THE
Next Wave

Vol. 23 | No. 2 | 2022

The National Security Agency's review of emerging technologies

ISSN 2640-1789 (Print) | ISSN 2640-1797 (Online)

NEXT-GENERATION HARDWARE FOR HIGH-PERFORMANCE COMPUTING

[Photo credit: iStock.com/NatalyaBurova]

THE Next Wave
The National Security Agency's review of emerging technologies

GUEST Editors' column

David J. Mountain & Robert J. Runser

[Photo credit: iStock.com/danbaily/Jessica Marx]

> ***It's not only merely dead, it's really most sincerely dead.***
>
> *~Computer engineer who shall remain anonymous~*

A brilliant observation by Gordon Moore in 1965 [1], and a precise technical analysis by Robert Dennard et al. in 1974 [2], paved the way for multiple decades of general purpose computing advancements by following one straightforward imperative—make transistors smaller. Consider this comparison:

- In 1971, Intel developed the 4004 microprocessor, which had a 4 bit data path, 740 kilohertz clock, and was fabricated on 10 micron technology.
- By 2005, Intel's Pentium microprocessor had a 64 bit data path, 3.8 gigahertz clock, and was fabricated on 0.09 micron technology.

That's an 80,000-times improvement in raw compute capability—a 40 percent increase per year—primarily driven by a reduction in transistor area of 12,000 times. This one rising tide lifted all boats, enabling computers as a general purpose technology to become ubiquitous in business, personal, and public sectors.

Alas, no exponential can continue forever; since early in the 21st century, a variety of issues, both technical (i.e., power dissipation, leakage currents, on-chip wire resistance) and non-technical (i.e., rising costs of fab facilities and masks, manufacturing and chip design complexity), have greatly eroded the value of complementary metal-oxide-semiconductor (CMOS) transistor scaling in improving computers. Fundamental research into new materials, novel architectures, and new domains for computation will drive advances in the years to come. These advances will be less predictable, require a greater diversity of solutions, and will demand highly creative ideas from the research community. NSA Research is an active participant, deeply involved in a wide variety of explorations, understanding and shaping the next wave of computing technology. We present some of our work in this special issue on computing hardware.

We start this issue with a look at the most basic technology for computing—materials. In "Beyond silicon: Novel materials heterostructures for future high-performance computing," Adam Friedman et al. describe an exciting array of novel materials and devices based on unique properties created by two-dimensional or topological effects that use spintronic, photonic, and magnetic effects to manipulate information.

In "Optical coprocessor generates bright future for probabilistic computing," John T. Daly moves up the computing stack to a functional design, describing how inherent properties of optics can be combined with mathematical techniques to perform multiplications. The emergence of nanoscale optics is examined as a path toward practical use of these ideas.

Probabilistic processing is further explored in "The Ising machine—A probabilistic processing-in-memory computer," by Lauren Huckaba. The basic Ising model

[1] Moore GE. "Cramming more components onto integrated circuits." *Electronics*. 1965;38(8). Available at: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf.

[2] Dennard RH, Gaensslen FH, Yu H, Rideout VL, Bassous E, LeBlanc AR. "Design of ion-implanted MOSFETs with very small physical dimensions." *IEEE Journal of Solid-State Circuits*. 1974;9(5):256–268. Available at: https://doi.org/10.1109/JSSC.1974.1050511.

and Ising problem are described, followed by a description of how coupled oscillator systems can be configured to solve optimization problems of interest.

We return to nanophotonics technology in “Cacheless computer architectures: 3D integration of optical interconnects and novel memory,” by Eric Cheng and S. J. Ben Yoo. Cleverly combining the technologies of silicon photonics, low-latency memory, and optical vias for vertical packaging is shown to be highly advantageous for high-performance data analytics.

Continuing our focus on applications, Roger Pearce and Geoffrey Sanders describe how exploratory data analytics can be deployed at unprecedented scale by using new solid-state devices in “Persistent memory as the substrate for HPC-scale graph analytics.”

Our final article, “Hardening the hardware supply chain: Standardized artifacts enable automated accountability” by Andrew Medak, reminds us of the need for supply chain security, and provides a model for how that can be accomplished even when vulnerabilities exist at many points along the way.

We thank the authors for their fantastic work and their willingness to provide a partial glimpse into NSA Research. We also want to thank Jessica for her efforts in “herding the technical cats” and bringing this issue into print. We hope you enjoy this issue of *The Next Wave,* and welcome aboard!

David J. Mountain
Advanced Computing Systems
Research Directorate, NSA

Robert J. Runser
Technical Director
Research Directorate, NSA

Contents

The Next Wave is published to disseminate technical advancements and research activities in telecommunications and information technologies. Mentions of company names or commercial products do not imply endorsement by the US Government. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the NSA/CSS.

This publication is available online at http://www.nsa.gov/thenextwave. For more information, please contact jsmarx@uwe.nsa.gov.

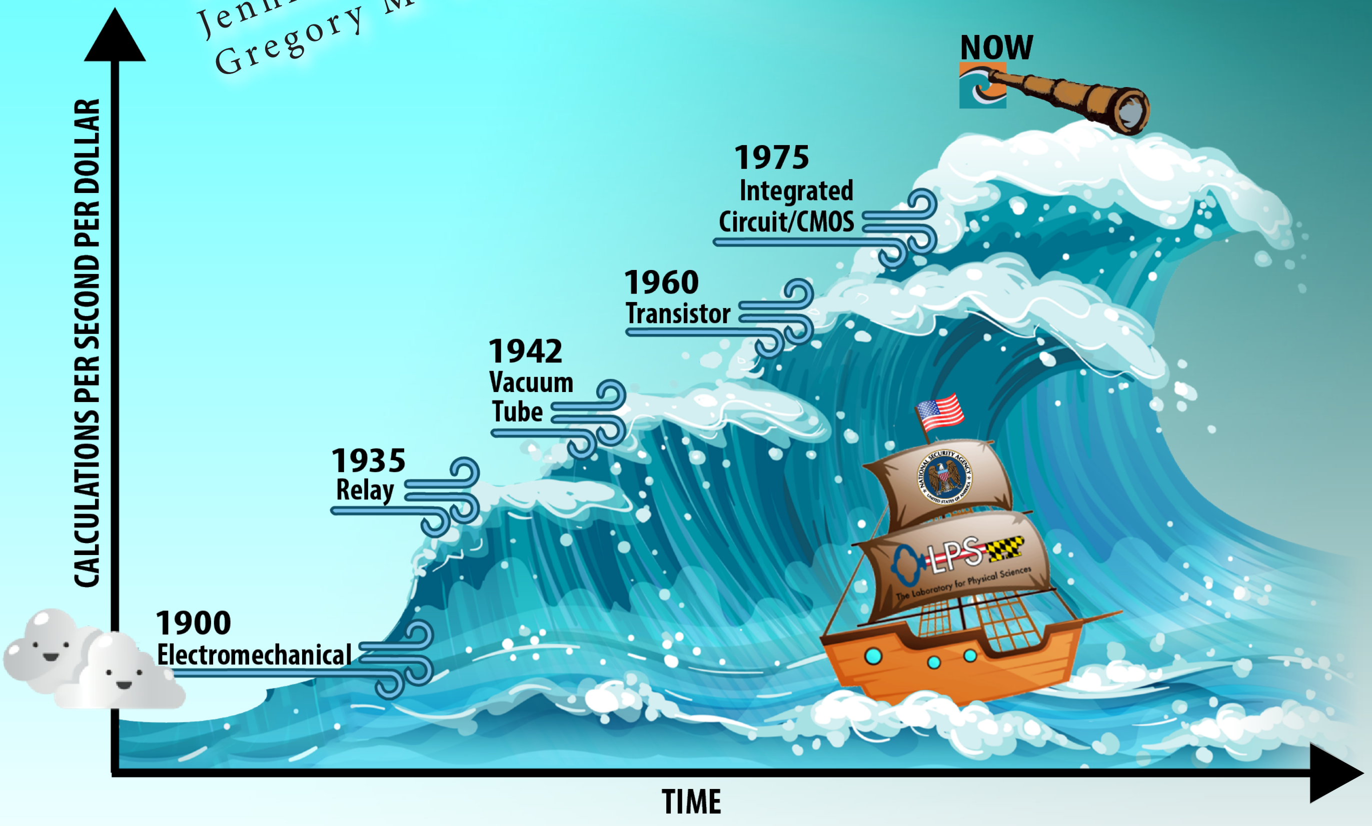
ISSN 2640-1789 (Print)
ISSN 2640-1797 (Online)

Vol. 23 | No. 2 | 2022

[Photo credit: iStock.com/Maxiphoto]

Beyond Silicon: Novel Materials Heterostructures for Future High-Performance Computing

Adam L. Friedman, Aubrey T. Hanbicki, Jennifer E. DeMell, Nicholas A. Blumenschein, Gregory M. Stephen



Return to CONTENTS

For the last 75 years, the operation of the majority of electronics has been based on manipulating electron charge in the elemental semiconductor silicon. The basic device operation of the silicon transistor is referred to as complementary metal-oxide-semiconductor (CMOS). While this paradigm was amazingly successful for generations, CMOS is reaching its physical limits and will be unable to keep pace with the speed, energy, and size requirements critical for data manipulation and the storage needs of the commercial, defense, and intelligence communities. Research into the next generation of materials and devices is therefore essential to enable future high-performance computing (HPC) platforms.

Consider that technologies are like waves (as in figure 1)—they slowly build up and then eventually crest while a new wave builds—a suitable analogy for this publication. Moore's law, the empirical observation that the number of transistors on a chip doubles approximately every two years, is the de facto driving force behind the current wave and has resulted in exponential growth in computing power. Troublingly, no new wave has been clearly identified. Without a firm path beyond the current paradigm—the Next Wave—critical computing needs will not be met.

Background

To enable future HPC systems, we need to imagine and create innovative solutions to fuel the next wave. This includes developing devices that incorporate alternate-state variables, for example, electron spin (i.e., spintronics), magnetism, or photonic devices that utilize inherent material properties besides electron charge to manipulate information. Ultimately, a "materials-by-design" solution will be feasible by combining different materials to create the necessary properties. Beyond the development of new devices, the entire advanced computing system problem space—material, device, architecture, etc.— must be holistically considered, a concept referred to as codesign.

FIGURE 1. The waves of technology from 1900 to now are much like waves in an ocean. A new wave increases the total calculations per second per dollar with each new technology. The wave builds up slowly, then quickly gains momentum and takes over, finally cresting while a new wave begins building. There is no clear next wave on the horizon. Technology users ride the waves like boats in the ocean. Researchers work to identify new waves before the prior waves crash onto them.

Research into alternate-state variables, coupled with breakthroughs in materials science that include entirely new materials classes with a host of advantageous properties, is generating a considerable amount of excitement in the HPC field for novel materials and devices. In particular, devices fabricated from novel two-dimensional (2D) materials, topological Dirac materials, or novel magnetic materials are expected to offer an avenue for lower-power, higher-performance memory and logic beyond Moore's law [1, 2]. However, basic research must be performed to identify the best materials and alternate-state variables to use from the available new classes. In addition, we must better understand how to make usable devices with combinations of these materials that optimize properties and solutions.

The Laboratory for Physical Sciences (LPS) has a novel materials and devices research program, and our objective in this program is to explore the properties of devices that incorporate emerging materials such as topological Dirac materials, 2D materials, and magnetic-phase-change materials with a goal of creating better capabilities (e.g., faster speed, lower-power, greater versatility/functionality) for memory, logic, and HPC beyond the paradigms established by Moore's law. This requires high-risk, high-reward research that is focused on understanding the unknown basic properties of promising new materials, determining exactly which properties can be exploited to the greatest effect, and designing, fabricating, and testing devices that use these properties and that can be quickly transitioned into technologies with the potential for disruptive, non-incremental discovery and implementation. In this article, we will briefly discuss two recent prototype devices developed at LPS: topologically enabled spintronic devices further exemplified by a cadmium arsenide (Cd_3As_2)/fluorographene heterostructure non-local spin valve, and a metamagnetic iron rhodium (FeRh) memory element.

Topologically enabled electronics

Topological materials have special properties enabled by their physical structure, or topology. Topological Dirac materials, a recently discovered class of materials, have the potential to enable spintronics as the defining technology of future electronic systems [3]. In these materials, the conduction and valence energy bands meet at a single "Dirac point," resulting in a host of exciting properties such as relativistic electronic transport and dissipationless spin transport.

The most well-known Dirac material is graphene, a 2D version of graphite. Another is bismuth selenide (Bi_2Se_3), a so-called topological insulator because only its surface conducts electricity [4]. Another promising new material is Cd_3As_2, which is a topological Dirac semimetal (TDS). This material can be tuned between multiple quantum phases (QPs), allowing for a truly multifunctional material [5]. While CMOS relies on toggling between charge states, a system made with Cd_3As_2 devices could reversibly switch between computing modes by toggling between the QPs—a process that is both faster and lower energy, while at the same time allowing inherent reprogrammability and multifunctionality.

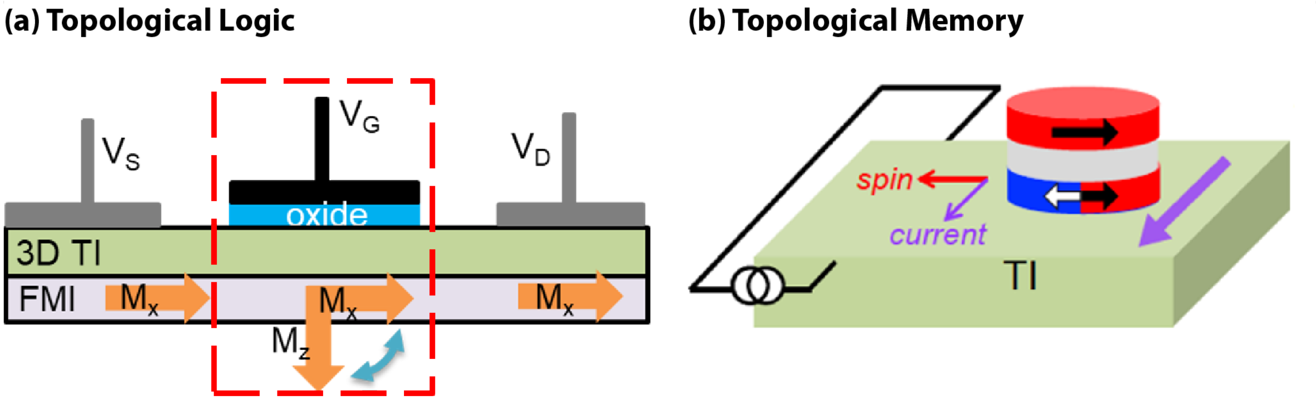
In a recent, exciting agency first, LPS received funding support from the Office of the Under Secretary of Defense, Research & Engineering as part of an Applied Research for the Advancement of Science and Technology Priorities (ARAP) program in collaboration with the Army Research Laboratory, Naval Research Laboratory, and the Air Force Research Laboratory. The purpose of the program is to design, fabricate, and prototype new Dirac material/topological memory and logic devices with technologically disruptive potential. Two identified pathways are through topological magnetoelectronic (TMET) logic and with magnetic topological memory (MTM).

The TMET project is concentrated on a device that works by using an external electric field to toggle the direction of a magnetic field in close proximity to a topological Dirac material, thereby using the magnetic interaction to switch a transistor-like device ON or OFF. This quantum transistor would operate at 1,000 times less power and 10 to 1,000 times faster speeds than today's state of the art. The MTM device operates by exploiting a unique property of a topological Dirac material—its spin is locked to its momentum. Using a process called spin-orbit-torque (SOT) to change the magnetic moment of an element, called a magnetic tunnel junction (MTJ), it provides the same functionality as commercial magnetoresistive random-access memory (MRAM). However, SOT-MTJs are significantly faster and use a fraction of the energy compared to commercial MRAM. The charts in figure 2 summarize the ARAP technologies.

For topologically enabled devices to become reality, we first must better understand the fundamental behavior of the materials. For example, how long does it take spins to scatter? How efficient is the spin-charge conversion? Recently, LPS was one of the first to answer these questions by investigating the spin Hall effect (SHE) in topological Dirac materials using a method previously used to investigate atomically heavy metallic materials. The SHE occurs when a charge current passes through a material with a high spin-orbit coupling (like a topological material), which functionally acts as an internal magnetic field, thereby producing a perpendicular spin current. The effect is reversable: a spin current also produces an orthogonal charge current via the inverse spin Hall effect (ISHE).

Thus, the SHE and ISHE directly measure how efficiently a material converts charge currents into spin currents and vice versa. In addition, these measurements can also tell us the spin diffusion length and spin relaxation time, which are measures of how far the spin travels before it scatters and how much time elapses between scattering events. A more robust spin current will directly lead to more robust information manipulation in future devices.

Return to CONTENTS


(a) Topological Logic

(b) Topological Memory

PROPERTY	HIGH-PERFORMANCE CMOS	TMET
Subthreshold slope	Typical: 70mV/dec Theoretical min @ RT: 60mV/dec	0.7–40mV/dec (Value limited by magnetism, not temp)
Operational power loss per 32-bit ALU	0.1 mW	1 μW
Device optimization	Reaching limits of engineering (e.g., FinFet, Chiplet)	New parameter space to optimize Structural device optimizations are transferable (e.g., FinFet)
High-frequency response	Requires complex HEMT material stacks	Naturally high-mobility channels

PROPERTY	CMOS NVM (E.G., FLASH, STT MRAM)	TOPOLOGICAL MRAM
Bit-switch energy cost	10^{-9} J	10^{-18} –10^{-15} J
Switching speed (RAM is > 1 GHz)	100–1000 MHz Only USB storage	1–10 GHz Ferromagnet 1 THz antiferromagnet Operable as RAM
Power dissipation if used as RAM	0.1–1 W/bit	10–10^4 nW/bit
Industrial maturity	End of Moore. Seventy years of advancements nearing end.	Non-topological MRAM Shipped in 2016 (Everspin) (Everspin, 1 Gb, 2019)

FIGURE 2. (a) In this schematic of the topological magnetoelectronic transistor (TMET) device, a 3D topological insulator Dirac material (3D TI) sits on a ferromagnetic insulator (FMI) with a magnetic moment that toggled using an oxide gate (red dotted line). The device toggled on/off by switching the direction of the magnetic moment in the FMI. The advantages are indicated in the chart by color: red fails to meet future requirements, yellow barely meets future requirements, and green fully meets future requirements. **(b)** In this schematic of the magnetic topological memory device, a charge current in a topological Dirac material is naturally transduced into a spin current. The spin current interacts with a magnet that is a layer in a magnetic tunnel junction (MTJ). MTJs are created by sandwiching an insulating material between two ferromagnetic materials. When the magnetic moments of the two ferromagnets are parallel/antiparallel, the resistance is low/high. The resistance state is written or read depending on the applied current. The advantages are indicated in the chart by color.

The SHE and ISHE measurements are performed as shown in figure 3. This geometry, called a Hall bar, is the basis for a wide variety of electronic measurements. When a conventional charge current passes between one set of side contacts (left), it produces a spin current in the central channel via the SHE. When this subsequent spin current reaches a second set of contacts (right), the spin current converts back into a charge current via the ISHE. Because the circuit does not directly connect these two sets of contacts, the resulting charge current in the contacts on the right manifests as a voltage. By applying a magnetic field along the initial charge current path, we can disrupt the flow of spins and measure the resulting effect on the voltage. The applied field causes the spins to precess, or rotate around an axis, reducing the measured voltage as a decaying oscillatory function of the applied field [6]. The voltage decrease is directly proportional to the spin-charge conversion efficiency of the material, and its decay gives us the spin diffusion length and spin relaxation time.

Metals like platinum (Pt) and tungsten (W) are currently used in spin-orbit torque-based MRAM designs and have spin-charge conversion efficiencies on the order of 0.05-0.1. The efficiency for topological materials can be more than 10 times larger. The topological insulator bismuth antimony (Bi$_{0.9}$Sb$_{0.1}$) has an efficiency as high as 52 [7].The efficiency is directly proportional to the current required to switch a bit: 10 times higher spin efficiency will result in 10 times less needed current and therefore 10 times less power. The spin diffusion length and relaxation time gives limits on the speed and dimensions of the devices, as any manipulation of the spins must occur before the spin information disappears. Based on these measurements, we can begin to select materials that optimize our device designs.

Cd$_3$As$_2$/Graphene spin valves: A novel topological Dirac material/2D heterostructure

Cd$_3$As$_2$ is an excellent candidate for topologically enabled electronics. Cd$_3$As$_2$ has a tunable quantum phase, a high mobility, microns-long spin diffusion lengths, and high spin-charge conversion efficiency. We have also identified a 2D material, fluorographene, as an excellent material to use to couple to the Cd$_3$As$_2$ to build a spintronic heterostructure. We want to combine these materials because when disparate materials (e.g., Dirac materials, 2D materials, and novel magnetic materials) form heterostructures, the individual properties of each material can overlap due to atomic-level proximity. As an example of how powerful this

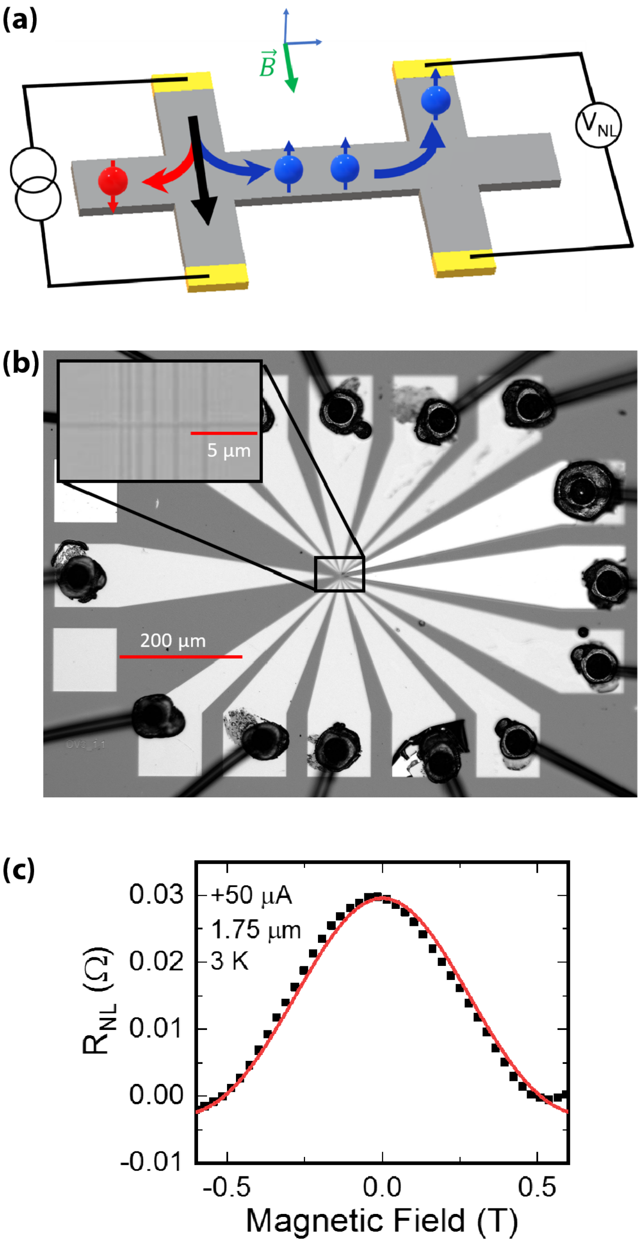


FIGURE 3. **(a)** This schematic of a spin Hall effect (SHE) device shows a charge current flowing through the contacts on the left, which generates a spin current. When the spin current reaches a second set of contacts outside of the charge current path (on the right of the schematic), it creates a measurable voltage. **(b)** This optical image is of a topological SHE device. (R_{NL} stands for non-local resistance.) **(c)** When a magnetic field is applied in plane, the spins begin to precess. Sweeping the field causes dephasing (black dotted curve). The data are fit to a model (red line) and important parameters can be extracted.

technique can be, consider a stack of 2D materials. These are single layer sheets ranging from one to five atoms thick, are mechanically flexible, can be grown in large areas, and can have various electronic states (e.g., semiconductor, metal, semimetal, superconductor, insulator, etc.). They are discretely stackable so they can be used to create true materials by design, where the final bulk heterostructure contains material properties made to order [8]. Two layers of graphene, atomically thin layers of carbon atoms, can be stacked on top of each other and, depending on the relative angle between them, can be insulating, metallic, or even superconducting.

