RCS 4
4th Rebooting Computing Summit
"Roadmapping the Future of Computing"
Summary Report

Washington Hilton
Washington, DC
December 9-11, 2015

Prepared By:
IEEE Rebooting Computing Committee

http://rebootingcomputing.ieee.org/

January 2016

Contents

Foreword

"IEEE Rebooting Computing" is an inter-society initiative of the IEEE Future Directions Committee to identify future trends in the technology of computing, a goal which is intentionally distinct from refinement of present-day trends. The initiative is timely due to the emerging consensus that the primary technology driver for 5 decades, Moore's Law for scaling of integrated circuits, is finally ending. How can we continue to project further improvements in computing performance in coming decades? We need to review the entire basis for computer technology, starting over again with a new set of foundations (hence "Rebooting Computing"). The need for new approaches has also been recognized by other organizations. The semiconductor industry's International Technology Roadmap for Semiconductors is now ITRS 2.0, with a new thrust that goes beyond Moore's Law scaling. Furthermore, the US Government has initiated several major programs in future computing, including the National Strategic Computing Initiative (NSCI), as well as a nanotechnology-inspired Grand Challenge for Future Computing.

The 1st Rebooting Computing Summit in Dec. 2013 (RCS 1), brought together decision makers from government, industry, and academia, to lay initial foundations for Rebooting Computing. This generated a vision of future computing based on three pillars of Energy Efficiency, Security, and Applications. RCS 2 in May 2014 focused on four technologies for further discussion, a mainstream approach of Augmenting CMOS, together with alternative approaches of Neuromorphic, Approximate, and Adiabatic Computing. RCS 3 in Oct. 2014 further addressed the topics of Parallelism, Security, Random Computing, and Human-Computer Interface. RCS 4, held in Washington DC, Dec. 9-11, 2015, continued this effort, elaborating four complementary tracks for enhancing future computer performance, consisting of Probabilistic, Neuromorphic, and Superconducting Computing, as well as Beyond CMOS System Integration.

In order to implement this program more effectively, Rebooting Computing executed a Memorandum of Understanding with ITRS in 2015, for the two organizations to work together to achieve common goals of advancing the future of computer technology. As part of this joint program, RC participated in ITRS Workshops in February and July 2015, and ITRS played a key role in RCS 4.

In addition, RC sponsored a special issue of IEEE Computer Magazine in December 2015, with seven articles on the theme of Rebooting Computing. These articles cover many of the same themes of RCS 4, and we recommend them as further reading.

Finally, the RC Web Portal (http://rebootingcomputing.ieee.org) contains information and presentations from all of the RC Summits, as well as ongoing programs, feature articles and videos, and news related to Rebooting Computing

Elie Track and *Tom Conte,* Co-Chairs, IEEE Rebooting Computing

Erik DeBenedictis and David Mountain, Co-Chairs, RCS 4

IEEE Rebooting Computing Committee

The RC Summits were sponsored and organized by the RC Committee, which consists of volunteers from nine IEEE Societies/Councils and two professional IEEE staff directors. Members and Participating Societies are listed below.

Participating IEEE Societies and Councils

Circuits and Systems Society (CAS), Council on Electronic Design Automation (CEDA), Computer Society (CS), Council on Superconductivity (CSC), Electron Devices Society (EDS), Magnetics Society (MAG), Nanotechnology Council (NTC), Reliability Society (RS) and Solid-State Circuits Society (SSC). Also, coordination with International Technology Roadmap for Semiconductors (ITRS) and Semiconductor Research Corp. (SRC).

Co-Chairs of RC Committee:

- Elie K. Track (CSC)
- Tom Conte (CS)

Other Committee Members:

- Dan Allwood (MAG)
- Neal Anderson (NTC)
- David Atienza (CEDA)
- Jesse Beu (CS)
- Jonathan Candelaria (EDS)
- Erik DeBenedictis (CS)
- Paolo Gargini (ITRS)
- Glen Gulak (SSC)
- Steve Hillenius (SRC)
- Bichlien Hoang, RC Program Director
- Scott Holmes (EDS, CSC)
- Subramanian (Subu) Iyer (EDS, CPMT, SSC)
- Alan M. Kadin (CSC)
- Arvind Kumar (EDS)
- Yung-Hsiang Lu (CS)
- David Mountain (EDS, CS)
- Oleg Mukhanov (CSC)
- Vojin G. Oklobdzijja (CAS)
- Angelos Stavrou (RS),
- Bill Tonti (RS), IEEE Future Directions
- R. Stanley Williams (EDS)
- Ian Young (SSCS)

Previous RC Summits

RCS 1: Future Vision and Pillars of Computing

The first Rebooting Computing Summit was held at the Omni Shoreham Hotel in Washington, DC, Dec. 11-12, 2013. This invited 37 leaders in various fields in computers and electronics, from industry, academia, and government, and included several plenary talks and smaller breakout discussion groups. The summary is available online at http://rebootingcomputing.ieee.org/RCS1.pdf. The consensus was that there is indeed a need to "reboot computing" in order to continue to improve system performance into the future. A future vision and three primary pillars of future computing were identified.

Future Vision of Intelligent Mobile Assistant

One future vision for 2023 suggested in RCS 1 consisted of ubiquitous computing that is fully integrated into the lives of people at all levels of society. One can think of future generations of smartphones and networked sensors having broadband wireless links with the Internet and with large computing engines in "the Cloud." More specifically, one may envision a wireless "intelligent automated assistant" that would understand spoken commands, access information on the Internet, and enable multimedia exchange in a flexible, adaptive manner, all the while maintaining data security and limiting the use of electric power. And of course, such a wireless assistant should also be small and inexpensive. Such a combination of attributes would be enormously powerful in society, but are not yet quite achievable for the current stage of computer technology.
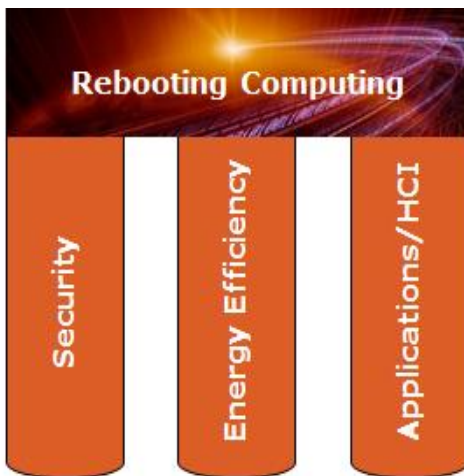
Three Pillars of Future Computing

RCS 1 further identified 3 "pillars" of future computing that are necessary to achieve this vision: Energy Efficiency, Security, and Human-Computer Interface.

Human/Computer Interface and Applications

A better Human/Computer Interface (HCI) is needed that makes more efficient use of natural human input/output systems, particularly for small mobile units. Improved natural language processing is just one key example. While there is a long history of text I/O, this is not really optimal. Wearable computers analogous to Google Glass may provide a glimpse into future capabilities.

Energy Efficiency

The small wireless units operate on battery power, and it is essential that they consume as little power as possible, so that the recharging is relatively infrequent. Some computing can be shifted to "the cloud," but enhanced performance requires substantial improvements in energy efficiency. In contrast, the data centers and servers in the cloud are wired, but their power consumption can be quite large, so that electricity bills are a major cost of operation. Improved energy efficiency is critical here, as well.

Security

With data moving freely between wireless units and computers in the cloud, encryption and computer security are critical if users can expect to operate without fear of data diversion and eavesdropping.

RCS 2: Future Computer Technology – The End of Moore's Law?

RCS 2 consisted of a 3-day workshop May 14 - 16, 2014, at the Chaminade in Santa Cruz, CA. The summary is available online at http://sites.ieee.org/rcsummit/rcs2/. The main theme of RCS 2 was on mainstream and alternative computing technologies for future computing, with four possible approaches identified. The format was similar to that for RCS 1, with a set of four plenary talks, followed by four parallel breakout groups culminating in outbrief presentations and concluding in a final plenary discussion. The primary conclusions were that focusing on energy efficiency and parallelism will be necessary to achieve the goals of future computing, with complementary roles for both mainstream and alternative technologies.

Augmenting CMOS

Silicon CMOS circuits have been the central technology of the digital revolution for 40 years, and the performance of CMOS devices and systems have been following Moore's law (doubling in performance every year or two) for the past several decades, together with device scaling to smaller dimensions and integration to larger scales. CMOS appears to be reaching physical limits, including size and power density, but there is presently no technology available that can take its place. How should CMOS be augmented with integration of new materials, devices, logic, and system design, in order to extend enhancement of computer performance for the next decade or more? This approach strongly overlaps with the semiconductor industry roadmap (ITRS), so RCS 2 coordinated this topic with ITRS.

Neuromorphic Computing

A brain is constructed from slow, non-uniform, unreliable devices on the 10 μm scale, yet its computational performance exceeds that of the best supercomputers in many respects, with much lower power dissipation. What can this teach us about the next generation of electronic computers? Neuromorphic computing studies the organization of the brain (neurons, connecting synapses, hierarchies and levels of abstraction, etc.) to identify those features (massive device parallelism, adaptive circuitry, content addressable distributed memory) that may be emulated in electronic circuits. The goal is to construct a new class of computers that combines the best features of both electronics and brains.

Approximate Computing

Historically computing hardware and software were designed for numerical calculations requiring a high degree of precision, such as 32 bits. But many present applications (such as image processing and data mining) do not require an exact answer; they just need a sufficiently good answer quickly. Furthermore, conventional logic circuits are highly sensitive to bit errors, which are to be avoided at all cost. But as devices get smaller and their counts get larger, the likelihood of random errors increases. Approximate computing represents a variety of software and hardware approaches that seek to trade off accuracy for speed, efficiency, and error-tolerance.

Adiabatic/Reversible Computing

One of the primary sources of power dissipation in digital circuits is associated with switching of transistors and other elements. The basic binary switching energy is typically far larger than the fundamental limit ~kT, and much of the energy is effectively wasted. Adiabatic and reversible computing describe a class of approaches to reducing power dissipation on the circuit level by minimizing and reusing switching energy, and applying supply voltages only when necessary.

RCS 3: Rethinking Structures of Computation

RCS 3 consisted of a 3-day workshop October 23-24, 2014, at the Hilton in Santa Cruz, CA. The summary is available online at http://rebootingcomputing.ieee.org/rc-summits/rcs3. RCS 3 addressed the theme of "Rethinking Structures of Computation", focusing on software aspects including HCI, Random/Approximate Computing, Parallelism, and Security. These are some of the conclusions.

