Rebooting Computing Summit

-- Event Summary

Omni Shoreham Hotel
Washington, D.C.
December 11 and 12, 2013

Prepared By:
Robin Wedewer
President, The Wedewer Group
4701 Paul Hance Road
Huntingtown, MD 20639
(410) 414-5718
robin@wedewergroup.com

December 2013

Contents

Foreword

The Future Directions Committee (FDC) is a committee of the IEEE Technical Activities Board (TAB). Through volunteers from IEEE's Societies and Councils, FDC seeks to identify multidisciplinary topics in which IEEE can play a unique role for catalyzing and crystallizing goals and activities which increase the efficiency of developing the needed technologies of the future. Rebooting Computing (RC) is an initiative of the FDC which proposes to rethink the computer through a holistic look that addresses all aspects of computing, "from soup to nuts." The RC committee consists of volunteers from seven IEEE Societies/Councils and two professional staff directors. After planning and deliberations over about a year, the RC committee organized a Summit, in December of 2013, bringing together a selection of thought leaders and decision makers from government, industry, and academia, to brainstorm ideas and lay initial foundations for Rebooting Computing. The Summit was not intended as a traditional workshop but rather a true holistic brainstorming exercise relying on the expertise of the attendees to define broad directions inspired by the known challenges of the field and allowing for a degree of imaginative thought for what the future could bring. A summary of the summit follows this foreword. This summary is not intended to be a technical report, nor does it intend to define with scientific or technical rigor what the elements and plans for Rebooting Computing are or should be. Rather, it reflects the discussions that took place at the Summit in a purposefully general manner, leaving the task of defining the specific options, plans, and follow-on activities to the subsequent meetings and reports that will follow. This first Summit was a beginning. The summary does report on the general concepts discussed, the goals, in general terms, that could be sought for the future of computing, along with possible pathways to achieve them. The intention of the RC committee is to engage the technical and scientific communities in a conversation about the best collaborative plans forward and, through IEEE activities of meetings, publications, and related events, gradually provide the key ingredients to accelerate the realization of the future of computing.

Elie Track and Tom Conte

Co-Chairs, IEEE Rebooting Computing

What Is Rebooting Computing?

"Rebooting Computing" is an initiative of the IEEE to identify new directions in computing, a goal intentionally different from refinement of existing directions. The initiative is timely due to the emerging consensus that "Moore's Law is Ending" or at least shifting direction.

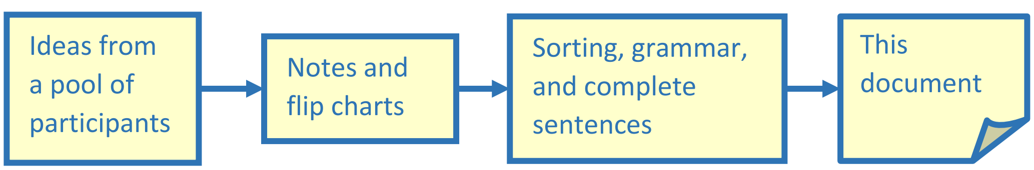
To identify new directions, the Summit used an unusual format. A group of invited thought leaders and decision makers were asked to "check their agendas at the door" and participate in a 1½ day session on new ideas. This is in contrast to the overwhelming majority of workshops and conferences were people work on a topic for months or years and then present a paper with a polished exposition of a specific idea.

The workshop organizers created a specific plan in advance. This included creating a trial technical organization of four "pillars," exploring questions for group discussion, identifying the four plenary speakers, and identifying a facilitator on IEEE staff and a technical writer.

The format creates interesting issues in writing and interpreting a summary of what was essentially a brainstorming session by an authoritative group of participants. The summit included group sessions and subsequent out briefs that gave ideas some degree of group consensus or acceptance, but without attribution to any specific participant. As a result, notes and writing on flip-charts have a degree of authority derived from the participants.

In other contexts, one of the organizers would reorganize the notes into a consistent report. This was not the chosen process because it would replace the authority of the participants with that of the person doing the reorganization.

This meeting summary was created through a process shown below, intended to retain the anonymized positions of the participants. A professional writer used notes she took during the meeting and the flip charts remaining after the meeting and wrote the initial meeting summary document, with instructions not to edit the technical ideas. The summary document was subsequently reviewed by the organizing committee and then offered to the participants for comment.



Readers should make their own decision on the meaning of this document, but we can offer a suggestion. The meeting participants tended to be influential people in government and industry. While they each have specific projects, programs, and views in their "day job," they met as a group and discussed new ideas. Ideas were then shared with other participants and only remained to be written into this document if they received the support of the group. This document might thus give a glimpse of what new ideas might be acceptable to thought leaders in the future.

Details of the Summit

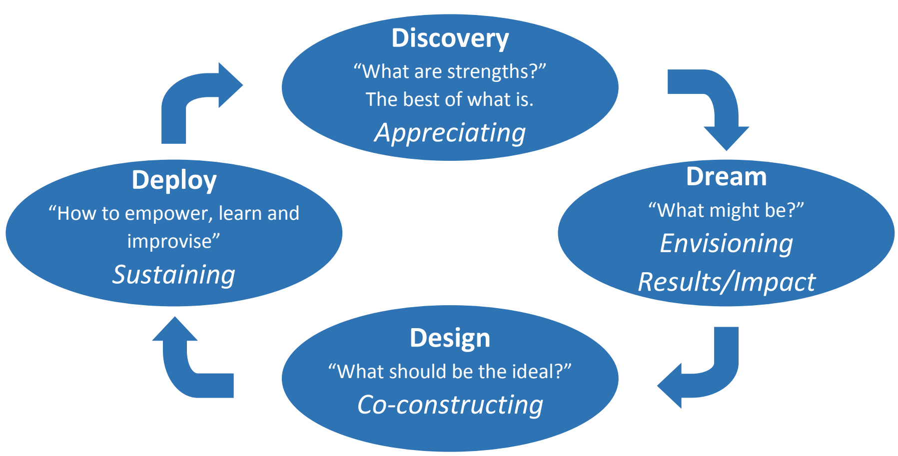
The term "Rebooting Computing" was first coined by Dr. Peter Denning in an earlier project focused on the educational aspect of inspiring the new generation of technologists and innovators who will create the next computing paradigm.

