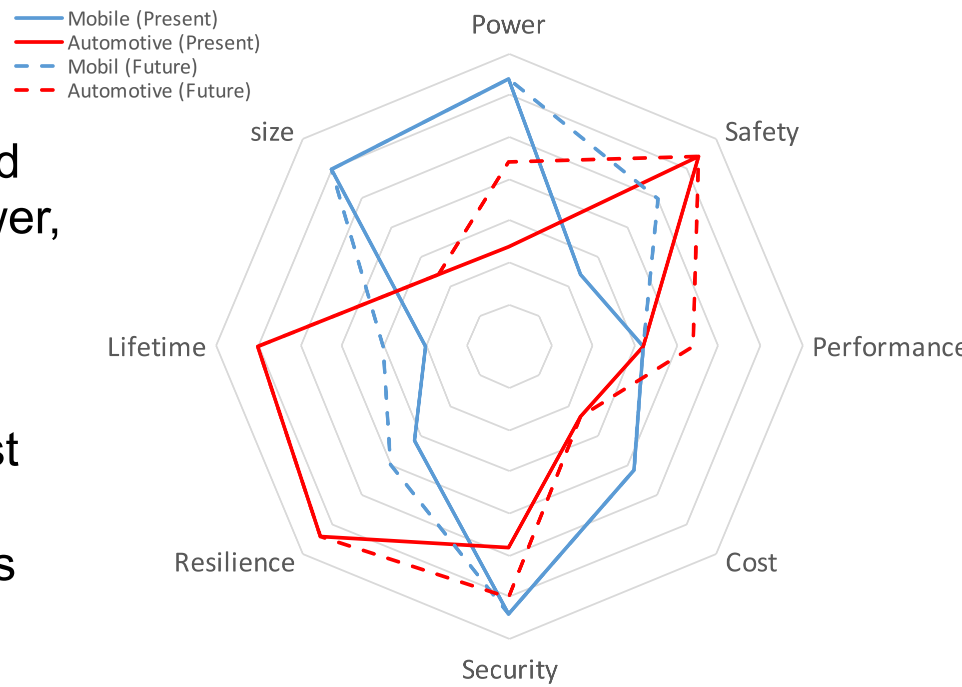
- **Holistic, hierarchically structured platform control**
  - As modular and decentralized as possible, …
  - … as much centralized as necessary
  - Scalability and dynamic elasticity
    - react to short/long term workload fluctuations
    - platform uncertainties
- **Orthogonal functional and control complexity**
  - Inspired by self-organization and emergent phenomena in nature
    - Self-awareness: introspection + ubiquitous sensing of platform and environment
    - Self-organization: local actions
- **Predictive guarantees for critical tasks / functions**
  - Long term planning and logistics

Self-Aware / Self-Organizing Information Processing Factory





Comparing Use Case Requirements



Automated Driving

- Safety, lifetime and resilience dominated
- Foreseeable future challenges with power, security and performance