4th Generation Computing

Computing is entering a new generation, characterized by world-wide networks coupling the Cloud with a variety of personal devices and sensors in a seamless web of information and communication. This is more than just the Internet or the Internet of Things; it also encompasses Big Data and financial networks. This presents new challenges, and will require new sets of tools on every level, with contributions needed from industry, academia, and government.

Dynamic Security for Distributed Systems

One key challenge is in the area of computer security. Current security systems represent a patchwork of solutions for different kinds of systems. What is needed is a universal, forward-looking set of protocols and standards that can apply to all parts of the distributed network, with a combination of simple hardware and software building blocks. These must also be dynamic and capable of being updated to reflect newly recognized system features and threats.

Ubiquitous Heterogeneous Parallelism

Parallelism will be a central feature of future computing, even if an alternative technology should take hold. This will be massive parallelism for high-performance computing, but even personal devices will be parallel in nature. In many cases, these parallel processors and memories will be heterogeneous and distributed. This represents a strikingly different paradigm than the conventional von Neumann machine, and may require rethinking many of the foundations of computer science.

Adaptive Programming

High-level programming needs to operate efficiently on a wide variety of platforms. This may require providing high-level information (e.g., on parallelism, approximation, memory allocation, etc.) that can be properly optimized by the compiler or system software. Furthermore, the system should learn to become more efficient based on the results of repeated operations and appropriate user feedback, i.e., it should exhibit long-term adaptive learning.

Vision of Future Human-Centric Computing

Prof. Greg Abowd (Georgia Tech) identified the new generation of Complementary Computing, where the boundary between computer and human is blurred. Others have asserted that a personal computing device should be programmed to act in the best interests of each individual. Finally, for an optimum human-centric computing system, the computing devices should be adapted to the needs and preferences of the individual human user, rather than the human adapting to the needs of the computer or the programmer. We have already seen the start of this revolution, but the ending is still being imagined.

RCS 4 Brief Meeting Summary

The fourth IEEE Rebooting Computing Summit (RCS 4), organized by the Rebooting Computing Initiative of the IEEE Future Directions Committee (FDC), was held on December 9-11, 2015 at the Washington Hilton, Washington, DC. RCS 4 included 73 invited participants from industry, academia, and government. RCS 4 built on RCS 1, 2, and 3, held in 2013 and 2014, with a theme of **"Roadmapping the Future of Computing: Discovering How We May Compute."** The agenda is shown in Appendix A.

RCS 4 Plenary Talks

RCS 4 began with introductions by RC co-chairs Tom Conte and Elie Track, and RCS 4 Co-Chairs, Erik DeBenedictis and David Mountain. The Summit was organized around 4 Technical Tracks, consisting of 3 primary tracks and a 4th extra track, with invited talks as shown below. The slides from these talks are available on the RC Web Portal http://rebootingcomputing.ieee.org.

Track 1: Approximate/Probabilistic Computing

> Laura Monroe, Los Alamos – Probabilistic and Approximate Computing
> Santosh Khasanvis, BlueRiSC – Architecting for Nanoscale Causal Intelligence

Track 2: Extending Moore's Law

> Kirk Bresniker, Hewlett Packard Labs – Memory Abundance Computing
> Philip Wong, Stanford – Computing Performance – N3XT 1000X
> Additional talks by Ian Young, Suman Datta, Matt Marinella, and Eli Yablonovitch

Track 3: Neuromorphic Computing/Sensible Machines

> Stan Williams, Hewlett Packard Labs – Sensible Machine Grand Challenge
> David Mountain, NSA – Neuromorphic Computing for NSA Applications
> Additional talk by John Paul Strachan

Extra Track 4: Superconducting Computing

> Marc Manheimer, IARPA – Cryogenic Computing Complexity Program

Reviews of Other Future Computing R&D Programs

The Summit included brief overviews of a range of other Future Computing programs sponsored by government and industrial consortia: ITRS 2.0, SRC, NSCI, OSTP Grand Challenge, DARPA, and IARPA

Poster Session

A Poster Session was held with 13 posters covering a wide range of topics related to these tracks and initiatives. See Appendix F for the Poster Abstracts.

IEEE Competition for Low-Power Image Recognition

Purdue Prof. Yung-Hsiang Lu described an IEEE prize competition, focusing on Low-Power Image Recognition using a mobile device, held in 2015 [Lu poster]. This involved presentation of a set of test images to the device, and a limited time to accurately identify the images. This will be held again in 2016; see http://lpirc.net/ for details.

Sensible Machine Grand Challenge After-Session

Finally, after the formal end of RCS 4 on Dec. 11th, a special meeting was held to continue discussion on the Sensible Machines Grand Challenge.

While the various tracks featured quite different approaches for Rebooting Computing, there was general agreement that there may be an important role for all of these in different parts of future computing technology. Exponential improvement in computing performance may continue, but not via a single transistor scaling rule as in Moore's Law in the past.

Technical Summary of RCS 4

There is a widespread concern that the traditional rate of improvement for mainstream computer technology (transistors and the von Neumann computer architecture/microprocessor) is in jeopardy, but there is hope that new approaches to computing can keep growth rates at historical levels. This section organizes ideas on this topic that were presented at RCS 4.

RCS 4 confirmed that roadmaps for transistors and the von Neumann computer architecture are essentially on track for about the next decade, with RCS 4 also giving considerably more clarity to some of the new approaches expected to dominate in the longer term. In summary, the semiconductor industry will drive transistors to a state of high maturity over the next decade while starting to manufacture initial versions of new non-transistor devices for the era beyond. The new devices are expected to support a different mix of computing capabilities, following evolving trends in the types of problems people want to solve.

The group of research interests represented at RCS 4 may collectively "reboot computing" by augmenting transistors with new devices that have both state (memory) and an energy efficient computational capability, and complemented by new general-purpose architectures that have been inspired from the brain. This new approach would seem consistent with existing industry plans, yet seems to be more ambitious and highlights a need for further research. In particular, co-design activities will become more important – iteratively improving algorithms, architectures, and technologies to provide improvements in power, performance, and cost at the application level over time.

Multiple Paths to the Future

The organizers structured the meeting around multiple alternative paths or road maps for the future of computing. As illustrated in Figure 1, the computer industry developed a stack of mutually-supporting technologies that have grown as a group since the 1940s. Continued growth will require adding some technology to the stack, but the new technology could appear at various levels. The organizing theme for RCS 4 is that new technology will be added at different levels and yield several viable solutions. It seems likely that today's CMOS-microprocessor systems will persist over the long term by addition of improved transistors and transition to 3D, but one or more of the other approaches may emerge and be economically successful as well. The task of RCS 4 was primarily to present and discuss the most promising alternative approaches.
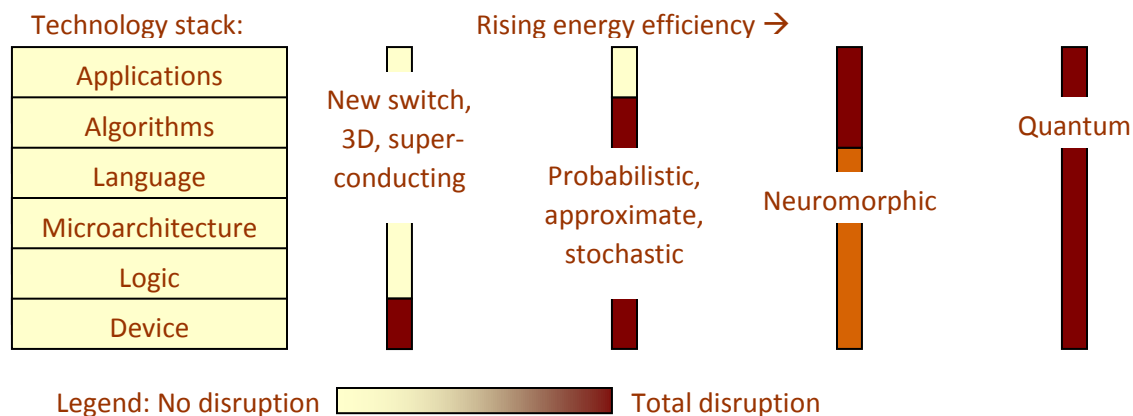

Figure 1: Multiple paths to the future in computing

RCS 4 included consideration of a variety of new technology approaches, including new switches and 3D architectures, superconducting, probabilistic, approximate, and neuromorphic computing. While Quantum computing has great potential, it was only mentioned briefly by the director of IARPA due to its low maturity.

Continued Evolution of Transistors (Track 2)

Paolo Gargini, chairman of the International Technology Roadmap for Semiconductors 2.0 (ITRS 2.0) provided a vision for transistor evolution at RCS 4 [Gargini Wed 10:45] based on papers at the IEDM conference at the same hotel earlier in the week.

The concern about transistor evolution focuses on the energy per Boolean logic operation in a computer, which is dependent on supply voltage V and wire capacitance C. The energy of a Boolean operation can be represented for the purposes of this section as CV^2, or the product of capacitance and the square of voltage. Figure 2 shows time graphs of C (red), V^2 (green), and their product CV^2 (blue), where the product is shown as the sum of the two graphs on a logarithmic scale.
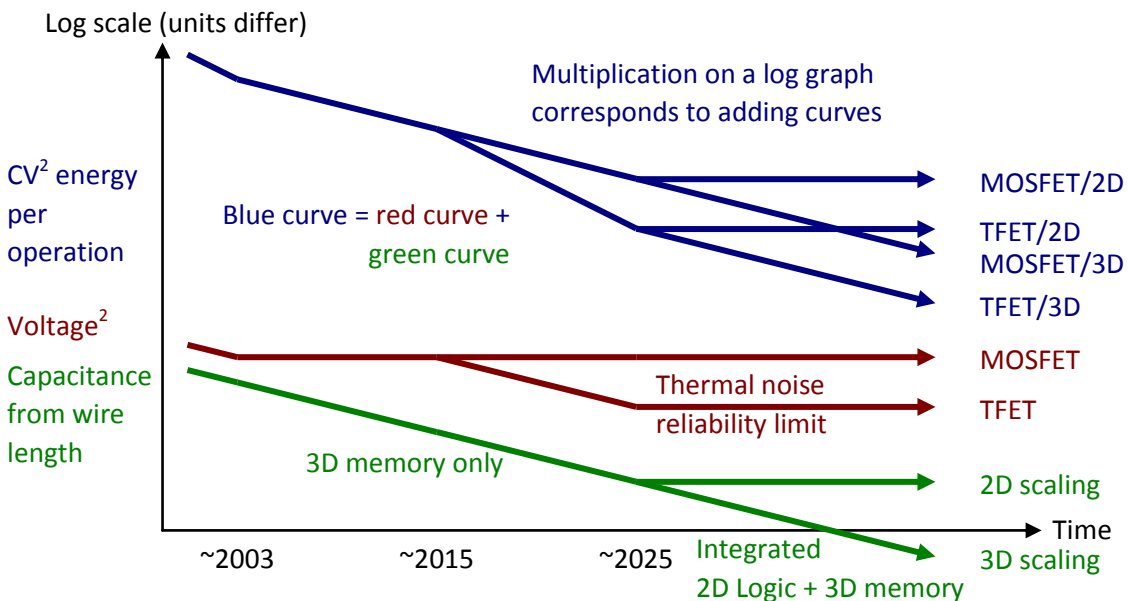
Figure 2: Energy per operation based on MOSFET, TFET (milliVolt switch), and 2D/3D

The red curve for V^2 in Figure 2 shows the scaling or time evolution of supply voltage in integrated circuits, with a potential split ~2015 (i. e. now) due to the development of a new transistor type (to be described below).

