Self-Aware Silicon

Arman Anzanpour [3], *Iman Azimi* [3], *Nikil Dutt* [2], *Maximilian Götzinger* [3],
Axel Jantsch [1], *David Juhasz* [1], *Hedyeh Kholerdi* [1], *Pasi Liljeberg* [3],
Amir Rahmani [1,2,3], *Nima Taherinejad* [1]

[1] TU Wien, Vienna, Austria
[2] UC Irvine, California
[3] University of Turku, Finland

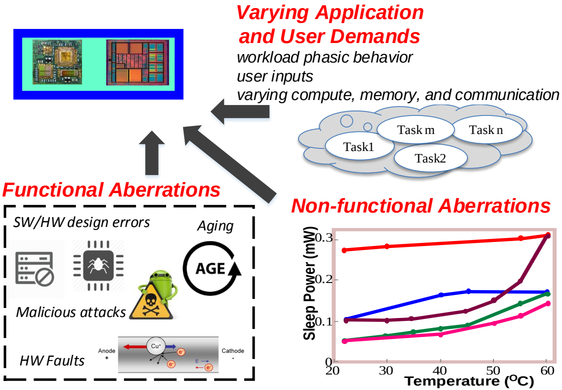
Chips on the Sands
Fortaleza, Brazil

28 August - 1 September 2017

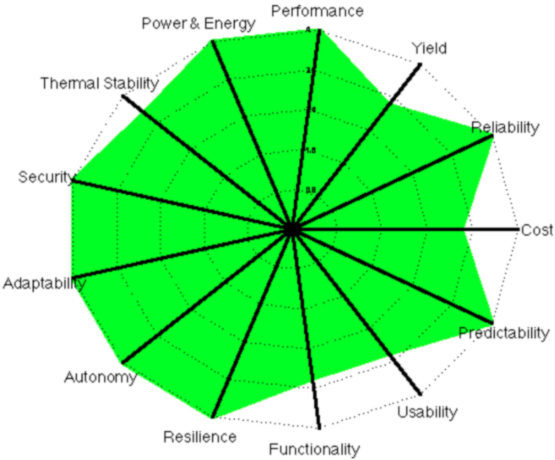
Outline

The Problem

- Large number of resources
- Many tight constraints
- Varying application demands, both within and between applications;
- Functional Aberrations:
  - Design errors or omissions;
  - Malicious attacks;
  - Aging;
  - Soft errors;
- Non-functional Aberrations:
  - Performance;
  - Power consumption;

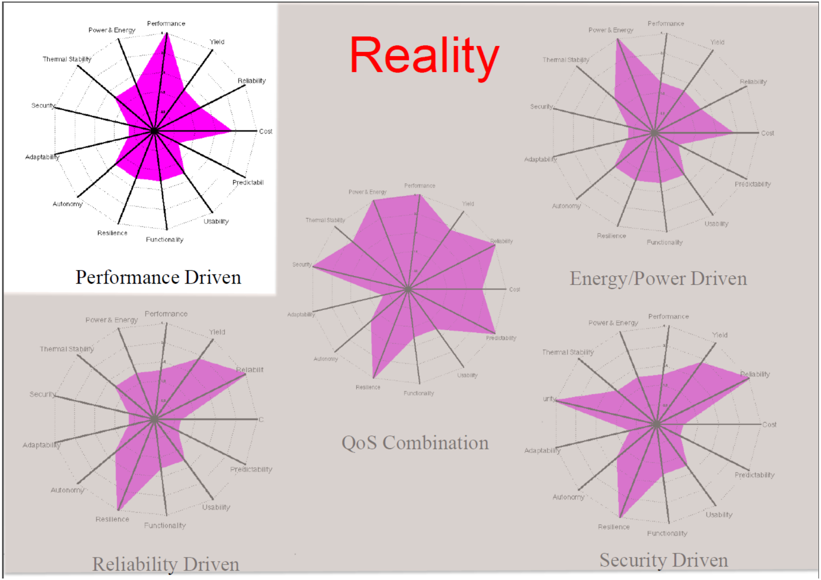


The SoC Radar

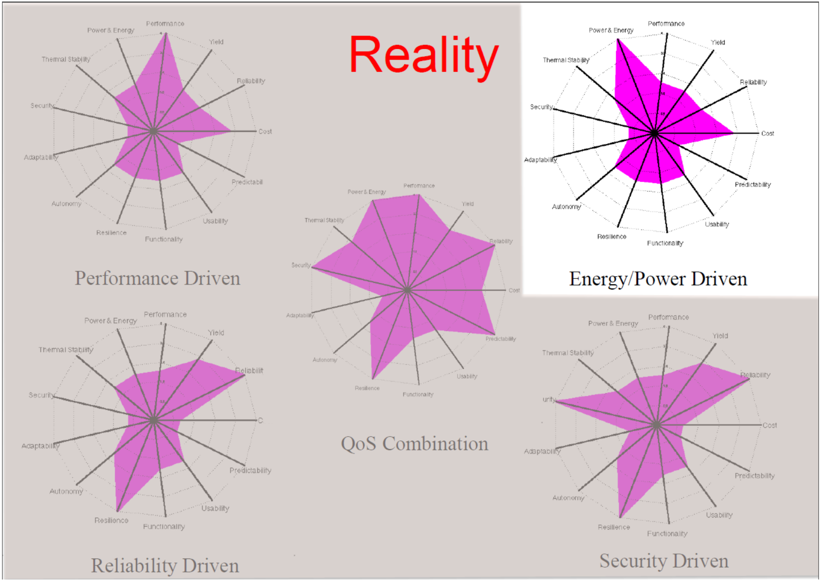


Santanu Sarma et al. “On-Chip Self-Awareness Using Cyberphysical-Systems-On-Chip (CPSoC)”. . In: *Proceedings of the 12th International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*. New Delhi, India, Oct. 2014