The fundamental spintronic device is the nonlocal spin valve (NLSV). Here, spin current in a channel is compared to a magnet moment in a detector contact. Depending on the relative orientation of the contact with the injected spin moment, there will be a high- or low-magnitude resistance in the device. Figure 4 shows an optical image of a NLSV along with a schematic representation. Two tunneling ferromagnetic contacts (center lines, fluorographene/magnesium oxide (MgO)/Permalloy (Py) are placed on top of a spin-transport channel, in this case Cd_3As_2[9]. Py is a magnetic alloy containing 80 percent nickel and 20 percent iron.

The spintronic behavior of the NLSV is entirely enabled by the fluorographene/MgO tunnel barrier. Electronic tunneling is a quantum mechanical effect occurring when electrons "tunnel" or pass through materials that classically should block them. As an analogy, one can think of a ball thrown at a brick wall. Classically, the ball will bounce back. However, quantum mechanically, the ball will occasionally go through and come out on the other side. For the NLSVs, this barrier is essential due to the electronic properties of the ferromagnet (FM) and the spin channel. The FM is a normal metal, but the spin channel is a semimetal; it conducts current but far less efficiently than the FM. Without the tunnel barrier, the spin channel could not handle the number of electrons trying to move through it, so they bounce off the interface and interfere with the ones that do get through. The tunnel barrier acts as a sort of flow regulator, limiting the number of electrons that get through and allowing for a much smoother flow of electrons into the device.

Graphene makes an excellent tunnel barrier due to its 2D nature. Graphene is conductive in-plane, but highly resistive out-of-plane. It can be discretely stacked (at 0.3 nanometers thick!) onto any surface, is self-healing, is pinhole free, and can further serve

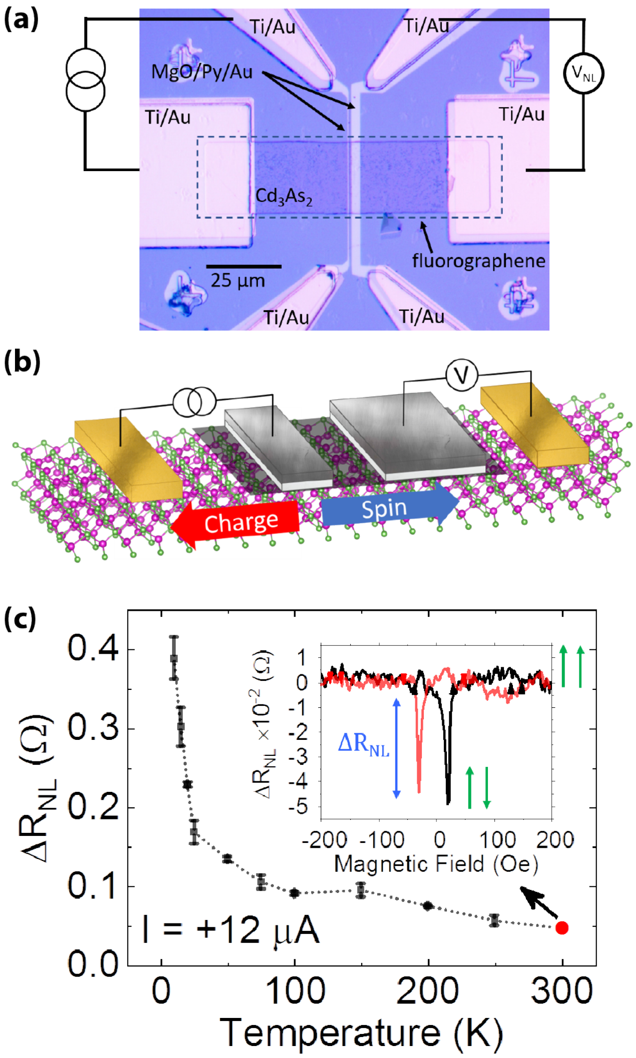


FIGURE 4. (a) In this optical image of a Cd_3As_2/fluorographene non-local spin valve, current flows between the nonmagnetic titanium/gold (Ti/Au) reference contact and the fluorographene/MgO/Py/Au ferromagnetic tunneling contact as shown on the left side of the device. A non-local voltage measured due to a pure spin current is then measured between a second set of similar contacts on the right of the device. **(b)** In this illustration, the fluorographene/MgO/Py/Au contacts are different widths to exploit shape anisotropy, allowing the magnetic moments in the contacts to switch at different fields, leading to the resistance states seen in the inset of **(c)** where the green arrows indicate a low/high magnitude resistance state when the relative magnetic moment orientations are parallel/antiparallel. The black/red curve in the inset is for sweeping the magnetic field from negative/positive to positive/negative. The magnitude of the resistance change is measured as ΔR_{NL}. The inset is at room temperature as indicated in the large curve of **(c).**

as a chemical diffusion barrier to prevent oxidation and unwanted alloying. Additionally, it can be grown in large area sheets inexpensively in a simple furnace reactor [10]. When graphene is exposed to xenon difluoride (XeF_2) gas, the fluorine ions bond to the graphene surface to create fluorographene. This layer is completely insulating and serves as an atomically thin barrier between the FM and the spin channel.

As a charge current passes from a FM contact, through the tunnel barrier, into the Cd_3As_2 channel and out of a reference contact [titanium/gold (Ti/Au)], a spin current is also produced that radiates outward. Because the spin current travels in all directions, unlike the charge current, it is detected in the second set of FM/Au contacts. The measured electrical resistance at the second FM will be higher when the magnetic moments of the two FMs are antiparallel compared to when they are parallel, giving the HIGH and LOW states necessary for digital memory. The magnetic moments are switched using an external magnet as described in the caption of figure 4. Alternatively, the spin current simply toggles on and off by decreasing the spin relaxation time in the channel using, for example, a gate voltage, allowing logic operations with the same device.

Recently we demonstrated high-quality spintronic switching in a Cd_3As_2/graphene NLSV heterostructure device from cryogenic temperatures up to room temperature. Our devices, the first NLSVs to utilize Cd_3As_2, operate with a 10 times larger signal than silicon-based devices. Moreover, this is just the first step. Our ongoing research includes controlling the quantum phase of the Cd_3As_2 in the device by changing the channel thickness, applying an electric field through external gating, or atomically doping the Cd_3As_2 films.

Metamagnetic iron rhodium

Another possible way to improve spintronic switching behavior and realize next-generation topologically enabled logic or memory elements is to incorporate ferromagnets that are more complex with greater functionality and improved material properties. One material of particular interest is FeRh. FeRh possesses a temperature-dependent metamagnetic phase change where it transitions from antiferromagnetic to ferromagnetic (AFM-FM) with increasing temperature [11]. The temperature at which this transition occurs

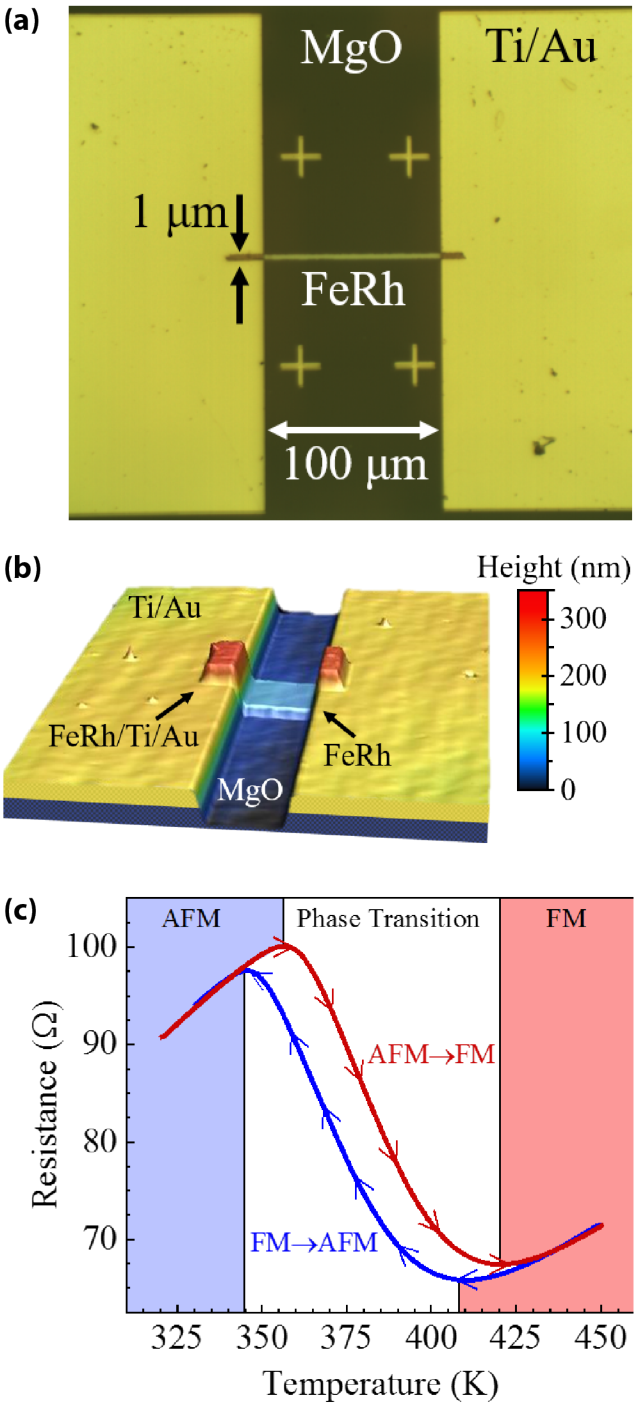


FIGURE 5. (a) This diagram is a top-view optical image of the fabricated device with FeRh wire width and length of 1 micrometer (µm) and 100 µm, respectively, and **(b)** is a 3D optical image of the device. **(c)** This graph shows the resistance measurement while varying the ambient temperature from 320–450 K. Red and blue curves represent heating and cooling cycles, respectively. Background shading colors denote temperature regimes at which the FeRh is antiferromagnetic (AFM, blue), ferromagnetic (FM, red), and in transition (white).

(T_{Cr}), can be fine-tuned using various fabrication techniques, such as substitutional doping [12] and patterning [13]. According to Pouillet's law, expansion of a unit cell will naturally cause its electrical resistance to decrease for a given length of material. Therefore, by careful manipulation of temperature, both the magnetic state and electrical resistance can be selected, allowing its use in switching device applications. This becomes even more attractive when considering the 350 femtosecond AFM-FM transition time, translating to an operating frequency of nearly 3 terahertz [14].

Optical images of simple FeRh devices that we fabricated by standard lithographic methods are shown in figures 5(a) and (b). The simplicity of fabrication is another advantage of working with FeRh. Figure 5(c) demonstrates the AFM-FM transition. The measurement begins at a starting temperature of 320 kelvins (K). Upon increasing the temperature (red curve), the wire resistance also increases. The AFM-FM transition begins once the FeRh temperature surpasses 365 K, accompanied by a decreasing resistance. The effect persists until the temperature reaches 420 K, indicating that the wire has fully transitioned into the ferromagnetic phase. The opposite effect occurs when cooling the wire (blue curve).

To make a useful device, the state must be controlled via current rather than temperature. According to the Joule-Lenz law, an electrical current through the wire will cause the FeRh temperature to increase until it thermally stabilizes. In figure 6(a), a pulsed current controls the temperature. The FeRh temperature and subsequent resistance are held to a constant value by a constant current (I_{Read}). Here, the FeRh will remain in the high-resistance AFM phase (OFF state). If the current amplitude is sufficiently increased, the wire temperature will rise past T_{Cr} and the FeRh will transition into the low-resistance ferromagnetic phase (ON state). As shown previously, the FeRh will remain in the ferromagnetic phase until cooled to below T_{Cr}. Therefore, applying a continuous current (I_{Read}) will maintain the present state of the device. Upon reducing the current, the FeRh transitions back into the high-resistance antiferromagnetic phase (OFF state). In figure 6(b), a pulsed current switches the FeRh back-and-forth between antiferromagnetic (OFF) and ferromagnetic (ON) states.

One possible application of a device with this hysteretic behavior is as a memristor, the basic memory component of many neuromorphic circuit designs. The device demonstrated here uses the magnetization

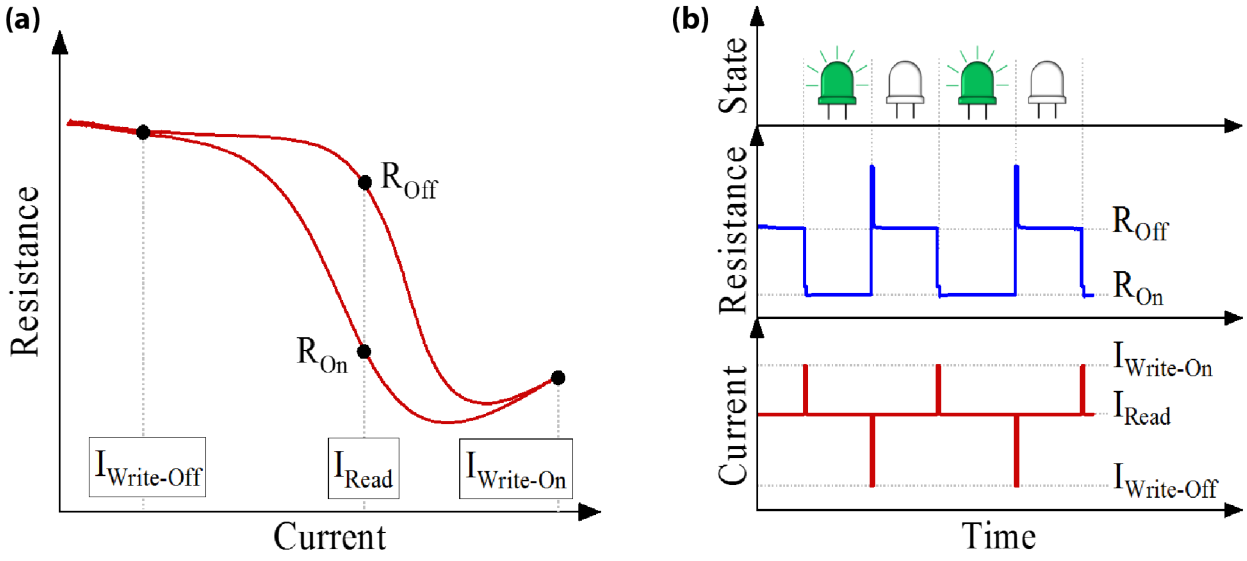


FIGURE 6. (a) This graph shows FeRh resistance as a function of current. Joule heating changes the wire temperature and subsequent phase, allowing for antiferromagnetic and ferromagnetic state control via electrical bias. A constant current I_{Read} will allow the FeRh temperature to stabilize. An FeRh transition is made with a short current pulse of $I_{Write\text{-}Off}$ or $I_{Write\text{-}On}$. **(b)** This graph shows the resistance and current profiles as a function of time. Repetitively switching between pulse current amplitudes of $I_{Write\text{-}On}$ and $I_{Write\text{-}Off}$ causes FeRh phase transitions and substantial change in the wire resistance. The green bulbs represent the ON state.

of the material as a state variable, which inherently provides reproducibility, endurance, and state retention in comparison to charge-based switching devices. The switching capability of these devices is estimated to be on the order of 1 picosecond [14]. For comparison, previously reported devices with more advanced architectures have switching times that range from 50 nanoseconds to more than 100 microseconds [15, 16]. Moreover, one could imagine incorporating a metamagnetic element into the previously described NLSV. The metamagnetic transition would switch off the spin current in the NLSV, thereby allowing transistor-like behavior. Alternatively, the metamagnetic transition could switch the device into a new operational mode.

Conclusion and a beginning: Vision for the future

Moore’s law began as an observation of the density of transistors in a circuit. Serving as a roadmap for the entire semiconductor industry, it has subsequently transformed into a self-fulfilling prophecy. It has been re-imagined, re-invented, and ultimately embraced as a mindset, an approach, and a philosophy. Device scientists and engineers have accomplished everything envisioned by Gordon Moore in 1965 [17] and are quickly approaching the physical limits which portend an end to this path.

One possible path toward a new scientific paradigm can be found by going back in history to even before the 1965 observation by Moore. Richard Feynman, in his famous speech “There’s plenty of room at the bottom” [18], given at the American Physical Society meeting in December 1959, is often credited with inventing the field of nanoscience. He revisited and repurposed a similar line of reasoning in a 1983 speech given at the Jet Propulsion Laboratory and, astonishingly, established the field of quantum computing [19]. Although his prophesies do not specify a pathway to advanced classical HPC, Feynman does establish a way forward by suggesting a holistic, outward-looking approach to device (co-)design. Summarizing this approach, Feynman said, “It would be interesting in surgery if you could swallow the surgeon” [18].

The novel materials and devices that we described here present the beginning of a new era in computing technologies. In his address to the American Physical Society, Feynman said:

> What could we do with layered structures with just the right layers? What would the properties of materials be if we could really arrange the atoms the way we want them?... I can’t see exactly what would happen, but I can hardly doubt that when we have some control of the arrangement of things on a small scale we will get an enormously greater range of possible properties that substances can have, and of different things that we can do.

Indeed, by choosing the best materials and combining them in new ways with operating modes in mind, we can optimize our future computing systems and maintain a critical quantitative edge for the future.

References

[1] Pesin D, MacDonald AH. “Spintronics and pseudo-spintronics in graphene and topological insulators.” *Nature Materials.* 2012;11:409–416. Available at: https://doi.org/10.1038/nmat3305.

[2] IEEE. The International Roadmap for Devices and Semiconductors. 2020. Available at: https://irds.ieee.org/editions/2020.

[3] Wang XL. “Dirac spin-gapless semiconductors: Promising platforms for massless and dissipationless spintronics and new (quantum) anomalous spin Hall effects.” *National Science Review.* 2017;4(2):252–257. Available at: https://doi.org/10.1093/nsr/nww069.

[4] Jamali M, Lee JS, Jeong JS, Mahfouzi‖ F, Lv Y, Zhao Z, Nikolić BK, Mkhoyan KA, Samarth N, Wang JP. “Giant spin pumping and inverse spin Hall effect in the presence of surface and bulk spin–orbit coupling of topological insulator Bi_2Se_3.” *Nano Letters.* 2015;15(10):7126–7132. Available at: https://doi.org/10.1021/acs.nanolett.5b03274.

[5] Goyal M, Galletti L, Salmani-Rezaie S, Schumann T, Kealhofer DA, Stemmer S. “Thickness dependence of the quantum Hall effect in films of the three-dimensional Dirac semimetal Cd_3As_2.” *APL Materials.* 2018;6(2):026105. Available at: https://doi.org/10.1063/1.5016866.

Return to CONTENTS

[6] Abanin DA, Shytov AV, Levitov LS, Halperin BI. "Nonlocal charge transport mediated by spin diffusion in the spin Hall effect regime." *Physics Review B.* 2009;79(3):035304. Available at: https://doi.org/10.1103/PhysRevB.79.035304.

[7] Khang NHD, Ueda Y, Hai PN. "A conductive topological insulator with large spin Hall effect for ultralow power spin–orbit torque switching." *Nature Materials.* 2018;17:808–813. Available at: https://doi.org/10.1038/s41563-018-0137-y.

[8] Friedman AL, Hanbicki AT, Perkins FK, Jernigan GG, Culbertson JC, Campbell PM. "Evidence for chemical vapor induced 2H to 1T phase transition in MoX_2 (X = Se, S) transition metal dichalcogenide Films." *Scientific Reports.* 2017;7:3836. Available at: https://doi.org/10.1038/s41598-017-04224-4.

[9] Stephen GM, Hanbicki AT, Schumann T, Robinson JT, Goyal M, Stemmer S, Friedman AL. "Room-temperature spin transport in Cd_3As_2." *ACS Nano.* 2021;15(3):5459–5466. Available at: https://doi.org/10.1021/acsnano.1c00154.

[10] Bae S, Kim H, Lee Y, Xu X, Park JS, Zheng Y, Balakrishnan J, Lei T, Kim HR, Song Y, Kim Y, Kim KS, Özyilmaz B, Ahn J, Hong BH, Iijima S. "Roll-to-roll production of 30-inch graphene films for transparent electrodes." *Nature Nanotechnology.* 2010;5:574–578. Available at: https://doi.org/10.1038/nnano.2010.132.

[11] Fallot M, Hocart R. "Sur l'apparition du ferromagnetisme par elevation de temperature dans des alliages de fr et de rhodium." *La Rev. Sci.* 1939;77:3.

[12] Le Graët C, Charlton TR, McLaren M, Loving M, Morley SA, Kinane CJ, Brydson RMD, Lewis LH, Langridge S, Marrows CH. "Temperature controlled motion of an antiferromagnet-ferromagnet interface within a dopant-graded FeRh epilayer." *APL Materials.* 2015;3:041802. Available at: https://doi.org/10.1063/1.4907282.

[13] Uhlíř V, Arregi JA, Fullerton EE. "Colossal magnetic phase transition asymmetry in mesoscale FeRh stripes." *Nature Communications.* 2016;7:13113. Available at: https://doi.org/10.1038/ncomms13113.

[14] Pressacco F, Sangalli D, Uhlíř V, Kutnyakhov D, Arregi JA, Agustsson SY, Brenner G, Redlin H, Heber M, Vasilyev D, Demsar J, Schönhense G, Gatti M, Marini A, Wurth W, Sirotti F. "Subpicosecond metamagnetic phase transition driven by non-equilibrium electron dynamics." 2021. Cornell University Library, available at: https://arxiv.org/abs/2102.09265.

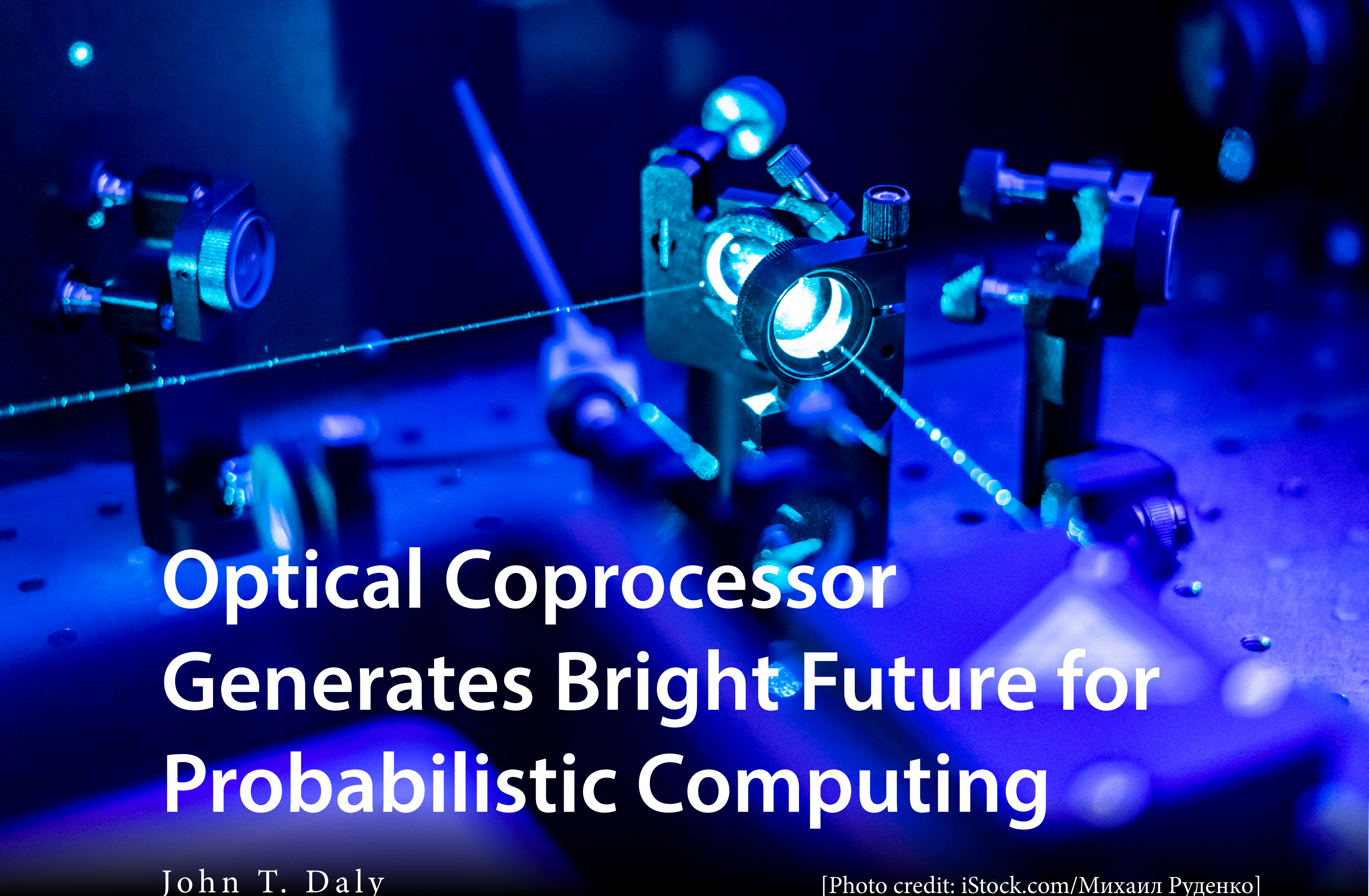
[15] Yang JJ, Strukov DB, Stewart DR. "Memristive devices for computing." *Nature Nanotechnology.* 2013;8:13–24. Available at: https://doi.org/10.1038/nnano.2012.240.