The current IEEE invitation-only summit lasted one and a half days, December 11-13, 2013, in Washington, DC. The discussion portion of the summit used the techniques of Appreciative Inquiry to allow participants to envision the future of computing.

Participants were also encouraged to use science fiction as inspiration to push the limits of today's thinking. Prior to the summit, Brian David Johnson, Intel's resident futurist and a leader in the method of science fiction prototyping, provided a selection of science fiction stories handpicked to best relate to our workshop. They are available at:

- "18% Happier," Adrian Ellis - http://uk.tomorrow-projects.com/2012/04/18-happier/
- "After Science," Karl Schroeder - http://uk.tomorrow-projects.com/2012/09/the-tomorrow-project-anthology-imagining-the-future-and-building-it/
- "Novella: Knights of the Rainbow Table," Cory Doctorow - http://uk.tomorrow-projects.com/2012/02/novella-knights-of-the-rainbow-table/
- "A Long Day," Gerardo Furtado - http://br.tomorrow-projects.com/en/2013/10/a-long-day/
- "For the Benefit of the People," Nathan David Smith - http://us.tomorrow-projects.com/2013/10/for-the-benefit-of-the-people/
- "Sustainable Suburbia," Zoe Calhoun - http://us.tomorrow-projects.com/green-dreams-winners/

The Summit meeting sessions used the Appreciative Inquiry (AI) method. Appreciative Inquiry focuses on strengths and building on those strengths to answer business questions. The process follows a four-D model of change: discovery, dream, design and deploy, as shown below.



Organizers and Participants

Co-Chairs:

Elie Track, President, Council on Superconductivity
Tom Conte, President Elect, Computer Society

Committee Members:

David Atienza, Council on Electronic Design Automation
Jonathan Candelaria, Electron Devices Society
Erik DeBenedictis, Computer Society
Glenn Gulak, Solid State Circuits Society
Scott Holmes, Electron Devices Society
Yung-Hsiang Lu, Computer Society
David Mountain, Electron Devices Society
Oleg Mukhanov, Council on Superconductivity
Vojin Oklobdzijja, Circuits and Systems Society
Angelos Stavrou, Reliability Society
Ian Young, Solid State Circuits Society
Bichlien Hoang, IEEE Future Directions

Facilitator:

Elena Gerstmann, IEEE Enterprise Planning and Development

Participants:

Susan Alexander, IARPA
Geoffrey Brown, National Science Foundation
Jonathan Candelaria, Semiconductor Research Corporation
Colin Cantlie, Defence Research and Development Canada (Government agency)
Andrew Chien, University of Chicago
Robert Colwell, DARPA
Tom Conte, Georgia Institute of Technology
Erik DeBenedictis, Sandia National Labs
Andre DeHon, University of Pennsylvania
Gary Delp, Mayo Clinic /Special Purpose Processor Development Group
Elena Gerstmann, IEEE
Rajesh Gupta, UC San Diego
Robert Hebner, The University of Texas at Austin, Center for Electromechanics
Peter Highnam, IARPA
Bichlien Hoang, IEEE Future Directions

Scott Holmes, Booz Allen Hamilton
Wen-mei Hwu, University of Illinois at Urbana-Champaign
William Joyner, Semiconductor Research Corp
Tracy Kimbrel, National Science Foundation
David Kirk, NVIDIA
Rob Leland, Office of Science and Technology Policy
Konstantin Likharev, Stony Brook University
Yung-Hsiang Lu, Purdue University
Marc Manheimer, IARPA
David McQueeney, IBM
David Mountain, NSA
Oleg Mukhanov, HYPRES, Inc.
Lucy Nowell, U.S. Department of Energy
Yale Patt, The University of Texas at Austin
Massoud Pedram, University of Southern California
Burton Smith, Microsoft Corp.
Marc Snir, Argonne National Laboratory
Thomas Sterling, Indiana University
William Tonti, IEEE NJ
Elie Track, Council on Superconductivity
Brian Van Essen, Lawrence Livermore National Laboratory
Jeffrey Welser, IBM

Speakers

Four plenary speakers provided opening talks the first morning of the Summit, followed by a video conference presentation by Brian David Johnson of Intel.

"Eeyore's Look at the End of Computing" - Bob Colwell, Director, Microsystems Technology Office, DARPA

Referencing the perennially gloomy Eeyore's (from the Winnie-the-Pooh story series by A.A. Milne) view of the world, Bob Colwell hypothesized that the end of Moore's Law will mean the end of computing as we currently know it. The end of Dennard scaling and the constraints of thermal dissipation limits will create a "race to the bottom" with cost the only remaining differentiator. As profit margins plummet, the traditional big money technology companies will abandon investing in new designs. New mobile platform app development cannot support the large companies. Servers will be built from commodity components. And it is unlikely that quantum computing will save the day anytime soon.