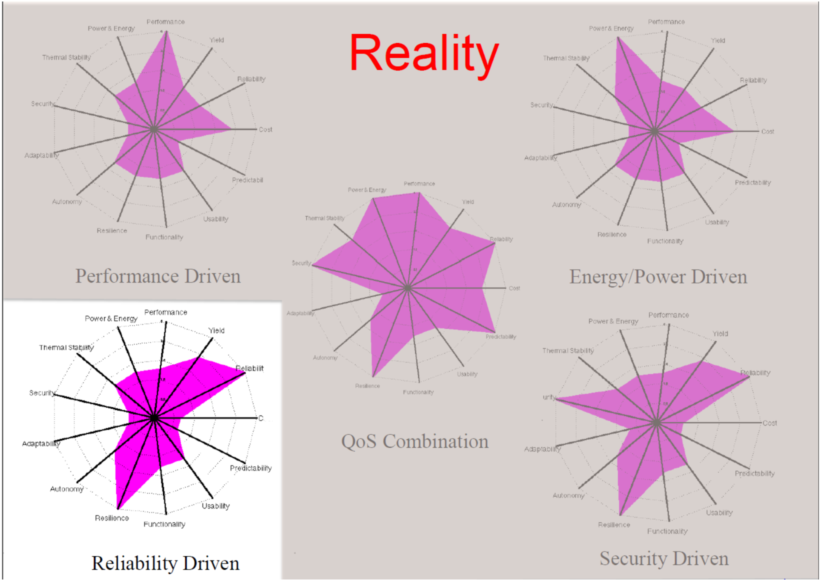
The SoC Radar



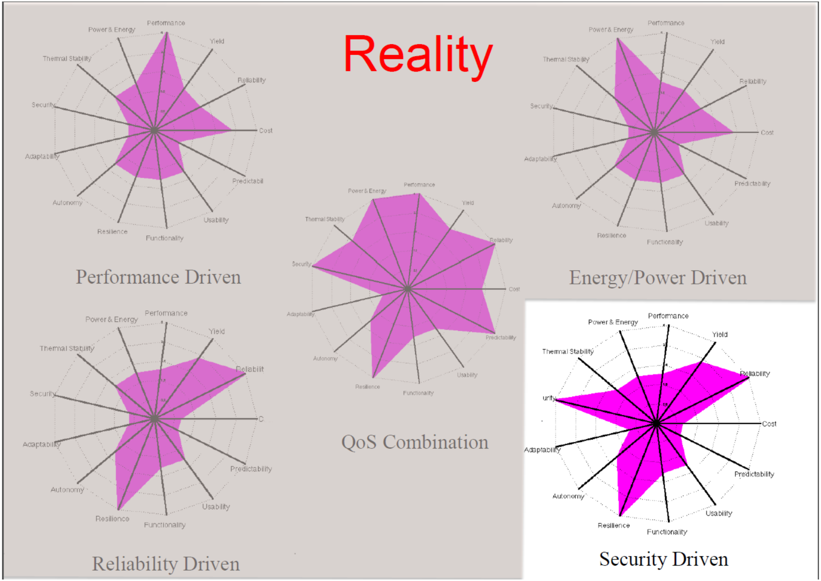
The SoC Radar



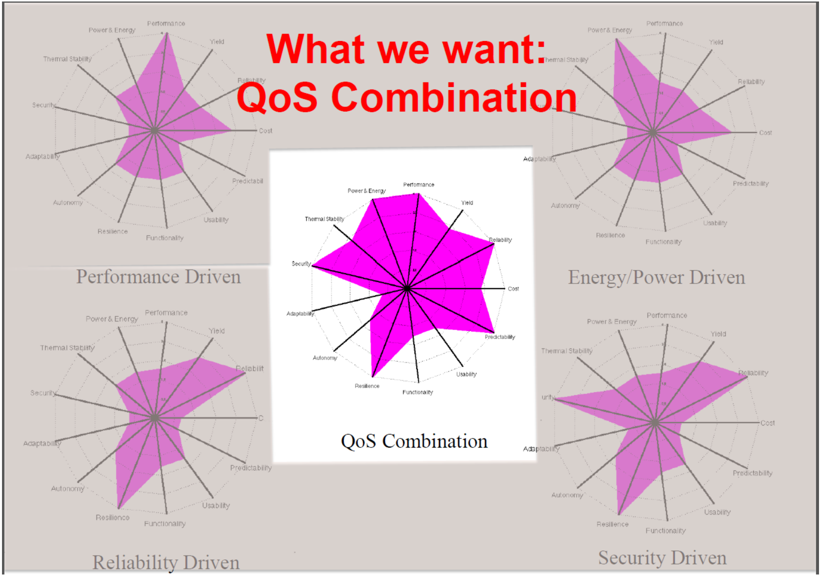
The SoC Radar



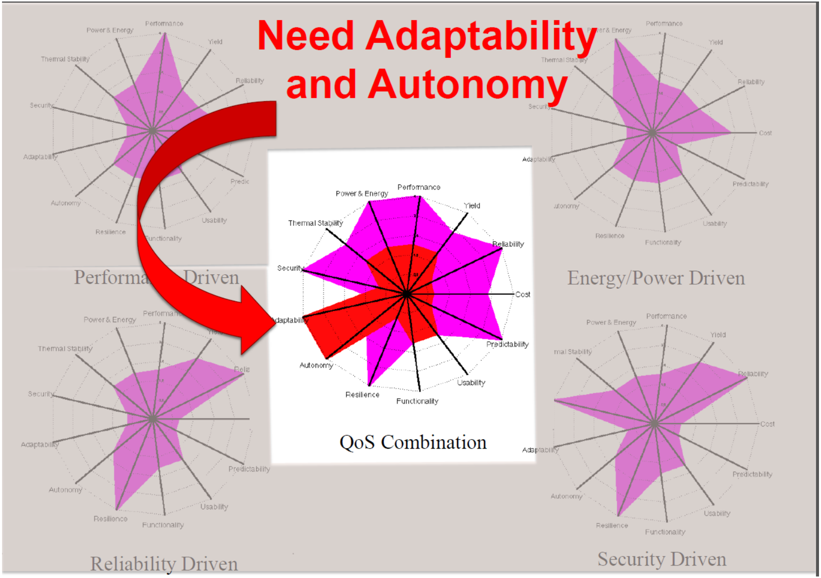
The SoC Radar



The SoC Radar



The SoC Radar



Autonomy and Adaptivity

Autonomy is the ability to operate independently, without external control.

Adaptivity is the ability to effect run-time changes and handle unexpected events.

What do we mean with Awareness?

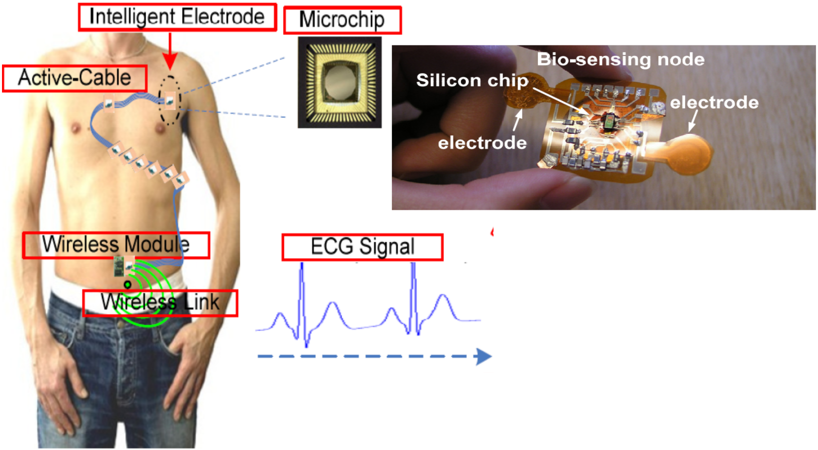
Does a bridge know when it is weakening? Is a thermometer aware of the temperature? Does a robot recognize its own limps? Does a surveillance system recognize its own camera? Is a human aware of his immune system? Is a human aware of her own arms?

Which Ingredients Lead to Awareness ?

- Data abstraction
- Disambiguation
- Desirability scale
- History
- Goals
- Attention
- Learning
- Introspection

Johan Moreelses “Der Alchemist”, 1630

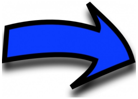
Awareness for Resource Constrained, Insect-like Gadgets

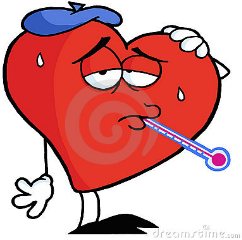


Courtesy of KTH

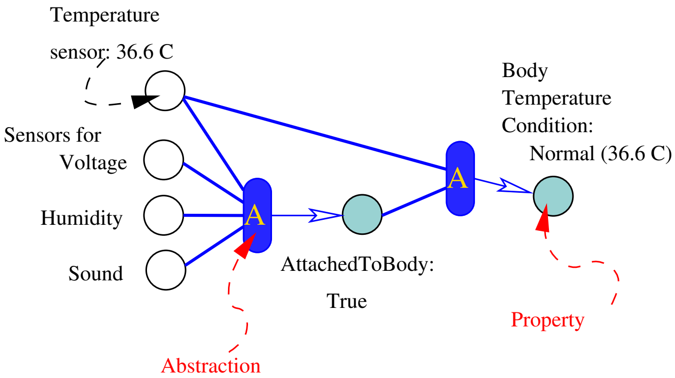
Abstractions and Models

Abstraction: Mapping of Measurements $\Rightarrow$ Properties



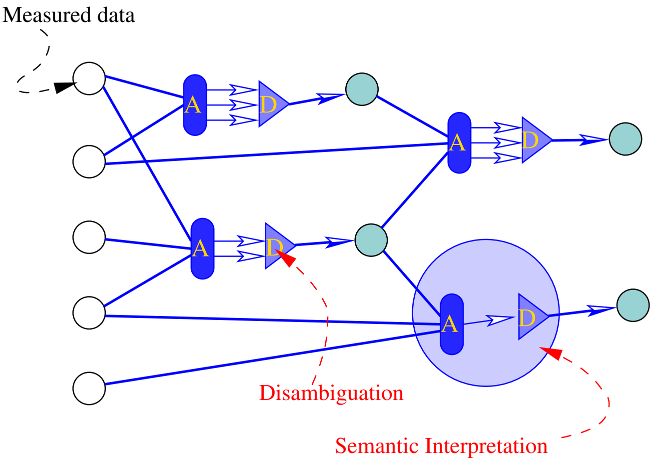


Abstractions and Models



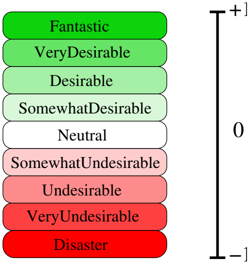
Disambiguation

Selection among several interpretations



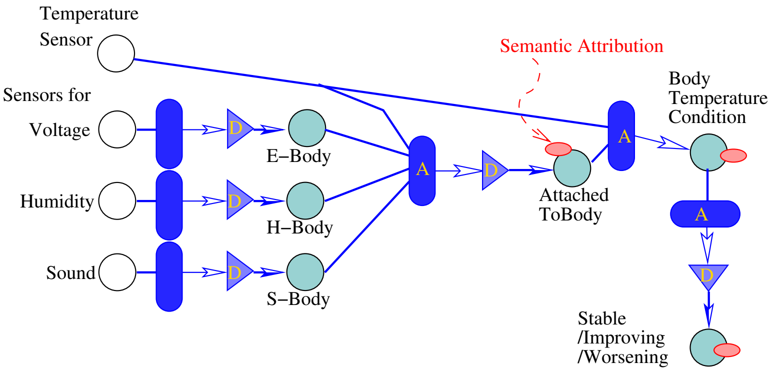
Desirability Scale

Desirability is the common, internal currency.



Semantic Attribution maps the values of a property to a point in the desirability scale.

BioPatch with Semantic Attribution



History

History of a Property The evolution of the values of a property.

Abstracted History The history stores abstracted values.

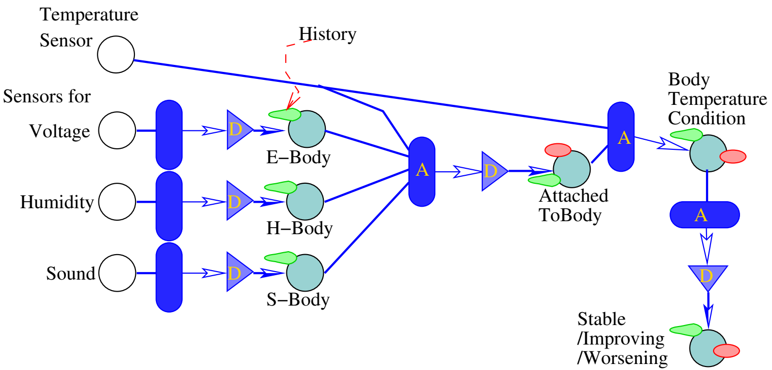
Attributed History The history is annotated with attributions.

Fading History If the property values are more abstracted the longer ago they have occurred.

Consolidating History Relevant memories are enforced, irrelevant memories are cleaned.

Evolving History Memories are adjusted to fit later observations.