[16] Prezioso M, Merrikh-Bayat F, Hoskins BD, Adam GC, Likharev KK, Strukov DB. "Training and operation of an integrated neuromorphic network based on metal-oxide memristors." *Nature.* 2015;521:61–64. Available at: https://doi.org/10.1038/nature14441.

[17] Moore GE. "Cramming more components onto integrated circuits." *Proceedings of the IEEE.* 1998;86(1):82–85. Available at: https://doi.org/10.1109/JPROC.1998.658762.

[18] Feynman RP. "There's plenty of room at the bottom [data storage]." *Journal of Microelectromechanical Systems.* 1992;1(1):60–66. Available at: https://doi.org/10.1109/84.128057.

[19] Feynman R. "Infinitesimal machinery." *Journal of Microelectromechanical Systems.* 1993;2(1):4–14. Available at: https://doi.org/10.1109/84.232589.



Optical Coprocessor Generates Bright Future for Probabilistic Computing

John T. Daly

[Photo credit: iStock.com/Михаил Руденко]

In 2006, a team of researchers published results for a new type of microprocessor architecture dubbed probabilistic complementary metal-oxide semiconductor (PCMOS) [1]. It used 30 times less power than conventional CMOS to perform computations. *MIT Technology Review* later recognized probabilistic computing as one of the 10 technologies "most likely to change the way we live" [2]. Probabilistic computers can solve complex problems by storing and processing states of zeros and ones that are indeterminate with some probability. They cannot always be guaranteed to provide the same solution to a given problem run more than once, but they can be built using hardware that runs extremely fast and at very low power. Spurred on by these early successes, the Defense Advanced Research Projects Agency (DARPA) Unconventional Processing of Signals for Intelligent Data Exploitation (UPSIDE) program began to explore applications of probabilistic computing to feature extraction from sensor imaging. Their image-processing pipeline used the physics of emerging probabilistic devices including analog nonvolatile memory. In the end, they demonstrated a 100-times performance increase and 1,000-times power efficiency improvement compared to traditional CMOS [3]. Considering these impressive results, why has probabilistic computing not changed the way we live like *MIT Technology Review* suggested? The answer to that question is multifaceted, much like the probabilistic devices themselves, but behind all the challenges in implementing probabilistic computing, there lies a common thread of nondeterminism.

Return to CONTENTS

Nondeterministic hardware is computer hardware that has the capacity to provide more than one answer for at least some, but not necessarily all, operations that have a single correct answer. In a word, it is unpredictable. Nondeterminism is the reason that every application of probabilistic computing to date has been highly specialized, both in terms of the problems it solves and the underlying technologies it uses to achieve nondeterminism. Only certain types of applications are amenable to unpredictability. The DARPA UPSIDE program achieved success with feature extraction from sensor imaging using a particular architecture and technology. What other technologies and applications stand to benefit from these types of physics-based approaches to computation? One such technology is optics. Optics is subject to thermodynamic noise by its basic physics, so it is reasonable to expect nondeterminism in an optical computing device. Optics is an inherently analog technology where both electron and photon noise contribute to the nondeterminism.

This article will explore the potential for using emerging nanophotonic devices to deliver optical arithmetic coprocessors for fast, energy-efficient computing. Such an approach exemplifies the high-performance computing (HPC) codesign methodology, whereby computer systems are purpose-built with specific application requirements in mind.

The application: Multiplication

Most digital computers implement standard precision multiplication using a straightforward cross-product scheme where every digit of the multiplicand multiplies every digit of the multiplier. Such a scheme is said to be quadratic in its complexity, since the number of operations required to complete the muliplication is proportional to the square of the size of the numbers. For small numbers, with fewer than a few hundred bits, this is entirely adequate, but for larger numbers, there are a variety of "subquadratic" algorithms that trade off additional complexity for reduced asymptotic complexity. Of these approaches, Fourier multiplication is of particular interest because of its natural connection to optics [4].

Fourier multiplication

By the convolution theorem, described in figure 1, we know we can write the Fourier transform of a convolution as a pointwise product of Fourier transforms [5]. In other words, by Fourier transforming the multiplier and multiplicand of a product, we can reduce the complexity of the multiplication from quadratic (i.e., $O(n^2)$) to linear (i.e., $O(n)$), but at the cost of calculating the Fourier transforms of the inputs and inverse Fourier transforms of the output. Since the Fourier transform is $O(n \log n)$, a relatively expensive calculation on a digital computer, Fourier multiplication is typically limited in its application to very large numbers (e.g., thousands of digits).

$$\mathscr{F}\{f\} = \int_{-\infty}^{\infty} f(x)\, e^{-ixt}\, dx$$

$$\mathscr{F}\{f * g\} = \mathscr{F}\{f\} \odot \mathscr{F}\{g\} \;\;\Rightarrow\;\; f * g = \mathscr{F}^{-1}\big\{\mathscr{F}\{f\} \odot \mathscr{F}\{g\}\big\}$$

FIGURE 1. The Fourier transform $F\{f\}$ of a continuous function $f(x)$ is defined above. The convolution theorem states that the Fourier transform of the convolution of two functions is equal to the pointwise product of their Fourier transforms. Since convolution in the discrete domain is equivalent to polynomial multiplication, this suggests a straightforward method of multiplying using a pointwise product and the forward and inverse Fourier transforms.

In order to enable our optical coprocessor to scale to the very large numbers, it will need an efficient method of accumulating larger numbers of low-precision multiplications into a single high-precision multiplication. A residue number system (RNS) is one way to build a bridge between low-precision and high-precision arithmetic [6]. The residue number system enables calculation with arbitrarily large numbers by "breaking down" the numbers into sets of smaller numbers using properties of residue arithmetic as illustrated in figure 2. Given a multiplication problem with very large numbers, RNS allows us to break that problem down into a collection of smaller, independent operations which can be implemented in parallel. To utilize this approach however, our optical coprocessor will need to perform many, many calculations of $(a \times b) \bmod m$ very efficiently for different values of m and all in parallel.

Montgomery multiplication

In the previous section, we described at a high level the Fourier transform approach to multiplication and how we intended to apply it to very large numbers by implementing them as a collection of lower-precision

$$(a+b) \bmod m = \big((a \bmod m)+(b \bmod m)\big) \bmod m$$
$$(a-b) \bmod m = \big((a \bmod m)-(b \bmod m)\big) \bmod m$$
$$(a\times b) \bmod m = \big((a \bmod m)\times(b \bmod m)\big) \bmod m$$

FIGURE 2. Residue number system (RNS) arithmetic makes use of a simple theorem of number theory that the remainder of a binary operation divided by *m* is equal to that same binary operation applied to the remainder of each of the operands. This holds true for the binary operators addition, subtraction, and multiplication.

modular multiplications all operating in parallel. In this section, we will look in more detail at the mathematics of combining Fourier multiplication and modular multiplication in the same optical coprocessor without using division, which would be extremely difficult to implement in optics. The solution is to reformulate the modular multiplication problem using Montgomery multiplication [7]. Montgomery recognized that it was often most efficient to rewrite the modular multiplication problem in a different "domain" where performing the division is much, much easier, as illustrated in figure 3.

With the Montgomery multiplication, we have all the necessary mathematical machinery to describe the operation of our optical coprocessor for easy and efficient multiplication of very large numbers. In the next section, we will consider why and how this approach can be implemented using optics technology.

The technology: Optical computing

Optical computing—performing calculations using photons instead of electrons—is an idea that has been around for decades [8, 9]. Like many competing non-CMOS technologies, it ultimately lost out to the exponential growth of transistor counts in CMOS. Optics, being fundamentally limited by the wavelength of light, cannot attain the density of transistors. Furthermore, as an analog technology that is typically limited in precision, optical computing does not compete with digital technologies in terms of computational accuracy. However, with the recent slowing of transistor scaling accompanied by advances in emerging nanophotonic devices, there is renewed interest in exploring the role of optics in computing.

The remainder of this section will describe a coprocessor design based on Fourier optics and discuss possibilities and challenges of improving that design with emerging nanophotonic technology.

Macro-scale optics

Fourier multiplication is known to be faster than traditional schoolbook multiplication, but it is typically not implemented in digital computing because of the time and energy overhead for performing the required Fourier transforms. However, optical computers can complete the Fourier transform with zero overhead using simple lenses and masks. An optically implemented Fourier transform can be used to simplify the multiplication of two n-digit numbers. To multiply with digital electronics, two n-digit numbers are represented as polynomials in powers of two. Digitally multiplying these to numbers has the same complexity as a convolution (i.e., $O(n^2)$). If we use $F\{f\}$ however, we can convert this convolution in the time domain to pointwise multiplications in the frequency domain (complexity $O(n)$). The long-range interference of coherent photons enables "on-the-fly" Fourier transform computation using lenses and masks. Figure 4 illustrates a standard four focal-lengths long (4-f) optical system [10] using the symmetric property of the transform and depicting lines as exemplar rays. The optical domain produces an $O(1)$ Fourier transform compared to digital electronics where the transform is a costly $O(n \log n)$.

$$\text{Given } c=(a\cdot b) \bmod m \text{ choose } r=2^k>m$$
$$\bar{x}=x\cdot r \bmod m \text{ and } M=-m^{-1} \bmod r$$
$$\Downarrow$$
$$\bar{c}=\frac{\bar{a}\cdot\bar{b}+m\cdot\big((\bar{a}\cdot\bar{b}\cdot M) \bmod r\big)}{r}$$

FIGURE 3. Montgomery multiplication provides a means to compute a modular multiplication problem without performing division by the modulus. Assuming that *m* is always odd, choosing an *r* > *m* that is a power of two allows us to replace division with simple bit shifting and masking in binary arithmetic. By precomputing *M* for a fixed set of moduli corresponding to the bases of our residue number system (see the previous section), Montgomery multiplication can be implemented very efficiently.

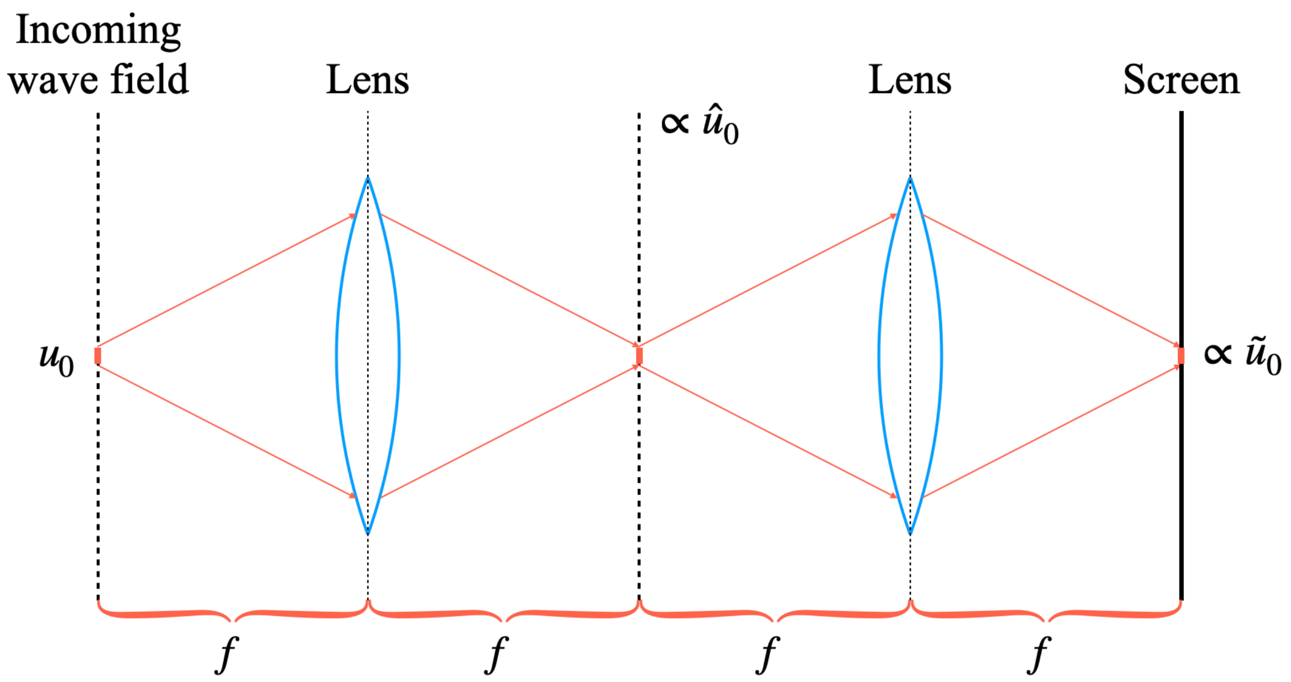


FIGURE 4. Figure representing a simple 4-f optical system. The wave field at the plane $\hat{u}_0$ located one focal length behind the first lens is the exact Fourier transform of the input wave field at plane u_0 By replicating this system with a second lens that performs the inverse transform, the wave field $\tilde{u}_0$ will be a mirror image of $\hat{u}_0$. This is the building block of the optical modular multiplication coprocessor.

While division is generally difficult in optics, division by a power of two is easily done by masking off low-order bits. An implementation of Montgomery modular multiplication, described in the previous section of this article, is a natural approach in optics. Accommodating arbitrarily large inputs requires an efficient architecture that benefits from smaller, fixed-size multipliers. The Fourier transform in the optical domain uses a static configuration of source planes, lenses, and a fixed location objective image plane. One such system designed and evaluated both in simulation and on an optical bench is illustrated in figure 5. The device performs the complete Montgomery multiplication in the Montgomery domain starting with two input values (step ① and ②) on the left, encoded as wave fields. The values are Fourier transformed and multiplied using a spatial light modulator (SLM) or equivalent device to impose spatially varying modulation to the light beam in the Fourier transform plane. This is how we accomplish the pointwise multiplication of the light intensity. Following the multiply of $\bar{a}$ x $\bar{b}$ (step ③), the result is divided into two paths by a beam splitter. The bottom path computes the *m* (($\bar{a}$ x $\bar{b}$ x *M*) mod *r*) term of the Montgomery multiplication using a sequence of fixed masks and filters (steps ⑧ through ⑫), while the top path applies phase corrections to $\bar{a}$ x $\bar{b}$ in order to correctly "add" it to the result of the bottom path (steps ④ through ⑦). The result is masked to achieve the final division by *r* required to recover the solution $\bar{c}$ (step ⑬).

Figure 6 demonstrates a technique to encode 16-bit unsigned integers into the wave field. Numerical values are represented by the intensity of points of light along a diagonal. Lighted points correspond to binary one and dark spaces correspond to zero. The layout avoids optical smearing and "cross-talk" between pixels in the two-dimensional Fourier transform since no pixel is horizontally or vertically aligned with any other pixel. After a "multiplication," the light intensities represents a range of values from dark to light that encode the result. Discrete convolution differs from multiplication in that it does not perform carries between digits. The entire contribution of each pointwise

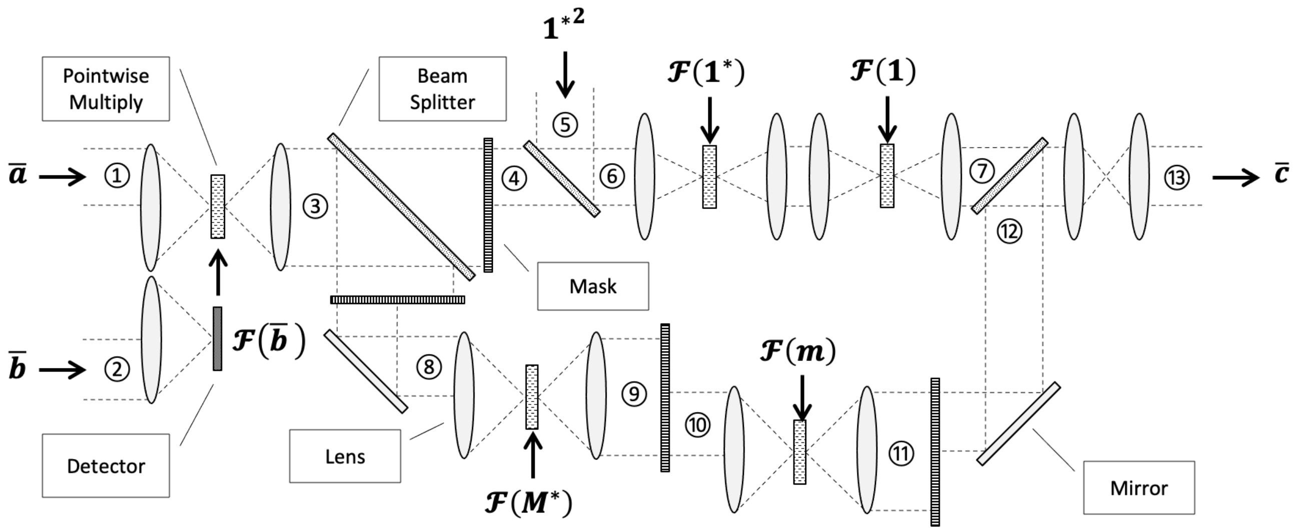


FIGURE 5. In this design of a macro-scale optical device for modular multiplication (a x b) mod m using the technique of Montgomery described in the previous section, values of $\bar{a}$, $\bar{b}$, and M are assumed to be precomputed. Asterisks indicate locations where the image plane is reversed (i.e., mirror image).

input digit multiplication is encoded in each output digit. The multiplier was evaluated with the help of collaborators at George Washington University using both the LightPipes simulator [11] and an optical bench prototype illustrated in figure 6. Simulation and prototype agreed with the theory to 16 bits of precision, thus providing an initial validation of the all-optical approach to multiplication. The prototype anecdotally verified the probabilistic nature of the device as well, since initial results were correct "only about" 80 percent of the time.

Despite the advantages of using an all optical system for multiplication, the bulkiness of macro-scale optical components, in particular the lenses, makes such a system too large (i.e., linear dimensions on the order of 10–100 centimeters) to be practical. Further, the performance of the macro-scale system is limited to the sub-megahertz regime by the speed of the SLM. In terms of energy efficiency, we are able to make an initial estimate based on our simulation results. Assuming that one photon per detector site is sufficient for room temperature detection [12], the end-to-end photon loss across the entire modular multiplication device in figure 5 is around 3,000 times. With a photon energy of approximately 10^{-18} Joules for 200-nanometer light, the projected efficiency of the device is about 1 nanojoule per operation for a 16-bit modular multiply. This is not particularly competitive compared to modern application-specific integrated circuit (ASIC) designs, where the same calculation implemented in 16-nanometer CMOS technology would consume closer to 10 picojoules [13]. Given the limitations of macro-scale optics, in the next section we will turn our attention to consideration of emerging nanoscale optical devices.

Promise and challenge of nanoscale optics

Nanophotonics using metamaterials enables the design and engineering of novel optical systems, overcoming the limitations of macroscale optics. Metamaterials are heterogeneous or highly structured nano-engineered media for next-generation optical systems. These are materials with electromagnetic responses that cannot be obtained from conventional media. For instance, surface plasmon polaritons (SPP) have a square-root dispersion enabling "very high k-vectors for a finite frequency." In other words, energy can be confined within the metamaterial with very little dispersion. These materials operate with sub-wavelength, nano-engineered heterostructures that allow customization of optical properties. Using metalenses and reconfigurable metasurfaces, a 4-f optical system can be constructed in tens of micrometers [14]. Where a macro-scale optical system's performance is limited by the configuration speed of the SLM to sub-megahertz performance, the small electrical capacitance of a nanoscale reconfigurable metasurface allows for processing at rates above the gigahertz range [15]. We can bound the required power by considering the minimum detectability at the photodetector, a value typically about tens of nanowatts for high-speed detectors. Assuming visible or near-infrared optical frequencies, and a bandwidth of tens of gighertz, we find a minimum optical power required for the intermediate result of a 16-bit modular multiply to be 6.5 microwatts. A 10-gigahertz system running at 6.5 microwatts yields a computational efficiency of around 1 femtojoule per 16-bit modular multiply. That is four orders of magnitude more energy efficient than the 10 picojoules per multiplication based on digital CMOS technology. Nanophotonics has potential to be a game changer for high-speed, energy-efficient computation.

The promise of nanophotonics is not without challenges though. Several of the most significant technical challenges are summarized below. They will need to be solved before devices like the one described in this article can fully be realized.

- **Attenuation through the optical system—** Several stages of the subsystem provide less than half the energy to the subsequent stage. Using a phase synchronous light source makes the noise less than it might be, but where the signal is at low levels, nonlinearities, path length differences, and shot noise cause additional loss in the signal-to-noise ratio.
- **Electrical-optical and optical-electrical conversions—**Moving data from electrical to optical domains expends energy as much as 10^{12} above the optical calculation energy. This means that chaining multiple operations in the optical domain will be critical, and accumulated inaccuracies need to be managed.
- **Manufacturing tolerances for metamaterials—** As with most emerging technologies, there are a variety of engineering challenges to be overcome in order to manufacture meta-devices at acceptable cost and yield. Feature variation will be a concern in a production-scale computing system and will need to be addressed as the technology continues to mature.

Conclusion

Emerging nanophotonic technology offers an enticing opportunity for multiple orders of magnitude improvement in arithmetic processing efficiency similar to the gains demonstrated by DARPA UPSIDE for image feature extraction. The potential of computing in excess of a billion operations per second while expending femtojoules per operation could provide significant performance advantages beyond the capabilities of digital CMOS. The future looks very bright for probabilistic computing based on optical coprocessors.

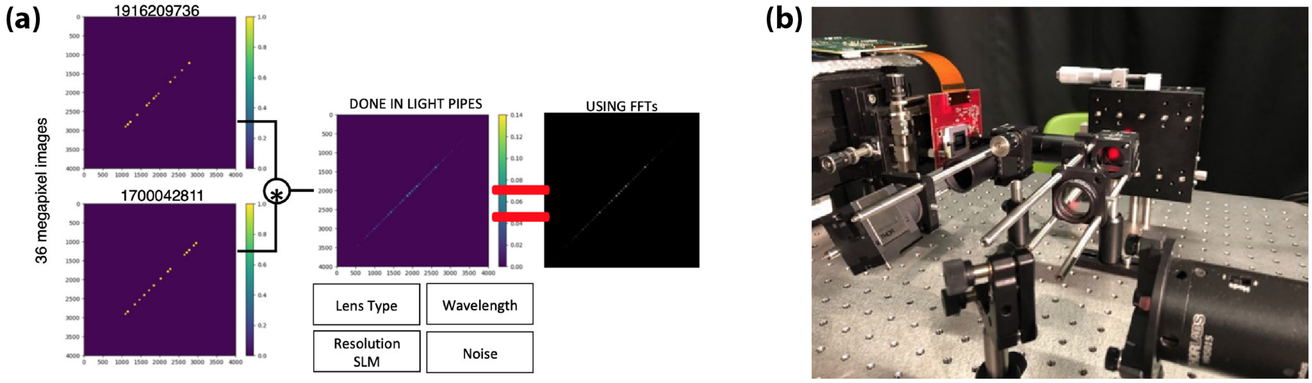


FIGURE 6. (a) Here is the result of a LightPipes simulation compared to an exact fast Fourier transform (FFT) calculation and (b) a physical test on an optical workbench. Experiments performed by George Washington University (GWU) validated a multiplier subcomponent using a spatial light modulator (SLM) and comparing the result to those calculated based on theory using FFTs. The results shows a bit-for-bit exact match from theory to simulation to experiment. [Photo credit: The SORGER Group at George Washington University (GWU).]

References

[1] Chakrapani LN, Akgul BES, Cheemalavagu S, Korkmaz P, Palem KV, Seshasayee B. “Ultra efficient embedded SOC architectures base on probabilistic CMOS (PCMOS) technology.” In: *Proceedings of the Design Automation and Test in Europe Conference (DATE);* 2006: pp. 1–6. doi: 10.1109/DATE.2006.243978.