But just as even Eeyore had good days, Colwell said that in the absence of a substitute for Moore's Law to drive development, there are developments that can help to move the industry forward. Possibilities included:

- Fix the "brittle SW [*software*] and systems" problem.
- Get serious about security, hackers, etc.
- Get serious about energy efficiency.
- Improve the human/computer interfaces.
- Exploit the Internet of Things.
- Create automatic help systems, such as with car maintenance, imminent battery failure, engine anomalies, home furnace efficiency, etc.
- Apply IBM Watson-like computers to real problems—climate change, energy shortages, overpopulation, pandemics, terrorism, security, water conservation, poverty, education, etc.
- Deal with the fact that the more pervasive a technology is, the more it gets taken for granted.

"Future Technology Directions" - Peter Highnam, IARPA

IARPA's mission is to invest in high-risk/high-payoff research that has the potential to provide the U.S. with an overwhelming intelligence advantage. Peter Highnam asserted, however, that "High-risk/high-payoff is not a free pass for stupidity."

IARPA's approach is to engage the best minds to work on research problems that 1) have clear, measurable, ambitious and credible goals; 2) Employ independent and rigorous test and evaluation; 3) Involve IC partners from inception to finish and 4) Span three to five years.

Highnam provided a brief overview of some of the technologies IARPA is exploring, including the possibility of superconducting computing, nominal system comparison, energy efficient cryogenic memory, cryogenic computing complexity (C3), superconductor circuit fabrication, Trusted Integrated Chips (TIC) program, cortical computing primitives and connectomics.

Highnam invited Summit participants to join and participate in two IARPA-supported crowd-sourced forecasting programs that are showing promise (and are addictive to participants):

- Geopolitical forecasting - http://goodjudgmentproject.com
- Science and technology forecasting - http://scicast.org

"Big Data and International Competition" - Rob Leland, Office of Science Technology Policy

Looking back in history to the 1940s with John von Neumann's contributions to computing that led to the initial "boot up" of computing, Rob Leland identified several conditions that made it possible, including:

- *There was a major conceptual advance.* Easily changed software was combined with fixed hardware to produce a flexible system. And there was a previous theoretical basis (Turing).
- *There was a technology opportunity.* A mature technology base (vacuum tubes), some experimental feedback (Feynman's experience during the war), and the need for invention (memory, layout, system engineering and reliability engineering)
- *There was a sense of urgency*. Lessons of WW2 (role of computing in cryptography, ballistics, first atomic bomb). Cold war driver.

Leland identified several strategic trends today that could become catalysts for a reboot, including:

- *The erosion of Moore's Law is creating a fundamental transition in technology.* In the future, the vast code base will have to work holistically on chips that are parallel processors.
- *The rise of Big Data.* Already huge impact – google search, human genome sequencing, climate data analysis, discovery of Higgs particle. The challenge will be to further leverage data by improving both sensing and processing.
- *International competition is creating a sense of urgency*. The US has historically held the lead, but other major international competitors are striving to close the gap.

Leland concluded by asking questions about the future, including:

- *What makes this urgent?* Is it the erosion of Moore's Law? Is there some other motivation as significant?
- *What is the technology opportunity?* Is it the creation of a new device technology and commodity curve? Is it a major shift in architecture, and if so what is that shift? Is it embracing some alternative style of computing, or some combination of all of these?

- *What is the conceptual breakthrough needed to reboot computing that will be comparable to the conjoining of hardware and software?* Is it the networked world allowing us to do science in a different way? Is it some architecture dramatically different from the von Neumann model? Is it a learning system that requires minimal software? Is it some new mathematical conceptualization of a key problem?

"IBM Research: The Journey to Watson" - David McQueeney, Vice President Software, IBM

Using examples from IBM's experiments with developing a chess-playing computer in 1997 to the more recent Jeopardy-playing Watson machine, David McQueeney explored the possibilities of future computing.

When IBM built a computer that could challenge the reigning World Chess Champion—and win—it was because they were able to build a system that could play using different chess-playing styles. Garry Kasparov, Russian chess Grandmaster and then World Chess Champion, had proven highly skilled at reading and learning from an opponent's playing style. IBM's ability to shift playing style from game-to-game provided a competitive advantage—and one that Kasparov asserted was cheating.

Looking for a new experiment, IBM finally embraced the challenge of building a machine that could play—and win—at Jeopardy. The challenge was much more than a simple question and answer game. It involved the more complex problem of finding solutions.

One of the challenges in developing Watson—the name of the Jeopardy-playing computer—was determining how much data is useful. Although the initial data libraries loaded into Watson were much larger, the final data set was less than 100GB. The metadata generated when all the problem solving techniques were run was far larger. It is the multiplier effect of the metadata that offers the greatest promise to future computing.

McQueeney discussed the possibilities of applying systems such as Watson to real-world problems. The possibilities would move computing into a new era. He described the first era in computing as the tabulating systems era, in which computers were used for the automation of tasks, shifting workers from menial work to more productive activities. The second era was the programmable systems era, in which processes and transactions were automated, enabling global enterprise and empowering individuals with computing power. We are now poised on the next era of computing—the cognitive systems era, in which human capabilities are scaled and magnified.

Such systems could be used to solve problems that plague many areas of society. Four hundred billion dollars are lost to cybercrime each year. There is a 44% misdiagnosis rate for some types of cancer. Only 22% of students worldwide graduate from high school. Cognitive computing could leverage an encyclopedic domain of knowledge, map emergent patterns and connections, analyze conflicting points of view and, ultimately, create new insights and find new value.

"Visions of the Future" - Brian David Johnson, Intel (*via video conference*)

"The future is the result of people who get together with an intent to change the future," asserted Brian David Johnson in his video conference address to Summit participants.