The green curve for C shows wire capacitance as the number of devices on a chip increases. Lower capacitance results from shortening wires due to a rising number of devices on chips of constant size, but device shrinkage is expected to end around 2025. The green curve thus shows capacitance flat-lining for the current 2D scaling scenario, but scaling could continue if 3D manufacture becomes practical because the tighter packing of devices in 3D will further shorten wires. 3D logic is problematic, as will be described below.

The blue curve for CV^2 shows how technology that (a) reduces supply voltage and (b) enables 3D manufacturing will create four scaling scenarios. The hope is that industry can shift from the current path of "MOSFET/2D" to "TFET/3D" to both assure a near-continuous improvement path as well as a more energy-efficient end point for transistor when transistor scaling stops.

Tunnel FETs and MilliVolt Switches

The preeminent form of logic has been Boolean Logic implemented with transistors in the role of switches. Reducing the size of MOSFET transistors has improved power efficiency so much that parasitic leakage current, technically called "kT/q sub threshold slope" now dominates. Leakage current is a property of MOSFETs irrespective of size, so Moore's Law will not help. Unchecked, this leakage current would mean chips could hold more transistors over time just as predicted by Moore, but power per transistor would remain constant. Microprocessors use an architectural remedy to avoid overheating, a remedy that would need to be used in a more extreme form over time. The remedy is to replace a growing fraction of a chip's logic with memory. Memory dissipates less power per unit area, so this reduces overall power per chip. This will make chips less capable than their potential, but it is not feasible to sell chips that overheat. The MOSFET branch of the red curve in Figure 2 started to flat-line around 2003, coincident with the emergence of multi-core processors.

Current developments reported at IEDM earlier in the week and then at RCS 4 reported progress on a potential MOSFET successor called the Tunnel FET (TFET). The TFET could become the first member of a class of proposed devices called milliVolt switches [Yablonovitch Fri 10:00] to reach production. The situation a year ago is that there was diligent search underway for transistors that, when used in a Boolean logic circuit, would have a sub threshold slope of less than kT/q = 60 mV/decade. The consensus of experts at the time was that this level of transistor performance is physically feasible and inevitable, but there were no experimental demonstrations and nobody had an idea of when the experiments would occur. However, IEDM included a handful of papers showing some critical experiments had occurred in the last year [Pandey 15]. Suman Datta summarized his results at RCS 4 [Datta Fri 10:30] showing experimental demonstration of 55 mV/decade for one of two types of TFET (NTFET), beating the 60 mV/decade by 5 mV (lower sub threshold slope is better).

While experimentally beating the limits of the MOSFET by 10% or so is tantalizing and may lead to commercial advances, Eli Yablonovitch gave a talk [Yablonovitch Fri 10:00] on more ambitious research goals that would be needed to fully realize the potential of milliVolt switches. The TFET curve in Figure 2 shows how further advances could allow a reduction in power supply voltage until the cumulative reduction of energy per operation reaches 10× to 100× and a thermal noise reliability limit is reached. This boost would make a difference in computer usage worldwide, but is still not enough to reestablish the expectations of Moore's Law. Eli Yablonovitch forsees additional long-term possibilities. Large power efficiency improvements are also possible from adiabatic and reversible computing, such as [Snider poster].

3D Manufacture

There is also progress in a partial transition of from 2D to 3D chips, another advance that will be important although not enough by itself to restore Moore's Law [Bresniker Thu 12:30][Wong Thu 12:30][Kumar poster]. In the last year or so, multiple vendors started selling memory and/or storage chips using cost-efficient layered manufacturing. The layered manufacturing is likely to extend Moore's Law into the third dimension, yet limited to memory.

The original Moore's Law essentially scales linear dimensions in the X-Y plane of an integrated circuit. The newer 3D scaling keeps dimensions fixed in the X-Y plane but increases the number of layers in the Z dimension. Combining both 2D and 3D scaling may enable historical scaling trends to continue while reducing the pace of technology development required for either factor.

Currently, there are competitively priced Solid State Disks (SSDs) available from consumer vendors [Amazon 15] comprised of 32 layers of Flash storage. The vendors boast that the next generation will be 48 layers [Samsung 15]. The rapid rate with which traditional single-layer chips became 32 layers and the sizes of the increases is reminiscent of Moore's Law. The combination of TFETs and 3D memory should allow more energy-efficient execution of existing software and new software of the current type, including on smartphones, servers, and supercomputers.

However, 3D for logic is somewhat more problematic. Overheating would be a problem due to just a 2D surface for heat removal from a 3D solid – even with TFET/milliVolt switches on the red curve of Figure 2. Manufacturing imperfections in memory can be addressed with Error Correcting Codes (ECC), which is much more difficult to apply to logic. 3D manufacturing would be of a definite benefit, but long term benefits would require advances in manufacturing and computer architecture to deal with heat and reliability issues.

The outbrief by Paolo Gargini [Fri 11:00] concluded that the advances described above for transistors and 3D should be sufficient to drive industry expansion over the next decade, at which time other devices now in the research pipeline would be ready (as described below).

Probabilistic Computing (Track 1)

Laura Monroe gave an overview talk on probabilistic and approximate computing [Monroe Thu 8:45], followed by Dave Mountain and Laura Monroe leading a track on these topics. These approaches build naturally on the results of track 2 above.

If it is assumed that TFETS and milliVolt switches will become part of the technology mix, pressure from the user community is expected to drive continued reduction of component size and component energy consumption until scaling is stopped by other issues.

The issues are believed to be known at this time, and fall into two categories: (a) Tolerances and defects in a given manufacturing technology will stop scaling due to errors resulting from too weak and faulty devices. Progress in manufacturing is expected to reduce this type of error over time, but progress in manufacturing cannot continue forever due to the discrete nature of atoms. (b) Thermal noise will cause an exponentially rising error rate as signal energy approaches kT, an effect that is fundamental to Boolean logic. Mitigating this effect with non-Boolean logic will be deferred to the later section on track 3.

Since scaling-induced errors rise continuously as opposed to having an abrupt onset, the ability to manage a moderate number of errors can extend scaling. If errors are not considered in advance, scaling would need to stop at the point where the chance (or certainty) of an error exceeds user-originated reliability requirements for an application. This is because any error at run time could propagate to become a system crash or an incorrect answer being given to the end user. However, scaling could continue further if the computer had the ability to manage one error per N operations (or memory bits) sufficiently well that the end user remained satisfied. The most effective method and the value of N vary by the problem being solved.

The methods considered by this track were approximate, probabilistic and stochastic computation. We distinguish between these, and in particular between approximate and probabilistic, which are often conflated.

Approximate computation is designed to come appropriately close to a correct answer, whether through use of reduced precision or through numerical methods. It may be deterministic. Probabilistic computation calls upon probabilistic reasoning on the underlying hardware or the data. It is non-deterministic by nature, so need not give the same results between runs. However, the results in a probabilistic calculation should average out to a correct result over repeated runs. Approximate and probabilistic compute methods thus are inherently different, but there are approaches that combine the two.

Approximate computing can be used for applications that can tolerate slightly inaccurate results. The decision to be made is the degree of approximation that may be tolerated. A typical example of approximate computing is a video playback where the human viewer may be willing to tolerate some inaccuracy in color reproduction in exchange for longer battery life. Another example is deterministic digital computation, which approximates floating point calculations to the precision allowed in the hardware or software.

Probabilistic computing applies when a computer system is expected to deliver accurate results to the user, yet the underlying components produce errors due to their own inaccuracy or due to custom-built non-determinism. The decision to be made here is the degree to which incorrect results can be tolerated, i.e., the probability whether and by how much the result will differ from the correct result.

One example of probabilistic computing is when the underlying computer hardware has had voltage scaled down so far that logic gates make too many mistakes for the system to meet stringent reliability requirements. Management of these errors often includes error detection codes for logic/memory, with detection followed by recovery and rerunning the erroneous computations. Another approach is to use fault-tolerant algorithms. For example, if an error occurs in an iterative algorithm that converges to the correct answer, an error may simply lead to more iterations before convergence. Finally, the calculation may simply be run and the results used if the application is sufficiently tolerant of the given probability of an incorrect result.

Stochastic computing is a form of probabilistic computing is where algorithms rely on random numbers, such as Monte Carlo simulations. In these algorithms, components that have been scaled so far that they produce random errors can be used as extremely energy efficient random number generators.

Approximate, probabilistic, and stochastic methods all require a good understanding of the underlying physics, methods for ascertaining which energy efficiency gains might be possible and at what cost [Anderson poster], and strategies for realizing systems that achieve maximum efficiency gains with minimum loss of computational efficacy.

New Devices and New Approaches to Computing (Tracks 1, 2, and 3)

Instead of computing being "rebooted" by some future discovery, RCS 4 raises the possibility that the key discoveries are being made independently and what is needed is to fuse them into a common approach. RCS 4 created a forum where one set of complementary approaches were discussed by their proponents. The defining characteristics of the approaches are illustrated in Figure 3 in a way that highlights their common features. Projections for the energy efficiency, density, and other benefits for these approaches are so much higher than the equivalent Boolean logic gate implementation that they may together have enough growth potential to restore an exponential improvement path like Moore's

law. These new approaches rely on the continued evolution of transistors, since they are also dependent on transistors.

