Wearable Cognitive Assistance

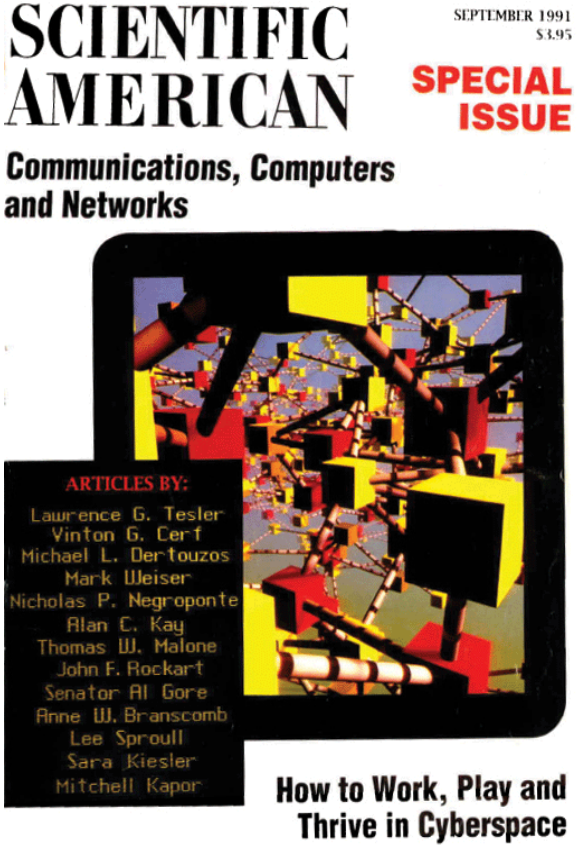
Vision, Reality & Challenges

Mahadev Satyanarayanan
School of Computer Science
Carnegie Mellon University

I. Vision

1991 to 2004

Mark Weiser (1952-1999)



The Computer for the 21st Century

Specialized elements of hardware and software, connected by wires, radio waves and infrared, will be so ubiquitous that no one will notice their presence

by Mark Weiser

The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.

Consider writing, perhaps the first information technology. The ability to represent spoken language symbolically for long-term storage freed information from the limits of individual memory. Today this technology is ubiquitous in industrialized countries. Not only do books, magazines and newspapers convey written information, but so do street signs, billboards, shop signs and even graffiti. Candy wrappers are covered in writing. The constant background presence of these products of "literacy technology" does not require active attention, but the information to be transmitted is ready for use at a glance. It is difficult to imagine modern life otherwise.

Silicon-based information technology, in contrast, is far from having become part of the environment. More than 50 million personal computers have been sold, and the computer nonetheless remains largely in a world of its own. It is approachable only through complex jargon that has nothing to do with the tasks for which people use computers. The state of the art is perhaps analogous to the period when scribes had to know as much about making ink or baking clay as they did about writing.

The arcane aura that surrounds personal computers is not just a "user interface" problem. My colleagues and I at the Xerox Palo Alto Research Center think that the idea of a "personal" computer itself is misplaced and that the vision of laptop machines, dynabooks and "knowledge navigators" is only a transitional step toward achieving the real potential of information technology. Such machines cannot truly make computing an integral, invisible part of people's lives. We are therefore trying to conceive a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish into the background.

Such a disappearance is a fundamental consequence not of technology but of human psychology. Whenever people learn something sufficiently well, they cease to be aware of it. When you look at a street sign, for example, you absorb its information without consciously performing the act of reading. Computer scientist, economist and Nobelist Herbert A. Simon calls this phenomenon "compiling"; philosopher Michael Polanyi calls it the "tacit dimension"; psychologist J. J. Gibson calls it "visual invariants"; philosophers Hans Georg Gadamer and Martin Heidegger call it the "horizon" and the "ready-to-hand"; John Seely Brown of PARC calls it the "periphery." All say, in essence, that only when things disappear in this way are we freed to use them without thinking and so to focus beyond them on new goals.

The idea of integrating computers seamlessly into the world at large runs counter to a number of present-day trends. "Ubiquitous computing" in this context does not mean just computers that can be carried to the beach, jungle or airport. Even the most powerful notebook computer, with access to a worldwide information network, still focuses attention on a single box. By analogy with writing, carrying a superlaptop is like owning just one very important book. Customizing this book, even writing millions of other books, does not begin to capture the real power of literacy.