[2] Jonietz E. “Probabilistic chips.” *MIT Technology Review.* 2008 Feb 18. Available at: https://www.technologyreview.com/technology/tr10-probabilistic-chips/.

[3] Hammerstrom D. “UPSIDE/cortical processor study: DISTAR approved,” presented at the *IEEE Rebooting Computing Summit IV;* 2015 Dec 7; Washington, DC. Available at: http://rebootingcomputing.ieee.org/images/files/pdf/RCS4HammerstromThu515.pdf.

[4] Timmel AN, Daly JT. “Multiplication with Fourier optics simulating 16-bit modular multiplication.” In: *Proceedings of 2018 IEEE International Conference on Rebooting Computing (ICRC);* 2018 Nov 7; McLean, VA: pp. 1–11. Available at: http://doi.org/10.1109/ICRC.2018.8638618.

[5] Schönhage A, Strassen V. “Schnelle multiplikation großer zahlen.” *Computing.* 1971;7:281–292. doi: 10.1007/DF02242355.

[6] Bajard JD, Eynard J, Merkiche N. “Montgomery reduction within the context of residue number system arithmetic.” *Journal of Cryptographic Engineering.* 2018;8(3):189–200. Available at: http://doi.org/10.1007/s13389-017-0154-9.

[7] Montgomery PL. “Modular multiplication without trial division.” *Mathematics of Computation.* 1985; 44(170):519–521. Available at: http://doi.org/10.1090/S0025-5718-1985-0777282-X.

[8] Sawchuk AA, Strand TC. “Digital optical computing.” *Proceedings of the IEEE.* 1984;72(7). doi: 10.1109/PROC.1984.12937.

[9] Feitelson DG. *Optical Computing: A Survey for Computer Scientists.* Cambridge (MA): MIT Press; 1988. ISBN: 978-0-262-56062-7.

[10] Candes E. MATH 262/CME 372: Applied Fourier Analysis and Elements of Modern Signal Processing. Lecture 17. 2021 Mar 11. Available at: https://statweb.stanford.edu/~candes/teaching/math262/Lectures/Lecture17.pdf.

[11] Jiale J, Minhao Z, Gupta P, Shengqi Y, Rubin SM, Garret G, Jin H. “A CAD tool for design and analysis of CNFET circuits.” In: *Proceedings of 2010 IEEE International Conference of Electron Devices and Solid-State Circuits (EDSSC);* 2010 Dec 15–17: pp. 1–4. Available at: http://doi.org/10.1109/EDSSC.2010.5713735.

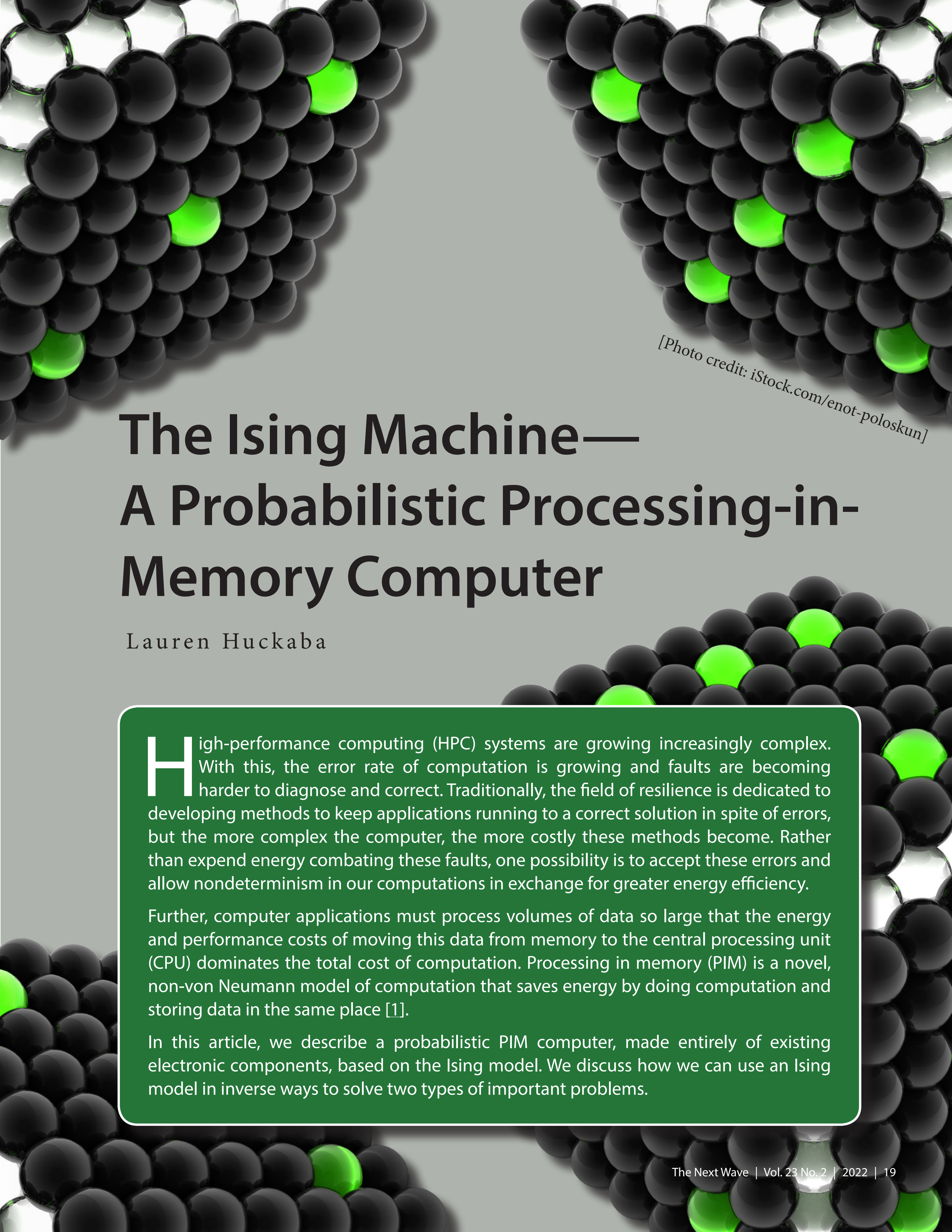
[12] Ma J, Masoodian S, Starkey D, Fossum E. “Photon-number-resolving megapixel image sensor at room temperature without avalanche gain.” *Optica.* 2017;4(12). doi: 10.1364/OPTICA.4.001474.

[13] Pohokar SP, Sisal RS, Gaikwad KM, Patil MM, Borse R. “Design and implementation of 16x16 multiplier using Vedic mathematics.” In: *International Conference on Industrial Instrumentation and Control (ICIC);* 2015 May 28–30: pp. 1174–1177. doi: 10.5120/15802-4641.

[14] Khorasaninejad M, Chen WT, Devlin RC , Oh J, Zhu AY, Capasso F. “Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging.” *Science.* 2016;352:1190–1194. Available at: http://doi.org/10.1126/science.aaf6644.

[15] Yao Y, Shankar R, Kats MA, Song Y, Kong J, Loncar M, Capasso F. “Electrically tunable metasurface perfect absorbers for utrathin mid-infrared optical modulators.” *Nano Letters.* 2014;14(11):6526–6532. doi: 10.1021/nl503104n.

Return to CONTENTS



[Photo credit: iStock.com/enot-poloskun]

The Ising Machine—A Probabilistic Processing-in-Memory Computer

Lauren Huckaba

High-performance computing (HPC) systems are growing increasingly complex. With this, the error rate of computation is growing and faults are becoming harder to diagnose and correct. Traditionally, the field of resilience is dedicated to developing methods to keep applications running to a correct solution in spite of errors, but the more complex the computer, the more costly these methods become. Rather than expend energy combating these faults, one possibility is to accept these errors and allow nondeterminism in our computations in exchange for greater energy efficiency.

Further, computer applications must process volumes of data so large that the energy and performance costs of moving this data from memory to the central processing unit (CPU) dominates the total cost of computation. Processing in memory (PIM) is a novel, non-von Neumann model of computation that saves energy by doing computation and storing data in the same place [1].

In this article, we describe a probabilistic PIM computer, made entirely of existing electronic components, based on the Ising model. We discuss how we can use an Ising model in inverse ways to solve two types of important problems.

The Ising model and the Ising problem

An Ising model is a mathematical model originally formulated to describe ferromagnetism in statistical mechanics. It consists of a lattice of spins in one of two states (see figure 1).

There is a measure of the surrounding magnetic field corresponding to each spin and a measure that denotes the interaction between each pair of spins. We call these measurements "weights." We can write down an expression for the total energy of the system in terms of the spin states and weights. The expression for total energy is known as the Hamiltonian (see figure 2).

In the physical world, an Ising system has the important property that once configured with a set of weights, the spins try to settle to a configuration that yields the lowest total energy. It achieves this state with some probability and this is where the nondeterminism comes in.

The goal of the Ising Problem is to fix a set of weights and find the configuration of spins that minimizes the Hamiltonian. The Ising Problem is a combinatorial optimization problem in which the set of local minima grows exponentially as a function of the number of spins, making it NP-hard. Many other combinatorial optimization problems, such as the traveling salesman problem and the MAX-CUT problem, can be mapped to the Ising model. Hence, the ability to efficiently solve the Ising Problem can potentially lead to solutions to a large class of other combinatorial optimization problems [2].

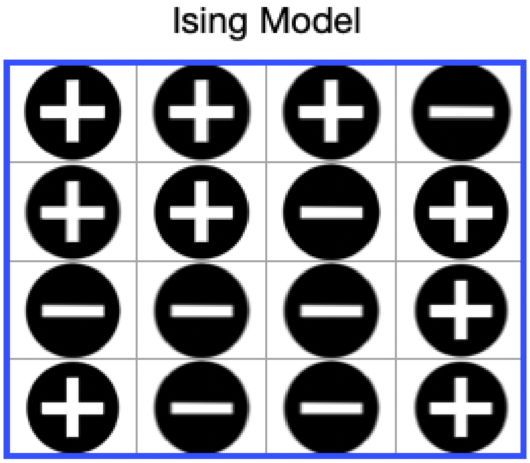


FIGURE 1. An Ising model is a lattice of spins, some positive, or taking the value 1, and some negative, or taking the value -1.

$$H(s) = \sum_{i=1}^{n} h_i s_i + \sum_{i,j=1}^{n} J_{ij} s_i s_j$$

FIGURE 2. In the equation above, $H(s)$ represents the Hamiltonian or total energy of the Ising system when in configuration s. Here, $s = (s_1,\ldots,s_n)$ is the configuration of spins, h_i is the measure of the magnetic field surrounding spin s_i, and J_{ij} is the measure of the interaction between s_i and s_j. We refer to the h_i's and J_{ij}'s as "weights" in this article.

Using an Ising model to perform arithmetic—the inverse Ising problem

The Ising problem above consists of fixing weights and determining the appropriate configuration of spins. Alternately, we can solve the inverse Ising problem by fixing a configuration of spins and finding the weights that minimize the Hamiltonian. In doing this, we can use an Ising model to do arithmetic.

We can fix a set of spins that corresponds to a correct arithmetic equation. Then we solve an optimization problem where we determine the weights that minimize the Hamiltonian. We also add constraint inequalities to our optimization problem to ensure that configurations corresponding to incorrect answers do not give a lower total energy. Even for small problems, say 3-bit multiplication, the number of constraint inequalities is quite large. For this reason, we do not give a concrete example in this article.

The weights found by solving the optimization problem can be used to tune an Ising machine or simulator built to solve the Ising problem. As described in the previous section, the machine will then try to settle to a configuration that gives the lowest total energy. In this case, that configuration is the one that corresponds to the correct answer to our multiplication problem. Because of the way we set our constraints above, getting a correct answer is more likely than getting an incorrect answer. In the following subsection, we model this process mathematically.

At first glance, it may seem as though the inverse Ising problem is easier to solve than the Ising problem

Return to CONTENTS

since Hamiltonian is clearly quadratic in the spin variables, but linear in the h's and J's. However, after setting up even a small inverse Ising problem, it becomes clear that the number of constraint equations grows exponentially in the number of spins. As such, it quickly becomes difficult to multiply numbers of more than a just a few bits in this way, and alas, the inverse Ising problem is NP-hard as well.

The Resilience and Probabilistic Computing team at the Laboratory for Physical Sciences (LPS) has spent the last few years working on the inverse Ising problem. We have a solution technique that involves an internally developed two-stage algorithm that first searches for a set of feasible parameters and then solves a system of constraints derived from the feasible parameters. Both stages have exponential complexity, but our team improved the solution time of a 3-bit multiplier from 120 days to under 10 minutes for a system of 32,000 constraints by reducing the problem to polynomial complexity. We also successfully solved the system of 267,000,000 constraints for the 4-bit multiplier. The total solve time was 27 days and used 5.5 terabytes of shared memory.

A mathematics illustration

In the absence of an actual Ising machine, we can compute a probability. The probability that the system settles to a certain configuration for a given set of weights is called the Boltzmann probability. It depends on the total energy of the system when in this configuration, as well as the noise present in the system (see figure 3).

$$\mathbf{Prob}_\beta(s) = \frac{e^{-\beta H(s)}}{\sum_\sigma e^{-\beta H(\sigma)}}$$

FIGURE 3. The equation above represents the probability that an Ising system settles to a configuration s. β represents the noise present. More specifically, $\beta = 1/(k_{B}.T)$, where T is the temperature of the system in kelvin and k_B is the Boltzmann constant. For the purpose of this article, we can think of β as simply the "noise term."

Figure 4 (on the following page) shows probabilities of solutions to 2-bit multiplication problems. After solving the inverse Ising problem, we calculated the Boltzmann probability for each configuration, both those corresponding to incorrect solutions and the configuration corresponding to the correct solution, and added some noise into our computation. This illustrates the results of simulating the nondeterminism present in an actual Ising machine. Properly setting our constraints described above gives us control over our nondeterministic computation so that there is hope of obtaining a correct answer. In this model, there is no need for error correction, which allows for a more energy-efficient computation than in a traditional digital machine.

Hardware

While the Ising model dates back to the 1920s, it was re-popularized much later by D-Wave Systems in an attempt to simulate quantum mechanical phenomena to speed up computation, including computation to solve the aforementioned combinatorial optimization problems. Recently, alternative classical methods to solve the Ising problem have emerged using optoelectronic parametric oscillators, memristor cross-bar arrays, electronic oscillators, and GPU-based algorithms [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12].

An analysis of an optoelectronic coupled oscillator system revealed the potential for a significant speedup over digital computing algorithms when the number of oscillators (nodes) is large enough [13]. Scaling up the optoelectronic oscillator Ising machine, however, remains challenging due in part to its high complexity and costly setup [5, 6, 13].

However, an all-electronic oscillator concept initially proposed by Wang and Roychowdhury introduces the idea of creating a similar system using readily available electronic components interconnected in a parallel fashion and is particularly well suited for chip-scale integration and scaling using present day technologies [2, 9].

Sync Computing and MIT Lincoln Laboratory (MIT-LL) built on this initial work and demonstrated a 4-node, fully-connected, differential LC (inductor-capacitor) oscillator-based analog circuit with standard electronic components which accurately maps to the Ising model. To the best of our knowledge, this is the first demonstration of an all-electronic oscillator-based Ising machine with multi-bit weights [2]. In [2], Chou et. al detail a statistical analysis that provides insight into the viability of these systems as computing platforms when scaled to larger node

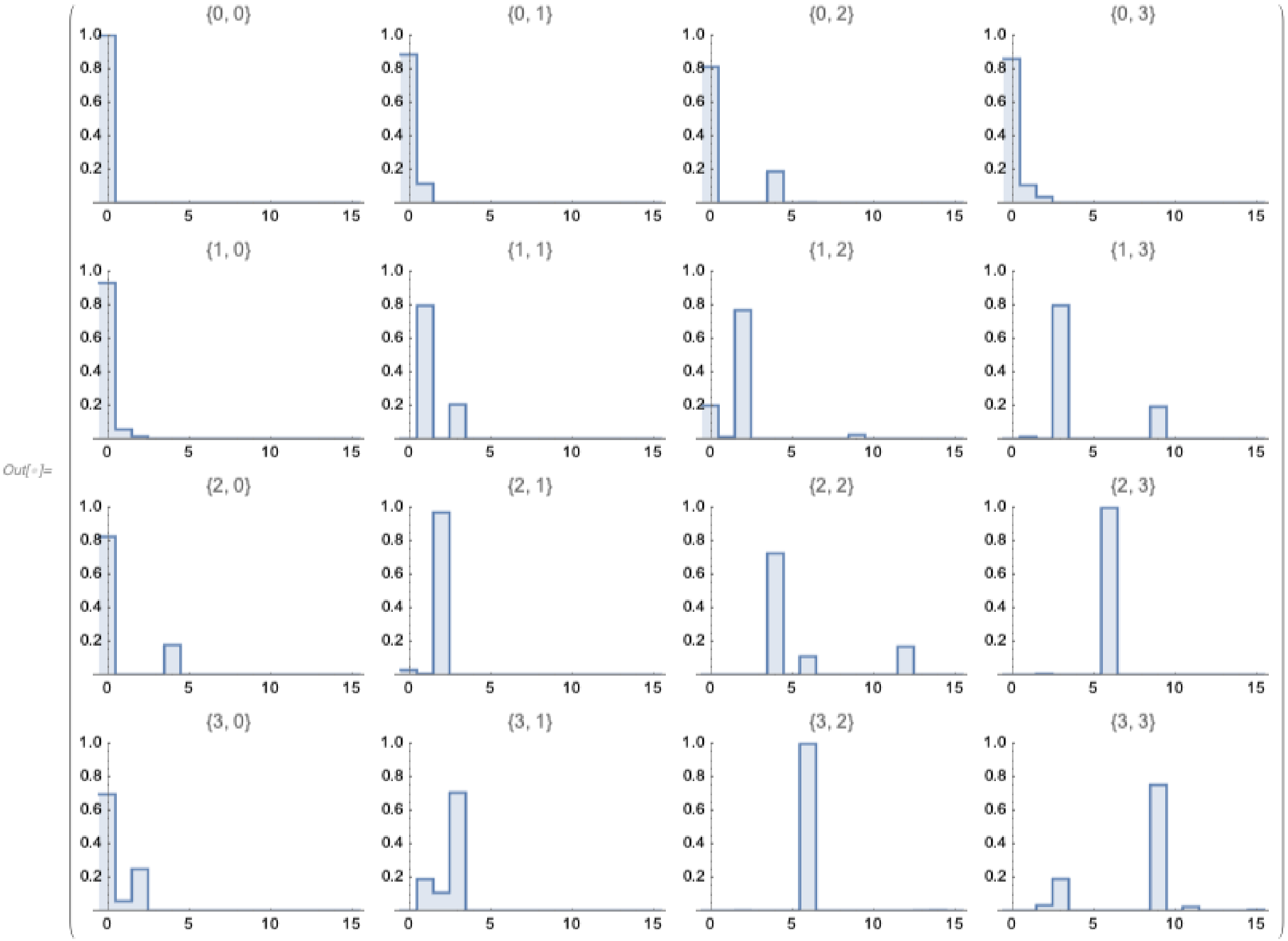


FIGURE 4. This illustrates the nondeterminism inherent in an Ising machine. The horizontal axis is possible answers and the vertical axis is the probability of a getting a certain answer. These are plots of the probabilities of getting different solutions to 2-bit multiplication problems. For example, the upper right corner represents 0 x 3 as denoted by the heading {0, 3}. We see here that we get the answer 0 most of the time, but that the answer is wrong sometimes. If we were to increase the noise term, *β*, for this calculation, we would get a wrong answer even more often.

counts. Figure 5 shows a circuit diagram of the LC oscillator circuit that employs a differential injection-locked frequency divider, the oscillators arranged in a cross-bar array, and the full breadboard system [2, 15]. Currently, Sync Computing is building a 16-node system. Figure 6 shows a photo of the printed circuit board.

Simulation—coupled oscillator system

A well-known benchmark optimization problem is the MAX-CUT problem from graph theory. Following an example in [2], we discuss a small MAX-CUT problem below using a simulation by MIT-LL of the coupled oscillator system.

Given an undirected graph, the MAX-CUT problem consists of finding a partition of that graph into two sets so that the number of edges between the two sets is as large as it can be. It has been shown previously that these graphs can be represented by a network of coupled nonlinear oscillators whose phase dynamics are described by the Kuramoto model, and that this model maps directly to the Ising Hamiltonian if the phases of these oscillators take values of either 0 or 180 degrees. As such, the Kuramoto model is the basis for the MIT-LL simulation [2, 9, 14].

An example of a 4-node (4-spin) system is shown in figure 7a (on page 24). This can also be thought of as a graph with four vertices such that every vertex is

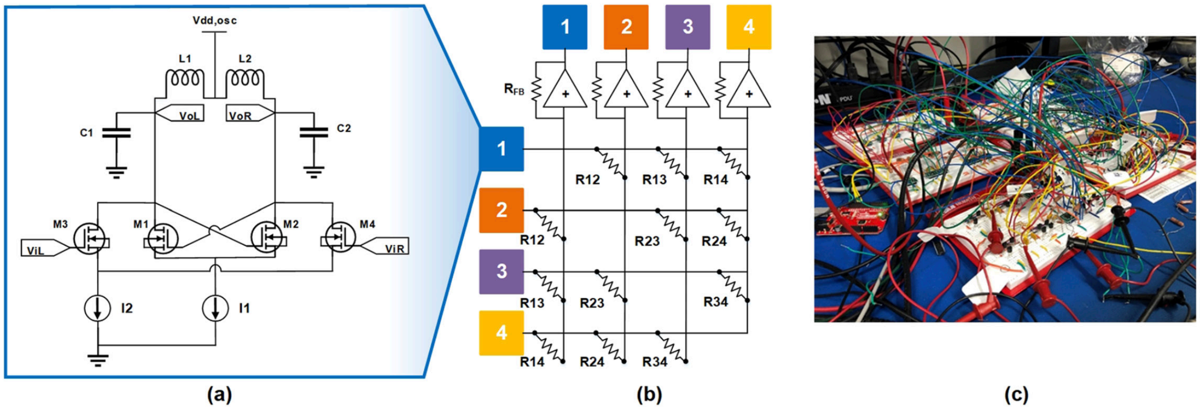


FIGURE 5. (a) This circuit diagram depicts the LC (inductor-capacitor) oscillator circuit. (b) In this diagram, the oscillators are arranged in a cross-bar array. (c) This photo shows the full breadboard system [2]. [Photo credit: MIT Lincoln Laboratory.]

connected to every other vertex. If we let J = 1 for all connections and h = 0, then the Ising problem, in this case, has six solutions shown in figure 7b. One of these solutions is shown in terms of phase in figure 7c. Here, the four spins were intentionally configured to an incorrect solution state and they settled at one of the six correct solution states as expected. Additionally, the system settles to the ground state within three oscillation cycles in this example. Figure 7d shows the results of running the simulation 1,000 times with random initial configurations. We see that the system settles to a correct solution state fairly uniformly [2].

The simulator and the inverse Ising problem

We (i.e., the LPS Resilience and Probabilistic Computing team) used the MIT-LL simulator to validate the results we obtained from solving the inverse Ising problem. We obtained weights from solving the inverse Ising problem, and we used those weights to tune the simulator. The results of using the simulator look similar to figure 4 where we computed Boltzmann probabilities and plotted the results. We found that using the simulator to validate the weights

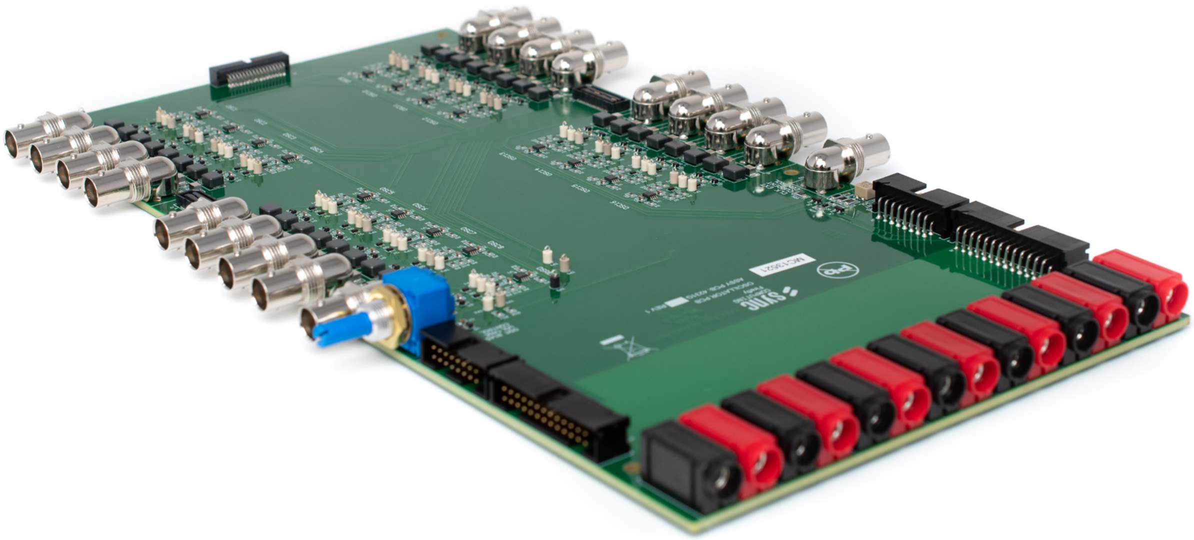


FIGURE 6. Photo of the printed circuit board of the 16-node system. [Photo credit: Sync Computing.]