Mobile Devices

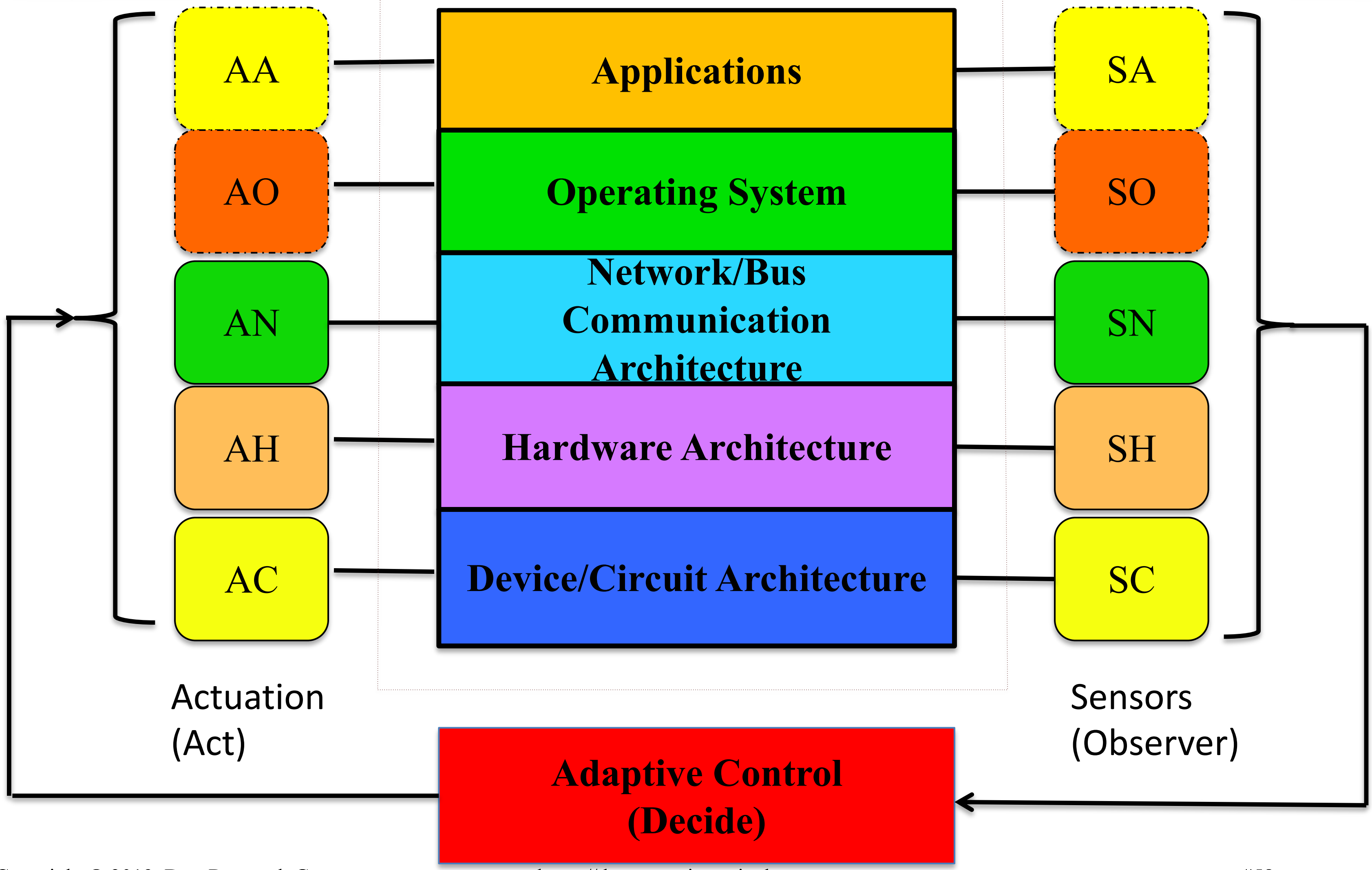
- Performance / W, size, security and cost are main challenges
- Safety in context of medical applications expected to rise

Trend: System evolution and convergence (SmartPhones as integrated devices in future cars) will lead to requirements convergence

Outline

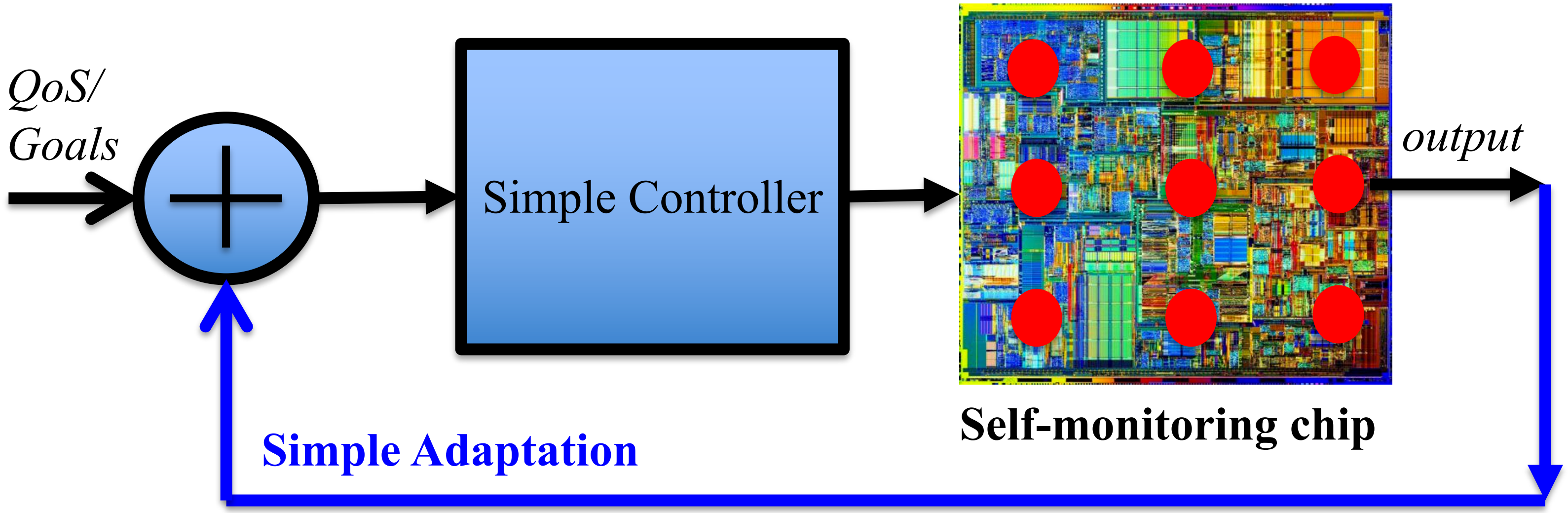
- Computational Self-Awareness
- Exemplar: Cyber-Physical Systems-on-Chip (CPSoC)
- Managing Self-Awareness and Autonomy
- Wrap-up