Furthermore, although ubiquitous computers may use sound and video in addition to text and graphics, that does not make them "multimedia computers." Today's multimedia machine makes the computer screen into a demanding focus of attention rather than allowing it to fade into the background.

Perhaps most diametrically opposed to our vision is the notion of virtual reality, which attempts to make a world inside the computer. Users don special goggles that project an artificial scene onto their eyes; they wear gloves or even bodysuits that sense their motions and gestures so that they can move about and manipulate virtual objects. Although it may have its purpose in allowing people to explore realms otherwise inaccessible—the insides of cells, the surfaces of distant planets, the information web of data bases—virtual reality is only a map, not a territory. It excludes desks, offices, other people not wearing goggles and bodysuits, weather, trees, walks, chance encounters and, in general, the infinite richness of the universe. Virtual reality focuses an enormous apparatus on simulating the world rather than on invisibly enhancing the world that already exists.

Indeed, the opposition between the

MARK WEISER is head of the Computer Science Laboratory at the Xerox Palo Alto Research Center. He is working on the next revolution of computing after workstations, variously known as ubiquitous computing or embodied virtuality. Before working at PARC, he was a professor of computer science at the University of Maryland; he received his Ph.D. from the University of Michigan in 1979. Weiser also helped found an electronic publishing company and a video arts company and claims to enjoy computer programming "for the fun of it." His most recent technical work involved the implementation of new theories of automatic computer memory reclamation, known in the field as garbage collection.

94 SCIENTIFIC AMERICAN *September 1991*

Key Insight of Mark Weiser

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it."

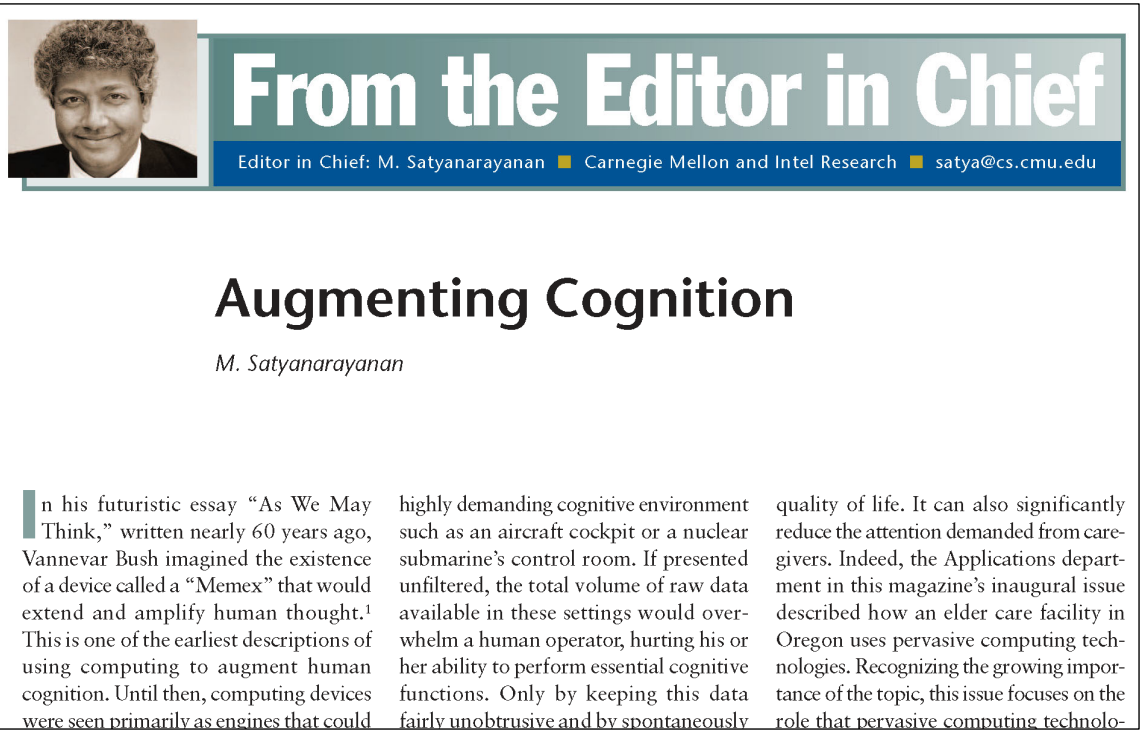
The Quest for Weiser's Vision

Creation of environments saturated with computing and communication, yet gracefully integrated with human users.