A: Memory (reading) [Marinella Fri 10:00]
B: Vector-matrix multiply [Strachan Fri 10:30]

C: Vector-matrix-transpose multiply [Agarwal 15]

Memory data;
A for $y = xA$

Memory address

x for $y = xA$

y for y = xA^T

Memory word

y for $y = xA$

x for y = xA^T

New, stateful devices

D. Neuron [Kim 15, Hasler 13]

Conductance is synapse weight

E. Rank-1 update [Agarwal 15], which is also
A. Memory (writing) x for $A = A + xy^\mathsf{T}$

y for $A = A + xy^\mathsf{T}$

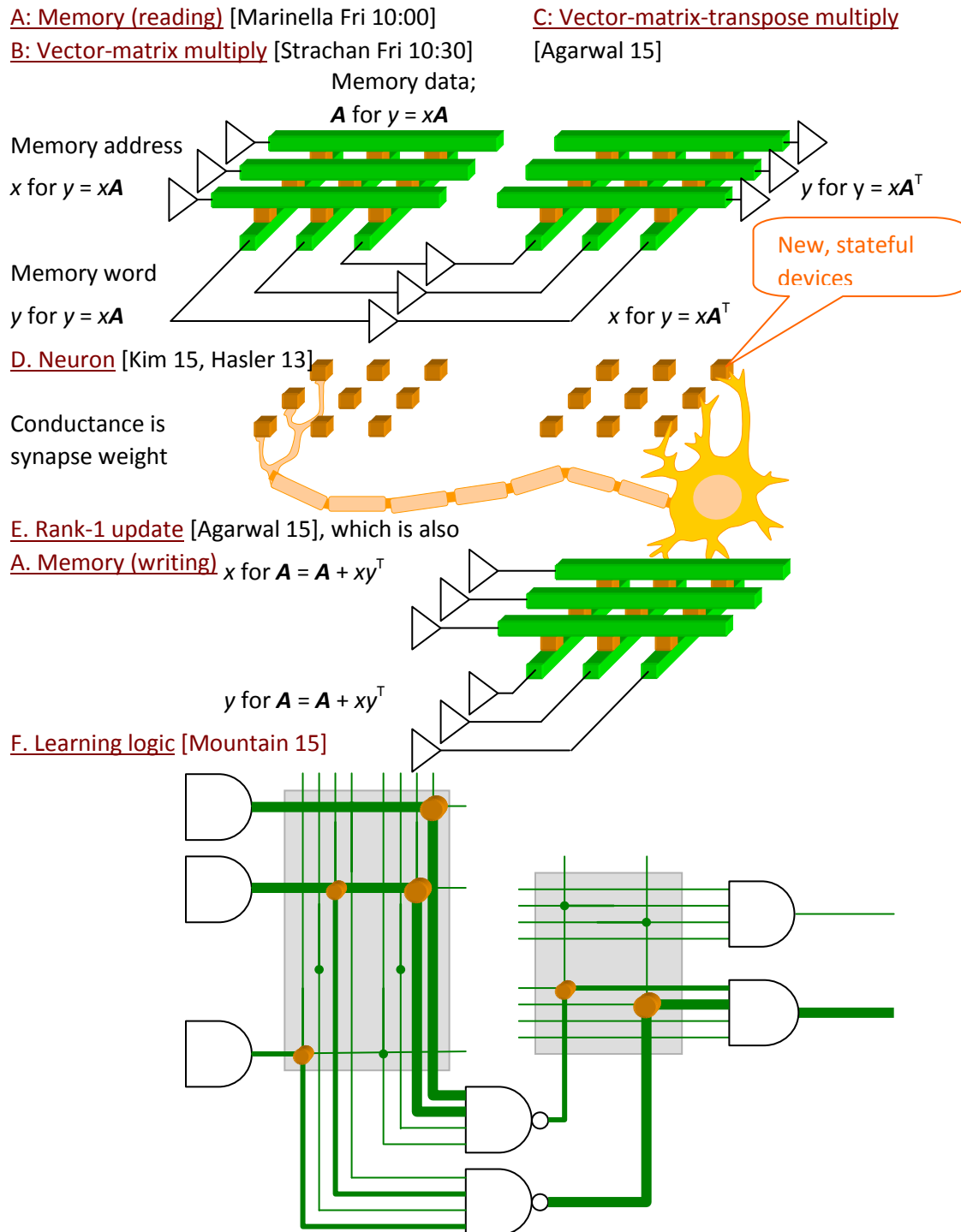F. Learning logic [Mountain 15]

Figure 3: Multiple usage modes for new state-containing devices

The common features across the examples in Figure 3 are as follows: Each uses a new state-containing device in addition to transistors. Furthermore, the building block common across Figure 3 is an array or crossbar, which contrasts to the Boolean logic gate that has been the common building block for logic. The original cross-device studies [Nikonov 13] and summarized by Ian Young [Young Thu 2:15] looked at non-transistor devices as replacements for the switches underlying Boolean logic gates, comparing the devices against CMOS by the energy and speed of the resulting circuit. The structures in Figure 3 are bigger and more complex than a single Boolean logic gate, meaning the functions are equivalent to hundreds, thousands, or in fact a quantity of Boolean logic gates that scales up over time. Energy efficiency can be much higher when a function is realized directly instead of being realized through the intermediate step of creating Boolean logic gates, an idea with a theoretical support [DeBenedictis poster].

The concept of state-containing devices deserves explanation: A transistor is described by equations, tables, or measurements that relate the voltages and currents in the leads. However, the behavior of a state-containing device will also be dependent on the data stored in the device. This data is called state and is typically the Boolean logic values TRUE and FALSE or memory bit values 0 and 1. For example, the current through a device could be higher when the device's state holds a 1 than when it holds a 0.

The contributions at RCS 4 that are described in Figure 3 are as follows:

Advanced memories

Dedicated memory is important due to the ubiquity of the von Neumann architecture and its division of computers into a processor and memory. Irrespective of new approaches, plain memory is expected to remain important even after computing is "rebooted." Figure 3A illustrates the baseline memory circuit, which reads by driving one row with a decoded address and senses the memory contents on the columns. Writing involves driving one row with a decoded address and driving the data to be written on the columns. The International Technology Roadmap for Semiconductors (ITRS) roadmaps memory devices such as Flash, the memristor, phase change memory, and various magnetic devices [Marinella Fri 10:00].

Historically (i. e. not in RCS 4), the memristor (a device) was renamed to a Resistive Random Access Memory (RRAM or ReRAM) device for its use in advanced memories (more on this below).

Neural Networks

Neural networks are often conceptualized as an "arrays of synapses," which are investigated as row-column arrays of cells that store synapse values at the crossings as illustrated in Figure 3D [Burr 15][Hasler 15][Marinella Fri 10:00b][Mountain 15][Mountain Thu 3:45] [Franzon poster][Vineyard poster]. All the synapses in an array can learn by driving the rows and columns appropriately, an operation expressed mathematically as a rank-1 matrix update where the state-containing devices comprise the matrix elements. To make a neural net perform (a neural network is said to be performing when it processes information without learning), the rows are driven with stimuli and results are read from the columns. Performance is mathematically equivalent to vector-matrix multiply. The devices at the cross points have changed over the years, becoming smaller, more precise, and more energy efficient. Memristors/RRAM and phase change memory have been used quite effectively for neuromorphic computing research. The brain-inspired approach for creating better computers in [Kumar poster] is shared with the references earlier in this paragraph, but the execution platform is different.

It is ironic but consistent with the point of this summary that the act of renaming "memristors" to "RRAM" associated the device with a specific application, which was promptly reversed by the use of the memristors in neural networks.

Matrix Algebra Engines

RCS 4 included a presentation on the "Dot Product Engine," [Strachan Fri 10:30], a memristor crossbar that performs vector-matrix multiplication at very high speed. The circuit in Figure 3B is the same as neuromorphic crossbars, but the usage has been generalized to be a component in non-neuromorphic systems, such as signal processors.

Vector-matrix multiplication sometimes works quite nicely in reverse. The roles of rows and columns can be trivially interchanged if the devices at the row-column intersections have two terminals (as shown throughout Figure 3 – although there are important devices that have three terminals and require a double layer of rows or columns). Such a change would require more complex row and column electronics, but the stored data would not change. This has the mathematical effect of transposing the matrix, for example leading to Figure 3B computing $y = x\boldsymbol{A}$ while Figure 3C computes $y = x\boldsymbol{A}^\mathrm{T}$.

The writing of a memory illustrated in Figure 3D is a special case of what is known in vector algebra as a "rank 1 update" and is essentially the delta learning rule [Widrow 60] in neuromorphic systems. A rank-1 update is defined as $\boldsymbol{A} = \boldsymbol{A} + yx^\mathrm{T}$, where A is a matrix, and vectors x and y are multiplied in an outer product (yielding a matrix). The delta rule is used in backpropagation in neural system where the outer product of the neural stimulation and the error is used to adjust synapses. In a memory, one of the vectors is the decoded address and the other is the data to be written.

The discussion above has reversed the simulation relationship between neural networks and some uses of supercomputers. Neural networks have been simulated on supercomputers for many years using matrix algebra subroutine packages. In a role reversal, this section showed how technology derived from neural networks could simulate the linear algebra subroutines that run on conventional computers. While not described at RCS 4, some of the attendees wrote a paper [Agarwal 15] analyzing the energy efficiency of a sparse coding algorithm on a crossbar like the ones in Figure 3. This analysis of an exemplary matrix algebra algorithm showed an energy efficiency improvement over an equivalent CMOS implementation.

Precision

Computation based on the approaches above would have precision limits, but RCS 4 also included a paper [Khasanvis 15][Khasanvis Thu 8:45] and attendees who have made research contributions [Nikonov 13] that address precision limits. The array structure common across Figure 3 has single, independent device at each intersection. While the devices may hold analog values, analog computing becomes increasingly difficult as precision increases. However, Santosh Khasanvis presented a talk on an architecture that uses multiple magnetoelectric devices to represent a single value at increased precision. Santosh's structure was different from an array. Magnetoelectric devices are also one of the advanced memory devices covered by ITRS [Marinella Fri 10:00], studied as a logic device [Nikonov 13], and analyzed theoretically [DeBenedictis poster]. There was not enough material at RCS 4 (or in the literature, for that matter) to more fully analyze high precision computation using emerging devices.

General Logic

RCS 4 also included a paper on an Ohmic Weave [Mountain 15], which is essentially a hybrid computer architecture of Boolean logic and artificial neural networks. Ohmic Weave can embed a Boolean logic diagram into a neural network as shown in Figure 3F, using the new memory devices in part to specify logical function and in part to specify how the logical functions are wired together. Ohmic Weave could lead to a future computing system that is manufactured with unallocated resources that would later become either the current style of digital logic, neurons in an artificial neural network, or perhaps the current style of digital logic based on the circuit learning its function instead of being designed or programmed by a human.