Key Take-Away 1:
Cross-Layer Physical/Virtual Sensing & Actuation



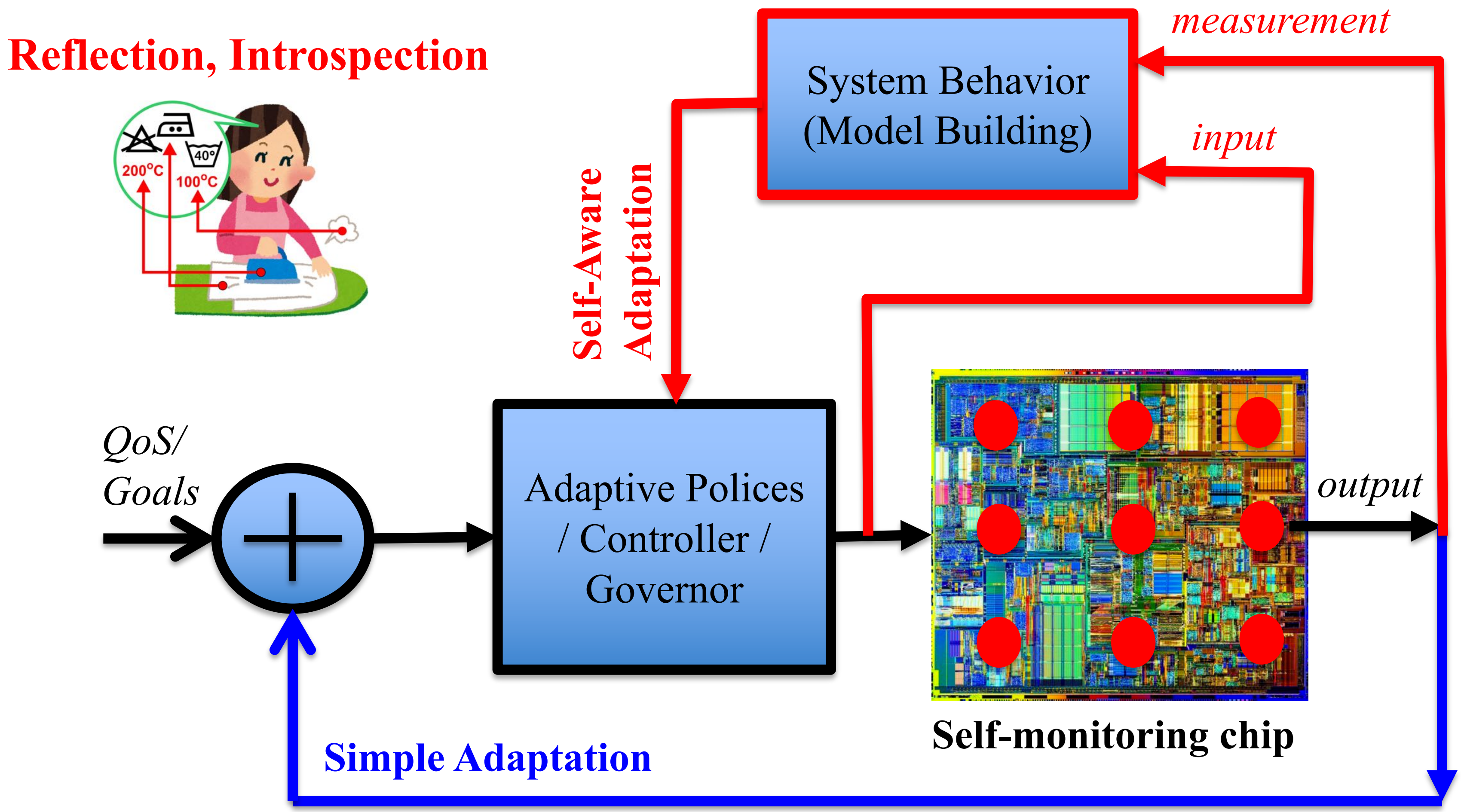
From today's chips

Reflexive, Reactive