A 30-year quest for this holy grail

1991	"Ubiquitous Computing"	original name given by Mark Weiser
late 1990s to early 2010s	"Pervasive Computing"	originally used by IBM also IEEE Pervasive Computing and many other uses
today	"Internet of Things" (IoT)	widely used today omission of human-centric viewpoint

2004 Thought Piece in IEEE Pervasive Computing



From the Editor in Chief

Editor in Chief: M. Satyanarayanan ■ Carnegie Mellon and Intel Research ■ satya@cs.cmu.edu

Augmenting Cognition

M. Satyanarayanan

In his futuristic essay "As We May Think," written nearly 60 years ago, Vannevar Bush imagined the existence of a device called a "Memex" that would extend and amplify human thought.[1] This is one of the earliest descriptions of using computing to augment human cognition. Until then, computing devices were seen primarily as engines that could highly demanding cognitive environment such as an aircraft cockpit or a nuclear submarine's control room. If presented unfiltered, the total volume of raw data available in these settings would overwhelm a human operator, hurting his or her ability to perform essential cognitive functions. Only by keeping this data fairly unobtrusive and by spontaneously quality of life. It can also significantly reduce the attention demanded from caregivers. Indeed, the Applications department in this magazine's inaugural issue described how an elder care facility in Oregon uses pervasive computing technologies. Recognizing the growing importance of the topic, this issue focuses on the role that pervasive computing technolo-

"For example, imagine a wearable computer with a head-up display in the form of eyeglasses and with a built-in camera for continuous face recognition. This would offer the essentials of an augmented-reality system to aid cognition. When you look at a person, his or her name could pop up, possibly with additional cues to guide your greeting. Such "magic glasses" could transform your environment."

“It is hard to predict when systems of this kind will become off-the-shelf products. At the moment, they do not exist even as lab prototypes. Fortunately, compelling visions of the future have a habit of becoming true sooner than most people think possible.”

II. Reality

2015 to 2021*

NSF Large CNS-1518865: *"Wearable Cognitive Assistance"*
co-PIs: Martial Hebert, Roberta Klatzky, Dan Siewiorek

* What happened during 2004 to 2015? Wait until Part III

"Towards Wearable Cognitive Assistance"
Ha, K., Chen, Z., Hu, W., Richter, W., Pillai, P., Satyanarayanan, M.
Proceedings of the Twelfth International Conference on Mobile Systems, Applications, and Services (MobiSys 2014), Bretton Woods, NH, June 2014

"Early Implementation Experience with Wearable Cognitive Assistance Applications"
Chen, Z., Jiang, L., Hu, W., Ha, K., Amos, B., Pillai, P., Hauptmann, A., Satyanarayanan, M.
Proceedings of WearSys 2015, Florence, Italy, May 2015

"An Empirical Study of Latency in an Emerging Class of Edge Computing Applications for Wearable Cognitive Assistance"
Chen, Z., Hu, W., Wang, J., Zhao, S., Amos, B., Wu, G., Ha, K., Elgazzar, K., Pillai, P., Klatzky, R., Siewiorek, D., Satyanarayanan, M.
Proceedings of SEC 2017, San Jose, CA, October 2017

"Towards Scalable Edge-Native Applications"
Wang, J, Feng, Z., George, S., Iyengar, R., Pillai, P., Satyanarayanan, M.
Proceedings of SEC 2019), Washington, DC, November 2019

"Ajalon: Simplifying the authoring of wearable cognitive assistants"
Truong An Pham,Junjue Wang,Roger Iyengar,Yu Xiao,Padmanabhan Pillai,Roberta Klatzky,Mahadev Satyanarayanan
Journal of Software Practice and Experience, Vol. 51, No. 8, May 2021

Wearable Cognitive Assistance

"Look and feel of AR, with functionality of AI"

"An Angel on Your Shoulder"
Project Gabriel
http://gabriel.cs.cmu.edu

Wearable UI with wireless offload to cloudlet

***Real-time cognitive engines* on cloudlet (microservices)**

- scene analysis
- object/person recognition
- speech recognition
- language translation
- planning, navigation
- question-answering technology
- voice synthesis
- real-time machine learning
- …