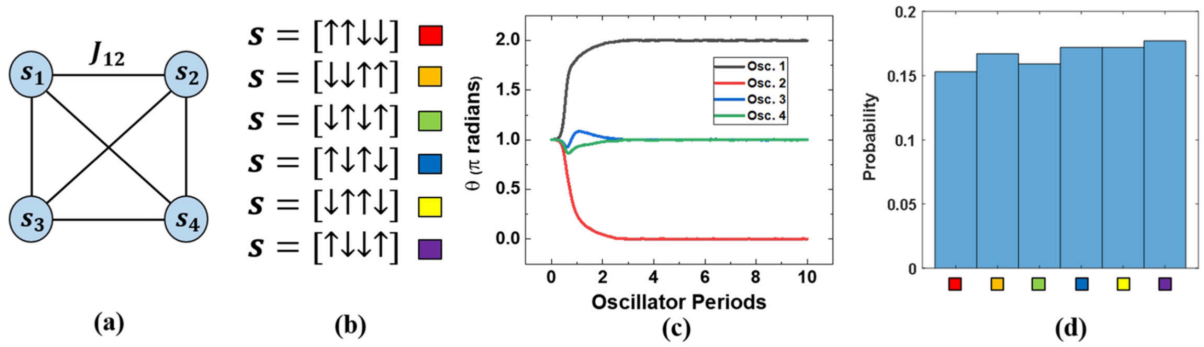


FIGURE 7. (a) This diagram depicts a fully connected 4-node system, for which (b) is the solution set where an up arrow is a positive spin and a down arrow is a negative spin. The graph in (c) shows the first solution in terms of phases, and (d) is a histogram of the results of running the simulation 1,000 times [2]. [Photo credit: MIT Lincoln Laboratory.]

was a step up from computing the Boltzmann probabilities. Our next step is to validate our results using hardware in place of the simulation.

Looking ahead

While quantum computers, like D-Wave, have the potential to solve these NP-hard combinatorial optimization problems, scaling up the number of quantum bits in these systems remains a great challenge. On the other hand, it is possible to build a probabilistic computer out of standard electronic components, as demonstrated by Sync Computing [2]. This allows for faster and more cost-effective scaling. It is reasonable to believe that this all-electronic Ising machine is scalable from four nodes to hundreds of nodes within just a few years. A limiting factor of such a machine is indeed physical space. For the more we scale up, the more oscillator circuits we must add. This does suggest that in order to build an Ising machine for practical use, we will want to explore additional technologies. However, an Ising machine with hundreds of nodes is enough to validate nontrivial results we obtain mathematically that are too large to validate in simulation. While thousands of nodes are necessary for practical use, this system is a step in the right direction and shows promise for a future that includes probabilistic computers.

References

[1] Ghose S, Boroumand A, Kim JS, Gomez-Luna J, Mutlu O. "Processing-in-memory: A workload-driven perspective." *IBM Journal of Research and Development.* 2019;63(6). Available at: https://doi.org/10.1147/JRD.2019.2934048.

[2] Chou J, Bramhavar S, Ghosh S, Herzog W. "Analog coupled oscillator based weighted Ising machine." *Scientific Reports.* 2019;9(14786). Available at: https://doi.org/10.1038/s41598-019-49699-5.

[3] Guo X, Merrikh-Bayat F, Gao L, Hoskins BD, Alibart F, Linares-Barranco B, Theogarajan L, Teuscher C, Strukov DB. "Modeling and experimental demonstration of a Hopfield network analog-to-digital converter with hybrid CMOS/memristor circuits." *Frontiers in Neuroscience.* 2015;9. Available at: https://doi.org/10.3389/fnins.2015.00488.

[4] Inagaki T, Haribara Y, Igarashi K, Sonobe T, Tamate S, Honjo T, Marandi A, Mcmahon PL, Umeki T, Enbutsu K, Tadanaga O, Takenouchi H, Aihara K, Kawarabayashi K, Inoue K, Utsunomiya S, Takesue H. "A coherent Ising machine for 2,000-node optimization problems." *Science.* 2016;354(6312):603–606. Available at: https://doi.org/10.1126/science.aah4243.

[5] Mcmahon PL, Marandi A, Haribara Y, Hamerly R, Langrock C, Tamate S, Inagaki T, Takesue H, Utsunomiya S, Yamamoto Y. "A fully programmable 100-spin coherent Ising machine with all-to-all connections." *Science.* 2016;354(6312):614–617. Available at: https://doi.org/10.1126/science.aah5178.

Return to CONTENTS

[6] Hamerly R, Inagaki T, Mcmahon PL, Venturelli D, Marandi A, Onodera T, Ng E, Langrock C, Inaba K, Honjo T, Enbutsu K, Umeki T, Kasahara R, Utsunomiya S, Kako S, Kawarabyashi K, Byer RL, Fejer MM, Mabuchi H, Englund D, Rieffel E, Takesue H, Yamamoto Y. "Experimental investigation of performance differences between coherent Ising machines and a quantum annealer." *Science Advances.* 2019;5(5). Available at: https://doi.org/10.1126/sciadv.aau0823.

[7] Shin JH, Jeong YJ, Zidan MA, Wang Q, Lu WD. "Hardware acceleration of simulated annealing of spin glass by RRAM crossbar array." In: *2018 IEEE International Electron Devices Meeting (IEDM);* 2018 Dec 1–5; San Francisco, CA. Available at: https://doi.org/10.1109/IEDM.2018.8614698.

[8] Cai F, Kumar S, Vaerenbergh TV, Liu R, Li C, Yu S, Xia Q, Yang JJ, Beausoleil R, Lu W, Strachan JP. "Harnessing intrinsic noise in memristor Hopfield neural networks for combinatorial optimization." 2019. Cornell University Library. Available at: https://arxiv.org/abs/1903.11194.

[9] Wang T, Roychowdhury J. "OIM: Oscillator-based Ising machines for solving combinatorial optimisation problems." 2019. Cornell University Library. Available at: https://arxiv.org/abs/1903.07163.

[10] Parihar A, Shukla N, Jerry M, Datta S, Raychowdhury A. "Computing with dynamical systems based on insulator-metal-transition oscillators." *Nanophotonics.* 2017;6:601–611. Available at: https://doi.org/10.1515/nanoph-2016-0144.

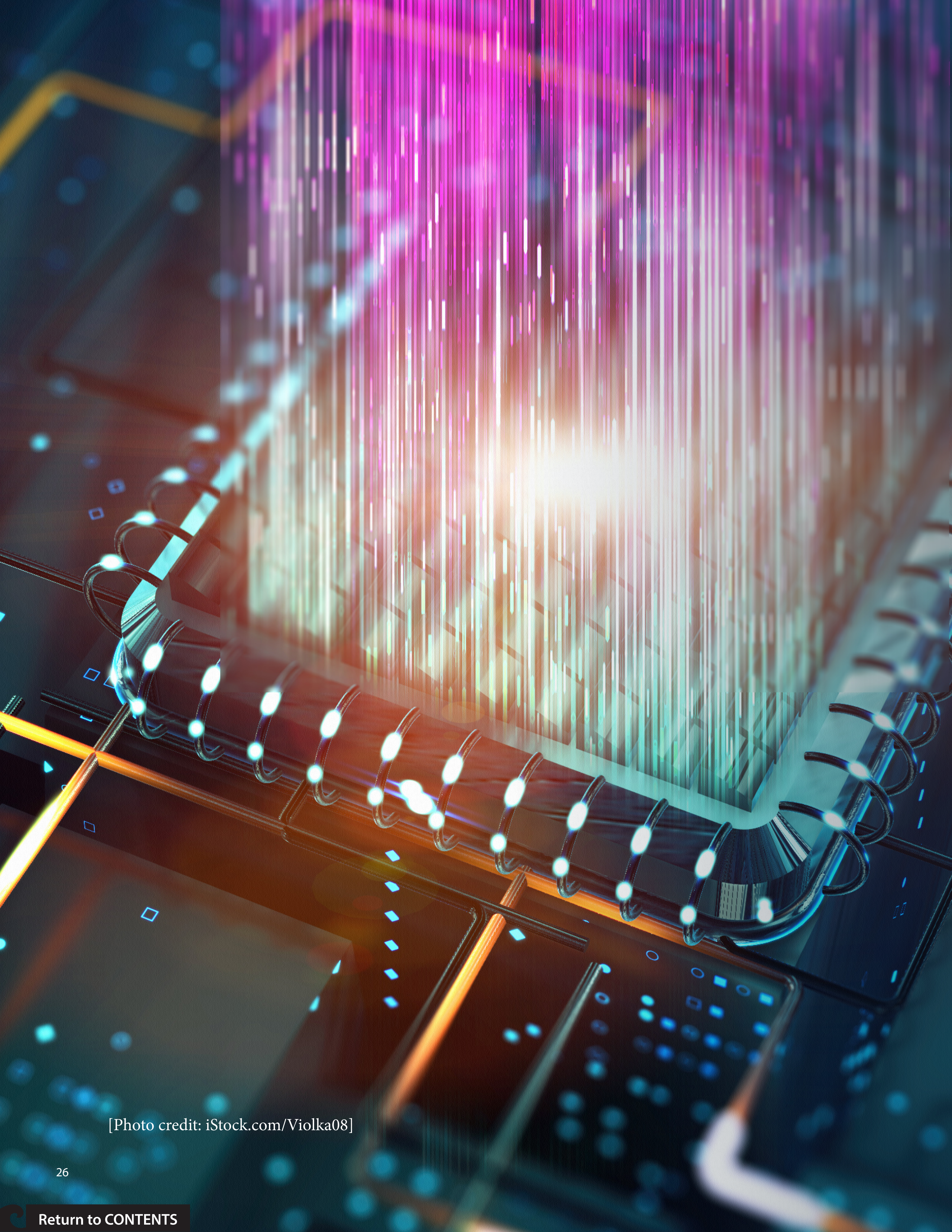
[11] King AD, Bernoudy W, King J, Berkley AJ, Lanting T. "Emulating the coherent Ising machine with a mean-field algorithm." 2018. Cornell University Library. Available at: https://arxiv.org/abs/1806.08422.

[12] Tiunov ES, Ulanov AE, Lvovsky AI. "Annealing by simulating the coherent Ising machine." *Optics Express.* 2019;27(7):10288–10295. Available at: https://doi.org/10.1364/OE.27.010288.

[13] Haribara Y, Utsunomiya S, Yamamoto Y. "A coherent Ising machine for MAX-CUT problems: Performance evaluation against semidefinite programming and simulated annealing." In: Yamamoto Y, Semba K, editors. *Principles and Methods of Quantum Information Technologies.* Japan: Springer; 2016. Pp. 251–262. Available at: https://doi.org/10.1007/978-4-431-55756-2_12.

[14] Acebrón JA, Bonilla LL, PérezVicente CJ, Ritort F, Spigler R. "The Kuramoto model: A simple paradigm for synchronization phenomena." *Reviews of Modern Physics.* 2005;77(1):137–185. Available at: https://doi.org/10.1103/RevModPhys.77.137.

[15] Rategh HR, Lee TH. "Superharmonic injection-locked frequency dividers." *IEEE Journal of Solid-State Circuits.* 1999;34(6):813–821. Available at: https://doi.org/10.1109/4.766815.



[Photo credit: iStock.com/Violka08]

Return to CONTENTS

Cacheless Computer Architectures: 3D Integration of Optical Interconnects and Novel Memory

Eric Cheng, S. J. Ben Yoo

As we near the end of traditional complementary metal-oxide semiconductor (CMOS) scaling, and systems are further constrained by various “walls” (e.g., power, memory, resilience), we are no longer seeing historical year-over-year exponential performance improvements. New technologies, architectures, and integration techniques are required to ensure future computing systems (from embedded to high-performance) continue to deliver the capabilities required by the scientific, business, and national security communities. Memory remains a significant limiter when it comes to the energy efficiency, scalability, performance, and productivity of computing systems. In particular, traditional memory hierarchies (i.e., the cache hierarchy) have been optimized over decades to map well to workloads with frequent, regular access and are not well-suited for the sparse, random-access workloads that are emerging and dominating key applications in the high-performance data analytics (HPDA) and graph analytics space.

Re-architecting the memory subsystem to better account for these sparse, irregular workloads and, in particular, flattening traditional memory hierarchies, can drastically reduce energy consumption, reduce memory-access latency, improve memory-access predictability, increase memory bandwidth, and enhance programmer productivity. This so-called “cacheless computer architecture” is enabled by advances in massively parallel silicon photonic wavelength-division multiplexed (WDM) interconnects, novel memory devices and architectures, and innovative electronic/photonic three-dimensional (3D) integration capabilities. This new architecture features 3D integration of optically-interconnected low-latency memory (LLM) to provide four times the memory capacity and two times the improvements in latency and energy efficiency as compared to traditional dynamic random-access memory (DRAM) while demonstrating a five times reduction in average memory access time and 60 percent reduction in access latency variability across a variety of application workloads.

Vision

Emerging applications in problem domains such as HPDA and graph analytics are placing ever-increasing demands on the computing systems available to the computing community and, in particular, stretching the capabilities of today's high-performance computing (HPC) systems. The global semiconductor industry is constantly developing new technologies to help provide the massive performance improvements and new capabilities required to meet these demands. However, effective hardware/software codesign along with intelligent integration of these available technologies into new system architectures are required to fully realize the potential for future compute systems. By leveraging advanced technologies such as massively parallel silicon photonic WDM interconnects, novel memory devices, and innovative electronic/photonic 3D integration capabilities, we can create a new "cacheless computer architecture" that is better optimized for new problem domains and provides better scalability for future HPC systems. This cacheless architecture implements a flattened memory architecture that can drastically reduce energy consumption, reduce memory access latency, improve memory access predictability, increase memory bandwidth, and enhance programmer productivity (see figure 1).

HPDA and graph analytics workloads are often dominated by memory operations (as opposed to more compute-dominated workloads) and exhibit random or irregular access to data that is sparsely distributed across a large memory footprint. With the introduction of silicon photonics (SiPh), it is possible to begin removing levels of the cache hierarchy (i.e., flattening the hierarchy) to reduce energy consumption and access latency while increasing memory bandwidth. This new architecture also improves memory access predictability, which is key to enhancing programmer productivity, simplifying application analysis, and supporting better algorithm design.

Technologies

This cacheless computer architecture is enabled by the convergence of several technologies (incubated over the past decade) that are effectively integrated into an overall system architecture. Some of these technologies have been fully demonstrated (i.e., in the fab), while others are nearing maturity (i.e., in the lab).

In the fab

As we enter a future of heterogeneous compute [i.e., architectures with a diverse mix of compute elements ranging from central processing units (CPUs) to graphics processing units (GPUs) to fixed-function accelerators that are coupled with diverse memory and storage solutions] and resource disaggregation (i.e., logical or physical clustering/separation of distinct resources that are connected via a network), the need for highly scalable interconnection networks that provide high bandwidth, low latency, all-to-all communication (i.e., every device on the network can directly communicate with every other device simultaneously) is crucial. Massively parallel SiPh using WDM have the ability to provide this scalable and high-bandwidth

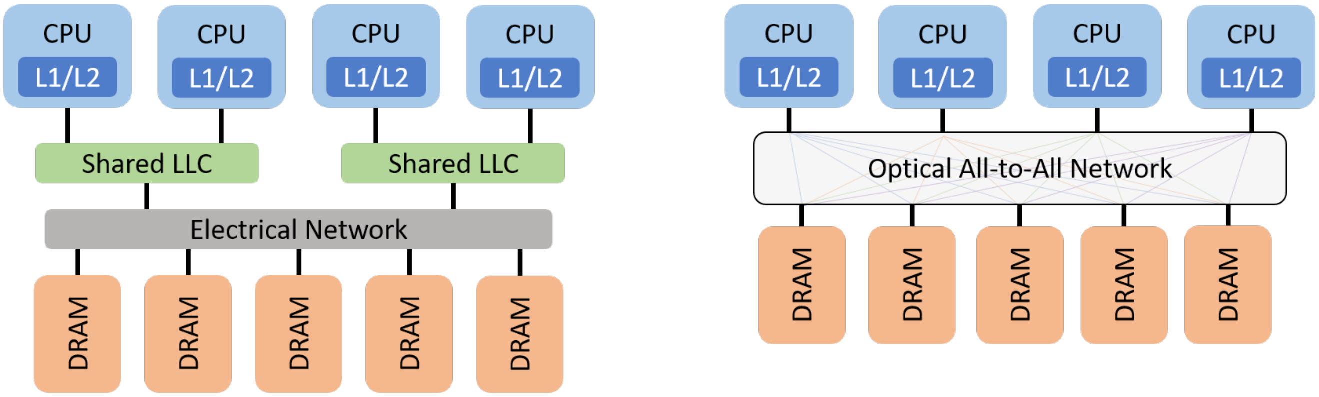


FIGURE 1. Comparison of a traditional architecture consisting of a shared last-level cache (LLC) electrically connected to DRAM (left) versus a (last-level) cacheless architecture consisting of compute elements optically connected to DRAM (right).

networking capability for computing systems. In particular, by leveraging arrayed waveguide grating routers (AWGRs) [1], we can achieve very compact interconnect fabrics that provide all-to-all connectivity between devices.

This optical interconnect fabric is created as follows: a laser source (typically an off-chip, external laser) is used to generate the various wavelengths required, which can be generated by a single frequency comb or by multiple individual lasers. The use of WDM allows for these multiple wavelengths to traverse over a single waveguide. Individual frequency-tuned modulators are used to encode data on each wavelength, and the corresponding frequency-tuned photodetector decodes the data from the corresponding wavelength. An AWGR uses this general concept to connect multiple input nodes in an all-to-all manner to multiple output nodes (i.e., every input is directly connected to every output).

Importantly, AWGRs do not just exist as a conceptual or theoretical design. Compact 8 x 8 silicon nitride (SiN) AWGRs have been fabricated and demonstrated in a compact 1 square millimeter (mm^2) footprint [2] (see figure 2). Physical demonstrations of scaled networks implementing 512 x 512 SiPh AWGRs have also been fabricated [3]. Such fabricated devices demonstrate the feasibility of realizing actual systems using SiPh as well as showcase the ability to provide much better scalability, single-hop distance-independent energy-efficient communication, and higher bandwidth communication as compared to using traditional electrical-only equivalents. As the number of nodes in a system grows, the hardware cost for implementing SiPh grows linearly as opposed to quadratically for the electrical equivalents, thus providing better scalability. Furthermore, tight integration of such SiPh technologies (see later sections) hold the potential for reducing memory access energy from order 2–4 picojoules (pJ) per bit down to order 1 pJ per bit and increasing aggregate memory bandwidth from 1 gigabyte (GB) per second to 1 terabyte (TB) per second as compared to current electrical-only technologies.

In the lab

Innovation in the interconnect fabric is not the only advanced technology required to realize a cacheless architecture. While the following technologies are arguably less mature, they have nonetheless demonstrated significant benefits and have been validated through detailed simulation and small-scale benchtop experiments.

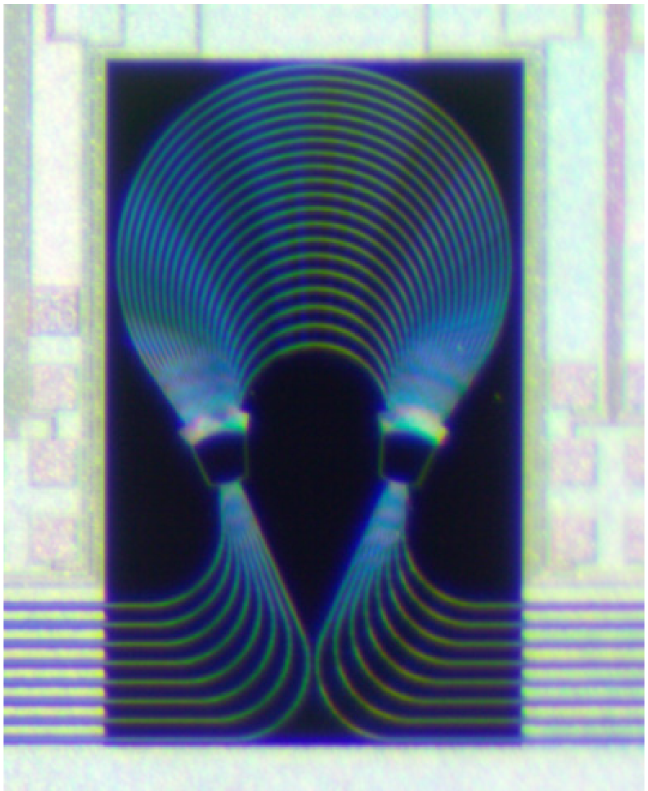


FIGURE 2. Fabricated 8 x 8 silicon nitride (SiN) arrayed waveguide grating router (AWGR).

Innovative electronic/photonic 3D integration techniques are necessary to provide the capabilities necessary to effectively couple SiPh technologies with conventional electronic devices. In particular, 3D stacking (i.e., stacking multiple silicon wafers/dies together to form a single integrated circuit) can greatly enhance the capabilities of a cacheless architecture by providing tight integration of compute and memory dies or by providing much greater memory capacity per packaged part. Traditional through-silicon vias (TSVs) are used to provide the connectivity across stacked dies but have bandwidth and scalability limitations due in part to TSV density constraints. An alternative is to use through-silicon optical vias (TSOVs), which are enabled by the use of ultra-compact vertical U-shaped couplers [4, 5]. TSOVs can provide additional bandwidth, reduce the via density, maintain low-energy communication across stacked

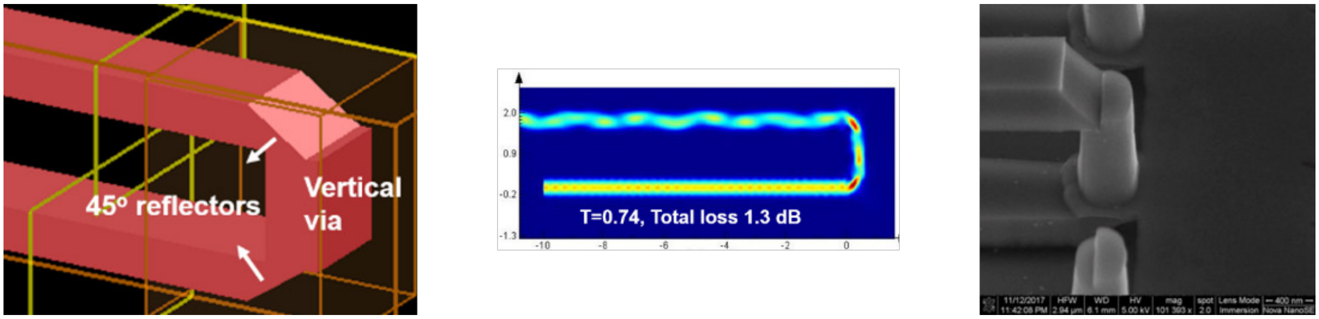


FIGURE 3. Projected view of an ultra-compact vertical U-shaped coupler (left), simulated optical propagation through the coupler (center), and scanning electron microscope (SEM) image of a fabricated coupler (right).

dies, and ensure scalability in future process nodes. Small-scale component fabrication of the U-shaped couplers, which are key to implementing TSOVs, have demonstrated the ability to achieve 1–2 micron (μm) pitches with detailed simulation demonstrating transmission losses of 1.3 decibels (dB) [4] (see figure 3). The combination of these small-scale demonstrators and detailed simulations help validate the feasibility of this electronic/photonic 3D integration technology.