Self-monitoring and **simple adaptation**

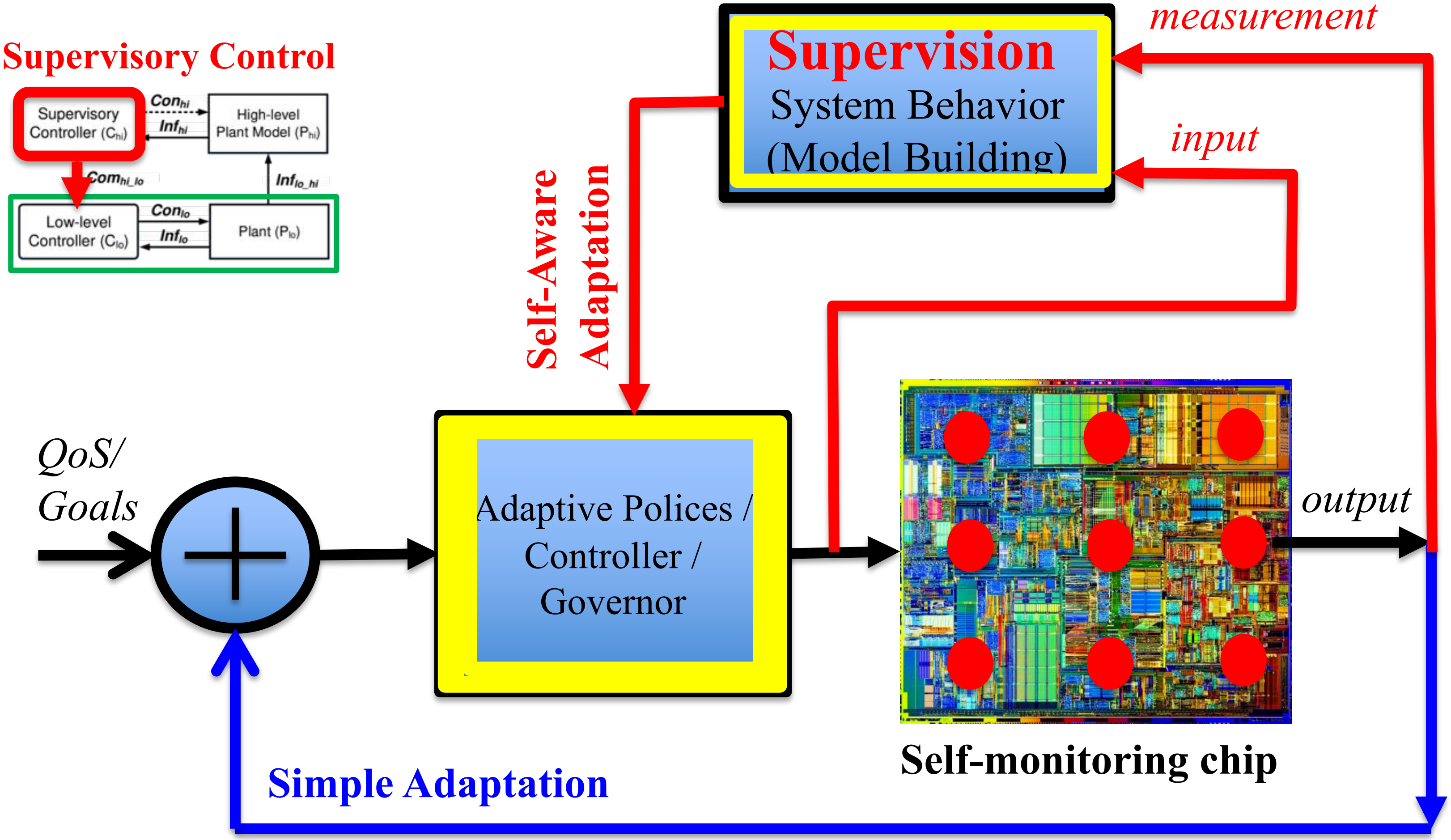
Key Take-Away 2:
Towards on-chip self-awareness