Low latency response is crucial

Human Cognition is Amazing

Fast, accurate and robust

- face detection under hostile conditions (low lighting, distorted optics) **< 700 ms**
- face recognition **370 ms – 620 ms**
- is this sound from a human? **4 ms**
- VR head tracking **< 16 ms**

To be "superhuman" we need to beat these speeds

Leave time for additional software processing (e.g. database lookup) to add value to user

Task-specific Assistance

Example: cooking

passive recipe display

versus active guidance

“Wait, the oil is not hot enough”

Inspiration: GPS Navigation Systems

Turn by turn guidance

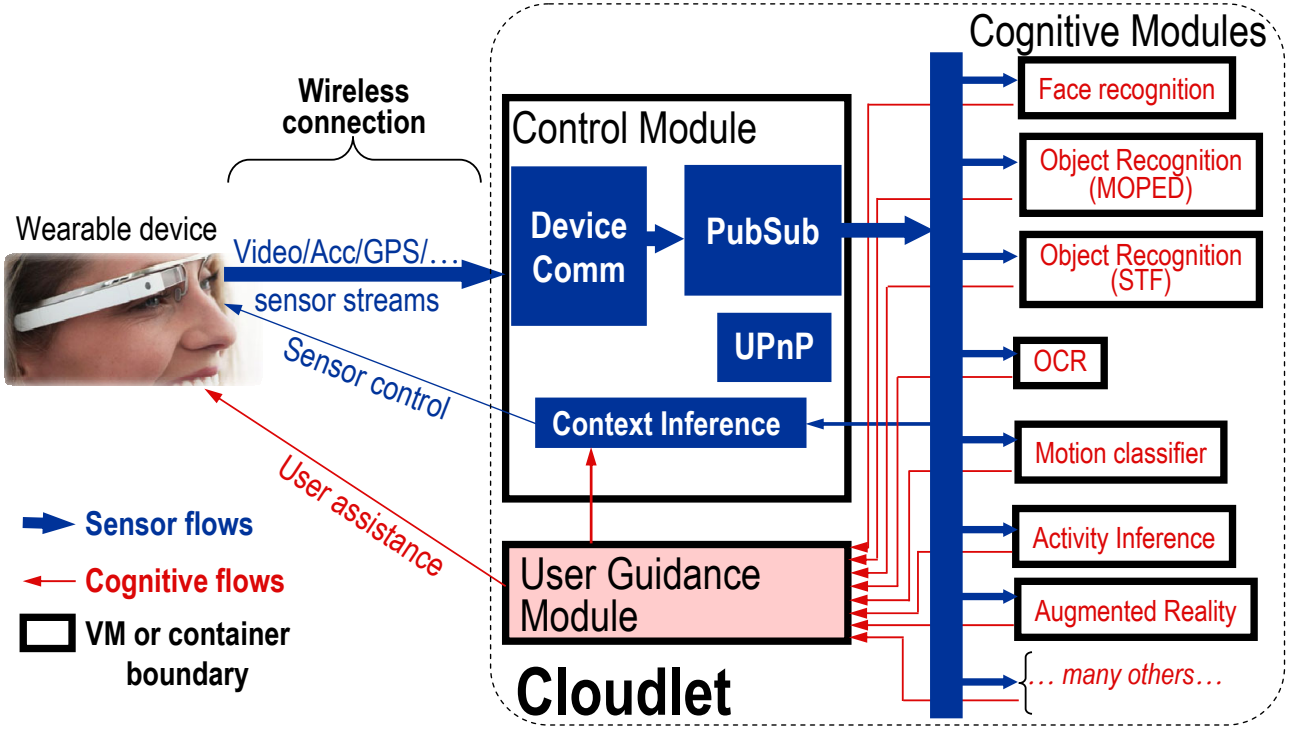
- **Ability to detect and recover**
- **Minimally distracting to user**

Uses only one type of sensor: location from GPS

Can we generalize this metaphor?