This ability to be agents for change gives Summit participants the opportunity to change IEEE, the industry and, ultimately, society. The ability to change the future is only constrained by the limits of our imaginations. Change comes from moving beyond the question "Can we…?" and toward the question "What can we…?" But perhaps most important is to ask the question "Why?"

Johnson believes strongly in the power of science fiction to prototype the future.

"A good story is about people," said Johnson. "A good science fiction story is about people. A good science fiction story helps imagine the human, social and legal implications of technology."

According to Johnson, science fiction also allows us to explore the future we want and the future we want to avoid. Science fiction also gives us a common language to talk about the future. For example, a story about a synthetic biologist can enable people who are not synthetic biologists to talk about synthetic biology. The story allows readers to imagine the future together.

The stories can empower people by changing the stories they tell themselves about the future, enable them to make different decisions and take new actions. By sharing those stories and having conversations with others, the multiplier effect can be transformative.

Johnson said that if we do our jobs correctly, we can touch the lives of everyone on the planet and that our goal should be to make their lives better.

Pillars of Computing: Setting the Stage for Dialog

Elena Gerstmann launched the discussion with the first of the four Ds in the Appreciative Inquiry process—"Discovery" and appreciating the current reality. Summit participants shared stories about experiences in computing that inspired in them a sense of awe. Participants then said the qualities of those stories were:

- New/novel
- Fascinating
- Pride
- Helpful/useful
- Fast
- Real
- Science fiction
- Beautiful
- Heroic
- Miraculous
- Unexpected
- Complex/simple
- Personal
- Surprise
- Impressive

Prior to the event, summit organizers met to discuss the organizing mechanism for the discussions. They identified three "pillars" of computing that describe very broad, functional areas that will likely serve as the driving forces to advance computing. These forces were: applications, energy efficiency and security.

The summit participants discussed alternative organizing mechanisms and additional pillars. A fourth pillar, human/computer interface, was added and then later folded back into applications for the purposes of the discussion. The group agreed that the security pillar would include trust, privacy, reliability, and robustness.

The final four pillars of computing for the purposes of this summit were:

- Applications
- Human/Computer Interface
- Energy efficiency and
- Security

Undoubtedly the pillars do not exist in isolation and there is considerable overlap. In addition, spanning all four pillars were issues related to who pays for innovation, the Internet of things/pervasiveness, performance and productiveness/ease of programming.

Computing in the Year 2023 or Beyond

The group collectively discussed how far out the group's vision should go. The organizers had developed the program with the presumption of a 10 year timeframe, meaning "2023." This led to group discussion and a show of hands vote. Ten years was not far enough out according to the vote, but the meeting moved on without further discussion. Participants presumably used the timeframe of "2023 and beyond."

Applications and the Human/Computer Interface (HCI)

Note: The two groups addressing applications and the human/computer interface (HCI) were separate on the first day of the discussion, which addressed the possibilities in 2023 and beyond. The groups were combined on the second day to discuss development of a potential roadmap.

The human/computer interface group described a potential future scenario: You wake up in the morning and go downstairs. As you arrive in the kitchen, a computer asks if you want your usual hot Earl Grey tea. You respond in the affirmative and the computer observes that your hip seems to be bothering you and asks if you want a doctor's appointment. After you respond you ask the computer, "Sarah (because the computer has a name!), I have been thinking about buying that house I looked at on the beach. Do you think it's likely to flood in the next 30 years?" The computer considers the question, consults the appropriate data and responds, "It is unlikely that property will flood within the next 30 years."

This scenario describes how computers and humans could potentially interact in the future. It would be analogous to having a best friend. The group envisioned a system that doesn't simply answer questions, but also anticipates needs. Interaction with the computer of the future will be as natural as talking with another human. These future systems will be able to differentiate between people and identify speakers. They can learn and grow.

The group also coined a term to describe how the computer of the future acts: "aug-mentor," a computer mentor that augments human capabilities. It would operate from a corpus of knowledge, including knowledge about the user. Possibly, it could grow with the individual from childhood to adult to old age. Such a machine could help compensate for the physical and mental losses associated with old age or other disabilities.

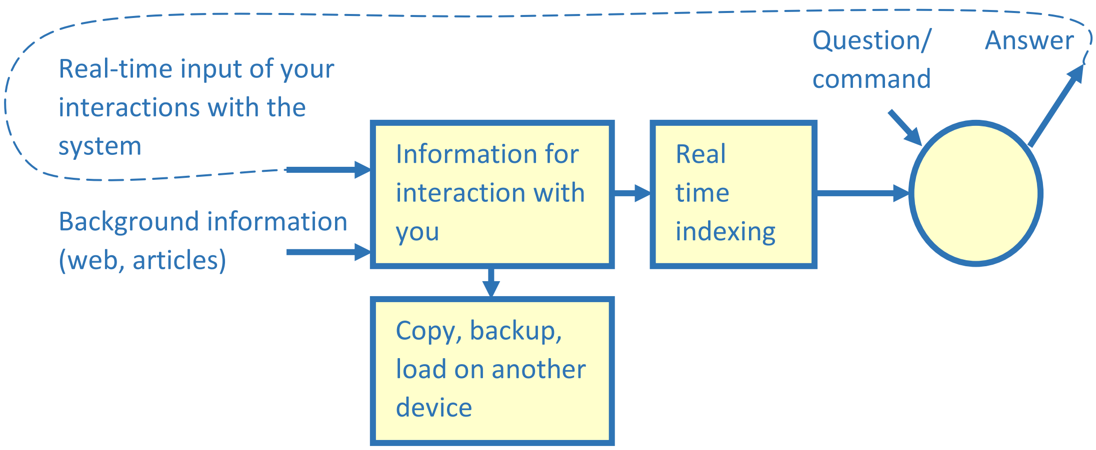
While the vision presented is similar to the influential Apple Knowledge Navigator video circa 1987, the group believes the vision can be more completely realized by 2023. Today, e-mail systems, smart phones, and so forth act as an assistant to us by sorting textual data by time, alphabetical order, and keywords. In contrast, cluster supercomputers like IBM Watson can parse the meaning of words and sentences and attempt to communicate with as people do conversationally with each other. A key aspect of the proposed vision is that Watson-level Natural Language Processing (NLP) would become available in a portable and consumer environment.