Key limiters to achieving a cacheless architecture using traditional memory devices and architectures [e.g., DRAM, high-bandwidth memory (HBM), etc.] include access energy, access latency, access granularity, memory bandwidth, and overall capacity. With the assistance of optical interconnect fabrics and innovative electronic/photonic integration capabilities described earlier, novel memory devices and architectures (such as LLM) can be constructed to help overcome these limitations. Building off the conceptual fine-grained DRAM (FG-DRAM) [6] design, an extension referred to here as LLM can provide improvement across all of these parameters.

LLM leverages dedicated optical buses to memory banks to reduce contention and improve bandwidth. The use of TSOVs helps reduce access latency, increase memory capacity (through the ability to integrate a greater number of memory dies), and minimize the number of vias required to effectively connect and provide I/O (input/output) to the memory stacks (see figure 4). In the age of disaggregation, (CPU) cores would be connected directly to shared, global memory. To reduce memory access latency in this configuration, LLM leverages a dedicated memory controller at each core connected to the memory controller blocks on the memory side. The compute-side memory controller (CMC) contains the read and write queues while the memory-side memory controller (MMC) contains the scheduling logic and command queues. The CMCs and MMCs communicate with one another via an all-to-all optical interconnect provided by the use of AWGRs. Detailed simulation has demonstrated that such an LLM design can be used to increase memory capacity by up to four times, reduce access latency and energy by two times, reduce average memory access time by five times, and reduce access latency variability by 60 percent as compared to conventional DRAM-based HBM.

Application impact

Each of the individual technology components used to realize a cacheless architecture demonstrate significant potential; however, intelligent integration and codesign is required to realize the full potential and understand the overall impact on application workloads. To more fully explore and evaluate the benefits of a cacheless computer architecture, a broad set of applications spanning traditional HPC workloads [e.g., Advanced Encryption Standard (AES), convolution, Fast Fourier Transform (FFT)] to HPDA/graph workloads [e.g., Breadth-First Search (BFS), PageRank] were evaluated.

As the motivating design point, a cacheless GPU design [i.e., a design that has removed the level 2 (L2) caches, uses conventional HBM, and leverages AWGRs] was compared to an equivalent multi-GPU system (i.e., a design with a conventional memory hierarchy, HBM, and is electrically connected) providing the same number of compute elements in both designs. Across the applications studied, a codesigned

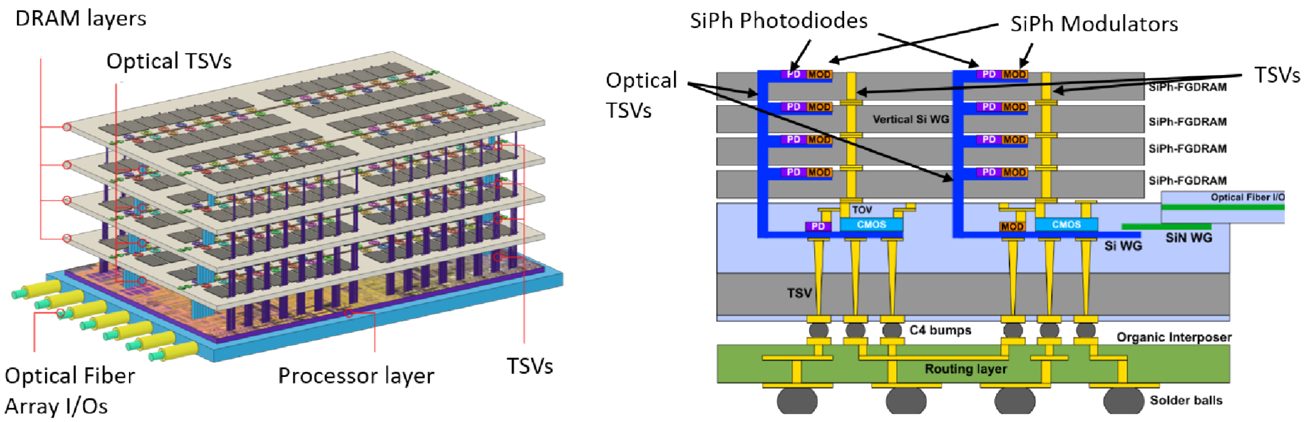


FIGURE 4. Projected view (left) and side view (right) of low-latency memory (LLM) with a SiPh base layer integrated via through-silicon optical vias (TSOVs) to the DRAM stacks.

cacheless GPU can reduce overall DRAM access latency by 10–55 percent, improve level 1 (L1) miss penalty by 2.3–5 times, and offer an overall application speedup of 1.1–1.8 times [7].

Specifically, if we focus on two applications, FFT (a traditional HPC application) and PageRank (an important graph workload), we can concretely identify the sources of improvement. The removal of L2 caches significantly improves access latency for FFT by 31 percent and for PageRank by 10 percent. L1 miss penalty improves on FFT by 3.2 times and on PageRank by 2.8 times. Taken together, this translates to roughly a 1.4 times speedup on FFT and a roughly 1.8 times speedup on PageRank for the cacheless GPU design.

Conclusion

A cacheless computer architecture holds great potential for greatly improving the capabilities of future compute systems and, in particular, for applications in the HPDA and graph analytics domains. By leveraging innovative 3D integration techniques, optical interconnect fabrics, and LLM, future systems can be better optimized for and perform more efficiently on emerging workloads that are characterized by sparse and irregular accesses to memory. Through a combination of physically fabricated test vehicles, subcomponent demonstrations, and detailed simulations, a cacheless computer architecture has been shown to not only be highly promising, but also well-within reach for system demonstrations at commercial foundries.

For a range of application workloads, an initial study of a cacheless computer architecture has demonstrated that it is possible to achieve a two times improvement in memory access latency and energy efficiency, a five times reduction in average memory access time, and a 60 percent reduction in access latency variability. Not only do these improvements have direct impact on the performance of applications, they also provide secondary benefits. A cacheless architecture, which provides more predictable application performance, affords an application programmer (or even an advanced compiler) a greater ability to easily reason about, analyze, and optimize applications for the cacheless architecture. This opens up the potential for additional performance gains. Finally, with the simpler and more predictable architecture, it also becomes easier to design algorithms that can better take advantage of the underlying hardware architecture (e.g., algorithms that more closely resemble underlying mathematical constructs as opposed to needing to worry about manually blocking, partitioning, and placing data to fit within the limits of conventional memory subsystems). This cacheless computer architecture opens up many possibilities for future computing systems.

References

[1] Grani P, Liu G, Proietti R, Yoo SJB. "Bit-parallel all-to-all and flexible AWGR-based optical interconnects." *Optical Fiber Communication Conference, OSA Technical Digest* (online). 2017. Available at: https://doi.org/10.1364/OFC.2017.M3K.4.

[2] Shang K, Pathak S, Qin C, Yoo SJB. "Low-loss compact silicon nitride arrayed waveguide gratings for photonic integrated circuits." *IEEE Photonics Journal*. 2017;9(5). doi: 10.1109/JPHOT.2017.2751003.

[3] Cheung S, Su T, Okamoto K, Yoo SJB. "Ultra-compact silicon photonic 512x512 25 GHz arrayed waveguide grating router." *IEEE Journal of Selected Topics in Quantum Electronics*. 2014;20(4). Available at: https://doi.org/10.1109/JSTQE.2013.2295879.

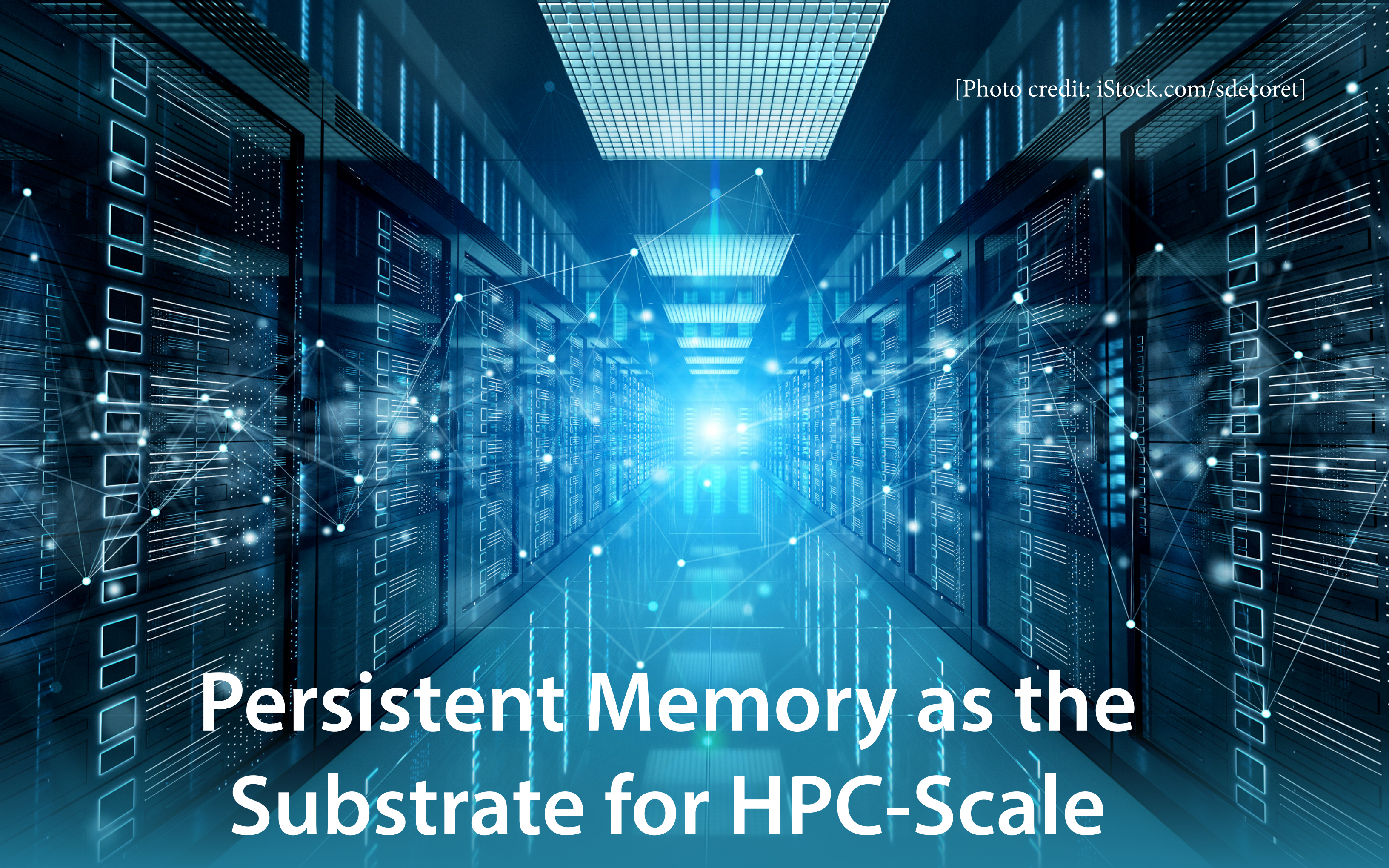
[4] Zhang Y, Ling YC, Zhang Y, Shang K, Yoo SJB. "High-density wafer-scale 3-D silicon-photonic integrated circuits." *IEEE Journal of Selected Topics in Quantum Electronics*. 2018;24(6). Available at: https://doi.org/10.1109/JSTQE.2018.2827784.

[5] Zhang Y, Samanta A, Shang K, Yoo SJB, "Scalable 3D silicon photonic electronic integrated circuits and their applications." *IEEE Journal of Selected Topics in Quantum Electronics*. 2020;26(2). Available at: https://doi.org/10.1109/JSTQE.2020.2975656.

[6] O'Connor M, Chatterjee N, Lee D, Wilson J, Agrawal A, Keckler SW, Dally WJ. "Fine-grained DRAM: Energy-efficient DRAM for extreme bandwidth systems." In: *MICRO-50 '17: Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture;* 2017 Oct: pp. 41–54. Available at: https://doi.org/10.1145/3123939.3124545.

[7] Fotouhi P, Fariborz M, Proietti R, Lowe-Power J, Akella V, Yoo SJB. "HTA: A scalable high-throughput accelerator for irregular HPC workloads." In: *ISC High Performance 2021:High Performance Computing*. Part of the *Lecture Notes in Computer Science* book series (LNCS, volume 12728). Springer International Publishing; 2021. Pp. 176–194. Available at: https://doi.org/10.1007/978-3-030-78713-4_10.

Return to CONTENTS



[Photo credit: iStock.com/sdecoret]

Persistent Memory as the Substrate for HPC-Scale Graph Analytics

Roger Pearce & Geoffrey Sanders, Lawrence Livermore National Laboratory[a]

The volume of data currently generated both by science and security applications and by the modern Internet-connected human experience has surpassed our ability to process and understand at adequate levels of fidelity. When deep historical or longitudinal analysis is required, the volume of data often requires heavy triage or filtering that can impede deep analysis. The promise of using high-performance computing (HPC) for such analysis is that a unified picture of a large, distributed data set is possible; however, tools to tackle enterprise-level data sets are still in research. Emerging within many HPC environments are high-capacity, high-bandwidth solid-state nonvolatile storage devices—including block- and byte-addressable persistent memories. Such memories, combined with HPC, are poised to revolutionize how data-intensive workloads are deployed. This article showcases many of the ongoing research efforts at Lawrence Livermore National Laboratory (LLNL) to enable data-intensive computing at unprecedented scales.

[a] This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and was supported in part by the LLNL-LDRD Program under Project No. 21-ERD-20 LLNL-JRNL-82847.

Performing exploratory data analytics is often the first step used by data scientists when faced with a new data set or analytic task, and it specifically aids in hypothesis generation and evaluation. The de facto standard among a large percentage of data scientists is Jupyter notebooks (i.e., interactive Python), in which relatively small data sets are manipulated using popular tools such as NumPy, SciPy, Pandas, or NetworkX on a desktop or laptop environment. Data scales of up to a few million data points are commonly processed in this environment, limited by the available main memory, or dynamic random access memory (DRAM), of the environment. When faced with data sets exceeding the memory capacity of a laptop or desktop, data scientists face a choice of reducing data volume through filtering and sampling or transitioning their workload to a distributed computing environment (e.g., Apache Spark [1] or Arkouda [2]). These tools have the ability to process big data workloads in fast memory, and they able to write and read intermediate representations that are larger than fast memory to file, albeit more slowly. In many data analysis settings (including HPC systems shared by multiple users, rolling back to intermediate state in a workflow, and interactive exploratory data analytics on massive data), more optimal balancing of this trade-off between speed and data volume is of high interest.

Solid-state persistent memory technologies have widespread adoption in the portable computing market (e.g., laptops, tablets, and smartphones) and are emerging on many new distributed computing systems (e.g., HPC and Cloud) in the form of node-local and rack-local persistent memory, most commonly in the form of peripheral component interconnect express (PCIe)-attached nonvolatile memory express (NVMe) solid-state drives (SSDs), and soon, byte-addressable memory-attached non-volatile dual in-line memory modules (NVDIMMs, e.g., Intel Optane).

This persistent memory bridges the gap between low-latency DRAM and high-latency spinning disk storage. As a prime example, LLNL's upcoming exascale-class supercomputer, El Capitan, will be outfitted with novel storage nodes, each consisting of 18 SSDs referred to as *rabbits* [3]. The design of El Capitan's rabbits is primarily to enable high-speed checkpointing of scientific simulations; however, the rabbits also provide unique opportunities to data scientists.

An open research challenge for the community of HPC data science tool builders is how can we leverage these emerging persistent memory technologies to enable data scientists to tackle their growing data volumes? Figure 1 illustrates our anticipated complex memory hierarchy for future HPC systems. For data-science applications, managing this complex memory hierarchy containing both ephemeral and persistent memory will be a key challenge. This fast-growing capacity of persistent memory, extending the reach of expensive and power-hungry main-memory DRAM, is ideal for staging data between data analysts' interactive operations and storing incremental state, should they choose to roll back a computation and modify a portion of their data workflow without completely restarting it. Prior experience indicates that emerging persistent memory devices are well suited for data-intensive computing tasks including graph analytics, and our broader LLNL team has developed memory allocators and run-times that allow applications to transparently operate out of persistent memory.

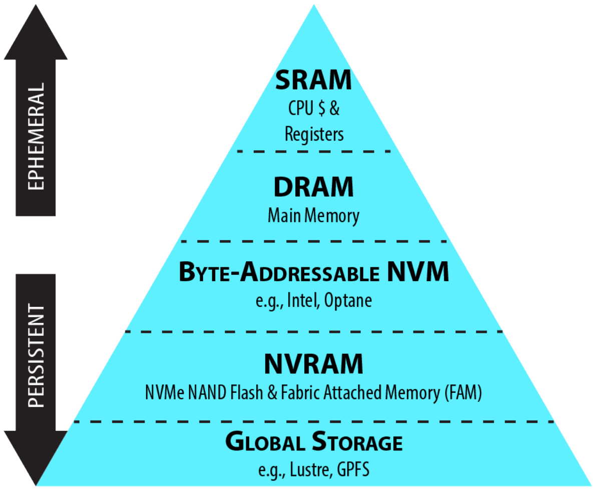


FIGURE 1. This pyramid represents the anticipated complex memory hierarchy for future HPC data science systems. Managing the complex memory hierarchy containing both ephemeral and persistent memory is a key open challenge.

An often overlooked but common theme among a variety of data analytics platforms is the need to persist data beyond the scope of a single execution. The task of ingesting data, indexing, and partitioning data in preparation of running an analytic, is often more expensive than the analytic itself. The promise of persistent data structures is that, once constructed, data structures can be re-analyzed and updated beyond the lifetime of a single execution without much slower file read/write access, and new forms of persistent memory are increasing the viability of our target use case—holistic relational data analysis, which jointly involves graph calculations and traditional data analysis with both topological data and metadata in place for enhanced exploratory data analytics.

Mmap versus traditional data serialization

Persisting data beyond the life of an application or process has traditionally used data serialization, and a plethora of middleware libraries have been developed to aid application developers in this process. A de facto standard in the HPC community for data storage is HDF5 [4], providing a portable self-described data format. When stored on traditional rotating media (e.g., a distributed file system such as Lustre), the overheads of heavyweight serialization are masked by the slow rotating media. However, on emerging nonvolatile random-access memory (NVRAM) devices with multiple orders of magnitude of lower latency than rotating media, such serialization overheads become noticeable.

Memory mapping (via the POSIX mmap() system call) is a mechanism by which the operating system (e.g., Linux/Unix) virtually maps a file's contents into the address space of a process and reads or writes pages to the file using the virtual memory page fault mechanism. Such mmap mappings can be larger than the physical main memory of the system, allowing applications to address data sets larger than physical main memory (often referred to as out-of-core or external memory). Prior research has demonstrated memory mapping as an ideal mechanism to access NVRAM for data-science applications [5]. It is our position that memory mapping persistent memory is the key to achieving interactive exploratory data analytics.

To assist software developers in the design of persistent data structures using persistent memory, Iwabuchi et al. at LLNL have developed Metall, a persistent memory allocator designed to provide developers with an application programming interface to allocate custom C++ data structures in both block storage and byte-addressable persistent memories [6, 7, 8]. Metall relies on a file-backed mmap mechanism to map a file in a file system into the virtual memory of an application. Metall's approach allows a C++ application to transparently *create, detach,* and *reattach* to persistent data structures without heavyweight serialization. Traditional serialization techniques continue to have their place for portable data archive reasons; Metall's approach is aimed at enabling lightweight manipulation of persistently stored data structures and not a replacement for portable serialization and archive.

Case study: Large-scale static graph analytics

Within the domain of graph analytics, we have utilized large NVRAM devices with memory map to scale to some of the largest graph data sets, both on single-node workstations [9] and distributed clusters [10]. The approach enables large volumes of graph data to spill out of main memory ("out of core") into large-capacity NVRAM devices attached to each compute node. Graph algorithms proceed to fetch portions of the out-of-core graph data on demand using the operating system's virtual memory paging system.

To benchmark algorithms and architectures for processing large graphs, the HPC community established the Graph500 [11] in 2010. The benchmark generates large synthetic scale-free graphs and measures the traversal time of Breadth-First Search (BFS) across the graph. The synthetic scale-free graphs are particularly challenging due to the skewed vertex degree distributions leading to load imbalances, and these challenges are representative of many real-world networks.

The effort focused research attention on many pressure points across the spectrum of computer architecture, system software, and algorithm engineering, with the goal of increasing capabilities in processing performance and data scales. In just a few short years, the Graph500 community effort led to significant advancements in graph processing capabilities.

Our team at LLNL focused on persistent memory technologies, specifically PCIe attached NAND Flash devices, as extended memory devices to enable the processing of some of the largest graphs; a summary of these results is shown in table 1 (on the following page). LLNL's largest result utilized the Sierra supercomputer that contains 1.6-terabyte NVMe SSDs per compute node; using 2,048 compute nodes and SSDs together, the result achieved the largest result to date, traversing 70-trillion edges in 17.43 minutes. Without using Sierra's node-local NVMe SSDs, eight times the compute nodes would have been required for this computation.

TABLE 1. LLNL's Graph500 Results (each supercomputer using NVRAM as extended graph storage)

Year	Machine	Compute Nodes	Graph Edges	Giga-Traversed Edges per Second (GTEPS)
2011	Kraken	1	275 Billion	0.053
2011	Leviathan	1	1 Trillion	0.053
2011	Hyperion	64	1 Trillion	0.601
2014	Bertha	1	2.2 Trillion	0.054
2014	Catalyst	100	17.6 Trillion	4.175
2018	Sierra	2,048	70.4 Trillion	67.258

Case study: Persisting & versioning of dynamic graph data structures

An important feature for exploratory data analytics is the ability to snapshot and persist versions of complex data structures. Like the requirement of lightweight detaching and reattaching to persistent data structures previously discussed, when significant compute time has been consumed building complex interconnected data structures, persisting consistent snapshots of them becomes invaluable. Examples of such use cases include preserving periodic snapshots (e.g., nightly snapshots) of live data streams or preserving the incremental steps with provenance within a complex data analytics pipeline.

To achieve such snapshot consistency, Metall employs an explicit, coarse-grained persistence policy in which persistence is guaranteed only when the heap is saved in a snapshot to the backing store. Youssef, Iwabuchi, et al. recently developed a technique to enable space-efficient snapshots that only stores the difference from previous snapshots instead of duplicating the entire persistent heap [12].

To illustrate this approach, the team dynamically constructed a graph consisting of the dynamic hyperlink structure of Wikipedia from January 2001 through July 2017. The graph contains a total of 1.8 billion temporal hyperlink edges. The persistent data structure used was a classic adjacency list implemented with Metall, and the experimental platform was a 48-core AMD EPYC machine with 256-gigabyte DRAM and 1.6-terabyte NVMe SSD for graph data structure storage.

As the Wikipedia hyperlink graph is ingested in temporal order, monthly snapshots are taken of the adjacency list data structure to mimic a data scientist preparing for a longitudinal study. Figure 2 shows the cumulative storage requirements to preserve the monthly snapshots, with and without the deduplication approach developed by Youssef [12]; in total, a 33.2 percent storage savings is achieved using deduplication. There is a performance trade-off with deduplication based on the granularity or block size, and figure 3 presents the trade-off between storage improvement and execution performance as a factor of Metall's persistent store block size.

This approach is not limited to graph analytics, and the goal of Metall's persistent snapshots is to enable data scientists to preserve the state of evolving data, preserve intermediate steps of a complex analytics pipeline, utilize intermediate steps within collaborators' workflows, or to simply allow an "undo" when performing exploratory data analytics on large data sets.