Gabriel Architecture

(PaaS for Wearable Cognitive Assistance)



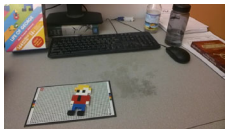
Baby Steps: 2D Lego Assembly

Very first proof-of-concept (September 2014)

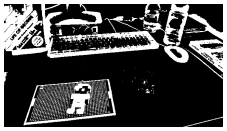
Deliberately simplified task to keep computer vision tractable

2D Lego Assembly **(YouTube video at http://youtu.be/uy17Hz5xvmY)**

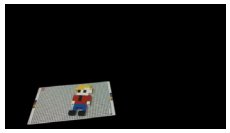
On Each Video Frame



(a) Input image



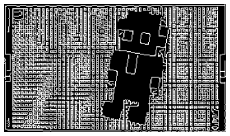
(b) Detected dark parts



(c) Detected board

(d) Board border

(e) Perspective corrected

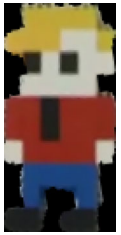


(f) Edges detected

(g) Background subtracted

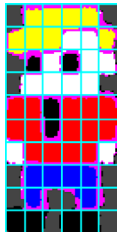
(h) Side parts added

(h) Lego detected



(i) Unrotated

(i) Color quantized



(j) Partitioned

[[0, 3, 3, 3, 3, 0],
[3, 3, 3, 1, 1, 3],
[0, 6, 1, 6, 1, 1],
[0, 1, 1, 1, 1, 0],
[4, 4, 6, 4, 4, 4],
[4, 4, 6, 4, 4, 4],
[1, 4, 4, 4, 4, 1],
[0, 5, 5, 5, 5, 0],
[0, 5, 0, 0, 5, 0],
[6, 6, 0, 6, 6, 0]]

(j) Matrix

(k) Synthesized

When Milliseconds Matter

Ping-pong assistant
(https://www.youtube.com/watch?v=_lp32sowyUA)

Assembling an IKEA Kit

IKEA kit assistant

(https://www.youtube.com/watch?v=qDPuvBWNIUs&index=5&list=PLmrZVvFtthdP3fwHPy_4d61oDvQY_RBgS)