The diagram below illustrates information flow. The human interface maintains a body of information for interaction with the user, such as the user's preferences in tea, real estate interests, and information or external origin like web pages and the latest newspapers. This information is indexed like the Watson Jeopardy engine and the user can ask questions of the computer – or give commands.

Past: Fixed dialog (telephone voicemail)

Present: Large knowledge base; indexed offline

Future: Online indexing of knowledge (learning)



Unlike the Watson technology, the proposed interface would index in real time. To play Jeopardy, Watson was loaded with historical data on Jeopardy questions offline. Once the data was indexed over a period of days, Jeopardy questions could be answered in real time by a 2500-node cluster. One aspect of the proposed advance is that the user's evolving history of interaction with the computer and data sources would be indexed in real time, creating a higher computational load that would be feasible only with "rebooted computer" technology.

Another possibility is that information in the machine could potentially be taken out and put into another device or even be transferred temporarily to other machines as the individual moves around in the world. For example, the computer in the car could differentiate between your preferences and your spouse's preferences.

The metric for success is a simple question: Would your 90-year-old grandmother be able to use it effortlessly?

Challenges with such an advanced interface would include knowing when the machine has all the relevant data. To make it work, the machine would need to be capable of understanding fine language nuances, such as the difference between the statements "The boy is ready to eat" and "The chicken is ready to eat."

The applications group also described a potential future in which cars drive themselves, tractors plow fields without human assistance and trenches are dug by big machines without operators. The interaction between the human controlling the machine and the physical environment is replaced by a computer controlling the machine and the physical environment.

Such advanced systems would require a very good understanding and models of human behavior. Progress on this front is currently slow.

The applications group also identified a deficit in seamless computing. Currently the technology is not stable enough. For example, users spend too much time managing, maintaining and upgrading laptop computers. Technology will need to advance to the point at which computers are continually updated and upgraded without user support.

Computers could also be employed for medical diagnoses in a triage system that could immediately separate people who need significant medical attention from those who do not. The "human computer" would deal only with the more complex problems. Robust sensors will be needed for this type of computing.

The human/computer interaction and applications group reported together the second day.

Two types of understanding will be needed to fulfill the vision of the future:

- Understanding of the physical requirements and how to make the right actions based on what the user wants to do and not do and
- Understanding how to use large repositories of data effectively.

Other requirements to fulfill the vision include movement toward:

- A system in which the work is done to a large extent by the system rather than by the programmer
- New hybrid architectures. The expectation is that the von Neumann-type machines will not be replaced, but that hybrid architectures will use different types of circuit architecture and different systems.
- Low precision computation

By End of 2014

- Conduct a workshop drawing conclusions and a vision for HCI as it applies to an assistant for

    a) science research – dealing with formulae and concepts

    b) "intelligence" analysis

    c) medicine, noting that IBM already has business in this segment but that there may be other aspects than the ones IBM is dealing with