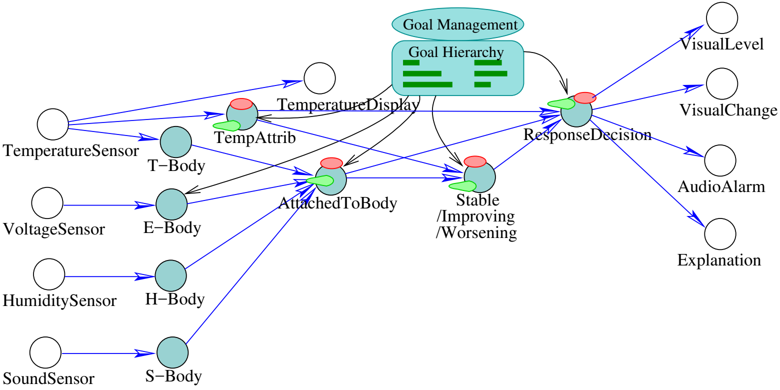
BioPatch with History



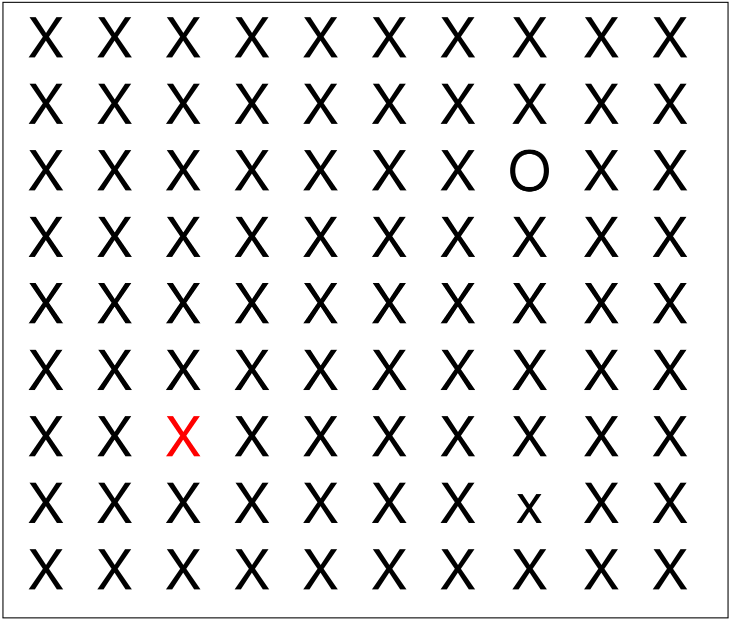
Expectations and Goals

- Expectations on Environment
- Expectations on Subject
- Sub-Goals
- Goals
- Purpose

Acting BioPatch



Attention



Attention


H. Rousseau

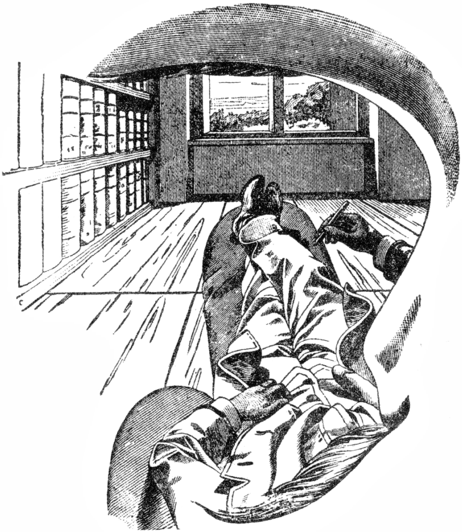
Attention

Introspection and Simulation

Self Inspection Engine

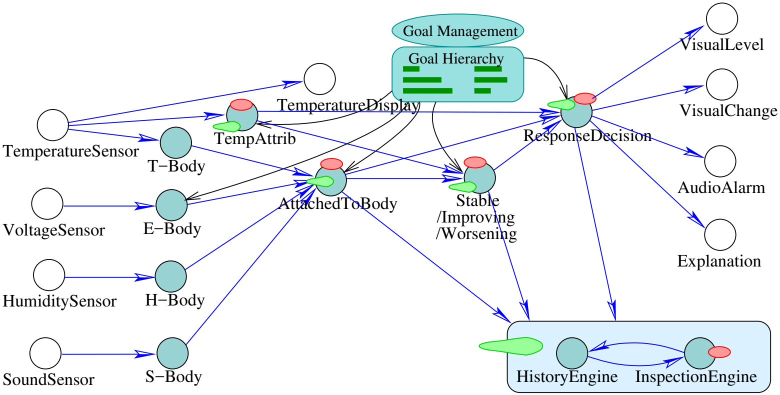
Model Transformation

Simulation

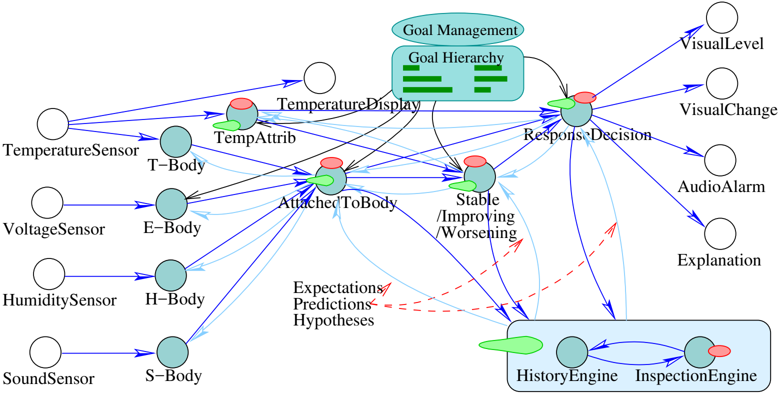


Ernst Mach "Innenperspektive", 1886

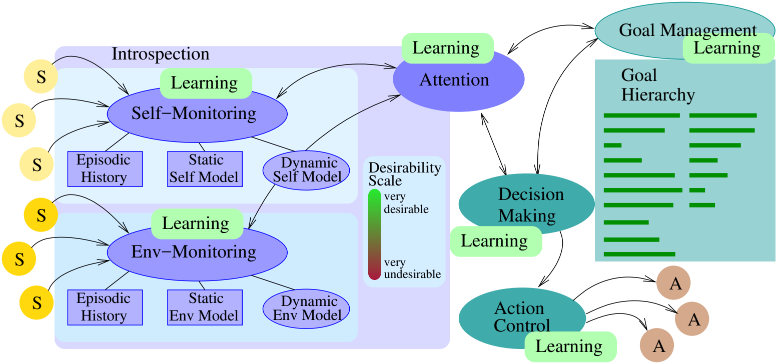
Self-inspecting BioPatch



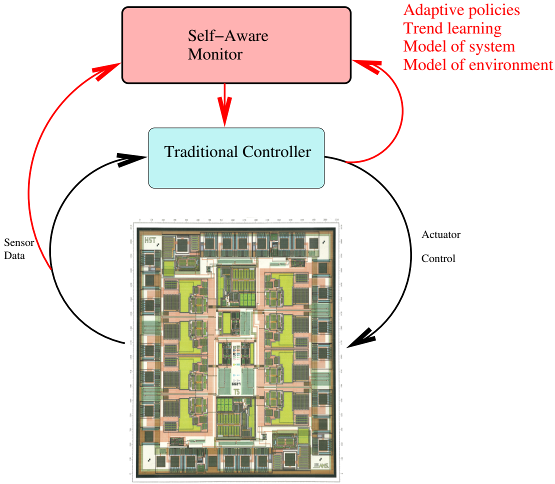
BioPatch with Top-down Prediction



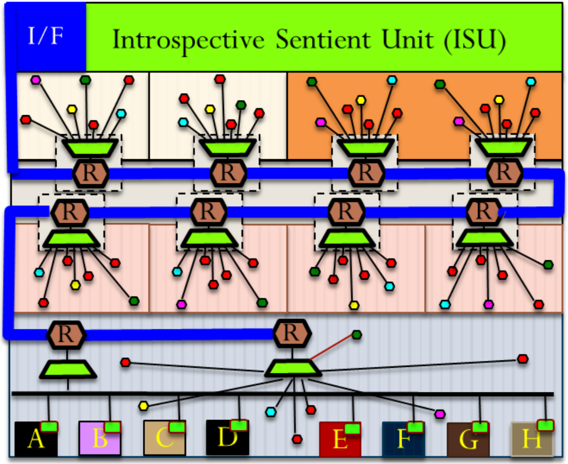
Self-Awareness Architecture



Cyber-Physical SoC

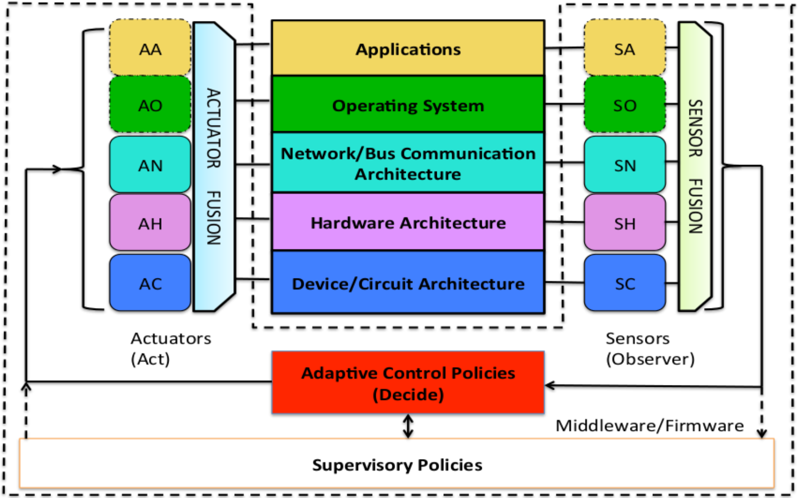


CPSoC - A Sensor Rich SoC Platform



Santanu Sarma et al. “CyberPhysical-System-On-Chip (CPSoC): A Self-Aware MPSoC Paradigm with Cross-Layer Virtual Sensing and Actuation”. In: *Proceedngs of the Design, Automation and Test in Europe Conference and Exhibition (DATE)*. Grenoble, France, Mar. 2015