Many Use Cases …

Assembly instructions

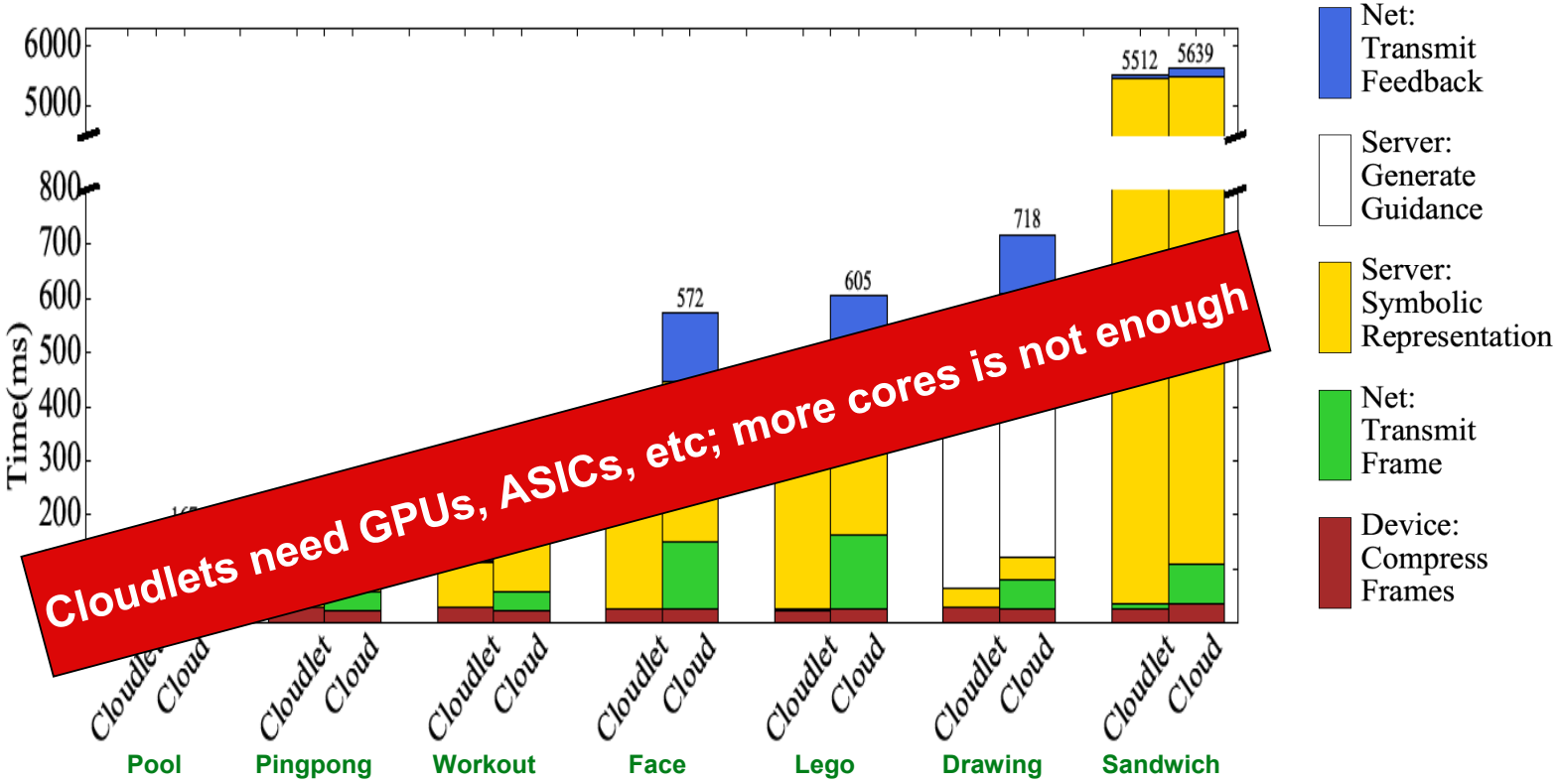
Industrial troubleshooting

Medical training

Self-Instrumentation

Strong willpower

Where Does Time Go?



Network time (green & blue) varies between cloudlet & cloud

Yellow (processing) is similar on cloudlet and cloud

Sandwich is huge outlier: deep neural network (DNN) classifier w/o GPU

Escalation to Human Expert

Perfecting software is a long and arduous process

How hard do we have to work on WCA applications?

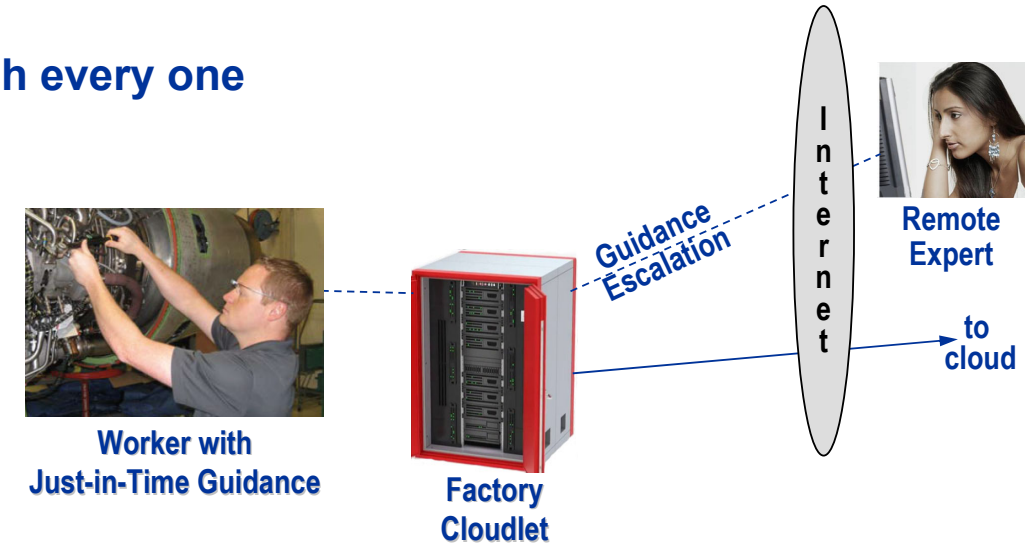
- **Many common errors by humans in a specific task**
- **But also rare errors, not often repeated**
- **Expensive to implement software to catch every one**

Solution: ***Escalate to a human via Zoom***
(exception handling)

Example: factory setting

WCA enables many workers per expert

Contrast with Microsoft 365 Dynamic Remote Assist (one expert per worker)



III. Challenges

2021 to ???

What Happened in the "Lost Decade" ?