    d) data exploration

    e) HCI in the control of robots or other items that impact their environment

By End of 2017

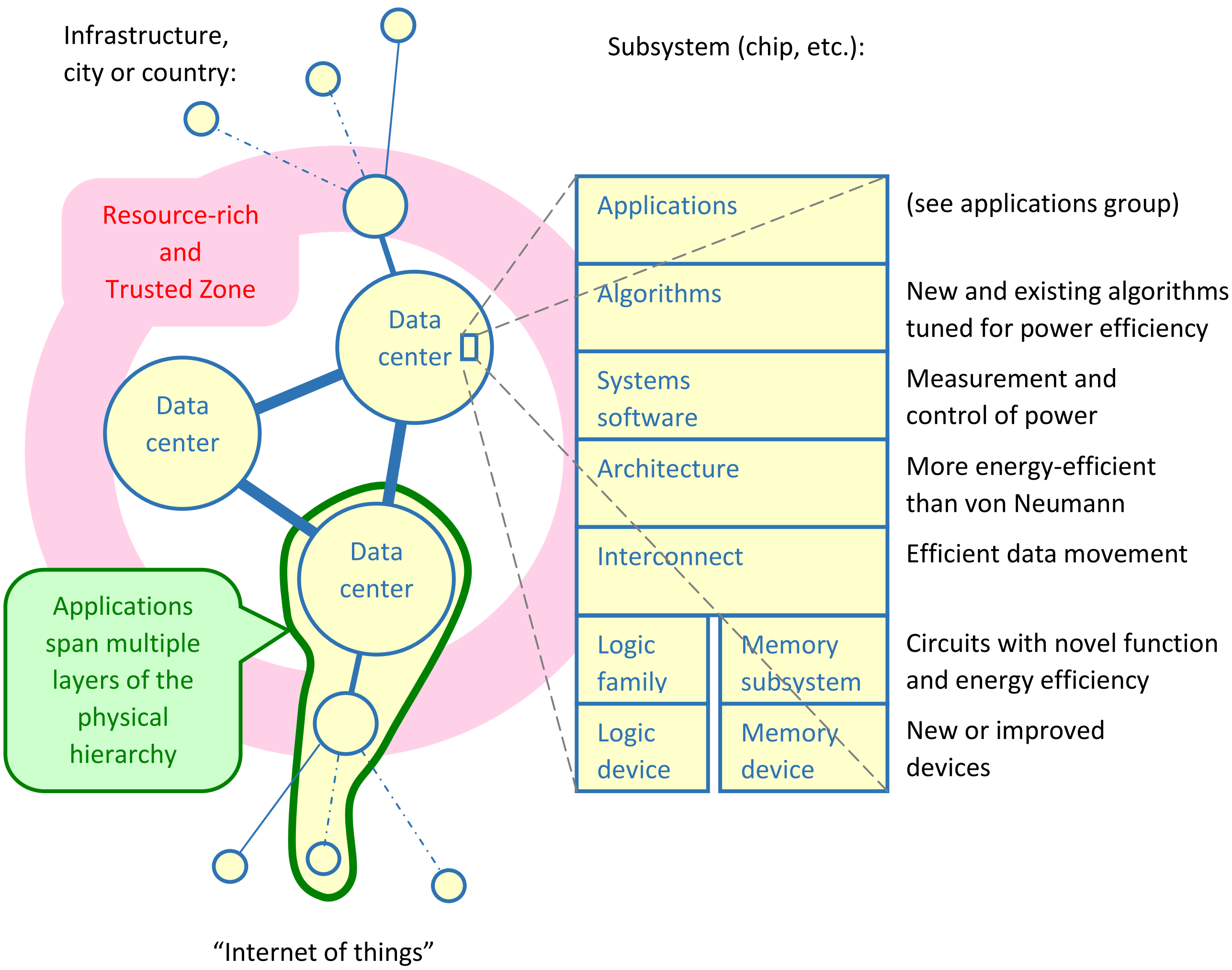
- Completion of seedling research on the basic algorithms that break down human communications into data and functions applicable to items a-e above and act on them. Some demos. This would be analogous to the block diagrams that show how the Watson Jeopardy engine processes Jeopardy questions[1]

[1] Some notes for this section were provided post-event by participants.

Energy Efficiency

In the future, smart devices ("things") at the periphery of the Internet will consume power in proportion to the level of performance that they deliver (known as the energy proportionality) and only as much as their energy harvesting rate or expected charging power allows them—energy source awareness. These devices are trustworthy, responsive and robust. They can access vast resources available in the cloud anytime and anywhere to enhance their own functionality, resiliency and performance. Ideally, all devices are powered by the environment without connection to the power grid.

The information and communication infrastructure (ICI) is comprised of large and powerful datacenters at the core and billions of smart devices at the edges, as illustrated on the left of the diagram below. Much of the time, an edge device will be connected to the core via an intermediate edge infrastructure, which we will call a resource-rich and trusted zone (RT zone). When entering into one of these zones, a device identifies itself and based on its privacy setting and access privileges shares information and data, and most important, its energy consumption and the ability to harvest energy from environment.



At the same time, its energy storage means (e.g., its rechargeable battery) will be fully replenished in a seamless manner. Within each zone, the devices communicate through local networks and consume much less energy than today's cellular networks. When a device is outside any zones, it maintains connectivity and has access to energy in much the same way that today's mobile devices do (i.e., via some type of cellular network, battery and/or energy harvesting means.)

Power to support devices will be highly distributed, local and less grid dependent. The net energy used for computing will be reduced, even as more computing is taking place.

Every location will have connectivity but not all connectivity will be equal. More connectivity will be available where there are large concentrations of people and thus large centers of power. There will be less connectivity—but still connectivity—at the “edges,” where energy harvesting may take place. Harvested energy may be transmitted to local devices.

Consistent with today’s trends towards “Beyond CMOS” and related efforts, the computational infrastructure within each chip will show advances over today’s CMOS technology. As illustrated on the right of the figure above, the technology stack from devices to applications may be redesigned for power efficiency.

The majority of energy today is in data movement rather than logic, so the low-level portion of the technology stack actually comprises several interacting technologies. Energy efficiency for data movement is possible using optics, for example. However, the most energy efficient data movement technologies require signal converters (such as electrical-to-optical and back) that limit their use to long-distance interconnect. To address energy efficiency in the lower-level portions of a computer requires efficiency from the combination of logic and its native interconnects. In this context, certain low-energy transistors are advantageous because the native signals have lower voltage and are thus the overall system is projected to be lower in energy. Logic and memory at both the device and family level will be optimized along with interconnect to produce a more energy-efficient core computational technology.

New computing architectures may emerge that support new types of computing, such as neuromorphic or pattern matching. Offloading computation becomes the norm, not exception. Computation will automatically find the most efficient location for execution.

A new model of parallel processing is expected to emerge to support power-aware applications across the ICI. Data at the edges may be sent to other places for more energy efficient processing, just a today’s parallel computers move data to other computational nodes to achieve higher performance. New systems software and applications will be required that distribute function across heterogeneous elements of varying performance and power efficiency.

Specialized computational devices will be embedded throughout the environment and perhaps even within the human body. Large data centers will continue to exist but perhaps will be geographically situated in cold climate regions for greater cooling efficiency.

Hardware in the future must be more robust and last longer, particularly if it is embedded. New means of hardware installation may be needed. For example, thousands of sensors could be embedded in a single shotgun-like blast. This type of ubiquitous hardware must be capable of failing in place without harm to the environment. It may be recyclable and biodegradable.