Self-monitoring and **Self-modeling**

[Sarma14, CODES+ISSS14]

Key Take-Away 3:
Supervisory Control & Coordination



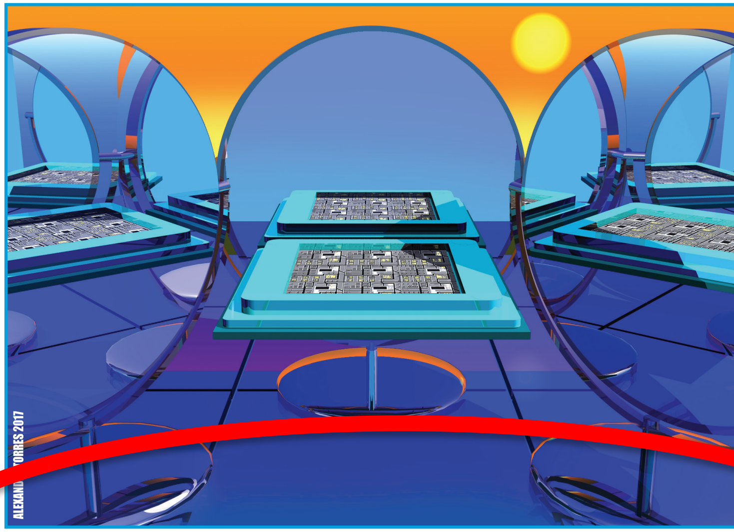
Supervisory Control & Coordination

[Rahmani18, ASPLOS 2018]

IEEE DESIGN & TEST November/December 2017 Special Issue on Self-Awareness in Systems on Chip 2017 VOLUME 34 NUMBER 6

NOVEMBER/DECEMBER 2017

IEEE Design&Test



Special Issue on Self-Awareness in Systems on Chip 2017

· Self-Awareness in Systems on Chip—A Survey
· Health Management for Self-Aware SoCs Based on IEEE 1687 Infrastructure
· KOCL: Power Self-Awareness for Arbitrary FPGA-SoC-Accelerated OpenCL Applications
· A Self-Aware Architecture for PVT Compensation and Power Nap in Near-Threshold Processors
· Self-Adaptive Timing Repair

CEDA IEEE Council on Electronic Design Automation · CAS IEEE Circuits and Systems Society · SSCS · tttc · IEEE

Contents

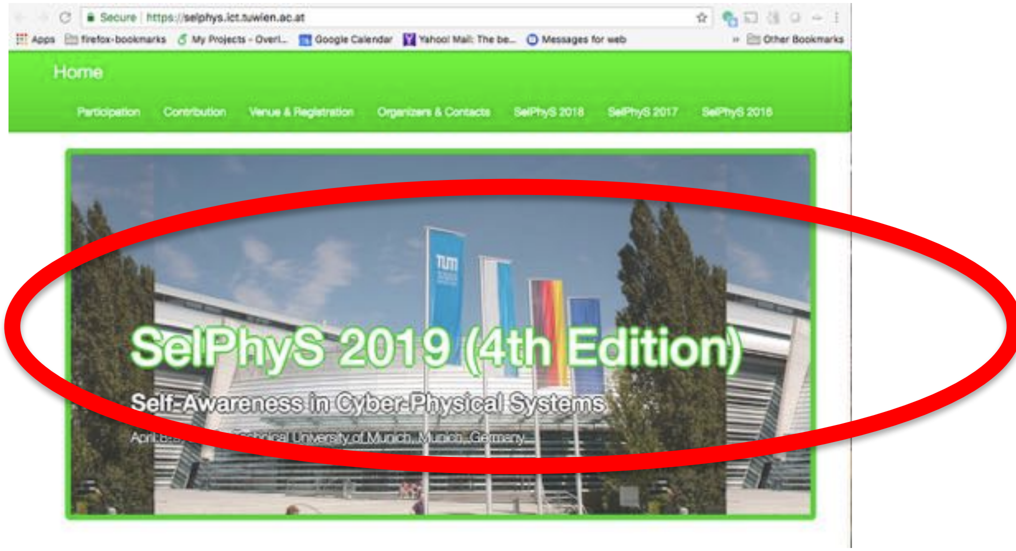
November/December 2017
Volume 34 Number 6