CPSoC - A Sensor Rich SoC Platform



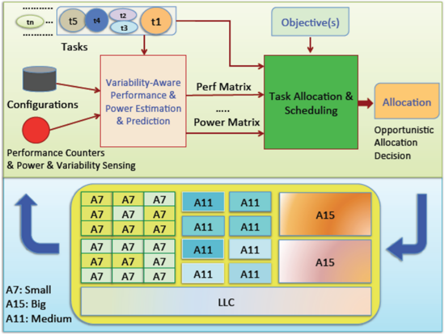
Nikil Dutt, Axel Jantsch, and Santanu Sarma. “Self-Aware Cyber-Physical Systems-on-Chip”. In: *Proceedings of the International Conference for Computer Aided Design*. invited. Austin, Texas, USA, Nov. 2015

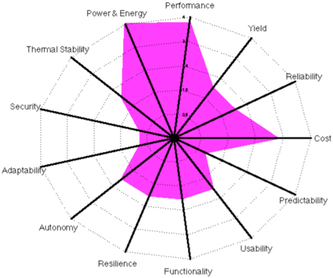
Sensing and Actuating at All Layers

Layers	Virtual/Physical Sensors	Virtual/Physical Actuators
Application	Workload, Power, Energy, Execution Time	Approximation, Algorithmic choice, Transformations
Operating System	System utilization, Peripheral states	Task allocation, Partitioning, Scheduling, Migration, Duty cycle
Network/Bus	Bandwidth, Packet/flit status, Channel status, Congestion	Adaptive routing, Dynamic BW allocation, Channel allocation, Flow control
Hardware Architecture	Cache miss rate, Access rate, IPC, Throughput, Resource utilization	Cache sizing, Issue width sizing, Reconfiguration, Resource provisioning
Circuit/Device	Circuit delay, Aging effects, Leakage, Temperature, Device faults	DVFS, Clock gating, Power gating

Santanu Sarma et al. "CyberPhysical-System-On-Chip (CPSoC): A Self-Aware MPSoC Paradigm with Cross-Layer Virtual Sensing and Actuation". In: *Proceedngs of the Design, Automation and Test in Europe Conference and Exhibition (DATE)*. Grenoble, France, Mar. 2015

Improvement of Energy Efficiency



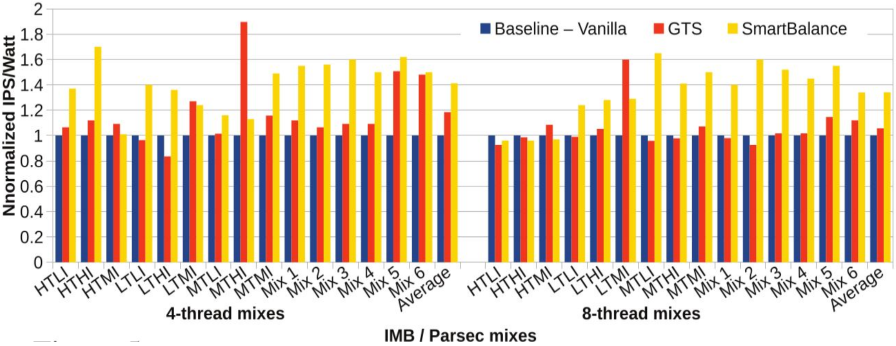


Goal:

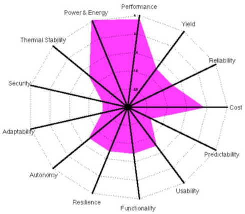
- **Energy Efficiency**

Santanu Sarma and Nikil Dutt. "Cross-Layer Exploration of Heterogeneous Multicore Processor Configurations". In: *VLSI Design Conference*. 2015

Improvement of Energy Efficiency

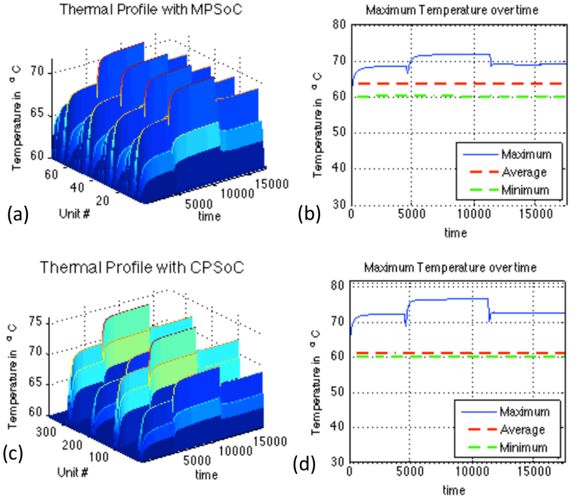


The benefit comes from actually measuring energy efficiency.



Santanu Sarma et al. “SmartBalance: A Sensing-Driven Linux Load Balancer for Energy Efficiency of Heterogeneous MPSoCs”. In: *Proceedings of the Design Automation Conference.* July 2015

Thermal-Aware Performance

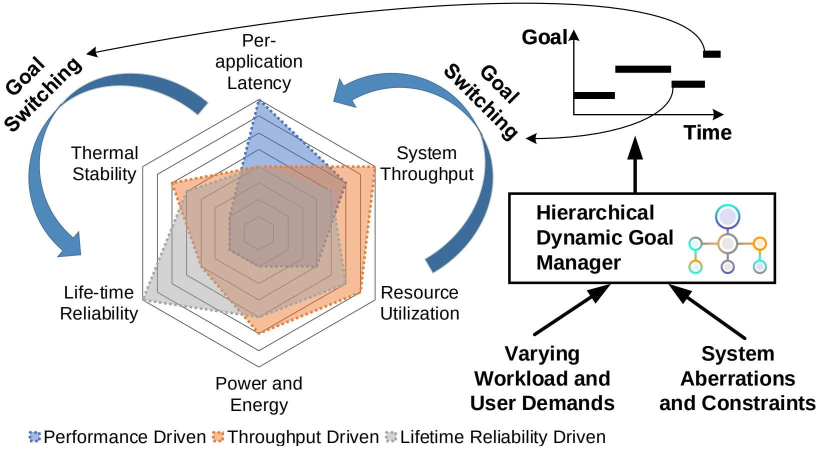


Throughput improvement by 70%-300% for same power and temperature.

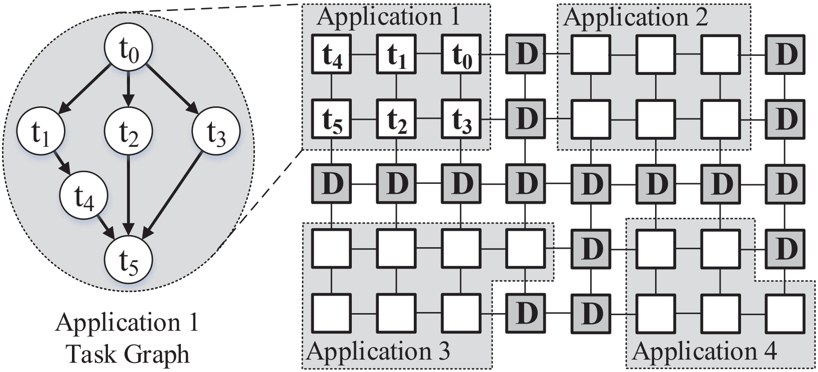
Benefit is due to accurate and fine-grain measurement and tight tracking.

Santanu Sarma et al. *CyberPhysical-System-On-Chip (CPSoC): Sensor-Actuator Rich Self-Aware Computational Platform*. Tech. rep. CECS Technical Report No: CECS TR–13–06. Irvine, CA 92697-2620, USA: Center for Embedded Computer Systems University of California, Irvine, May 2013

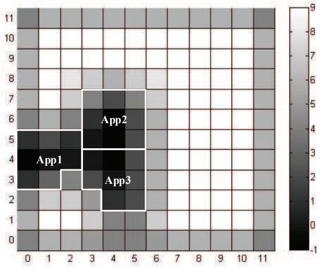
Goals for Dynamic Task Mapping

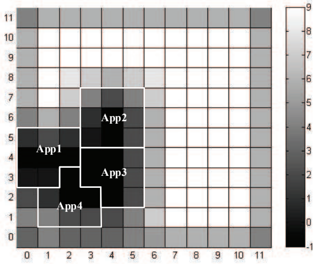


Dynamic Task Mapping



Example 1: Performance Driven Task Mapping

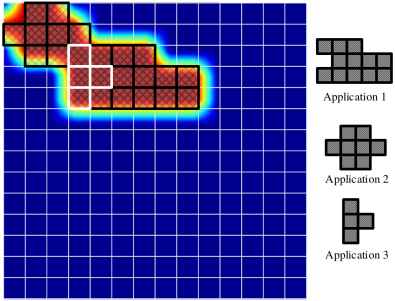


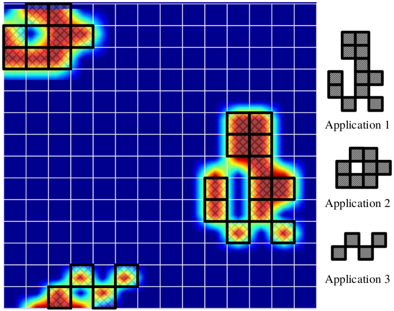


MapPro prefers compact and contiguous regions.