Success is demonstrated by the fact that:

- Users hardly ever have to charge devices.
- The phrases “Can you hear me now?” and “Can you see me now?” are forgotten.
- Users never have to think about synching devices.

- Users never have to ask for updates, because they are always taking place.
- Information is automated. For example, when going to a meeting, a user's device will automatically load directions, contact information, needed documents, etc. By preloading information, the device can choose more energy-efficient methods than today's on-demand methods that may be much less efficient.
- Devices are context aware and disable functions that are inappropriate or inefficient. For example, in a movie theater, a phone call cannot be answered. As a result, the mobile phone stops (or reduces the frequency) communication with towers.

Achieving this vision and the proposed ICI architecture requires:

1. Development of a transformative multidisciplinary theory of design, optimization, and runtime management to improve the ICI in light of the emerging new devices, human user interfaces, and computing/storage/networking technologies and develop scientific foundations and practical solutions for achieving energy proportionality and energy source awareness in the devices, the edge infrastructure, and the internet core.
2. Design of the architectural blueprints of different classes of new devices (deeply embedded application-specific devices, to portable devices for general use, to smart sensors and data aggregators) with the requisite (new) user interfaces while realizing energy proportionality and energy source awareness.
3. Development of a hardware/software exploration platform that will enable system designers and/or application developers to explore the tradeoff space of performance, energy efficiency, trustworthiness, and cost early in the design process.

In particular, some of the needed technologies are:

1. A theory and principles of designing perfect systems from imperfect components (e.g., energy proportional systems from non-energy-proportional components, reliable systems from non-reliable components).
2. Design tools and technologies that can derive the 'optimal' meta-architecture for a class of new devices based on factors such as a desired level of performance and security, energy proportionality, and energy source awareness. The goal is to exploit heterogeneity when useful but avoid excessive proliferation of hardware solutions (i.e., adopt a platform-based design approach where each platform architecture targets a wide range of applications). New computational paradigms such as approximate computing, neuro-fuzzy classifiers can be explored depending on the target application and system.
3. Runtime environments that enable adaptation of a device's operating point (in the energy-efficiency vs. performance vs. trust space) to varying environmental and workload conditions and service level agreements.
4. A heterogeneous network infrastructure that supports very high bandwidth wireless communication and networking, high end-to-end task throughput, and low response time in the RT zones. The interaction with heterogeneous networks (in particular cellular networks with pico-and femto base stations, the use of distributed antenna architectures, and similar modern PHY-layer aspects) must also be investigated.
5. Wireless (fast) power charging capability within the RT zones. This should be done in light of the type of energy storage element used in each device and considering the aging effects and safety issues.

6. Realizing energy efficiency gains of new architectural/circuit solutions (including static provisioning of heterogeneous compute/storage resources and dynamic management of their power/performance states), device engineering (e.g., optimizing new sub-10nm multi-gate CMOS devices), advanced heat removal technologies (which would allow us to tolerate higher power densities).[2]

By the end of 2014:

Organize workshops to start the development of

- efficient networks and highly distributed computation environments
- programming models for offloadable applications based on energy resources
- virtualization of networks to make task migration easy
- hardware components that have low sleep power and are energy proportional
- These workshops should include people developing pervasive computing systems

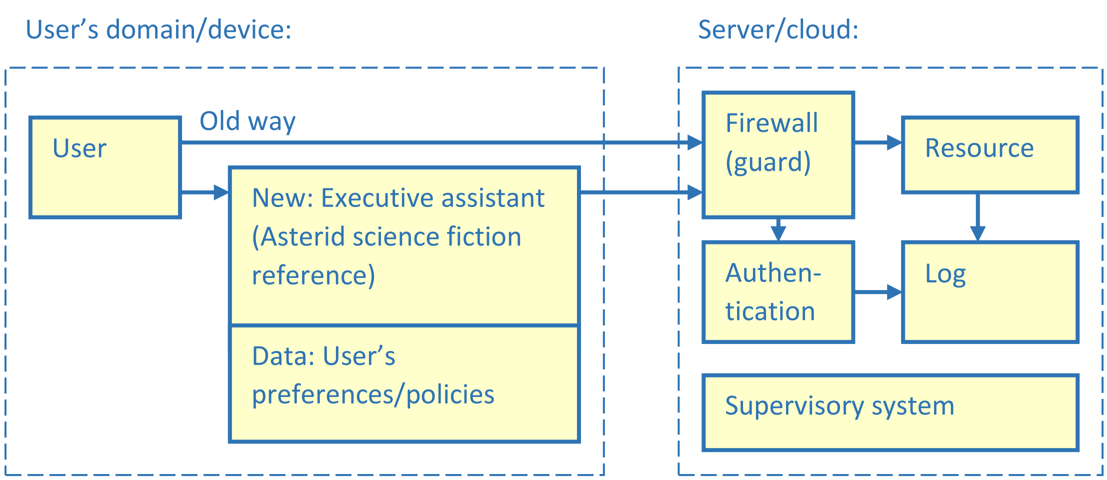
By the end of 2017

Standards for processor, network, and programming models in highly distributed environment. New design of data centers that are energy-proportional. Solutions to retrofit existing data centers.

Security

In the year 2023 and beyond, the concepts of security and trust are closely intertwined. Computer systems will be capable of doing what a human individual can do if that person had all the information and understanding needed to make an informed decision and actions. Systems follow the executive assistant mode, as shown below. The assistant knows when to follow rules, when to break the rules, and when to ask a human operator for more information. Trust is implicit in this new model of security because users trust that the device has his or her best interests at heart and trust the device to do the right thing.