Many types of artificial neurons are a generalization of logic gates, thus forming the technical basis of this approach. More specifically, setting neural synaptic weights in a specific ways allows a neuron to perform a Boolean logic function such as the NAND shown in Figure 3F. Artificial neurons are more general than Boolean logic gates in the sense that the synaptic weights are learned or trained, making a group of artificial neurons roughly equivalent to the combination of Boolean logic gates plus the interconnect wire. A Field Programmable Gate Array (FPGA) is similar.

However, Ohmic Weave has a learning capability beyond what is possible in Boolean logic networks or FPGAs. Some of the synapses would become strong connections through learning that become the thick wires illustrated in Figure 3F and which control the circuit. However, a neural network contains more information than just what has been learned. Neural networks also contain information observed in the environment or during training that has not been consistent enough to actually create new behavior, but which may speed the learning of new behaviors later. Figure 3F shows this additional in formation as wires that are too thin or weak to control the circuit, but which may influence the circuit learning new behavior later. This shows how Ohmic Weave may replace both a logic circuit and some of the activities of the logic design engineer.

As mentioned above, the structures in Figure 3 can have energy efficiency benefits over implementation of equivalent functions using Boolean logic gates. Thus, the Ohmic Weave is in part a demonstration of how lessons learned from the study of brains could be used to make more energy efficient computers. The demonstration in the RCS 4 paper was an AES encryptor implemented with neurons performing complex Boolean logic functions, and a malware detector implemented as a neural network.

Sensible Machine and Grand Challenge

The collection of ideas in Figure 3 could create a new approach to computing when viewed all at once, which is very nearly the definition of the OSTP Nano-Inspired Grand Challenge for Future Computing announced October 20, 2015. This Grand Challenge followed a Request for Information (RFI) from OSTP in June 2015 that Stan Williams and about 100 other people responded to. Stan's response titled the "Sensible Machine" was the technical idea or template for this Grand Challenge. Lloyd Whitman of OSTP was the lead on defining the Grand Challenge, and gave a talk on it [Whitman Thu 5:00]. Stan Williams also gave a talk on his idea [Williams Thu 3:45].

Given the importance of Federal Government sponsorship, the RCS 4 organizers made last-minute adjustments to the agenda after the Grand Challenge announcement. Synergy between the Grand Challenge, the organization of RCS 4, and this document should be seen as deliberate.

The definition of the Grand Challenge in [Whitman Thu 5:00] and elsewhere, "[C]reate a new type of computer that can proactively interpret and learn from data, solve unfamiliar problems using what it has learned, and operate with the energy efficiency of the human brain," and clarifying text [Whitehouse 15] seems to fit quite well with the exposition presented above. The objective is to make a new type of computer with new capabilities to learn from massive amounts of data and solve problems. The direct connection to the human brain is through energy efficiency, but indirectly the expectation is that neuroscience and neuromorphic computing could be used as inspiration for the development of new computational techniques.

Superconducting Technologies

In a fourth track, Marc Manheimer, program manager for the IARPA Cryogenic Computing Complexity (C^3) program, provided an overview of a computing approach based on SuperConducting Electronics (SCE) [Manheimer Thu 10:15][Manheimer poster] and based on [Holmes 15][Kadin poster]. While C^3 is based on completely new technology at the low level, it parallels research directions in the larger industry quite well. SCE is a computing approach where the electronics are cooled to nearly absolute zero, causing the wires to become superconductors where they lose all resistance. Two-terminal Josephson Junctions (JJs) are used in lieu of transistors in Boolean logic circuits. The C^3 program includes research on both JJ-based logic circuits and cryogenic versions of some of the state-containing memory devices in Figure 3.

Computer logic based on SCE has been a possibility for decades, yet shifts in the way transistors are likely to scale may be providing an opportunity for this approach to move into production. If the computer industry accepts segmentation of technology as suggested above, SCE could become an option for large computer installations such as supercomputers and server farms. The limitation to large installations is due to economies of scale for cooling.

The plot in Figure 4 [Frank 14] shows a basis for segmenting logic technology. The energy versus speed plot shows many crossing curves for transistorized options, yet all fall behind the Pareto frontier added by the current authors as a heavy red line. Energy can be traded off for speed in transistorized Boolean logic circuits, but all such circuits are limited by certain features common to transistors.

Superconducting electronic circuits based on Single Flux Quantum (SFQ) signaling are not subject to the energy-speed tradeoff, creating an opportunity for extremely high speed circuits annotated on the right of Figure 4. Other circuits made of JJs and superconducting wires can implement Boolean logic functions with ultra high energy efficiency, leading to the opportunity annotated at the bottom of Figure 4.

A limitation on the minimum size of superconducting electronics has been a criticism in the past, yet shifting electronics to 3D may make this criticism moot. Superconducting electronics needs feature sizes greater than about 100 nm in order for the quasi-particles that carry information to have space to move freely. This 100 nm "coherence length" is an order of magnitude larger than the projected minimum feature size for transistors of 10 nm or thereabouts. However, shifting electronics to 3D would make the feature size limitation of superconducting electronics much less of a problem while making ultra high energy efficiency much more of an advantage.
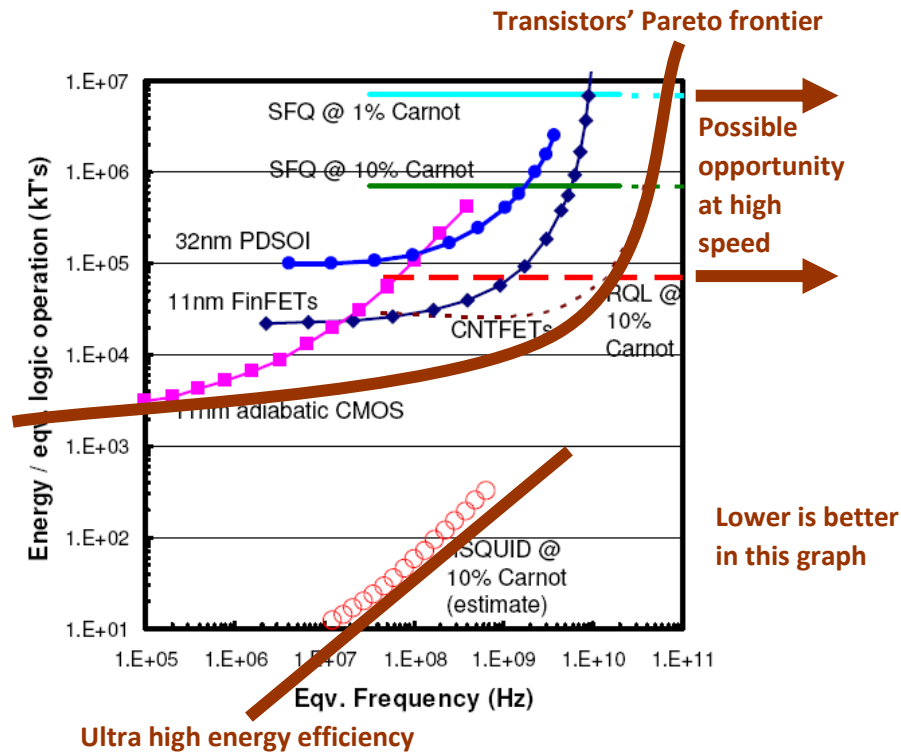
Figure 4: Superconducting technology in context [Frank 14]

National Scale Programs

RCS 4 included brief presentations by program managers and other leaders across multiple funding agencies, including NSCI [Koella Thu 11:15] OSTP [Whitman Thu 5:00] DARPA [Hammerstrom Thu 5:15] and IARPA [Matheny 5:30]. In addition, several non-government organizations supporting computing-related research gave overviews of their activities ITRS 2.0 [Gargini Thu 10:45] SRC [Joyner Thu 11:00] and IEEE [Conte Thu 8:30].

Conclusions and Looking Ahead

The Future of Computing

The ideas above start to define a path forward. Transistor-like devices used as switches in Boolean Logic and von Neumann computers will continue to improve for a decade, allowing continuation of "Moore's Law" in that timeframe. At the same time, new systems will develop based on arrays of new types of state-logic devices arranged into arrays that will process stored data very efficiently, including learning from data. These new systems will boost the performance of computers and supercomputers – but not in the traditional direction. Computer applications that rely on fast single processors with low or modest memory requirements may be reaching a performance plateau. However, the end-state of that plateau may include unfamiliar technologies such as probabilistic and superconducting technologies. However, applications for servers and supercomputers that currently rely on big data may grow with a reinvigorated Moore's Law. Applications that learn may emerge for the first time with an exponential growth path. A key software factor will be the ability to capture the behavior of today's computer programmers, operators, and data analysts and teach the behaviors to new learning computers.

RCS Publications, Roadmaps, and Future Conferences

One of the goals of the RC Committee and the participants is to publish a White Paper or article summarizing the conclusions of the RCS series of Summits. The venue of such a report might be in a journal such as *IEEE Computer*, or alternatively in a new journal such as the *IEEE Journal of Exploratory Solid-State Computational Devices and Circuits*. In addition, these summits could lead to the establishment of an annual international conference on Rebooting Computing, which will bring together engineers and computer scientists from a wide variety of disciplines, to help promote a new vision of Future Computing. Finally, there is interest in developing industry-wide roadmaps and standards that can guide future development of computer systems in the same way that ITRS guided device development during the Moore's Law era.