Copublished by the IEEE Council on Electronic Design Automation, the IEEE Circuits and Systems Society, the IEEE Solid-State Circuits Society, and the Test Technology Technical Council

ISSN: 2168-2356

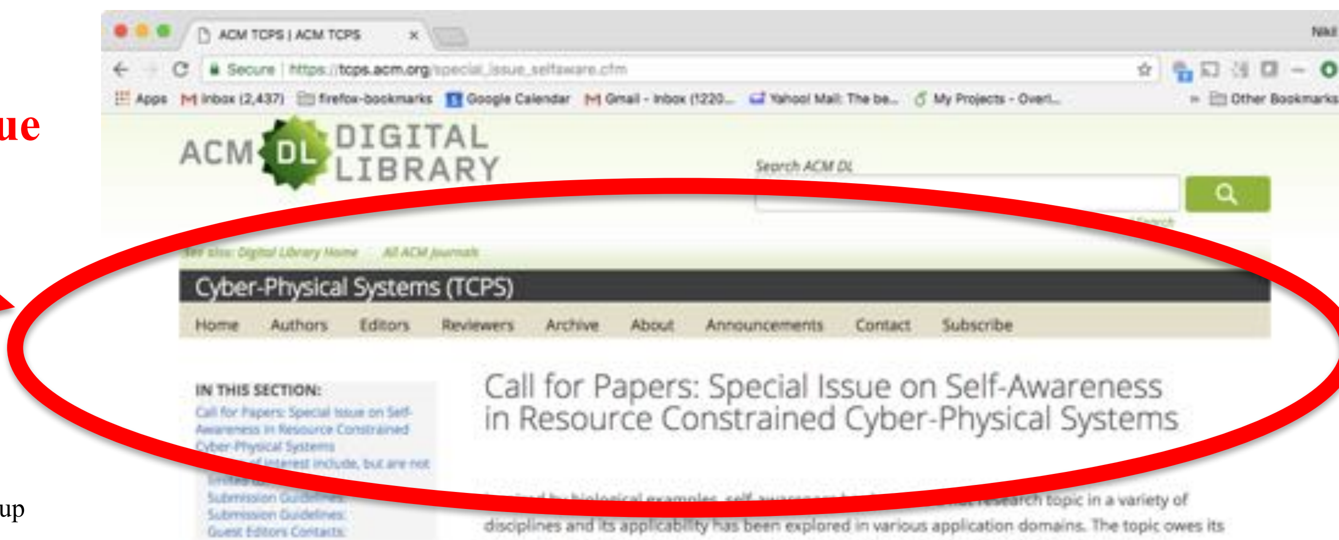
UCI CECS

SelPhyS 2019 and ACM TCPS Special Issue



April 8-9 2019
@ TU Munich

ACM TCPS Special Issue
on Self-Awareness in
CPS



Ongoing Challenges

- Self-trained models
  - Add feedback for error correction
  - Challenging for models that are non-linear and/or based on heuristics
- Machine learning
  - Replacement for analytical/heuristic-based models ?
  - Unsupervised machine learning to mine sensing data and find patterns for optimizing policies or creating new ones
- Policy supervisors
  - Provide formal or stronger guarantees

Acknowledgements

- Dutt Research Group (IPF Team)
  - Bryan Donyanavard, Tiago Mück, Amir Rahmani, Santanu Sarma, Kasra Moazzemi, Biswadip Maity, Saehanseul Yi, Caio Batista de Melo, Kenneth Stewart
- Collaborating Faculty:
  - UCI: Fadi Kurdahi
  - Germany: TUM: Prof. Andreas Herkersdorf, TUB: Prof. Rolf Ernst
  - Austria: TU Wien: Prof. Axel Jantsch
- NSF/DFG Information Processing Factory (IPF) project

Questions?

Dutt Research Group: http://duttgroup.ics.uci.edu/