2004-2015

Pause for foundational technologies to catch up

Wearable Cognitive Assistance

DeepFace

skype Translator

Siri

Watson

DNN-based Cognitive Algorithms

Cloudlets

Google Glass

Edge Computing

Wearable Hardware

Microsoft Hololens

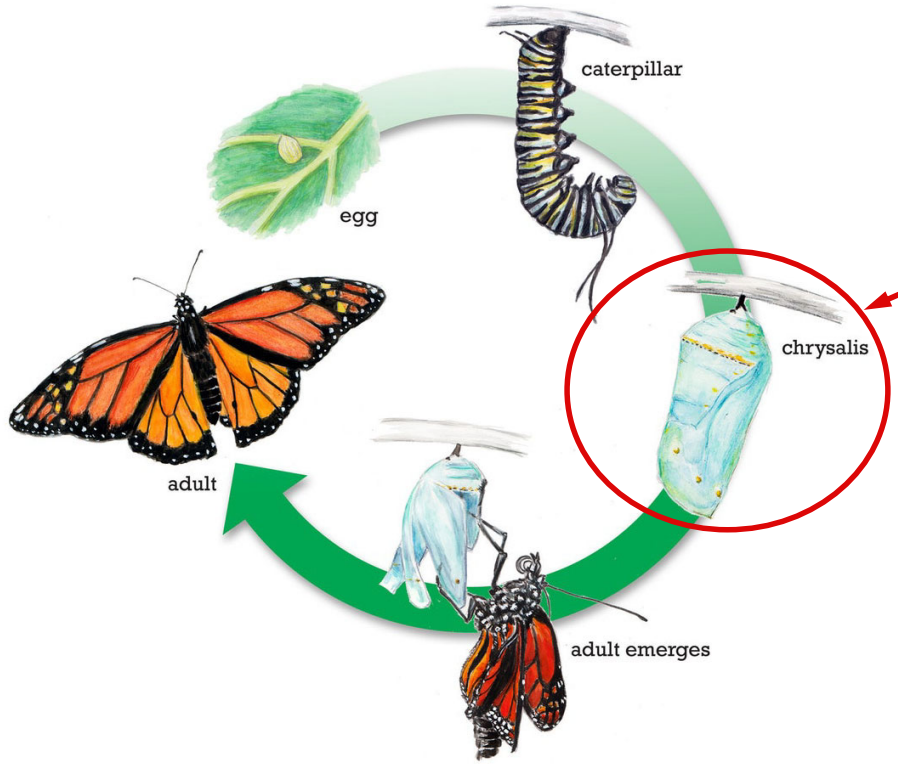
openEDGEcomputing

5G

ETSI MEC

Intel Vaunt

We may need a similar pause (hopefully not a decade!) for foundational technologies to catch up before use of Wearable Cognitive Assistance becomes widespread



Dormant Period

What Needs to Happen?

1. **Widespread use of Edge Computing** (chicken or egg?)
2. **Wearable devices** (chicken or egg?)
3. **5G wireless networks** (happening anyway)
4. **Continued improvements in ML-based computer vision** (happening anyway)
5. ***Software tools to simplify and speed up development*** (who will do this?)

State of WCA Development Today

15 units of course credit for a Master's student in Computer Science
→ ~200 person-hours of time

An already-experienced WCA developer will likely take less time for this task (half? two-thirds?)

For comparison: a high school student can create a decent Web site in a few hours today on her first try

Tools have to be created/improved to yield one to two orders of magnitude speedup