Appendices

Appendix A: Agenda for Rebooting Computing Summit 4 (RCS4)

9-11 December, 2015 – Washington Hilton, Washington, DC

Duration	Time	Activity		
		Wednesday, December 9, 2015		
3:00	6:00 PM	Reception		
	9:00 PM	End reception		
		Thursday, December 10, 2015		
0:15	8:30 AM	Review of impetus for IEEE RC initiative, review of RC summits (3 pillars, complementary nature of various approaches, etc.). Tom Conte/Elie Track		
1:15	8:45 AM	Track 1: Probabalistic/random/approximate big picture and experimental results L. Monroe; S. Khasanvis (tent.)		
0:15	10:00 AM	Break		
0:30	10:15 AM	Extra Track: Superconductive electronics/C^3 Marc Manheimer		
0:15	10:45 AM	Review of other initiatives in this area – ITRS 2.0 Paolo Gargini		
0:15	11:00 AM	Review of other initiatives in this area – SRC William Joyner		
0:15	11:15 AM	Review of other initiatives in this area – NSCI William Koella		
1:00	11:30 AM	Lunch (after a brief announcement of LPIRC 2016)		
1:15	12:30 PM	Track 2: 3D integration and new devices big picture and experimental results Kirk Bresniker; H. S. P. Wong		
0:30	1:45 PM	Track 1: Co-facilitators Dave Mountain, Laura Monroe	Track 2: Beyond CMOS Benchmarking I. Young, plus discussion	Track 3: Co-facilitators Erik DeBenedictis, Yung-Hsiang Lu
0:30	2:15 PM			
0:30	2:45 PM			
0:30	3:15 PM	Break		
1:15	3:45 PM	Track 3: Neuromorphic/Sensible Machine big picture and experimental results Stan Williams; Dave Mountain		
0:15	5:00 PM	Review of other initiatives in this area – OSTP Grand Challenge Lloyd Whitman		
0:15	5:15 PM	Review of other initiatives in this area – DARPA Dan Hammerstrom		
0:15	5:30 PM	Review of other initiatives in this area – IARPA Jason Matheny		
0:30	5:45 PM	Break (needed for set up by hotel) and *** **GROUP PICTURE** ***		
0:45	6:15 PM	Posters (in same room as reception)		
2:00	7:00 PM	Reception starts in poster area		
	9:00 PM	End reception		
		Friday, December 11, 2015		
0:30	8:30 AM	First working group review		
1:00	9:00 AM	Track 1: Co-facilitators: Dave Mountain, Laura Monroe	Track 2: Moore's law	Track 3, continued
0:30	10:00 AM		E3S Eli Yablonovitch	Neuromophic tech. Matt Marinella
0:30	10:30 AM		Steep Slope Transistors S. Datta	Dot Product Engine J. P. Strachan
1:00	11:00 AM	Second working group review		
0:30	12:00 PM	Lunch		
0:00	12:30 PM	RCS 4 Adjourns		
5:30	12:30 PM	Associated IEEE/RC "Sensible Machine" Grand Challenge group meeting		
	6:00 PM	Sensible Machine group meeting adjourns		

Note: Matt Marinella actually gave a talk memory in track 2 (and attended track 3 as well).

Appendix B: RCS 4 Participants

John Aidun	Sandia National Laboratories
Neal Anderson	UMass Amherst
Marti Bancroft	MBC
Mustafa Baragoglu	Qualcomm
Herbert Bennett	AltaTech
Kirk Bresniker	Hewlett Packard Labs
Geoffrey Burr	IBM Almaden
Dan Campbell	GTRI
Tom Conte	Georgia Tech
Stephen Crago	USC - ISI
Shamik Das	Mitre
Suman Datta	Univ. of Notre Dame
Barbara De Salvo	CEA – LETI (France)
Erik DeBenedictis	Sandia National Laboratories
Gary Delp	Mayo Clinic
Carlos Diaz	TSMC
Michael Frank	Sandia National Laboratories
Paul Franzon	North Carolina State Univ.
Paolo Gargini	ITRS
Kevin Gomez	Seagate
Tim Grance	NIST
Wilfried Haensch	IBM Yorktown Heights
Jennifer Hasler	Georgia Tech
Kenneth Heffner	Honeywell
Bichlien Hoang	IEEE
Thuc Hoang	NNSE - DoE
Scott Holmes	IARPA
Wen-Mei Hwu	Univ. of Illinois
William Joyner	SRC
Alan Kadin	Consultant
Andrew Kahng	UC San Diego
Santosh Khasanvis	BlueRISC
David Kirk	NVIDIA
Will Koella	NSA
Dhireesha Kudithipudi	Rochester Inst. of Technology
Arvind Kumar	IBM
Rakesh Kumar	Univ. of Illinois
Hai Li	Univ. of Pittsburgh
Ahmed Louri	George Washington Univ.
Yung-Hsiang Lu	Purdue Univ.
Mark Lundstrom	Purdue Univ.
Marc Manheimer	IARPA
Matthew Marinella	Sandia National Laboratories
Jason Matheny	IARPA
LeAnn Miller	Sandia National Laboratories

Chris Mineo	Lab. For Physical Sciences
Laura Monroe	Los Alamos
David Mountain	NSA
Robert Patti	Tezzaron Semicond.
Robert Pfahl	iNEMI
Wolfgang Porod	Univ. of Notre Dame
Rachel Courtland Purcell	IEEE Spectrum
Shishpal Rawat	Intel
Chuck Richardson	iNEMI
Curt Richter	NIST
Heike Riel	IBM Research
Stefan Rusu	TSMC
David Seiler	NIST
Gregory Snider	Notre Dame University
Roger Sowada	Honeywell
John Spargo	Northrop Grumman
John Paul Strachan	Hewlett Packard Labs
Jack Yuan-Chen Sun	TSMC
Elie Track	IEEE Council on Superconductivity
Wilman Tsai	TSMC
Jeffrey Vetter	Oak Ridge National Lab
Craig Vineyard	Sandia National Laboratories
Lloyd Whitman	OSTP
Stan Williams	Hewlett Packard Labs
Philip Wong	Stanford
Eli Yablonovitch	UC Berkeley
Ian Young	Intel

Summary of the approximate, probabilistic, and stochastic computing breakout sessions (Track 1).

Breakout session attendees represented a variety of interests and experience in this subject
David Mountain and Laura Monroe, co-facilitators

Tom Conte	Steve Crago	John Daly
Neal Anderson	Elie Track	John Aidun
Jeff Vetter	Chris Mineo	Rakesh Kumar
Kurt Richter	Gary Delp	Thuc Hoang
Arvind Kumar	Bill Harrod	Roger Sowada

Key takeaways from a (somewhat) structured discussion over the two days:

Identifying applications to drive R&D efforts is highly effective. Applications can be broken down into two major categories:

Single applications such as streaming analytics or image recognition – these are applications for an end user.

Foundational applications such as iterative solvers, BLAS (basic linear algebra solvers), etc. – these are libraries or components that tend to be used in a large number of applications.

The driving applications may be very different for each type of computing in this track.

Developing a taxonomy and language to describe these approaches to computing and their components is important – standards and metrics are part of this; there needs to be a way to describe and quantify trade-offs.

Fault models are crucial for probabilistic computing. Rates of faults, distribution of fault types, propagation vectors, etc.

Need to think about all parts of the computing architecture – can these approaches help with data movement issues?

Scientists that explore the natural world deal with approximations all the time. How can we leverage their knowledge?

Neal Anderson noted a current lack of high-level theoretical guidance on what gains are possible in principle through probabilistic computing, including costs and savings for specific computational problems and input characteristics, and suggested that such guidance would be helpful as the field progresses.

An example of such guidance would be answers to questions like:

"Given a computational problem P, input/data statistics S (possible inputs and their probabilities), and a deterministic solution C (hardware and, where applicable, program/algorithm), does there exist a probabilistic solution at reliability R that could provide an X-fold increase in Y within a specified penalty Z?" (Here Y, Z are things like energy consumption, run time, circuit complexity, etc.)

The group also completed an initial road mapping exercise, based on the following assumptions:

Approximate computing is 'ready to go,' while probabilistic computing needs a little maturation. Some reasonable level of investment will be made in these approaches to make progress.

Year 1 milestones
Create a community of interest, initial tasks would be:
- Develop a framework and language for describing and evaluating ideas and accomplishments.
- Develop kernels, benchmarks, metrics to drive explorations and evaluations.
- Develop a modeling-simulation environment and hardware testbed based on CMOS technology. This is probabilistic computing centric. Mod-sim goals for years 1, 2, 3 – Toy environment, PhD usable environment, 'production' level environment.
- Develop an approximate BLAS library that enables precision vs. performance, energy vs. resilience trade-offs.

Year 2 milestones
- Develop new algorithms that leverage approximate computing approaches (such as Monte Carlo, machine learning, etc.)
- Develop a production quality toolchain for implementing approximate computing routines.
- Specify an ISA (instruction set architecture) and functional units of value for approximate computing
- Develop a strong working relationship with the beyond CMOS device community to support longer range efforts. This is probabilistic computing centric.

Longer term milestones
- Apply random algorithms to probabilistic hardware and show improvement in metrics of value.
- Develop advanced hardware prototypes that implement specialized microarchitectures for application development and evaluation.
- Develop an initial information theory of probabilistic and stochastic computing.
- Build a hardware testbed for probabilistic computing that incorporates 'beyond CMOS' technology.
- Demonstrate an approximate-computing centric system level implementation.

Appendix D: Group Outbrief on Beyond CMOS

Summary by Paolo Gargini

- Moore's Law (i.e., doubling of transistors every 2 years) will continue for the next 5-10 years.

- FIN FET transistors were introduced into manufacturing in 2011; due to their vertical on-the-side structure TFETs provide higher packing density than planar CMOS transistors.

- The NRI was initiated in 2005 with the goal of finding the "next switch".

- In 2010 a selected group of possible new switches was identified.

- TFET transistors were identified in the breakout as the most likely candidates to replace or work in conjunction with FIN FET beyond 2020. Multiple papers on TFET were presented at 2015 IEDM on Dec 7-9. TFET transistors based on 2D materials developed at E3S Center represent a real breakthrough.

- Memory devices are reaching fundamental 2D-space limits.

- Leading Flash companies are introducing 3D flash memory in production in 2016 packing 32 to 48 layers.

- Logic devices will also convert to this 3D architecture in the next decade.

- The next generation of scaling, succeeding Geometrical Scaling (1975-2002) and Equivalent Scaling (2013~2025) has been named 3D Power Scaling. 3D architecture and minimal power consumption are the main features of this scaling method.

- Reduction of power consumption in logic devices will allow logic/memory 3D architecture to dominate the latter part of the next decade.

- 3D architecture will allow insertion of multiple logic and memory devices in the cross point nodes.

- Resistive memory and carbon nanotubes are also considered viable candidates for 3D memory implementation.

- Significant progress has also been accomplished in magneto-electric devices. These devices, often spin based, combine the mobility of electrical charges with the memory features of magnets. Possible co-location of logic and memory operations may be possible with these types of devices.

- New materials are the key enablers of all these new devices and architectures.

- Lack of adequate facilities capable of processing full flow device is a major limiter.