There were two science fiction references. The first was "Limit of Vision" by Linda Nagata, a book where the human characters were augmented with artificial neurons called "asterids" that bonded to their skin



2 Some notes for this section were provided post-event by participants.

and interacted with their brains to enhance cognitive ability. The idea was that the asterids represented the tight bonding of the executive assistant to the human, with the asterids using their information-processing capabilities to enhance security. A second reference was Polar Express movie and the uncanny valley feeling that it invokes. The primitive animation in the Polar Express movie is known to invoke a feeling of disquiet in viewers (this scenario could connect to computer security in several ways, but the editor does not know what was intended).

The computer system in 2023 works correctly at all times despite possible hardware faults, software errors and in the face of malicious interference. When small errors accumulate, the system informs the user that attention and maintenance are required.

Systems are expected to be designed with special attention to mitigating high consequence failures. The consequences of failing to identify mistrust vastly outweighs the consequence of correctly trusting something or transaction, as evidenced by the Toyota unintended acceleration issue some years back. However, cascading failures may be avoided with separate supervisory systems, as shown in the figure. These supervisory systems could be based on spatial/temporal clues and provable methods of assuring safety and progress. For example, a dishwasher controller could be programmed with an invariant such as requiring soap before running.

Furthermore, the computer will durable, with reliability at the level of an appliance (refrigerator).

Success is demonstrated by the fact that:

- In the human/computer interface tradeoff, humans win.
- Antivirus companies are out of business.
- Liability phrases are left out of contracts.
- If the computer fails, users blame themselves and not device.
- Rules are applied.
- News articles report that incidents of data theft have dropped to zero.
- Users are no longer annoyed by gadgets asking incessant questions.
- Criminals give up trying to attack computers and go find different targets.
- Federal funding for security research has dropped by 90 percent.
- The following policy issues are expected to be resolved by 2023:
    - Issues of agency
    - Issues of public privacy - tracking
    - Opt in Opt out
    - Hermit vs oversharing

Dream state of 2023

The dream state of 2023 will start with the completion of ideas already proposed, such as Usable Security: How to Get It, Butler Lampson, CACM, Vol. 52 No. 11, Pages 25-27 http://cacm.acm.org/magazines/2009/11/48419-usable-security-how-to-get-it/fulltext. The diagram above includes pertinent aspects of Lampson's vision on the server side.

The security system will include outgrowths of today's Trusted Platform Modules (TPMs), including immutable roots of trust and a chain of trust. The security system will authenticate and check with state- and rule- based invariants. Processors will require certification.

Achieving this vision will include the following technical advances:

1. Immutable roots of trust, including:
   - Hardware Trusted Platform Module (TPM)
   - Authentication of Trusted agent (a superset of identification)
   - Capability granted associated with authentication

2. Validation based on spatial/temporal information

3. Calculus of proof of safety and progress

4. Supervision by an alternatively implemented function

5. Validation of operations will be required, one such way follows this flow chart:

6. Hardware-implemented functions
   - Processors that bounds check the stack/ call/return
   - Immutable constraints
   - Physical monitoring & intervention

7. Limited complexity separate validated checker

Before Dec 31, 2014

An independent body develops to collect data about computer crashes and provide a level-of-trust rating for computers. This would come from targeted teams collecting data. Communications of the rating would not be like a cryptic energy star rating, but rather a way to expose the level of certification *a la* car safety or movie ratings.

Before Dec 31, 2017

Two independent but interoperable implementations of the rule based system leading to a progressive standards effort.[3]

[3] Some notes for this section were provided post-event by participants.

Summit Wrap-Up and Comments

At the conclusion of the Summit, participants were asked to provide feedback on three major questions:

- How do we broaden the conversation with people who did not attend but who could contribute?
- What comes next?
- What can you do to further the next rebooting of computing?

Participants contributed the following remarks:

- Several people raised concerns about the lack of age, geographic, gender and even technological diversity. Suggestions to remedy this for the next Summit included:
  - Participants from this Rebooting Computing Summit can be charged with inviting and bringing a younger colleague to the next event.
  - Create and distribute the format for a five-hour dialogue on Rebooting Computing that could be hosted by chapters and other groups, with the results reported back to the organizing committee.
  - Engage IEEE student branches. Faculty advisors could be encouraged to steer students who need to complete fourth-year projects toward activities that support the Rebooting Computing effort.
  - Reach out to people in related communities to participate. For example, researchers in the energy community could help with problems related to distributed production and control. Telecommunications is also set to undergo a sea change as a result of advances in computing. Other industries could include mining.
  - Consider inclusion in the next Summit of people with greater understanding of forces outside of computing that may have an impact on the future. Economists, for example, could provide information about economic constraints. "We need to follow the money."
- Provide a longer lead-time to put the next event on participants' calendars and block out more time for deeper dialogue.
- Provide greater focus to the next Summit so that there is a more coherent vision and message. Have clear answers to the questions: What do organizers want as a result of Rebooting Computing? Will it be a change of behavior? Funding? Change in curriculum?
- Consider expanding the timeframe beyond 2023—or at least providing greater clarification of scope for dialog exercises.
- Summit participants could work within IEEE to create and host events and mini-workshops, contests and challenges to describe a problem to be solved.
- IEEE should consider partners outside of IEEE to broaden the scope of the dialog and increase the momentum. One possibility could be a government agency that could also provide support through grant funding.
- Consider assembling a task force to create a technology road map for the future.
- Establish Rebooting Computing as a recognizable brand.