Looking forward: HPC-scale property graphs

As we anticipate the future needs of data scientists, an emerging trend is the need to support HPC-scale topological relationships with rich unstructured metadata at vertices and edges, commonly referred to as property graphs. Often, these relational data sets are stored as several database tables whose fields relate data in table A to table B in explicit and implicit ways. Network scientists wishing to understand important higher-order behavior within these data sets make modeling decisions to represent the collection of tables as a large graph. Figure 4 (on page 38) illustrates a prototypical example of a cybersecurity property graph based on a graph representation inspired by the Defense Advanced Research Projects Agency (DARPA) Operationally Transparent Cyber (OpTC) data set [13]. In this example, one table represents

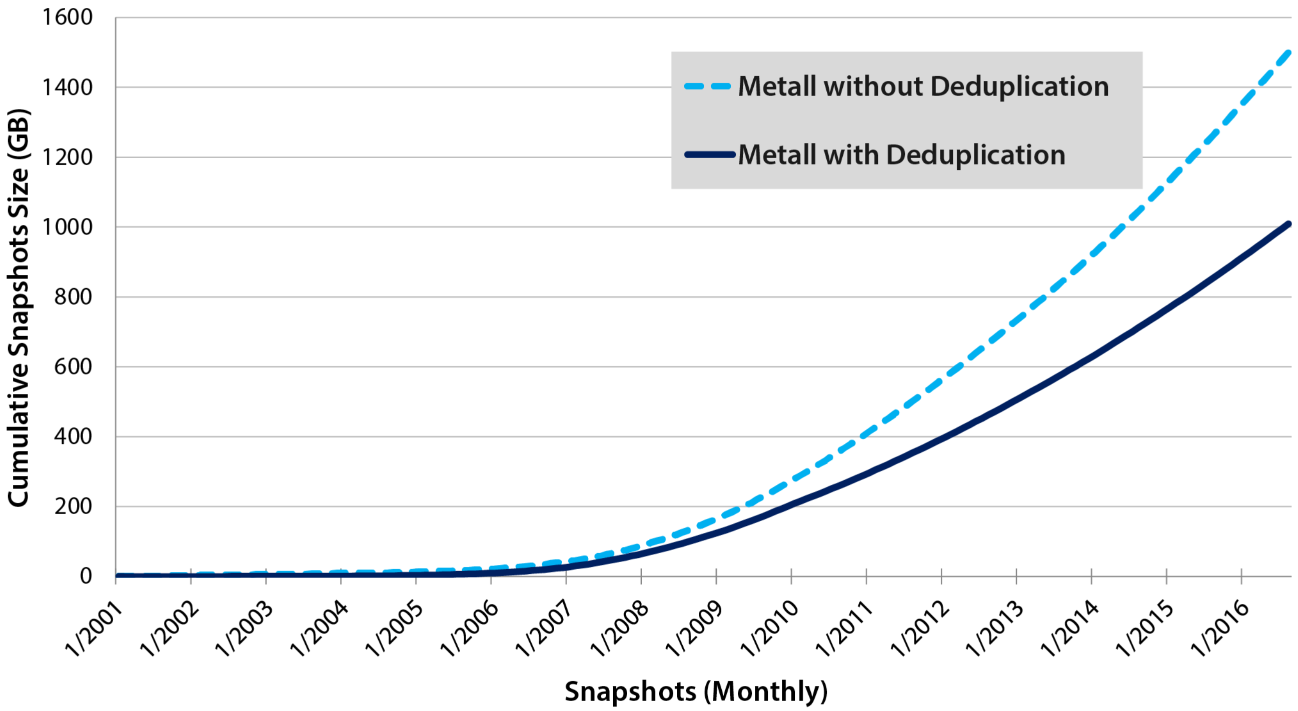


FIGURE 2. Cumulative storage size of monthly snapshots of the 16-year Wikipedia dynamic hyperlink graph; Metall's deduplication features save 33.2 percent storage [12].

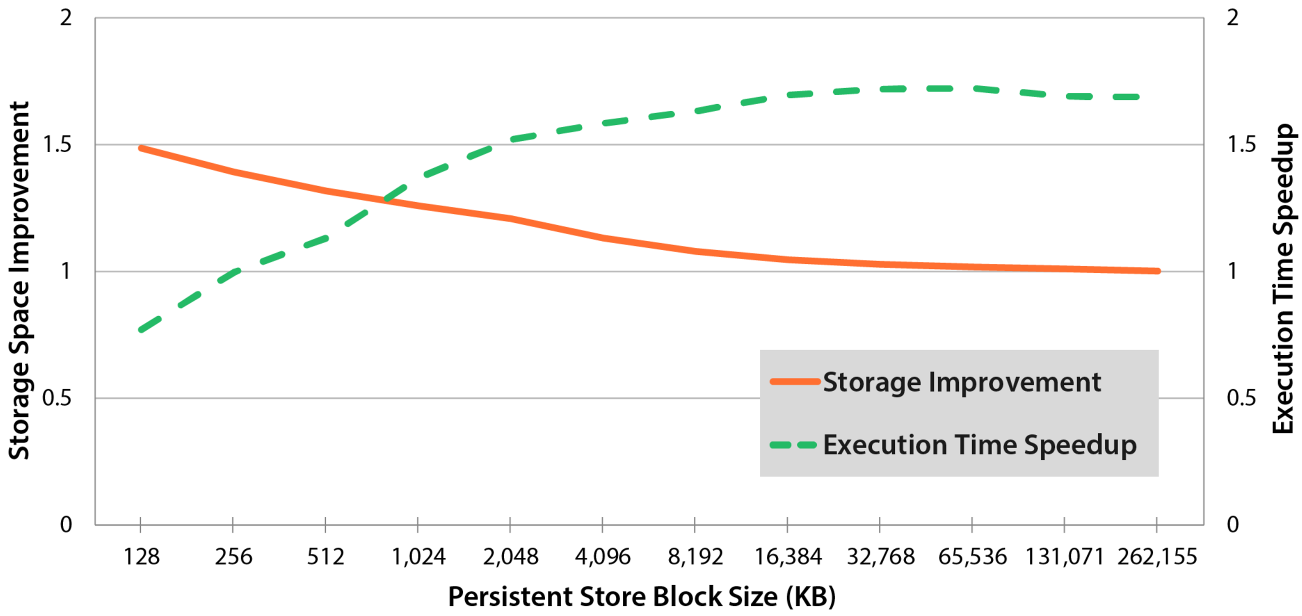


FIGURE 3. With Metall's deduplication, there is a trade-off between storage and time costs to store monthly snapshots of the 16-year Wikipedia dynamic hyperlink graph [12].

information about operating system processes and a second table represents network connections. The existence of edges, the direction of edges, and edge weights are commonly functions of the many fields in each table.

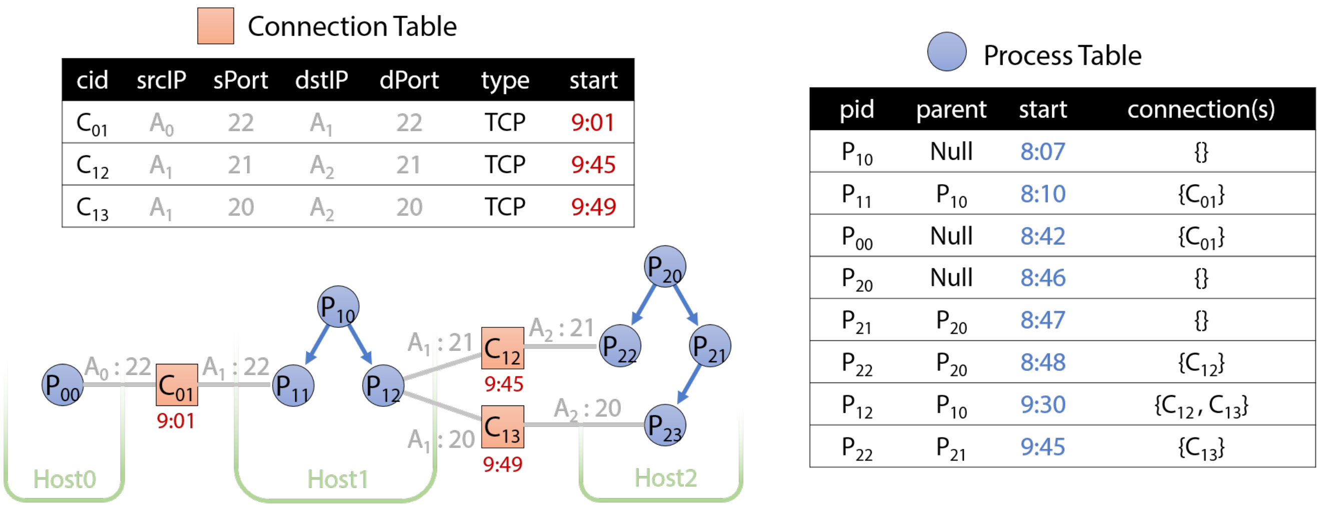
In most of today's large-scale graph frameworks where a complex graph algorithm can be efficiently executed, it is common to have the following cyclic-modeling analytic workflow:

1. Use a model to construct the graph topology,
2. Run a possibly expensive graph analytic,
3. Use the output to realize better modeling decisions could have been made,
4. Repeat many times (redefining the graph modeling and rerunning the analytic) until converging to a desirable analysis.

This process is often slow and can take weeks to resolve as cycles 1 through 3 may involve recomputing entire computational workflows, reading and writing entire passes of the massive data to disk, interpreting output and intermediate results, and communicating results or consulting with the domain scientists on how to improve the graph model. Utilizing persistent memory in the ways we have advocated facilitates keeping the metadata in place with the topology which, in turn, allows more reuse of intermediate results, greatly enhanced interactivity at scale, more potential for multiuser data interactivity, and paves the way for artificial-intelligence-based techniques that can refine such workflows automatically.

In our LLNL research projects, we have recently demonstrated the power of utilizing property graphs on massive relational data sets. Steil et al. prototyped an HPC capability that allows network scientists to survey the types of triangles (three-vertex cycles, where type can be defined by the user as a function of what metadata is involved) present within their relational data sets [14]. The framework is designed so that a network scientist can tailor a custom callback function that is invoked when a triangle is detected, reading vertex and edge metadata, and counting the specific type of triangle. Such surveys scale to a 128-billion-edge WWW hyperlink graph, with the triangle type being defined as unique combinations of three top-level domains (website URLs) involved in each triangle.

These property graph capabilities allow network scientists to more directly begin studying how metadata is important in the formation of higher-order network structures within their massive relational data sets. A common desire is to understand the importance of more general higher-order graph patterns (motifs) within massive data. Today, this is often done in a cumbersome cyclic-modeling analytic workflow, where vertex/edge labels are decided upon, a graph



🟧 Connection Table

cid	srcIP	sPort	dstIP	dPort	type	start
C_{01}	A_0	22	A_1	22	TCP	9:01
C_{12}	A_1	21	A_2	21	TCP	9:45
C_{13}	A_1	20	A_2	20	TCP	9:49

● Process Table

pid	parent	start	connection(s)
P_{10}	Null	8:07	{}
P_{11}	P_{10}	8:10	$\{C_{01}\}$
P_{00}	Null	8:42	$\{C_{01}\}$
P_{20}	Null	8:46	{}
P_{21}	P_{20}	8:47	{}
P_{22}	P_{20}	8:48	$\{C_{12}\}$
P_{12}	P_{10}	9:30	$\{C_{12}, C_{13}\}$
P_{22}	P_{21}	9:45	$\{C_{13}\}$

FIGURE 4. In this illustration, topological data and metadata are fused into a property graph inspired by DARPA's OpTC data set. The property graph represents a cybersecurity scenario in which Host0 uses Secure Shell (SSH) to look around on Host1, and Host2 uses File Transfer Protocol (FTP) to transfer large files to/from Host1.

pattern-matching algorithm (such as PruneJuice [15]) is applied, and various motifs are explored and compared with other metadata to gauge importance. Often this cycle needs to be repeated with different definitions of edges and vertex/edge labels to discover a meaningful pattern.

We envision persistent memory will be the substrate that enables topological and metadata analytics to coexist and enable future data scientists to interact and analyze property graphs at unprecedented scales. Persistent snapshots of complex data structures will accelerate the iterative and often cyclic nature of exploratory data analytics, and ultimately, improve data scientists' time to solution.

References

[1] Zaharia M, Xin RS, Wendell P, Das T, Armbrust M, Dave A, Meng X, Rosen J, et al. "Apache spark: A unified engine for big data processing." *Communications of the ACM.* 2016;59(11):56–65. Available at: http://dx.doi.org/10.1145/2934664.

[2] Merrill M, Reus W, Neumann T. "Arkouda: Interactive data exploration backed by chapel." In: *Proceedings of the ACM SIGPLAN 6th on Chapel Implementers and Users Workshop.* 2019 Jun. Available at: https://doi.org/10.1145/3329722.3330148.

[3] Trader T. "Livermore's El Capitan supercomputer to debut HPE 'Rabbit' near node local storage." *HPC Wire.* 2021 Feb 18. Available at: https://www.hpcwire.com/2021/02/18/livermores-el-capitan-supercomputer-hpe-rabbit-storage-nodes/.

[4] Folk M, Heber G, Koziol Q, Pourmal E, Robinson D. "An overview of the HDF5 technology suite and its applications." In: *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases.* 2011 Mar. pp. 36–47. Available at: https://doi.org/10.1145/1966895.1966900.

[5] Essen BV, Hsieh H, Ames S, Pearce R, Gokhale M. "DI-MMAP—A scalable memory-map runtime for out-of-core data-intensive applications." *Cluster Computing.* 2015;18(1):15–28.

[6] Iwabuchi K, Youssef K, Velusamy K, Gokhale M, Pearce R. "Metall: A persistent memory allocator for data-centric analytics." 2021. Cornell University Library, arXiv: 2108.07223.

[7] Iwabuchi K, Lebanoff L, Gokhale M, Pearce R. "Metall: A persistent memory allocator enabling graph processing." In: *2019 IEEE/ACM 9th Workshop on Irregular Applications: Architectures and Algorithms (IA3).* 2019 Nov 18. Pp. 39–44. Available at: https://doi.org/10.1109/IA349570.2019.00012.

[8] Github. LLNL/metall. Available at: https://github.com/llnl/metall.

[9] Pearce R, Gokhale M, Amato NM. "Multithreaded asynchronous graph traversal for in-memory and semi-external memory." In: *SC'10: Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis.* 2010 Nov 13–19. Available at: https://doi.org/10.1109/SC.2010.34.

[10] Pearce R, Gokhale M, Amato NM. "Scaling techniques for massive scale-free graphs in distributed (external) memory." In: *2013 IEEE 27th International Symposium on Parallel and Distributed Processing.* 2013 May 20–24. Available at: https://doi.org/10.1109/IPDPS.2013.72.

[11] Graph 500. Available at: http://graph500.org/.

[12] Youssef K, Iwabuchi K, Feng W, Pearce R. "Privateer: Multi-versioned memory-mapped data stores for high-performance data science." In: *2021 IEEE High Performance Extreme Computing Conference (HPEC).* 2021 Sep 20–24.

[13] Weir C, Arantes R, Hannon H, Kulseng M. (2021). "Operationally Transparent Cyber (OpTC)." *IEEE DataPort.* Datasets. Available at: https://ieee-dataport.org/open-access/operationally-transparent-cyber-optc.

[14] Steil T, Reza T, Iwabuchi K, Priest BW, Sanders G, Pearce R. "TriPoll: Computing surveys of triangles in massive-scale temporal graphs with metadata." In: *SC21: International Conference for High Performance Computing, Networking, Storage and Analysis.* 2021 Nov. No. 67. Pp. 1–12. Available at: https://doi.org/10.1145/3458817.3476200.

[15] Reza T, Ripeanu M, Tripoul N, Sanders G, Pearce R. "PruneJuice: Pruning trillion-edge graphs to a precise pattern-matching solution." In: *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis.* 2018 Nov 11–16. Pp. 265–281. Available at: https://doi.org/10.1109/SC.2018.00024.

Hardening the Hardware Supply Chain: Standardized Artifacts Enable Automated Accountability

Andrew Medak

Counterfeit, substandard, and malicious electronic components pose a risk to the US government as well as the broader US economy. Today, the Department of Homeland Security has identified 16 critical infrastructure sectors, “namely systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters” [1]. The Cybersecurity and Infrastructure Security Agency emphasizes that “the Information Technology Sector is central to the nation’s security, economy, and public health and safety as businesses, governments, academia, and private citizens are increasingly dependent upon Information Technology Sector functions” [2].

[Photo credit: iStock.com/Olivier Le Moal]

Return to CONTENTS

Most cybersecurity solutions begin protection after an individual computing device is up and running. There exists a security gap that relies on a belief that the configuration of the device is trustworthy. That gap is today covered using tamper-indicative security tape and handwritten signatures. There is no feasible way for a typical end-user to know whether an information technology (IT) device contains the same components that the manufacturer installed into the device. Building and maintaining local trust in the device hardware is critical to extending trust to anything running on top of that hardware. This article presents one application of standards to increase the integrity of the hardware supply chain and introduces a path forward to increasing the confidence in this approach.

A recent executive order has made it a priority to modernize the cybersecurity of the federal government by incorporating zero trust capabilities. "Zero Trust Architecture embeds comprehensive security monitoring; granular risk-based access controls; and system security automation in a coordinated manner throughout all aspects of the infrastructure in order to focus on protecting data in real-time within a dynamic threat environment" [3]. The National Institute of Standards and Technology (NIST) further expands that a zero touch "approach is primarily focused on data and service protection but can and should be expanded to include all enterprise assets" and that an aspect of the architecture requires the identification of assets owned [4]. This article will additionally document how assurance over the hardware supply chain within the zero trust architecture is achievable today.

When a computing device is procured, the bill of materials (BOM) will say what components are within the device. That sheet of paper could be used in a process that includes physically opening the device and manually identifying components against the BOM. The handwritten signature on that BOM provides some attribution that someone was confident in the configuration of the device. This process exposes the internal components of the device to an attack and would be time-prohibitive to conduct on every device. The NSA's Laboratory for Advanced Cybersecurity Research (LACR) has developed standards-based means to cryptographically validate the hardware configuration of a device and its components. These standards specify artifacts that enable automated and scalable validation of the hardware supply chain via an acceptance test that can be integrated into a procurement system of any size. The coverage of this acceptance test includes 100 percent of devices procured before they ever touch a trusted network. The validation process is being used to detect counterfeit devices and components, firmware alterations, and security properties of each device. The output of this acceptance test is a digital certificate that can be used to prove to the intended network that the device passed validation and is in a trustworthy state. This certificate also enables a local device identity on the network. This capability increases confidence in the integrity of the device from the production line to the desktop and is adaptable to authorized changes throughout the life cycle of that device.

Roots of trust

The goal is to establish trust in a new device by verifying components came from trusted sources and by verifying the endorsements made by those trusted sources against the current state of the device. Before anything can be verified, the device requires one component that can enable cryptographic functions within the device and act as a beachhead—on which the first notion of trust can be established.

One technique uses the Trusted Platform Module (TPM) [5], a cryptographic standard maintained by the Trusted Computing Group. Other cryptographic options could be used in place of the TPM. Those other solutions must be able to perform a similar set of functions in order to become that beachhead of trust in the device. The TPM can be implemented in a variety of ways to best fit the size and criticality of the device [6]. Ultimately, LACR recommends that a security expert review these options and select the best fit for the application. Physical and firmware TPMs are very common in the IT landscape today. Microsoft has required the inclusion of a TPM for new devices to be shipped with the Windows logo for desktop editions of Windows 10 since 2016 [7]. They took that a step further this year with Windows 11 by making the TPM 2.0 a requirement "in order to run Windows" [8]. It's important to remember that the purpose of any root of trust is to enable cryptographic functions within the device. Cybersecurity capabilities are built on top of that root of trust.

An aspect of building confidence in the root of trust is understanding how it is specified to work, by critically reviewing the code, and by gauging confidence in that solution within the cybersecurity community. The standards and code for the TPM are open to public review. There is an industry of cyber experts that are capable of critical review and are implementing TPM-based technology. And while some vulnerabilities have been found, they have either been fixed or require extraordinary access to a device. Part of this capability is meant to significantly raise the bar in terms of difficulty of executing an attack. At the same time, it can enable rapid mitigation once a threat is identified.

Beyond doing homework to understand the root of trust, the device owner also needs to be able to verify the authenticity of its manufacturer. The TPM fills this role through an Endorsement Key. This key can be created at any time during the life cycle of the TPM. The strongest assurance is possible when the key is created by the manufacturer of the TPM and is accompanied by an artifact that contains signed assertions about the key as well as the TPM. This Endorsement Key Certificate enables the device owner to verify the TPM was built by a particular manufacturer, that the TPM possesses the Endorsement Key described in the certificate, and that the key meets the desired algorithm and bit-length requirements. Those algorithm and bit-length requirements should again be reviewed by a security expert to meet the criticality of the device. The TPM manufacturer signs the Endorsement Key Certificate [9] using a certification path that includes a root certificate it publishes. The Endorsement Key is generally a restricted decryption key, and the TPM has functions that enable the device owner to challenge it to prove possession of the key. The ability for the TPM to prove it possesses a specific Endorsement Key endorsed by the exact TPM manufacturer expected by the device owner is critical to extending trust outside the cryptographic chip. This establishes a very strong binding to the root of trust [10].

Extending trust from the root

Extra time was spent discussing the provisioning of the root of trust because that first verification step is critical to the operation of everything that comes afterward. Additional verification will take place, and it all extends from this first step. If an error was found at this step, no additional verification should take place.

One of NIST's tenets of zero trust states, "The enterprise collects as much information as possible about the current state of assets, network infrastructure and communications and uses it to improve its security posture" [4]. The next step to securing the hardware supply chain is to extend trust from the root to artifacts that can be used to verify the current state of the hardware configuration of the device. This is where that digital hardware BOM comes in.

Earlier in this article, it was stated that one could open the device and manually identify components against a printed BOM. Information printed on the label of each component includes identifiers such as manufacturer names, model numbers, and serial numbers. Those identifiers can be compared against the printed list. That process would rely on the accuracy of the BOM as well as confidence in the information on the label of each component. These same identifiers are delivered digitally to the device by the firmware installed on those components. This information is used for the purpose of enabling management of each component by an operating system. Encoding the information into an artifact endorsed by the manufacturer who installs those components enables validation of a digital hardware bill of materials (HBOM) that is analogous to the validation of the Endorsement Key Certificate.

The Trusted Computing Group's Platform Certificate [11] is one option for a digital artifact that can be used as an HBOM. The Platform Certificate encapsulates hardware component identifiers that make up a device and is endorsed by a supply chain entity. In this context, a supply chain entity is anyone who adds a component to or removes a component from the device at any point in its supply chain. It could be the original equipment manufacturer, value-added resellers, enterprise IT services, or even an individual. Each entity creates a certification path that they use to endorse artifacts and delivers the root of that certification path separately from the device so that the device owner can select the supply chain entities they trust to have participated in the construction of their device. The Platform Certificate also encapsulates a link to the root of trust; in our case, a link to the Endorsement Key Certificate of a TPM. The trustworthiness of that link is dependent on trust in the supply chain entity as well as verification of the root of trust as discussed in the previous section. Figure 1 shows the relationship of the certificate with a device. The Platform Certificate can be delivered in multiple ways, including on the device or via a blockchain [12]. The certification path should be delivered separately from the Platform Certificate. For several years, some TPM manufacturers have posted their certification paths publicly on their websites [13, 14].

Component identification

Together, the Endorsement Key Certificate and the Platform Certificate introduce a framework to perform device identification with a component list that includes strong binding to a root of trust, accountability to manufacturers, and is entirely verifiable by software. A zero trust architecture can compare the real-time configuration of a device against these endorsed artifacts.

One challenge to this capability is ensuring the HBOM artifact encapsulates reliable component identifiers. The primary source of platform-independent digital hardware identifiers for a component is the component firmware. Any services that collect those identifiers for management of a component, including system firmware or an operating system, can change the formatting in a way that creates a dependency on that service. Understanding how those services report identifiers is essential prior to using them as a source of identifiers for inclusion into the Platform Certificate. Otherwise, identifiers collected on Windows may not match those collected on Linux despite both sets referencing the same physical component.

Therefore, it is important that standards are developed that define how to translate component identifiers from reliable sources for encoding into the endorsed artifact. Those same standards can be programmed into verification software to ensure this capability is scalable. To this end, the Trusted Computing Group is publishing standards to make its Platform Certificate interoperable. There are already two published standards for encoding identifiers: one for sourcing information from System Management Basic Input/Output System (SMBIOS) [15], and another to cover Peripheral Component Interconnect Express (PCIe) [16] components. Suggestions are welcome for additional protocols to target.

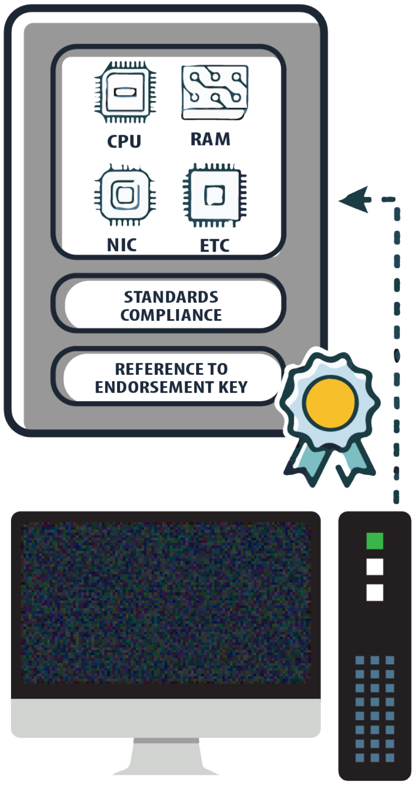


FIGURE 1. The platform certificate is an endorsed artifact that contains facts about the hardware configuration of a computing device. Trust is extended from the root of trust to the certificate.