Summary Roadmap

Present 2015-2030 Roadmap

- FinFET good for 5+5 years to 2025
- 2005. NRI: New Switch
  - Technology reuse: cousins of CMOS
  - Explore magneto-electric effects
- 2010. TFET looks promising, effort increased
  - New magneto-electric devices proposed
  - Memory options on the raise
- 2015. TFET good for 2020-2025 coexistence with CMOS
  - Increase effort on magneto-electric likely winner(s)
  - 3D memory in production
  - Resistive memory very promising
- 2020. Magneto-electric earliest insertion
  - Some new memory makes it big

Appendix E: Group Outbrief on Neuromorphic Computing

This brief section has been added for purposes of consistency. The group lead for the neuromorphic track was the principal author of the "technical summary of RCS 4." Ideas for the neuromorphic group outbrief have been integrated into that section.

Abstracts for Posters Presented at the 4th IEEE Rebooting Computing Summit

---

Revealing Fundamental Efficiency Limits for Complex Computing Structures: The FUELCOST Methodology

Neal G. Anderson, Ilke Ercan*, and Natesh Ganesh
University of Massachusetts Amherst

The energy efficiency of computation doubled every ~1.57 years from the dawn of digital computation to around 2010 (Koomey's Law), after which progress has slowed. Restoration of exponential efficiency scaling over the long term will likely be achievable only through the development of new computing technologies based on unconventional computing strategies and paradigms. Given the major investment that will be required to develop any new computing paradigm, and the critical importance of energy efficiency, evaluation of alternative computing paradigms should include limiting efficiency as an integral component.

In this poster, we describe an evolving physical-information-theoretic methodology—the FUELCOST methodology—that enables determination of the FUndamental Efficiency Limits of complex COmputing STructures. This methodology is based on a fundamental physical description of the *dynamics of information* as it is processed by computing hardware, as opposed to a physics-based description of the *dynamics of computing hardware* as it processes information (e.g. as in standard models and simulations). This enables isolation of fundamental sources of inefficiency that are deeply rooted in physical law and incurred at different levels of abstraction in complex computing systems. We discuss the underlying theoretical approach; previous studies of various computing structures (finite-state automata, simple processor architecture), logic blocks and functions (ALUs, decoders, adders), and nanocircuit implementations (both transistor- and non-transistor-based); progress toward full synthesis, integration, and automation of the multi-level evaluation methodology; and exploratory application directions (digital/discrete-analog neuromorphic, approximate, and Brownian approaches).

* Present Address: Boğaziçi University

_____

Adapting to Thrive in a New Economy of Memory Abundance

Kirk Bresniker and R. Stanley Williams
Hewlett Packard Labs

Processing technology has eclipsed memory technology for the past six decades, but processor-centric architectures are reaching their terminal efficiency. We can reboot computing on the basis of abundant memory enabled by emerging device physics, which will make computation, communication, and memory more efficient. This approach also provides a unique opportunity to address novel security threats with modern, systemic solutions.

_____

Improving Energy Efficiency via Nonlinear Dynamics and Chaos

Erik P. DeBenedictis[1], Neal G. Anderson[2], Michael P. Frank[1], Natesh Ganesh[2], R. Stanley Williams[3]
[1]*Sandia National Laboratories,* [2]*University of Massachusetts Amherst,* [3]*Hewlett Packard Labs*

The Boolean logic abstraction offers intellectual elegance and reduces design effort, but may also limit energy efficiency. This poster gives one example where a new circuit based on a new MeRAM device theoretically improves energy efficiency by several orders of magnitude over accepted projections of Boolean logic gates. A

route to improved energy efficiency was demonstrated for a simple "learning machine," but generalization to other problems is beyond the scope of this poster.

_____

Cortical Processing

Paul Franzon
North Carolina State University

Cortical Processing refers to the execution of emerging algorithms relying on probabilistic spatial and temporal recognition. In this work we are building processors customized towards execution of these algorithms. Examples of these algorithms include Cogent Confabulation, Hierarhical Temporal Memory, and Long Short Term Memory. Customization features include Processor in Memory and functional accelerators. Improvements in performance/power of up to 10^5 have been demonstrated over GPUs.

_____

Superconducting Computing in Large-Scale Hybrid Systems

Alan M. Kadin[1], D. Scott Holmes[2], and Mark W. Johnson[3]
[1]Consultant for Hypres, Inc.; [2]Booz Allen Hamilton; [3]D-Wave Systems, Inc.

The past, present, and future of superconducting computing will be discussed, based on the feature article in the December issue of IEEE Computer Magazine. Specific systems addressed will include processors for digital radio receivers, quantum annealing, neural simulators, and ultra-low-power adiabatic computing.

_____

Scaling Up of Neuromorphic Computing Systems using 3D Wafer Scale Integration

Arvind Kumar[1],Zhe Wan[2,3], Subramanian Iyer[3], and Winfried Wilcke[4]
[1]IBM TJ Watson Research Center, [2]IBM Albany Nanotech, [3]UCLA, [4]IBM Almaden Research Center

The cognitive era is just beginning, with hopes of computing machines that can solve "unstructured" computational problems, such as sensing, learning, and inferring; detecting patterns and anomalies; and predicting and discovering. These types of data-centric applications require a fundamental shift from the von Neumann architecture which has defined computing systems since the 1940s. Inspired by the brain, we propose a radically different architecture consisting of a large number of highly interconnected simple processors intertwined with very large amounts of low-latency memory. This memory-centric architecture can be realized using 3D wafer scale integration, which provides massive interconnectivity through very high bandwidth directly between processors. Combined with mature CMOS technologies, it provides a path toward early realization of a highly scaled-up neuromorphic computer. The natural fault tolerance and lower power requirements of neuromorphic processing make 3D wafer stacking particularly attractive.

_____

Low-Power Image Recognition Challenge (LPIRC)

Yung-Hsiang Lu
Purdue University

Many mobile systems (smartphones, electronic glass, autonomous robots) can capture images. These systems use batteries and energy conservation is essential. This challenge aims to discover the best technology in both image recognition and energy conservation. Winners will be evaluated based on both high recognition accuracy and low power usage. Image recognition involves many tasks. This challenge focuses on object detection, a basic routine in

many recognition approaches. The first LPIRC was held in June 2015 and the top two winners presented their solutions in the International Conference on Computer Aided Design in November 2015. The second LPIRC is planned for June 2016 in Austin Texas.

_____

Cryogenic Computing Complexity (C3) Program

Marc Manheimer
IARPA

The ultimate goal of the Intelligence Advanced Research Projects Activity (IARPA)'s Cryogenic Computing Complexity (C3) program is to demonstrate a complete superconducting computer including processing units and cryogenic memory. IARPA expects that the C3 program will be a five-year two-phase program. Phase one, which encompasses the first three years, primarily serves to develop the technologies that are required to separately demonstrate a small superconducting processor and memory units. Phase two, which is for the final two years, will integrate those new technologies into a small-scale working model of a superconducting computer. Program goals are presented, and the approaches of the phase-one teams are reviewed.

_____

Acceleration of Neural Algorithms using Nanoelectronic Resistive Memory Crossbars

Matthew J. Marinella, Sapan Agarwal, David Hughart, Steve Plimpton, Ojas Parekh, Tu-Thach Quach , Erik Debenedictis, Ron Goeke, Pat Finnegan, Derek Wilke, Denis Mamaluy, Harry Hjalmarson, Brian Tierney, Dave Henry, Alex Hsia, Brad Aimone, and Conrad James
Sandia National Laboratories

The size and "depth" of deep neural algorithms are currently limited by available hardware. It is typically not practical to run simulations that require more than one week to run, and hence the neural field is limited to problems that can be run in this length of time with a modern supercomputer (typically 50k-3M CPU/GPU cores). Although impressive results training deep networks using modern GPU clusters have recently been reported, training much larger deep networks and datasets is highly desirable. Numerous groups are making progress in the short term toward this goal, though the development of highly efficient GPU, FPGA, and ASIC cluster architectures, which will likely increase the size of these networks by as much as two orders of magnitude in the short term. For the longer term, we are exploring the use of emerging nanoelectronic resistive memory technologies, which could provide as much as eight orders of magnitude improvement over implementing the same algorithm on a modern CPU. We will report an overview and share recent results from our effort at Sandia to create a neural algorithm accelerator, which includes multidisciplinary work ranging from basic materials science, device fabrication and characterization, through the architecting, theoretical modeling, and simulation of this accelerator.

_____

Energy Recovery and Recycling in Computation: Reversible Adiabatic Logic

Gregory L. Snider[1], Ismo K. Hänninen[1], César O. Campos-Aguillón[1], Rene Celis-Cordova[2], Alexei Orlov, and Craig S. Lent
[1]*University of Notre Dame,* [2]*Tecnológico de Monterrey, Mexico*

Energy use in computation has become the dominant challenge in the design of computational logic and systems. Here energy is dissipated to heat in two ways, by passive dissipation due to leakage, and by active dissipation caused by the processing of information. As supply voltages are lowered to reduce the active dissipation, the passive dissipation increases, so resent research has concentrated on reducing the passive dissipation. Even if

passive dissipation is eliminated, active dissipation in conventional computation will still set a lower limit on total dissipation, limiting future progress.

Recent experiments testing the Landauer principle have shown that, as predicted, there is a minimum limit of dissipation, $k_BT \ln(2)$, if information is destroyed, and that dissipation can be less than $k_BT \ln(2)$, with no lower limit, if information is not destroyed. Since these experiments have shown that ultra-low energy dissipation is possible, the question becomes how to extend these results to real computing systems. One approach is to use reversible logic implemented with adiabatic circuits to avoid information destruction, so that energy can be recovered and loss is minimized in state transitions. In such a system the energy needed to process information is sent to the logic by power clocks, and then returned from the logic when the computation is complete. To achieve overall energy savings the energy returned must be recycled and reused in the next computation, rather that dissipated to heat in the clock generator.

This poster presents reversible adiabatic circuits designed using adiabatic CMOS as a test bed. As an existing technology, adiabatic CMOS can be used to evaluate the performance and active power dissipation of circuits. Simple test circuits and a simple reversible adiabatic microprocessor will be presented. To recycle the energy used in computation, MEMS resonators are proposed as clock circuits. Molecular quantum-dot cellular automata (QCA) is presented as a beyond-CMOS paradigm that maps well onto reversible adiabatic computational systems.