Mohammad-Hashem Haghbayan et al. “MapPro: Proactive Runtime Mapping for Dynamic Workloads by Quantifying Ripple Effect of Applications on Networks-on-Chip”. In: *Proceedings of the International Symposium on Networks on Chip*. Vancouver, Canada, Sept. 2015

Example 2: Throughput- and Power-Constrained Task Mapping

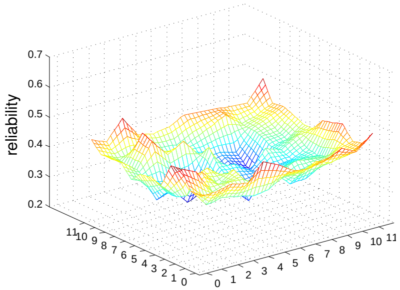


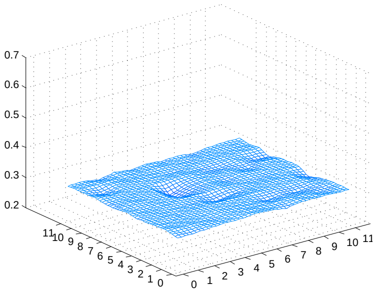


The patterning algorithm disperses mapped cores to maximize the Thermal Safe Power budget.

Anil Kanduri et al. "Dark Silicon Aware Runtime Mapping for Many-core Systems: A Patterning Approach". In: *Proceedings of the International Conference on Computer Design (ICCD)*. New York City, USA, Oct. 2015, pp. 610–617

Example 3: Lifetime-Reliability-Driven Task Mapping





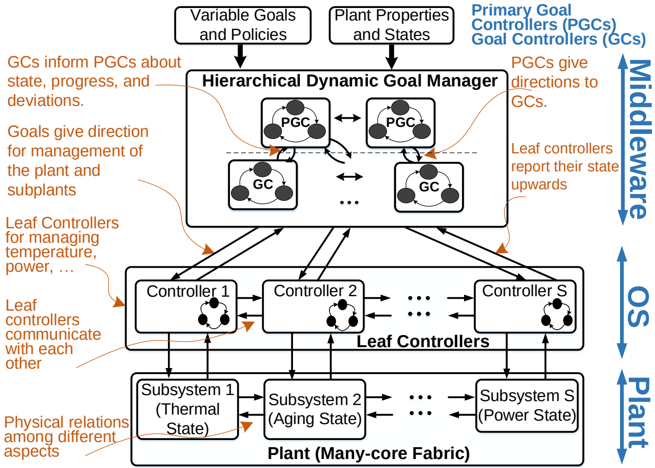
MapPro:
lifetime=5.52 years

Reliability aware mapping:
lifetime=12 years

The plots show the reliability of cores at the end of the system's lifetime.
The end of the system's life is reached when the reliability of one core drops below 30%.

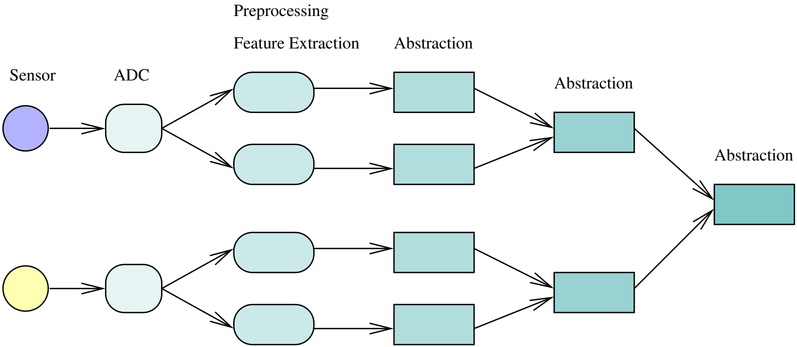
M. H. Haghbayan et al. "A lifetime-aware runtime mapping approach for many-core systems in the dark silicon era". In: *Design, Automation Test in Europe Conference Exhibition (DATE)*. Mar. 2016, pp. 854–857

Hierarchical Goal Mangement

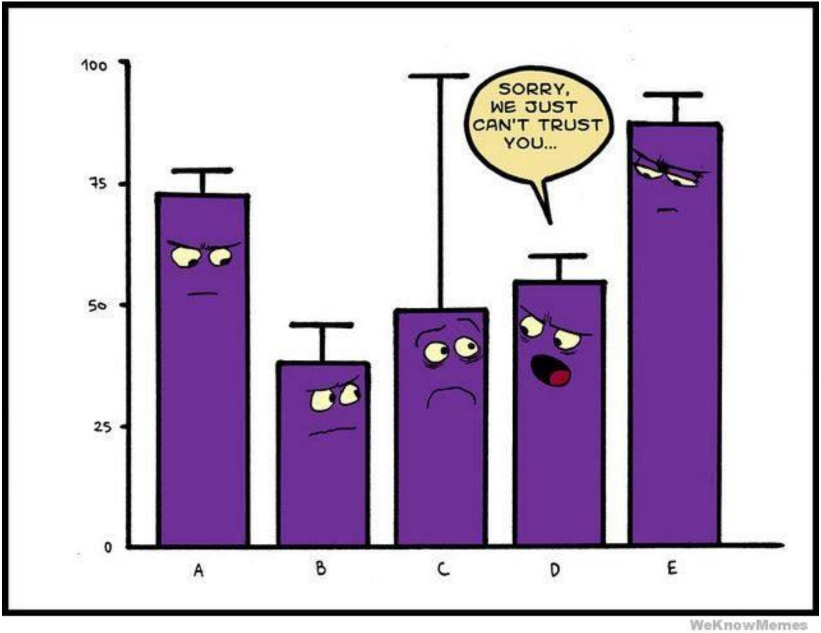


- The system's requirements changes over its lifetime.
- Different objectives are invoked at different time.

Observation Pipeline



Data and Meta-Data



Data and Meta-Data

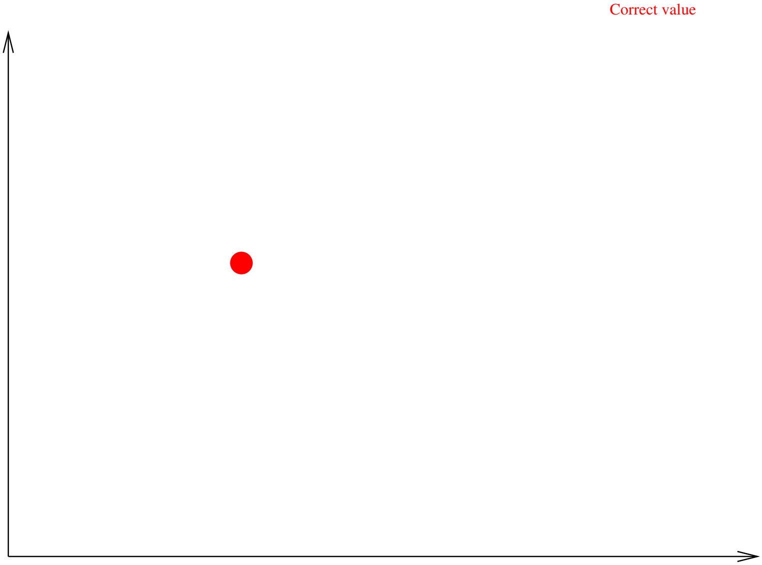
Accuracy Systematic errors, a measure of statistical bias.

Precision Random errors, a measure of statistical variability.

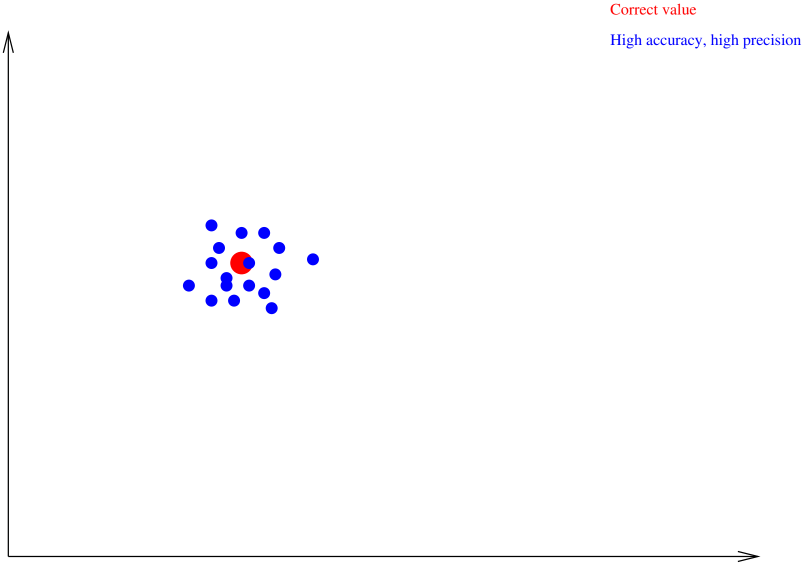
Data Reliability The extent to which a measuring procedure yields the same results on repeated trials.

Relevance The quality of being important for the matter at hand.