Has to be done in the context of real use cases, not toys

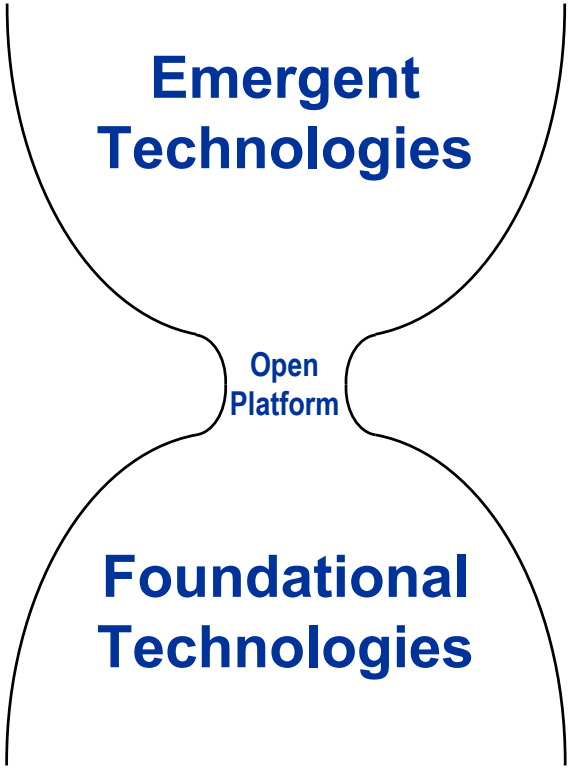
IKEA

≈ 25 parts

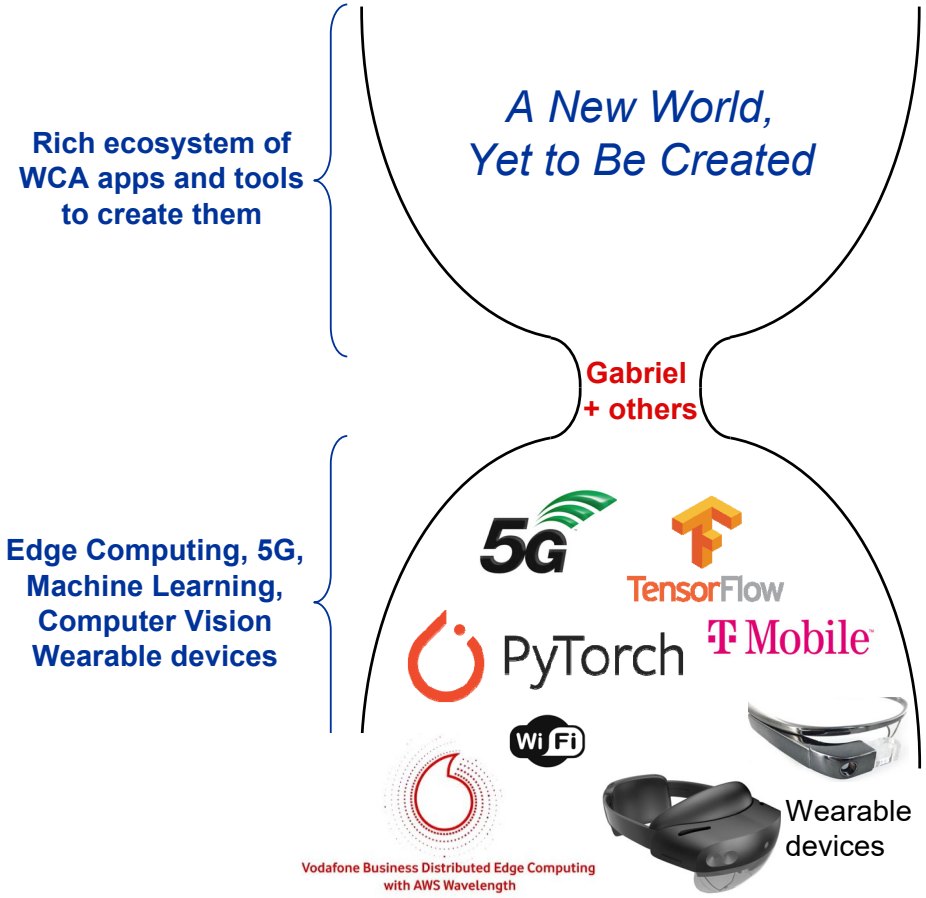
RÅSKOG $29.99

Utility cart, gray-green, 13 3/4x17 3/4x30 3/4 "

Hourglass Metaphor



The Butterfly Waiting to Emerge



Ongoing Partnerships

AutoDesk

- **Digital Triplets for Born-Digital Objects**

Deloitte

- **Customer-driven Use Cases of Wearable Cognitive Assistance**
- **The Smart Factory @ Wichita**

What is a Digital Triplet?

"Digital Twin" → high-fidelity software implementation of a physical object

- **used in simulations, visualizations, etc as stand-in for the real object**
- **extent of fidelity varies, depending on goal of twin**

"Digital Triplet" → highly accurate software detector of a physical object

Our Research Hypothesis

"For born-digital objects, it is feasible to fully automate the creation of digital triplets for wearable cognitive assistance. Zero manual effort is needed to achieve the necessary accuracy."

Closing Thoughts

Humans are the standard against which AI is measured

Evolution took 10^9 years to evolve humans and their specialized neural circuits

We can't wait that long! We want progress in 10 years at most

Wireless edge offload is key to this accelerated progress

Allows real-time use of resources far beyond what could be worn or carried by humans

- **larger**
- **heavier**
- **heat-dissipative**
- **energy-hungry**