———————————————————

The Dot-product engine: programming memristor crossbar arrays for efficient vector-matrix multiplication

John Paul Strachan, Miao Hu, J. Joshua Yang, Emmanuelle Merced-Grafals, Noraica Davila, Catherine Graves, Eric Montgomery, R. Stanley Williams
Hewlett Packard Labs

Vector-matrix multiplication dominates the computation time and energy for many workloads, particularly neural network algorithms and linear transforms (e.g, the Discrete Fourier Transform). We developed the Dot-product Engine (DPE), an enhanced memory array that exploits the fundamental relationship between row voltage and column current, to realize an analog multiply-accumulate unit with high power efficiency and throughput. We first invented a conversion algorithm to map arbitrary matrix values appropriately to memristor conductances in a realistic crossbar array, accounting for device physics and circuit issues to reduce computational errors. Accurate device resistance programming in large arrays is enabled by closed-loop pulse tuning and access transistors. To validate our approach, we simulated and benchmarked one of the state-of-the-art neural networks for pattern recognition on the DPEs. The result shows no accuracy degradation compared to software approach (99% pattern recognition accuracy for MNIST data set) with only 4 Bit DAC/ADC requirement, while the DPE can achieve a speed-efficiency product of 1,000x to 10,000x compared to a comparable digital ASIC.

———————————————————

Neural Computing at Sandia National Laboratories

Craig M. Vineyard, Erik Debenedictis, James B. Aimone, Michael L. Bernard, Kristofor D. Carlson, Frances S. Chance, James C. Forsythe, Conrad D. James, Fred Rothganger, William M. Severa, Ann E. Speed, Stephen J. Verzi, Christina E. Warrender, and John S. Wagner
Sandia National Laboratories

An understanding of how the brain operates in performing computations has for decades served as an inspiration to build and design computers. Physical limits in computer hardware, advances in neuroscience, and the success of artificial neural network software have led to an emergence in neural inspired computing approaches. Over the course of a decade, leveraging neural principles for computational benefit at Sandia National Laboratories has

resulted in a variety of research focusing upon neural theory, modeling and simulation, and neural inspired application development.

At the intersection of computability and neuroscience understanding, neural theory research strives to yield a formal understanding of various aspects of the brains operation for computational benefit. An analysis of the tradeoffs associated with computations performed and space requirements yields insights into the low power operating regime of brains. An information theoretic analysis of neural ensembles yields insights into representation and encoding paradigms. And neural ensemble research has investigated adaptive encodings.

In addition to developing theories regarding neural computing, we have also led several modeling and simulation efforts. Leveraging the high performance computing expertise at Sandia National Laboratories we have developed and analyzed several large scale neural models, advocated uncertainty quantification and sensitivity analysis in neural models, and developed a language and tool for describing large-scale dynamical systems.

And finally, the insights gained by an increased understanding of neural theory and building modeling and simulation capabilities has allowed us to develop a variety of neural applications. These include the development of neural inspired machine learning algorithms, a neural modeling approach to decision making, neural circuit approaches for information encoding and retrial, and the development of neural inspired computer architectures.

_____

References

[Agarwal 15] Agarwal S, Quach T, Parekh O, Hsia AH, DeBenedictis EP, James CD, Marinella MJ and Aimone JB (2015). Energy Scaling Advantages of Resistive Memory Crossbar Based Computation and its Application to Sparse Coding. *Front. Neurosci.* **9**:484. doi: 10.3389/fnins.2015.00484

[Amazon 15] "Samsung 850 PRO 1 TB 2.5-Inch SATA III Internal SSD (MZ-7KE1T0BW), $469, which is a 32-layer technology," downloaded 12/21/2015.

[Anderson poster] Neal G. Anderson, Ilke Ercan, and Natesh Ganesh, "Revealing Fundamental Efficiency Limits for Complex Computing Structures: The FUELCOST Methodology," RCS 4 poster http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Andersonposter.pdf

[Bresniker Thu 12:30] Kirk Bresniker, "Adapting to Thrive in a new economy of memory abundance," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4BresnikerThu1230.pdf

[Bresniker poster] "Adapting to Thrive in a New Economy of Memory Abundance" http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Bresnikerposter.pdf

[Conte Thu 8:30] Tom Conte, "IEEE Rebooting Computing Summit #4," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4ConteThu830.pdf

[Datta Fri 10:30] Suman Datta, "Recent Progress in Steep Slope Transistors," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4YoungDattaEliMattThu145.pdf

[DeBenedictis poster] Erik DeBenedictis, Neal Anderson, Michael Frank, Natesh Ganesh, R. Stanley Williams, "Improving Computational Energy Efficiency: A New Approach Using Nonlinear Dynamics and Chaos," RCS 4 poster. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4DeBenedictisposter.pdf

[Franzon poster] Paul Franzon, "Cortical Processing," RCS 4 poster.

[Frank 2014] David J. Frank, IBM, "Reversible Adiabatic Classical Computation – an Overview," RCS 2 presentation (NOTE THIS IS RCS 2 not RCS 4.)

[Gargini Wed 10:45] Paolo Gargini, "ITRS 2.0," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4GarginiThu1045.pdf

[Hammerstrom Thu 5:15] Dan Hammerstrom, "name," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4HammerstromThu515.pdf

[Hasler 13] Hasler, Jennifer, and Bo Marr. "Finding a roadmap to achieve large neuromorphic hardware systems." *Frontiers in neuroscience* 7 (2013).

[Holmes 15] Scott Holmes, Alan Kadin, and Mark Johnson, "Superconducting Computing in Large-Scale Hybrid Systems," Computer, Dec 2014 pp. 34-42.

[Joyner Thu 11:00] William Joyner, "SRC," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4JoynerThu1200.pdf

[Kadin poster] A. Kadin, D. Holmes, M. Johnson, "Superconducting Computing in Large-Scale Hybrid Systems," RCS 4 poster http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Kadinposter.pdf

[Khasanvis Thu 8:45] "Probabilistic computing," RCS 4 presentation.
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4KhasanvisThu845.pdf

[Khasanvis 15] Santosh Khasanvis, et. al., "Architecting for Causal Intelligence at Nanoscale," Computer, Dec 2015, pp. 54-62.

[Kim 15] S. Kim, et al., "NVM Neuromorphic Core with 64k-cell (256-by-256) Phase Change Memory Synaptic Array with On-Chip Neuron Circuits for Continuous In-Situ Learning," IEDM 2015.

[Koella Thu 11:15] Williams Koella, "NSCI," RCS 4 presentation.

[Kumar poster] A. Kumar, Z. Wan, S. Iyer, W. Wilcke, "Scaling Up of Neuromophic Systems using 3D Wafer Scale Integration," RCS 4 poster.
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Kumarposter.pdf

[Lu poster] Y. Lu, et. al., "Low-Power Image Recognition Challenge (LPIRC)," RCS 4 poster.
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Luposter.pdf

[Matheny Thu 5:30] Jason Matheny, "IARPA," RCS 4 verbal presentation.

[Manheimer Thu 10:15] "Cryogenic Computing Complexity (C3)"
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4ManheimerThu1015.pdf

[Manheimer poster] Marc Manheimer, IARPA, "Cryogenic Computing Complexity (C3)"
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Manheimerposter.pdf

[Marinella Fri 10:00] Matt Marinella, "Overview of Emerging Research Memory Devices," RCS 4 Presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4YoungDattaEliMattThu145.pdf

[Marinella Fri 10:00b] Matt Marinella, "Acceleration of Neural Algorithms using Emerging Nanoelectronic Devices," RCS 4 Presentation.
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4MarinellaFri1000.pdf

[Marinella poster] "Acceleration of Neural Algorithms using Nanoelectronic Resistive Crossbars," RCS 4 poster. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Marinellaposter.pdf

[Monroe Thu 8:45] Laura Monroe, "Probabilistic and Approximate Computing," RCS 4 presentation.
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4MonroeThu845.pdf

[Mountain 15] D. Mountain, et. al, "Ohmic Weave: Memristor-based threshold gate networks," IEEE Computer, December 2015, pp. 65-71.

[Mountain Thu 3:45] David Mountain, "Neuromorphic Computing: Out approach to developing applications using a new model of computing," RCS 4 presentation.
http://rebootingcomputing.ieee.org/images/files/pdf/RCS4MountainThu345.pdf

[Nikonov 13] Nikonov, Dmitri E., and Ian Young. "Overview of beyond-CMOS devices and a uniform methodology for their benchmarking." *Proceedings of the IEEE* 101.12 (2013): 2498-2533.

[Pandey 15] Pandey, R., et al. "Demonstration of p-type In 0.7 Ga 0.3 As/GaAs 0.35 Sb 0.65 and n-type GaAs 0.4 Sb 0.6/In 0.65 Ga 0.35 As complimentary Heterojunction Vertical Tunnel FETs for ultra-low power logic." *VLSI Technology (VLSI Technology), 2015 Symposium on*. IEEE, 2015.

[Samsung 15] see
http://www.samsung.com/global/business/semiconductor/minisite/SSD/global/html/ssd850pro/3dvnand.html, downloaded 12/21/2015.

[Snider poster] Greg Snider, et. al., "Energy Recovery and Recycling in Computation: Reversible Adiabatic Logic," RCS 4 poster. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Sniderposter.pdf

[Strachan Fri 10:30] John Paul Strachan, "The Dot-Product Engine (DPE): exploring high efficiency analog multiplication with memristor arrays," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4StrachanFri1030.pdf

[Vineyard poster] C. Vineyard, et. Al., "Neural Computing at Sandia National Laboratories," RCS 4 poster. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4Vineyardposter.pdf

[Whitehouse 15] Lloyd Whitman, Randy Bryant, and Tom Kalil, "A Nanotechnology-Inspired Grand Challenge for Future Computing," on White House blog https://www.whitehouse.gov/blog/2015/10/15/nanotechnology-inspired-grand-challenge-future-computing.

[Whitman Thu 5:00] Lloyd Whitman, "The NNI and the Nanotechnology-Inspired Grand Challenge for Future Computing," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4WhitmanThu500.pdf

[Widrow 60] Widrow, B., and M.E. Hoff, Jr., 1960. "Adaptive switching circuits", IRE WESCON Convention Record, pp. 96-104

[Williams Thu 3:45] Stan Williams, "Rebooting Computing Summit 4: Sensible Machine," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4WilliamsThu345.pdf

[Wong Thu 12:30] H. S. Philip Wong, "Computing Performance: The N3XT 1000x," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4WongThu1230.pdf

[Yablonovitch Fri 10:00] Eli Yablonovitch, "Moore's Law for energy: Searching for a milli-Volt switch," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4YoungDattaEliMattThu145.pdf

[Young Thu 2:15] Ian Young, "Mapping the Pathways to 'Beyond CMOS' Technology for Computation," RCS 4 presentation. http://rebootingcomputing.ieee.org/images/files/pdf/RCS4YoungDattaEliMattThu145.pdf