Firmware verification

As inferred previously, component identification relies on accurate component identifiers coming from component firmware. This is dependent upon information that could be altered by firmware updates. To increase confidence in these identifiers seen by the device, tracking of firmware configuration is required—in a similar way to the hardware configuration.

The Trusted Computing Group describes the attestation of firmware integrity measurements, “When a Platform with a TPM boots, executable components may perform integrity measurements of other components and extend these in the TPM’s Platform Configuration Registers (PCRs) before passing execution control to the newly measured components. Changes to the values in the PCRs would then indicate changes to the measured components” [10]. To accomplish firmware verification, each supply chain entity captures changes in firmware measurements into an endorsed artifact called a Reference Integrity Manifest (RIM) [17]. Then at each subsequent point in the supply chain, the real-time state of the firmware can be verified against those measurements endorsed into the RIM.

Component attestation

Component identification and firmware validation can easily detect mistakes during the production of a device and provide accountability to the supply chain entity responsible. An increased level of sophistication is required, but there still exists a chance for an attacker to swap hardware components and alter firmware to report the identifiers expected in the platform certificate. Without additional validation methods, a swapped component with undefined traits could be carried by a device, and it would pass the supply chain acceptance test.

The hardware supply chain claims covered by the standards outlined in this article will be enhanced with the introduction of technologies that enable components to carry their own roots of trust. Standards like Device Identifier Composition Engine (DICE) [18], Security Protocol and Data Model (SPDM) [19], Enhanced Privacy ID (EPID) [20] enable component manufacturers to deliver endorsed identity certificates and keys, even on resource-constrained components. Referencing those identities within the Platform Certificate will enable an ability to identify all components of a device as well as establish a strong cryptographic binding to them via a nonce challenge.

Component attestation will significantly increase the time and access required to perform a hardware supply chain attack. At that point, confidence in the security of this acceptance test is dependent on trust in the endorsing supply chain entities and the security analysis of the standards that extend from those roots of trust. All the standards used in this method are open. Constant public scrutiny and security analysis is required to maintain trust in these standards.

Use cases

Supply chain acceptance test

Supply Chain validation of every new device is the primary use case that we've focused on for this technology. The acceptance test described in this article elevates supply chain assurance from checking physical tape seals to Top Secret-grade cryptography [21]. It should be used to verify the components of a device are authentic before allowing the device to connect to a trusted network. Figure 2 illustrates the acceptance test. From the left, one truck is carrying devices that possess TPM-based artifacts and that are accepted at a facility. From the right, a different truck is carrying devices to the facility that do not pass the acceptance test.

The acceptance test is extensible in that it allows for multiple supply chain entities to participate—including the device owner—as long as they are authorized to change the hardware configuration of a device. There could be more than one Platform Certificate and/or more than one RIM involved in the validation of a single device. For example, an original equipment manufacturer may only add a processor to a TPM-enabled motherboard. That entity would endorse a base Platform Certificate that included component identifiers for those two components. There might not even be enough of a device yet to create a RIM. Another entity may add additional components to the device and that entity would endorse a Delta Platform Certificate with those component identifiers. Each supply chain entity must include any hardware or firmware changes made to the device for it to pass the acceptance test. Ultimately, the device owner is empowered to select the entities they trust to have participated in the supply chain of the device.

The device owner could also be a supply chain entity. They may wish to add memory or remove a hard disk. In these cases, they would need to create and protect a local certification path to endorse their own Platform Certificates or RIMs as necessary to track changes to the device.

Local device identity

One output of the acceptance test is a certificate signed by a local certificate authority confirming that the device passed validation. Specifically, that the root of trust within the device possessed a particular key endorsed by the manufacturer of the root of trust and that all of the supply chain checks built on top of that key also passed. This local Certificate Authority is controlled by the device or network owner. This certificate

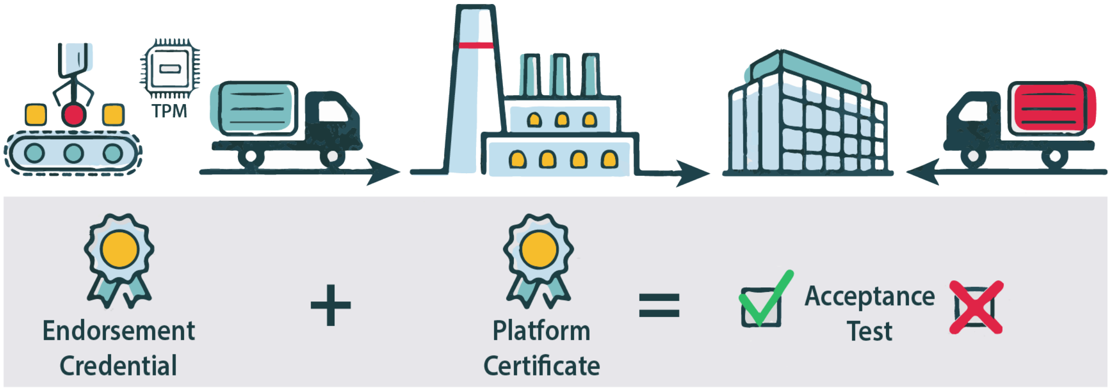


FIGURE 2. The supply chain acceptance test ensures only devices that have the appropriate artifacts (illustrated on the left) and can pass validation will continue through the supply chain; whereas, those without the appropriate artifacts (illustrated on the right) fail the test and are not accepted.

binds the device to its root of trust and also to the local network.

Additional keys and certificates can be issued during the same provisioning procedure that enable device identity. The Institute of Electrical and Electronics Engineers (IEEE) explains that "A Secure Device Identifier (DevID) is cryptographically bound to a device and supports authentication of the device's identity. An Initial Device Identifier (IDevID) provided by the supplier of a device can be supplemented by Local Device Identifiers (LDevIDs) facilitating enrollment (provisioning of authentication and authorization credentials) by local network administrators" [22].

An IDevID Certificate is proof that the endorser felt comfortable creating and certifying a key that can be used as an identity for the life of the device. An LDevID Certificate declares that all the artifacts along the entire supply chain were verified against the current state of the device—if it is signed by a local Certificate Authority that has appropriate policies in place and is trusted by the network owner.

Part of migrating to a zero trust architecture is the need to identify assets owned by the enterprise [4]. The Platform Certificate and the LDevID provide a strong cryptographically verifiable identification of an asset and the components that make up that asset.

Opinions differ on whether all applications require a thorough supply chain check before creating the device identity. It could save the time necessary to run the acceptance test by using an IDevID pre-built into the device. But I think local verification of the artifacts should be required in all situations. No matter if it's a Top Secret network or a personal network at home. Nobody who cares about the security of their network installs a new computer knowing it will cause a vulnerability. It is possible to design this supply chain risk mitigation to make the provisioning procedure as simple as plugging in a device and starting it. This device identity has multiple applications from within low-level management services all the way up to user applications. The confidence and scalability of the acceptance test is too strong to pass on.

For example, NSA recommends Port Security on all Network Devices. "Port security 802.1x device authentication should be enabled to force clients to authenticate before they are allowed onto the network" [23]. This strategy introduces public key infrastructure (PKI) trust mechanisms to only allow devices that possess signed certificates the enterprise trusts with access to the network. The LDevID created during the provisioning process of a device can be used in conjunction with the IEEE 802.1x authentication protocol to allow a device to cryptographically prove its identity to network devices to attain network transport access. If the LDevID key is stored within a TPM, verification of that root of trust ensures a strong cryptographic binding between the identity and the device [24]. No matter the size or purpose of the network, the zero trust architecture will greatly benefit from only creating local device identities for devices that have passed hardware supply chain validation.

Asset management and security monitoring

Security monitoring does not end with the acceptance test. “The enterprise monitors and measures the integrity and security posture of all owned and associated assets. No asset is inherently trusted” [4]. These supply chain artifacts are intended to represent the hardware configuration of a device from its time on a production line through the end of its life. The acceptance test is intended to be scalable and run remotely. It enables continuous monitoring of 100 percent of the devices on a network. Once a device is added to a trusted network, the network owner may set a policy to verify the hardware configuration of that device as often as they wish. All of this requires strong trust in the device as early in the supply chain as possible.

Recovery from a supply chain attack

Verifying the integrity of the supply chain increases confidence there were no unauthorized components added or removed from the device. If an attacker were to be successful in hiding malicious hardware, or if they gained access to the private keys of the certification path of a supply chain entity, then every entity after that point in the supply chain would have to contend with configuration changes endorsed by a potentially trusted entity. There could be a situation where a network owner needs an ability to recover from the supply chain attack. That recovery will involve locating all devices on the network that contain compromised components.

As devices go through the acceptance test, their component information is stored into an asset management system within the zero trust architecture. If a faulty or malicious component is discovered, the network owner has an ability to identify and locate all the devices on the network that have that same component. Additionally, supply chain artifacts for those devices will enable attribution to find where in the supply chain compromised components were installed. Recovery time from a supply chain attack is greatly improved since the acceptance test is meant to provision every device.

Open-source applications

LACR has released open-source projects on GitHub. These projects are used to demonstrate our provisioning protocol and to ease the burden on manufacturers as they build infrastructure.

The Host Integrity at Runtime and Startup (HIRS) Attestation Certificate Authority [25] consists of a client and server application. It performs the full provisioning protocol on a TPM-enabled device. This includes:

- Verifying the Endorsement Key Certificate was signed by a trusted TPM manufacturer,
- Ensuring the Endorsement Key meets our security requirements,
- Performing a nonce check to ensure the TPM contains the specific Endorsement Key endorsed by the TPM manufacturer,
- Verifying the Platform Certificate(s) were signed by trusted supply chain entities,
- Comparing hardware component information within the Platform Certificate(s) against the current state of the device and TPM,
- Verifying the RIM(s) were signed by trusted supply chain entities,
- Comparing measurements within the RIM(s) against the TPM quote and event log, and
- Creation and certification of keys within the TPM that enable proof of supply chain validation on the local network as well as additional uses for those keys in post-provisioning applications.

The Platform Attribute Certificate Creator (PACCOR) [26] is a tool for creating platform certificates. The main goals of the project are:

- To ease the burden on entities who wish to create platform attribution certificates according to the Trusted Computing Group specification,
- To demonstrate how details from a platform can be gathered in a quick, automated manner, and
- To demonstrate that component identification can be accomplished regardless of Operating System.

PACCOR works on Windows and Linux to

- Gather all component identifiers automatically,
- Gather the Endorsement Key Certificate from the TPM,
- Generate and sign a platform certificate, and
- Validate the signature on an attribute certificate.

The tcg_rim_tool [27] can be used to create NISTIR 8060 compatible Software Identification (SWID) tags that adhere to the Trusted Computing Group PC Client RIM specification. It also supports the ability to digitally sign the Base RIM file as the HIRS Attestation Certificate Authority (ACA) will require a valid signature in order to upload any RIM file.

The PC Client RIM Specification utilizes the TPM Event Log as a Support RIM type. It was useful to have a tool for inspecting the contents of the TPM event log. The tcg_eventlog_tool [28] parses the binary Event Log, prints events as human readable output, and provides hexadecimal events which can be used as test patterns. It can also be used to compare event logs to find details on what events caused mismatches.

Sample scenario

To help illustrate the acceptance test, here's an outline of what happens when a shipment of devices arrives at a customer site. Keep in mind that the test scales to any number of devices within the shipment. The customer site could be anywhere from a home to a warehouse.

Each device in the shipment is pulled aside and

- Assigned a local barcode or name.
- Connected to a server hosting the acceptance test.
- Scanned for artifacts to be verified by the acceptance test,
    - This could include a Platform Certificate, RIM, and component attestation identities.
- If the acceptance test is HIRS, performs the full provisioning protocol outlined earlier.
- If the device is successfully validated using the acceptance test,
    - It is given one or more certified local device identities from the ACA and
    - Enrolled in network asset management databases and monitoring services.

Throughout its life cycle, the device could be modified by the customer. This could be a hardware or firmware change to any of the components within the device.

In that case, the customer would

- Create additional artifacts as necessary to document the modification,
- Sign those artifacts using a local CA to endorse the modification, and
- Include those new artifacts so that testing the device will see that modification was accepted.

Results and commercial adoption

Our open-source applications have been shown to work in demonstration and piloting capacities of increasing scale. The latest pilot included 96 desktop and laptop computers from two manufacturers. All 96 were tested to prove that the hardware configuration of each device was consistent from the production line of the original equipment manufacturer to the loading dock of the customer. This capability also works on any network device that can support a TPM, including servers and routers. Additional pilots helped to show interest in this technology, increase the capability of the software, and prove its readiness to scale even more.

Major suppliers of computing equipment to the Department of Defense (DoD) are investing in this capability. And these manufacturers have been making progress on implementation in both research and production despite the pandemic. Hewlett Packard Enterprise (HPE) is including Platform Certificates and IDevIDs on their server products as of June 2021 [29]. Dell began offering Platform Certificates on their server products at the end of 2020 [30]. Intel expanded their Transparent Supply Chain service last year to include servers as well as some of their client devices [31]. Any vendors that supply computing equipment to the DoD should feel encouraged to begin implementing the generation of these supply chain artifacts into their production infrastructure.

Conclusion

Zero trust is an attractive catch phrase that is defined in many ways today. The name on its own can give a false sense that trust is being removed from the security architecture. As seen throughout this article, zero trust relies on the continuous re-evaluation in confidence that only authorized code, hardware, resources, and people are included in the architecture. Trust is assigned to entities via the digital signatures they use to endorse their products. People still need to continuously vet all cryptographic solutions and standards, as well as the supply chain entities, that are included in their architecture. I like the simplicity of this definition

IBM is using in some of their marketing, "A zero trust approach aims to wrap security around every user, every device, every connection—every time" [32].

Hardware validation is a difficult problem because additional capabilities can lay dormant until activated—sometimes after all verification checks have been passed. One thing we can do is verify these hardware components were manufactured and installed by expected entities. The standardized artifacts highlighted in this article assign accountability to supply chain entities for those components. They enable a path to identify and verify the trustworthiness of devices remotely on a network. They also provide a means to recover from a hardware-based attack and locate components on the network if they are found to be corrupted.

With increased strong integrity capability, commercial adoption, and ongoing research of enhancements, this capability is ready for prime time. The acceptance test provides critical evaluation of the integrity of a device, can be performed on every device procured with supply chain artifacts, and can be performed prior to placing any device on a trusted network.

The US DoD has included the TPM as a requirement on new computer assets procured by DoD components since 2014. Inclusion of a TPM on an asset is only half the answer. The other half is to implement cybersecurity capabilities that may rely on the TPM as a root of trust. In fact, NSA is tasked to "identify use cases and implementation standards and plans for DoD to leverage TPM functionality fully to enhance IT device security, including platform integrity verification, platform identification and authentication, and enhanced encryption" [33]. NIST is hosting a Supply Chain Assurance project that will produce SP 1800-34, a Cybersecurity Practice Guide to Validating the Integrity of Computing Devices [34]. That series of documents should help NSA with its task.

This capability has reached a stage where vendors have built infrastructure within their factories to make this a reality. Continuing to pilot and scale up shows vendors that there is interest in this ability for the owner of a device to perform this kind of verification. It is available to large enterprises and personal home networks through open-source standards and applications. Because the capability is being built on open standards, everyone has an opportunity to learn about how it works, its security implications, and start to use it.

Once NIST and NSA publish their recommendations for this technology, requirements on the validation of the hardware supply chain via the acceptance test could benefit the new DoD Cloud contract [35] to enable trustworthiness in the hardware on which those platforms will run.

Additionally, the zero trust architecture that is being developed to improve the nation's cybersecurity [3] could be expanded to include the hardware supply chain. Many of the requirements for enhancing software supply chain security can be directly applied to hardware as well. In particular, providing consumers a HBOM for each device directly enables the employment of automated tools to maintain trusted hardware supply chains, thereby ensuring the integrity of every device.

References

[1] The White House, Office of the Press Secretary. "Presidential Policy Directive 21: Critical Infrastructure Security and Resilience (PPD-21)." 2013 Feb 12. Available at: https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil.

[2] Cybersecurity and Infrastructure Security Agency (CISA). "Information Technology Sector." Available at: https://www.cisa.gov/information-technology-sector. [Accessed 2021 Jul 12.]

[3] The White House. "EO 14028: Improving the Nation's Cybersecurity." *Federal Register.* 2021 May 12;86(93). Available: https://www.govinfo.gov/content/pkg/FR-2021-05-17/pdf/2021-10460.pdf. [Accessed 2021 Aug 2.]

[4] Rose S, Borchert O, Mitchell S, Connelly S. "NIST SP 800-207: Zero trust architecture." 2020. Available at: https://csrc.nist.gov/publications/detail/sp/800-207/final.

[5] Trusted Computing Group. "Trusted Platform Module Library Specification, Family "2.0," Level 00, Revision 01.59." 2019 Nov. Available at: https://trustedcomputinggroup.org/resource/tpm-library-specification/. [Accessed Jul 2021.]

[6] Trusted Computing Group. "TPM 2.0 A Brief Introduction." 2019 Jun. Available at: https://trustedcomputinggroup.org/wp-content/uploads/2019_TCG_TPM2_BriefOverview_DR02web.pdf.

[7] Microsoft. "TPM 2.0 Compliance for Windows 10." 2018 Nov 29. Available at: https://docs.microsoft.com/en-us/windows/security/information-protection/tpm/tpm-recommendations#tpm-20-compliance-for-windows-10. [Accessed 2021 Jul 29.]

[8] Microsoft. "Compatibility for Windows 11." 2021 Jun 24. Available at: https://docs.microsoft.com/en-us/windows/compatibility/windows-11/. [Accessed 09 Jul 2021.]

[9] Trusted Computing Group. "TCG EK Credential Profile For TPM Family 2.0; Level 0 Version 2.4 Revision 3." 2021 Jul 16. Available at: https://trustedcomputinggroup.org/wp-content/uploads/TCG_IWG_EKCredentialProfile_v2p4_r3.pdf.

[10] Trusted Computing Group. "TCG TPM v2.0 Provisioning Guidance Version 1.0 Revision 1.0." 2017 Mar 15. Available at: https://trustedcomputinggroup.org/wp-content/uploads/TCG-TPM-v2.0-Provisioning-Guidance-Published-v1r1.pdf.

[11] Trusted Computing Group. "TCG Platform Certificate Profile Version 1.1 Revision 19." 2020 Apr 11. Available at: https://trustedcomputinggroup.org/wp-content/uploads/IWG_Platform_Certificate_Profile_v1p1_r19_pub_fixed.pdf.

[12] Cabre E, Dodson T. "Secure your business: End-to-end supply chain traceability" [Intel white paper]. 2019. Available at: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/transparent-supply-chain-ethereum-blockchain-white-paper.pdf. [Accessed 2021 Aug 22.]

[13] Infineon. "OPTIGA TPM certificates." Available at: https://www.infineon.com/cms/en/product/promopages/optiga_tpm_certificates/. [Accessed 2021 Aug 29.]

[14] STMicroelectronics. "TN1330 STM TPM EK certificates." 2020 Jun. Available at: https://www.st.com/resource/en/technical_note/tn1330-st-trusted-platform-module-tpm-endorsement-key-ek-certificates-stmicroelectronics.pdf. [Accessed 2021 Aug 29.]

[15] Trusted Computing Group. "SMBIOS-based Component Version 1 Revision 01." 2021 Feb 18. Available at: https://trustedcomputinggroup.org/wp-content/uploads/SMBIOS-Component-Class-Registry_v1.01_finalpublication.pdf.

[16] Trusted Computing Group. "PCIe-based Component Class Version 1 Revision 18." 2021 Oct 27. Available at: https://trustedcomputinggroup.org/wp-content/uploads/TCG_PCIe_Component_Class_Registry_v1_r18_pub10272021.pdf.

[17] Trusted Computing Group. "TCG Reference Integrity Manifest (RIM) Information Model Version 1.01 Revision 0.16." 2020 Nov 12. Available at: https://trustedcomputinggroup.org/wp-content/uploads/TCG_RIM_Model_v1p01_r0p16_pub.pdf.

[18] Trusted Computing Group. "Dice Attestation Architecture r23." 2021 Mar 1. Available at: https://trustedcomputinggroup.org/wp-content/uploads/DICE-Attestation-Architecture-r23-final.pdf.

[19] Distributed Management Task Force (DMTF). "Security Protocol and Data Model (SPDM)." 2021 May 24. Available at: https://www.dmtf.org/sites/default/files/standards/documents/DSP0274_1.1.1.pdf.

[20] Intel. "A Cost-Effective Foundation for End-to-End IoT Security" [Intel white paper]. 2016. Available at: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/intel-epid-white-paper.pdf.

[21] National Security Agency (NSA). "Commercial National Security Algorithm Suite." Available at: https://apps.nsa.gov/iaarchive/programs/iad-initiatives/cnsa-suite.cfm. [Accessed 2021 Jul 29.]

[22] Institute of Electrical and Electronics Engineers (IEEE). "IEEE Standard for Local and Metropolitan Area Networks—Secure Device Identity." 2018 Jun 14. Available at: https://standards.ieee.org/standard/802_1AR-2018.html.

[23] National Security Agency (NSA). "Deploying Secure Unified Communications/Voice and Video over IP Systems." 2021 June. Available at: https://media.defense.gov/2021/Jun/17/2002744054/-1/-1/1/CTR_DEPLOYING%20SECURE%20VVOIP%20SYSTEMS.PDF.

[24] Trusted Computing Group. "TPM 2.0 Keys for Device Identity and Attestation | Trusted Computing Group." Available at: https://trustedcomputinggroup.org/resource/tpm-2-0-keys-for-device-identity-and-attestation/. [Version 1r12, accessed 2021 Oct 8.]

[25] NSACyber. "GitHub - nsacyber/HIRS: Trusted Computing based services supporting TPM provisioning and supply chain validation concepts. #nsacyber." Available at: https://www.github.com/nsacyber/hirs. [Accessed Jul 2021.]

[26] NSACyber. "GitHub - nsacyber/paccor: The Platform Attribute Certificate Creator can gather component details, create, sign, and validate the TCG-defined Platform Credential. #nsacyber." Available at: https://www.github.com/nsacyber/paccor. [Accessed Jul 2021.]

[27] NSACyber. "HIRS/tools/tcg_rim_tool at master · nsacyber/HIRS · GitHub." Available at: https://github.com/nsacyber/HIRS/tree/master/tools/tcg_rim_tool. [Accessed Jul 2021.]

[28] NSACyber. "HIRS/tools/tcg_eventlog_tool at master · nsacyber/HIRS · GitHub." Available at: https://github.com/nsacyber/HIRS/tree/master/tools/tcg_eventlog_tool. [Accessed Jul 2021.]

[29] Hewlett Packard Enterprise (HPE). "Device Identity and Component Attestation comes to HPE Gen10 Plus servers." 2021 Jun 22. Available at: https://www.hpe.com/us/en/pdfViewer.html?docId=a00114960. [Accessed 2021 Jul 26.]

[30] Dell. "Dell Technologies Secured Component Verification." Available at: https://www.delltechnologies.com/en-us/solutions/openmanage/secure-component-authentication.htm. [Accessed 2021 Jul 20.]

[31] Intel. "Intel Transparent Supply Chain." Available at: https://tsc.intel.com/. [Accessed 2021 Jun 29.]

[32] IBM. "Zero Trust Security Solutions | IBM." Available at: https://www.ibm.com/security/zero-trust. [Accessed 2021 Jul 30.]

[33] US Department of Defense. "Department of Defense INSTRUCTION Cybersecurity." 2014 Mar 14. Available at: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodi/850001_2014.pdf.

[34] National Institute of Standards and Technology (NIST) National Cybersecurity Center of Excellence (NCCoE). "Supply Chain Assurance." Available at: https://www.nccoe.nist.gov/projects/building-blocks/supply-chain-assurance. [Accessed 2021 Aug 29.]

[35] US Department of Defense. "Future of the Joint Enterprise Defense Infrastructure Cloud Contract." 2021 Jun 7. Available at: https://www.defense.gov/Newsroom/Releases/Release/Article/2682992/future-of-the-joint-enterprise-defense-infrastructure-cloud-contract/. [Accessed 2021 Jul 26.]

SCAN ME

National Security Agency Central Security Service

Defending Our Nation. Securing The Future.