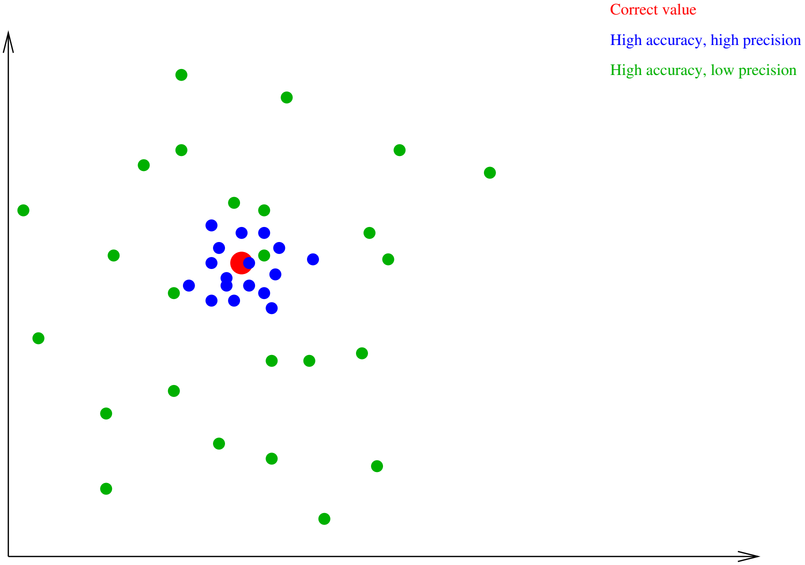
Accuracy and Precision



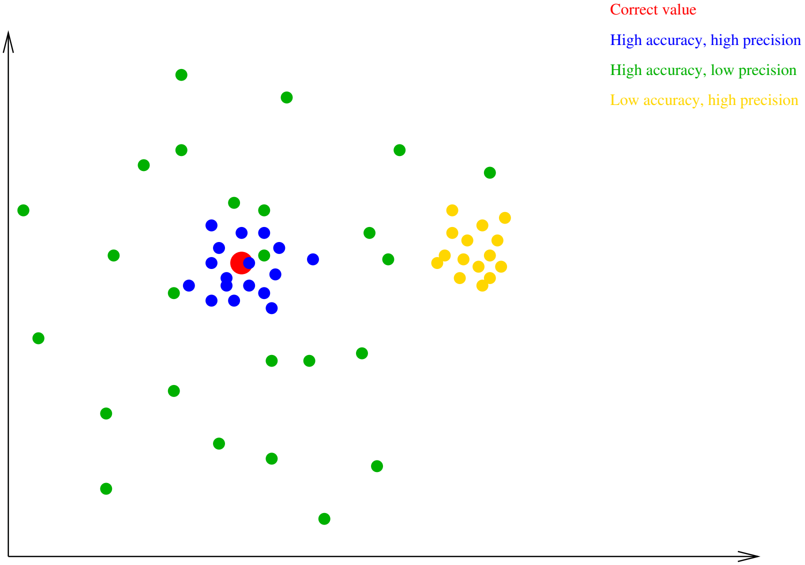
Accuracy and Precision



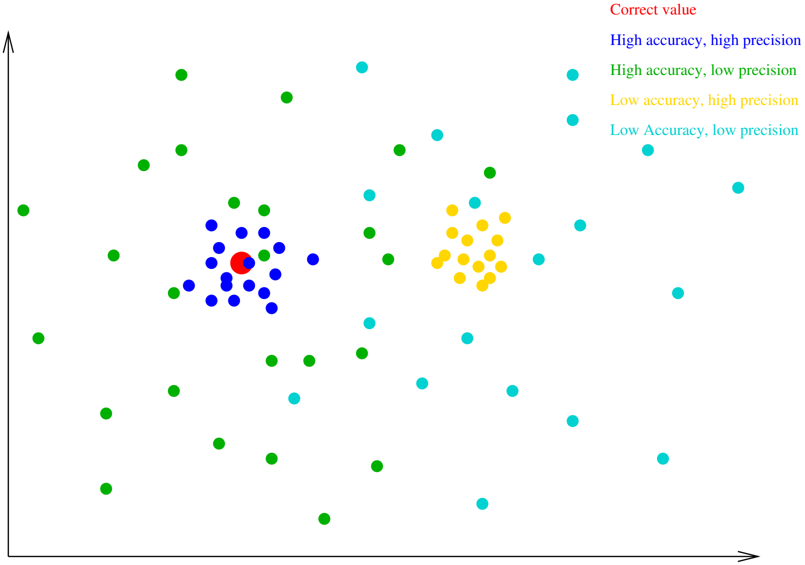
Accuracy and Precision



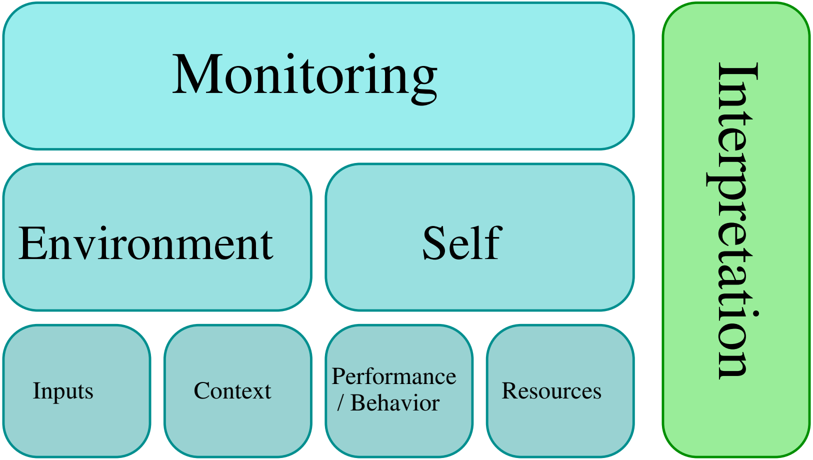
Accuracy and Precision



Accuracy and Precision

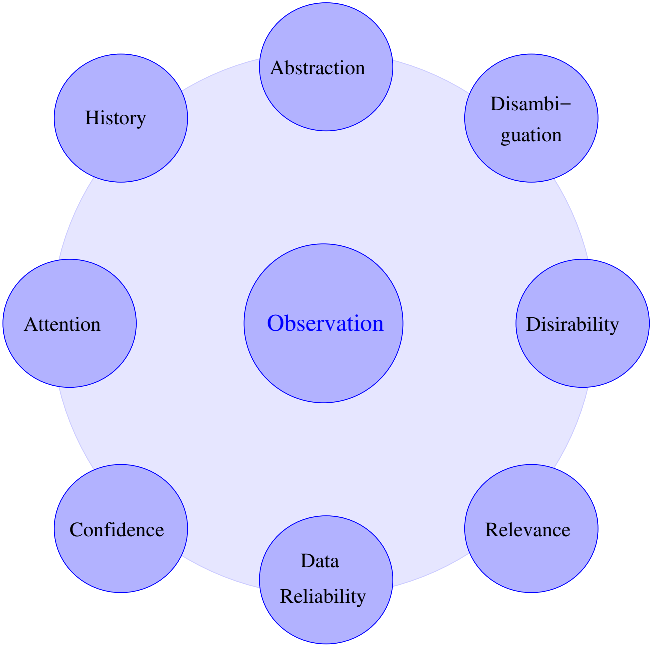


Comprehensive Observation



Nima TaheriNejad, Axel Jantsch, and David Pollreisz. “Comprehensive Observation and its Role in Self-Awareness - An Emotion Recognition System Example”. In: *Proceedings of the Federated Conference on Computer Science and Information Systems*. Gdansk, Poland, Sept. 2016

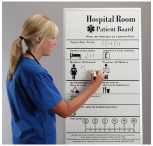
Observation Circle



Early Warning Score

Score	3	2	1	0	1	2	3
Heart rate[1]	<40	40–51	51–60	60–100	100–110	110–129	>129
Systolic BP[2]	<70	70–81	81–101	101–149	149–169	169–179	>179
Breath rate[3]		<9		9–14	14–20	20–29	>29
SPO$_2$ (%)	<85	85–90	90–95	>95			
Body temp.[4]	<28	28–32	32–35	35–38		38–39.5	>39.5

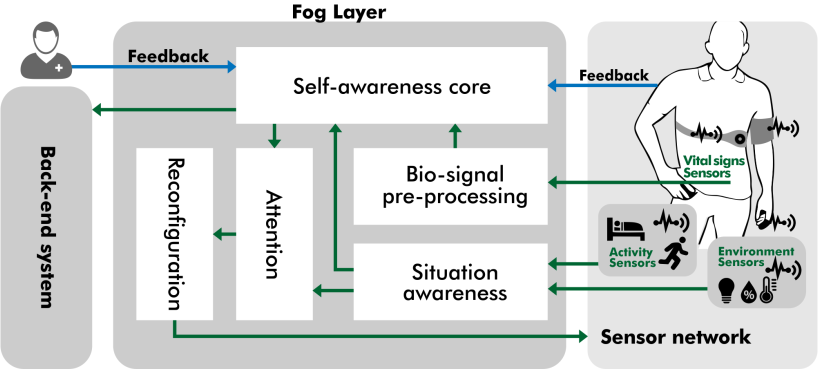
[1]beats per minute, [2]mmHg, [3]breaths per minute, [4] $^\circ$C



EWS Improvement

- Data reliability:
  - Values in reasonable scope
  - Changes in reasonable scope
  - Consistency between sensors
- Situation awareness
- Power efficiency

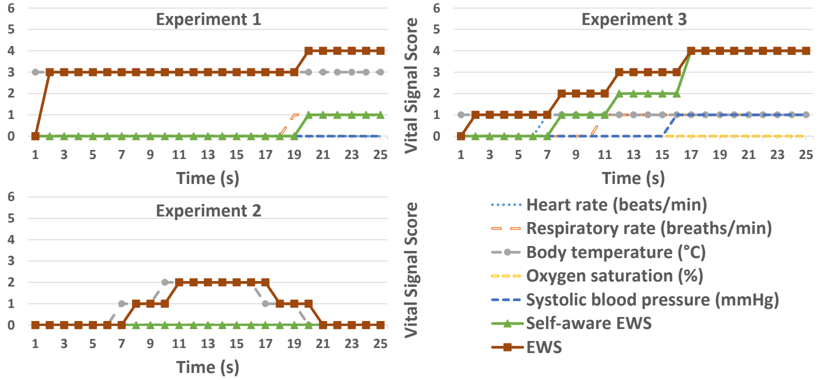
Enhanced Early Warning Score



Arman Anzanpour et al. "Self-Awareness in Remote Health Monitoring Systems using Wearable Electronics". In: *Proceedings of Design and Test Europe Conference (DATE)*. Lausanne, Switzerland, Mar. 2017

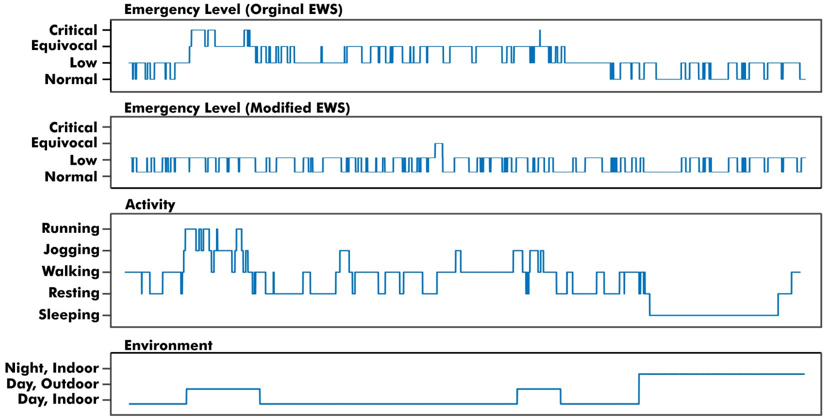
Enhanced Early Warning Score - Data Reliability

1. Check on the reliability of sensed values
2. Check on the reliability of value changes
3. Check on consistency between sensor data



Enhanced Early Warning Score - Situation Awareness

1. Consider the activity mode of person
2. Consider time of day
3. Consider location



Enhanced Early Warning Score - Power Efficiency

1. Prioritize different situations

Enhanced Early Warning Score - Power Efficiency

1. Prioritize different situations
2. Distinguish different modes of urgency

Emergency Level:	Score:0 Normal				Score:1-3 Low				Score:4-6 Medium				Score>6 High			
	Indoor		Outdoor		Indoor		Outdoor		Indoor		Outdoor		Indoor		Outdoor	
	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
Sleeping	E	E	E	E	C	D	D	D	B	C	C	C	A	A	B	B
Resting	D	D	D	D	C	C	C	C	B	B	B	B	A	A	B	B
Walking	C	C	C	C	B	C	C	C	B	B	B	B	A	A	A	B
Jogging	C	C	C	C	B	B	B	C	B	B	B	B	A	A	A	B
Running	C	C	C	C	B	B	B	B	B	B	B	B	A	A	A	A

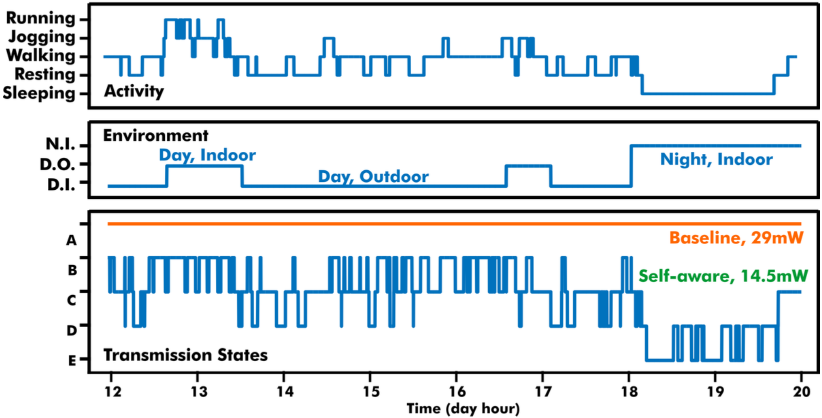
Enhanced Early Warning Score - Power Efficiency

1. Prioritize different situations
2. Distinguish different modes of urgency
3. Define sensing activity for each mode

State	Respiration Rate Activity	Blood Pressure	Heart Rate, SpO2, and Body Temp.	Transmission Power Consumption
A	Continuous	Every hour in day Disabled in night	Every sec.	29 mW
B	2 min continuous 8 min OFF	Every hour in day Disabled in night	Every sec.	26.8 mW
C	2 min continuous 3 min OFF	Every 3 hours in day Disabled in night	Every min.	12.5 mW
D	2 min continuous 8 min OFF	Every 3 hours in day Disabled in night	Every min.	7 mW
E	2 min continuous 18 min OFF	Disabled	Every min.	4.3 mW

Enhanced Early Warning Score - Power Efficiency

Over a day half the energy can be saved.

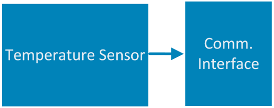


Enhanced Early Warning Score Summary

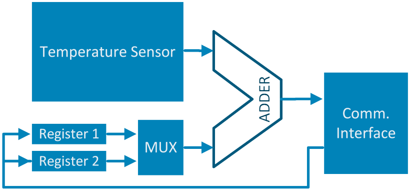
- Considering data reliability improves quality of observation;
- Considering sitation improves quality of observation;
- Collecting needed data only improves efficiency.

Attention Based Temperature Measurement

- How many temperature measurements are required in an MPSoC?
- It varies over several orders of magnitude depending on activity and current temperature.



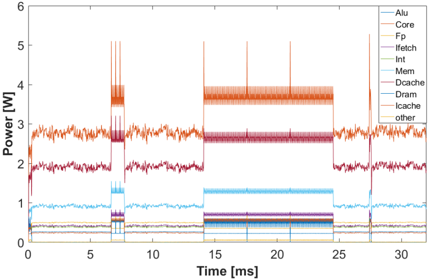
Conventional Architecture

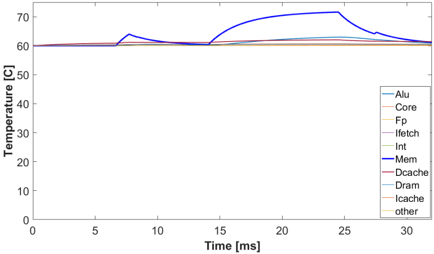


Proposed Architecture

Nima TaheriNejad, M. Ali Shami, and Sai Manoj P. D. “Self-aware sensing and attention-based data collection in Multi-Processor System-on-Chips”. In: *15th IEEE International New Circuits and Systems Conference (NEWCAS).* June 2017, pp. 81–84

Attention Based Temperature Measurement





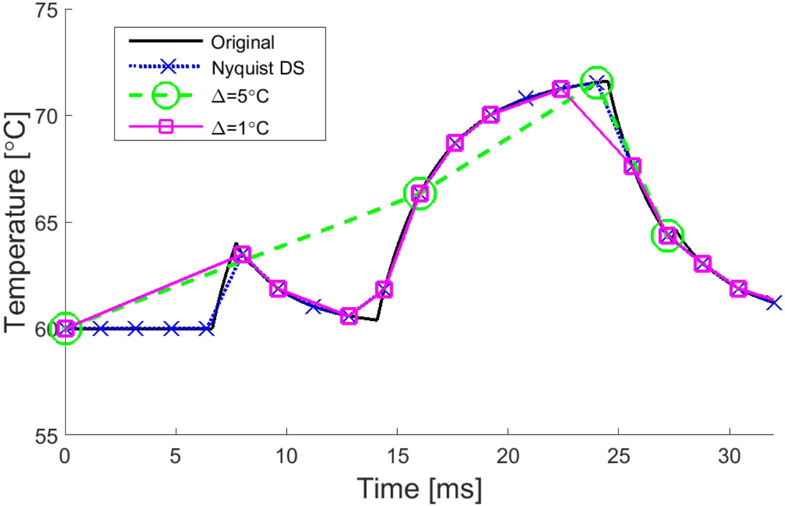
Intel Nehalem processor, running Barnes from SPLASH-2 Benchmarks, using Snipersim and Hotspot.

Attention Based Temperature Measurement

- When only differences $> \Delta = 1, 2, 5^{\circ}$C are reported, 7 out of 10 sensors send only 1 value in this experiment.
- Reduction of temperature reports for Memory, ALU and D-Cache:

Unit	$\Delta = 1$	Imp.	$\Delta = 2$	Imp.	$\Delta = 5$	Imp.
Memory	13	35%	9	55%	4	80%
ALU	4	80%	2	90%	1	95%
D-Cache	2	90%	2	90%	1	95%
All others	1	95%	1	95%	1	95%

Attention Based Temperature Measurement



Attention Based Temperature Measurement

- Rate of temperature reporting can be significantly reduced and fine tuned;
- Can depend on
  - relative difference,
  - absolute difference,
  - absolute value,
  - system level mode;
- Potential benefits:
  - reduced processing,
  - reduced communication,
  - reduced measurements.

Challenges with Self-aware, Autonomous, Adaptive SoCs

- How to assess and ensure the quality of sensor data?
- How to express “correctness”?
- How to validate a smartly adapting system?
- How to perform tradeoff analysis for smartness features?
- How to quantify uncertainty, dynamicity, and variability in the platform, the environment, and the applications?

Challenges with Self-aware, Autonomous, Adaptive SoCs

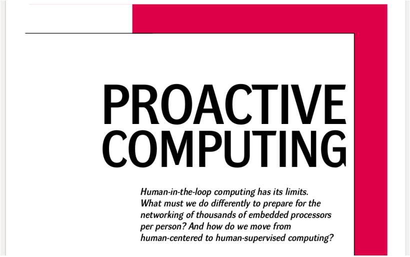
- How to reconcile autonomy with safety critical and real-time systems?
- How to develop efficient learning algorithms?
- How to formally model and formulate the goal management?
- How to verify it w.r.t. convergence, efficiency, robustness, QoS guarantees, etc.?
- How to handle a dynamic hierarchy of goals?

Challenges with Self-aware, Autonomous, Adaptive SoCs

- How to make goal management lightweight?
- How to scale self-awareness?
- How to detect and handle design errors?
- How to detect and handle malicious attacks?
- .....

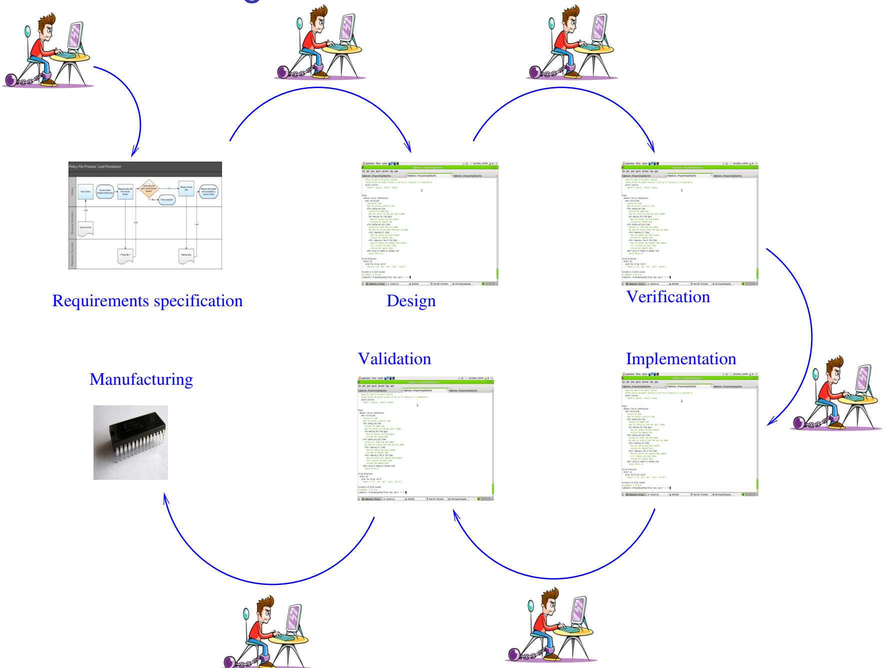
Let's Get Out

- Let's get physical
- Let's get real
- Let's get out

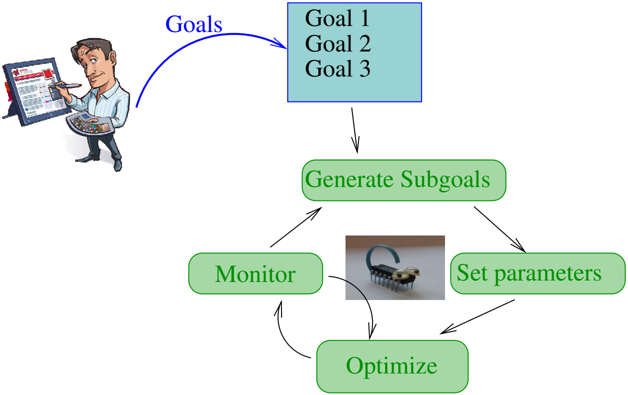


David Tennenhouse. "Proactive Computing". In: *Communications of the ACM* 43.5 (May 2000), pp. 43–50

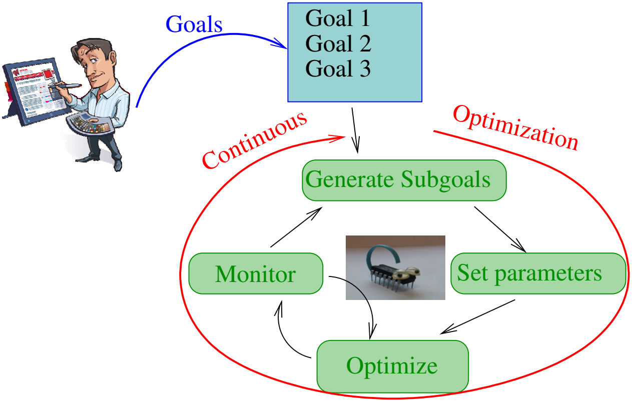
Traditional Design Flow



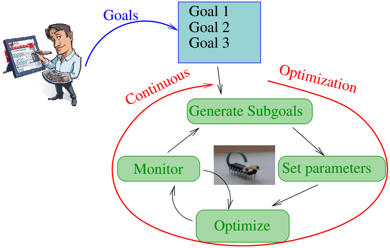
Design of Self-Aware Chips



Design of Self-Aware Chips



Design of Self-Aware Chips



For that to work we need Methods:

- to guarantee behavior,
- to guarantee performance,
- formulate and manage goals,
- for customized and efficient learning.

Questions ?

References I

Arman Anzanpour et al. "Self-Awareness in Remote Health Monitoring Systems using Wearable Electronics". In: *Proceedings of Design and Test Europe Conference (DATE)*. Lausanne, Switzerland, Mar. 2017.

Nikil Dutt, Axel Jantsch, and Santanu Sarma. "Self-Aware Cyber-Physical Systems-on-Chip". In: *Proceedings of the International Conference for Computer Aided Design*. invited. Austin, Texas, USA, Nov. 2015.

Nikil Dutt, Axel Jantsch, and Santanu Sarma. "Towards Smart Embedded Systems: A Self-Aware System-on-Chip Perspective". In: *ACM Transactions on Embedded Computing Systems, Special Issue on Innovative Design Methods for Smart Embedded Systems* 15.2 (Feb. 2016). invited, pp. 22–27.

Nikil Dutt, Amir M. Rahmani, and Axel Jantsch. "Empowering Autonomy through Self-awareness in MPSoCs". In: *Proceedings of the IEEE NEWCAS Conference*. Strasbourg, France, June 2017.

Mohammad-Hashem Haghbayan et al. "MapPro: Proactive Runtime Mapping for Dynamic Workloads by Quantifying Ripple Effect of Applications on Networks-on-Chip". In: *Proceedings of the International Symposium on Networks on Chip*. Vancouver, Canada, Sept. 2015.

References II

M. H. Haghbayan et al. “A lifetime-aware runtime mapping approach for many-core systems in the dark silicon era”. In: *Design, Automation Test in Europe Conference Exhibition (DATE)*. Mar. 2016, pp. 854–857.

Axel Jantsch and Kalle Tammemäe. “A Framework of Awareness for Artificial Subjects”. In: *Proceedings of the 2014 International Conference on Hardware/Software Codesign and System Synthesis*. CODES ’14. New Delhi, India: ACM, 2014, 20:1–20:3.

Anil Kanduri et al. “Dark Silicon Aware Runtime Mapping for Many-core Systems: A Patterning Approach”. In: *Proceedings of the International Conference on Computer Design (ICCD)*. New York City, USA, Oct. 2015, pp. 610–617.

S. Kounev et al., eds. *Self-Aware Computing Systems*. Springer, 2017.

Peter R. Lewis et al. “Architectural Aspects of Self-aware and Self-expressive Computing Systems”. In: *IEEE Computer* (Aug. 2015).

Peter R. Lewis et al., eds. *Self-Aware Computing Systems: An Engineering Approach*. Springer, 2016.

References III

Santanu Sarma et al. *CyberPhysical-System-On-Chip (CPSoC): Sensor-Actuator Rich Self-Aware Computational Platform*. Tech. rep. CECS Technical Report No: CECS TR–13–06. Irvine, CA 92697-2620, USA: Center for Embedded Computer Systems University of California, Irvine, May 2013.

Santanu Sarma et al. "On-Chip Self-Awareness Using Cyberphysical-Systems-On-Chip (CPSoC)". In: *Proceedings of the 12th International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*. New Delhi, India, Oct. 2014.

Santanu Sarma et al. "CyberPhysical-System-On-Chip (CPSoC): A Self-Aware MPSoC Paradigm with Cross-Layer Virtual Sensing and Actuation". In: *Proceedngs of the Design, Automation and Test in Europe Conference and Exhibition (DATE)*. Grenoble, France, Mar. 2015.

Santanu Sarma et al. "SmartBalance: A Sensing-Driven Linux Load Balancer for Energy Efficiency of Heterogeneous MPSoCs". In: *Proceedings of the Design Automation Conference*. July 2015.

Santanu Sarma and Nikil Dutt. "Cross-Layer Exploration of Heterogeneous Multicore Processor Configurations". In: *VLSI Design Conference*. 2015.

References IV

David Tennenhouse. "Proactive Computing". In: *Communications of the ACM* 43.5 (May 2000), pp. 43–50.

Nima TaheriNejad, Axel Jantsch, and David Pollreisz. "Comprehensive Observation and its Role in Self-Awareness - An Emotion Recognition System Example". In: *Proceedings of the Federated Conference on Computer Science and Information Systems*. Gdansk, Poland, Sept. 2016.

Nima TaheriNejad, M. Ali Shami, and Sai Manoj P. D. "Self-aware sensing and attention-based data collection in Multi-Processor System-on-Chips". In: *15th IEEE International New Circuits and Systems Conference (NEWCAS)*. June 2017, pp. 81–84.